A Symphony of Redistributive Instruments

ISBN: 978 90 361 0378 7

Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **572** of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

A Symphony of Redistributive Instruments

Een symfonie van herverdelingsinstrumenten

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op

vrijdag 21 februari 2014 om 13.30 uur

door

FLORIS TOBIAS ZOUTMAN
geboren te Apeldoorn.

ERASMUS UNIVERSITEIT ROTTERDAM

Promotiecommissie:

Promotor:	Prof.dr. B. Jacobs
Overige leden:	Prof.dr. R. Boadway
	Prof.dr. L.M. Jacquet
	Prof.dr. B. Visser

Preface

This dissertation is the result of almost five years of very hard work. Doing my coursework and writing my dissertation has changed me fundamentally. I find that I have gained a much deeper understanding of economics in general, and public finance specifically. In the beginning I could not come up with any research ideas but now my mind is constantly engaged in coming up with new research ideas and with methods to execute them. I, like many of my colleagues, find that it is impossible to separate work from private life when you are an academic. At any moment in time a new and brilliant research idea may pop into your mind, and it is obviously impossible to relax until that new research idea has been solidified in a finished paper. Of course, there is no such thing as a finished paper, since there are always small imperfections in any paper no matter how much you polish it. The back of my mind, if not also the fronts and sides, is therefore now always engaged in research. Even my dreams are often related to economic issues.[1] In interactions with non-economists I often have the feeling that I need to flip a switch. Much of the economics lingo that I use every day is entirely indiscernible to people that have received no training in economics. For example, even when I explicitly state that I am a hyperbolic discounter, my non-economist friends will still buy me that extra beer that I know I will regret the next morning.

When I stepped into the office of Bas Jacobs for my job interview I had no idea that I was on my way to becoming an entirely different person. In fact, I had no idea who Bas Jacobs was. All I knew was that Ruud de Mooij would be my second adviser. I had met Ruud De Mooij during a student conference the year before my application. He presented some work about optimal taxation based on research done by Nobel prize winners James Mirrlees and Peter Diamond and John Bates Clark medalist Emmanuel Saez. At that moment I could not really understand why the optimal tax schedule looked the way it did, but I was mesmerized by the magic that Ruud presented nevertheless.

[1] I remember vividly how Netherlands lost its triple A status in one of my dreams, and how in my dream this turned out not to be the worst thing that could happen.

Now, five years later I feel I can call myself an expert in public finance. Unfortunately, Ruud left the Erasmus University to work for the IMF a year after I joined the Erasmus University, but I quickly found out that Bas knew just as much about optimal taxation as Ruud, and his enthusiasm for the subject was contagious. Although I am very broadly interested in economics and could have probably written a dissertation in another field altogether, I am very happy that Bas pushed me into public finance. The main theme of my dissertation is redistribution between rich and poor and I think this is one of the most exciting, important and policy-relevant topics in economics.

I know with absolute certainty that doing a PhD was the correct choice for me. Despite all the hard work and frustration that comes with it, at heart, I feel, I am a true academic. My true passion lies in identifying deep questions that shape our economy, and answering them using scientific methods. Although, I know all about the frustration that comes with working out a research question, I cannot help but come up with new ones and answering them using the best tools available to me, such that hopefully, possibly in the far future, my answers disseminate to choices made by policy makers. I find it is impossible to fully convey this passion to people that have not been hit by it, but I am very happy to be part of the small scientific community in which this passion is highly valued. In retrospect it would have been very hard, if not impossible for me to develop my full potential anywhere else. As an economist I have turned my passion into my occupation, and I enjoy going to work almost every day.

This dissertation is not just the product of my hard work. I have been in the fortunate circumstance to have received help from a very large group of colleagues, family and friends. I want to take this opportunity to thank them. The listing below will undoubtedly be incomplete, and I apologize in advance for the people I have forgotten to mention.

First, and foremost I want to thank my girlfriend Eva. We have only known each other for a year and a half, but I do not see how I could have written this dissertation without her. Her complete confidence in me, kept me going during that most stressful period right before the dissertation deadline. She gave me courage during my search for a job and she stood by me when I was in the hospital for surgery. Most of all, she has given more joy to my life than I could have imagined before I met her.

I also would like to thank my adviser Bas. Bas is both a great academic and a true perfectionist. He taught me everything I needed to know to become an economist and he always challenged me to get the best out of myself. I can truly say I learned more about economics during the coffee breaks with Bas, than I learned through all the courses I took in undergrad and graduate school. He has also introduced me to the academic society by pushing me to go to present my work whenever I could. I learned that it is vital for any young economist to visit seminars and conferences, both to present your work and to hear others present their work. In addition, his input as a co-author in three chapters of this dissertation has been of vital importance. Finally, I am happy to call Bas my friend, and I fondly remember all the dinners and drinks we had, where we could unwind from our hard work.

Aart Gerritsen and Hendrik Vrijburg are my paranymphs today. During most of my time at the Erasmus University Aart was also my office mate. Aart and I met during our Research Master in Groningen and in the classes we solved problem sets together. In Rotterdam we started our PhD in the same year and with the same adviser. Aart has the amazing ability to immediately see through to the core of any argument. Whenever you want to know whether a new research idea has any merit, you can either go through the lengthy process of working it out and presenting it to a large audience and seeing if they agree with it, or you can simply discuss your idea with Aart and see if he likes it. If the idea is wrong or your logic is flawed, Aart will detect it right away. Any idea that does not pass the Aart-test does not have to be executed. I have also often benefited a lot from Hendrik's vast knowledge of public finance, as well as his creative ideas and his Frysian clearheadedness, during our (research) meetings.

Egbert Jongen and Sander Renes have each been a co-author to two of the chapters in this thesis. It is truly a pleasure to work with them. Their knowledge, as well as their creativity have been indispensable in these chapters. In addition, Egbert has introduced me to the wondrous world of CPB Netherlands Bureau for Economic Policy Analysis, where many great economists somehow manage to translate the abstract language of academic economics into policy advice aimed at politicians and policy makers. The spirit of cooperation within the organization has been an inspiration to me. Sander is a true economic omnivore. He knows something about each and every topic in economics, and he can seemingly effortlessly combine concepts between different fields of economics. The discussions I have had with him about each and every chapter of this thesis have been invaluable in the completion of this dissertation.

During the fall semester of 2012 I have visited UC Berkeley on invitation of Emmanuel Saez. Emmanuel is the true superstar in the field of public finance. He has written seminal articles in almost all subfields. His articles are cited multiple times in all of the chapters in this thesis. My visit to Berkeley enabled me to observe one of the best research institutes in economics from the inside, and I have learned a lot about what it takes to execute a research idea using state of art methodology from the staff and students in Berkeley. Even more important, during my first class at Berkeley I met Eva.

My time as a PhD-candidate at the Erasmus University would have been much more bleak without the support of my colleague students. I fondly remember all the downtime we enjoyed, having drinks, dinners and all sorts of other activities. I would like to thank especially (in no particular order) Kyle Moore, Olivier Herlem, Rei Sayag, Barbara Sadaba, Saskia ter Ellen, Ruben de Bliek, Sun Pengfei, Tommi Tervonen, Wim Riedijk, Paul Steffens, Joeri Sol, Alexei Parakhonyak, Andrei Dubovik, Andrej Mondom, Frederik Hogenboom, Alexander Hogenboom, Oke Onemu, Eran Raviv, Heiner Schmittdiel, and Lerby Ergun.

In addition, I would like to thank all the (academic) staff at the Erasmus University as well as the various colleagues I have met during international seminars, conferences and workshop for their help on several occasions, especially, (again in no particular order), Robin Boadway, Laurence Jacquet, Felix Bierbrauer, Bauke Visser, Otto Swank, Lorenzo Pozzi, Robert Dur, Jurjen Kamphorst, Dana Sisak, Dominik Sachs, Dirk Schindler, Guttorm Schjelderup, Ruud de Mooij, Katherine Cuff, Andreas Peichl, Olivier Bargain, Rick van der Ploeg, Casper de Vries, Etienne Lehmann, Danny Yagan and Coen Teulings.

Last, but not least, I would like to thank my parents who have always stood by me and gave me moral and financial support during my studies.

Contents

List of Tables

List of Figures

Chapter 2

Chapter 1

Introduction

One of the most important trends in developed countries is the rise of economic inequality. In the US the top 1 percent currently earns a little more than 22 percent of all income.[1] In the period 1993-2012 real income of the top income earners has increased by 86 percent, whereas the bottom 99 percent saw an increase of only 6.6 percent. To put these numbers in some perspective, the income loss associated with the increase in inequality is larger than the income loss associated with the Great Recession for the bottom 99 percent. Although no other country has seen inequality rise as much as in the US, inequality has risen in almost all OECD countries (see e.g. Atkinson *et al.*, 2011).

The main causes behind the trend of rising inequality are technological and financial innovation, and globalization. The IT and internet revolution has brought large productivity gains for highly skilled workers, while the productivity of low-skilled workers has remained roughly constant. Financial innovation created new lines of credit for entrepreneurs and new insurance opportunities for firms, but the spoils have not trickled down to the poor in our society. Globalization has lead to increasing specialization in high-skilled sectors in developed countries, at the expense of the workers in the low-skilled sectors whose jobs were moved to countries with lower wages. Because of these developments, average income in the developed world has increased tremendously, but those at the bottom of society saw their opportunities deteriorate.

The rise in inequality presses governments in developed countries to respond. Protest movements around the world call for government intervention that should redistribute resources from the fortunate to the less fortunate. However, such government intervention comes at an economic cost. If the government increases the tax rate on the top 1 percent income earners, in order to redistribute the income to the poor, incentives to earn income become weaker. Income earners at the top may respond by working less, or less

[1] Source: Updated tables of Piketty and Saez (2003), all income including capital gains

efficiently, investing more resources in avoiding taxes, or even by moving their company to jurisdictions with a lower tax burden. The rich are not the only ones whose incentives are affected by an increase in redistribution. Students face lower returns to education, since the government will tax a larger part of their income once they start working. Also, the poor themselves have a lower incentive to work if government benefits are phased out with income. Therefore, a trade-off exists: measures that increase equity between rich and poor will generally do so at the expense of economic efficiency. If governments do not carefully design their tax-benefit system, measures taken to redistribute income may weight excessively on the economy, and actually worsen the situation for the poor.

To make matters worse, there is ample reason to believe that the costs and difficulty of redistribution have increased over the recent decades. Globalization has increased the mobility of firms and tax payers, thereby allowing them to locate in jurisdictions with lower tax rates. Governments in turn have to compete to attract them by lowering their tax rates. Financial innovation has made it easier for firms and tax payers to invest in financial products that are taxed favorably. The developments that lead to an increase in average productivity, and to an increase in inequality, also made it more costly for society to redistribute.

In addition, the increasing complexity in modern societies begs the question of who is deserving of government support. In particular, the number of single households, with and without children, and of dual income households has increased tremendously over the last decades. Should we redistribute to singles, or are dual earning families more deserving of income support? An added difficulty arises in the responsiveness of secondary income earners to taxation. If privileges, such as child care subsidies, of this group are removed secondary earners will likely decrease their labor supply or exit the labor market entirely. Also, many welfare states redistribute income towards the elderly. Traditionally, the elderly were among poorest members of society. However, during recent decades the average income in this group has risen tremendously, begging the question whether transfers to the elderly should not be moderated.

It may seem as if these global developments are unstoppable and countries should simply accept the rise of inequality that comes with them. Yet, this dissertation shows that we can adapt the welfare state to make it better equipped for the 21st century. Welfare states use a large number of instruments to redistribute from the fortunate to the less fortunate, such as tax instruments, subsidies, and transfers. However, they are not using them effectively. Policies that aim to increase the lot of the poor do not end up with the proper recipients, income tax rates are set too high at some income levels and too low at other levels, some very effective instruments for redistribution have not been considered, and no proper economic framework exists with which governments can assess the effectiveness of redistribution in various dimensions. As a result, governments redistribute less than they could, and they do so at a higher economic cost than necessary. This dissertation gives governments policy advice that enables them to reform their cacophony of redistributive instruments into a symphony, where each instrument play its intended melody and full harmony between the various redistributive instruments is achieved.

Optimal Labor Income Tax Rates in The Netherlands

In the second chapter of this dissertation my co-authors Bas Jacobs, Egbert Jongen and I study the optimal redistributive tax rates on labor income for the Netherlands. In the orchestra of redistributive instruments, surely the tax on labor income is the unrivaled soloist. One of the first questions asked during any course on public economics is why labor income is taxed. The typical first answer from students will be: 'To finance government expenditure.' But it has long been recognized by economists that this answer is not satisfactory. If the government simply wanted to finance its expenditure it would be better to introduce a lump-sum or head tax as was proposed by British conservative leader Margaret Thatcher. Unlike the tax on labor income, a head tax would not affect the behavior of households, and hence it would ensure that an efficient market outcome is achieved. Such a head tax is not popular because most believe that people with more labor income should contribute more to government expenditure than those with less labor income. Hence, the sole reason for having a tax on labor income is to distribute the burden of government expenditure to those with the best means to afford it. Since a large part of government expenditure exists of transfers paid to the poor, the labor income tax is effectively the main instrument that redistributes from rich to poor.

Under its legal definition, the tax on labor income is the largest contributor to government revenue in most OECD countries. However, economists typically take an even broader definition of the labor income tax. Throughout this thesis the marginal tax rate on labor is defined as the increase in net payment to the government if labor income increases by one euro. Under this definition, social insurance contributions are also part of the tax rate on labor, since contributions increase with labor income. In addition, many subsidy arrangements are phased out with labor income, such that an individual receives less subsidies from the government if his labor income goes up. Hence, these phase-outs also add to the tax rate on labor income. Even, indirect taxes on consumption, such as the value added tax are can be viewed as taxing labor, since the total amount of indirect taxes paid to the government increases strongly with labor income. Under this economic definition, the vast majority of government revenue, as well as a large part of government expenditure is part of the labor income tax.

Since the tax on labor income is such an important redistributive instrument, it is imperative to make sure this instrument plays the proper redistributive melody. This melody is determined by the trade-off between efficiency and equity, and can be calculated using optimal tax theory. The formula for the optimal tax rates on labor income was first presented in a Noble-prize winning article by James Mirrlees in 1971. Subsequently, Nobel Prize winner Peter Diamond and John Bates Clark Medalist Emmanuel Saez reinterpreted the formula such that we can use available data to calculate optimal tax rates for each country.

Chapter two is the first exercise that calculates the optimal labor income tax rates for a country in continental Europe. Although similar exercises have already been done for the US and the UK, the optimal tax rates in those countries are likely to be very different, because income inequality is much larger in Anglo-Saxon countries than in continental European countries. We use state-of-the-art estimates on labor supply elasticities and the income distribution in the Netherlands to calibrate the optimal tax model to Dutch data.

Obviously, the level of the optimal tax rates depends, in part, on political preferences. Left-wing parties will value redistribution more than right-wing parties, and will as a result opt for higher tax rates and higher transfers to the poor. In our approach, we abstain from taking a political position by calculating the optimal tax rates for the most extreme right -wing preferences and the most extreme left-wing preferences. If the current tax rates are set above what is optimal under extreme left-wing preference, or below what is optimal under extreme right-wing preferences, the current tax rates are suboptimal. The government can redistribute more at lower economic costs. By doing so it can increase the harmony of the redistributive symphony, independent of the listener.

Our findings show that the current Dutch tax system redistributes too much to the middle-income, and too little to the poor workers. This situation could be rectified by measures that increase the tax burden for middle-income earners, thereby boosting both equity and efficiency in the economy. We also show that raising the tax rate for top income earners will lead to less equity and less efficiency. Furthermore, we show that the optimal tax rate should vary by income level. Therefore, the introduction of a flat-tax will work counter productively. A more intricate melody is necessary to redistribute efficiently.

In the third chapter of this thesis Bas Jacobs, Egbert Jongen and I use optimal tax theory to elicit the preferences for redistribution of Dutch political parties. In a process unique in the world, all major Dutch political parties provide CPB Netherlands Bureau for Economic Policy Analysis with detailed reform proposals for the income tax rates in every national election. With optimal tax theory we can use this data to extract the welfare weight each political party gives to each income group in society, if the proposed tax rates are optimal to the political party. One would expect left-wing parties to give more weight to the poor, whereas right-wing parties are more likely to give welfare weight to the rich. In the previous chapter we saw that the current tax system redistributes too much to middle income groups. In this chapter we show that Dutch political parties want to maintain this status quo. Most of the welfare weight is given to the middle-income group for all political parties, independent of their stated ideology.

There are several reasons why political competition may drive parties to attract middle-income levels. Most importantly, this group is the largest in the economy. Therefore, alienating the middle-income group may be quite costly in terms of vote, and in terms of coalition forming after the elections. This analysis is of importance to public economics, since there exists no working theory on how political competition affect the tax schedules implemented by political actors. The results in this chapter provide the first empirical approach to elicit the preferences of political parties. We show that political opportunism leads political parties to overweight middle-income groups, thereby guiding future theories on the effect of politics on redistributive systems.

Duet of Labor Income Taxation and Monitoring

In the fourth chapter Bas Jacobs and I study a duet between the labor income tax and a new instrument: monitoring of labor effort. In the optimal tax model developed by Mirrlees the government can observe the labor income earned by individuals, but it cannot observe how much labor effort was provided to earn the income. It stands to reason that in order to make the same income level, a low-educated minimum wage employee must make much more working hours than a university graduate. Therefore, if the government could monitor the hours worked of each individual it could rewarding hard-working individuals or punishing those that shirk. This allows the government to redistribute more to those individuals that have the highest need for it, at lower economic cost.

Of course, using the extra monitoring instrument comes at a cost. The government has to thoroughly audit individuals in order to detect how many hours they have worked. Such audits may be expensive, since individuals may have an incentive to overstate the amount of hours they have worked. However, countries like the UK, Ireland and New-Zealand have already adopted a strategy similar to the one we are proposing by conditioning tax credits on the number of hours worked. Therefore, we know it is possible to use this instrument.

In this chapter we characterize the optimal formula for the monitoring schedule as a function of income. The formula shows that the government should monitor those individuals that face a high marginal tax rate on labor income, and those with a high labor supply elasticity. In addition, we show that monitoring raises optimal tax rates, thereby allowing governments to redistribute more income at lower costs.

Simulations on US data show that monitoring is particularly important at the bottom of the income-distribution, since these groups face a high marginal tax rate on their labor income in the optimum. In our base scenario the welfare gain of monitoring amounts to 1.4 percent of GDP. In addition, we show that monitoring may raise both equity and efficiency of the economy. The redistributive melody of the labor income tax attains a higher level of harmony when it is accompanied by monitoring of labor effort.

An Intermezzo of the Capital Income Tax

The fifth chapter is an intermezzo played by the capital income tax. In this chapter I study the effect of capital income taxation on savings and portfolio composition. Economists who study redistributive taxation are divided on whether capital income should be taxed or not. The traditional argument against capital taxation is that the households with large capital income are the households that have saved a lot of their labor income in previous years. Since the labor income was already taxed, taxation of capital income amounts to the government taxing the labor income of saving-prone households twice. Capital taxation will therefore lead to households saving too little of their income. In addition, households may invest too much of their savings into tax-favored assets such as owner-occupied housing and pension funds.

However, the result that capital should not be taxed has received a lot of criticism. If households differ in their investment skills, the government may want to redistribute from households with good investment skills to households with inferior investment skills through the capital income tax. In addition, the labor income tax may be better able to redistribute income if the government also taxes capital income. For example, top managers may have the ability to avoid labor income taxation by reporting their income as capital income. This motive for tax avoidance is reduced if the government also taxes capital income. In addition, young people may face the choice of investing in human capital or physical capital. Since the proceeds from human capital are taxed through the labor income tax, young people may under invest in human capital, unless physical capital is also taxed. Finally, too much pension savings may incentivize households to retire from the job market early. Therefore, it may be essential for governments to accompany their labor income tax with an instrument that taxes capital income. However, the level of the capital income tax rate crucially depends on how strongly household's savings and portfolio choices are affected by the tax. The stronger the effect of the capital income tax rates, the lower should be the tax rate on capital income.

Unfortunately, it is very difficult to measure the effect of taxation on household portfolio and savings choices. Due to tax progressivity, different households face different tax rates on their capital income. However, these households do not only differ in their tax rate, but also in many other characteristics such as, labor income, education, patience and risk preferences. In addition, effective capital income tax rates may differ over time due to shocks in the interest rate and the stock markets, and it is difficult to disentangle the direct effect of these shocks on savings and portfolio choices from the indirect effect that goes through the tax rate. To make matters worse, there is almost no reliable data on household's portfolio composition.

However, in 2001 the Netherlands went through a major capital income tax reform which altered the capital income tax rate for almost every household in the country. The reform can be studied as a quasi-experiment. Different households faced different shocks in their effective marginal capital tax rate. In addition, different asset categories were reformed in a different way. For this study Statistics Netherlands compiled a large database on household's portfolio composition before and after the reform. By comparing the same households before and after the reform, and comparing between households whose marginal tax rate went up, and households whose marginal tax rate went down, I can isolate the effect of the tax rate on households investment behavior.

Contrary to the current consensus in public economics, I find that capital income taxation has a relatively small impact on household's investment behavior. The effect of capital taxation on portfolio composition and overall savings is small. This indicates that governments should tax capital income at higher rates to redistribute more income.

The Symphony of All Redistributive Instruments

It is the government's task to redistribute from the fortunate to the less fortunate when insurance markets fail to do so. Chapters two to four considered redistribution between households with high labor income and households with low labor income and the fifth chapter looked at redistribution between households with high capital income and households with low capital income. However, there are many more dimensions in which the government redistributes. For instance, in the market for health insurance, most governments interfere by redistributing from the healthy to the sick, and in the labor market governments offer unemployment insurance to those that have recently lost their job.

Up to now, optimal taxation has only been able to study the different redistributive dimensions in isolation. The mathematical complexities of solving for the optimal combined tax and insurance instruments in a model where households differ in multiple dimensions have prevented public economists from making progress. A joint characterization of these instruments is important, because there may be a strong correlation between the different redistributive dimensions. Those that have lower labor earnings are also more likely to lose their job, more likely to get sick, and more likely to have low capital income. A model that can only study each of these redistributive dimensions in isolation will therefore likely redistribute too little to the least fortunate in society. In the sixth chapter of this thesis Sander Renes and I make a first step to finding the formula that orchestrates the full symphony of redistributive instruments.

We overcome the technical complexities that other economists struggled with by taking a so-called 'first-order' approach. The first-order approach calculates for each type in each dimension, the optimal redistribution between that type, and a type that is marginally different. For example, we calculate how much labor income should optimally be redistributed between someone with a labor income of 10,000 euros and somebody with labor income of 10,001 euros. Since we do this for for each type in the economy and all dimensions, we derive a formula for all redistributive instruments. Under some strict mathematical conditions provided in the chapter, we derive a formula that portrays the optimal redistributive symphony, using all available instruments.

Our formula has an intuitive interpretation and shows many similarities to optimal policies in simpler models where the government redistributes in only one dimension. It shows the optimal tax/subsidy rate on each good can be evaluated by the cost and benefits associated to the redistribution. In addition, we show that if a government wants to redistribute in several dimensions, it should also tax several goods. This may seem obvious, but in many models of optimal taxation the government only interferes in the economy by taxing labor income. This result fails once the government desires redistribution in multiple dimensions. This sheds new light on several government interventions, such as rent subsidy, capital taxation, and differential commodity taxation which are often undesirable in uni-dimensional models, but may be desirable in multi-dimensional models of optimal taxation. A redistributive symphony should use the entire orchestra.

The final chapter of this thesis discusses an important implicit assumption taken by public economists who study non-linear taxation. Calculating the non-linear tax rates that optimally redistribute from the fortunate to the less fortunate in a market economy is a task that is technically very complex. Therefore, economists, starting with the founding father of the optimal taxation James Mirrlees, study a problem that is somewhat simpler. In particular, in their models the choices taken by economic actors are rather limited. A typical example would be that a household can only increase its labor effort, and hence its labor income, if it also increases its level of savings in stride. This assumption simplifies the analysis significantly.

Placing such restrictions on constituents in a market-economy is undesirable. The freedom to choose would be severely limited, and market frictions may increase. Therefore, it is simply postulated in all of the preceding chapters of this thesis, as well as a large part of the literature, that the solution found in these simplified models will work equally well in a setting where constituents have full freedom to make their own decisions. However, it remains unclear whether this is indeed the case.

We show through an example that such a link between the simpler problem and the market economy does not always exist. Freedom of choice may lead to outcomes that the government deems entirely undesirable. However, the link does exist provided markets function properly, and the preferences of the constituents in the economy, and the government are aligned. Most of the tax systems designed in the literature, as well as the prescriptions provided in this thesis are therefore implementable in a market economy. We thus fill an important caveat in the models for studying redistributive taxation, and show that a full symphony of redistributive instruments can indeed work in a market economy.

Chapter 2

Optimal Redistributive Taxes and Redistributive Preferences in the Netherlands[1]

2.1 Introduction

What is the optimal structure of tax rates in a redistributive income tax system? This is a simple and policy-relevant question, but the answer is quite difficult. In his Nobel-prize winning article James Mirrlees (1971) wrote: "One would expect that in any economic system where equality is valued, progressive income taxation would be an important instrument of policy. [...] but there is virtually no relevant economic theory to appeal to, despite the importance of the tax (p.175)". Mirrlees has solved the theoretical problem of how to determine the optimal non-linear income tax and he concluded: "The problem seems to be a rather difficult one, even in the simplest cases"(p.175). Due to its analytical complexity, Mirrlees, and subsequently many authors, resorted to numerical simulations of the model to shed light on the shape of the optimal tax schedule.

[1]This chapter is based on Zoutman *et al.* (2013a). We would like to thank Miriam Gielen from CPB Netherlands Bureau for Economic Policy Analysis for her assistance with calculating the marginal tax rates used in this paper. We also thank seminar and congress participants at Tinbergen Institute Rotterdam, IIPF Michigan, CESifo Munich, Helsinki Center for Economic Research and UCL Louvain-la-Neuve for useful comments and suggestions. In addition I would like to express my gratitude to Olivier Bargain, Laurence Jacquet, Etienne Lehmann, Bertholt Leeftink, Erzo Luttmer, Luca Micheletto, Andreas Peichl, Pierre Pestieau, Rick van der Ploeg, Paul Tang, Coen Teulings, Emmanuel Saez, Laurent Simula, Trine Vattø and Danny Yagan for their comments, suggestions or assistance with this paper. All remaining errors are our own.

This study presents simulations of the optimal non-linear income tax for the Netherlands. To the best of our knowledge it is the first ever to do so.[2] Besides the intensive margin of labor supply, we will also allow for an extensive margin as in Jacquet *et al.* (2010), which combines Mirrlees (1971) with Diamond (1980). Recent advances in the empirical labor-supply literature point to the importance of the extensive margin for labor-supply decisions, see Blundell *et al.* (2011). This is also true for the Netherlands as Mastrogiacomo *et al.* (2013) have shown. Allowing for the participation margin has important implications for the setting of optimal income tax rates and the design of in-work tax credits, as stressed by Saez (2002b). Finally, by using the inverse optimal-tax approach developed in Bourguignon and Amedeo Spadaro (2010), we will derive implicit social welfare weights of the actual tax-benefit system had it been optimized. This allows us to detect inconsistencies in the current tax-benefit system and helps us finding welfare-improving tax reforms.

Our study sheds light on many policy questions that are currently fiercely debated in the Netherlands. For example, Should the Netherlands introduce a flat tax as proposed by Bovenberg and Teulings (2005) and Wetenschappelijk Instituut voor Het CDA (2009)? Should the tax rate at the top of the income distribution be raised as suggested by some political parties in the 2012 election platforms, *e.g.* by the Labor Party (PvdA), the Green Left and the Socialist Party?[3] Should social-assistance benefits be reduced, as proposed by *e.g.* the Christian Democratic Party (CDA), and the conservative-liberal party VVD[4], or increased, as proposed by the Socialist Party? Should the Netherlands increase the earned-income tax credit, as proposed by *e.g.* the conservative-liberal party VVD, the Christian Union (CU), the social-liberal party D66, the Labor Party, the Green Left and the Socialist Party? Furthermore, how should rent assistance, health-care subsidies, and subsidies to families with dependent children be phased out with income?[5]

The relevance of this paper extends beyond the Dutch case. The Netherlands is a country with a large amount of redistribution via the welfare state, which resembles other

[2] Jacobs (2008) calculates an optimal top marginal tax rate of 50% using empirical estimates by Atkinson and Salverda (2005) for the Pareto parameter for the top tail of the Dutch income distribution. However, he does not analyze the full optimal income tax schedule for the Netherlands.

[3] The Labor Party and Green Left propose to raise the income tax rate from 52 to 60% for taxable income beyond 150 thousand euro. The Socialist Party wants to raise the tax rate to 65%, see CPB and PBL (2012).

[4] In contrast to American-English use of the term 'liberal', this word has no 'left-wing' connotation in the Netherlands (and in many other European countries). To emphasize this distinction we use the adjective 'conservative'. In addition, there is a more left-leaning liberal party in the Netherlands, D66, which we label as 'social-liberal'.

[5] For a complete overview of the proposals for income dependent taxes and subsidies by the political parties in the 2012 elections, see CPB and PBL (2012).

European and Scandinavian countries in many respects.[6] Moreover, many policy questions in the Netherlands are discussed elsewhere. Most of the literature, however, has mainly focused on the Anglo-Saxon countries, in particular the US and the UK (Mirrlees, 1971; Tuomala, 1984; Diamond, 1998; Saez, 2001; Jacquet *et al.*, 2010). Our analysis reveals that there are some notable differences in the optimal non-linear tax schedules in comparison to those for the US and the UK.

Our main findings are the following. For the model with only an intensive labor supply margin, our calculations reveal that the current tax system is highly suboptimal. The optimal marginal tax schedule is U-shaped with decreasing marginal tax rates up to median income. However, marginal taxes are roughly increasing over the entire income distribution in the current tax system. The optimal top rate for the most redistributive (Rawlsian) social preferences is almost equal to the current top rate. For any social welfare function attaching a positive welfare weight to the top-income earners, the top rate is set beyond the top of the Laffer curve. Raising the top-rate to 55 or 60% lowers social welfare by both reducing redistribution and economic efficiency. In addition, we find that current marginal tax rates for the low-income groups and the average tax rate for middle-income groups are too low compared to the optimal non-linear tax schedule.

However, when the extensive margin of labor supply is included in the analysis, we find that optimal marginal tax rates are substantially lowered, especially for the low-income earners. Intuitively, by raising participation tax rates, marginal tax rates discourage participation. Because tax revenue declines when participation falls, marginal taxes should optimally be lower. Participation responses are especially important for the lower end of the earnings distribution. For high-income earners the participation responses to income taxation are relatively weak, since not many high-income earners will stop working when the marginal taxes slightly increase. With both intensive and extensive margins, the actual tax schedule is much closer to the optimum than with only an intensive margin. Nevertheless, the optimal tax schedules become more U-shaped when the participation margin is included, since marginal tax rates are lowered especially in the bottom half of the earnings distribution. In addition, compared to the current tax system, optimal marginal tax rates at the bottom are still higher under both Rawlsian and utilitarian social preferences. As regards the top rate, no real changes are found as optimal top rates are not sensitive to the participation margin.

[6] A principal-component analysis places the Netherlands in the group with Scandinavian countries, see Dekker and Ederveen (2003). Additionally, Bargain *et al.* (forthcoming) demonstrate that inequality aversion in the Netherlands ranks the fourth highest in Europe.

Our findings suggest that the current political system is either not able to redistribute income in the most efficient way or is not maximizing a standard social welfare function exhibiting declining social welfare weights. Policies to lower average taxes for the working poor, for example by raising the earned-income tax credit (EITC), have the potential to raise social welfare. The current government does not redistribute sufficient income to the 'working poor' in comparison with the 'non-working poor'. Whether participation is ultimately taxed or subsidized on a net basis critically depends on the social preferences for income redistribution. Intuitively, in-work subsidies redistribute resources from non-employed to employed workers, and this raises inequality between employed and non-employed workers. Only when the government has relatively weak redistributive social preferences, net participation subsidies are optimal.

The optimal tax system is heavily non-linear, thereby discarding the proposals for a flat tax. Indeed, our calculations suggest that the optimal flat tax always reduces efficiency, equity or both in comparison with the optimal non-linear tax. For the model with only an intensive margin the welfare losses of an optimal flat tax compared to the optimal non-linear tax are 0.4% of GDP for utilitarian social objectives, and increasing until 9% of GDP for Rawlsian social objectives. These findings reveal that the flat tax is a particularly tight strait jacket when social preferences are more redistributive. Intuitively, the flat tax employs no information on individual earnings and income redistribution cannot be effectively targeted to the individuals with the lowest incomes. The flat tax requires much higher marginal tax rates to obtain a given amount of income redistribution. Therefore, the equity-efficiency trade-off worsens and Okun (1975)'s 'leaking bucket' becomes a sieve.

Finally, by computing the social welfare weights implied by the current tax-benefit system, we indeed find that social welfare weights do *not* monotonically decline; welfare weights first increase and then decrease, become slightly negative for the top-income earners, and are zero in the limit. Social welfare weights discontinuously drop with about one third when individuals earning no income start participating in the labor market. Hence, the current political system does not seem to maximize a standard social welfare function. Instead, the current government redistributes resources away from the working poor towards the middle-income groups. Moreover, it soaks the top-income earners as much as it can, and it even penalizes them by setting too high marginal taxes. Finally, the current political system strongly prefers transferring resources to the non-working rather than to the working poor. This could be taken as evidence that political-economy considerations matter a lot in shaping actual tax schedules. Moreover, the non-working poor are apparently seen as much more deserving of income support than the working poor, for reasons that remain unclear to us.

This paper is structured as follows. Section 2 provides a review of the literature and discuss our paper's contributions. Section 3 derives the model used in our simulations. In Section 4 we introduce the data, compute marginal and participation tax rates, estimate the Pareto parameter for the Dutch income distribution, calibrate the utility function, and estimate the distributions of skills and participation costs. In Section 5 we compute the optimal non-linear income tax for the Netherlands for models with only an intensive labor supply margin and with both an intensive and an extensive labor supply margin. In Section 6 we calculate the social welfare weights that are implied by the current Dutch tax and benefit system. In Section 7 we discuss the limitations of our analysis and provides the policy conclusions. Various Appendices contain less essential technical derivations.

2.2 Earlier Literature

Our study aims to contribute to the scientific literature on optimal taxation by presenting advanced simulations of the Mirrlees (1971) model, which is extended with an extensive labor supply margin and income effects along the lines of Jacquet *et al.* (2010). From his own simulations, Mirrlees concluded: "[P]erhaps the most striking feature of the results is the closeness to linearity of the tax schedules (p.206)". However, subsequent research has shown that this conclusion was premature, the result depended heavily on functional form assumptions for the utility function (Cobb-Douglas) and the income distribution (log-normal). Tuomala (1984) uses a different utility function, which allows for a more realistic elasticity of taxable income and finds declining optimal marginal tax rates with income. Diamond (1998) finds a U-shape for optimal marginal tax rates, using a Pareto distribution for the top incomes. The log-normal distribution implies an upper tail for the income distribution that is too 'thin', resulting in optimal top tax rates that are too low. Saez (2001) also finds a U-shape for optimal marginal tax rates, extending the analysis of Diamond (1998) by *e.g.* allowing for income effects. Our paper contributes in a number of important ways to the existing literature on optimal-tax simulations.

A substantial part of this paper is devoted to estimating the joint distribution of ability and participation costs. In contrast to earlier papers that assumed synthetic skill distributions (Mirrlees, 1971; Tuomala, 1984), we estimate the skill distribution using the structural method pioneered by Saez (2001) and Bourguignon and Spadaro (2000). In particular, we assume that individuals maximize a particular utility function, which is defined over consumption and labor supply, subject to a non-linear budget constraint.

We carefully reconstruct the individuals' budget constraints, taking into account income-dependent transfers, numerous tax credits, indirect taxes, and welfare benefits. The first-order conditions and the household budget constraints enable us to retrieve the non-observable skill level for each household. In doing so, we will rely on advanced, recent Dutch estimates for the elasticity of taxable income (Jongen and Stoel, 2013a), the intensive labor-supply elasticity and the extensive labor-supply elasticity (Mastrogiacomo *et al.*, 2013). Additionally, this study very precisely estimates a Pareto distribution for the top of the Dutch skill distribution. The Pareto parameter is estimated to be around 3.35, which is among the highest found in the literature, suggesting that it is lonely at the top in the Netherlands, see the overviews in Heady (2010) and Atkinson *et al.* (2011).

We confirm the findings of Saez (2001) that the optimal tax schedules feature a U-shape. Marginal tax rates at lowest income groups are very high, in the order of 70-80%. Marginal tax rates decline towards the middle-income groups, increase again after middle-income groups, and converge to a constant of about 50% for the top-income earners. The increase in the marginal tax rates after modal income is, however, much more limited than in Saez (2001), which is due to the very thin top tail of the earnings distribution in the Netherlands. Indeed, for the Rawlsian social welfare function, we find only a tiny increase in marginal tax rates, in contrast to the US. For the same reason, optimal top rates in the Netherlands are much lower than those for the US (Saez, 2001) or the UK (Brewer *et al.*, 2010).

We further contribute to the analysis of optimal income tax simulations with both intensive and extensive labor-supply margins. In contrast to Jacquet *et al.* (2010), we estimate the distribution of participation costs using the first-order conditions for labor-market participation, which are supplemented with data on participation rates and participation taxes under the current tax-benefit system. Like Jacquet *et al.* (2010), we assume that idiosyncratic participation costs/benefits are separable from leisure and consumption and that the distribution of participation costs is normal. We estimate the parameters of the distribution of participation costs such that participation rates match with skill-specific employment rates. Participation rates in the Netherlands run from 0.51 for low-educated individuals to 0.85 for high-educated individuals. In this way, we are able to fully determine the non-observed joint distribution of ability and participation costs.

Our findings contrast sharply with the baseline simulation of Jacquet *et al.* (2010). When including the extensive margin, they find that the optimal non-linear tax schedule shifts downwards across all income levels compared to the optimal schedule without the extensive margin. We find that marginal taxes mainly fall in the bottom part of the skill distribution. We believe that this is due to the different specifications of participation costs employed in both studies. Jacquet *et al.* (2010) do not estimate the distribution of participation costs, but assume that i) participation rates have some non-linear relation with ability, and ii) participation elasticities have a linear relation with ability. These relationships are not empirically estimated. As a result, participation rates are very similar for low- and high-skill types, rising from 0.7 for the lowest to 0.8 for the highest skill types. Consequently, the optimal tax schedule shifts down in Jacquet *et al.* (2010) for all income levels. Our specification of participation costs implies that net participation costs are much lower for higher-skilled individuals, since they have much higher participation rates. The participation elasticities in our model (about 0.25) are calibrated on empirical values. We believe that this explains why optimal tax schedules mainly shifts down for low-skilled individuals, and not for the high-skilled individuals, when including the participation margin in the Mirrlees model.[7]

The inverse optimal-tax problem is analyzed in Bourguignon and Spadaro (2000), Bourguignon and Amedeo Spadaro (2010), Blundell *et al.* (2009), Bargain and Keane (2010) and Bargain *et al.* (2011). Inconsistencies of actual tax systems compared to optimal tax systems derived from standard social welfare functions are found in most of the literature. Bourguignon and Amedeo Spadaro (2010) find monotonically declining welfare weights for singles in France when considering only the intensive margin of labor supply. However, welfare weights turn negative for high-income earners. In addition, when they include the extensive margin, welfare weights are not monotonically declining. In particular the working poor receive a negative social welfare weight, while both the unemployed and the middle-income earners obtain a positive weight. They conclude that policy makers underestimate the tax distortions on the extensive margin. Blundell *et al.* (2009) consider single mothers in Germany and the UK analyzing both intensive and extensive responses. They also find that weights are not monotonically decreasing with a dip in the welfare weight for the working poor. Similar results are found in Bargain and Keane (2010) and Bargain *et al.* (2011), where the authors consider singles in respectively Ireland, and 17 European countries and the US.

We also detect numerous inconsistencies in the current tax-benefit system. In particular, our analysis demonstrates that social welfare weights are not continuously declining, but increasing up to modal income, which could be explained by political-economy considerations. After modal income, they decline as expected, but even turn negative for the very high income earners, indicating that tax rates are set beyond the top of the

[7]A sensitivity analysis of Jacquet *et al.* (2010) indeed confirms that when participation elasticities fall with income, the effect on the optimal tax schedule will be more pronounced at the lower income levels.

Laffer curve. Moreover, welfare weights discontinuously drop for workers moving from non-participation to participation, suggesting that income is taxed too highly for the low-income earners.

2.3 Theory

In this section we first introduce the Mirrlees (1971) model of optimal labor income taxation with intensive labor-supply responses. Then, we analyze the model of Jacquet *et al.* (2010) with both intensive and extensive labor-supply responses. Finally, this section derives how to determine the social welfare weights implicit in our current tax-benefit system using the inverse optimal-tax method.

2.3.1 Intensive Margin

Individuals

We follow the optimal-tax literature by supposing that heterogeneity in individual types derives from their exogenous ability to earn income (and their participation costs when the participation decision is included, see below). The fundamental insight of Vickrey (1947) and Mirrlees (1971) is that earnings ability is not observable by the government. Due to the non-observability of ability the government needs to resort to distortionary tax instruments, most importantly taxes on labor income, to redistribute income. Taxing labor income is distortionary because it not only taxes the return to ability, but also the fruits of labor effort. Hence, income redistribution leads to the well-known trade-off between equity and efficiency.[8]

Ability is distributed according to probability density function $f(n)$ and corresponding cumulative distribution function $F(n)$, with support $\mathcal{N} \equiv [\underline{n}, \overline{n})$. The upper bound $\overline{n}$ can be infinite. n denotes the number of efficiency units of labor. We follow Mirrlees (1971) by assuming perfect substitution between skill types on the labor market. Hence, by normalizing the wage rate per efficiency unit of labor to unity, we can associate n with the wage rate per hour worked of individual n. Gross labor earnings of an individual with ability n are given by $z_n \equiv nl_n$ where l_n denotes the normalized labor supply of an individual with ability n.

[8]As long as ability differences are the only source of heterogeneity, and preferences of individuals are homogeneous and weakly separable, only the non-linear income tax will be employed for income redistribution (Atkinson and Stiglitz, 1976).

All net income is consumed, hence the individuals' budget constraint is given by:

$$c_n = z_n - T(z_n), \quad \forall n, \tag{2.1}$$

where c_n denotes consumption and $T(z_n)$ the tax schedule as a function of gross labor income. $T'(z_n) \equiv \mathrm{d}T(z_n)/\mathrm{d}z_n$ is the marginal tax rate. All individuals have identical preferences over consumption c_n and labor l_n, which are represented by a separable, continuous and twice-continuously differentiable utility function:[9]

$$u_n \equiv v\left(c_n\right) - h\left(l_n\right), v', h' > 0, \quad v'', -h'' \leq 0, \quad \forall n, \tag{2.2}$$

where $v(\cdot)$ is a concave function representing the utility of consumption, and $h(\cdot)$ is a convex function representing the disutility of labor effort. By substituting the budget constraint (2.1), into the utility function (2.2), the maximization problem of the individual can be stated as:

$$\max_{z_n} v\left(z_n - T(z_n)\right) - h\left(\frac{z_n}{n}\right), \quad \forall n, \tag{2.3}$$

The first-order condition (FOC) of this problem is:

$$\left(1 - T'(z_n)\right) v'(c_n) = \frac{h'(l_n)}{n}, \quad \forall n. \tag{2.4}$$

The marginal benefits of earning an additional euro on the labor market, as represented by the left-hand side, are equated to the marginal utility cost of labor required to earn the additional euro of income, as represented by the right-hand side. As can be seen, the marginal benefits of work are decreasing in the tax rate.

The allocation is said to be incentive compatible if the following first-order incentive-compatibility constraint holds:

$$\frac{\mathrm{d}u_n}{\mathrm{d}n} = \frac{l_n h'(l_n)}{n}, \quad \forall n. \tag{2.5}$$

This condition can be derived from totally differentiating utility with respect to ability and using the first-order condition for labor supply.

The incentive-compatibility constraint (2.5) is a necessary constraint. However, each incentive-compatible allocation must also respect second-order sufficiency conditions for utility maximization (Mirrlees, 1976). This is the case if, in addition, the Spence-Mirrlees

[9] The assumption of separability in the utility function is made in all simulation studies in the literature. Numerically, it is very difficult, if not impossible, to simulate the optimal tax schedule if the utility function is non-separable. See e.g. Mirrlees (1971) and Saez (2001).

and monotonicity constraints are satisfied:

$$\frac{\mathrm{d}\left(\frac{h'(l_n)}{nv'(c_n)}\right)}{\mathrm{d}n} \leq 0, \quad \forall n, \tag{2.6}$$

$$\frac{\mathrm{d}z_n}{\mathrm{d}n} > 0, \quad \forall n. \tag{2.7}$$

These conditions imply that the utility function features the single-crossing property. Hence, at each bundle of gross and net income, individuals with a higher ability have incentives to self-select into the bundles with higher net and gross income. The Spence-Mirrlees condition is satisfied by most utility functions used in the literature, including the ones that are used in our simulations. The second condition states that income should increase monotonically with ability.[10] Hence, the second condition ensures that self-selection of higher ability types into higher consumption-earnings bundles will also occur. From the monotonicity condition we can derive that it is never optimal to have higher marginal tax rates than 100%, otherwise the monotonicity condition would be violated, since it implies that $\frac{\mathrm{d}c_n}{\mathrm{d}n} > 0$, see Mirrlees (1976).

In our simulations we will use the first-order approach using (2.5), assuming that the second-order conditions will be satisfied. After having derived the optimal allocation, we will check ex post whether the sufficiency conditions (2.6) and (2.7), are indeed met, which is always the case.

Government

The objective of the government is to maximize social welfare. Social welfare is assumed to be described by a Samuelson-Bergson social welfare function, which is a concave sum of individual utilities:

$$\int_{\mathcal{N}} W(u_n) f(n) \mathrm{d}n, \quad W' > 0, \quad W'' \leq 0, \tag{2.8}$$

Redistribution is socially desirable if either the social marginal value of utility (W') or the private marginal value of income (u_c) are decreasing, i.e., $W'' < 0$ or $u_{cc} < 0$.[11] The government has to respect the economy's resource constraint:

$$\int_{\mathcal{N}} (z_n - c_n) f(n) \geq R, \tag{2.9}$$

[10] See Ebert (1992) for a detailed discussion on this issue.

[11] In the extreme case, where both $W'' = 0$ and $u_{cc} = 0$, the optimal tax problem becomes trivial, as there is no social desire for redistribution and the government will finance all of its spending through non-distortinary lump-sum taxes.

where R denotes exogenous government expenditure. As long as the economy's resource constraint and the household budget constraints are met, also the government budget constraint is satisfied by Walras' law.

Optimal Income Taxation

The optimal allocation is found by maximizing the social welfare function, (2.8), subject to the resource constraint, (2.9), and the incentive compatibility constraint, (2.5). The Appendix derives that the optimal schedule of marginal income taxes then satisfies the following ABC-formula:

$$\frac{T'(z_n)}{1-T'(z_n)} = A_n B_n C_n, \quad \forall n, \tag{2.10}$$

where:

$$A_n \equiv \frac{1}{\varepsilon_n^c}, \quad \varepsilon_n^c \equiv -\frac{\partial l_n}{\partial T'(z_n)} \frac{1-T'(z_n)}{l_n}, \quad \forall n, \tag{2.11}$$

$$B_n \equiv \frac{v'(c_n) \int_n^{\overline{n}} \frac{1-g_m}{v'(c_m)} f(m)\,\mathrm{d}m}{1-F(n)}, \quad g_n \equiv \frac{W'(u_n)v'(c_n)}{\lambda} \quad \forall n, \tag{2.12}$$

$$C_n \equiv (1+\varepsilon_n^u)\frac{(1-F(n))}{f(n)n}, \quad \varepsilon_n^u \equiv \frac{\partial l_n}{\partial n}\frac{n}{l_n}, \quad \forall n, \tag{2.13}$$

where ε_n^c is the compensated *tax* elasticity of labor supply, ε_n^u is the uncompensated wage elasticity of labor supply, and g_n is the social marginal value (in monetary terms) of providing individual n a unit of resources. We shall refer to g_n as the social welfare weight of individual n.

At each point of the income distribution, marginal equity gains and efficiency losses of the marginal tax rate are equalized. Intuitively, the function of the marginal tax rate at an income level z_n is to raise tax revenue from all individuals above z_n. The marginal tax rate at z_n redistributes resources from individuals above z_n to the government. In turn, the government can use this revenue to raise the uniform transfer $-T(0)$ in the tax system. A higher marginal tax rate at z_n thus increases the average tax burden above z_n and lowers the average tax burden on individuals below z_n.

A_n represents the efficiency costs of having a marginal tax at income level z_n. If the marginal tax rate at income level z_n is increased, individuals with income z_n have an incentive to decrease their labor supply. This behavioral response is captured by the compensated elasticity of labor supply with respect to the tax rate ε_n^c.

B_n represents the average redistributional gain of having a marginal tax at income z_n. B_n is equal to the revenue of a euro increase in taxes on individuals above z_n, minus the monetized value of the welfare loss g_n due to extracting an additional euro revenue from these individuals. The difference is represented by the term $1-g_n$. B_n averages this difference over all individuals with an income above z_n.

Term C_n gives weights to terms A_n and B_n via the distribution of earnings ability. The optimal tax rate is determined by the number of individuals paying the marginal tax rate $(1-F(n))$ and the number of individuals whose labor supply choice is distorted $(nf(n))$. The more individuals above income level z_n, the larger the redistributive gains of a higher marginal tax. The more individuals at skill level n, or the larger their wage rates, the larger is the tax base, and, therefore, the larger are the efficiency losses of a higher marginal tax rate.

If we would express the optimal-tax formula in terms of earnings densities, rather than the densities of the ability distribution, the C_n-term would collapse to $C_n = \frac{1-\tilde{F}(z_n)}{\tilde{f}(z_n)z_n}$, where $\tilde{F}(z_n) \equiv F(n)$ is the cumulative earnings distribution, $\tilde{f}(z_n)$ is the earnings density at z_n, and $z_n\tilde{f}(z_n) = (1+\varepsilon_n^u)nf(n)$, see Saez (2001). Hence, the C_n term is entirely determined by the shape of the empirical earnings distribution $\tilde{f}(z_n)$.

There is no closed-form solution for the optimal tax rate. Nevertheless, a few properties of optimal tax schedules can be established analytically. We already derived that the optimal marginal tax rate is never above 100% at any income level. In addition, the marginal tax rate is never below 0%, see the ABC-formula. Indeed, a marginal tax rate below 0% redistributes income in the wrong direction, and thereby lowers social welfare (Seade, 1982).[12] Sadka (1976) and Seade (1977) show that the marginal tax rate at the bottom and the top of the skill distribution should be equal to zero if the skill distribution has a finite top and all individuals provide positive work effort. Intuitively, there are no redistributional gains and only distortions associated with marginal taxes at the endpoints, so that marginal tax rates are zero.[13] As is shown in Diamond (1998), the result of the zero tax rate at the top does not apply if (the top of) the skill distribution is Pareto distributed. Later, we will demonstrate that this is also the case for the Netherlands. Similarly, the zero marginal tax at the bottom is positive, and generally very large, when there is an atom of non-working individuals, which is observed in the real world, and assumed in most simulations. No further analytical results can be obtained. Therefore, many authors have resorted to simulations of the optimal non-linear tax schedule. That is what we will do in the remainder of this paper, after we introduced the extensive margin.

[12] Note that a zero marginal tax rate at the bottom or the top does not imply that there is no redistribution on average. The amount of redistribution is determined by the average tax individuals pay over all their income, not by the marginal tax they pay over their last euro of income.

[13] Especially, the zero top rate has attracted a lot of attention. However, its practical applicability is limited, because it is a very local result, see Tuomala (1984).

2.3.2 Extensive Margin

In the Mirrlees model individuals can only adjust their labor supply on the intensive margin. They can decide to work more or less, but they cannot decide to enter or exit the labor market entirely. In contrast, Diamond (1980) derives the optimal tax schedule where individuals can only adjust their labor supply along the extensive margin, but not on the intensive margin. Saez (2002b) and Jacquet *et al.* (2010) combine the Mirrlees-model with the Diamond-model to analyze the optimal non-linear income tax and the optimal participation tax. In this paper, we will follow the analysis of Jacquet *et al.* (2010) to find the optimal tax schedule with both intensive and extensive labor-supply responses for the Netherlands.

Individuals

The extensive margin is introduced through a random participation model. Each individual has an individual-specific participation utility cost φ of entering the labor market, which reflects the individuals' outside options such as household production or income from the black labor market. We also allow some individuals to have a negative disutility of participation. This could be related to a social stigma of being unemployed. We assume that the disutility of participation is unobservable to the government. φ follows a probability density function conditional on ability n given by $k(\varphi|n)$. The corresponding cumulative distribution function is $K(\varphi|n)$. The support, also potentially conditional on n, is given by $[\underline{\varphi}^n, \overline{\varphi}^n]$.

Individuals can decide not to participate and receive unemployment benefits b. We assume that the government can verify the employment status of an individual, and hence, condition non-employment benefits on it. The utility of a non-employed worker is equal to $v(b)$. b can be different from the transfer $-T(0)$ implied by the tax schedule. The utility of an employed individual with ability n and discrete participation cost φ is given by:

$$U_n \equiv v(c_n) - h(l_n) - \varphi, \quad \forall n. \tag{2.14}$$

An individual decides to participate in the labor market if the maximum utility derived of participation is at least as large as the utility of the unemployment benefits:

$$u_n - \varphi \geq v(b), \quad \forall n. \tag{2.15}$$

where $u_n \equiv v(c_n) - h(l_n)$. The individual will participate if his/her utility cost of working is sufficiently low, or if his/her ability to earn income is sufficiently high.

The participation tax is the net extra amount of tax an individual pays if he/she decides to participate and earns gross income level z_n. The participation tax consists of two components. First, when working the individual is subject to the tax schedule $T(z_n)$, and, second, the individual loses his/her benefits b. The total participation tax is therefore $T(z_n)+b$. A higher participation tax naturally discourages participation. We do not constrain the participation tax to be positive, and, therefore, the government is also allowed to give a participation subsidy when this raises social welfare, i.e., $T(z)+b<0$ would imply an in-work tax credit.

The incentive-compatibility constraint (2.5) is unaltered by the introduction of the participation costs. Intuitively, a worker with ability n has to incur participation cost φ irrespective of whether the worker self-selects in the consumption-income bundle for type n or decides to mimic a worker of type m to obtain the consumption-income bundle intended for type m.

Government

The government's objective is a weighted sum of the utility of non-employed and the utility of employed workers:

$$\int_{\mathcal{N}} \left(\int_{\underline{\varphi}^n}^{u_n - v(b)} W(u_n - \varphi) k(\varphi|n) \mathrm{d}\varphi f(n) + W\left(v(b)\right)\left(f(n) - \tilde{k}(n)\right) \right) \mathrm{d}n. \tag{2.16}$$

The bounds of the inner integral are given by equation (2.15). Therefore, all individuals with φ in $[\underline{\varphi}^n, u_n - v(b)]$, given by $K\left(u_n - v(b)|n\right)$, participate, and all individuals with φ in $(u_n - v(b), \overline{\varphi}^n]$ do not participate. There are $f(n)$ individuals with ability n, and, hence, the fraction of individuals in the population that work at skill level n is given by $\tilde{k}(n) \equiv K\left(u_n - v(b)|n\right) f(n)$. The fraction of non-employed individuals at skill level n is, therefore, $f(n)-\tilde{k}(n)$, as can be seen in the second term of equation (2.16).

Correspondingly, the economy's resource constraint is modified to:

$$\int_{\mathcal{N}} \Big((z_n - c_n)\tilde{k}(n) - (f(n) - \tilde{k}(n))b \Big) \mathrm{d}n \geq R. \tag{2.17}$$

First, note that at each skill level n only the working fraction of the population $\tilde{k}(n)$ produces output z_n and consumes c_n. Second, at each skill level n, the non-working population $f(n) - \tilde{k}(n)$ does not produce anything and consumes its unemployment benefits b.

Optimal Income Taxation

The adjusted ABC-formula for optimal taxation in the presence of intensive and extensive labor-supply responses is given by – see Appendix for the derivation:

$$\frac{T'(z_n)}{1-T'(z_n)} = A_n B_n C_n, \quad \forall n, \tag{2.18}$$

where:

$$A_n \equiv \frac{1}{\varepsilon_n^c}, \quad \varepsilon_n^c \equiv -\frac{\partial l_n}{\partial T'(z_n)} \frac{1-T'(z_n)}{l_n}, \quad \forall n, \tag{2.19}$$

$$B_n \equiv \frac{v'(c_n) \int_n^{\overline{n}} \left(\frac{1-g_m^P}{v'(c_m)} - \kappa_m \left(b + T(z_m)\right)\right) \tilde{k}(m)\mathrm{d}m}{\tilde{K}(\overline{n}) - \tilde{K}(n)}, \tag{2.20}$$

$$\kappa_n \equiv \frac{K'\left(u_n - v(b)|n\right) f(n)}{\tilde{k}(n)}, \quad g_n^P \equiv \frac{\int_{\underline{\varphi}^n}^{u_n - v(b)} \frac{W'(u_n-\varphi)v'(c_n)}{\lambda} k(\varphi|n)\mathrm{d}\varphi}{K\left(u_n - b\right)}, \quad \forall n, \tag{2.21}$$

$$C_n \equiv (1+\varepsilon_n^u) \frac{\tilde{K}(\overline{n}) - \tilde{K}(n)}{n\tilde{k}(n)}, \quad \forall n, \tag{2.22}$$

where g_n^P is the social welfare weight given to *employed* workers, and κ_n is the semi-elasticity of participation with respect to a utility increase for the employed. $\tilde{k}(n)$ is the fraction of employed with ability level n, $\tilde{K}(n)$ is the fraction of employed in the population with ability n or less, and $\tilde{K}(\overline{n})$ is the total fraction of workers participating.

Term A_n and its interpretation are unaltered by the introduction of the extensive margin. In term C_n all occurrences of the distribution of earnings ability ($f(n)$ and $1-F(n)$) have been replaced by the distribution of employed workers ($\tilde{K}(\overline{n}) - \tilde{K}(n)$ and $\tilde{k}(n)$). Intuitively, weights to term A_n and B_n should be given on the basis of the number of employed workers, because non-employed workers do not pay the marginal tax rate.

The largest difference with the model without an extensive margin is found in term B_n. The extensive margin reduces the average revenue available for redistribution, because a higher marginal tax rate results in revenue losses by discouraging labor-force participation. Suppose the government increases the tax rate at income level z_n such that all individuals above z_n need to pay one additional euro extra tax. Mechanically this raises the tax revenue for all individuals with income z_n or larger by 1 euro. In addition, the government inflicts a welfare loss on all taxed individuals as represented in normalized welfare weights g_n^P. Finally, a higher marginal tax increases the total tax paid by working individuals, and, thereby, reduces the attractiveness of participation compared to non-participation. The government loses revenue as some individuals decide to exit the labor market, stop paying taxes and start collecting non-employment benefits. The decline in revenue is thus determined by the participation elasticity, κ_n, which governs the reduction in participation, and by the net participation tax $T(z_n) + b$. Term B_n is the average of the difference, $1 - g_n^P - \kappa_n(T(z_n) + b)$, over all employed workers with an income level above z_n.

In addition, term B_n may change due to a second-order effect. The optimal marginal tax rates decrease due to the participation margin, which is captured by $\kappa_n(T(z_n) + b)$. Therefore, this reduces the amount of income redistribution. In turn, the decrease in redistribution might raise the value of redistribution B_n at some income levels as g_n^P decreases when individuals are taxed less. This second-order effect might actually lead to an increase in the marginal tax rate at some income levels. B_n decreases at low- and medium-income levels, because the participation elasticity κ_n is large among these income groups. However, B_n increases at high-income levels because the participation elasticity at these income levels is typically small. If g_n^P falls enough to offset the effect of $\kappa_n(T(z_n) + b)$, then B_n rises and the tax rate at high-income earners could increase. This will never occur under Rawlsian preferences, since social welfare weights for employed workers g^P are constant and equal to zero.

Optimal Participation Taxation and Transfers

Above, we derive the optimal marginal tax rates. However, the government also optimizes the optimal non-employment benefits b. The latter determines the optimal participation taxes at each point in the income distribution. The optimal non-employment benefit b is set such that the following equation is satisfied – see Appendix for the derivation:

$$\int_{\mathcal{N}} \kappa_m(T(z_m) + b)\tilde{k}(m)\mathrm{d}m = \int_{\mathcal{N}} \frac{(1 - g_m^P)}{v'(c_m)}\tilde{k}(m)\mathrm{d}m, \quad \forall n. \tag{2.23}$$

This equation equates total distortions on participation with the total gains of income redistribution from workers to non-workers. The left-hand side gives the distortion in participation of providing a higher non-employment benefit b, which is captured by the participation elasticity κ_n, times the participation tax $T(z_n) + b$, aggregated over all households. The right-hand side gives the total distributional benefits of providing higher non-employment benefits. Distributional benefits occur if $g_n^P < 1$ at skill level n, while they yield distributional losses if $g_n^P > 1$ at skill level n. Hence, the larger are the distributional benefits of transferring resources to the non-working part of the population, the larger will be the optimal participation distortions.

The optimal intercept of the tax function $T(0)$ is determined implicitly by ensuring that the weighted average of the marginal social welfare weights sum to one (as we have derived before), see the Appendix:

$$\frac{(g_0 - 1)}{v'(b)}(1 - \tilde{K}(\overline{n})) = \int_{\mathcal{N}} \frac{(1 - g_m^P)}{v'(c_n)} \tilde{k}(m)\mathrm{d}m, \tag{2.24}$$

where $g_0 \equiv W'(v(b))v'(b)/\lambda$ denotes the marginal social welfare weight of non-employed individuals. This equation ensures that the marginal euro is valued equally by the public and private sector. Equivalently, this equation states that the marginal cost of public funds equals one at the optimal tax system. Distributional benefits cancel against deadweight losses at the optimal tax system, see Jacobs (2013).

2.3.3 Inverse Optimal-Tax Problem

In the last part of this paper, we will calculate the social welfare weights g_n^P at each skill level n using the inverse optimal-tax problem, see also Bourguignon and Spadaro (2000), Bourguignon and Amedeo Spadaro (2010), Blundell *et al.* (2009), Bargain and Keane (2010) and Bargain *et al.* (2011). That is, by supposing the government has maximized social welfare by optimally designing its tax-benefit system, we can use the current tax-transfer system to back out the social welfare weights corresponding to the social welfare function.

The implied social welfare weights for employed workers can be found by solving the optimal tax formula, (2.18), for normalized welfare weights g_n^P. First, we rewrite equation (2.18) for B_n:

$$B_n = \frac{T'(z_n)}{1 - T'(z_n)} \frac{1}{A_n C_n}. \tag{2.25}$$

Second, insert the definition for A_n, B_n, and C_n from equations (2.19), (2.20), and (2.22), and simplify:

$$\int_n^{\overline{n}} \left(\frac{(1 - g_m^P)}{v'(c_m)} - \kappa_m(b + T(z_m)) \right) \tilde{k}(m)\, \mathrm{d}m = \frac{T'(z_n)}{1 - T'(z_n)} \frac{\varepsilon_n^c n \tilde{k}(n)}{(1 + \varepsilon_n^u)\, v'(c_n)}. \tag{2.26}$$

Next, differentiate both sides of the equation with respect to n, and apply Leibniz' rule to the left-hand side:

$$g_n^P = 1 - \kappa_n(b + T(z_n))v'(c_n) + \frac{\mathrm{d}}{\mathrm{d}n} \left[\frac{T'(z_n)}{1 - T'(z_n)} \frac{\varepsilon_n^c n \tilde{k}(n)}{(1 + \varepsilon_n^u)\, v'(c_n)} \right] \frac{v'(c_n)}{\tilde{k}(n)}. \tag{2.27}$$

We cannot obtain an analytical solution for the differential on the right-hand side. Therefore, we numerically approximate the expression in our calculations.

In addition, we can derive the weight of the unemployed by solving equation (2.24) for g_0:

$$g_0 = 1 + \frac{v'(b)}{(1 - \tilde{K}(\overline{n}))} \int_{\mathcal{N}} \frac{(1 - g_m^P)\tilde{k}(m)\mathrm{d}m}{v'(c_m)}. \tag{2.28}$$

Applying equations (2.27) and (2.28) to our data yield the welfare weights that are implied by the current Dutch tax-benefit system. We calculate welfare weights on the intensive margin by setting κ_n equal to zero and replacing all instances of $\tilde{k}(n)$ with $f(n)$ in equations (2.27) and (2.28).

2.4 Data and Calibration

To compute the optimal non-linear tax-benefit system we need the following ingredients: i) the distribution of skills and participation costs/benefits, which determine the amount of income inequality and the number of non-participating individuals; ii) the utility function, which governs the behavioral impact of taxes and transfers; and iii) the social welfare function, which gives the social preferences to redistribute income. In this section, we define labor income, and determine the corresponding earnings distribution. Since there are only few observations on earnings for the top tail of the earnings distribution, we estimate the top of the income distribution with a Pareto distribution, which gives an excellent fit for top incomes. We then define marginal tax rates, and determine the distribution of marginal tax rates. We use the data on labor income and marginal tax rates together with a utility function consistent with empirical studies to retrieve the ability distribution. Together with data on participation by level of education (as a proxy for skill) we use these to calibrate the distribution of idiosyncratic participation costs/benefits. We, finally, determine the revenue requirement of the government and the transfer to non-employed.

2.4.1 Income Distribution

Following Brewer *et al.* (2010), we use labor costs rather than gross wages, because the former includes all premiums paid by employers and employees.[14] Most of these premiums eventually flow back to workers in the form of deferred payments in the states of

[14] Although labor costs are already a broad definition of individual compensation, there are still some types of compensation missing like the use of a lease car, favorable mortgage-loan rates, and so on. We do not have data on these types of fringe benefits.

unemployment, disability or retirement. As long as employers' premiums are a constant fraction of gross wages, using either gross wages or labor costs to calculate the ability distribution only affects the mean of the skill distribution, but not its basic shape. However, in the Netherlands labor costs are not proportional to gross wages, since premiums are collected only over earnings between certain thresholds, where the thresholds differ for the different premiums.

We use the data from the Dutch Income Panel Investigation (IPO in Dutch) from 2002 to determine the earnings distribution in the Netherlands.[15,16] The data are gathered by Statistics Netherlands from individual tax returns. The sample consists of 175,876 individuals in 2002. We consider individuals aged 23 until 65. We ignore individuals that are in school or studying, because their earnings are not a good indicator of their earning ability. We also exclude all individuals with a non-positive gross labor income, because we cannot determine their earnings capacity. Our final data set consists of a sample of 94,859 individuals.

Figure 2.1 gives a Gaussian kernel density estimate of the income distribution up to 200,000 euro (99% of the sample in 2002). The solid line gives labor costs and the dashed line gross wages. Mean labor costs are (approximately) 35,000 euro, and the median is 31,000 euro, so the earnings distribution is skewed to the right. The mode is 33,000 euro. Table 2.1 gives the descriptive statistics for the distribution of labor costs and gross wages. Gross wages are lower than labor costs because the latter includes employers' premiums.

2.4.2 Estimation of the Pareto Tail

The kernel estimate is an accurate estimate of the true density of income for most income levels. However, because the sample does not include many observations in the right tail, we make a distributional assumption for this part of the distribution. Like many papers in the optimal-tax literature we assume the right tail to be Pareto distributed, see for example Saez (2001) and Jacquet *et al.* (2010). Also, Clementi and Gallegati (2005a) and Clementi and Gallegati (2005b) find evidence of a Pareto distributed right tail in Germany, Italy, the US and the UK. Below we will demonstrate that the Pareto distribution fits the top-income data extremely well.

Different sources of income are taxed under separate regimes in the Netherlands. This separate tax treatment could bias our estimates for the Pareto tail. Therefore, we explore

[15] IPO data have previously been used by Atkinson and Salverda (2005) to determine the top income share (up to 1999).

[16] We use IPO 2002, because it has been checked by several researchers, and has been cleaned of various mistakes.

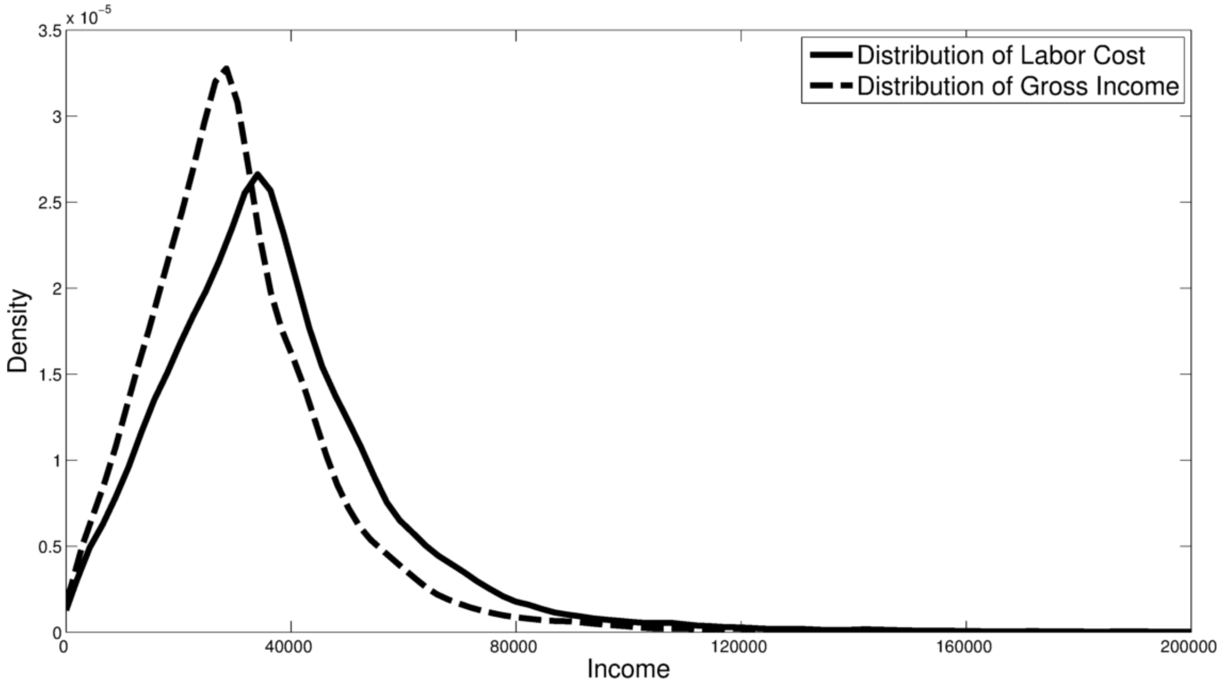


Figure 2.1: A Kernel Density Estimate of Income in the Netherlands in 2002

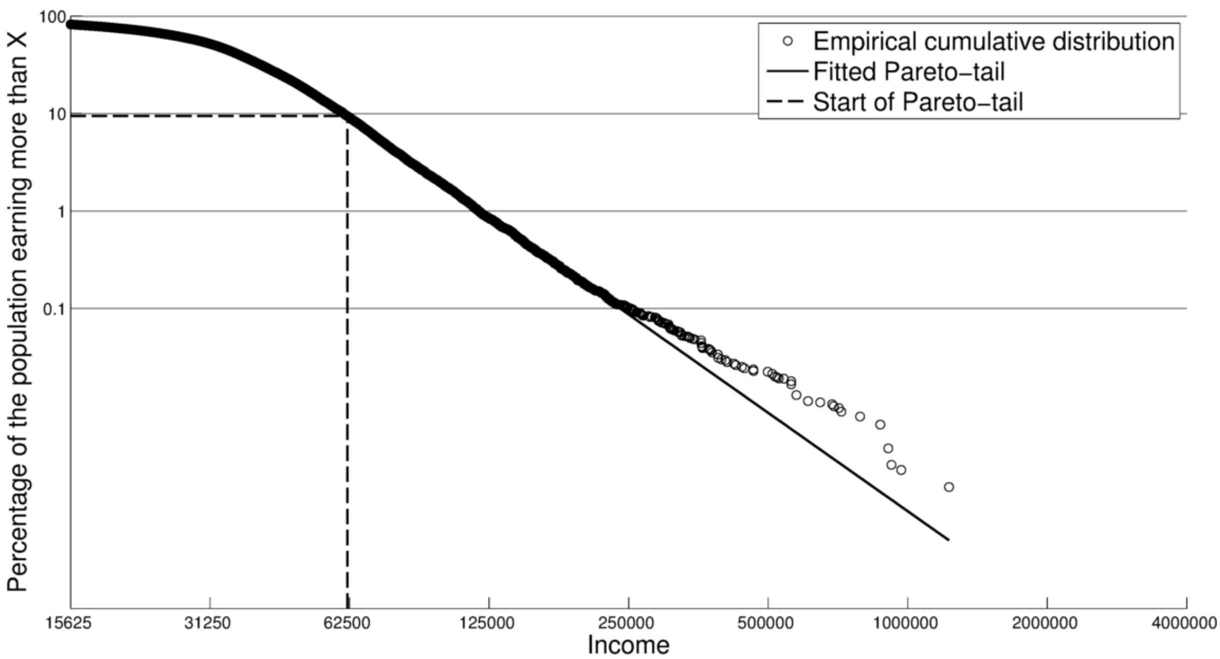


Figure 2.2: The Fit of the Pareto Tail

various income definitions in our estimations. Labor income of workers, fictitiuous labor income of self-employed and fictitiuous labor income of director-shareholders of closely-held firms are all taxed under the progressive labor income tax ('Box-1'). Capital income of director-shareholders of closely-held companies in excess of fictitiuous labor income, retained profits, dividends and capital gains on shares which form a dominant holding are taxed at a 25% rate ('Box-2').[17] Director-shareholders of closely-held companies will therefore allocate income over Box-1 and Box-2 to minimize their tax burden. As a result, part of Box-2 income might be considered as income from labor effort. Table 2.1 also gives descriptive statistics for the sum of Box-1 and Box-2 income. In our main analysis we will focus on labor income taxation (in Box-1).

The cumulative distribution function of the Pareto distribution is given by:

$$F(z) = 1 - \left(\frac{\hat{z}}{z}\right)^{\alpha}, \tag{2.29}$$

where α is the Pareto parameter, z is gross income, and $\hat{z}$ denotes the cut-off level after which the Pareto distribution applies. We estimate the parameters of the Pareto tail using the method developed by Clauset *et al.* (2009). In particular, for a given $\hat{z}$, we choose α such that it maximizes the likelihood function. Subsequently, we choose $\hat{z}$ by minimizing the Kolmogorov-Smirnov (KS) statistic. The KS-statistic measures the maximum distance between the estimated and the empirical cumulative distribution function.[18]

Table 2.1: Summary Statistics for the Income Distribution in 2002

Income definition	Mean	Median	Standard deviation
Labor costs	34,487	31,253	26,350
Gross wages	28,691	25,892	22,202
Box-1 and Box-2 income[a]	34,774	31,309	27,286

[a]See the main text for the definition of Box-1 and Box-2 income.

Table 2.2 provides the estimation results for the Netherlands using IPO 2002. We report the estimates of the Pareto parameter with the corresponding asymptotic standard errors and bootstrapped 95% confidence intervals, the estimates for the starting point of the Pareto distribution with the corresponding bootstrapped 95% confidence intervals, and the R^2-measures of fit. For our preferred definition of income, labor costs, we find a Pareto parameter of 3.35 with a 95% confidence interval of $[3.24, 3.44]$. The starting point is estimated at approximately 62,000 euro with a 95% confidence interval of approximately

[17]See Bovenberg and Cnossen (2001) for an overview of the system of tax boxes in the Netherlands.

[18]Using simulated data, Clauset *et al.* (2009) show that this method outperforms other methods proposed in the literature.

Table 2.2: Estimates of the Pareto Parameter α and Starting Point $\hat{z}$ in 2002

Income definition	α	SE[a]	CI[b]	$\hat{z}$	CI[b]	R^2
Labor cost	3.35	0.037	[3.24,3.44]	61,793	[56,624,77,182]	.995
Gross wages	3.22	0.029	[3.15,3.32]	45,040	[39,565,61,388]	.997
Box-1 and Box-2 income[c]	3.18	0.029	[3.08,3.26]	55,448	[48,115,77,238]	.997
Labor cost 2006	3.30	0.061	[3.17,3.48]	61,943	[56,926,73814]	.991

[a]Asymptotic standard errors.
[b]Bootstrapped 95% confidence interval.
[c]See the main text for the definition of Box-1 and Box-2 income.

[57; 77]. The fit of the Pareto tail is extremely good, with an R^2 of .995. Since $1-F(z) = (\hat{z}/z)^{\alpha}$, we can write for the earnings distribution $\ln(1-F(z)) = \alpha \ln \hat{z} - \alpha \ln z$. Therefore, a plot on a log-log scale with $1 - F(z)$ on the vertical axis and labor earnings on the horizontal axis should be a straight line with a slope of $-\alpha$ if the tail is Pareto distributed. Figure 2.2 shows this plot for our sample. The dots are one minus the empirical cumulative distribution function for each earnings level. The dashed line is the estimated Pareto tail. As can be seen, the relationship between the logarithm of z and the logarithm of one minus the empirical cumulative distribution function is indeed extremely close to being linear.

We investigate whether ignoring Box-2 income leads us to overestimate the Pareto parameter. According to IPO data, only .39% of individuals has income in Box-2.[19] Table 2.1 shows that the mean of the sum of Box-1 and Box-2 income is still higher than mean labor costs, even though it ignores employers' premiums, and the same is true for the median. Nevertheless, the point estimates of the Pareto parameter using gross wages and the sum of Box-1 and Box-2 income are very close, although somewhat lower, at 3.22 and 3.18, respectively. The starting points are estimated to be lower, since mean labor costs are 20% higher than mean gross wages. For gross wages the point estimate of the starting point of the Pareto tail (approximately 45,000 euro) is very close to the starting point of the top income tax bracket (approximately 48,000 euro) in 2002 (more details on the parameters of the Dutch tax system are provided later). Therefore, it is lonely at the top in the Netherlands.

Our empirical findings are very close to the estimates from Atkinson and Salverda (2005), although the latter are based on aggregate household income, including capital income, whereas we use individual, labor income. Atkinson *et al.* (2011) report Pareto-parameter estimates for 20 countries. Like Atkinson and Salverda (2005), these estimates

[19]The data on income in Box-2 are right censored at 250,000 euro. 62 of the 489 individuals (13%) with Box-2 income in IPO are right censored. This may further lead us to somewhat underestimate the thickness of the right tail of the income distribution.

are all based on aggregate tax statistics.[20] Clearly, Pareto parameters vary much across countries. Notably, the Netherlands features the highest Pareto parameter of all studies covered in this study. In a separate study for Denmark, Kleven and Kreiner (2006) report the Danish Pareto parameter to be 3.5, which is the highest estimate in the world that we are aware of.

2.4.3 Marginal tax rates

Marginal tax rates measure the difference between the increase in gross and net income when an individual's gross income increases by a small, marginal amount. However, determining the additional amount of net income after an increase in earnings is a complex task.

First, the effective marginal tax rate depends on a large number of income-dependent tax credits and subsidies, many of which in turn depend on household characteristics like the composition of the household and the age of the children. We use the sophisticated tax-benefit calculator MIMOSI of CPB to calculate the effective marginal tax rates.[21] MIMOSI contains all the relevant details of the tax and benefit system in the Netherlands.

Second, one needs to include indirect taxes. Indirect taxes distort the choice over income and leisure just like direct taxes do. According to Statistics Netherlands (2013) (net) indirect taxes were 11.7% of total consumption in 2002. Bettendorf *et al.* (2012) demonstrate that indirect taxes are very close to proportional in total consumption in the Netherlands. Hence, we assume indirect taxes are proportional to consumption in the model.[22]

Third, we need to determine the net income component of premiums. Most studies treat all premiums as taxes, *e.g.* Saez (2001), Gruber and Saez (2002), Brewer *et al.* (2010), and in their study of marginal tax rates in the Netherlands Gielen *et al.* (2009). However, this is not correct for the Netherlands. Individual benefits (unemployment, disability, pension) are linked to individual contributions made by either employees or their employers. Hence, not all premiums are taxes, and one needs to treat premiums for unemployment, disability and pensions as deferred wage income. However, determining the marginal tax rate on deferred wage incomes is rather complicated. First, replacement

[20]The estimates by Atkinson *et al.* (2011) are based on total income, including not only labor income, but also capital income. Capital gains are excluded from their (and our) income definition. Naturally, the estimates gathered in Atkinson *et al.* (2011) are only as good as the aggregate income tax statistics from which they are computed. These authors provide an extensive discussion of the potential caveats.

[21]See Gielen *et al.* (2009) for a recent analysis of changes in marginal tax rates over the past decade using MIMOSI.

[22]Note that consumption equals disposable income in our static model.

rates during unemployment, disability and retirement (state and occupational pensions combined) are around 70 percent of earned income. Hence, individuals typically experience a substantial drop in income, and hence a drop in the marginal tax rate, when they receive the deferred wage income. Therefore, we assume that deferred wage incomes are taxed in one tax bracket below the current bracket (except for individuals in the first tax bracket). Second, it is hard to determine whether premiums are used to redistribute from high-income to low-income earners, and thereby contribute to the marginal wedge. The unemployment and disability schemes redistribute income from high income to low income workers.[23] However, the pension scheme redistributes income from low-income to high-income individuals.[24] Due to these complexities we decided ignore redistribution in premiums. Third, assets accumulated for (and blocked until) retirement are not subject to wealth taxes. Given sufficient separability in preferences between consumption and labor (as we assume), the exemption for the wealth tax does not directly affect labor supply incentives. Therefore, we also decided to ignore intertemporal considerations in premiums. In the end we then treat premiums for unemployment, disability and pensions as wage income, taking into account that the effective marginal tax rate is typically lower when they receive the income.

A potentially important element missing from our calculations is the tax deductibility of interest on mortgage loans. Specifically, we assume that the mortgage interest payments stay constant when there is a marginal change in income. However, individuals that earn more income are more likely to own a (more expensive) house. Hence, one could argue that the mortgage-rent deduction acts as an indirect subsidy on labor. However, this reasoning assumes a perfectly elastic supply of housing. When housing supply is not perfectly elastic, as in the Netherlands, larger demand for housing translates into higher housing prices,

[23] De Koning *et al.* (2006) calculate the implicit redistribution in unemployment insurance from low income to high income individuals, using panel data for 12 years. They divide the population in three skill groups, all of equal size. Over a period of 12 years, the lowest 33% of the population uses 46% of all benefit days used, whereas the highest 33% uses only 20% of all benefit days used. De Koning *et al.* (2006) also calculate the implicit redistribution in disability insurance from low-income to high-income individuals. Over a period of 12 years, the lowest 33% of the population uses 59% of all benefit days used, whereas the highest 33% uses 17% of all benefit days used.

[24] Bonenkamp (2009) calculates redistribution in occupational pensions in the Netherlands on a lifetime basis. He calculates the present discounted value of pension contributions and pension benefits for four skill groups by gender (and cohort). The inter-educational redistribution for the lowest skill groups is − 17% of premiums for low skilled males and − 13% of premiums for low skilled females. High-skilled males receive a subsidy of 6% on a lifetime basis, and high-skilled females 1% (controlling for 'cross-gender' redistribution).

which reduce the incentive to supply labor.[25] Based on the Dutch situation, we roughly calculate that the effect of housing subsidies on the total tax wedge is relatively minor (roughly 2.5%, see previous footnote). Therefore, we decided to ignore housing subsidies in the calculation of marginal taxes.

Despite the latter limitation, our calculations of marginal tax rates are much more advanced than in any other studies. For example, Saez (2001) and Jacquet *et al.* (2010) assume a linear tax system to recover the ability distribution. However, marginal tax rates are quite nonlinear, as we will demonstrate below, and this could bias the estimation of the ability distribution.

There is substantial variation in marginal tax rates at each income level, see Figure 2.3 where we made a scatter plot of marginal taxes against income. However, the model only works with a single marginal tax rate at each income level, hence we use a kernel estimate to smooth out the differences. Figure 2.4 gives the kernel estimate for effective marginal tax rates in the Dutch income distribution for all workers participating in the labor market. To understand the patterns in Figure 2.4, Table 2.3 provides some parameters of the Dutch tax system in 2002.

Table 2.3: Tax Brackets and Tax Credits in 2002

	Start	End	Percentage	Maximum amount
Tax brackets				
First tax bracket	0	15,331	32.35	4,960
Second tax bracket	15,331	27,847	37.85	4,737
Third tax bracket	27,847	47,745	42.00	8,357
Fourth tax bracket	47,745	∞	52.00	∞
Tax credits				
General tax credit	0	∞	0	1,647
Earned-income tax credit				
- First part	0	7,692	1.73	133
- Second part	7,692	15,375	10.62	949
Single parent tax credit	0	∞	0	1,301
Earned-income single-parent tax credit	0	30,256	4.30	1,301

[25] Denote the elasticity of housing demand by $\varepsilon^d \equiv \frac{\mathrm{d}h^d}{\mathrm{d}p^*}\frac{p^*}{h^d}$, and the elasticity of housing supply by $\varepsilon^s \equiv \frac{\mathrm{d}h^s}{\mathrm{d}p}\frac{p}{h^s}$, where $p^* \equiv (1-s)\,p$ denotes the net housing price, p the gross housing price and s the housing subsidy. Then, standard tax-incidence analysis shows that $\frac{\mathrm{d}p^*}{p^*} = -\frac{\varepsilon_s}{(\varepsilon_s+\varepsilon_d)}\frac{\mathrm{d}s}{s}$. Van Ewijk *et al.* (2006) estimate that $\varepsilon_s = \varepsilon_d = 0.7$, hence the net housing price falls 0.5% when the subsidy increases with 1%. On average, net housing expenditures are about 25% of total income (CPB, 2010b), and this pattern is rather flat over all income groups. The tax advantage in housing is about 20% of total housing costs, so this reduces the tax wedge on labor by at most 2.5% (1/2 × 20% × 25%).

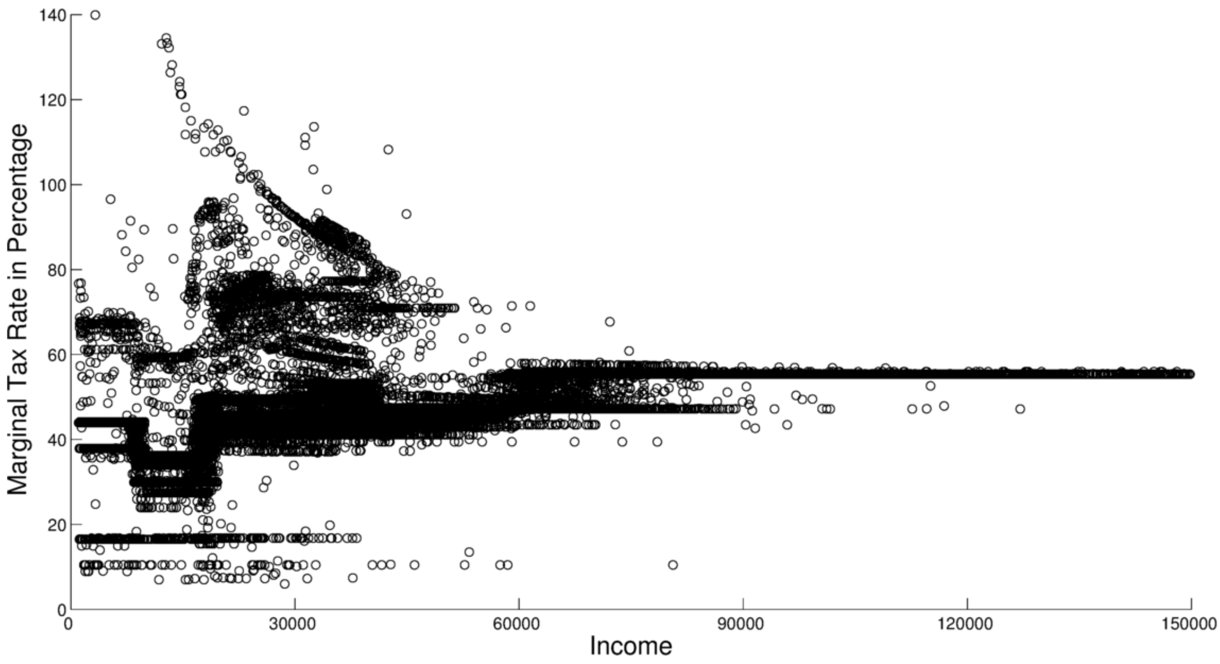


Figure 2.3: Scatterplot of Marginal Tax Rates by Income

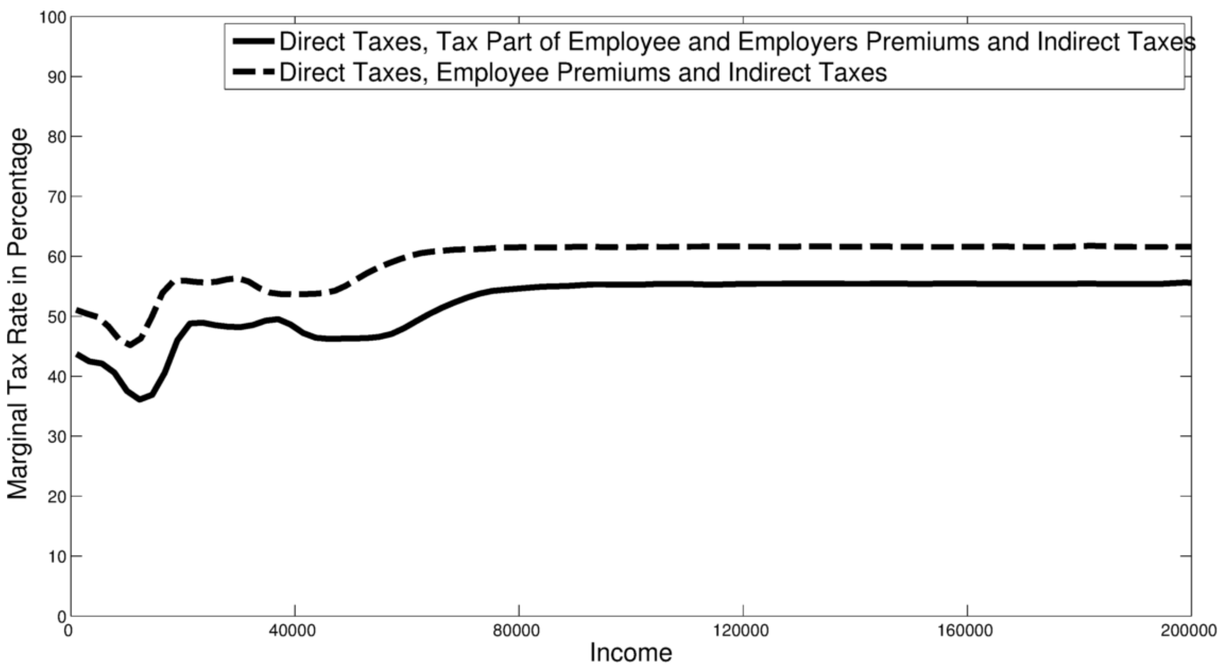


Figure 2.4: Kernel of Marginal Tax Rates by Income

In 2002, the Dutch tax system had four tax brackets for labor income, based on individual (not household) income, with rates rising from somewhat below 33% at the bottom to 52% at the top. This explains why marginal tax rates are typically lower for individuals with low income than for individuals with high income.

But there are also a number of noticeable deviations from these rates, which result mostly from targeted subsidies and tax credits. For the lowest incomes, marginal tax rates are initially significantly higher than the first tax bracket because a number of income-support schemes are phased out with income, in particular rent subsidies and a general child tax credit.[26] Marginal tax rates are much lower in the segment where the earned-income tax credits (EITCs) are phased in (see Table 2.3). The end of the phase-in range for the EITCs (almost) coincides with the start of the second tax bracket, and marginal tax rates jump up by some 15%-points between 15,000 and 20,000 euro.

Another noticeable jump can be observed for individuals with a gross income close to 40,000 euro. Individuals below this income threshold are eligible for the public health-insurance scheme with relatively low contribution rates, whereas individuals above this threshold are required to have private health-care insurance with relatively high premiums. For some households close to the threshold this results in very high marginal tax rates.[27]

2.4.4 Government Revenue Requirement and Benefit Level

We assume that the government has to collect 9.5% of total output to finance government consumption net of income redistribution. This is the sum of expenditures on public administration, police, justice, defense and infrastructure minus non-tax revenues (for example from the sales of natural gas) as a percentage of GDP in 2002 (CPB, 2010a, Annex 9).[28] This is in the same order of magnitude as Tuomala (2010), who assumes (maximum) government consumption of 10% of GDP. With the revenue requirement set at 10% of total labor income, the tax system is budgetary neutral with a benefit level of approximately 12,000 euro. This is somewhat higher than the current level of net welfare benefits in 2002 amounting to 9,014 euro for a single-person household. However, we ignored some other forms of social assistance at the local level ('Bijzondere Bijstand'), exemptions from local taxes, and transfers in kind (discounts for arts, public transport,

[26] The exact subsidy levels and taper rates vary with household characteristics other than income, and are therefore not reported in Table 2.3.

[27] In 2003 this health-care system has been replaced by a uniform, obligatory basic health-insurance scheme, which is financed by a payroll tax and 'lump-sum' premiums paid by individuals. Individuals can voluntarily top up the basic health-insurance scheme with supplementary insurance packages.

[28] We have experimented with different levels of the government's revenue requirement, but we do not find that this induces significant changes in the optimal marginal tax schedules.

etc.), training, public employment, and labor-market programs, which also act as support schemes for the non-employed.

2.4.5 Elasticity of Income with Respect to Marginal Taxes

An important determinant of optimal income tax rates is the elasticity of the tax base. We use recent Dutch estimates of the participation elasticity and the elasticity of taxable labor income to calibrate the extensive and intensive margin responses of the tax base in the model. We discuss these estimates and the calibration method below.

Elasticity of Labor Supply

Traditionally, economists have analyzed at the impact of taxes on labor supply to measure the distortions from income taxation. Table 2.4 gives an overview of recent estimates of labor-supply elasticities in the Netherlands. Mastrogiacomo *et al.* (2013) estimate a structural discrete-choice model for a number of subgroups using data for the period 1999-2005. Our calibration year (2002) is in the middle of this sample. These authors present estimates for the uncompensated wage elasticity of total hours worked, the participation rate and hours per worker.

Table 2.4 reveals that the total uncompensated labor-supply elasticity of men in couples is rather small. Elasticities are larger for women in couples, in particular when small children are present. Single parents have the highest labor-supply elasticities, and elasticities of singles are in between single parents and individuals in couples. Looking at the decomposition of these elasticities into participation (extensive) and hours per worker (intensive) responses, we find that most of the response is on the participation margin.[29] These findings are in line with the findings of related empirical studies for the Netherlands, see Mastrogiacomo *et al.* (2013, Table 15).

The ranking of the elasticities by household types and the extensive versus the intensive margin are in line with the findings of empirical studies abroad. Once more, see Mastrogiacomo *et al.* (2013, Section 5) and the excellent overview in Bargain *et al.* (forthcoming). Weighting the elasticities of Mastrogiacomo *et al.* (2013) for the different

[29] Chetty (2012) has recently argued that optimization frictions may mask part of responses on the intensive margin. Furthermore, the elasticities reported in Mastrogiacomo *et al.* (2013) are so-called unconditional intensive-margin elasticities. These are simulated by increasing gross wages by 10%, which is common in the structural discrete-choice literature on labor supply. Simulated in this way, the elasticities capture both the response in hours by those already working and a composition effect because new entrants may work different hours than those already working. The conditional intensive-margin elasticity of those already working is actually higher than the unconditional intensive-margin elasticity, though still smaller than the extensive-margin elasticity.

groups by their respective sizes on the Dutch labor market, we obtain an average total-hours elasticity of 0.30, an average participation elasticity of 0.25, and an hours-per-worker elasticity of 0.06.

Table 2.4: Estimates of the Uncompensated Labor-Supply Elasticity

Group	Obs.	Total hours		Participation		Hours per worker	
		Men	Women	Men	Women	Men	Women
Couples with children	72,000	0.14***	0.50***	0.14***	0.38***	0.01	0.12***
		(0.01)	(0.03)	(0.01)	(0.03)	(0.01)	(0.02)
Couples w/o children	72,000	0.07***	0.27***	0.07***	0.22***	0.00	0.05***
		(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
Singles	24,000	0.39***	0.47***	0.33***	0.39***	0.06***	0.08***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Single parents	24,000	0.45***	0.62***	0.32***	0.43***	0.12***	0.18***
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Source: Mastrogiacomo *et al.* (2013, Table 15, Table A.12). Bootstrapped standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1.

Table 2.5: Estimates of the Uncompensated Elasticity of Taxable Labor Income

Group	>10,000	50,000–100,000	>50,000
All workers	0.24***	0.26***	0.46***
	(0.01)	(0.04)	(0.07)
Observations	157,510	11,346	12,196
Singles	0.36***	0.36*	0.58*
	(0.04)	(0.21)	(0.30)
Observations	18,061	507	530

Source: Jongen and Stoel (2013a, Table 4, Table 5) and additional estimates for singles using the same data set (details available on request).
*** p<0.01, ** p<0.05, * p<0.1.

Unfortunately, Mastrogiacomo *et al.* (2013) do not consider income elasticities, as data on unearned income are lacking. However, some other recent studies use data on unearned income to estimate the unearned-income elasticity for the Netherlands. We convert these unearned-income elasticities into income elasticities using the average share of unearned income in total income in the descriptive statistics. For Vermeulen (2005) we then obtain an income elasticity ranging from -0.01 and -0.02 for single men and women to -0.10 and -0.14 for men and women in couples, resepctively. For Bloemen (2009) we find an income elasticity of -0.10 to -0.18 for men in couples and -0.12 to +0.10 for women in couples. Finally, for Bloemen (2010) we find an income elasticity of -0.22 and -0.16 for men and women in couples, respectively. The average earned income elasticity for men is –0.12, and for women, ignoring the positive value for married women in Bloemen (2009), is –0.11.

Elasticity of Taxable Income

Following the seminal work by Feldstein (1995) recent empirical studies have looked at the elasticity of taxable income (ETI). The ETI may capture a wider range of behavioral responses to income taxes, such as changes in work effort, occupational choice, human capital investment, tax avoidance, tax evasion, and migration. ETI studies typically focus on the employed, hence ETI-estimates are conceptually close to the intensive-margin elasticities in the labor-supply literature.

Table 2.5 presents recent estimates of the elasticity of taxable labor income in the Netherlands from Jongen and Stoel (2013a). They use data for the period 1999-2005 and exploit the 2001 tax reform to estimate the elasticity of taxable labor income. In their base specification, the estimated ETI is 0.24 for all workers. This baseline only employs workers with taxable labor income above 10,000 euro to circumvent problems with strong mean reversion in incomes at the bottom of the income distribution (Gruber and Saez, 2002). The estimated ETI for higher incomes (ranging from 50,000–100,000 euro) is slightly larger than for all workers (>10,000 euro), though the difference between the estimates is not statistically significant. Including also the highest income earners (>50,000 euro) leads to a much larger ETI-estimate, but this estimate is rather imprecise due to a relatively small number of observations.[30] These ETI-estimates are in line findings abroad. In their overview paper Saez *et al.* (2012) suggest a range of 0.12 to 0.40.

For singles, Jongen and Stoel (2013a) find a somewhat larger ETI. Again, the ETI for incomes between 50,000-100,000 euro is is very close to the ETI for all workers, and the ETI for the highest incomes (>50,000 euro) is much larger, though imprecisely estimated.

Jongen and Stoel (2013a) also lack information on unearned income and therefore do not estimate income elasticities ('income effects') for the elasticity of taxable labor income. In their overview Saez *et al.* (2012) suggest that income elasticities are rather small. The income elasticity for taxable income ranges from essentially zero (Kleven and Schultz, forthcoming) to -.14 in Gruber and Saez (2002).

[30] Furthermore, the estimated ETI for incomes above 50,000 euro is sensititve to the controls for exogenous income growth whereas the estimated ETI for all workers and for 50,000-100,000 euro is rather stable for different controls for exogenous income growth, see Jongen and Stoel (2013a),.

Elasticities in Simulations

In our simulations we consider a baseline scenario, and two alternative scenarios, based on different estimates for the elasticities. The key assumptions in these scenario's are summarized in Table 2.6. In the baseline case we assume a compensated wage elasticity of earnings supply equal to .35, an income elasticity of .10, and hence an uncompensated wage elasticity of earnings supply equal to .25 based on the findings in Table 2.5 on recent ETI-studies for the Netherlands. Furthermore, we assume a participation elasticity of .25, also based on recent evidence for the Netherlands, see Table 2.4. We also consider a robustness scenario with 50% higher elasticities, i.e. a compensated wage elasticity of .53, an income elasticity of .15, an uncompensated wage elasticity of .38, and a participation elasticity of .38. And, for completeness, we also consider the opposite case with 50% lower elasticities: .18 for the compensated wage elasticity, .05 for the income elasticity, .13 for the uncompensated wage elasticity and .13 for the participation elasticity. These robustness checks facilitate our comparisons with Saez (2001), Brewer *et al.* (2010) and Jacquet *et al.* (2010).

Table 2.6: Elasticities Used in the Simulation

	Compensated wage elasticity	Income elasticity	Uncompensated wage elasticity	Participation elasticity
Baseline scenario	0.35	0.10	0.25	0.25
Low-elasticity scenario	0.18	0.05	0.13	0.13
High-elasticity scenario	0.53	0.15	0.38	0.38

Table 2.7: Employment Rates for Different Education Levels

Level of Education	Net Employment Rate	Share in Population
Only elementary school	36.90	11.99
Some high school	53.50	25.79
High school	56.80	10.26
Low-level college	71.20	15.84
Mid-level college	79.10	14.95
Bachelor degree	80.40	13.88
Master degree or higher	84.40	7.28

2.4.6 Utility and Welfare Functions

We consider optimal tax rates for the two polar, Benthamite (utilitarian) and Rawlsian (maxi-min) cases of the social welfare function:

$$\begin{aligned} \text{Bentham} &: \int_{\mathcal{N}} W(U_n)\mathrm{d}F(n) = \int_{\mathcal{N}} U_n \mathrm{d}F(n), \\ \text{Rawls} &: \int_{\mathcal{N}} W(U_n)\mathrm{d}F(n) = U_{\underline{n}}. \end{aligned} \tag{2.30}$$

Recall that $U_n \equiv u_n - \varphi$, which equals u_n when only the intensive labor-supply margin is included.

We assume a functional form for the utility function, which encompasses most of the utility functions encountered in the literature:

$$u_n = \frac{c_n^{1-\alpha}}{1-\alpha} - \gamma \frac{l_n^{1+1/\varepsilon}}{1+1/\varepsilon}, \qquad \alpha, \gamma, \varepsilon > 0. \tag{2.31}$$

Our utility function allows for income effects and is also used by Mankiw *et al.* (2009). When $\alpha = \frac{1}{\varepsilon}$ this specification is in line with the CES functions used by Mirrlees (1971) and Tuomala (1984). α and ε are calibrated so as to match the compensated and uncompensated elasticity of the scenarios described in Table 2.6, which is in the spirit of Chetty (2006).[31] Parameter γ is an innocuous scaling parameter, which hardly affects the resulting optimal tax rates. We adjust it to keep the mean of the ability distribution fixed in the different scenarios.

Table 2.8: Calibrated Parameters for the Utility Function

Parameter values	Base	Low Elasticity	High Elasticity
α	0.46	0.48	0.45
ε	0.38	0.18	0.60
γ	1982	13503	1082
μ_k	55.95	0.00	82.42
σ_k	271.27	511.00	189.98

Table 2.8 displays the values of the parameters for the utility function. As can be seen, parameter α is almost constant in all scenarios. Hence, the elasticity of the marginal utility of income is the same across the simulations. Therefore, the difference in optimal

[31] As long as the ratio $\varepsilon^c/\varepsilon^u$ is fixed, the calibrated α is almost the same for different elasticities. This is a useful property, since then we can isolate the effect of a change in the elasticities without changing the redistributional concerns. All our scenarios therefore have the same ratio $\varepsilon^c/\varepsilon^u$.

tax rates in the scenarios should be attributed to the differences in the elasticities, and not a changed preference for redistribution via the curvature of the private utility function.

2.4.7 Determination of the Ability Distribution

The determination of the distributions of ability and participation costs/benefits is not straightforward, since they are not directly observable. We assume that the data on earnings and participation are a choice process that follows from our assumed utility function and the distribution of participation costs/benefits. Given observed choices for earnings and labor force participation, and assuming separability between the leisure and consumption component of utility and participation costs/benefits, we are able to identify the distributions of skills.

In particular, conditional upon participation in the labor market, and using the definition of gross labor earnings $z_n \equiv nl_n$, we can invert the first order condition for optimal labor supply (2.4) to express ability n as a function of marginal tax rates and income. The solution for ability is:

$$n = \left(\frac{\gamma z_n^{1/\varepsilon}}{(1 - T'(z_n))c_n^{-\alpha}} \right)^{\frac{\varepsilon}{\varepsilon+1}}. \tag{2.32}$$

Using information on gross earnings z_n, consumption c_n, and marginal tax rates $T'(z_n)$ we are able to compute ability of each working individual in the data-set. Note that the consumption level follows from the difference between gross earnings and total taxes paid: $c_n = z_n - T(z_n)$.

2.4.8 Calibration of the Distribution of Participation Costs and Benefits

We estimate the distribution of participation costs using information on the employment rate and the participation elasticity. Ideally, we would like to have data on the employment rate for each level of ability n, but no such data are available. However, we do have data on employment rates by 7 levels of education. These data are given in Table 2.7. By assuming that the cumulative distribution of education corresponds to the cumulative distribution of ability, the education-specific employment rates allow us to estimate the distribution of participation costs by skill type. Since we assume that everyone has the same utility function, we also assume that the distribution of the disutility of participation is independent of ability, i.e. $k(\varphi|n) = k(\varphi)$. Under this assumption, the theoretical model predicts an ability-specific participation rate $\hat{E}(n_1, n_2)$ for all individuals between skill levels n_1 and n_2 for any pair $\{n_1, n_2\}$, with $n_2 > n_1$ equal to:

$$\hat{E}(n_1, n_2) = \frac{\tilde{K}(n_2) - \tilde{K}(n_1)}{F(n_2) - F(n_1)} = \frac{\int_{n_1}^{n_2} \tilde{k}(m)\,\mathrm{d}m}{F(n_2) - F(n_1)} = \frac{\int_{n_1}^{n_2} K(u_m - v(b))\, f(m)\,\mathrm{d}m}{F(n_2) - F(n_1)}. \quad (2.33)$$

In addition, the elasticity of participation with respect to the gross wage rate has recently been estimated, as discussed in section 4.5.1. Our model can also be used to predict the value of this participation elasticity.

Note that labor-force participation at ability level n is given by $K(u_n - v(b))$. In addition, the gross wage rate is equal to n. Therefore, the participation elasticity ε_n^P at skill level n with respect to the gross wage rate is given by:

$$\varepsilon_n^P = \frac{\partial K(u_n - v(b))}{\partial n} \frac{n}{K(u_n - v(b))} = \frac{\partial u_n}{\partial n} \frac{nk(u_n - v(b))}{K(u_n - v(b))}, \quad (2.34)$$

where the final step follows from the envelope theorem. Hence, the predicted average participation elasticity in the economy is given by:

$$\hat{\varepsilon}^P = \int_{\underline{n}}^{\overline{n}} \varepsilon_m^P f(m)\mathrm{d}m = \int_{\underline{n}}^{\overline{n}} \frac{\partial u_m}{\partial m} \frac{mk(u_m - v(b))}{K(u_m - v(b))} f(m)\mathrm{d}m. \quad (2.35)$$

In equations (2.33) and (2.35) u_n, $\frac{\partial u_n}{\partial n}$, $v(b)$, $F(n)$ and $f(n)$ can be inferred from the data. Furthermore, we assume that $k(\varphi)$, is normally distributed with mean μ_k and standard deviation σ_k: $\varphi \sim N(\mu_k, \sigma_k^2)$. Finally, assume that data exist on both $E(n_1, n_2)$ and $\hat{\varepsilon}^P$. In that case, we can write the error term between the model-predicted employment rate and the true employment rate, and the model-predicted participation elasticity and the true elasticty by:

$$\epsilon_{n_2} = \hat{E}(n_1, n_2) - E(n_1, n_2), \quad (2.36)$$

$$\epsilon_p = \hat{\varepsilon}^P - \varepsilon^P. \quad (2.37)$$

We choose parameters μ_k and σ_k such that the absolute value of the weighted error terms is minimized using non-linear least squares. The weighting procedure in estimating the distribution of participation costs is simple. Table 2.7 provides observations of skill-specific employment rates for seven different education levels. We only have one estimate of the participation elasticity. Hence, we give each error-term for the skill-specific employment rate a weight equal to 1, and we give the error term for the participation elasticity a weight equal to 7.

We do need to take into account a selection bias in our estimation of the participation-cost distribution. In the data we only observe the density of ability *conditional on employment* $f(n|I)$, where I is an indicator variable equal to 1 when an individual is employed, and zero otherwise. However, our model predicts a specific relationship between skill and employment; better skilled individuals are more likely to participate, because they receive a higher wage. Since we are interested in the unconditional skill density $f(n)$, we need to reweigh the skill distribution taking this selection bias into account. This correction is very similar to the procedure described in Heckman (1979).[32]

Bayes' Law provides the relationship between the two densities:

$$f(n) = \frac{f(n|I)\,p(I)}{p(I|n)}. \tag{2.38}$$

The unconditional probability of employment $p(I)$ equals the total employment rate in the population $E(\underline{n}, \overline{n})$, which can be derived from Table 2.7. The conditional probability of employment conditional on ability then equals the cut-off level of the disutility of participation below which all individuals with ability n work: $p(I|n) = K(u_n - v(b)|n)$. Using these results, we can recalibrate the skill distribution to get a correct measure of the skill distribution for the entire population:

$$f(n) = \frac{f(n|I)\,E(\underline{n}, \overline{n})}{K(u_n - v(b)|n)}. \tag{2.39}$$

Note that we can only adjust for the estimation bias if we have the unconditional distribution $k(\varphi)$. However, we can only find the distribution $k(\varphi)$ through non-linear least squares if we have derived $f(n)$, which needed to be determined in the first place. We resolve this indeterminacy as follows. First, we guess that the unconditional distribution of ability $f(n)$ is equal to the conditional distribution of ability $f(n|I)$. Based on this initial guess, we estimate the parameters of the distribution $k(\varphi)$. After retrieving the distribution $k(\varphi)$ we can update our initial estimate for $f(n)$. With the updated estimate for $f(n)$ we can again re-estimate the parameters of $k(\varphi)$, etc. We exit the updating procedure when the parameters of $k(\varphi)$ converge. Table 2.8 provides the estimated values of γ, μ_k and σ_k, which are hard to interpret, as they depend on the cardinal properties of the utility function.

[32]If we would not correct for this selection bias, our estimates would underestimate the number of individuals at the low end of the earnings distribution with 60% and overestimate the number of individuals at the top end with 30%.

2.5 Optimal Tax Schedule

Having discussed the calibration of the model, we now turn to the optimal tax profiles for different elasticities and different social welfare functions. We first consider the optimal tax schedules when individuals can only respond on the intensive margin, and subsequently consider the optimal tax schedules when individuals can respond both on the intensive and the extensive margin. However, we start with a discussion of the optimal top rate, which is virtually identical in both cases, as non-participation is basically not relevant for individuals with a top income.

2.5.1 Top Rate

For both social welfare functions, the marginal tax rate converges to a constant at the top. Table 2.9 reports the resulting optimal top rates for the different assumptions about the elasticities and the social welfare function.

Table 2.9: Optimal Effective Marginal Top Rates

Uncompensated/compensated elasticity	.13/.18	.25/.35	.38/.53
Rawlsian	65%	56%	49%
utilitarian	60%	48%	40%

Source: Figures taken from simulations, see later in paper.

The Rawlsian (maxi-min) government aims to maximize tax revenue from the top income earners, it wants to 'soak the rich' by setting their tax rate at the top of the Laffer curve. Table 2.9 shows that the current effective top tax rate of 55.4% (which includes indirect taxes) is virtually identical to the baseline value (55.6%). Increasing the current top rate from 52% to 60% (excluding indirect taxes) – as some political parties have suggested – thus results in revenue losses. Higher top rates then result in both less income redistribution and larger deadweight losses. Consequently, both equity and efficiency are reduced. Only at a low elasticity of the tax base, a higher top rate could be optimal.

A Benthamite (utilitarian) social welfare function attaches a positive welfare weight to high-income earners; the euro extracted from the top-income earners results in a utility loss, which is valued by the government. The monetized valued of this utility loss needs to deducted from tax revenues to determine the optimal top rate, see the theory section. In this case, the optimal effective marginal top rate is 48% in the baseline. Hence, the current effective top rate of 55.4% would be set too high. Again, only with very low elasticities a higher top rate would be optimal under utilitarian social preferences.

2.5.2 Intensive Margin

Next, we consider the entire profile of optimal marginal tax rates. Figures 2.5 and 2.6 show the optimal linear and non-linear tax schedule under Rawlsian and utilitarian preferences, respectively. For comparison the graphs also show the actual tax schedule.

For the Rawlsian social welfare function, we find that optimal marginal tax rates are generally decreasing. After modal income there is a very tiny increase in marginal tax rates. This contrasts with Saez (2001) and Brewer *et al.* (2010) who find an inverse U-shape for a Rawlsian social welfare function. The intuition is that the skill distribution behaves differently in the Netherlands in comparison with the US or the UK.

Note that with the Rawlsian social welfare function, the B-term of the optimal tax formula is unity, see Section 3. The A-term does not play a role, since the elasticity is constant across the entire earnings distribution. Hence, all the changes in optimal taxes after modal income are generated by the C-term, which is determined by the earnings distribution. The top tail of the earnings distribution in the Netherlands is much thinner than in both the US and the UK. Hence, setting higher marginal tax rates produces relatively small distributional benefits compared to deadweight losses. Indeed, the plot for the Rawlsian optimal tax schedule implies that the C-term becomes roughly constant after modal income. The U-shape in the optimal marginal tax rates under the utilitarian government is almost entirely driven by the B-term, as the C-term remains relatively constant and the A-term does not play a role. Since average distributional benefits are always rising with income, B rises with income, hence marginal tax rates increase after modal income (Diamond, 1998). These findings are in line with Saez (2001) and Brewer *et al.* (2010).

Optimal utilitarian tax rates are set below the current tax rate everywhere, except at the bottom of the income distribution. Even with utilitarian social preferences marginal tax rates at the bottom are too low. Higher marginal tax rates at the bottom allow the government to redistribute more income from lower-middle incomes to the lowest incomes. The lower marginal tax rates everywhere else indicate that the utilitarian government is less interested in income redistribution amoung all other groups in comparison to the current system. Hence, apparently social preferences implicit in the actual system are not very utilitarian (we consider the social welfare weights implicit in the actual system more closely below).

Under Rawlsian social preferences optimal tax rates are typically higher than in the current tax schedule, except at the top. We see that the difference between actual and optimal tax rates is largest at the bottom of the income distribution, where the optimal tax rate is close to 100%. The efficiency loss of the high marginal tax rate at the bottom

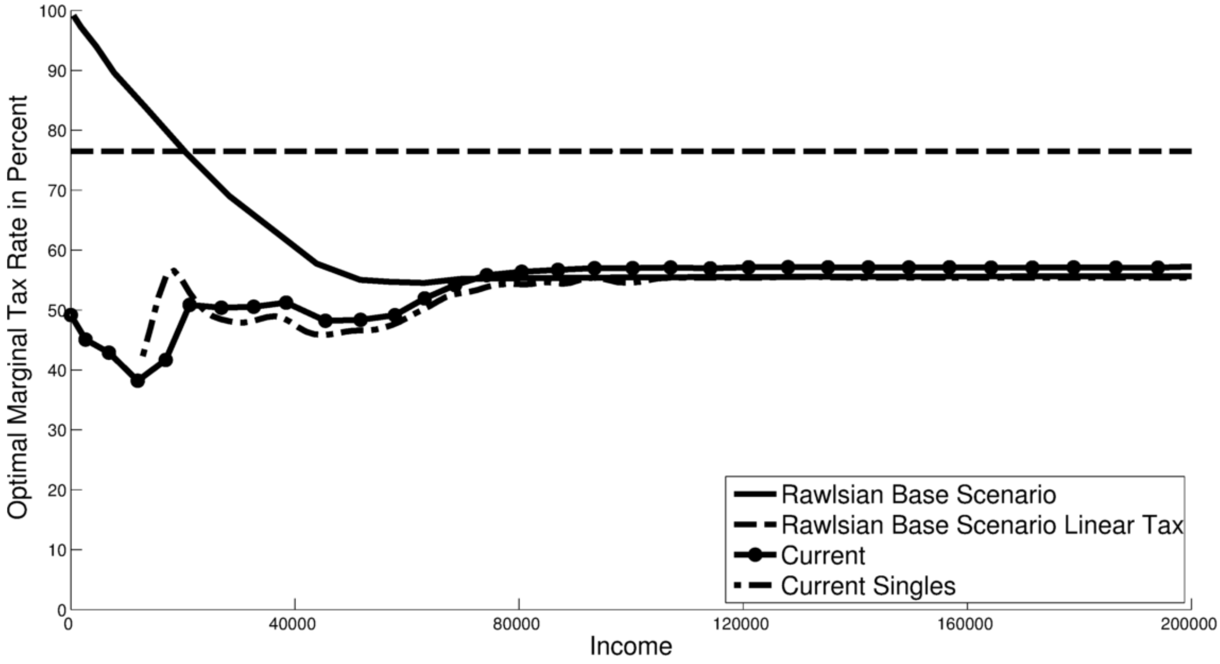


Figure 2.5: The Optimal Tax Schedule with Rawlsian (Maxi-Min) Social Preferences

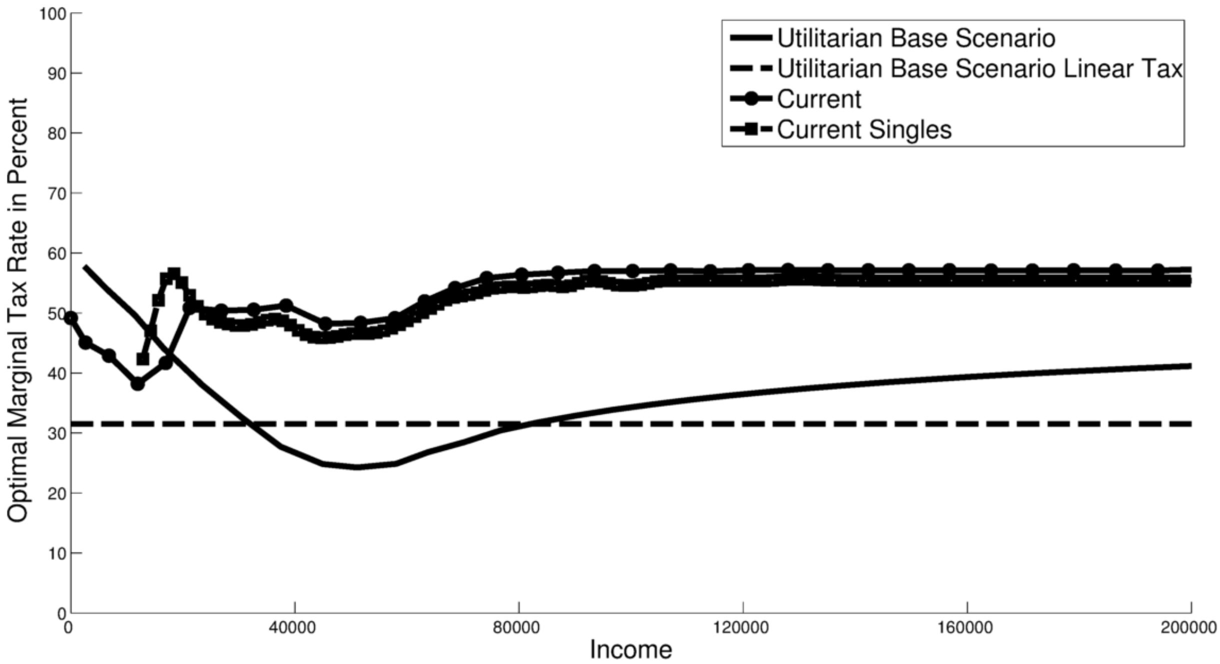


Figure 2.6: The Optimal Tax Schedule for Benthamite (Utilitarian) Social Preferences

is small, because it only affects a small group of individuals. Hence, the optimal tax rate at the bottom is an efficient way to redistribute income from the rich and middle income groups to the poor. Below we will see if this conclusion remains valid once we allow for an extensive margin decision. A Rawlsian government sets marginal tax rates that are generally declining for the earnings distribution. In addition, it basically 'soaks' all the middle and high incomes to maximize government revenue, so as to give the highest feasible transfer to the poorest people in society. Indeed, such a transfer can only be financed if it is phased out through very high marginal tax rates.

A comparison with Saez (2001) and Brewer *et al.* (2010) further shows that optimal marginal tax rates are generally lower in the Netherlands than in the UK and the US. Abilities are more equally distributed in the Netherlands than in the US and the UK so that the gains of redistribution are typically lower. Distortions of income redistribution are similar, since elasticities of labor supply are comparable. Hence, and optimal taxes are lower in the Netherlands.

We are not only interested in marginal tax rates but also in total taxes. Who gains and who loses under each tax-benefit system? Figure 2.5.2 illustrates the allocations that follow from the optimal non-linear tax schedules and the current tax schedule. In addition, it shows the laissez-faire allocation in which gross income equals net income. The intercept is the transfer the government provides to individuals without gross income. The slope equals 1 minus the marginal tax rate. Individuals left of the laissez-faire allocation receive net income support from the government, whereas individuals right of the laissez-faire allocation are net tax payers.

With Rawlsian social objectives we find a much higher optimal transfer to non-employed individuals than in the actual system. In addition, individuals who earn up to about 15,000 euros are better off under the optimal Rawlsian tax schedule than they currently are. And, up to about 22,000 euros individuals receive net-income support from the government. The utilitarian government provides about the same transfer to non-employed individuals as in the actual tax-benefit system. Individuals below the average income level are worse of under the utilitarian allocation than the actual system, and individuals above the average income are better off.

Figure 2.8 plots the average tax rates. Both the current tax schedule and the two optimal tax schedules are strictly progressive as can be witnessed from the fact that the average tax is strictly increasing. The optimal utilitarian social planner would increase average taxes at the bottom and decrease taxes at the top. The difference between the current average tax rate and the optimal utilitarian average tax rate is largest at the top. On the other hand, the Rawlsian social planner would decrease the average tax

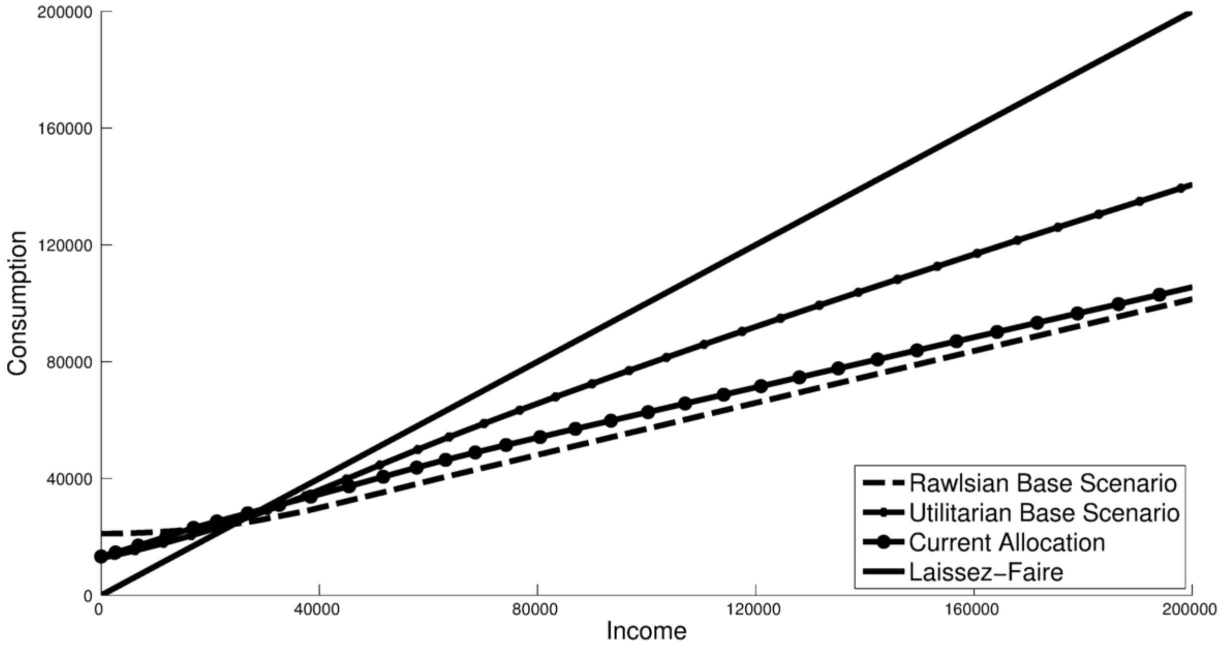


Figure 2.7: The Allocation for the Different Social Preferences

rate for the lowest income earners and increase the average tax rate for the middle- and top-income earners. The difference is largest for middle-income earners. In addition, the average tax burden for high-income earners up to 200,000 euro is still higher under the optimal Rawlsian tax schedule even though the marginal tax rate at the top of the current tax schedule is set at the top of the Laffer-curve. Only the individuals earning more than 250,000 euro will face a lower average tax rate, but less than 0.1% of the population has an income above that level in the Netherlands.

Sensitivity Analysis

Figures 2.9 and 2.10 display the optimal tax rates under a scenario with low elasticities, $\varepsilon^c = 0.18$ and $\varepsilon^u = 0.13$, and a scenario with high elasticities, $\varepsilon^c = 0.53$ and $\varepsilon^u = 0.38$. As noted before, when the elasticity is very low, and we employ a Rawlsian social welfare function, the top rate is too low. But for the high-elasticity case it is too high. These figures also demonstrate that the actual marginal tax rate at the bottom is always too low, even if the elasticity of taxable income is large and we assume a utilitarian social welfare function. Below we consider whether this conclusion still holds when we introduce an extensive margin.

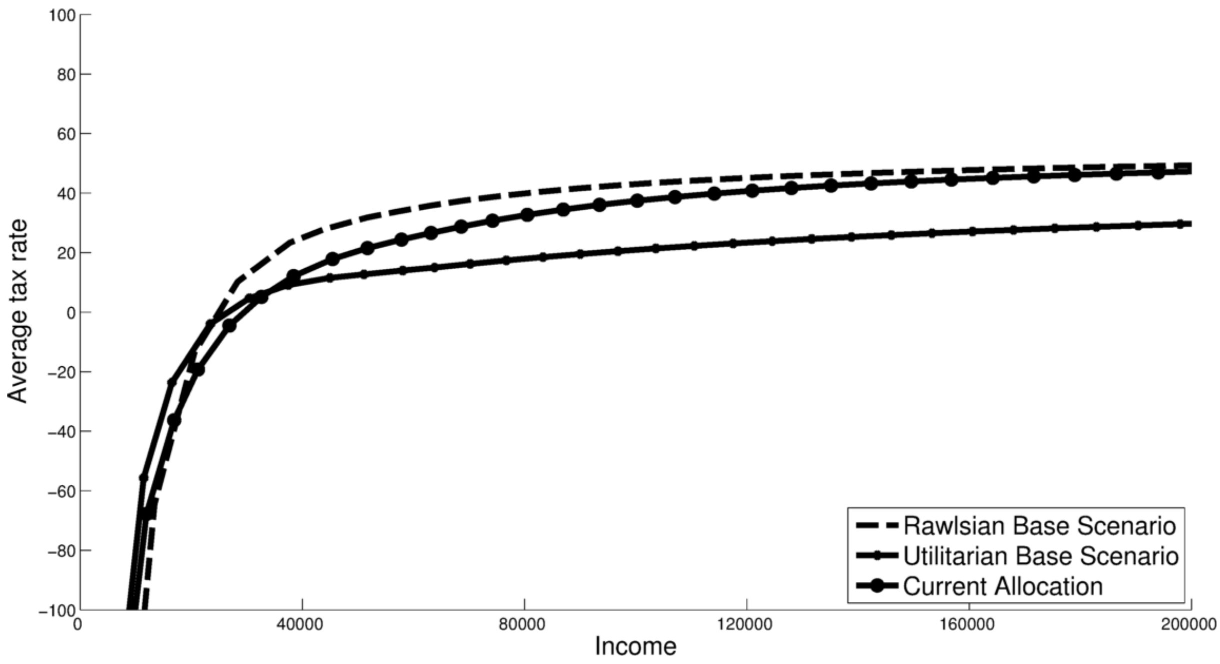


Figure 2.8: The Average Tax Rates for Different Social Preferences

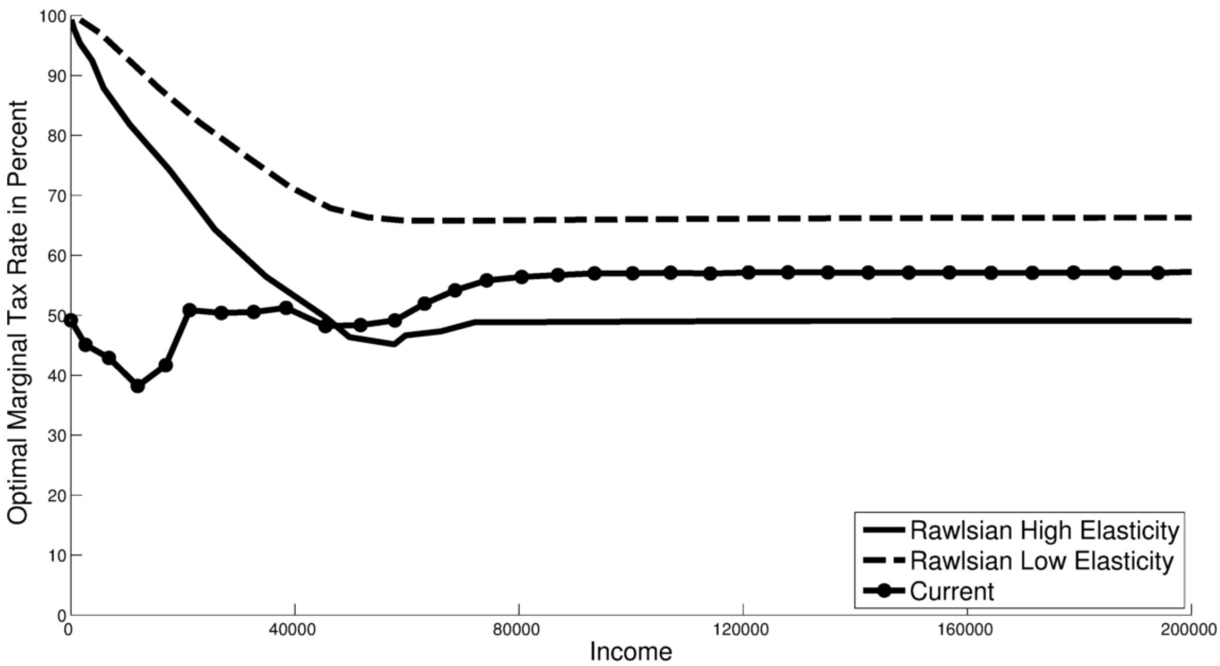


Figure 2.9: The Optimal Tax Schedule with Rawlsian Social Preferences. Low: $\varepsilon^c = 0.18$ and $\varepsilon^u = 0.13$. High: $\varepsilon^c = 0.53$ and $\varepsilon^u = 0.38$.

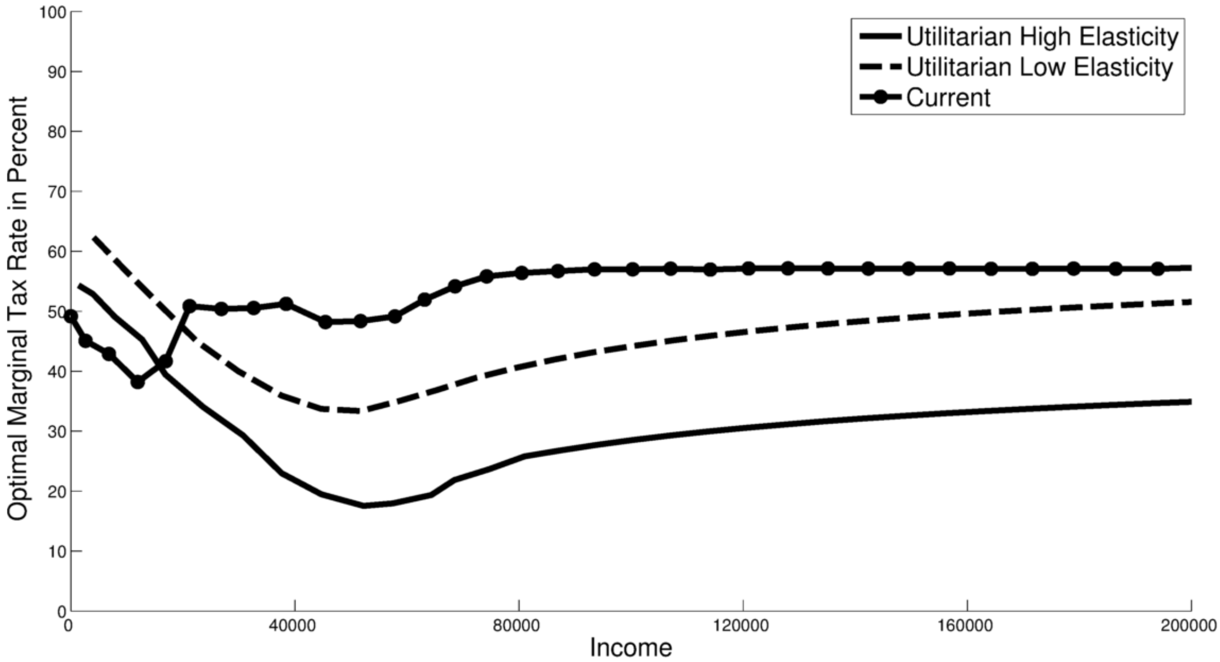


Figure 2.10: The Optimal Tax Schedule with Utilitarian Social Preferences. Low: $\varepsilon^c = 0.18$ and $\varepsilon^u = 0.13$. High: $\varepsilon^c = 0.53$ and $\varepsilon^u = 0.38$.

2.5.3 Intensive and Extensive Margin

Figures 2.11 and 2.12 give the optimal marginal tax rates under both intensive and extensive labor-supply responses. From these figures we see that the introduction of an extensive labor-supply response hardly affects the optimal top rate. High-income earners do not really respond on the extensive margin. The extensive labor-supply response reduces the marginal tax rates especially for low- and middle-income earners.

For low-income earners the optimal marginal tax rate drops significantly due to the introduction of the extensive margin, especially under Rawlsian social preferences. A high marginal tax rate at the bottom increases the average tax rate for middle-income earners. This induces middle-income earners to leave the labor market, which reduces government revenues. Hence, marginal tax rates are optimally set lower. Nevertheless, when compared to the actual system, marginal tax rates at the bottom should still be higher, even under utilitarian social preferences. For middle-income earners we see that the marginal tax rates in the current tax system are lower than actual rates, even under Rawlsian preferences. From this we could conclude that in the current tax system the marginal tax rates for middle-income groups are too high.

Our findings contrast sharply with those of Jacquet *et al.* (2010). The main difference between their simulations and ours is that the participation elasticities in our model are highly non-linear and hump-shaped with income, see Figure 2.13. Although the average

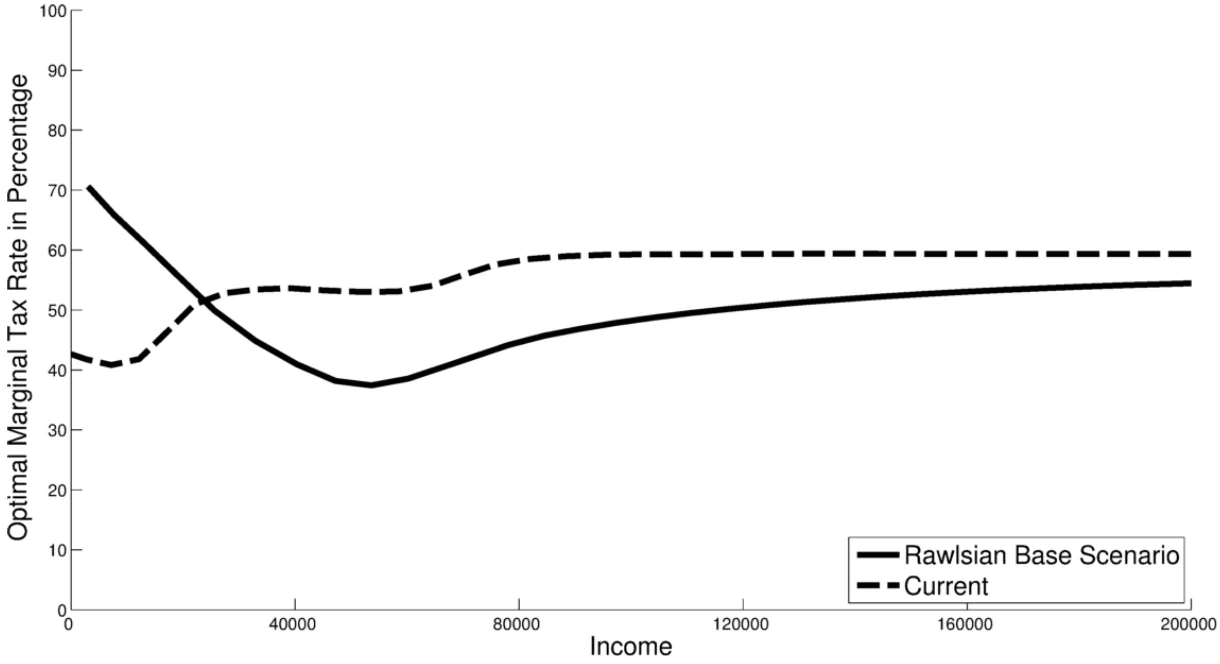


Figure 2.11: The Optimal Tax Schedule with Rawlsian Social Preferences, with Intensive and Extensive Labor-Supply Responses

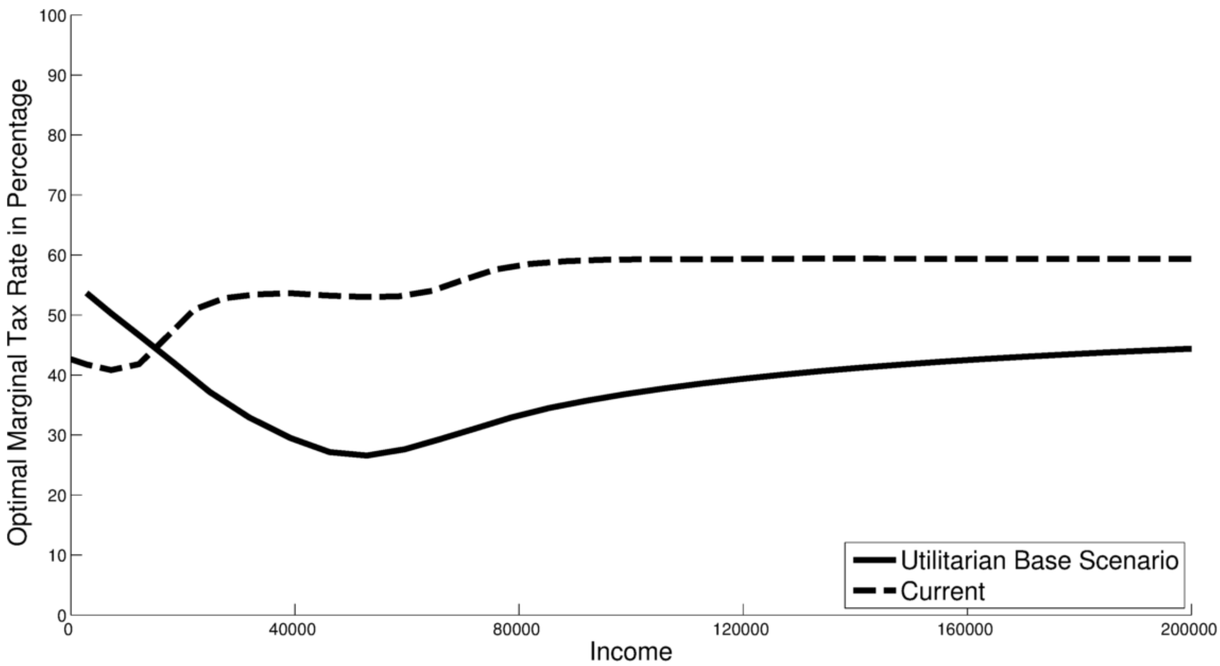


Figure 2.12: The Optimal Tax Schedule with Utilitarian Social Preferences, with Intensive and Extensive Labor-Supply Responses

participation elasticity is calibrated at .25 in the baseline, the participation elasticity is endogenously determined by the distribution of participation costs. We can fit observed education-specific employment rates only when participation elasticities are low for the lowest-skilled workers (due to high non-employment benefits) and for the highest-skilled workers (due to very high earnings compared to non-employment benefits).

In the baseline simulations of Jacquet *et al.* (2010), the participation elasticity is roughly flat over the entire earnings distribution. These authors do not estimate the distribution of participation costs and calibrate the model to real-world data, but plainly assume that participation elasticity declines from 0.5 to 0.4 from the lowest to the highest income level. As a result, introducing an extensive margin results in much lower optimal tax rates over the entire earnings distribution, including the top. We think that our estimation of the distribution of participation costs is better founded, and produces empirically more plausible participation elasticities, see also the literature review.

Figure 2.14 shows the participation tax corresponding to the optimal tax schedules in 2.11 and 2.12, and the actual schedule. Recall from Section 2 that the participation tax $(T(z_n) + b)$ measures the transfer to the government if an individual decides to enter the labor market, pay taxes and forgo non-employment benefits.

For a utilitarian social welfare function a positive participation subsidy (about 2,250 euro) is optimal for the workers earning a very low income. Such a subsidy redistributes resources to the working poor, which still have a large social welfare weight. For a Rawlsian social welfare function, it is always optimal to tax participation on a net basis, even for those workers with low earnings. The participation tax is then about 9,600 euro. The Rawlsian government only cares about the worst-off in society, which are the non-employed. As a result, redistribution to the working poor does not raise social welfare, as it implies less redistribution towards the non-working poor. The participation tax with the Rawlsian government is thus quite high even for very low earnings.

Figure 2.15 gives the optimal average tax rates. Both the optimal utilitarian tax schedule and the optimal Rawlsian tax schedule are no longer strictly progressive. The average tax rate slightly decreases above median income levels. Participation of these groups is very important for government revenue. In particular, the participation elasticity is highest among the middle-income groups, see Figure 2.13. By slightly decreasing the average tax rate, the government boosts labor-force participation, which yields larger government revenue. As expected, the utilitarian social planner increases the tax burden for poor individuals and decreases the tax burden for the rich. The Rawlsian social planner increases the tax burden for all participating individuals, in order to reach maximum support levels for the unemployed.

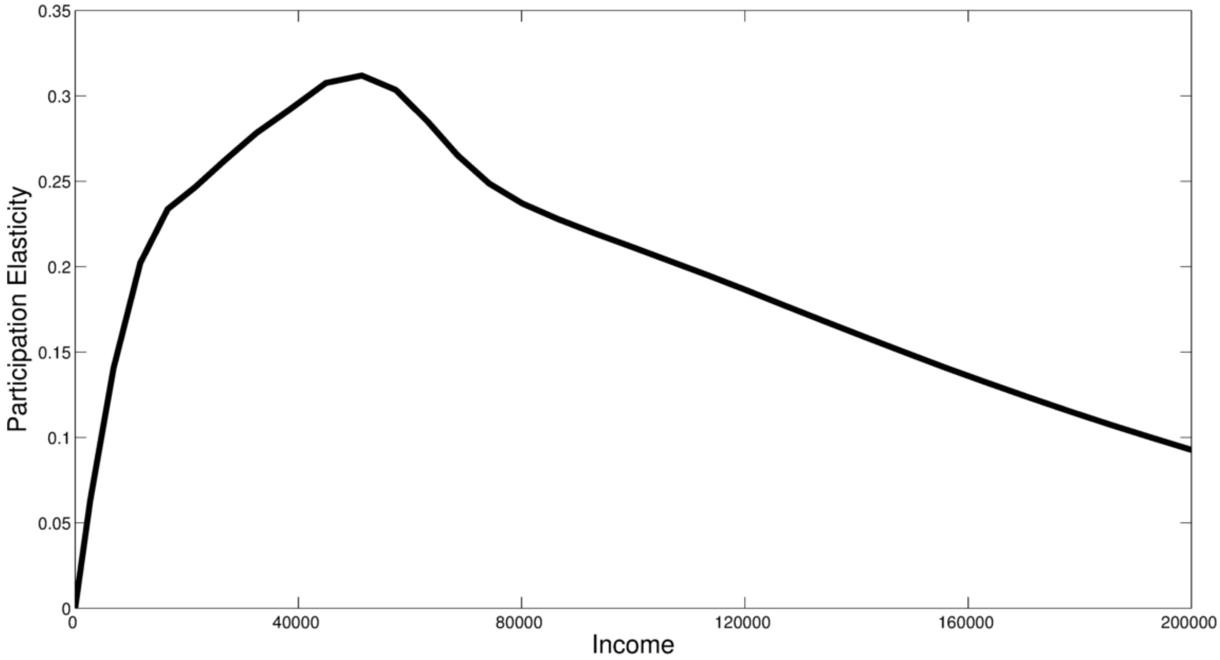


Figure 2.13: Participation Elasticity by Income

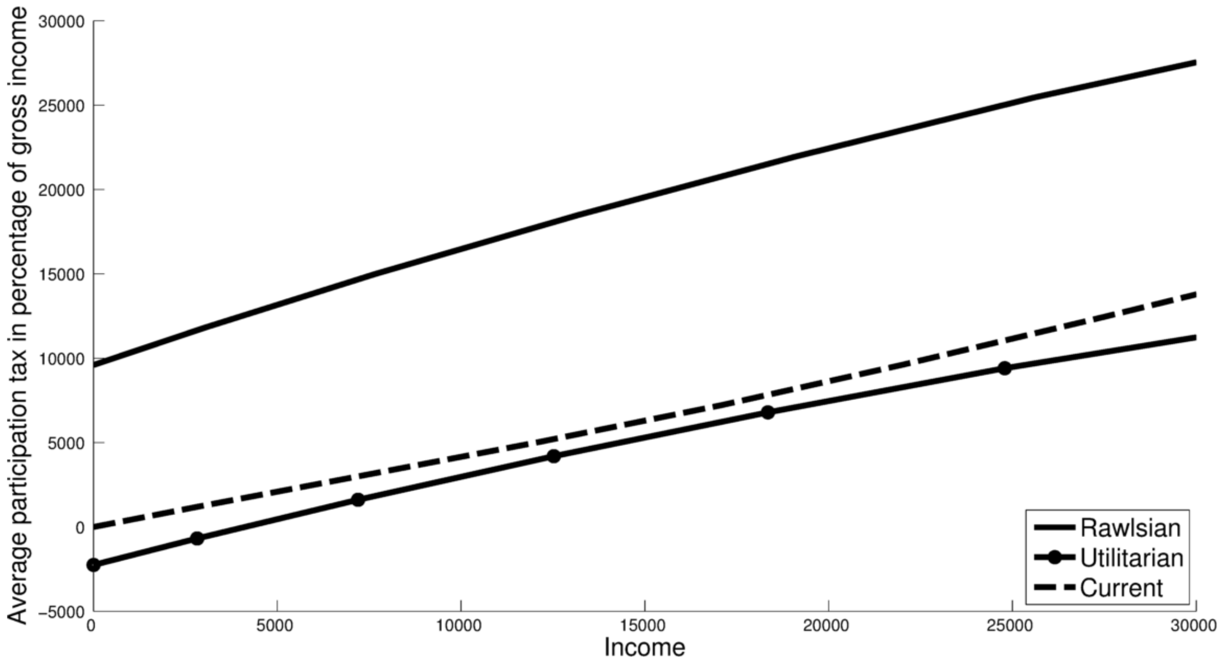


Figure 2.14: The Participation Tax $T(z) + b$ with Utilitarian and Rawlsian Preferences

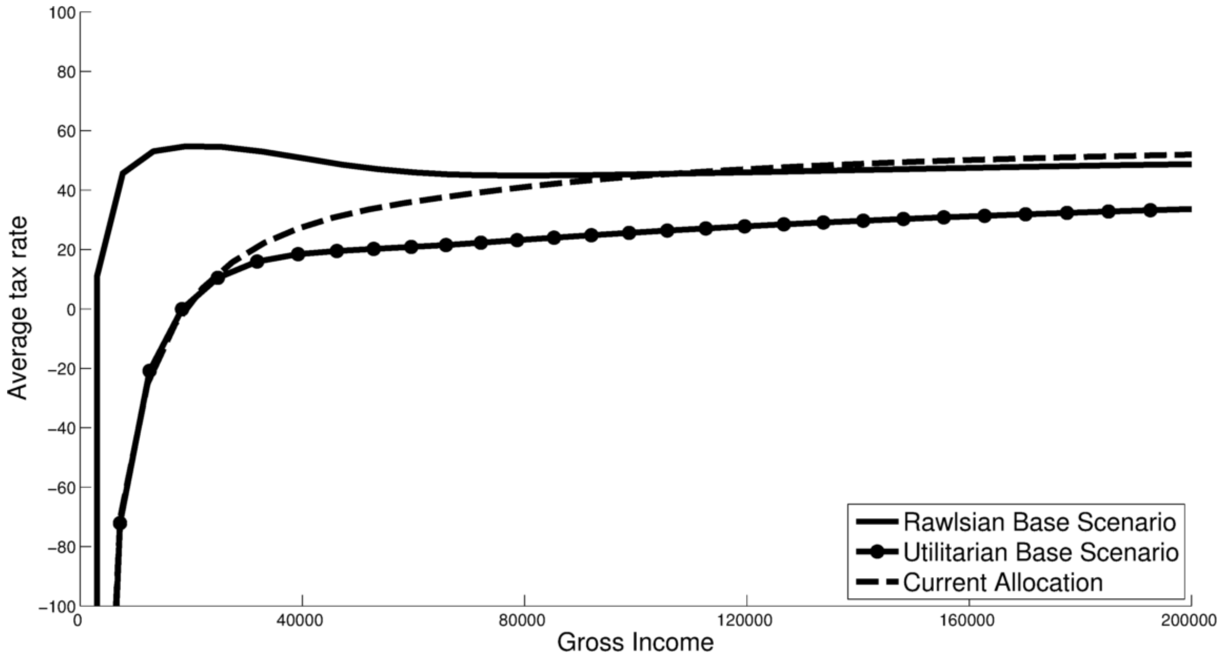


Figure 2.15: The Average Tax Rates for Different Social Preferences

Figure 2.16 depicts the cumulative employment rate up to each ability level for the current tax schedule and the two optimal tax schedules. As can be seen, total employment decreases by about 20% when the Rawlsian tax schedule would be implemented. On the other hand, the optimal utilitarian tax schedule increases total employment by around 10%. The slope of both lines is about equal, which indicates that the large differences in employment are caused by differences in benefit levels. The difference in marginal tax rates is less important.

Sensitivity Analysis

We should note that the results with an extensive margin in general should be interpreted with the appropriate care. We have only limited knowledge on participation rates by skill, and on participation elasticities by skill. Consequently, there is significant uncertainty surrounding our estimates for the distribution of participation costs. Therefore, we also conducted alternative simulations with different elasticities. Figures 2.17, 2.18 give the optimal tax rates under a scenario with low intensive elasticities($\varepsilon^c = 0.18$ and $\varepsilon^u = 0.13$) and a low extensive elasticity ($\varepsilon^P = 0.13$), and a scenario with high intensive elasticities ($\varepsilon^c = 0.53$ and $\varepsilon^u = 0.38$) and a high extensive elasticity ($\varepsilon^P = 0.38$).

From Figures 2.17 and 2.18 we conclude once more that the current tax rate at the top is above (below) the optimal tax rate if the elasticity of taxable income is high (low) and the government has a utilitarian (Rawlsian) preferences. The result that the tax rate

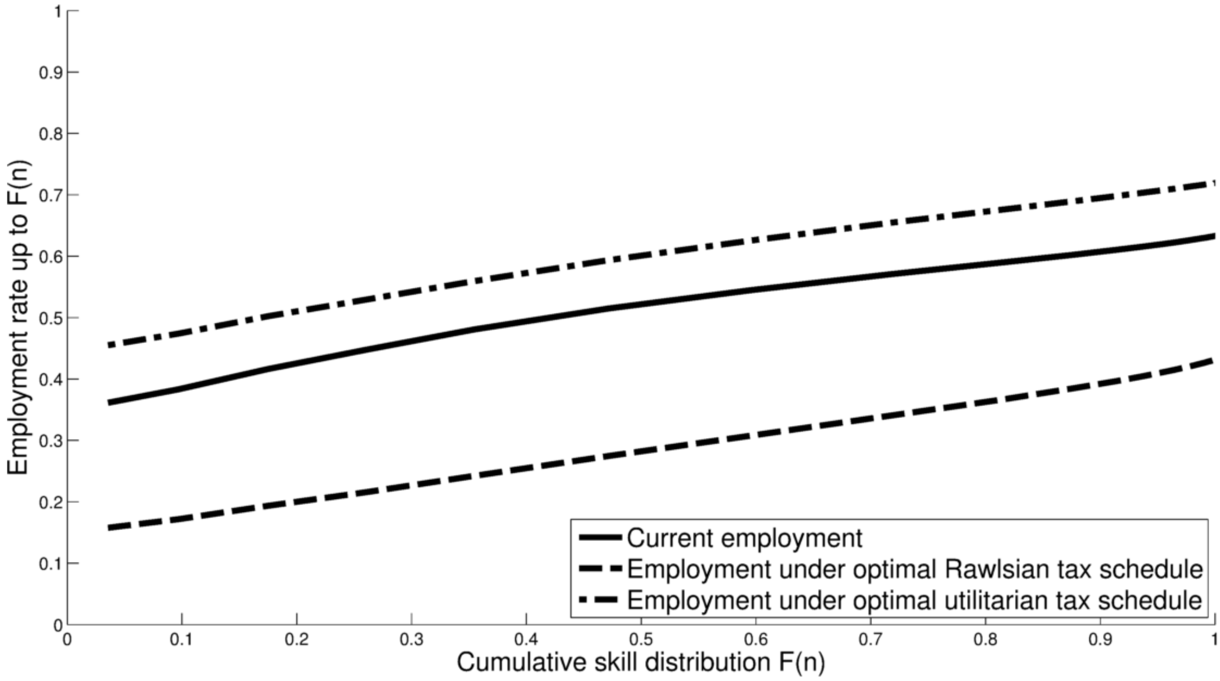


Figure 2.16: Cumulative Employment under the Current and Optimal Tax Schedules

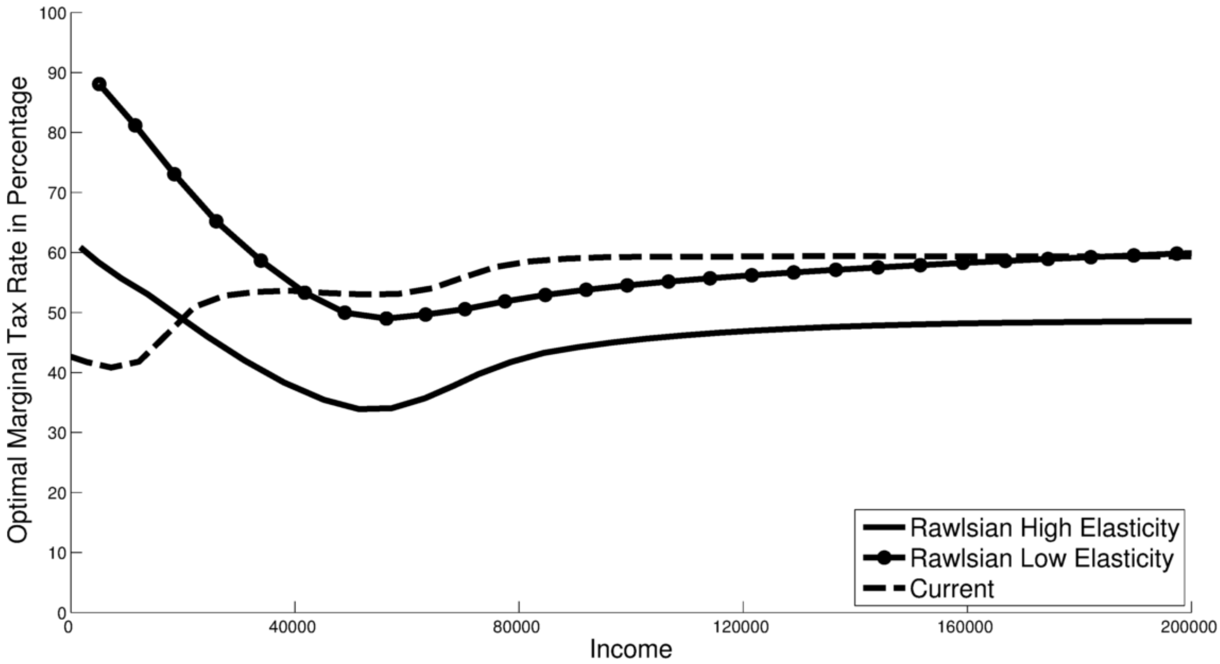


Figure 2.17: The Optimal Tax Schedule under Rawlsian Social Preferences, with Intensive and Extensive Labor-Supply Responses. Low: $\varepsilon^c = 0.18$, $\varepsilon^u = 0.13$ and $\varepsilon^P = 0.13$. High: $\varepsilon^c = 0.53$, $\varepsilon^u = 0.38$ and $\varepsilon^P = 0.38$.

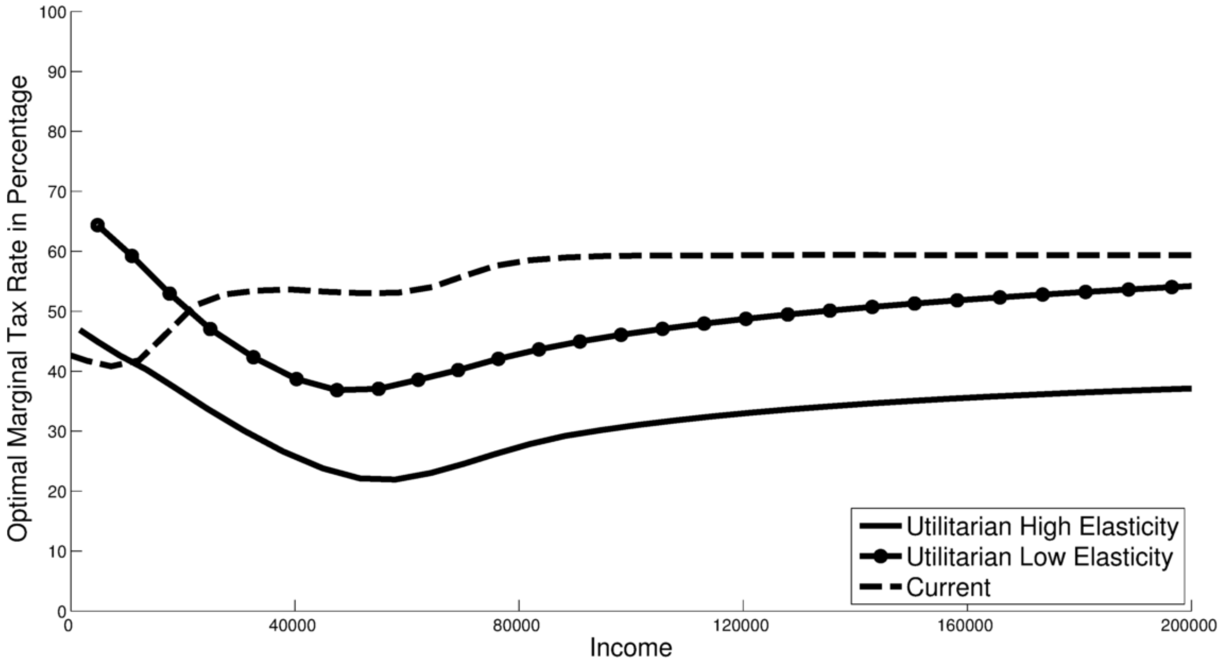


Figure 2.18: The Optimal Tax Schedule with Utilitarian Social Preferences, with Intensive and Extensive Labor-Supply Responses. Low: $\varepsilon^c = 0.18$, $\varepsilon^u = 0.13$ and $\varepsilon^P = 0.13$. High: $\varepsilon^c = 0.53$, $\varepsilon^u = 0.38$ and $\varepsilon^P = 0.38$.

at the bottom is too low remains valid even for a utilitarian social welfare function and a high elasticity.

2.5.4 Single Earners vs. All Earners

We explore the robustness of one of our main conclusions – that current marginal tax rates are too low for the low-income earners – by focusing at single-earning individuals. Although our analysis applies to the average of all tax payers, it might not apply to all types of tax payers. Since single-earner households are generally more reliant on income-dependent support than other tax payers are, their current marginal tax rates at low incomes are larger than for the average earner.

The kernel estimate of the non-linear tax schedule indeed masks a lot of heterogeneity, as the scatter plot in Figure 2.3 reveals. Single earners are located at the upper-left corner of Figure 2.3, but their marginal tax rates are completely smoothed out in the kernel estimate of the tax schedule. This is due to a large group of secondary earners, in the bottom-left corner, having low earnings, and facing low marginal tax rates, since they are not receiving income-dependent support.

Therefore, we not only plotted the current effective marginal tax schedule for all tax payers in the graphs with optimal tax schedules, but also the current effective marginal tax schedule for single earners, see Figures 2.5, 2.6, 2.11 and 2.12. We thus smoothed the marginal tax rates for single-earning individuals only.[33] The tax schedules of the average income earner and the single-income earner are very close, but the tax schedule for the single-income earner at very low earnings indeed features higher marginal tax rates.

Consequently, when designing policy reforms, one should keep in mind that the single-earning individuals already face marginal tax rates that are closer to the optimal non-linear schedule than most other tax payers. Moreover, the marginal tax rates of the current tax schedule would be somewhat too high for single earners under utilitarian social preferences, with a very weak social preference for redistribution. Still, marginal taxes would still be much too low for single earners for a very redistributive Rawlsian social welfare function. Hence, even for moderately redistributional concerns our conclusion would survive.

A final caveat is that primary earners have much lower labor-supply elasticities than secondary earners, see also the review of our labor-supply stimates. This may also undermine our conclusion that low-income earners face too low marginal tax rates, since we assumed that all individuals have the same labor-supply elasticity. Quite some elastic secondary earners in small part-time jobs could be located at the lower end of the earnings distribution. Hence, optimal taxes could be lower than in our simulations.[34] Nevertheless, the ETI-estimates discussed earlier also revealed that the elasticities of taxable income are roughly flat over the entire earnings distribution. Hence, it remains unclear whether our simulations are indeed biased.

2.6 Flat Tax

The flat income tax consists of a flat tax rate, which finances a non-individualized lump-sum transfer $(-T(0))$ in the model with an intensive margin only. In the model with both labor-supply margins the flat tax finances both the transfer for the working population $(-T(0))$ and the non-employment benefit (b).

We derive that a flat tax is clearly not desirable. The optimal linear tax rates are always higher than the income-weighted marginal tax rates under the optimal non-linear

[33] We did not re-estimate the skill distribution and re-compute optimal tax schedules for single earners. The reason is that the actual Dutch tax schedule is individualized and not dependent on whether individuals are single earners or not. Hence, the optimal tax schedule would remain the same. Moreover, if we re-estimate the skill distribution based on the marginal tax schedule for single-earners and then re-compute optimal tax schedules, the resulting tax schedules are indistinguishable from the reported ones, since the marginal tax schedules for single-earners are very close to the marginal tax schedules for all income earners.

[34] Our simulation model cannot cope with preference heterogeneity resulting in different labor-supply elasticities for different groups of income earners. Consequently, future research should explore the sensitivity of our conclusions with respect to more elastic secondary earners.

schedule with an intensive margin only, as the dashed lines in Figures 2.5 and 2.6 demonstrate. The optimal flat tax is still higher when both intensive and extensive margins are included for the Rawlsian government, but almost correspond to the weighted average of non-linear tax rates with a utilitarian government, see Figures 2.11 and 2.12. Moreover, the amount of income transferred to the working and non-working poor is always lower under the optimal flat tax compared to the non-linear tax in all simulations, but one, see Table 2.10. Only in the utilitarian case with both intensive and extensive labor-supply margins the optimal transfer provided to the non-working poor is slightly higher under the flat tax. Hence, the flat tax entails either less efficiency or less equity or both.

Intuitively, in order to organize a given amount of redistribution, the linear tax always requires higher marginal tax rates, since the lump-sum transfers are provided to everyone, irrespective of income. The flat tax cannot precisely target transfers to different income groups so that the leaking bucket of Okun is leaking more when a flat income tax is employed rather than the non-linear income tax. The flat income tax is therefore an inferior instrument for income redistribution.

Table 2.10: Comparison Optimal Non-linear Tax with Optimal Flat Tax and Current Tax System

	Intensive margin only			Intensive + extensive margin		
	$-T(0)$	DWL	Welfare loss	$-T(0)$	b	DWL
Rawlsian						
Optimal non-linear	21,105	0.82	0.000	1,891	11,490	0.27
Optimal linear	18,587	1.00	0.090	795	10,801	0.49
Current	13,268	0.37	0.278	8,086	8,086	0.41
utilitarian						
Optimal non-linear	12,689	0.15	0.000	9,055	6,800	0.17
Optimal linear	9,220	0.16	0.004	4,045	6,879	0.18
Current	13,268	0.37	0.075	8,086	8,086	0.41

We calculate the marginal deadweight losses of the optimal non-linear tax, the optimal linear tax and the current tax-benefit system in Table 2.10.[35] Clearly, the marginal deadweight losses are always lower under an optimal non-linear tax system in comparison to the optimal flat tax. Moreover, our simulations with both extensive and intensive labor-supply margins demonstrate that moving from the current tax-benefit system to the optimal non-linear tax lowers the marginal deadweight loss from about 41 cent per

[35]The general formula for the marginal deadweight loss under a non-linear income tax is derived in the appendix and equals: $\int_{\mathcal{N}} \varepsilon_n^c \frac{T'(z_n)}{1-T'(z_n)} n l_n \tilde{k}(n)\, \mathrm{d}n \left(\int_{\mathcal{N}} n l_n \tilde{k}(n)\, \mathrm{d}n\right)^{-1}$.

euro, to 27 cent per euro in the Rawlsian case to 17 cent per euro in the utilitarian case. Hence, reforming the current tax-benefit system towards the optimal non-linear system almost halves the marginal distortions of the tax system for any social preference for redistribution.

We also compute the welfare loss of the flat tax system in comparison with optimal non-linear tax system, see Table 2.10. Similarly, we also calculate the welfare loss of the current tax system in comparison with the optimal non-linear tax system. Due to computational complexities, we were only able to do so with model with an intensive margin.[36] Our welfare measure is the compensating variation: how much resources can be taken out of the economy with optimal non-linear taxes to achieve the same level of social welfare as in the economy with an optimal flat tax or the current tax-benefit system? The welfare cost of moving from the optimal non-linear taxes to the optimal flat tax with utilitarian social preferences is 0.4% of GDP, which is relatively modest. But, one needs to recall that the utilitarian government is only weakly redistributive. The welfare loss for the Rawlsian government is a very large 9% of GDP. Intuitively, when social preferences are more redistributive, the flat tax is more of a strait jacket to the government to achieve its redistributional objectives. The welfare difference between the optimal non-linear tax system and the current tax-benefit system is 7.5% of GDP under utilitarian social preferences and an astonishing 28% of GPD under Rawlsian social preferences. This, again, demonstrates the sub-optimality of the current tax-benefit system in achieving social objectives. An important caveat is in order here. These welfare analyses are conducted for the model with an intensive margin only. Hence, they should be interpreted as an upper bound of the potential welfare losses of not implementing the optimal non-linear tax schedule.

2.7 Social Welfare Weights of the Current Tax-Benefit System

In the previous sections, we have derived the optimal tax schedule for given social preferences. In this section, we invert the question: under what social preferences is the current tax schedule optimal? Under any Bergson-Samuelson social welfare function, welfare weights are monotonically declining in income. In addition, all welfare weights are non-negative. However, political-economy considerations might induce politicians to set a tax schedule which attaches more weight to the middle-income groups. Policy mak-

[36] In a future version of this paper we hope to report these results.

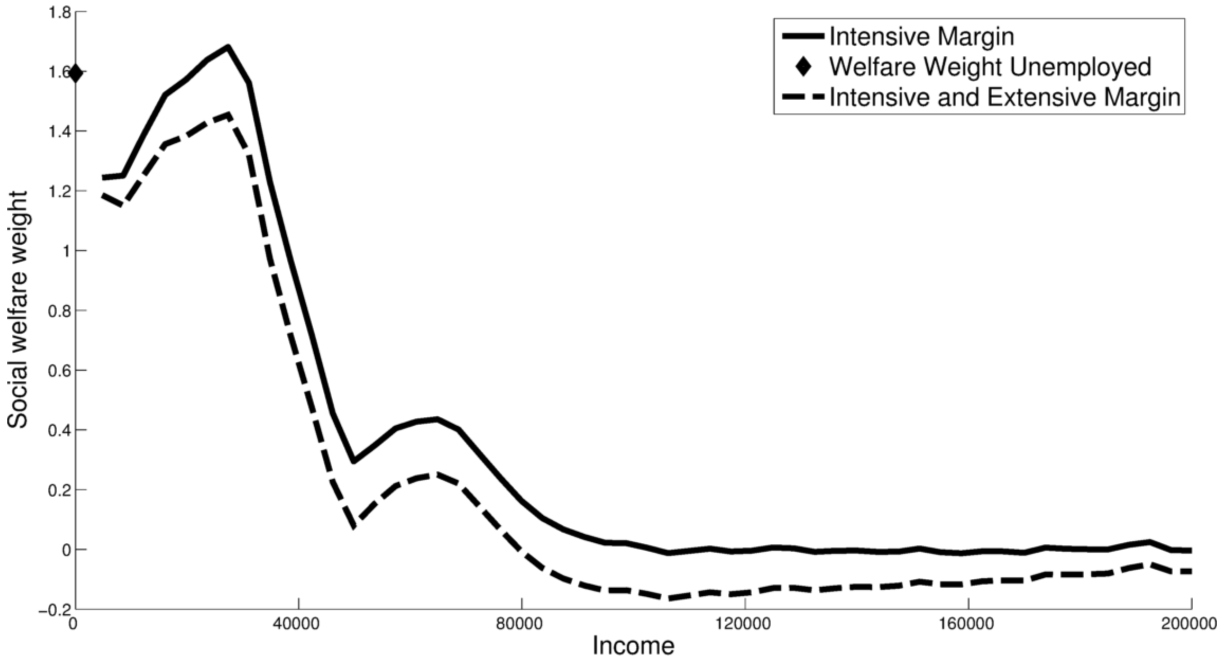


Figure 2.19: Social Welfare Weight Intensive Labor Supply Responses and Under Both Responses

ers might also have under- or overestimated distortions associated with the current tax schedule. We investigate whether inconsistencies in social welfare weights are present in the Netherlands.

2.7.1 Results

In figure 2.7.1, we plot the social welfare weights implied by the current tax-benefit system had it been optimized. We do so for the model with an intensive margin only and the model with both intensive and extensive margins. The social welfare weights under the extensive margin reveal the same patterns as the social welfare weights with an intensive margin only, although the increase in the welfare weights between low-income workers and middle-income workers is less prominent. As can be seen, social welfare weights under both labor-supply responses are generally below the social welfare weights under the intensive margin. For a given tax rate, distortions are larger if individuals can also respond on the extensive margin, and social welfare weights are lower. There are three clear inconsistencies in these patterns of social welfare weights.

First, in both graphs the social welfare weights are increasing until modal income. Indeed, the political system attaches the largest social welfare weights to the middle-income groups. This implies that the current government positively values taxing the working poor to redistribute more resources towards the middle-income groups. This is inconsistent with any standard social welfare function, which attaches a lower welfare weight to middle-income earners than to the working poor. These results suggest that political-economy considerations can be important in explaining current tax schedules. Indeed, the densely populated middle-income groups constitute the largest fraction of the Dutch electorate.

Second, for the high-income levels, the welfare weights are slightly negative, and return slightly above zero in the very limit. Apparently, the current government values penalizing the high-income earners. The reason is that the current tax rate in the top bracket is set beyond the top of the Laffer curve for most top-income tax payers.[37] Such a policy produces no redistributional benefits and only distortions. Negative welfare weights at the top of the income distribution are in line with findings in Bourguignon and Amedeo Spadaro (2010) for France.

Third, there is a large discontinuous drop in the welfare weight for the poor as they start working and earning income. In particular, the current government values a euro transferred to the non-working poor 1.5 times as high as transferring the same euro to the working poor. This is an anomaly as it suggests that the government views the non-working poor as much more deserving of income support than the working poor even if they have the same income. Low welfare weights for the working poor have been consistently found in other studies, see e.g. Bourguignon and Amedeo Spadaro (2010) and Bargain *et al.* (2011).

Figures 2.20 and 2.21 show the welfare weights in the case of a high and low labor-supply elasticity. As can be seen from the figures, welfare weights become close to monotonically decreasing if the elasticity of taxable income is very low. In addition, all welfare weights will then be positive. A possible explanation for the anomalies in our baseline simulation is that policy-makers underestimate the efficiency costs of taxes and do not optimize tax-benefit systems accordingly. On the other hand, if the earnings-supply elasticity is high, the non-monotonicity in the welfare weights is even more striking. Also, the welfare weights at the bottom are much lower and welfare weights for top-income earners are even more negative.

[37]Only for tax-payers with incomes above 200,000 euro the social welfare weights turn marginally positive. This explains why the revenue-maximizing top-rate is still marginally above the current top rate.

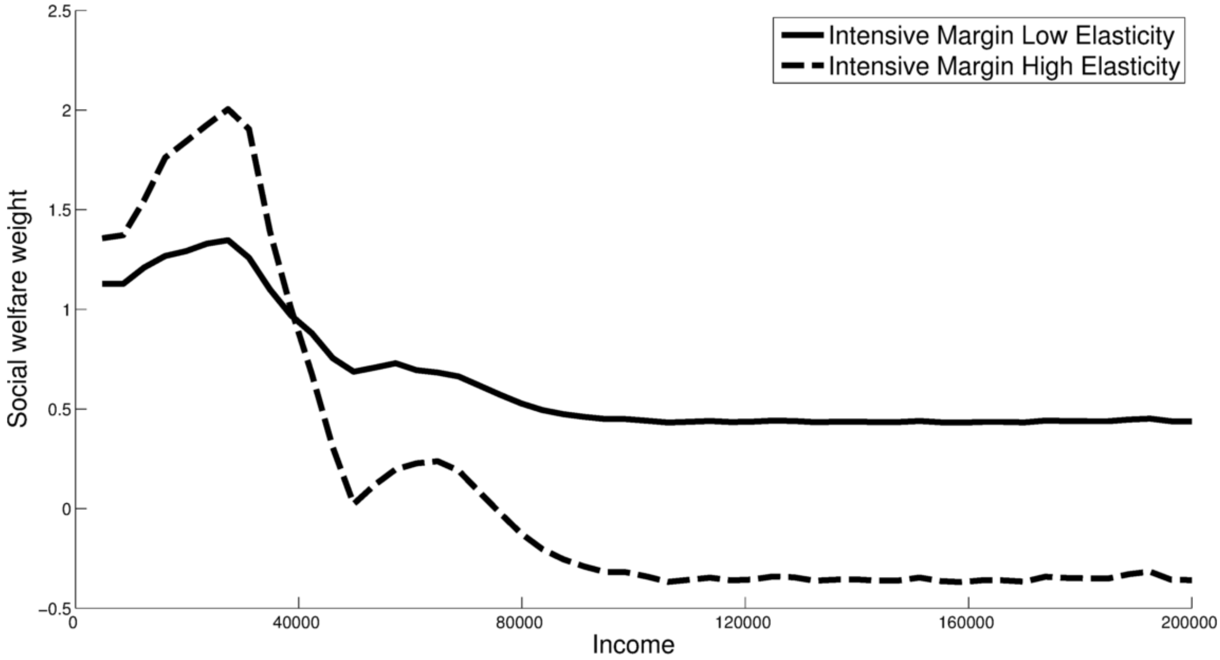


Figure 2.20: Social Welfare Weight Under Intensive Labor Supply Responses

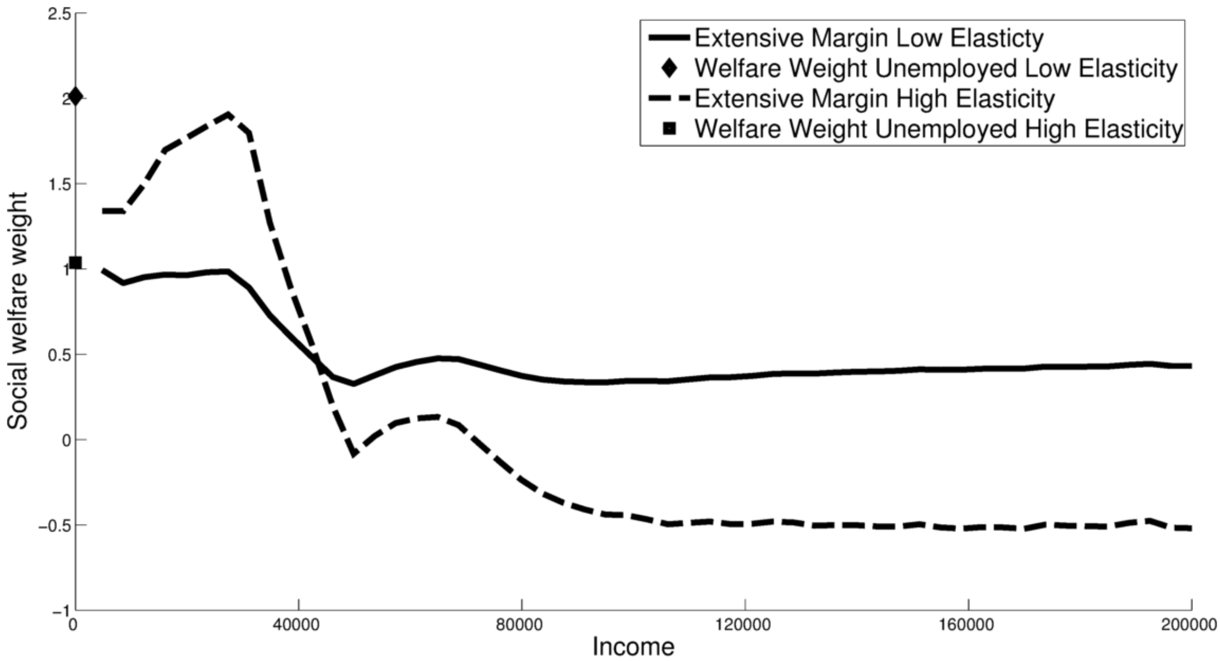


Figure 2.21: Social Welfare Weight Under Both Intensive and Extensive Labor Supply Responses

2.8 Directions for Future Research

We assumed that all worker types are perfect substitutes and that there are, therefore, no general-equilibrium effects on the wage structure as a result of redistribution policy, or otherwise. Rothschild and Scheuer (2013) extend the Stiglitz (1982) model of optimal income taxation with endogenous wages to an infinite number of skill types. These authors demonstrate that redistributive governments should exploit general-equilibrium effects on the wage structure by setting less progressive marginal tax schedules. Optimal tax rates would increase at the bottom and decrease high at the top. When applied to the Netherlands, this would presumably render the current tax-benefit system even more sub-optimal than our analysis has demonstrated.

Recent developments in behavioral economics point to a number of potential weaknesses of our analysis. It might be that individuals are engaged in 'rat races' (Akerlof, 1976) and 'keeping up with the Joneses' (Layard, 1980). Distortionary income taxes then not only entail deadweight losses, but also yield benefits by taming the rat race or correcting status-seeking behavior. Total distortions of income taxation are then smaller, and optimal taxes increase. See also Kanbur *et al.* (2006). By the same token, Alesina *et al.* (2005) argue that there could be rivalry in leisure as well. This raises distortions of income taxation, since not only labor-supply choices are distorted, but also a 'leisure multiplier' is put in motion. Hence, taxes should optimally be set lower. Gerritsen (2013) shows that utility-maximizing individuals might not maximize well-being, and, hence, suffer from internalities. He finds that marginal tax rates should be lowered for the poor to give them stronger incentives to work more, whereas they should be increased to the rich, to give them stronger incentives to enjoy more leisure.

In our analysis we ignore that many individuals live in multi-person households. We thereby ignore, for example, intrahousehold redistribution and economies of scale. An analysis of optimal family taxation, following the lead by *e.g.* Boskin and Sheshinski (1983), Apps and Rees (1998), Schroyen (2003), Alesina *et al.* (2011) and Kleven *et al.* (2009), is beyond the scope of this paper. Indeed, it would be useful to explore conditioning tax schedules on the income of primary and secondary earners, since the latter are typically more elastic. This is left for future research.

2.9 Conclusions

This study analyzed the optimal redistributive tax and transfer system in the Netherlands using realistically calibrated models with both intensive and extensive margins of labor supply. We found that the optimal non-linear tax schedule features a U-shape. This contrasts sharply with the current schedule of effective marginal tax rates in the Netherlands; tax rates are gradually increasing with income. Although the optimal marginal tax rates at the bottom fall significantly when an extensive margin is introduced, we find that marginal tax rates are too low at the bottom of the earnings distribution compared to the current tax schedule for all the social welfare functions we analyzed. Higher marginal tax rates until modal income thus help to redistribute more income towards the working and non-working poor. Marginal tax rates for the middle-income earners are too high. Also, the top tax rate appears to be set too high, and even on the wrong side of the Laffer-curve. The observed patterns in marginal taxes suggest that the middle-income earners are undertaxed, at the expense of the top-income earners and the working poor.

A central finding in all our simulations is that the working poor should pay much lower average taxes. However, this does not imply that they receive a net subsidy to work. A large participation subsidy is found only under weakly redistributive social objectives. Already for moderately redistributive preferences we find that there should always be a net participation tax for the working poor.

A flat income tax schedule is never found to be optimal. Indeed, all simulations demonstrate the inferiority of the flat tax to redistribute income. Under an optimal flat tax, marginal tax rates are higher, transfers/benefits are lower or both. Hence, the equity-efficiency trade-off worsens substantially. Simulations of the model with an intensive margin only demonstrate that an optimal flat tax gives substantial welfare losses compared to the optimal non-linear tax, running from 0.4% of GDP for utilitarian to 9% of GDP for Rawlsian social preferences. The flat tax is a particularly costly strait jacket for strongly redistributive governments. The marginal deadweight loss of the current tax system (41 cents per additional euro revenue) is roughly cut in half when the optimal non-linear tax schedule would be implemented for any social desire to redistribute income. A flat tax renders the leaking bucket of Okun (1975) a sieve. Hence, political discussions about a flat tax are an economic non-starter.

The social welfare weights underlying the current tax-benefit system give rise to similar conclusions. Dutch social welfare weights are increasing with income until median income. The government thus prefers transferring resources to middle-income earners rather than the working poor. Moreover, social welfare weights for top-income earners are slightly negative. This implies that the current government likes to penalize top-income earners by setting too high marginal tax rates. Finally, the government attaches a much larger welfare weight to the non-working poor than the working poor. Why the working poor are apparently less deserving of income support than the non-working poor – even if they have the same income – remains unclear to us.

The policy implications of our research are clear. The government should lower the tax burden on the working poor, by raising the tax burden on the middle- and higher-income groups. This can raise social welfare under all standard social welfare criteria we analyzed. This is typically not a Pareto improvement, since middle- and higher-income earners need to pay higher taxes. However, tax reforms are feasible where the welfare gains for the low-income groups outweigh the welfare losses for the middle- and higher-income groups. Substantial increases of the EITC therefore appear to be socially desirable. By exactly how much is a political judgement. The top rate should not be increased further, as it would only increase deadweight losses while reducing tax revenue available for income redistribution.

Chapter 3

Revealed Social Preferences of Dutch Political Parties[1]

"Don't tell me what you value, show me your budget, and I will tell you what you value."
Joe Biden – US Presidential Elections, September 15, 2008

3.1 Introduction

The quote from Vice-President Joe Biden of the US appeals to many economists who prefer revealed over stated preferences. In this paper we try to go beyond the rhetoric of the political debate, and study what political parties really want in terms of income redistribution. To this end, we use unique data on the proposed tax-benefit system of Dutch political parties in their election campaigns. We use the inverse optimal-tax method to derive the social welfare weights that Dutch political parties attach to different income groups. This allows us to analyze whether all parties care more about the poor(er) than the rich(er), if they care about all income groups, and who cares the most about whom.

Revealing the implicit social preferences of tax-benefit systems is an exciting new research area in optimal income taxation. For quite some time after the seminal contribution by Mirrlees (1971), optimal tax theory remained rather theoretical and provided little guidance to actual tax policy. However, at the turn of the century Diamond (1998),

[1]This chapter is based on Zoutman *et al.* (2013b). We thank Nicole Bosch for her assistance in calculating the effective marginal tax rates used in this chapter. We have benefited from comments and suggestions by Olivier Bargain, Etienne Lehmann, Erzo Luttmer, Andreas Peichl, Emmanuel Saez, Paul Tang, Danny Yagan and seminar and congress participants at CPB Netherlands Bureau for Economic Policy Analysis, IIPF Michigan, University of California Berkeley, CPB Workshop on Behavioural Responses to Taxation and Optimal Tax Policy. All remaining errors are our own.

Saez (2001) and Saez (2002b) greatly increased its policy relevance. In particular, Saez (2001) showed that optimal tax rates can be determined when we know the elasticity of the tax base, the distribution of gross earnings, and the social preferences for redistribution. In principle, both the taxable income elasticities and earnings distribution can be determined empirically.[2] However, the social preference for redistribution is ultimately a political question on which economists have little to say. Indeed, researchers can only determine plausible ranges of optimal marginal tax rates within the boundaries determined by the Rawlsian and utilitarian social welfare functions. Comparing the resulting optimal tax schedules with actual schedules may reveal whether the actual system is optimal, and where there might be room for improvement in terms of social welfare.

A somewhat less ambitious, but equally revealing strategy is to invert the optimal tax problem and look for the social preferences that make a given tax-benefit system optimal. By deriving the social welfare weights in this way, anomalies in actual tax-benefit schedules can be detected, and welfare-improving tax reforms can possibly be identified. By using this strategy one circumvents the necessity to assume an unknown social welfare function. Pioneering work on the dual approach for tax-benefit systems has been done by Bourguignon and Spadaro (2012).[3] They reveal the implicit preferences for income redistribution in the French tax-benefit system, using the inverse optimal problem of Saez (2001) with an intensive decision margin (hours or effort), and the inverse optimal problem of Saez (2002b) with both an intensive and an extensive decision margin (not only hours or effort, but also participation). For the model with only an intensive margin they find that social welfare weights are always decreasing, but they turn negative at the top.[4] They obtain these results both when considering only singles and when considering all workers (averaging income for couples). When they introduce an extensive margin, they find that social welfare weights are no longer monotonically declining, and can also turn negative for the working poor when participation elasticities are high (larger than .5).

Blundell *et al.* (2009) consider the social welfare weights of single mothers in the UK and Germany, allowing for both the intensive and extensive decision margin. Their

---

[2] *E.g.* Brewer *et al.* (2010) for the UK, Jacquet *et al.* (2013) for the US and Zoutman *et al.* (2013a) for the Netherlands, all recover the ability distribution using the income distribution, actual marginal tax rates and the required elasticities.

[3] Studying the 'dual' problem of optimal taxation has a longer history, see *e.g.* Stern (1977), Christiansen and Jansen (1978), Ahmad and Stern (1984) and Decoster and Schokkaert (1989). However, only recently have researchers been able to use detailed micro data on incomes and corresponding marginal tax rates to study the social preferences implicit in tax-benefit systems.

[4] The social weights turn negative even though they do not include indirect taxes (close to 20% of income net of direct taxes), as noted by Bourguignon and Spadaro (2012). Including indirect taxes in marginal tax rates would make the social weights even more negative at the top.

analysis goes a step further than Bourguignon and Spadaro (2012) in that they estimate rather than calibrate the behavioural elasticities, using micro data and a discrete-choice model for labor supply. For both Germany and the UK they find that social welfare weights are not monotonically decreasing with income, as the working poor get a lower weight than middle incomes. For Germany they find a negative social weight, for working single mothers with a low income and children younger than school-age.[5]

Bargain and Keane (2010) perform a similar analysis for singles in Ireland, and further extend the analysis by looking at social welfare weights at (four) different points in time (ranging from 1987 to 2005). They find that the resulting social welfare weights are remarkably stable over time, despite some significant policy changes. They do not find negative social welfare weights. However, they again find that social welfare weights are not monotonically declining, where the working poor get a lower weight than middle income earners.

Finally, Bargain *et al.* (2011) do a similar analysis for singles in 17 European countries[6] and the US. They find that social welfare weights are always positive, although they are not monotonically declining for low income groups, which is in line with the studies considered above. They further find that there are significant differences in social welfare weights between groups of countries (the US vs. Continental and Nordic Europe vs. Southern Europe), but rather similar social welfare weights for countries within a particular group.

We build on these previous analyses and consider whether optimal tax theory can be equally revealing for the social welfare weights implicit in the actual Dutch tax-benefit system, and the social welfare weights of Dutch political parties. Since 1986, in a process unique in the world, all major Dutch political parties provide CPB Netherlands Bureau for Economic Policy Analysis (CPB) with their detailed reform package for the tax-benefit system for every election for national parliament. CPB then calculates and reports the income, budgetary and behavioral effects of these reform packages, which then play an important role in the run up to the national elections, but also during the negotiations to form a new government after the elections..[7] These data provide us a with an opportunity to study the redistributive preferences of the political parties.

[5] They do not find negative social weights for top incomes. However, behavioral responses in their model are only in hours worked, which are very low at the top. This probably understates the tax base response to changes in the marginal tax rate at the top as suggested by the literature on the elasticity of taxable income (Feldstein, 1999). The same is true for Bargain and Keane (2010) and Bargain *et al.* (2011).

[6] Including the Netherlands.

[7] See CPB and PBL (2012) for the analysis of the 2012 elections, and the contributions in Graafland and Ros (2003) for the pros and cons of this exercise.

We invert the optimal tax model of Zoutman *et al.* (2013a), which builds on Jacquet *et al.* (2013) and allows for both an intensive (hours or effort) and an extensive (participation) decision margin. We apply this model to the tax-benefit systems as proposed by political parties in the 2002 elections. We use data for 2002 because this is the same year for which Zoutman *et al.* (2013a) recover the ability distribution, using detailed micro data on the income distribution and marginal tax rates. We focus on the proposals by the four largest political parties in the Dutch parliament after the 2012 elections that fit into the 'left-wing' and 'right-wing' taxonomy regarding preferences for redistribution. We ignore the smaller political parties, which might be more special-interest or one-issue parties, and the populist party of Pim Fortuyn, since they did not submit a tax-benefit plan to CPB in the 2002 elections.

Our main findings are as follows. In line with prior expectations, all parties attach a larger social weight to the poor than to the rich. Furthermore, again in line with expectations, we find that left-wing parties give a higher social weight to the poor and a lower social weight to the rich than right-wing parties do. However, we also uncover a number of anomalies. All parties, including the right-wing liberals, attach a negative social weight to the rich. Hence, all parties set the top tax rate beyond the 'Laffer rate' at which the government completely 'soaks the rich'. We argue that political parties underestimated the elasticity of the tax base with respect to top tax rates, since CPB – that judged the party platforms – did as well. Bourguignon and Spadaro (2012) call the resulting negative social welfare weights non-Paretian. Hence, there is a potential Pareto improvement: lower top tax rates can be used to increase the utility of some without reducing the utility of others. We further find that all parties give a lower social weight to the working poor than the non-working poor. None of the political parties proposes to substantially reduce the participation tax rate via an EITC targeted at the working poor. Often, lower marginal tax rates have been advocated to promote labor-force participation. However, lower marginal tax rates do not necessarily reduce participation taxes, which determine participation decisions. Another anomaly is that social welfare weights are increasing from the working poor to the middle-income groups, rather than decreasing. Indeed, in the Netherlands support schemes are phased out at a relatively high and dense part of the income distribution, so as to redistribute income towards the middle incomes. A plausible explanation is that political parties redistribute to middle incomes to attract additional votes. Our analysis suggests that this is particularly relevant for the left-wing parties. They attach relatively more weight to middle incomes than to low incomes. What is also surprising is how close the revealed social welfare weights are across political parties. Clearly, this is the result of rather small differences in tax-benefit systems.

All the anomalies we detected are consistent with important political-economy theories. First, the increasing social welfare weights until the middle-income groups can be understood by standard political models of income redistribution, since the support of middle-income voters is crucial to get elected (Romer, 1975; Roberts, 1977; Meltzer and Richard, 1981). Second, the patterns of the social welfare weights – increasing to modal incomes and sharply decreasing thereafter – are in line with Director's law, where the middle-income groups form a successful coalition against the low-income and high-income groups (Stigler, 1970). Third, the high welfare weights for the middle-income groups could be explained by two-dimensional political competition. Even left-wing parties may sacrifice on their redistributive goals if this helps to achieve larger electoral success by attracting more voters on other ideological positions (Roemer, 1998). Fourth, post-election considerations could explain the strong status-quo bias in announced tax-benefit plans. Political parties may deliberately want to avoid highly pronounced party positions, since they need to form a coalition government with other parties after the elections. Fifth, the status-quo bias and the persistence of various anomalies could also be explained by collective-action problems. Vested interests could be effective in blocking welfare-improving tax-benefit reforms if the benefits of these reforms are dispersed and the costs of the reforms are concentrated at the vested interests (Olson, 1982).

The outline of the paper is as follows. First, in Section 2 we outline the optimal tax model we use in the analysis, and then invert the optimality conditions to get an expression for the implicit social welfare weights. In Section 3 we discuss the calibration of the model and illustrate the inverse method by revealing the social welfare weights in the baseline. In Section 4 we then turn to the political parties. We first give a brief overview of the political parties in the 2002 elections, and outline the reform packages they propose for the tax-benefit system. Next, in Section 5 we present the implicit social welfare weights of the proposed systems. Section 6 offers a number of explanations for the anomalies we uncover. Section 7 concludes.

3.2 Optimal Tax Model and the Optimal-Inverse Method

The optimal-inverse method inverts the famous ABC-formula for optimal tax rates of Diamond (1998) building on the seminal work of Mirrlees (1971). To understand the resulting expressions for the social welfare weights it is instructive to first go through the expressions for optimal marginal tax rates. Furthermore, for clarity we start with a simplified version of the model to illustrate some key mechanisms that will play an important role in the full model as well.

3.2.1 A Simplified Model

First, we consider the expression for optimal marginal tax rates when individual utility is linear in income, features a constant elasticity in earnings supply, and there is no extensive margin. This case was previously studied in Diamond (1998). The full derivation of the model is given in the Appendix.

Optimal Tax Rates for Given Social Preferences for Redistribution

When the utility function is quasi-linear and iso-elastic in labor effort, the expression for optimal marginal tax rates is given by:

$$\frac{T'(z_n)}{1-T'(z_n)} = A_n B_n C_n, \tag{3.1}$$

where:

$$A_n = \frac{1}{\varepsilon}, \tag{3.2}$$

$$B_n = \frac{\int_n^{\overline{n}} (1-g_m) f(m)\,\mathrm{d}m}{1-F(n)}, \tag{3.3}$$

$$C_n = (1+\varepsilon)\frac{(1-F(n))}{nf(n)}, \tag{3.4}$$

where $T'(z_n)$ is the marginal tax rate at income level z_n, n is the earnings ability of an individual of type n, ε is the elasticity of labor supply (or effort), $\overline{n}$ is the top of the earnings ability distribution, $F(n)$ is the cumulative density function of the distribution of earnings abilities and $f(n)$ the corresponding probability density function. Finally, g_n is the social marginal value (in money units) of providing individual n one additional unit of income. Below we refer to g_n as the social welfare weight of an individual with earnings ability n.

A_n shows that when the elasticity of the tax base ε is higher, then optimal marginal tax rates should be lower. Intuitively, the efficiency costs of redistribution go up. B_n is the weighted average of 1 minus the social marginal value of one unit of income for individuals above n. At higher n the gains from redistribution are higher provided g_n falls monotonically. Finally, the C_n term gives weights to the A_n and B_n terms using the distribution of earnings abilities. The more individuals are above n, $1-F(n)$ is large, the larger are the redistributive gains of increasing marginal tax rates at n. However, the larger the tax base at n, $nf(n)$, the larger the efficiency costs of increasing marginal tax rates at n. If we would express the optimal-tax formula in terms of earnings densities, rather than the densities of the ability distribution, the C_n-term would collapse to $C_n = \frac{1-\tilde{F}(z_n)}{\tilde{f}(z_n)z_n}$, where $\tilde{F}(z_n) \equiv F(n)$ is the cumulative earnings distribution, $\tilde{f}(z_n)$ is the earnings density at z_n, and $z_n\tilde{f}(z_n) = (1+\varepsilon)nf(n)$, see Saez (2001). Hence, the C_n term is entirely determined by the shape of the empirical earnings distribution $\tilde{f}(z_n)$.

The famous U-shape result for optimal marginal tax rates from Diamond (1998) comes from the terms C_n and B_n, since he assumes ε to be constant. When the top of the income distribution follows a Pareto distribution then C_n will be constant, and the rise in B_n explains why it is optimal to have rising marginal tax rates towards the top. At low incomes, optimal marginal tax rates should be falling, the numerator $(1 - F(n))$ in C_n is falling, whereas the denominator $nf(n)$ is rising. Indeed, the redistributive gains of high marginal tax rates at low n are relatively large, whereas the corresponding efficiency losses are relatively low.

Inverse Optimal-Tax Method

Next, we invert the ABC-formula to get an expression for the social welfare weights in terms of the other variables. Rewriting (1) gives:

$$\int_n^{\overline{n}} (1 - g_m) f(m) \,\mathrm{d}m = \left(\frac{T'(z_n)}{1 - T'(z_n)}\right)\left(\frac{\varepsilon}{1+\varepsilon}\right) nf(n). \tag{3.5}$$

Taking the derivative with respect to n on the left and right hand side, and suppressing the z_n term in T', we obtain g_n as a function of the other variables:

$$g_n = 1 + \left(\frac{\varepsilon}{1+\varepsilon}\right)\left(\frac{T'}{1-T'}\right)\left(\varepsilon^{\frac{T'}{1-T'}} + \varepsilon^{nf}\right), \tag{3.6}$$

where:

$$\varepsilon^{\frac{T'}{1-T'}} = \frac{\partial\left(\frac{T'}{1-T'}\right)}{\partial n}\frac{n}{\frac{T'}{1-T'}} = \frac{T''}{T'}n, \tag{3.7}$$

$$\varepsilon^{nf} = \frac{\partial\left(nf(n)\right)}{\partial n}\frac{n}{nf(n)} = 1 + n\frac{f'(n)}{f(n)}. \tag{3.8}$$

Note from (3.6) that all the right-hand side parameters and variables can be determined empirically.

In the analysis below, we are particularly interested in whether i) the social welfare weights are monotonically declining in ability (income), so that parties always care more about individuals that have a lower ability, whether ii) the social welfare weights are always positive, otherwise there might be room for a Pareto improvement, and whether iii) there are jumps in the welfare weights, so that there are large differences in social weights for individuals that differ only marginally in their ability level. This too suggests the potential for social welfare-improving reforms. To answer these questions it is instructive to take a closer look at some of these terms at different parts of the ability distribution.

First, we consider whether the social welfare weights will be falling or rising at the lower and upper end of the ability distribution. $\varepsilon^{nf(n)}$ is the elasticity of the tax base with respect to ability. For the lower part of the distribution, where the density is increasing, this elasticity is unambiguously positive. From equation (3.6) we see that, ceteris paribus, high marginal tax rates will lead to high social welfare weights in the lower part of the ability distribution, and vice versa for low marginal tax rates. Suppose that marginal tax rates in the bottom are relatively low, which is what we find for the Netherlands. This implicitly implies that the tax-benefit system cares relatively little about individuals with a relatively low ability. Indeed, the efficiency losses from higher marginal tax rates at low ability are relatively small, the tax base $nf(n)$ is relatively small at low n, but apparently the gains from redistribution from higher ability individuals above n to individuals at n is not big enough to warrant higher marginal tax rates. Next, consider the top part of the ability distribution. When the top part of the distribution follows a Pareto distribution $\frac{nf(n)}{1-F(n)} = a$, with a the (constant) Pareto parameter, then the elasticity $\varepsilon^{nf(n)}$ will be negative and equal to $-a$. From equation (3.6) we then see that, ceteris paribus, rising marginal tax rates will lead to falling social welfare weights in the upper part of the ability distribution.

Second, from (3.6) we can also determine when social welfare weights are positive. This will be particularly relevant for the top of the ability distribution. For the top of the income distribution marginal tax rates are constant and the ability distribution follows a Pareto distribution. In this case the social welfare weight becomes:

$$g_n = 1 - a\left(\frac{\varepsilon}{1+\varepsilon}\right)\left(\frac{T'}{1-T'}\right),$$

where a is the Pareto parameter. Rewriting this condition we find that social welfare weights at the top remain positive as long as:

$$T' < \frac{1+\varepsilon}{1+(1+a)\varepsilon}.$$

With equality this is the ‘Laffer rate’, beyond which an increase in marginal tax rates at the top reduces tax revenue. Indeed, top rates beyond this rate are what Bourguignon and Spadaro (2012) call non-Paretian. In that case, a reduction in the marginal tax rate at the top increases both individual utility at the top, and increases tax receipts, which can be used to make other people better off.

Finally, the term $\varepsilon^{\frac{T'}{1-T'}}$, the elasticity of the tax wedge with respect to ability, is of some interest as well. In particular, this term plays a key role close to spikes in marginal tax rates. Below we will see that some party proposals lead to very high marginal tax rates over very small income intervals. For individuals at an income level in the upward part of the spike the elasticity is very high, and hence the welfare weight is very high as well, ceteris paribus. For individuals at an income level just in the downward part of the spike the elasticity is very low, and hence the welfare weight is very low as well, ceteris paribus. This reflects the fact that the government apparently wants to redistribute income to people just below the spike and wants to take money from people just above the spike. This too seems an anomaly, since then there are large differences in social welfare weights for individuals that differ only slightly in ability.

3.2.2 The Full Model

The simplified model above ignores income effects, which are generally found to be small but not zero, and also ignores the decision whether or not to participate. Indeed, in the Mirrlees model individuals can only adjust their labor supply on the intensive margin. They can decide to work more or less, but they cannot decide to enter or exit the labor market entirely. In contrast, Diamond (1980) derives the optimal tax schedule where individuals can only adjust their labor supply along the extensive margin, but this model cannot handle labor supply responses on the intensive margin. Saez (2002b) and Jacquet *et al.* (2013) combine the Mirrlees model with the Diamond model to analyze the optimal non-linear income tax and the optimal participation tax. We follow the analysis of Jacquet *et al.* (2013) to derive the optimal tax schedule under both intensive and extensive labor supply responses. The full derivation of the model is given in the Appendix, here we only present and discuss the optimality conditions.

Optimal Tax Rates

The extensive margin is introduced using a random participation model. When individuals participate they incur an idiosyncratic utility cost or benefit on top of the changes in income and leisure, denoted by φ. This component can reflect individual specific benefits or costs from non-participation, *e.g.* household production, or from participation, *e.g.* social contacts at work. Like ability, the disutility of participation is unobservable to the government, but the participation decision is observable, together with labor income.

The participation decision of an individual will depend on the participation tax. The participation tax consists of two components. First, when the individual starts working he or she loses benefits b. Second, when the individuals starts working at z he or she faces taxes $T(z)$. The total participation tax is therefore $b + T(z)$.

The resulting ABC-formula for optimal marginal tax rates with both an intensive and extensive decision margin, and with income effects, becomes:

$$\frac{T'(z_n)}{1-T'(z_n)} = A_n B_n C_n, \tag{3.9}$$

as before, but with:

$$A_n = \frac{1}{\varepsilon_n^c}, \tag{3.10}$$

$$B_n = \frac{\int_n^{\overline{n}} \left(\frac{v'(c_n)}{v'(c_m)} \left(1 - g_m^P\right) - \kappa_m \left(b + T\left(z_m\right)\right)\right) \tilde{k}\left(m\right) \mathrm{d}m}{\tilde{K}\left(\overline{n}\right) - \tilde{K}\left(n\right)}, \tag{3.11}$$

$$C_n = (1+\varepsilon_n^u)\frac{(\tilde{K}\left(\overline{n}\right) - \tilde{K}\left(n\right))}{n\tilde{k}\left(n\right)}, \tag{3.12}$$

where ε^c is the compensated labor-supply elasticity with respect to the marginal tax rate T', ε_n^u is uncompensated labor-elasticity with respect to the wage rate n, g_n^P is now the social welfare weight of an employed worker with ability n, κ_m is the (semi-)elasticity of participation, $\tilde{k}\left(n\right)$ is the fraction of employed with ability level n, $\tilde{K}\left(n\right)$ is the fraction of employed in the population with ability n or less, and $\tilde{K}\left(\overline{n}\right)$ is the total fraction of employed.

The term A_n and its interpretation remain unaltered by the introduction of the extensive margin.[8] In term C_n all occurrences of the distribution of earnings ability have been replaced by the distribution of employed workers (non-employed workers do not pay the marginal tax rate). The main difference with the model without an extensive margin is found in term B_n. The extensive margin reduces the average revenue available for redistribution, because a higher marginal tax rate also lowers participation. The amount of revenues lost is determined by the participation elasticity, κ_m, and by the participation tax $T(z) + b$. Term B now is the weighted average of the difference, $1 - g_m^P - \kappa(T(z) + b)$, over all employed workers with an income level above z. Finally, we note that in the utility function with income effects the marginal utility of consumption is not constant, hence the appearance of the term $\frac{v'(c_n)}{v'(c_m)}$ in B_n.

[8]This results from the random participation structure, where idiosyncratic costs/benefits of participation are separable in the utility function.

In the model with an extensive margin the government also optimizes social assistance benefits b. The optimal b is defined by (see the Appendix for the derivation):

$$\int_{\mathcal{N}} \kappa_m (T(z_m) + b)\tilde{k}(m)\mathrm{d}m = \int_{\mathcal{N}} \frac{(1 - g_m^P)}{v'(c_m)}\tilde{k}(m)\mathrm{d}m, \quad \forall n. \tag{3.13}$$

This equation implicitly defines the total, aggregate participation distortion over the entire working population. The left-hand side gives the distortion in participation of providing a higher non-employment benefit b, which is captured by the participation elasticity κ_n, times the participation tax $T(z_n)+b$, aggregated over all households. The right-hand side gives the total distributional benefits of providing higher benefits to non-employed.

The optimal intercept of the tax function $T(0)$ is then implicitly defined by ensuring that the weighted average of the marginal social welfare weights equals one (see the Appendix):

$$\frac{(g_0 - 1)}{v'(b)}(1 - \tilde{K}(\overline{n})) = \int_{\mathcal{N}} \frac{(1 - g_m^P)}{v'(c_n)}\tilde{k}(m)\mathrm{d}m, \tag{3.14}$$

where $g_0 \equiv W'(v(b))v'(b)/\lambda$ denotes the marginal social welfare weight of non-employed individuals. This equation ensures that the marginal euro is valued equally by the public and private sector as all the social welfare weights g_n sum to one. Equivalently, this equation states that the marginal cost of public funds equals one at the optimal tax system. Indeed, distributional benefits of redistribution cancel against deadweight losses at the optimal tax system, see Jacobs (2013).

Inverse Optimal-Tax Method

Again, we can invert the optimality condition for marginal tax rates to recover the social welfare weights. Rearranging (3.9) gives:

$$v'(c_n)\int_n^{\overline{n}} \left(\frac{(1 - g_m^P)}{v'(c_m)} - (b + T(z_m))\kappa_m \right) \tilde{k}(m)\,\mathrm{d}m = \frac{T'(z_n)}{1 - T'(z_n)} \frac{\varepsilon_n^c}{1 + \varepsilon_n^u} n\tilde{k}(n). \tag{3.15}$$

Solve for the social welfare weights by differentiating both sides of the equation with respect to n and rearrange to arrive at:

$$g_n^P = 1 - \varepsilon_n^p + \frac{\varepsilon_n^c}{1 + \varepsilon_n^u} \frac{T'}{1 - T'} \left(\varepsilon^{\frac{T'}{1-T'}} + \varepsilon^{n\tilde{k}} + \varepsilon^{\frac{\varepsilon^c}{1+\varepsilon^u}} - \varepsilon^{v'} \right). \tag{3.16}$$

where $\varepsilon^{\frac{T'}{1-T'}}$ is defined as before, $\varepsilon^{n\tilde{k}}$ replaces ε^{nf} of the intensive model, and:

$$\varepsilon_n^p = \kappa_n v'(c_n)(T(z_n)+b), \tag{3.17}$$

$$\varepsilon^{n\tilde{k}} = \frac{\partial\left(n\tilde{k}(n)\right)}{\partial n}\frac{n}{n\tilde{k}(n)}, \tag{3.18}$$

$$\varepsilon^{\frac{\varepsilon^c}{1+\varepsilon^u}} = \frac{\partial\left(\frac{\varepsilon_n^c}{1+\varepsilon_n^u}\right)}{\partial n}\frac{n}{\frac{\varepsilon_n^c}{1+\varepsilon_n^u}}, \tag{3.19}$$

$$\varepsilon^{v'} = \frac{\partial(v'(n))}{\partial n}\frac{n}{v'(n)}. \tag{3.20}$$

New is ε^p, which is the (semi-)elasticity of participation with respect to the participation tax. For a given optimal average participation tax rate, the higher the participation elasticity, the lower the implied welfare weight given to participants.

$\varepsilon^{\frac{\varepsilon^c}{1+\varepsilon^u}}$ is also new, it is the elasticity of the efficiency cost of taxation with respect to ability. For a given tax rate an increase in the efficiency cost implies that individuals below n receive a higher welfare weight than individuals above n. Intuitively, the government does not tax individuals at n at a higher level, even though the efficiency cost of taxing these individuals is smaller than the efficiency cost of individuals with an ability above n.

Finally, $\varepsilon^{v'}$ denotes the elasticity of the marginal utility of consumption with respect to ability. This expression corrects for income effects. If $\varepsilon^{v'}$ is large this implies that the marginal utility of consumption decreases a lot with ability.

The final part of the model is the welfare weight given to the non-participants g_0. To obtain this welfare weight rewrite (3.14) to get:

$$g_0 = 1 + \frac{v'(b)}{\left(1-\tilde{K}(\overline{n})\right)}\int_{\underline{n}}^{\overline{n}}\frac{\left(1-g_n^P\right)}{v'(c_n)}\tilde{k}(n)\,\mathrm{d}n. \tag{3.21}$$

3.3 Calibration and Social Welfare Weights in the Baseline

To illustrate the method we derive the social welfare weights implicit in the actual 2002 tax-benefit system. Using data on the income distribution, marginal tax rates, employment rates, and recent estimates of the elasticity of the tax base on both the extensive (participation) and intensive (hours per week, or effort more generally) margin we first recover the underlying ability distribution of workers and the distribution of idiosyncratic participation costs/benefits. The ability distribution and the participation cost/benefit distribution, together with the marginal tax rates, then determine the social welfare weights in the baseline.

3.3.1 The Income Distribution and Marginal Tax Rates

We define income as gross wages, and marginal taxes as the difference between a) the increase in gross wages and b) the increase in net disposable income net of indirect taxes, over gross initial income when we increase gross wages by 3%. The income data are from the *Inkomenspanelonderzoek 2002*, which contains detailed information on household characteristics necessary to calculate marginal tax rates corresponding to the individual incomes. Figure 3.1 plots a kernel density estimate of the income distribution. We have relatively few observations on the top tail of the income distribution. In Zoutman *et al.* (2013a) we study the top income distribution more closely. We find that the Pareto distribution gives an excellent fit to the top income distribution. Moreover, the Pareto distribution is estimated to start at approximately the top tax bracket, which contains the top 8% of incomes. For gross wages we estimate a Pareto parameter of 3.2. This rather high value compared to other countries indicates that it is lonely at the top in the Netherlands. The estimated Pareto parameter is in line with other studies using Dutch data, see Atkinson and Salverda (2005) and Atkinson *et al.* (2011).

The marginal tax rates are calculated using the tax-benefit calculator MIMOS-2 of CPB. MIMOS-2 takes into account all income-dependent subsidies and tax credits to calculate so-called effective marginal tax rates. We then calculate the effective total marginal tax rates by including (marginal) indirect taxes on marginal income net of direct taxes.[9] Figure 3.2 gives the kernel estimate for the corresponding effective marginal tax rates in the Dutch income distribution for all workers. Figure B.1 in the appendix gives a scatterplot of the marginal tax rates. There is quite some variation in these marginal tax rates at each income level, in particular for lower incomes. However, the model only works with a single marginal tax rate at each income level. Therefore we use a kernel estimate to smooth out the variation in individual marginal tax rates at each income level, and across individuals at different income levels.

To understand the patterns in Figure 3.2, Table 1 gives some parameters of the Dutch tax system in 2002. In 2002, the Dutch tax system had four tax brackets for labor income, based on individual (not household) income, with rates rising from somewhat below 33% at the bottom to 52% at the top. This explains why marginal tax rates are typically lower for individuals with low income than for individuals with high income.

[9] Denote the effective direct marginal tax rate by t_d, the marginal indirect tax rate by t_i and the effective marginal tax rate by t_e. We calculate the effective total marginal tax rate as $t_e = \frac{t_d + t_i}{1 + t_i}$.

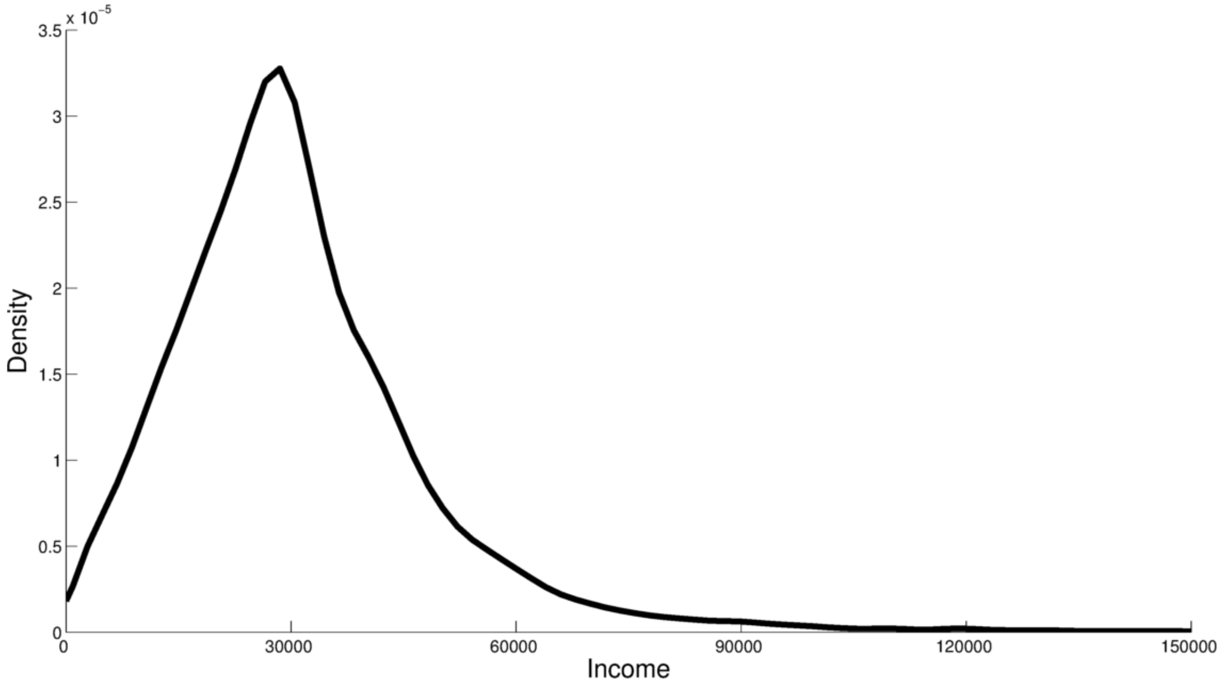


Figure 3.1: Kernel Density Estimate of Gross Wage Income in the Netherlands, 2002

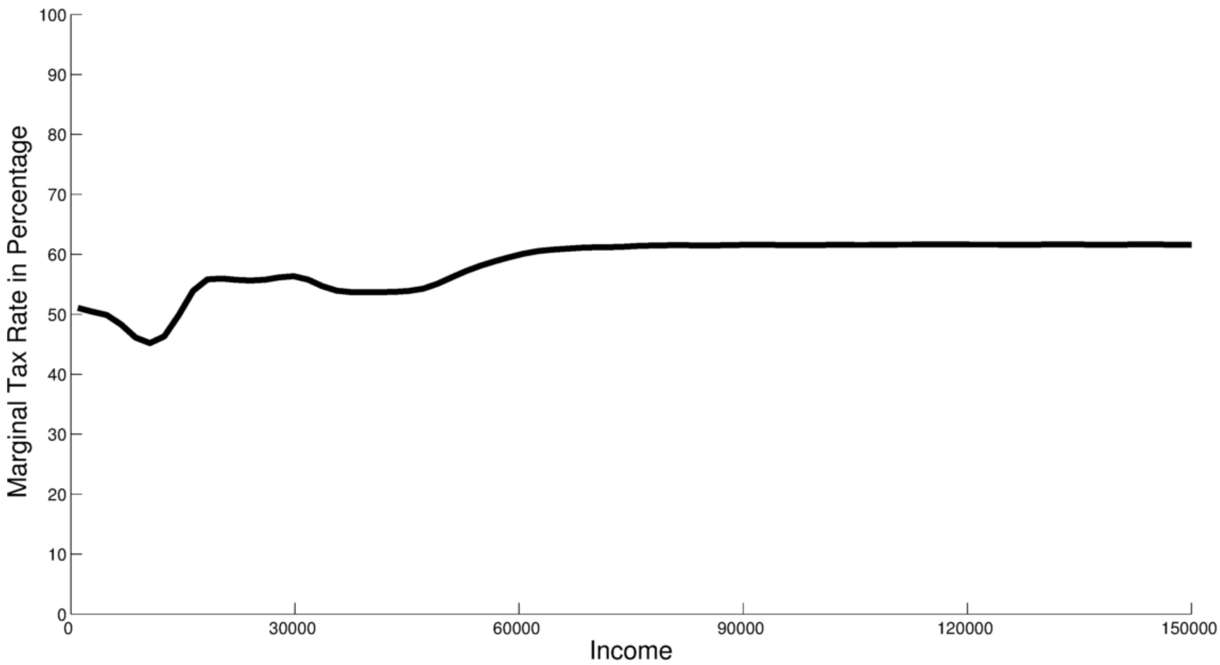


Figure 3.2: Kernel Density Estimate of Total Effective Marginal Tax Rates in the Netherlands, 2002

Table 3.1: Tax Brackets and Tax Credits in 2002

	Start	End	Percentage	Maximum amount
Tax brackets				
First tax bracket	0	15,331	32.35	4,960
Second tax bracket	15,331	27,847	37.85	4,737
Third tax bracket	27,847	47,745	42.00	8,357
Fourth tax bracket	47,745	∞	52.00	∞
Tax credits				
General tax credit	0	∞	0	1,647
Earned-income tax credit				
- First part	0	7,692	1.73	133
- Second part	7,692	15,375	10.62	949
Single parent tax credit	0	∞	0	1,301
Earned-income single-parent tax credit	0	30,256	4.30	1,301

There are also a number of noticeable deviations from these rates, which result from targeted subsidies and tax credits. For the lowest incomes, marginal tax rates are initially higher than the first tax bracket because a number of income-support schemes are phased out with income, in particular rent subsidies and a general child tax credit.[10] Then, there is a segment where marginal tax rates are lower due to the phase-in of the earned income tax credit (EITC). The end of the phase-in range for the EITC (almost) coincides with the start of the second tax bracket (about 15,000 euro), and after that marginal tax rates rise substantially up to some 40,000 euro.[11] Finally, total marginal tax rates are higher than direct marginal tax rates because of indirect taxes. Using publicly available input-output tables of Statistics Netherlands we calculate that indirect taxes on private consumption were 11.7% of private consumption in 2002. We assume that these indirect taxes are proportional to net income. Bettendorf *et al.* (2012) show that indirect taxes are close to proportional to consumption in the Netherlands.

Table 3.2: Elasticities Used in the Simulation

	Compensated wage elasticity	Income elasticity	Uncompensated wage elasticity	Participation elasticity
Baseline scenario	0.35	0.10	0.25	0.25
Low-elasticity scenario	0.18	0.05	0.13	0.13
High-elasticity scenario	0.53	0.15	0.38	0.38

Table 3.3: Employment Rates by Level of Education

Level of education	Net employment rate	Share in population
Elementary school	36.90	11.99
Some high school	53.50	25.79
High school	56.80	10.26
Low-level college	71.20	15.84
Mid-level college	79.10	14.95
Bachelor degree	80.40	13.88
Master degree or higher	84.40	7.28

Table 3.4: Calibrated Parameters for the Utility Function

Parameter values	Base	Low Elasticity	High Elasticity
α	0.46	0.48	0.45
ε	0.38	0.18	0.60
γ	1981.67	13503.12	1082.11
μ_k	55.95	0.00	82.42
σ_k	271.27	511.00	189.98

3.3.2 Utility Function and Elasticity of the Tax Base

For the behavioral responses we use the following utility function that can reproduce the observed tax base elasticities:[12]:

$$u = \frac{c^{1-\alpha}}{1-\alpha} - \gamma \frac{l^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}, \qquad \alpha, \gamma, \varepsilon > 0. \tag{3.22}$$

[10]The exact subsidy levels and taper rates vary with household characteristics other than income, and are therefore not reported in Table 3.1.

[11]Actually, there is an additional jump for individuals with a gross income close to 40 thousand euro. Individuals below a threshold could enter the public health insurance scheme with relatively low premium rates, whereas individuals above this threshold are forced to take private health insurance with relatively high premium rates. For some households close to the threshold this results in very high marginal tax rates. In 2003 this health care system has been replaced by a uniform, obligatory basic health insurance scheme, which is financed by a payroll tax and 'lump-sum' premiums paid by individuals. Individuals can voluntarily top up the basic health insurance scheme with additional insurance packages.

[12]This utility function is also used by Mankiw *et al.* (2009). When $\alpha \to 1$ this utility function converges to the logarithmic Utility Type-II used by Saez (2001). When $\alpha = \frac{1}{\varepsilon}$ this specification is in line with the CES-functions used by Mirrlees (1971) and Tuomala (1984).

α and ε are calibrated so as to match the compensated and uncompensated elasticities, discussed below.[13] Parameter γ is an innocuous scaling parameter we calibrate to keep the mean of the ability distribution fixed in the different scenarios (*e.g.* different elasticities) that we consider.

To calibrate the parameters of the utility function we employ empirical estimates of the elasticity of the tax base. We follow Zoutman *et al.* (2013a) and base the extensive-margin elasticity on recent estimates for the labor-supply elasticity of Mastrogiacomo *et al.* (2013) on the extensive margin. Mastrogiacomo *et al.* (2013) estimate extensive margin labor-supply elasticities for individuals in a number of different household types. The weighted average of the extensive margin elasticities over these household types is 0.25, which is our target value for the extensive-margin elasticity. However, we also want to match the participation rates by skill level, given in Table 3.3. Our algorithm optimizes the parameters of the participation-cost distribution to minimize the distance between the predicted and observed participation rates and the predicted and observed extensive-margin elasticity. See Zoutman *et al.* (2013a) for further details.

On the intensive margin we use recent estimates of the elasticity of taxable income in the Netherlands presented in Jongen and Stoel (2013b). The elasticity of taxable income captures responses in hours worked, but also other responses to changes in marginal tax rates. We prefer this broader concept of the intensive margin response to the more narrow concept of hours worked per worker. The elasticity of taxable income is estimated to be (approximately) 0.25 as well. This is an uncompensated elasticity. Based on the few estimates that are available in the literature, we calibrate the income elasticity to 0.10 on average, resulting in an average compensated intensive margin elasticity of 0.35. Table 3.2 gives an overview of our preferred elasticities in the baseline, and the elasticities we use in a sensitivity analysis where we decrease (low-elasticity scenario) or increase (high-elasticity scenario) the elasticity of the tax base. Table 3.4 gives the corresponding preference parameters.

Conditional upon participation in the labor market, and using the definition of gross labor earnings $z \equiv nl$, we can invert the first order condition for optimal labor supply (19) in the Appendix, for the utility functions above, to express ability n as a function of

[13] It turns out that as long as the ratio $\varepsilon^c/\varepsilon^u$ is fixed, the calibrated α is almost similar for different elasticities. This is a useful property, since the decrease in marginal utility of consumption is then similar, and we can isolate the effect of a change in the elasticities from the redistributional concerns. All our scenarios have the same ratio $\varepsilon^c/\varepsilon^u$.

marginal tax rates and income:

$$n = \left(\frac{\gamma z^{\frac{1}{\varepsilon}}}{(1 - T'(z))\, c^{-\alpha}} \right)^{\frac{\varepsilon}{\varepsilon+1}}. \tag{3.23}$$

The ability distribution follows the basic shape of the income distribution, but of course with some differences due to *e.g.* the differences in marginal tax rates. For instance, high incomes are 'pushed to the right' when we map their income into the ability distribution.

3.3.3 Government Budget Constraint

Finally, we assume that the government has to collect 9.5% of total output to finance government consumption (the benefits of which we ignore in the utility function for simplicity).[14] This is the sum of expenditures on public administration, police, justice, defense and infrastructure minus non-tax revenues (from *e.g.* the sales of natural gas) as a percentage of GDP in 2002 (CPB, 2010a, Annex 9). With the government revenue requirement set at 9.5% of output, the tax system is budgetary neutral with a social assistance level of approximately 12,000 euro. This is somewhat higher than the current level of net welfare benefits in 2002 amounting to 9,014 euro for a single-person household. However, we ignored some other forms of social assistance at the local level ('Bijzondere Bijstand'), exemptions from local taxes, and transfers in kind (discounts for arts, public transport, etc.), training, public employment, and labor-market programs, which also act as support schemes for the non-employed.

3.3.4 Social Welfare Weights in the Baseline

Figure 3.3 gives the resulting social welfare weights in the baseline, using equations (3.16) for the employed and (3.21) for the non-employed. Figure 3.4 gives the key determinants of the welfare weights for the employed from equation (3.16).

Some results are in line with a 'standard' social welfare function. Welfare weights are generally higher for low income individuals than for high income individuals. Indeed, the non-employed get the highest welfare weight, and the top incomes get the lowest welfare weights.

However, we also find some anomalies. First, we find that the working poor have a much lower social welfare weight than the non-working poor. Hence, participation tax rates in the baseline appear too high for the lowest income groups. Second, the social

[14] Tuomala (2010) uses a similar share of total output.

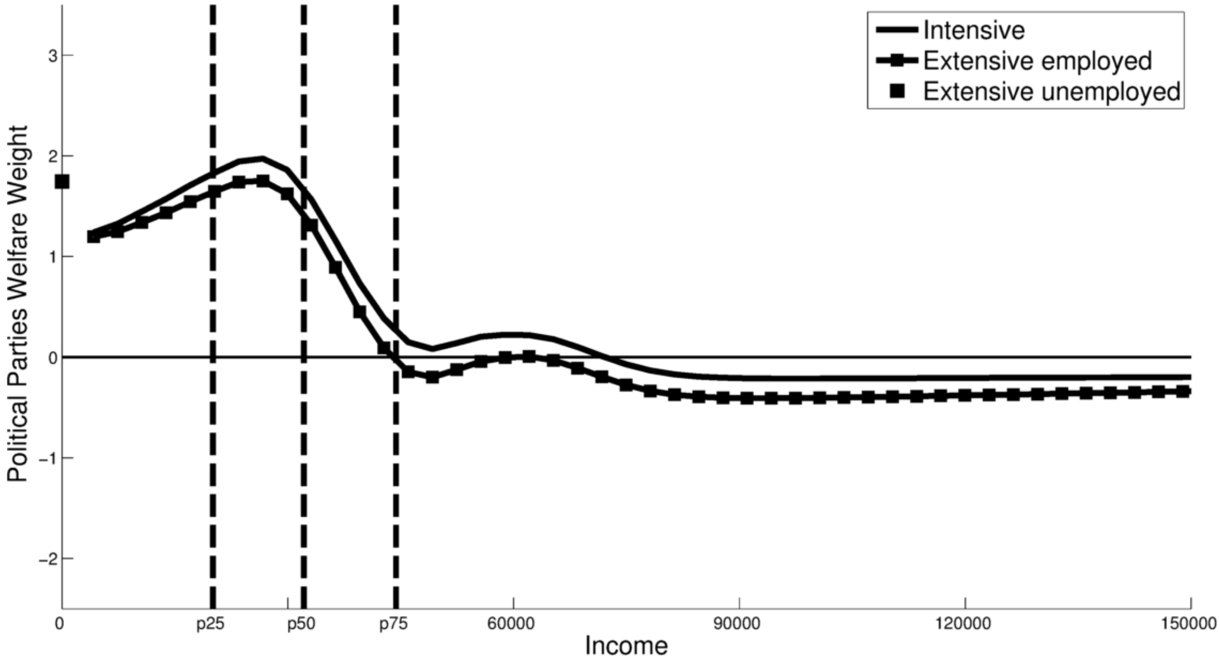


Figure 3.3: Social Welfare Weights in the Baseline

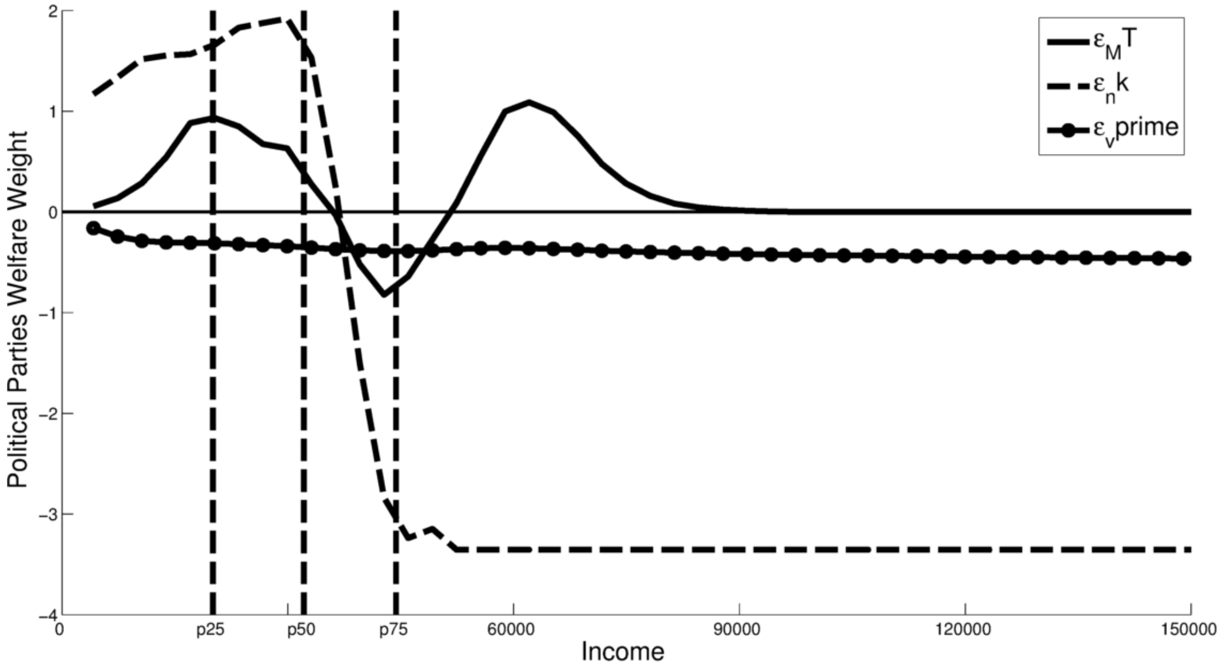


Figure 3.4: Determinants of Social Welfare Weights in the Baseline

welfare weights are not monotonically declining. Specifically, the working poor get a lower weight than workers with a median income. This is because marginal tax rates are rather low and rising at low incomes, which is reflected in the term $\varepsilon\frac{T'}{1-T'}$ in figure 3.4. Hence, there is 'too much' redistribution towards middle-income groups at the expense of low- and high-income groups. Also, higher up the income distribution, close to 60 thousand euro, welfare weights rise again somewhat with income. This is because of the sudden drop and subsequent rise in marginal tax rates before and after 60 thousand euro respectively, reflected again in the term $\varepsilon\frac{T'}{1-T'}$ in Figure 3.4. Third, for top incomes the welfare weights are negative because the tax rate in the top bracket is set beyond the 'Laffer rate' which maximizes tax revenue at the top.

These anomalies are actually in line with the findings of related studies on other countries. Bourguignon and Spadaro (2012) and Bargain *et al.* (2011) also find relatively low social welfare weights for the working poor, whereas Bourguignon and Spadaro (2012) also find negative social welfare weights for the top incomes in France.

Below we consider the social welfare weights implicit in the proposed tax-benefit system by political parties for the 2002 elections, to determine e.g. who they care about most, how much the welfare weights differ across the political parties and to what extent the anomalies above are mitigated, or amplified by their proposals.

3.4 Tax-Benefit Systems Proposed by Political Parties

Having outlined the method, we now take a closer look at the reform proposals for the tax-benefit system of the political parties. We study the proposals for the 2002 elections. We start with a short introduction to the participating political parties in the 2002 elections, and subsequently consider the proposals of the four biggest parties (in the 2010 elections) in more detail.

3.4.1 Political Parties in 2002

Table 3.5 provides an overview of the political parties that received votes in the 2002 elections. The table is taken from Graafland and Ros (2003), and for additional perspective we added the result of the most recent elections, in 2012. The parties are ordered from top to bottom according to their seats in parliament before the elections in 2002 (the period 1998-2002). The Dutch parliament contains 150 seats. The seats are awarded through a system of party-list proportional representation. That is, if a party gets $x\%$ of the votes in the country it is awarded with $1.5x$ seats.

Preceding the 2002 elections were two periods of so-called 'purple' cabinets (Kok-I from 1994-1998, and Kok-II from 1998-2002, named after prime minister Wim Kok), consisting of the 'left' oriented $PvdA$, 'right' oriented VVD and the smaller progressive democrats $D66$. They had a whopping 97 of a total of 150 seats in parliament preceding the 2002 elections. However, in a short period of time Pim Fortuyn and his populist party LPF became very popular indeed. Pim Fortuyn himself was murdered in the run up to the 2002 elections, but his party still got 26 seats in parliament following the 2002 elections. They formed a coalition together with CDA and VVD, which fell apart less than one year later, and the 'traditional' parties CDA, VVD and $D66$ formed a new coalition. Since the beginning of the century many coalitions have been unstable. In 2013, the ruling coalition consists of VVD and $PvdA$. In the analysis below on the four largest political parties in the Dutch parliament after the 2012 elections that fit into the 'left-wing' and 'right-wing' taxonomy regarding preferences for redistribution. We ignore the smaller political parties, which might be more special-interest or one-issue parties, and the populist party of Pim Fortuyn, they did not submit a tax-benefit plan to CPB in the 2002 elections. We then consider, from 'left' to 'right': the left-wing socialist party SP, the social democratic party $PvdA$, the christian democratic party CDA and the liberal conservative party VVD.

3.4.2 Reform Proposals

To determine the social welfare weights of the political parties we make use of the reform packages submitted for analysis to the CPB in 2002.[15] Clearly, the proposed policy reforms are not all related to redistribution. Below we outline the key policy changes that seem most relevant for our analysis, the proposed changes in direct taxes, indirect taxes, corporate taxes and benefits.

SP

First consider the proposed changes in direct taxes by the socialist party SP. The SP abolishes health-care premiums. To finance this operation, the SP raises the tax rate in the first tax bracket by 2.3 percentage points, and the tax rate in the second, third and fourth tax bracket by 3 percentage points. As part of the initial health-care premiums

[15] CPB (2002b) gives an extensive overview of the proposed policy changes and the resulting effects in Dutch. A brief English summary can be found in CPB (2002a).

Table 3.5: Political Parties in the 2002 National Elections[a]

Name	Acronym	Profile	Seats before 2002 election	Seats after 2002 election	Seats after 2012 election
Partij van de Arbeid	PvdA	Social democrat	45	23	38
Volkspartij voor Vrijheid en Democratie	VVD	Conservative liberal	38	24	41
Christen Democratisch Appèl	CDA	Christian democrat	29	43	13
Democraten 66	D66	Social liberal	14	7	20
GroenLinks	GL	Environmental progressive	11	10	4
Socialistische Partij	SP	Socialist (left wing)	5	9	15
ChristenUnie	CU	Protestant orthodox	5	4	5
Staatkundig Gereformeerde Partij	SGP	Protestant orthodox	3	2	3
Lijst Pim Fortuyn	LPF	Anti political establishment	-	26	-
Leefbaar Nederland	LN	Anti political establishment	-	2	-

[a]Source: Graafland and Ros (2003) and www.tweedekamer.nl.

are a fixed amount, and hence do not enter effective marginal tax rates, this operation in part increases marginal tax rates. The SP also introduces a fifth tax bracket for income above 213,000 euro, with a rate of 72%. The SP further introduces an additional earned income tax credit (EITC), which is phased in up to the annual minimum wage (15,800 euro in 2002), with a maximum of 1,017 euro, and is phased out between 130% (20,540 euro) and 170% (26,860 euro) of the annual minimum wage. Finally, the SP makes the general subsidy per child per year (*Kinderbijslag*) income dependent. Lower incomes receive a higher general subsidy per child per year, at 45,000 euro the subsidy is cut in half, and at 90,000 euro the subsidy is abolished. This leads to high marginal tax rates for individuals close to these thresholds (see below).

Next to changes in direct taxes, the SP raises environmental levies (2.6 bln euro), and raises corporate taxation (3.5 bln euro). We incorporate the additional environmental levies in indirect taxes. Furthermore, assuming that capital is mobile and labor is immobile, we assume that the increase in corporate taxes is passed on entirely to wages. For simplicity we incorporate this as a percentage-point rise in all marginal tax rates at all income levels.

Finally, regarding benefits to the non-employed, the SP wants to raise social assitance benefits by 5%.

Figure 3.5 gives the resulting kernel of effective marginal tax rates for the SP[16], and for comparison we also include the baseline. Higher indirect taxes and corporate taxes increase effective marginal tax rates across the board. The phase-in of the EITC somewhat limits the rise in marginal tax rates at the bottom, but the phase-out range leads to a significant rise in marginal tax rates around 25,000 euro. We also clearly see the additional spike at 90,000 euro where child subsidy is abolished. This leads to very high marginal tax rates for some households close to the threshold, which still show up in the smoothed kernel estimate.[17] We only plot incomes up to 150,000 euro, otherwise it becomes hard to discern what happens for the major part of the income distribution. However, this implies that we do not show that marginal tax rates jump up dramatically beyond 213,000 euro, where the new 72%-bracket kicks in.

PvdA

Next, we consider the tax-benefit reform proposed by the social-democratic party *PvdA*. Again, we start with the proposed changes in direct taxes. The *PvdA* integrates public

[16]Figure B.2 in the appendix gives a scatterplot of the individual marginal tax rates before we apply the kernel estimator.

[17]The spike at 45 thousand euro is less clear in the smoothed kernel, but is clearly visible in the scatterplot Figure B.2 in the appendix.

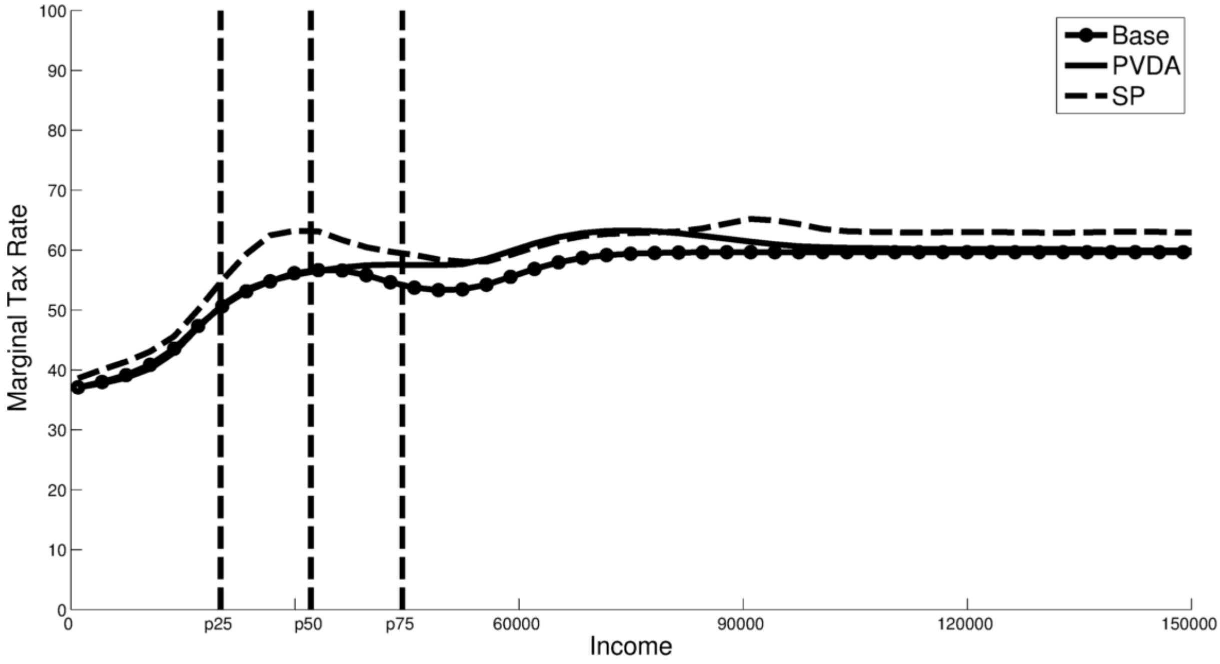


Figure 3.5: Kernel of Marginal Tax Rates: Social Democratic *PvdA* and (Left-Wing) Socialists *SP*

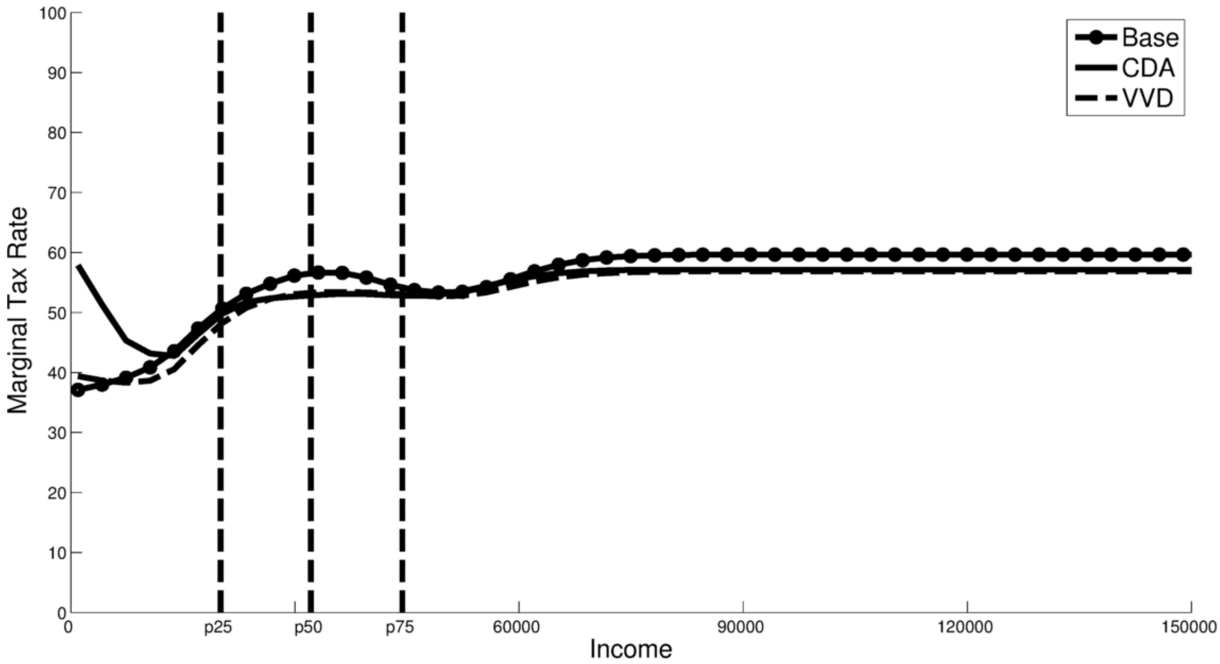


Figure 3.6: Kernel of Marginal Tax Rates: Christian Democratic *CDA* and Liberal Conservative *VVD*

and private health insurance. They raise health-insurance premiums, but reduce the first tax bracket by 1.5 percentage points to compensate the lowest incomes. Overall, marginal tax rates hardly change in this operation. The *PvdA* introduces an additional EITC, with a phase-in range between 90 and 100% of the minimum wage, a maximum of 353 euro (much lower than the *SP* above), and a phase-out range between 180 and 240% of the minimum wage. In addition, they phase-out the pre-existing EITC between 240 and 400% of the minimum wage.

The *PvdA* raises indirect taxes via environmental levies (3.5 bln euro), and lowers corporate taxes (–1.4 bln euro), which we incoporate via a percentage-point drop in marginal tax rates across the board.

The party leaves the level of social assistance benefits basically unchanged.

Figure 3.5 shows the kernel of resulting effective marginal tax rates for the *PvdA*[18], along with the baseline and the *SP*. Higher indirect taxes increase effective marginal tax rates across the board, which is somewhat mitigated by the reduction in corporate tax rates. The phase-out of the additional EITC (between 180 and 240% of the minimum wage), and subsequently the pre-existing EITC (between 240 and 400% of the minimum wage) leads to a rise in marginal tax rates over a long range beyond 30,000 euro.

CDA

We then move on to the more conservative christian democratic party *CDA*. The *CDA* also integrates public and private health insurance. The *CDA* increases the tax rates in the first and second bracket by 1.3 percentage points. Furthermore, they lower the starting point of the fourth tax bracket by 4,440 euro (from 47,745 euro to 43,305 euro), effectively raising marginal tax rates over this income range. The receipts of the increases in the first, second and third tax bracket are used to introduce an income-dependent subsidy for health-care costs and an income-dependent subsidy for children, both targeted at low-income families. The CDA reduces the effective top rate by 1.9 percentage points. Furthermore, the *CDA* introduces a small additional EITC, with a maximum of 72 euro, which is not phased-out (as opposed to the left-wing parties).

The *CDA* leaves indirect taxes, corporate taxes and the level of social assistance benefits virtually unchanged.

Figure 3.6 gives the resulting effective marginal tax rates for the *CDA*, and the baseline for comparison.[19] We see a noticeable increase in marginal tax rates for the lowest incomes,

[18]Figure B.3 in the appendix gives the scatterplot of individual marginal tax rates before we apply the kernel estimator.

[19]Figure B.4 in the appendix gives the scatterplot of the individual marginal tax rates before we apply the kernel estimator.

which is the result of the phasing-out of health-care subsidies and subsidies for parents with low incomes. Furthermore, the reduction in the top rate reduces marginal tax rates for top incomes.

VVD

The conservative-liberal party VVD reduces the tax rate in the first tax bracket by 0.4 percentage points. The VVD also introduces an EITC, with a maximum of 232 euro, and like the CDA they do not phase it out. The VVD reduces the top tax rate more than the CDA, by 3 percentage points.

The VVD slightly increases indirect taxes via environmental levies (0.2 bln euro), reduces corporate taxes (–2.3 bln euro), and leaves the level of social assistance benefits virtually unchanged.

Figure 3.6 gives the resulting effective marginal tax rates for the VVD.[20] Except for the lowest incomes, the reform package of the liberals reduces marginal tax rates, in particular for individuals with a high income.

3.5 Social Welfare Weights of Political Parties

In Section 3 we recovered the distributions of ability and idiosyncratic participation costs/benefits. Furthermore, we calibrated the individual utility functions to reproduce the elasticity of the tax base on the extensive and intensive margin. We now combine these distributions and individual utility functions with the effective marginal tax rates of the political parties outlined above, again using equations (3.16) for the employed and (3.21) for the non-employed.

3.5.1 Left-Wing Parties

Figure 3.7 gives the resulting implicit welfare weights for the left-wing parties SP and $PvdA$, and for comparison the implicit welfare weights of the actual system in 2002. The differences are due to the differences in the proposed marginal tax rates and social assistance benefits. Figure 3.8 gives the elasticity of the marginal tax rate for the employed from equation (3.16), which is helpful in explaining the resulting differences in the social welfare weights.

[20]Figure B.5 in the appendix gives a scatterplot of the individual marginal tax rates before we apply the kernel estimator.

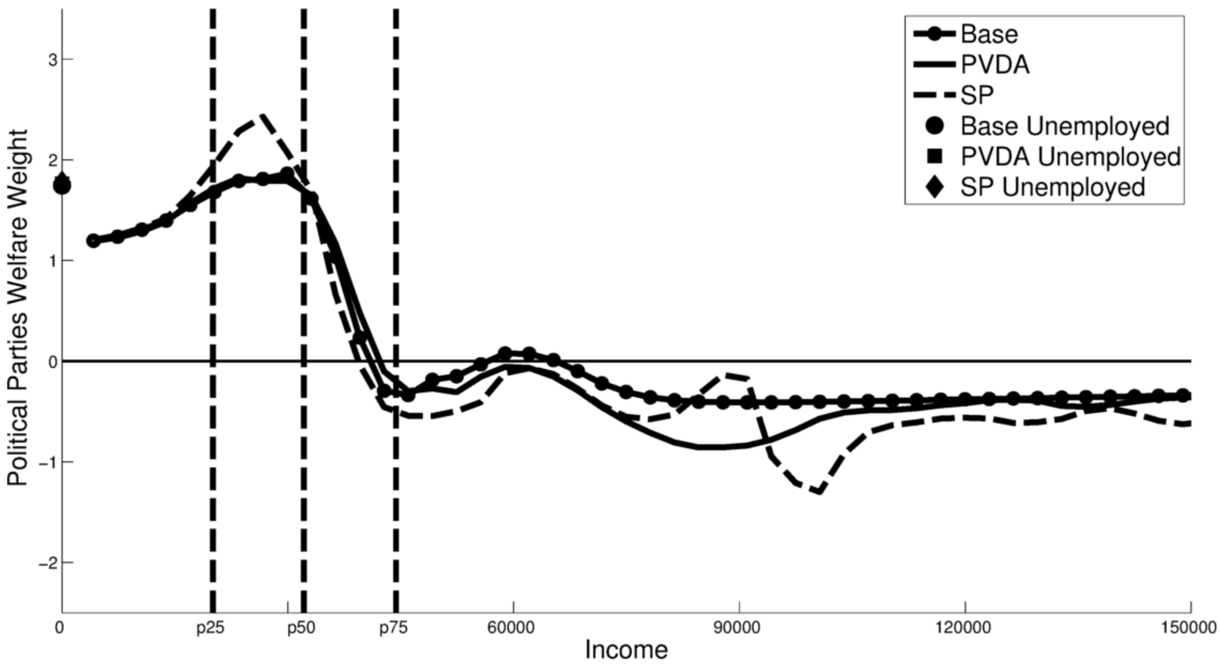


Figure 3.7: Social Welfare Weights: Social Democratic *PvdA* and (Left-Wing) Socialists *SP*

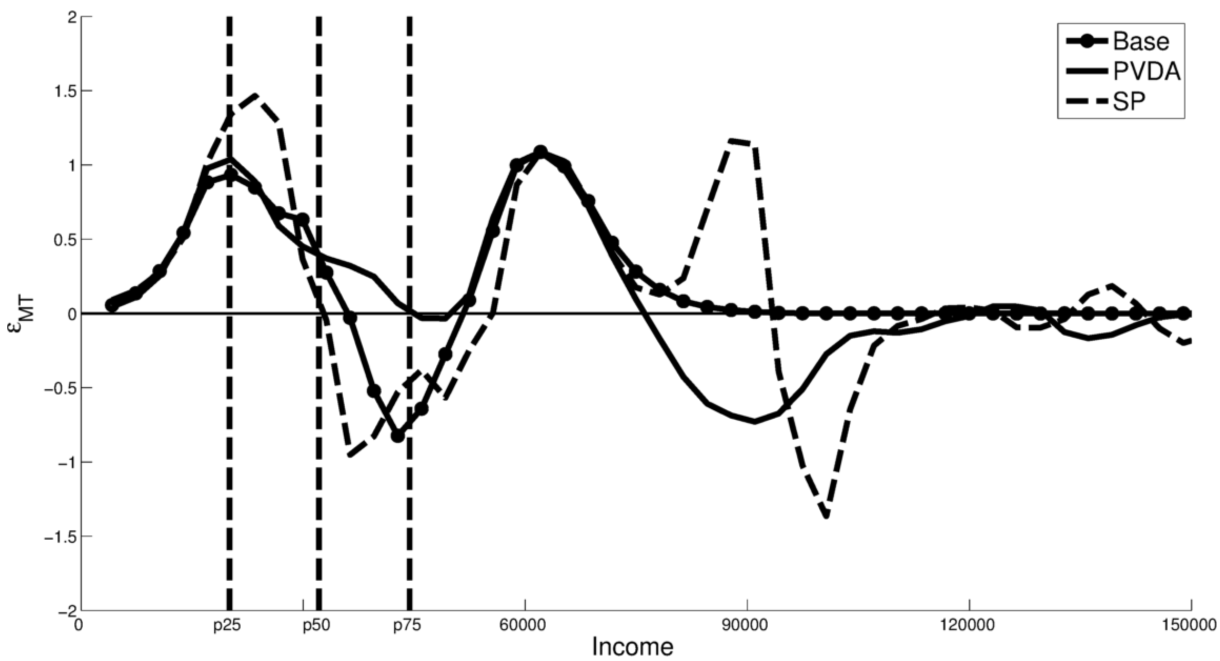


Figure 3.8: Elasticity of Marginal Tax Rates: Social Democratic *PvdA* and (Left-Wing) Socialists *SP*

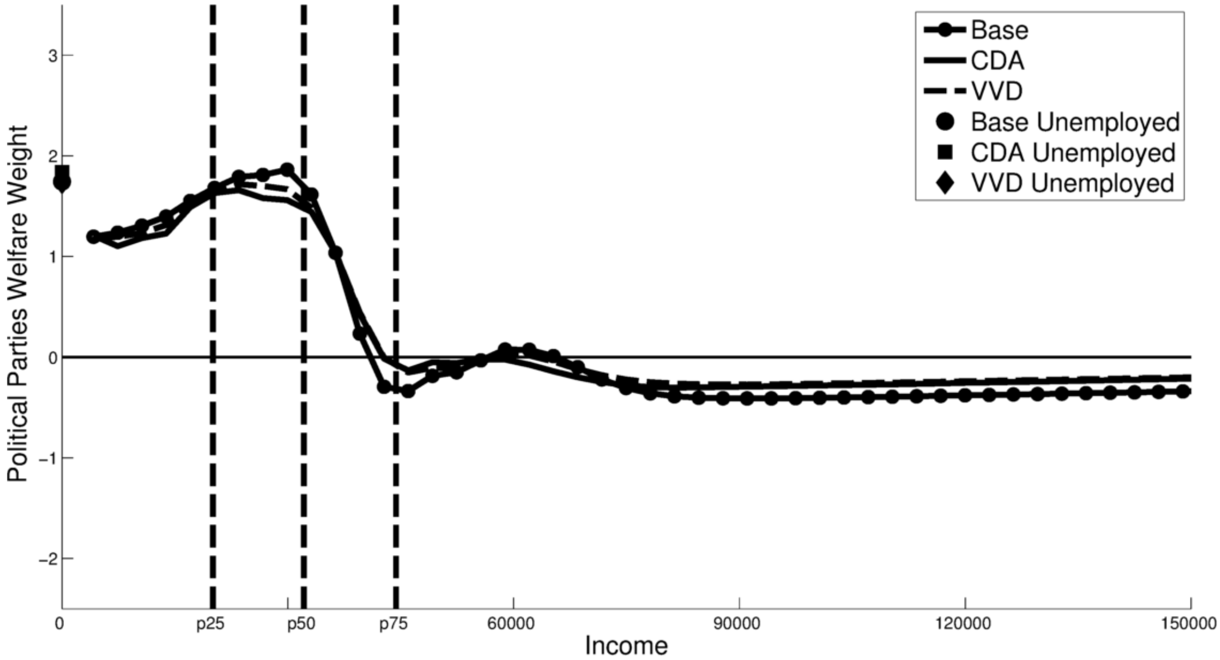


Figure 3.9: Social Welfare Weights: Christian Democratic *CDA* and Liberal Conservative *VVD*

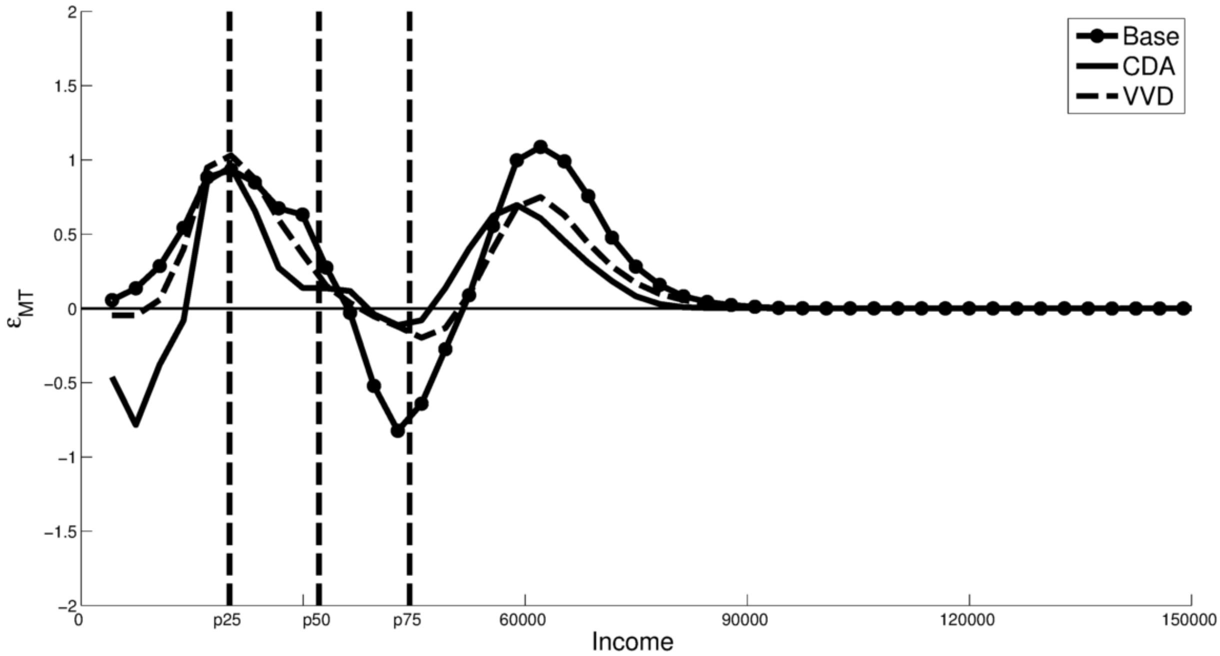


Figure 3.10: Elasticity of Marginal Tax Rates: Christian Democratic *CDA* and Liberal Conservative *VVD*

As we might expect, the left-wing parties grosso modo give a higher weight to the poor than the rich. However, the SP increases the anomaly in the baseline of giving more weight to incomes close to middle incomes than the working poor. This is the result of the phase-out of the EITC just above the minimum wage, which is also reflected in the rise in the elasticity of the marginal tax rate before 25 thousand euro in Figure 3.8. Indeed, the proposed changes by the SP in the tax-benefit systems seem to favor the middle incomes the most. This perhaps suggests that the SP deviates from its ideological preference, so as to attract votes from the large mass of voters in the middle.

For high incomes, the proposals of the left-wing parties exacerbate the pre-existing anomaly of negative social welfare weights, pushing the top rate further beyond the Laffer rate. Presuming that these parties do not have non-Paretian preferences, this suggests that they overestimate the thickness of the Pareto tail of the income distribution, or they underestimate the elasticity of the top tax base with respect to marginal tax rates.

For the $PvdA$ it is also surprising that they give more weight to the very rich than the rich, weights increase between 80,000 and 110,000 euro. This is the result of the phasing out of the EITC and the higher health-care premiums, which lower the weights of the middle- to high-income earners.

Finally, for the SP we see a spike close to 90,000 euro that mirrors the spike in the marginal tax rate, also visible from the marginal tax rate elasticity in Figure 3.8.[21] When a party phases out a subsidy from one euro to the next it apparently cares much more about the person just before the threshold, and much less about the person just after it. This seems rather odd, but of course this just reflects the sudden withdrawal of a subsidy, instead of a smooth phase-out.

3.5.2 Right-Wing Parties

Figure 3.9 shows the implicit social welfare weights for the proposed tax-benefit systems of the right-wing parties CDA and VVD. The first thing that strikes us is that their redistributive preferences are almost identical. Indeed, going back to Figure 3.7, we see that the proposed tax rates are quite similar. At the lower end we see that the CDA gives a little bit more weight to the lowest incomes, but otherwise similar weights as the baseline, as does the VVD. Hence, the proposals of these parties do not exacerbate, but still prolongate the anomaly of giving more weight to middle incomes than to the working poor.

[21]The spike from the withdrawal of part of t he child subsidy at 45,000 euro is less prononounced, but still visible in Figure 3.8.

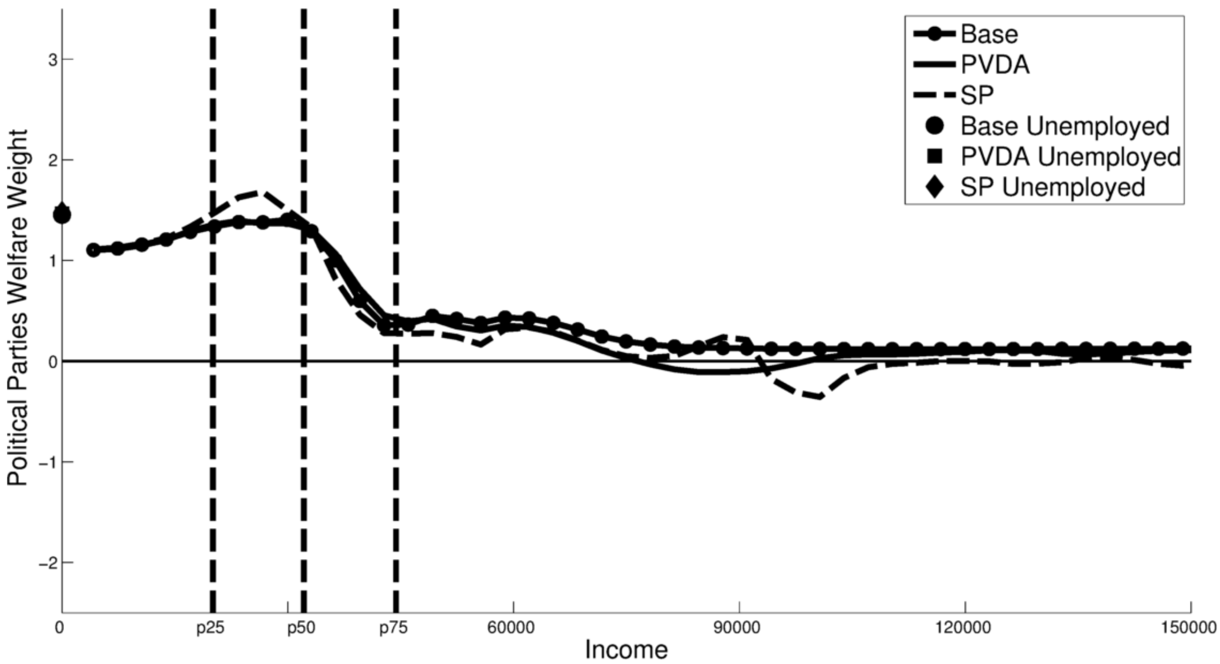


Figure 3.11: SWW Lower Elasticity of Tax Base: Social Democratic *PvdA* and (Left-Wing) Socialists *SP*

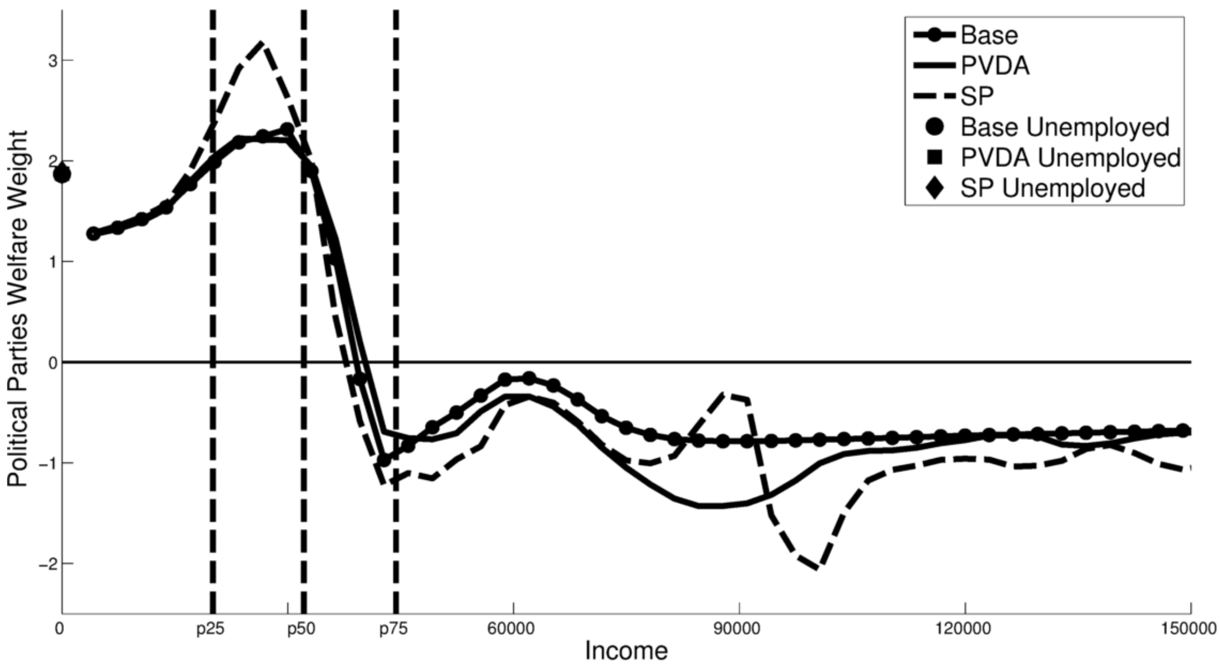


Figure 3.12: SWW Higher Elast. of Tax Base: Social Democratic *PvdA* and (Left-Wing) Socialists *SP*

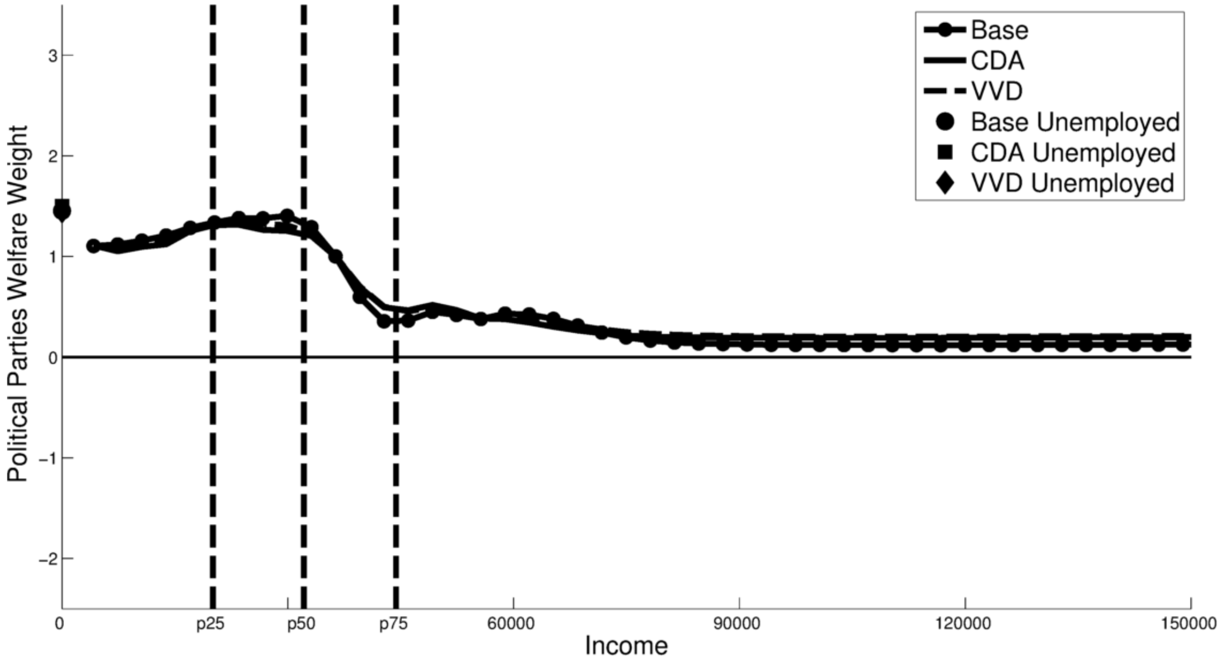


Figure 3.13: SWW Lower Elast. of Tax Base: Christian Democratic *CDA* and Liberal Conservative *VVD*

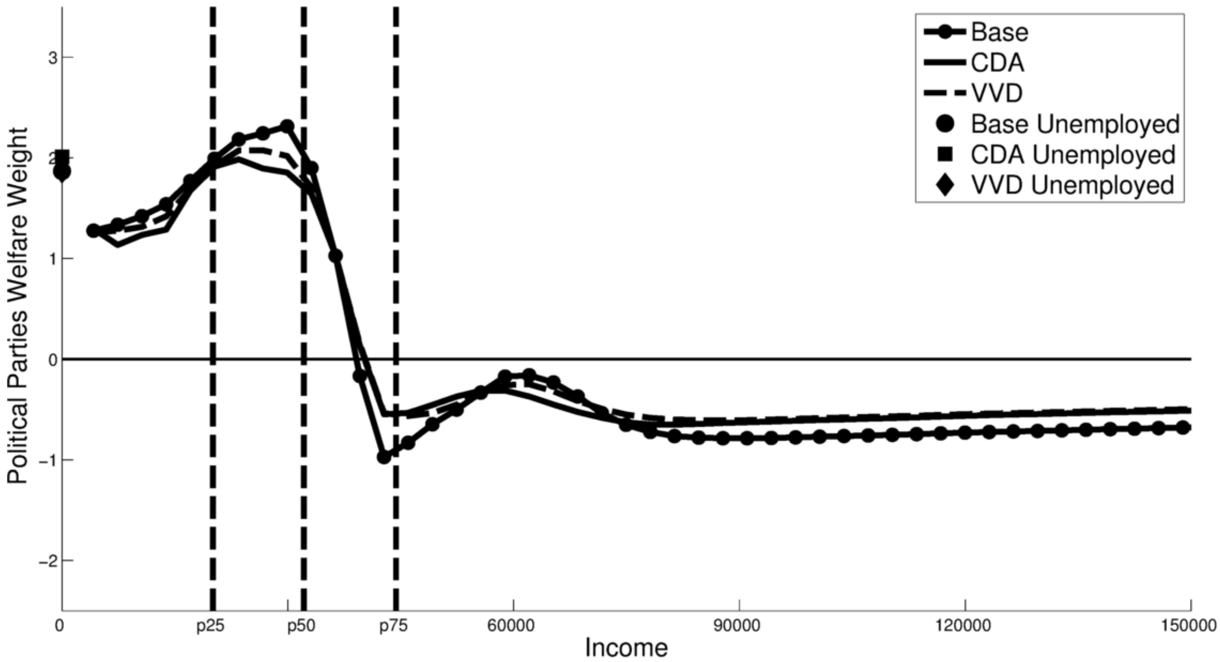


Figure 3.14: SWW Higher Elast. of Tax Base: Christian Democratic *CDA* and Liberal Conserv. *VVD*

We further see that although the right-wing parties lower the top rate, they still set the top rate beyond the Laffer rate, so that the social welfare weights remain slightly negative for top incomes.

What is also striking, looking at both the left-wing and the right-wing parties, is that gross modo the welfare weights are rather similar. This is a further indication that perhaps political-economy considerations play a role in the tax-benefit proposals.

3.5.3 Sensitivity with Respect to Elasticities

There is some uncertainty regarding the elasticity of the tax base. Hence, it is important to know how sensitive the resulting welfare weights are to changes in the assumed elasticities.

Figures 3.11 and 3.12 give the resulting welfare weights for the left-wing parties when the elasticity on both the intensive and extensive margin is 50% lower or 50% higher, respectively. When we assume a lower elasticity of the tax base, we still see the drop in the social welfare weight when we move from the non-employed to the working poor. Also, the social welfare weights are still increasing from low to middle incomes. For the majority of top incomes the social welfare weights are now close to zero, though for some parties it is still slightly negative. Hence, even assuming a much lower elasticity than in the baseline the top rates are still close to or beyond the Laffer rate. When we assume a higher elasticity than the base instead, the anomalies are amplified. The social welfare weights rise much faster in the lower part, and weights are much more negative at the top. All in all, the results are quantitatively affected by the change in elasticities, but the anomalies largely remain.

Figures 3.13 and 3.14 show the social welfare weights for the right-wing parties assuming lower or higher elasticities, respectively. When the elasticities are 50% lower, we still see the drop in social welfare weights going from the non-employed to the working poor, and we still see that social welfare weights increase from low to middle incomes. However, top incomes now get a positive weight from the rught-wing parties. Hence, perhaps the proposed top rates can be rationalized if the right-wing parties believe that the elasticity at the top is in fact much lower than in the baseline (based on recent estimates). When the elasticity of the tax base is higher, the anomalies become more pronounced, also for the right-wing parties.

3.6 Discussion

The analysis above reveals a number of anomalies in the redistributive preferences of the baseline and of the proposals of the political parties. Social welfare weights are increasing until modal incomes and they are negative for top incomes. Furthermore, the social welfare weights look remarkably similar across parties and compared to the baseline. Below we discuss how we could rationalize these findings, both from an economic point of view and from a political point of view.

3.6.1 Economic Interpretations

A plausible explanation for the negative welfare weights at the top even for the right-wing parties is that they did not have the right information on the elasticity of the tax base at the top. In the 2002 elections the CPB still assumed – conservatively – that the elasticity of the tax base at the top was only 0.10. In that case, increasing the top tax rate from the baseline value of 52% still generates some tax revenue (up to a limit) and reducing the top tax rate is costly in terms of lost tax revenue. Recent estimates of the elasticity of taxable income in the Netherlands by Jongen and Stoel (2013b) show that the elasticity of the tax base is actually in the order of 0.25 at the top. For this elasticity, and given the high Pareto parameter in the Netherlands, increasing the top tax rate beyond its current level of 52% already results in lower tax revenue.

Furthermore, regarding the top rates, we also assumed that utilities are not interdependent. However, it might be that individuals are engaged in 'rat races' (Akerlof, 1976) and 'keeping up with the Joneses' (Layard, 1980). In that case, if one individual decides to supply more labor or effort, negative externalities result as the utilities of other individuals fall due to a loss in relative status or income. Indeed, when there is rivalry in consumption, distortionary income taxes not only have costs, but also benefits in order to tame the rat race or to correct status-seeking behavior. The distortions of redistribution are then smaller and optimal tax rates are higher, see also Kanbur *et al.* (2006).[22]

The discontinuous jump down in the social weights when moving from the non-working poor to the working poor suggests that political parties underestimate the extensive-margin responses. Indeed, to align the welfare weights of the working poor more with the non-working poor policies such as the EITC would be needed. In order to promote labor-force participation, many policy makers argued for years that marginal tax rates

[22]However, by the same token, Alesina *et al.* (2005) argue that for one individual it becomes more attractive to enjoy more leisure if other individuals also enjoy more leisure. This rivalry in leisure thus exacerbates the distortions of income taxation, since not only labor supply choices are distorted, but also a 'leisure multiplier' is put in motion.

should be reduced at the lower end of the earnings distribution. What this paper shows is that such a policy lowers income redistribution towards the poor, while not necessarily raising participation. Indeed, only when the participation tax rates decline, not marginal tax rates, will participation increase.

One explanation for the anomaly that the social welfare weights rise at the bottom of the income distribution might be that we ignore that many individuals live in multi-person households. There is a small body of literature looking at family taxation, *e.g.* Boskin and Sheshinski (1983), Apps and Rees (1998), Schroyen (2003), Alesina *et al.* (2011) and Kleven *et al.* (2009). However, to the best of our knowledge there is not yet an inverse optimal-tax method for families. By ignoring household composition we ignore, for example, intra-household redistribution and economies of scale. Secondary earners typically have low income, but high consumption. Furthermore, we ignore differences in the labor supply elasticity of primary and secondary earners. It is a stylized fact in empirical labor economics that secondary earners have a (much) higher labor-supply elasticity than primary earners (Mastrogiacomo *et al.*, 2013). The difference between income and consumption for secondary earners and the higher elasticity of secondary earners may explain in part why marginal tax rates are kept relatively low at lower incomes. Failing to take this into account may generate social welfare weights at the bottom that are too low.

Admittedly, our model is still rather stylized. However, it remains to be seen how adding more realism to the model will alter the qualitative results of our analysis, in particular the finding that the social welfare weights of the working poor are lower than the welfare weights of the middle incomes. Indeed, the same anomaly can also be found in studies that focus only on singles or single mothers, or use more detailed information on tax base elasticities for subgroups.

3.6.2 Political Interpretations

There are also other explanations for the anomalies that we find, which may reflect political constraints in setting the tax-benefit system. Indeed, the anomalies we detect are consistent with a number of political-economics theories that figure prominently in the political-economics literature.

First, the rise in social welfare weights from the working poor to the middle-income groups and the sharp drop in these weights thereafter can be understood by standard political models of income redistribution (Romer, 1975; Roberts, 1977; Meltzer and Richard, 1981). In these models the median voter determines the amount of income redistribution. Consequently, the political system gears income redistribution towards the median voter. The tax-benefit system in the Netherlands is determined by coalition governments. Political-economy models with coalition governments are notoriously hard to solve in theory. However, the underlying mechanism of the basic median voter model is intuitively appealing. Middle incomes get a high weight relative to the rest because political parties have to attract enough votes from this densily populated group.

Second, the patterns of the social welfare weights – increasing to modal incomes, decreasing thereafter and turning negative for the high-income groups – are consistent with Director's law (Stigler, 1970). According to this theory, the middle-income groups can form a successful, stable political coalition to extract resources from both the low-income and the high-income groups, that cannot align their political interests. This is indeed what we observe in our analysis.

Third, left-wing parties might sacrifice on their ideological preference to redistribute income. Roemer (1998) shows that the poor – having a larger electorate – may not want to soak the rich through very redistributive tax systems. He develops a model of two-dimensional political competition where political parties position themselves on their redistributive preference and some non-economic ideological preference, such as religion. Even left-wing parties may then sacrifice on their redistributive goals if this helps to achieve larger electoral success by attracting more voters on their non-economic, ideological position.

Fourth, post-election considerations could explain the large status-quo bias that we observe in our analysis. Indeed, political parties may not want to deviate too much from the status quo given that they need to form a coalition government with other political parties after elections are held. Coalition agreements are more difficult to achieve if there has been a very polarized political campaign based on sharp ideological differences. See also Persson and Tabellini (2000).

Fifth, the status-quo bias and the persistence of various anomalies could also be explained by collective-action problems. The costs of the tax-benefits reforms that remove our anomalies are concentrated in the densely populated middle-income groups, whereas the benefits of reforms are dispersed among the electorate at large. Vested interests among the middle-income groups, could, therefore, be effective in blocking welfare-improving tax-benefit reforms (Olson, 1982).

Finally, it remains somewhat of a mystery to us that the non-working poor are apparently considered to be much more deserving by Dutch political parties than the working poor. We are unaware of political theories that could explain this anomaly.

3.7 Conclusions

In this paper we have used the inverse optimal-tax method to reveal the redistributive preferences of Dutch political parties in the 2002 elections. We have shown that there a number of pre-existing anomalies in the tax-benefit system of 2002. Indeed, social welfare weights for the working poor are much lower than welfare weights for the non-working poor. Social welfare weights are rising from the working poor towards middle-income workers. Finally, top incomes have a negative social welfare weight.

These anomalies are somewhat exacerbated by the proposals of the left-wing parties, and somewhat mitigated by the proposals of the right-wing parties. However, even more striking is the similarity of the social welfare weights across the party proposals, and how none of the parties has monotonically declining social welfare weights with income. We argue that these anomalies can be explained by political-economy considerations.

Although we put in quite some effort incorporating essential elements – such as the participation decision – into our analysis, our model is admittedly still rather stylized. In future work we hope to tackle some of the difficulties with multi-person households. Furthermore, it would be interesting to study whether the political constraints can be included into the model, and how well they can explain the anomalies detected in this paper.

Chapter 4

Optimal Redistribution and Monitoring of Labor Effort[1]

> "Informational frictions are a specification of a particular type of technology. For example, when we say "effort is hidden", we are really saying that it is infinitely costly for society to monitor effort. The desired approach would be to devise optimal tax systems for different specifications of the costs of monitoring different activities and/or individual attributes. To be able to implement this approach, we need to ... extend our modes of technical analysis to allow for costs of monitoring other than zero or infinity." Kocherlakota (2006, pp. 295-296)

4.1 Introduction

Redistribution of income is one of the most important tasks of modern welfare states. However, redistribution is expensive as it distorts the incentives to provide work effort. As a result, there is a trade-off between equity and efficiency. On a fundamental level, Mirrlees (1971) demonstrates that the trade-off between equity and efficiency originates from an information problem. Earnings ability and labor effort are private information and the government cannot condition redistributive taxes and transfers on earnings ability. Therefore, the government cannot distinguish individuals that are unable to work from

[1]This chapter is based on Zoutman and Jacobs (2013). We would like to thank Luca Micheletto, Jean-Marie Lozachmeur, Aart Gerritsen, Katherine Cuff and Dirk Schindler for useful suggestions and comments on an earlier version of this paper. Furthermore, this paper benefited from comments and suggestions made by participants at the 66th IIPF Conference in Uppsala; the CPEG Conference, Quebec; and seminar participants at the Erasmus School of Economics. All remaining errors are our own. The Matlab programs used for the computations in this paper are available from the authors on request.

individuals that are unwilling to work. Hence, redistribution from high-income to low-income earners inevitably distorts the incentives to provide work effort.

In practice, labor effort is not completely non-verifiable, as assumed by Mirrlees (1971). Indeed, some welfare states do condition the tax burden on some measure of labor supply. For example, in the UK low-income individuals receive a tax credit if they work more than 30 hours. This policy can only be implemented if the government is able to verify hours worked. Similar restrictions apply to in-work tax credits in Ireland and New Zealand, see also OECD (2011). Clearly, the assumption that work effort and earning ability are not verifiable is a too strong assumption. In the real world, the government does verify work effort of some individuals to some extent, albeit at a cost. Consequently, the government can – to some extent – separate shirking high-ability individuals from hard-working low-ability individuals.

This paper extends Mirrlees (1971) by allowing the government to operate a monitoring technology. The monitoring technology allows the government to verify labor effort of an individual at a positive, but finite cost. If an individual is monitored, the government perfectly verifies his/her labor effort and can thus deduce a worker's ability. The government can set the *monitoring schedule* as a function of gross income. That is, the probability that an individual is monitored depends (possibly non-linearly) on his gross labor earnings. Monitoring work effort provides incentives to individuals to adjust their labor supply in a direction that the government desires. When individuals are monitored, they receive an exogenous penalty. The penalty is increasing in ability as high-ability individuals are required to earn more income. The penalty is decreasing in earnings, as a higher level of earnings indicates a higher labor effort for given ability.

The role of the penalties should not be taken too literally. An entirely equivalent reformulation is that individuals would receive a bonus or a tax credit when they have a lower ability or higher earnings.

Each individual is aware of the monitoring schedule and the penalty function before making labor-supply decisions. Hence, individuals can alter their monitoring probability and penalties by adjusting their labor effort. The total *wedge* on labor effort consists of the explicit income tax rate and an implicit subsidy on labor effort due to monitoring. Monitoring of effort acts as an implicit subsidy on labor supply for two reasons. First, the expected penalty decreases with labor effort, since the penalty is assumed to be decreasing in gross earnings. Second, the monitoring intensity, and therefore the probability of receiving a penalty, may decrease with gross earnings, depending on the shape of the monitoring schedule. For a given tax rate, monitoring can thus reduce the distortions of the income tax on labor effort, thereby increasing both equity and efficiency.

The government maximizes social welfare by optimally setting the non-linear monitoring intensity, alongside the optimal non-linear income tax. In our model, first-best can generally not be obtained. Because the penalty function is exogenous, penalties are generally not sufficient to ensure that all individuals supply the required level of labor effort, and hence, some monitored workers will receive a penalty. If the government would be able to optimize the penalty function a trivial first-best outcome would result by either raising the penalty to infinity[2] or adjusting the penalty function such that the implicit subsidy on work exactly off-sets the explicit tax on work. We solve for the optimal non-linear tax and monitoring schedules by decentralizing the optimal, incentive-compatible direct mechanism that induces truthful revelation of ability types. We do not deviate from Mirrlees (1971) that individuals always truthfully report earnings.[3]

The schedule of optimal non-linear labor wedges is affected in two important ways in comparison to Mirrlees (1971). First, an increase in the labor wedge reduces labor supply, and hence, increases marginal penalties for monitored individuals. Therefore, a higher monitoring intensity reduces the efficiency costs of the labor wedge. Second, a decrease in labor supply directly increases the actual penalty. This increases within-ability inequality between monitored and non-monitored individuals, since the monitored individuals receive a penalty, whereas the unmonitored individuals do not. Therefore, higher marginal taxes result in distributional loss due to monitoring activities. The net effect of monitoring on the optimal wedge is thus theoretically ambiguous.

In Mirrlees (1971) tax rates at, or above, 100 percent can never be optimal. In contrast to Mirrlees (1971), we demonstrate that marginal tax rates could optimally be larger than 100 percent due to optimal monitoring. In particular, individuals may exert positive work effort even if the marginal income tax rate is above 100 percent, as long as the total wedge on labor remains below 100 percent. This could explain why effective marginal tax rates of close to, or even higher than, 100 percent are observed in real-world tax-benefit systems in the phase-out range of means-tested benefits. See Immervoll (2004), Spadaro (2005), Brewer *et al.* (2010) and OECD (2011) for examples in OECD countries.

The non-linear monitoring schedule is set so as to equate the marginal cost of monitoring to the marginal efficiency gain associated with monitoring at each gross income level. The efficiency gain of monitoring is increasing in the distortion created by the wedge on

[2] See also Schroyen (1997), Mirrlees (1997) and Mirrlees (1999).

[3] We realize that the assumption of truthful reporting of earnings is not always realistic due to, for example tax evasion and avoidance. This issue has been discussed in, amongst others, Cremer and Gahvari (1996), Schroyen (1997) and Chander and Wilde (1998). In most developed countries, however, firms are required to report gross labor earnings directly to the tax authorities, which prevents underreporting of earnings for a very large fraction of labor earnings (see e.g. Kleven *et al.*, 2011).

labor. Therefore, the optimal monitoring intensity increases with both the total labor wedge and the labor-supply elasticity.

Unfortunately, there is no closed-form solution for the optimal tax and monitoring schedules. Therefore, we resort to numerical simulations based on some realistic calibration of the model on US data. Our simulations demonstrate that the optimal tax schedule follows a U-shape, which closely resembles the simulations of Saez (2001). Moreover, the monitoring schedule also follows a U-shape. This confirms that the monitoring intensity should indeed be large when tax distortions on labor supply are large. The simulations demonstrate that the marginal tax rates with monitoring are generally larger than without monitoring. Hence, monitoring always results in more redistribution of income from high- to low-ability individuals, despite the inequality within-ability groups that results due to monitoring and penalizing individuals.

Strikingly, our simulations demonstrate that the optimal tax rate at the bottom end of the income scale is substantially above 100 percent. This implies that the implicit subsidy on work due to monitoring is very effective in reducing the total tax wedge on labor effort at the lower end of the income scale. Indeed, the optimal monitoring probability is close to one at the bottom, but it drops substantially towards middle-income levels. There is a slight increase in the monitoring probability towards the top, as tax rates increase. Therefore, we conclude from our simulations that monitoring is most important at the bottom of the income distribution. Strongly redistributive governments should therefore optimally employ a high monitoring intensity at the low end of the income scale, for example, via job-search requirements, benefit sanctions, work bonuses, and active labor-market programs. Moreover, our findings suggest that in work-dependent tax credits for low-income earners, like those in the UK, Ireland and New Zealand, are indeed part of an optimal redistributive tax policy.

The welfare gains of monitoring are shown to be large. Compared to the optimal non-linear tax schedule without monitoring, monitoring increases average labor earnings by 1.35 percent in our baseline simulation. Moreover, the transfer increases by about 4 percent. The monetized welfare gain of monitoring is about 1.4 percent of total output. The optimal monitoring probability does not exceed 20 percent anywhere except at the lower end of the income distribution. In our baseline simulations, the cost of monitoring and the average penalty are both only very small fractions of average labor earnings. Extensive sensitivity analyses demonstrate that the results are robust to parameter changes in the monitoring technology, on which little empirical evidence exists.

The setup of the paper is the following. In the next section we give a brief overview of the related literature. The third section introduces the model and derives the conditions for first- and second-order incentive compatibility. The fourth section derives the optimality conditions for monitoring and redistribution. The fifth section presents the simulations. Finally, the sixth section concludes.

4.2 Literature Review

Our model builds upon two strands in the mechanism-design literature. Mirrlees (1971), Diamond (1998), and Saez (2001) develop the theory of the optimal non-linear income tax under the assumption that both effort and ability are completely private information, implicitly assuming that verification of either effort or ability is infinitely costly. On the other hand, the literature on costly state verification develops principal-agent models where the outcome of a project is a function of both the state of the world and the action of the agent (see, e.g., Mirrlees, 1999, 1976, Holmstrom, 1979, and Townsend, 1979). The outcome is observed, but the action and the state of the world can only be verified through costly monitoring. Monitoring can then improve the ex-ante utility of both the principal and the agent. We apply the theory of costly state verification to the Mirrlees (1971) model and show that monitoring of effort can increase welfare significantly.

In a related paper, Armenter and Mertens (2013) study the effect of optimal monitoring of ability types on the optimal tax schedule. They analyze a dynamic model of optimal taxation where the government can use a monitoring technology to establish the ability of an agent. In their model, the monitoring intensity is exogenous, while penalties are endogenous. In equilibrium, individuals do not misreport their ability, and are, therefore, never penalized. Indeed, the economy is shown to converge to first best in an infinite-horizon setting. We instead analyze the case where monitoring is endogenous and penalties are exogenously given. Because penalties are exogenously given, individuals may misreport their ability type in equilibrium. Consequently, our model does not converge to a first-best outcome. An advantage of allowing for an endogenous monitoring intensity is that we do not need to worry about a tax-riot equilibrium in which all individuals misreport their type when they expect other individuals to do the same (Bassetto and Phelan, 2008).

The effect of monitoring has also been studied in the literature on tax evasion and unemployment insurance. The literature on tax evasion (see, e.g., Allingham and Sandmo, 1972, Sandmo, 1981, Mookherjee and Png, 1989, Slemrod, 1994, Cremer and Gahvari, 1994, 1996, Chander and Wilde, 1998, and Slemrod and Kopczuk, 2002) extends the Mirrlees (1971) framework by allowing individuals to under-report their earned income to the tax authorities.[4] Compared to the standard Mirrlees (1971) model, income taxation is more distortionary, because it not only reduces labor supply, but also increases tax evasion. However, the government can monitor individuals by auditing their tax returns and fine them when they evade taxes. In a two-type economy with non-linear taxation and monitoring Cremer and Gahvari (1994, 1996) show that the welfare-maximizing policy is to levy a positive marginal tax rate on the bottom type and a zero tax rate at the top. All individuals reporting income below a threshold level should be monitored with positive probability. The tax rate and the monitoring schedules are strategic complements for the government, because a higher tax rate induces an increase in tax evasion, thereby increasing the social value of monitoring.

In our model the only choice variable of individuals is their labor effort.[5] The monitoring instrument is therefore aimed at measuring effort instead of evasion. We extend the literature by considering optimal non-linear tax and monitoring under a continuum of skill types. This allows us to derive an elasticity-based formula for the optimal non-linear tax and monitoring schedule in the spirit of Diamond (1998) and Saez (2001). Moreover, we can determine the shape of non-linear tax and monitoring schedules over the entire income distribution through simulations.

In the literature on unemployment insurance, Ljungqvist and Sargent (1995a,b) study the effect of monitoring on equilibrium employment in welfare states.[6] In their model, unemployed workers may receive a job offer each period. In the absence of monitoring, the benefits induce workers to decline an inefficiently large number of job offers. Monitoring can help raising efficiency by punishing those workers who decline job offers. Simulations using Swedish data demonstrate that welfare states with large benefits and progressive taxation can have low equilibrium unemployment rates, provided the monitoring probability and sanctions are sufficiently large. In a model of optimal income redistribution with search, Boadway and Cuff (1999) determine the welfare-maximizing monitoring probability and demonstrate that it is increasing in the level of the benefits. Boone and Van Ours (2006) and Boone *et al.* (2007) develop a search model where the government can actively monitor and sanction job-search effort. They show that monitoring and sanctioning may be more effective in reducing unemployment than cutting the replacement rate. In addition, they show that monitoring may be effective, even when the duration of

[4] A comprehensive survey of the literature can be found in Slemrod and Yitzhaki (2002).

[5] An alternative interpretation would be that individuals exogenously supply labor, but can use a costly evasion technology.

[6] A lot has been written on optimal unemployment insurance, see Fredriksson and Holmlund (2006) for a survey of this literature. However, this literature typically does not consider monitoring of search effort.

unemployment benefits is limited. This literature has focused on monitoring the search effort of unemployed workers. We contribute to this literature by studying the effect of monitoring on employed workers.

Finally, we contribute to the literature on optimal non-linear tax simulations (see, for example, Mirrlees, 1971, Tuomala, 1984, Saez, 2001, Brewer *et al.*, 2010 and Zoutman *et al.*, 2013a). We show that monitoring can lead to significant improvements in both equity and efficiency.

4.3 Model

4.3.1 Households

The setup of our model closely follows Mirrlees (1971). Individuals are heterogeneous in their earnings ability, n, which denotes the productivity per hour worked. Ability is distributed according to cumulative distribution function $F(n)$ with support $[\underline{n}, \overline{n}]$, where $\overline{n}$ could be infinite. The density function is denoted by $f(n)$. Workers are perfect substitutes in production and the wage rate per efficiency unit of labor is constant and normalized to one. n therefore corresponds to the number of efficiency units of labor of each worker. Gross labor income of an individual is the product of his/her ability and his/her labor effort $z_n = nl_n$.

Individuals derive utility from consumption c_n and disutility from labor effort l_n. Net income is what remains of gross-labor income after taxes $T(z_n)$ and possibly a sanction, P, of not supplying enough labor effort. All net income is consumed.

$\pi(z_n)$ is the probability that an individual with earnings z_n is monitored by the government. It denotes the fraction of monitored individuals with gross earnings z_n. $\pi(z_n)$ is also referred to as the *monitoring intensity.* We assume the government receives a perfect signal of the individual's labor effort l_n if an individual is monitored. The ability of a monitored individual can then be inferred from the production relation: $n = z_n/l_n$.

Monitored individuals will receive a penalty or bonus depending on their observed earnings z_n and ability n:

$$P \equiv P(z_n, n), \quad P, P_n, -P_z \geq 0, \quad P_{zn} \leq 0, \quad \forall n, z. \tag{4.1}$$

We will refer to $P(\cdot)$ as the penalty function. The penalty function $P(\cdot)$ is exogenously given, and assumed to be continuous and twice differentiable in both arguments. We restrict penalties to be non-negative. Alternatively, we could have framed the model in terms of work bonuses rather than penalties, where individuals receive a bonus (or negative penalty) when they earn a higher income. Such a formulation would be mathematically equivalent to the current setup.

For given gross income z_n, penalties are assumed to be larger for individuals with higher ability n: $P_n > 0$. Intuitively, a low-ability individual has to work harder to earn a given gross income than a shirking high-ability individual. Furthermore, penalties decrease in gross income ($P_z < 0$), since an individual individual supplies more effort to earn a higher income, for a given level of ability. In addition, we make the logical assumption that the marginal penalty, $-P_z$, increases in ability, such that $P_{zn} \leq 0$. Hence, the government provides stronger work incentives to those with higher earnings potential. Later, we will derive that this assumption helps to ensure incentive compatibility.

As an example, a special case of the penalty function is $P(z_n, n) = \hat{P}\left(\frac{z_n}{n}\right) = \hat{P}(l_n)$, with $\hat{P}'(\cdot) < 0$. In this case, individuals receive penalties (bonuses) strictly on the basis of the amount of their work effort. With this penalty function, marginal penalties are increasing in ability if the elasticity of the marginal penalty is smaller than one:

$$P_{zn}(\cdot) \leq 0 \Longleftrightarrow -\frac{\mathrm{d}\hat{P}'(l_n)}{\mathrm{d}l_n}\frac{l_n}{\hat{P}'(l_n)} \leq 1. \tag{4.2}$$

That is, if a one-percent decrease in labor effort does not lead to an increase in the marginal penalty of more than one percent. Figure 2 displays an example of penalties that are only a function of labor effort. As can be seen, the penalty decreases quadratically in actual earnings up to $l_n = l_n^*$, after which it remains constant at 0. Such a penalty function will be used in the simulations later.

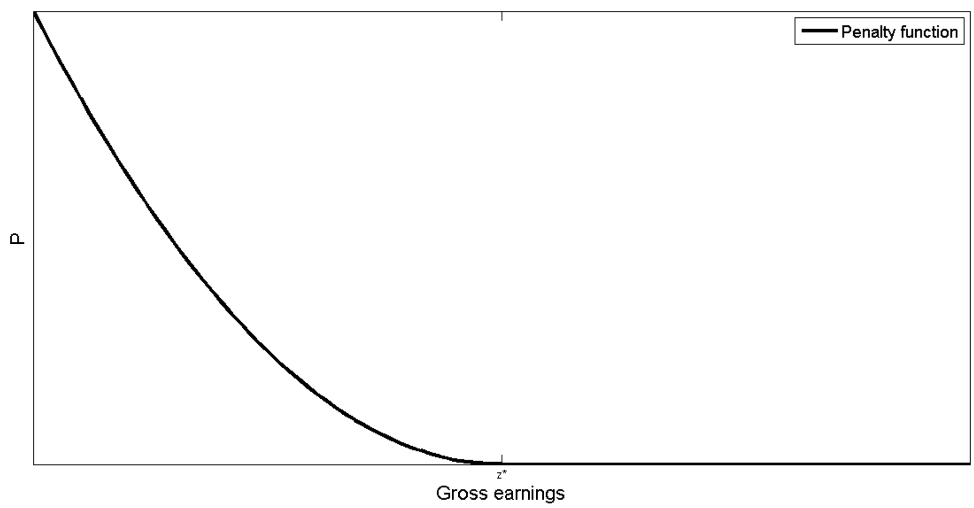


Figure 4.1: Example of a Penalty Function

We assume that the penalty function $P(z_n, n)$ is exogenous and outside the control of the redistributive government. The reason is that if $P(z_n, n)$ would be optimally set, the results would become trivial. In particular, if the penalty function could be optimized, the optimal penalty for any deviation of work effort from the first-best level would be infinity, so that all individuals would choose to perform the first-best labor effort. We believe that constraining $P(z_n, n)$ can be defended on legal grounds as the legal system imposes limitations on the government's ability to use infinite penalties. A more thorough discussion on these issues can be found in Schroyen (1997), Mirrlees (1997), and Mirrlees (1999).

What is not outside the control of the tax authority is the determination of the tax rate and monitoring probability. An individual with ability n is monitored with probability $\pi(z_n)$ and not monitored with probability $1-\pi(z_n)$. The consumption of an unmonitored individual is given by $c_n^U \equiv z_n - T(z_n)$. The consumption of a monitored (and penalized) individual is given by $c_n^P \equiv z_n - T(z_n) - P(z_n, n)$.

Individuals are assumed to maximize expected utility subject to their budget constraints in monitored and unmonitored states. We follow Diamond (1998) by assuming that all individuals have an identical, quasi-linear expected utility function:

$$\begin{aligned} u\left(z_n, n\right) &\equiv \pi(z_n)c_n^P + (1-\pi(z_n))c_n^U - v(l_n), \quad v'(\cdot) > 0, \quad v''(\cdot) > 0, \\ &= z_n - T(z_n) - \pi(z_n)P(z_n, n) - v(z_n/n), \quad \forall n, \end{aligned} \tag{4.3}$$

where we substituted the household budget constraint and $l_n = z_n/n$ in the second line.

The first term in the first line represents the non-monitoring probability times the consumption of an individual that is not monitored. The second term in the first line is the monitoring probability times the consumption of an individual that is monitored. The last term in the first line is the disutility of labor effort. An important analytical advantage of this quasi-linear-in-consumption utility function is that individuals are risk-neutral.[7]

Individuals choose the optimal amount of gross income based on their productivity n, the tax function $T(\cdot)$, the monitoring function $\pi(\cdot)$, and the penalty function $P(\cdot)$. An income level z_n is incentive compatible if it maximizes $u\left(z_n, n\right)$. The first-order condition

[7] We could allow for risk-aversion in the utility function. In that case we are only able to solve for the optimal non-linear tax and monitoring schedules if the social welfare function is utilitarian. Intuitively, the problem becomes analytically untractable if the government has a different degree of risk-aversion – which is implied by a non-utilitarian social welfare function – than households. Without risk aversion, this problem is always absent and we can allow for any degree of inequality aversion in the social welfare function.

for optimal labor supply is given by:

$$v'(z_n/n) = n\left(1 - T'(z_n) - \pi'(z_n)P(z_n,n) - \pi(z_n)P_z(z_n,n)\right), \quad \forall n. \tag{4.4}$$

On the right-hand side, we see that policy drives a wedge between the private and social benefits of labor supply. The *total labor wedge* $\mathcal{W}_n$ is given by

$$\mathcal{W}_n \equiv \frac{n - v'(z_n/n)}{n} = \underbrace{T'(z_n)}_{\text{explicit tax}} + \underbrace{\pi'(z_n)P(z_n,n) + \pi(z_n)P_z(z_n,n)}_{\text{implicit tax}}, \quad \forall n. \tag{4.5}$$

In a laissez-faire equilibrium the right-hand side of eq. (4.4) equals n and the total labor wedge $\mathcal{W}_n$ is zero. The total labor wedge consists of the explicit marginal tax on labor (T') and the *implicit* marginal tax (subsidy) on labor due to monitoring ($\pi'P + \pi P_z$). If $T' + \pi'P + \pi P_z > 0$, the redistributive tax and monitoring policy reduces optimal labor effort below the laissez-faire level, and vice versa if it is smaller than zero. The wedge is naturally increasing in the explicit marginal rate T'. The labor wedge is increasing in marginal monitoring probability π', such that $P > 0$. π' gives the marginal increase in the monitoring probability as a function of gross earnings. If the monitoring probability increases (decreases) with income, this reduces (increases) the incentive to exert effort, because a higher labor income increases (decreases) the probability of receiving a penalty. Therefore, an increase in the marginal monitoring probability decreases the incentive to exert work effort.

Proposition 1 shows that without loss of generality we can assume that expected consumption $\mathcal{C}(z_n) \equiv z_n - T(z_n) - \pi(z_n)P(z_n,n)$ is non-decreasing in earnings z_n. Consequently, the total labor wedge $\mathcal{W}_n$ can never be larger than one, i.e. larger than 100 percent.

Proposition 1 *All implementable continuous allocations can be implemented through a continuous non-decreasing expected consumption function $\mathcal{C}(z_n)$, $\forall n$. If $\mathcal{C}(z_n)$ is continuous and differentiable, the wedge $\mathcal{W}_n$ can never exceed* 1.

Proof. The proof directly follows Mirrlees (1971). Let $\tilde{\mathcal{C}}(z)$ be any continuous expected consumption function. The individual maximization problem is given by:

$$z_n = \arg\max_{z_n} \tilde{\mathcal{C}}(z_n) - v(z_n/n), \quad \forall n. \tag{4.6}$$

Now consider function $\mathcal{C}(z_n) = \max_{\tilde{z}_n \leq z_n} \tilde{\mathcal{C}}(\tilde{z}_n)$. Clearly, $\mathcal{C}(\cdot)$ is non-decreasing and continuous, because $\tilde{\mathcal{C}}(\cdot)$ is continuous. Now, consider the maximization problem:

$$\max_{z_n} \mathcal{C}(z_n) - v(z_n/n) = \max_{z_n} \left[\max_{\tilde{z}_n \leq z_n} \tilde{\mathcal{C}}(\tilde{z}_n) \right] - v(z_n/n), \qquad \forall n. \tag{4.7}$$

Assume z_n is the solution to problem (4.6). The solution to this second maximization problem must also be z_n. To see this evaluate $\mathcal{C}(\cdot)$ at z_n: $\mathcal{C}(z_n) = \max_{\tilde{z}_n \leq z_n} \tilde{\mathcal{C}}(\tilde{z}_n)$. Either $\mathcal{C}(z_n) = \tilde{\mathcal{C}}(z_n)$ or $\mathcal{C}(z_n) = \mathcal{C}(\bar{z}_n)$ with $\bar{z}_n < z_n$. In the first case, maximization problems (4.7) and (4.6) are equivalent, and hence, they must have the same solution. In the second case, because $v'(\cdot)$ is strictly increasing in z_n, $\bar{z}_n$ must give a higher value to the objective function in eq. (4.6) than does z_n. Hence, we arrive at a contradiction, because z_n could not have been the solution to problem (4.6) in the first place. Therefore, without loss of generality we can focus on non-decreasing functions $\mathcal{C}(\cdot)$. Now, suppose $\mathcal{C}(\cdot)$ is differentiable and consider its derivative.

$$\mathcal{C}'(z_n) = 1 - T'(z_n) - \pi'(z_n)P(z_n, n) - \pi(z_n)P_z(z_n, n) = 1 - \mathcal{W}_n, \qquad \forall n. \tag{4.8}$$

$\mathcal{C}(z_n)$ is non-decreasing if its derivative is greater than or equal to zero: $\mathcal{C}'(z_n) \geq 0 \Leftrightarrow \mathcal{W}_n \leq 1$. ■

Proposition 1 has an intuitive interpretation. Suppose, an individual has a budget constraint such that expected consumption is decreasing in gross income over some interval. Then, this individual will never choose his gross income in this interval, because he can work less and consume more, both yielding higher utility. Consequently, the government can never increase social welfare by setting the wedge $\mathcal{W}_n$ above 1. The explicit marginal tax rate $T'(z_n)$, however, could be above 1, provided that monitoring implies a sufficiently large implicit marginal subsidy on work, i.e. $\pi P_z + \pi' P < 0$, such that the overall wedge remains below 1. This is the case if the expected penalty decreases sufficiently fast in labor effort so that $-\pi' P > \pi P_z$. Therefore, monitoring can give incentives to provide work effort, even if the tax schedule reduces the incentives to work.

4.3.2 Government

The government designs an optimal income tax system and monitoring schedule so as to maximize social welfare, subject to resource and incentive constraints. The government's objective function is a concave sum of individual utilities:

$$\int_{\underline{n}}^{\overline{n}} G\left(u(z_n)\right) \mathrm{d}F(n), \quad G'(\cdot) > 0, \quad G''(\cdot) < 0. \tag{4.9}$$

$G(\cdot)$ is the social welfare function. Redistribution from high-income individuals to low-income individuals raises social welfare because the government is inequality averse. Due to quasi-linearity of private utility there is no social desire to redistribute income if the social welfare function is utilitarian. The government is constrained in its ability to redistribute income, because the ability of individuals is private information. However, the government can infer the ability of an individual from costly monitoring activities or it can induce self-selection by sacrificing on redistribution.

The total cost of monitoring is given by:

$$\int_{\underline{n}}^{\overline{n}} k(\pi(z_n))\mathrm{d}F(n), \quad k(0) = 0, \quad k'(\cdot), k''(\cdot) > 0. \tag{4.10}$$

The cost of monitoring is increasing and convex in the monitoring probability π. Since there is a perfect mapping between skill n and labor earnings z_n, we can also write $\pi(\cdot)$ as a function of the skill level n, where we use the short-hand notation $\pi(z_n) = \pi_n$. However, $\pi'(z_n) \equiv \frac{\mathrm{d}\pi_n}{\mathrm{d}z_n}$ always denotes the derivative of monitoring with respect to gross earnings.

The economy's resource constraint implies that total labor earnings equal aggregate consumption plus monitoring costs:

$$\int_{\underline{n}}^{\overline{n}} z_n \mathrm{d}F(n) = \int_{\underline{n}}^{\overline{n}} \left((1 - \pi(z_n))c_n^U + \pi(z_n)c_n^P + k(\pi(z_n))\right) \mathrm{d}F(n). \tag{4.11}$$

By defining unpenalized consumption as $c_n \equiv c_n^U = c_n^P + P(z_n, n)$, we can write for aggregate consumption:

$$\int_{\underline{n}}^{\overline{n}} \left((1 - \pi(z_n))c_n^U + \pi(z_n)c_n^P\right) \mathrm{d}F(n) = \int_{\underline{n}}^{\overline{n}} \left(c_n - \pi(z_n)P(z_n, n)\right) \mathrm{d}F(n). \tag{4.12}$$

Hence, using eq. (4.12) the economy's resource constraint (4.11) can be rewritten as:

$$\int_{\underline{n}}^{\overline{n}} \left(z_n + \pi(z_n)P(z_n, n)\right) \mathrm{d}F(n) = \int_{\underline{n}}^{\overline{n}} \left(c_n + k(\pi(z_n))\right) \mathrm{d}F(n). \tag{4.13}$$

We do not need to consider the government budget constraint, since it is automatically implied by Walras' law if the individual budget constraints and the economy's resource constraint are satisfied.

The timing of the model is as follows:

1. The government announces the exogenously given penalty function, as well as the optimal non-linear income tax and monitoring schedules.
2. Each individual optimally chooses the amount of labor effort.
3. The government observes the labor incomes chosen by each individual and taxes income and monitors individuals accordingly. The government penalizes all monitored individuals if they were found to be deviating from their required gross income level.
4. Individuals receive utility from consumption and leisure.

By the revelation principle any indirect mechanism can be replicated with an incentive-compatible direct mechanism (Myerson, 1979; Harris and Townsend, 1981). Therefore, we can find the optimal second-best allocation by maximizing welfare subject to feasibility and incentive-compatibility constraints. We can decentralize the optimal second-best allocation as a competitive market outcome through the non-linear tax and monitoring schedules.

4.3.3 First-Order Incentive Compatibility

By using the envelope theorem we can derive a differential equation for the indirect utility function u_n which is a necessary condition for incentive compatibility. The next subsection derives the conditions under which the first-order condition is indeed sufficient. The incentive compatibility constraint is found by totally differentiating eq. (4.3) with respect to n:

$$\frac{\mathrm{d}u_n}{\mathrm{d}n} = \frac{\partial u\left(z_n, n\right)}{\partial n} + \frac{\partial u\left(z_n, n\right)}{\partial z_n}\frac{\mathrm{d}z_n}{\mathrm{d}n} = \frac{l_n v'\left(l_n\right)}{n} - \pi(z_n)P_n(z_n, n), \qquad \forall n, \tag{4.14}$$

where $\frac{\partial u(z_n,n)}{\partial z_n} = 0$ due to the individual's first-order condition in eq. (4.4). Thus, if the optimal allocation satisfies eq. (4.14), individuals' first-order conditions for utility maximization are also satisfied.

4.3.4 Second-Order Incentive Compatibility

Without further restrictions we cannot be certain that the optimal allocation derived under the first-order incentive compatibility constraint (4.14) is also implementable. An implementable allocation should satisfy additional requirements to ensure that the first-order approach also respects the second-order conditions for utility maximization. The next Lemma summarizes the requirements for second-order incentive compatibility.

Lemma 1 *Second-order conditions for utility maximization are satisfied under the first-order approach if the following conditions hold at the optimal allocation for all n:*
i) single-crossing conditions on the utility and penalty functions are satisfied:

$$\frac{\partial(v'(l_n)/n)}{\partial n}+\pi(z_n)P_{zn}(z_n,n)+\pi'(z_n)P_n(z_n,n)\leq 0, \tag{4.15}$$

ii) z_n is non-decreasing in ability:

$$\frac{\mathrm{d}z_n}{\mathrm{d}n}\geq 0. \tag{4.16}$$

Proof. The second-order condition for the utility-maximization problem (4.3) is given by:

$$\frac{\partial^2 u(z_n,n)}{\partial z_n^2}\leq 0, \qquad \forall n. \tag{4.17}$$

This second-order condition can be rewritten in a number of steps. Totally differentiating the first-order condition (4.4) gives:

$$\frac{\partial^2 u(z_n,n)}{\partial z_n^2}\frac{\mathrm{d}z_n}{\mathrm{d}n}+\frac{\partial^2 u(z_n,n)}{\partial z_n\partial n}=0, \qquad \forall n. \tag{4.18}$$

Substitution of this result in eq. (4.17) implies that the second-order condition is equivalent to:

$$\frac{\partial^2 u(z_n,n)}{\partial z_n\partial n}\left(\frac{\mathrm{d}z_n}{\mathrm{d}n}\right)^{-1}\geq 0, \qquad \forall n. \tag{4.19}$$

Differentiating the first-order condition (4.4) with respect to n yields:

$$\left(\frac{\partial(v'(l_n)/n)}{\partial n}+\pi(z_n)P_{zn}(z_n,n)+\pi'(z_n)P_n(z_n,n)\right)\left(\frac{\mathrm{d}z_n}{\mathrm{d}n}\right)^{-1}\leq 0, \qquad \forall n. \tag{4.20}$$

The inequality holds if all conditions of the Lemma are satisfied. ■

The single-crossing condition and the monotonicity of gross earnings are well-known from the Mirrlees model (Mirrlees, 1971; Ebert, 1992). The single-crossing condition ensures that – at the same consumption-earnings bundle – individuals with a higher ability have a larger marginal willingness to provide work effort. In our model, the single-crossing condition contains three elements. The first is the standard Spence-Mirrlees condition on the utility function, i.e. $\frac{\partial(v'(l_n)/n)}{\partial n}$. If this term is negative, the marginal disutility of work for individuals with a higher ability level is lower. Most utility functions considered in the literature exhibit this property, including our own. The second term is determined by $P_{zn}(z_n, n)$. As discussed before, the assumption that $P_{zn} \leq 0$ implies that marginal penalties do not decrease in ability. Hence, high-ability individuals are more likely to self-select in higher income-consumption bundles if the marginal penalty for earning less income increases with ability. The third term concerns the slope of the monitoring schedule, $\pi'(z_n)P_n(z_n, n)$ and its sign is determined by the the monitoring schedule, since $P_n > 0$. If the marginal monitoring probability decreases in gross earnings ($\pi'(z_n) < 0$) individuals will work harder in order to decrease the probability of being monitored and penalized. Whereas the first two terms feature uncontroversial signs, the last term is determined by the endogenous monitoring schedule. Hence, high-ability individuals can be induced to self-select into higher income-consumption bundles, unless the monitoring probability increases too fast with ability.

A second requirement to induce self-selection is that gross earnings are indeed increasing with ability at the optimal schedule. Consequently, a tax schedule that provides higher income to higher ability individuals induces self-selection of higher ability types into higher income-consumption bundles.

In the remainder we assume that all the conditions derived in Lemma 1 hold at the optimal allocation. In our simulations, we check the second-order sufficiency conditions ex-post and we always confirm that they are respected.

4.4 Optimal Second-Best Allocation with Monitoring

The optimization problem with monitoring can be specified formally as:

$$\max \int_{\underline{n}}^{\overline{n}} \left[(1-\pi_n) G(c_n - v(z_n/n)) + \pi_n G(c_n - P(z_n, n) - v(z_n/n)) \right] f(n) \mathrm{d}n, \quad (4.21)$$

$$\text{s.t.} \quad \int_{\underline{n}}^{\overline{n}} \left[z_n + \pi_n P(z_n, n) - c_n - k(\pi_n) \right] f(n) \mathrm{d}n = 0, \quad (4.22)$$

$$\frac{\mathrm{d}u_n}{\mathrm{d}n} = \frac{z_n v'(z_n/n)}{n^2} - \pi_n P_n(z_n, n), \quad (4.23)$$

$$u_n = c_n - \pi_n P(z_n, n) - v(z_n/n), \quad \forall n, \quad (4.24)$$

$$\pi_n \geq 0, \quad \forall n. \quad (4.25)$$

The final constraint assumes that the probability of monitoring cannot be smaller than zero. We assume that the cost of monitoring is sufficiently large to ensure that the constraint $\pi_n \leq 1$ is never binding.

The Hamiltonian function of this problem is given by:

$$\begin{aligned} \mathcal{H} \equiv & \left[(1-\pi_n) G(u_n^U) + \pi_n G(u_n^P) + \lambda (z_n + \pi_n P(z_n, n) - c_n - k(\pi_n)) \right] f(n) \quad (4.26) \\ & -\theta_n \left(\frac{z_n v'(z_n/n)}{n^2} - \pi_n P_n(z_n, n) \right) + \mu_n (u_n - c_n + \pi_n P(z_n, n) + v(z_n/n)) + \eta_n \pi_n, \end{aligned}$$

c_n, z_n and π_n are control variables. u_n is a state variable with θ_n as its associated co-state variable. $u_n^U \equiv c_n - v(z_n/n)$ and $u_n^P \equiv u_n^U - P(z_n, n)$ denote the utility of the unpenalized and penalized individuals, respectively. μ_n is the Lagrange multiplier for the definition of utility. λ is the Lagrange multiplier of the economy's resource constraint. η_n is the Kuhn-Tucker multiplier of the non-negativity constraint on π_n.

The first-order necessary conditions are given by:

$$\frac{\partial \mathcal{H}}{\partial c_n} = 0 : \left[(1-\pi_n)\, G'(u_n^U) + \pi_n G'(u_n^P) - \lambda\right] f(n) - \mu_n = 0, \qquad \forall n, \tag{4.27}$$

$$\begin{aligned}\frac{\partial \mathcal{H}}{\partial z_n} = 0 : &\left[-(1-\pi_n)\, G'(u_n^U)\frac{v'(\cdot)}{n} - \pi_n G'(u_n^P)\left(\frac{v'(\cdot)}{n} + P_z(\cdot)\right) + \lambda\left(1 + \pi_n P_z(\cdot)\right)\right] f(n) \\ &- \theta_n\left(\frac{v'(\cdot) + z_n v''(\cdot)/n}{n^2} - \pi_n P_{zn}(\cdot)\right) + \mu_n\left(\frac{v'(\cdot)}{n} + \pi_n P_z(\cdot)\right) = 0, \qquad \forall n,\end{aligned} \tag{4.28}$$

$$\frac{\partial \mathcal{H}}{\partial \pi_n} = 0 : \left[G(u_n^P) - G(u_n^U) - \lambda\left(k'(\pi_n) - P(\cdot)\right)\right] f(n) + \theta_n P_n(\cdot) + \mu_n P(\cdot) + \eta_n = 0, \qquad \forall n, \tag{4.29}$$

$$\frac{\partial \mathcal{H}}{\partial u_n} = \frac{\mathrm{d}\theta_n}{\mathrm{d}n} : \frac{\mathrm{d}\theta_n}{\mathrm{d}n} = \mu_n, \qquad \forall n, \tag{4.30}$$

$$\eta_n \pi_n = 0, \qquad \eta_n \geq 0, \qquad \pi_n \geq 0, \qquad \forall n, \tag{4.31}$$

$$\lim_{n \to \underline{n}} \theta_n = \lim_{n \to \bar{n}} \theta_n = 0. \tag{4.32}$$

Compared to the analysis of Mirrlees there are two new first-order conditions. Eq. (4.29) states the optimal monitoring condition, and eqs. (4.31) state the Kuhn-Tucker conditions for the non-negativity constraint on π_n.

4.4.1 Optimal Wedge on Labor

Proposition 2 gives the conditions for optimal income redistribution.

Proposition 2 *The optimal net marginal wedge on labor $\mathcal{W}_n$ at each ability level satisfies:*

$$\frac{\mathcal{W}_n}{1 - \mathcal{W}_n} = A_n B_n C_n - D_n, \qquad \forall n, \tag{4.33}$$

where

$$A_n \equiv 1 + \frac{1}{\varepsilon_n} + \pi_n \frac{n}{v'(z_n/n)} \varepsilon_{P_z}, \tag{4.34}$$

$$B_n \equiv \frac{\int_n^{\overline{n}} (1 - g_m) f(m) \mathrm{d}m}{1 - F(n)}, \tag{4.35}$$

$$C_n \equiv \frac{1 - F(n)}{n f(n)}, \tag{4.36}$$

$$D_n \equiv -P_z(z_n, n) \frac{n}{v'(z_n/n)} \sigma_n, \tag{4.37}$$

$\sigma_n \equiv \frac{(1-\pi_n)\pi_n\left(G'(u_n^P)-G'(u_n^U)\right)}{\lambda} > 0$ *is a measure for the welfare cost of inequality between penalized and unpenalized individuals at ability level* n*,* $\varepsilon_n \equiv \left(\frac{l_n v''(l_n)}{v'(l_n)}\right)^{-1} > 0$ *is the compensated wage elasticity of labor supply,* $\varepsilon_{P_z} \equiv -\frac{n}{P_z}\frac{\partial P_z}{\partial n} > 0$ *is the elasticity of the marginal penalty with respect to ability, and* $g_n \equiv \frac{(1-\pi_n)G'(u_n^U)+\pi_n G'(u_n^P)}{\lambda} > 0$ *is the average, marginal social value of income, expressed in money units, for individuals at ability level* n*.*

Proof. Integrate eq. (4.30) using a transversality condition from eq. (4.32). If follows that $\theta_n = \lambda \int_n^{\overline{n}} (1-g_m)f(m)\mathrm{d}m$. Substitute this result and eq. (4.27) in eq. (4.28), use eq. (4.5), and simplify to obtain the Proposition. ■

The A_n-term is related to the inverse of the efficiency cost of a labor wedge at income level z_n. The second term in A_n, $1/\varepsilon_n$, is the inverse of the labor-supply elasticity and it enters because the deadweight loss of the wedge increases in the labor-supply elasticity. The third term represents the efficiency gains of monitoring. Penalties are more effective in seperating high- and low-ability individuals if marginal penalties are strongly increasing in ability, that is, if the elasticity of the marginal penalty with respect to ability, ε_{P_z}, is larger. This effect is stronger if the monitoring intensity π is larger. In addition, the third term decreases in the marginal disutility of labor v'/n. Intuitively, the benefit of increasing earnings, in order to reduce the marginal penalty, is smaller if the disutility of earning that additional income is larger. Hence, in comparison to the optimal wedge without monitoring (cf. Diamond, 1998; Saez, 2001), monitoring reduces the efficiency cost of taxation if marginal penalties increase in ability.

The B_n-term measures the equity gain of an increase in the wedge at income level z_n. The first term, 1, captures the revenue gain of a larger marginal labor wedge at n, such that individuals with an income level above z_n pay one unit of income extra tax. The welfare loss of extracting one unit of income from the individuals above n is g_m for all individuals $m \geq n$. Therefore, $\int_{\underline{n}}^{n} (1-g_m)dF(n)$ measures the redistributional gain of the labor wedge at n. B_n is the conditional average welfare gain of the wedge levied at n. The B_n-term is not directly affected by monitoring. Since welfare weights g_n are always declining with income, B_n always rises with income, see also Diamond (1998).

C_n is the inverse relative hazard rate of the skill distribution. Its numerator is the fraction of the population whose net income is decreased by increasing the wedge and its denominator captures the size of the tax base that is distorted by the wedge. Hence, the numerator in C_n gives weights to average equity gains in B_n and the denominator to average efficiency losses in A_n – as in the model without monitoring. The numerator of C_n always declines with income; there are fewer individuals paying marginal taxes if the tax rate is increased at a higher income level. Hence, for a given B_n the total distributional benefits of raising the labor wedge fall as the income level rises. For a unimodal skill distribution the denominator of C_n always increases with income before the mode, since both n and $f(n)$ are rising. Thus, labor wedges always decrease with income before modal income. After the mode, $f(n)$ falls, although n continues to rise with income. Hence, it depends on the empirical distribution of n whether C_n rises or falls with income after modal income. For most empirical distributions, C_n appears to rise after the mode and converges to a constant at the top. See also Diamond (1998), Saez (2001) and Zoutman *et al.* (2013a).

Finally, D_n measures the welfare loss associated with within-ability inequality. If the labor wedge increases, earnings at n decrease. Therefore, the penalty at n increases, which in turn increases inequality between monitored and unmonitored individuals. σ_n measures the marginal welfare cost of this within-ability inequality. The effect of a wedge on within-ability inequality is increasing in the marginal penalty, $-P_z$. It decreases in the marginal cost of earning one unit of additional income, v'/n, because, again, the relative effect of the penalty function on gross earnings decreases if the disutility of labor earnings increases. D_n increases in the monitoring probability for $\pi_n < .5$ because the within-skill variance of monitoring is increasing in π_n for $\pi_n < .5$. Finally, D_n is increasing in the concavity of the welfare function, because the difference in welfare weights between penalized and unpenalized individuals, $\frac{G'(u_n^p)-G'(u_n^u)}{\lambda}$, is larger if the government is more inequality averse.

We can summarize the impact of monitoring on optimal labor wedges as follows. Monitoring decreases the efficiency cost of setting a higher labor wedge, but introduces within-ability inequality. Therefore, the total effect of monitoring on the optimal labor wedge is theoretically ambiguous. Our simulations below demonstrate that the efficiency gains of monitoring outweigh the distributional loss due to inequality between monitored and non-monitored individuals.

We can derive the non-linear tax function, which implements the second-best allocation as the outcome of decentralized decision making in a competitive labor market. Substituting eq. (4.4) into eq. (4.33) yields:

$$\frac{T'(z_n)+\pi'(z_n)P(z_n,n)+\pi(z_n)P_z(z_n,n)}{1-T'(z_n)-\pi'(z_n)P(z_n,n)-\pi(z_n)P_z(z_n,n)}=A_nB_nC_n-D_n, \quad \forall n. \tag{4.38}$$

Thus, when we know the optimal monitoring schedule $\pi(z_n)$, this equation implicitly defines the optimal non-linear income tax function $T(z_n)$.

4.4.2 Optimal Monitoring

The next proposition derives the optimal monitoring schedule.

Proposition 3 *The optimal level of monitoring at each ability level follows from:*

$$k'(\pi_n) + \Delta_n - g_n P(\cdot) \geq \frac{\frac{\mathcal{W}_n}{1-\mathcal{W}_n} + D_n}{A_n} n P_n(\cdot) \qquad \forall n, \tag{4.39}$$

where $\Delta_n \equiv \frac{G(u_n^U)-G(u_n^P)}{\lambda}$ *is the welfare difference between a penalized and an unpenalized individual expressed in money units. If* $\pi_n > 0$*, the equations hold with equality.*

Proof. Substitute eq. (4.27) in eq. (4.29), rearrange terms, employ the definitions for B_n and C_n, and use the fact that $\eta_n \geq 0$. Finally substitute eq. (4.33) for $B_n C_n$ to obtain the expression. By eq. (4.31) η_n only equals zero if $\pi_n > 0$ and therefore the equation holds with equality if $\pi_n > 0$. ∎

The first term on the left-hand side in condition (4.39) is the marginal cost of raising the monitoring intensity. The second and third terms on the left-hand side jointly represent the welfare effect of a compensated increase in the monitoring probability. That is, the welfare effect of an increase in the monitoring probability, while keeping expected utility at skill level n unchanged. The second term represents the uncompensated, direct welfare loss of an increase in the monitoring probability. If the monitoring probability increases, there will be more penalized and less unpenalized individuals. Therefore, the loss is equal to the welfare difference between penalized and unpenalized individuals. The third term represents the welfare gain associated with the compensation to keep expected utility unchanged if the monitoring probability is increased. The compensation at ability level n requires a transfer of P and its associated welfare effect is thus given by $g_n P$. In Lemma 2 we derive how the compensated welfare effect of monitoring changes with the monitoring probability for given levels of utility in monitored and unmonitored states.

Lemma 2 *The compensated welfare effect of the monitoring probability is decreasing in* π_n*, positive if* $\pi_n = 0$ *and negative if* $\pi_n = 1$ *for given levels of utility in penalized and unpenalized states.*

Proof. By a first-order Taylor expansion around u_n^U we can write Δ_n as:

$$\Delta_n = \frac{G(u_n^U) - G(u_n^P)}{\lambda} = \frac{G'(u_n^U)(u_n^U - u_n^P) + R(P)}{\lambda} = \frac{G'(u_n^U)P}{\lambda} + R(P). \tag{4.40}$$

where $R(P)$ is a second-order remainder term. Similarly, a first-order Taylor expansion around u^P yields:

$$\Delta_n = \frac{G'(u_n^P)P}{\lambda} - \hat{R}(P), \tag{4.41}$$

where $\hat{R}(P)$ is again a second-order remainder term. By concavity of G both remainder terms are positive for $P > 0$: $R(P), \hat{R}(P) > 0$. Now multiply eq. (4.40) with $(1 - \pi_n)$ and eq. (4.41) with π_n and add them to find:

$$\Delta_n - g_n P = (1 - \pi_n)\, R(P) - \pi_n \hat{R}(P). \tag{4.42}$$

The right-hand side gives the compensated welfare effect of the monitoring probability, which is decreasing in π_n, always positive if $\pi_n = 0$, and always negative if $\pi_n = 1$, ceteris paribus. ■

The right-hand side of eq. (4.39) represents the marginal benefits of monitoring. The benefits of monitoring increase in $-P_n$. This term can be interpreted as the power of the penalty function. The penalty function is more powerful if penalties increase strongly in ability. In addition, the marginal benefits of monitoring increase if labor-supply distortions are larger, i.e. if the labor wedge $\frac{\mathcal{W}_n}{1-\mathcal{W}_n}$ is larger or if the efficiency cost of taxation is larger, as captured by $1/A_n$. The benefits of monitoring also increase in within-ability inequality D_n. Intuitively, as more monitoring leads to higher labor effort, the expected penalty decreases. Hence, monitoring helps to reduce within-ability inequality.

From Proposition 3 follows that the government does not engage in monitoring if and only if (evaluated at a no-monitoring equilibrium with $\pi_n = 0$):

$$k'(0) + \Delta_n - g_n P(\cdot) \geq \frac{\frac{\mathcal{W}_n}{1-\mathcal{W}_n} + D_n}{A_n} n P_n(\cdot), \quad \forall n. \tag{4.43}$$

That is, if the marginal cost of monitoring are higher than the marginal benefits for all types. By evaluating eq. (4.33) at $\pi_n = 0$ it easily follows that the optimal allocation is the allocation derived in Mirrlees (1971). Mirrlees (1971) is thus a special case of our model where monitoring is prohibitively expensive, such that the government never optimally monitors.

4.4.3 Boundary Results

In the next Proposition we derive the optimal wedge and monitoring probability at the bottom and the top of the ability distribution.[8]

Proposition 4 *If the income distribution is bounded at the top, $\overline{n} < \infty$, the optimal wedge and monitoring probabilities at the extremes are:*

$$\mathcal{W}_{\underline{n}} = \mathcal{W}_{\overline{n}} = \pi_{\underline{n}} = \pi_{\overline{n}} = 0. \tag{4.44}$$

If the penalties are zero at the first-best levels of earnings, marginal tax rates are also zero at the endpoints:

$$T'(z_{\underline{n}}) = T'(z_{\overline{n}}) = 0. \tag{4.45}$$

Proof. From eq. (4.33) follows that $\left(\frac{\mathcal{W}_n}{1-\mathcal{W}_n} + D_n\right)/A_n = B_n C_n$. The transversality conditions (4.32) imply $B_{\underline{n}} C_{\underline{n}} = B_{\overline{n}} C_{\overline{n}} = 0$. At the extremes, the optimal monitoring condition (4.39), therefore simplifies to: $\Delta_n - g_n P + k'(\pi_n) \geq 0$. Evaluate this expression at $\pi = 0$:

$$\Delta_n - g_n P + k'(0) = R(P) + k'(0) \geq 0. \tag{4.46}$$

where $R(P) > 0$ is a second-order remainder term, and the second step follows from Lemma 2. The condition is always satisfied at $\pi_n = 0$. Hence, $\pi_n = 0$ is optimal at the extremes. The optimal wedges in eq. (4.33) at the extremes are zero, because the product $B_n C_n$ is zero by the transversality conditions, and D_n is zero, since $\pi_n = 0$. If the penalties are zero when labor supply is at a first-best level, then $P(\cdot) = 0$ at the endpoints, since labor effort is undistorted because the wedges are zero. Using $\pi_n = P(\cdot) = 0$ in eq. (4.5) then demonstrates that $\mathcal{W}_{\underline{n}} = \mathcal{W}_{\overline{n}} = T'(z_{\underline{n}}) = T'(z_{\overline{n}}) = 0$. ■

Proposition 4 establishes that the zero wedge result at the bottom and top result of the model without monitoring carries over to the model with monitoring (Sadka, 1976; Seade, 1977). Intuitively, the wedge at n redistributes income from individuals above n to the government, and, hence indirectly to individuals below n. There are no individuals above $\overline{n}$ and no individuals below $\underline{n}$. Therefore, there are no benefits associated to a positive wedge at these points of the ability distribution. However, the wedge does distort the labor-supply decision. Hence, the optimal wedge must be zero. Because the wedge is zero, there is no efficiency gain of monitoring. As a result, the optimal monitoring probability is also zero.

[8] Due to the absence of income effects in labor supply, bunching at zero labor earnings is not an issue in deriving the boundary results, see also Seade (1977).

However, marginal tax rates at the endpoints do not necessarily need to be zero. This critically depends on the penalty function. In particular, if the marginal monitoring probability is non-zero at the end-points ($\pi'(z_n) \neq 0$) and the expected penalty is positive, marginal tax rates at the endpoints have to be non-zero in order to compensate for the distortion caused by the change in monitoring intensity. In particular, marginal tax rates at the endpoints should be positive (negative) if $\pi'(z_n)P(\cdot) < 0$ (> 0). However, if penalties are zero if earnings at the end-points correspond to the first-best levels of earnings, then marginal tax rates at the end-points are zero as well.

4.5 Simulations

In this section we use numerical simulations to establish the shape of the optimal tax schedule. The simulations require four main ingredients: the ability distribution, the individual preferences, the social preferences and the monitoring technology. First, we use the skill distribution from Mankiw *et al.* (2009). The hourly wage is used as a proxy for earnings ability. We follow Mankiw *et al.* (2009) by assuming that wage rates follow a log-normal distribution, which is extended with a Pareto distribution for the top tail of the wage distribution. In addition, we assume that there is an exogenous fraction of 5 percent disabled individuals having zero earning ability ($\underline{n} = 0$), which is also based on Mankiw *et al.* (2009). The earnings distribution is estimated from March 2007 CPS data. This resulted in a mean log-ability of $m = 2.76$ and a standard deviation of log ability of $s = 0.56$. The Pareto tail starts at the top 1 percent of the earnings distribution and features a Pareto parameter of $\alpha = 2$. The latter is in accordance with estimates of Saez (2001).

Second, a description of individual preferences is needed. For the purpose of our simulations it is convenient if optimal labor effort is restricted between zero and one. In addition, we follow the literature in assuming a constant elasticity of taxable income (see, e.g., Saez, 2001). The following utility function abides both features:

$$u(c_n, l_n) = c_n - \frac{n}{1 + 1/\varepsilon} l_n^{1+1/\varepsilon}, \qquad \varepsilon > 0. \tag{4.47}$$

ε is the (un)compensated elasticity of taxable income. This utility function was used before in Brewer *et al.* (2010).[9] We follow the empirical literature estimating the elasticity of taxable income (see, e.g., Saez *et al.*, 2012) and set $\varepsilon = 0.25$.

[9] We slightly deviate from the model in the previous sections by introducing heterogeneity in the preference for labor and making the penalty function depended on ability (equation 52). This does not affect our results qualitatively, but it does simplify the numerical simulations.

The third ingredient is the social welfare function. We assume an Atkinson social-welfare function featuring a constant elasticity of relative equality aversion β:

$$\begin{aligned} G(u_n) &= \frac{u_n^{1-\beta}}{1-\beta}, \quad \beta \geq 0, \quad \beta \neq 1, \\ G(u_n) &= \ln(u_n), \quad \beta = 1. \end{aligned} \tag{4.48}$$

The utilitarian objective is obtained by assuming $\beta = 0$. A Rawlsian social welfare function results if $\beta \to \infty$. The baseline assumes a moderately redistributive government with $\beta = 0.99 \approx 1$. In the robustness analysis we also consider less redistributive governments ($\beta = 0.5$) and more redistributive governments ($\beta = 1.5$).

Finally, we need to make specific assumptions on the monitoring technology, and the penalty function. Unfortunately, no empirical evidence is available that guides us to calibrate these functions. However, our theoretical model provides some restrictions on the choice of the functions. Also, we perform robustness checks on the parameter choices we have made for these functions.

In our theoretical model, the cost of monitoring needs to be increasing and convex in the monitoring intensity π. We assume that the cost of monitoring is quadratic:

$$k(\pi_n) = \frac{\kappa}{2}\pi_n^2, \quad \kappa > 0. \tag{4.49}$$

where κ is a cost parameter indicating the marginal cost of a higher monitoring probability. In the baseline we assume $\kappa = 1$. In the robustness analysis we vary κ between 0.25 and 4.

In our baseline simulations, we assume that required labor effort l_n^* equals:

$$\begin{aligned} l_n^* &= 1, \quad \forall n > \underline{n}, \\ l_n^* &= 0, \quad n = \underline{n} = 0. \end{aligned} \tag{4.50}$$

Therefore, all working individuals, i.e. those with positive earning ability ($n > \underline{n}$), are required to perform first-best labor effort. Individuals that cannot work ($n = \underline{n} = 0$) are not required to work. Required labor earnings z_n^* are a linear function of required labor effort l_n^*:

$$z_n^* = n, \quad \forall n. \tag{4.51}$$

Each monitored individual that faces a positive labor wedge is also subject to a penalty. Consequently, monitoring will be effective in boosting labor effort at all income levels. In the robustness analysis we analyze the case where required work effort is only half of the first-best effort level, i.e. $l_n^* = 0.5$ for $n > \underline{n}$.

We assume that the penalty function is quadratic in labor effort l_n and is given by:

$$P = \frac{p}{2}\left(\min\left\{0, l_n - l_n^*\right\}\right)^2, \qquad p > 0, \tag{4.52}$$

where p is a parameter determining the severity of the penalty. The penalty is also a function the 'reference level' of labor effort l_n^*. If working individuals, i.e. those with positive earning ability ($n > \underline{n}$), are working less than the reference level of labor effort, they will receive a penalty, and increasingly so if their labor effort deviates more from the reference level of work effort. Hence, each monitored individual that faces a positive labor wedge is also subject to a penalty. And, monitoring will be effective in boosting labor effort at all income levels. Individuals that cannot work ($n = \underline{n} = 0$) are not required to work.

In the baseline we set $p = 3$. In the robustness checks we employ values of $p = 1$ and $p = 5$. In the baseline, the reference level of work effort l_n^* equals the first-best level of work effort:

$$\begin{aligned} l_n^* &= 1, \quad \forall n > \underline{n}, \\ l_n^* &= 0, \quad n = \underline{n} = 0. \end{aligned} \tag{4.53}$$

In the robustness analysis we analyze the case where reference work effort is only half of the reference effort level in the baseline, i.e. $l_n^* = 0.5$ for $n > \underline{n}$.

The government-revenue requirement is exogenous and set to 10 percent of labor earnings in the baseline specification without monitoring, following Tuomala (1984) and Zoutman *et al.* (2013a). The choices for all the parameters can be found in Table 1.

In the table, the first column on the right-hand side gives the base value of the parameter. In addition, we perform robustness checks with high and low parameter values for the welfare function, all parameters of the penalty function and all parameters of the monitoring technology to analyze the sensitivity of our results.

The numerical procedure we use to solve for the optimal allocation is a so-called shooting method. We solve the differential equations (4.14) and (4.30) numerically for given initial values $\theta_{\underline{n}}$, $u_{\underline{n}}$, and λ. Subsequently, we "shoot" for initial values until we meet boundary conditions (4.13) and (4.32). The wedge, tax, and monitoring schedule can be found using eq. (4.38). A more detailed explanation of the numerical procedure can be found in the Appendix.

Table 4.1: Calibration for Simulations

Parameter	Description	Base value	High value	Low value
m	Mean log ability	2.76	N/A	N/A
s	Standard deviation log ability	0.56	N/A	N/A
α	Pareto parameter	2.00	N/A	N/A
d	Fraction of disabled individuals	0.05	N/A	N/A
ε	Compensated elasticity	0.25	N/A	N/A
r	Government revenue as fraction of GDP	0.10	N/A	N/A
κ	Cost of monitoring	1.00	0.25	4.00
p	Penalty parameter	3.00	5.00	1.00
l^*	Required labor effort	1.00	N/A	0.50
β	Relative inequality aversion	1.00	1.50	0.50

4.5.1 Results

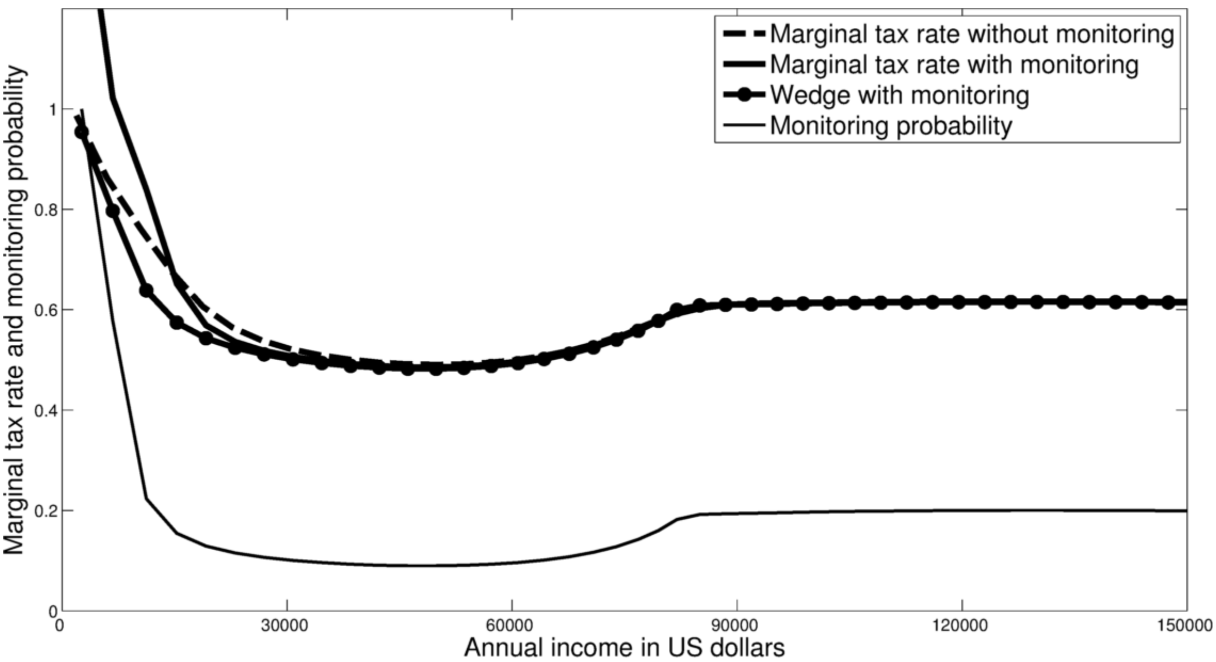


Figure 4.2: The Optimal Wedge, Tax and Monitoring Schedules in the Baseline Scenario
Note: Baseline parameter values of the model can be found in Table 1.

Figure 4.2 gives the optimal wedge, tax and monitoring schedules as a function of yearly income in US dollars. The fat solid line represents the optimal tax schedule with monitoring. The dashed line is the optimal tax schedule without monitoring. The circled line is the optimal total labor wedge with monitoring. And, the thin solid line is the optimal monitoring schedule. Recall that the optimal tax schedule coincides with the optimal labor wedge if there is no monitoring.

As can be seen, the optimal labor wedge follows a U-shape both with and without monitoring. Distortions are extremely large at the bottom of the labor market, relatively small for middle-income levels and somewhat higher at the top. The shape of these schedules is largely explained by the B_n and C_n terms in eq. (4.33). The B_n-term is strictly increasing with income as the welfare loss of taxing away one unit of income unit from individuals above z_n is strictly decreasing in z_n, see our previous discussion. The C_n-term follows a U-shape. At the bottom of the earnings distribution, the density of tax payers is small, and hence, efficiency costs of marginal taxes are low. In addition, the redistributional benefits of a higher marginal tax rate are large as it is paid by almost the entire population. Towards middle-income levels, the efficiency cost increases as the population density increases, whereas the redistributive benefits decrease as fewer individuals are paying a higher tax rate. After modal income marginal tax distortions decline more rapidly than distributional benefits of marginal taxes, hence marginal taxes increase. In the Pareto tail of the earnings distribition, the ratio of marginal distributional benefits and marginal efficiency costs of taxes becomes constant, and the tax wedge converges to a constant. These results are entirely in line with previous simulations performed in e.g. Saez (2001), Brewer *et al.* (2010), and Zoutman *et al.* (2013a).

Recall from the previous section, that the effect of monitoring on the labor wedge was theoretically ambiguous. However, in our simulations we see that the efficiency gain of monitoring in reducing labor distortions outweighs the distributional cost of raising within-skill group inequality. Therefore, monitoring decreases the optimal wedge, especially at low-income levels. The optimal monitoring schedule also follows a U-shape. In eq. (4.39) the labor wedge determines the shape of the monitoring schedule, as the other elements of the monitoring schedule do not exhibit a very strong dependence on income. The monitoring intensity decreases very steeply at the bottom of the income distribution. This gives individuals a strong incentive to increase their labor effort. At middle-income levels the monitoring intensity is relatively low. The monitoring intensity increases slightly towards top-income levels. However, the effect of monitoring on the labor wedge and the tax schedule is very small at higher income levels.

The optimal tax schedule exhibits extremely large tax rates at the bottom of the earnings distribution. Indeed, the government can levy tax rates above 100 percent at the lowest income earners. The sharp decrease in the monitoring intensity works as an implicit subsidy on work effort and partially off-sets the high explicit tax on labor supply. The poverty trap found in many countries (see, e.g., Spadaro, 2005, Brewer *et al.*, 2010 and OECD, 2011) can thus be optimal in the presence of monitoring. Indeed, there may not be a poverty trap if the monitoring schedule provides sufficient incentives, even if the tax-benefit system itself does not provide incentives to supply labor.

Note that the optimal wedge and monitoring probability at the top do not equal zero, as was derived in Proposition 4 for a bounded income distribution. Mirrlees (1971), Diamond (1998), and Saez (2001) show theoretically that the optimal wedge converges to a constant if the right tail of the ability distribution is Pareto distributed. Our simulations confirm that this result holds as well in the model with monitoring. In addition, we find that the optimal monitoring probability also converges to a positive constant.

4.5.2 Sensitivity Analysis

In this subsection we present the sensitivity analysis of the results obtained in the previous subsection. We especially explore the sensitivity of our simulation outcomes with respect to the monitoring technlogy and penalty function, on which little empirical evidence is available.

Figure 4.3 summarizes the simulations when the cost of monitoring is decreased ($\kappa = 0.25$) or increased ($\kappa = 4$). As expected, the monitoring schedule moves up if the monitoring cost decreases and down if the cost increases. However, the optimal tax schedule largely remains unaffected. From the optimal tax expression in eq. (4.38) we can infer that monitoring increases the optimal tax rate if the allocation remains unchanged. However, the allocation changes, since an increase in the monitoring probability increases revenue from taxation for any given tax rate. Therefore, the redistributive benefit of a marginal tax decreases as the same time. In our simulations, these two effects roughly cancel out and the optimal tax rates remain largely unaffected.

Figure 4.4 gives the optimal tax and monitoring schedules when the penalty parameter is decreased ($p = 1$) or increased ($p = 5$). As can be seen, for very low levels of income, both an increase and a decrease in the penalty parameter lead to a decrease in the monitoring intensity. This may seem *prima facie* a counter-intuitive outcome, but can be explained. An increase in the penalty parameter raises the effectiveness of monitoring, but it also increases within-skill level inequality. For low levels of income, the first effect dominates when penalties decrease, whereas the second effect dominates when penalties increase. However, beyond about 10,000 dollars of income, within-ability inequality becomes less relevant, and therefore, monitoring intensities always increase when the penalty parameter rises.

From the optimal tax formula in eq. (4.38) follows that an increase in the penalty parameter affects the optimal tax rate through six channels. First, an increase in the

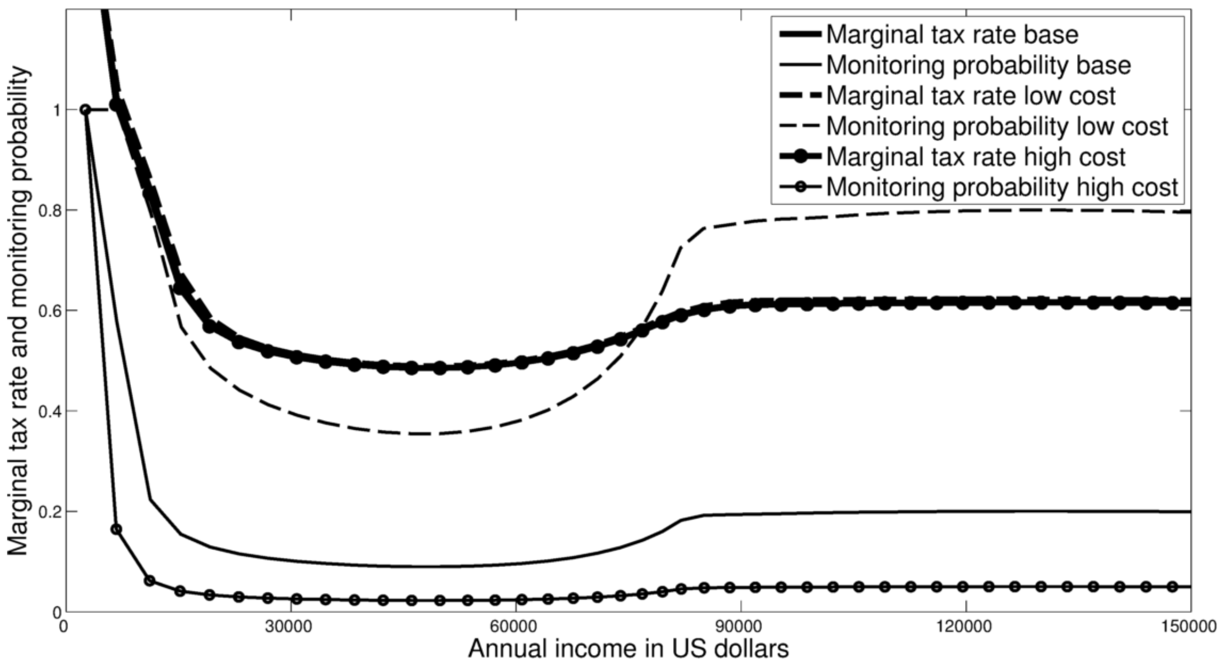


Figure 4.3: Optimal Tax and Monitoring Schedules for High ($\kappa = 4$) and Low ($\kappa = 0.25$) Marginal Cost of Monitoring
Note: All other parameters take baseline values, see Table 1.

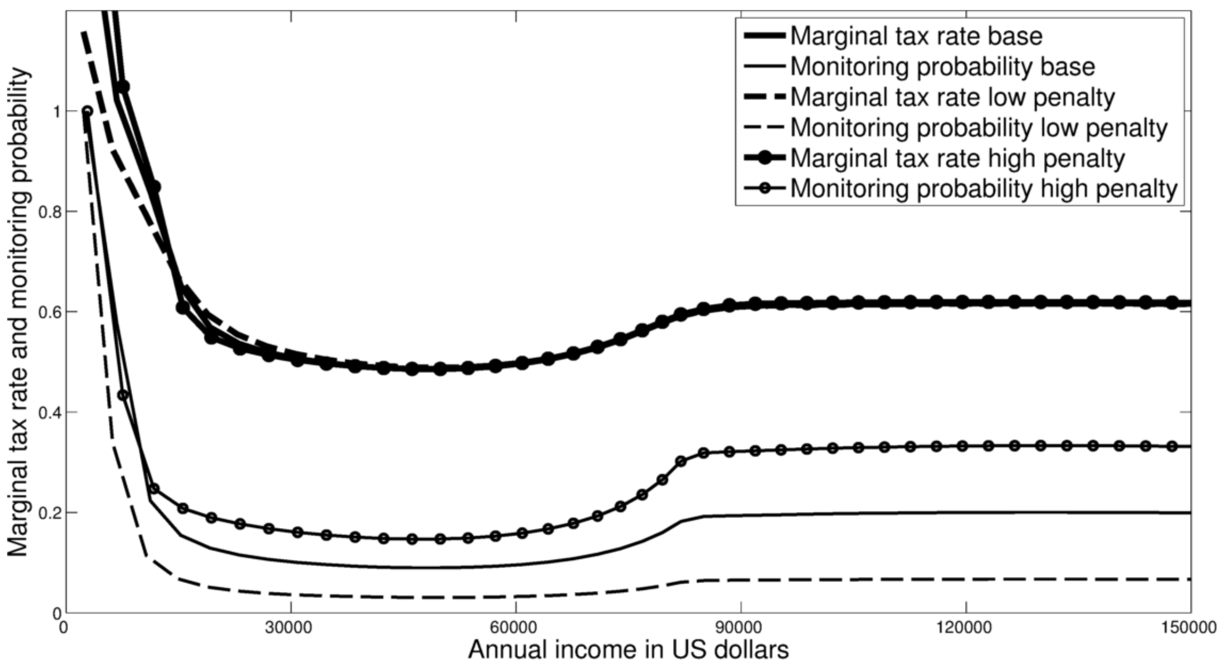


Figure 4.4: Optimal Tax and Monitoring Schedules for Strong ($p = 5$) and Weak ($p = 1$) Penalties
Note: All other parameters take baseline values, see Table 1.

marginal penalty raises the marginal tax rate for a given wedge. Second, an increase in the penalty itself may increase or decrease the optimal marginal tax rate for a given wedge depending on the sign of $\pi'(z_n)$. Third, an increase in the convexity of the penalty function decreases the efficiency cost of a wedge. Fourth, the penalty affects the monitoring probability, although the effect is ambiguous. Fifth, an increase in the penalty increases within skill-level inequality, which decreases the optimal wedge. Finally, the allocation itself is affected, but it is a priori unclear whether higher penalties lead to more or less redistribution. The simulation outcomes confirm these theoretical ambiguities. The net effect is positive for very low income levels, negative for medium-income levels, and negligible for higher income-levels.

Figure 4.5 illustrates the effect of a decrease in the reference level of labor effort ($l_n^* = 0.5$). As can be seen, the monitoring probability very quickly drops to zero, because all individuals find it in their best interest to work at least the reference amount of labor hours without monitoring. Surprisingly, the tax schedule remains virtually unaffected. This outcome demonstrates monitoring effort is most important at the bottom of the earnings distribution, where the labor wedge is highest. Still, marginal tax rates can be substantially above 100 percent at the bottom of the earnings distribution.

Finally, in Figure 4.6 we simulated the optimal tax and monitoring schedules for a higher degree of inequality aversion ($\beta = 1.5$) and a lower degree ($\beta = 0.5$) of inequality aversion. As can be seen, the optimal tax rate increases in inequality aversion as should be expected, although the difference at the bottom of the income distribution is small. Intuitively, monitoring decreases the distortion of a higher tax rate, but it also creates within skill-group inequality. The poorest individuals in society are the low-income individuals who are penalized. Hence, within-ability inequality is particularly costly if the government is strongly inequality-averse. For low levels of income, both an increase and a decrease of inequality aversion decrease the optimal monitoring intensity. At higher levels of income, within-skill group inequality aversion is less important, and the monitoring intensity unambiguously increases with inequality aversion as labor wedges are set higher when redistributive desires are stronger.

4.5.3 Allocations and Welfare

Clearly, monitoring is part of the optimal redistributive tax-benefit system. But, how important is monitoring for the optimal allocation and welfare? Table 4.2 reports the average monitoring cost, $\bar{k}/\bar{z}$, the average penalty $\bar{P}/\bar{z}$, the penalty for the lowest working individual, $P(\underline{n})/\bar{z}$, the transfer paid out to individuals having zero earnings, $-T(0)/\bar{z}$,

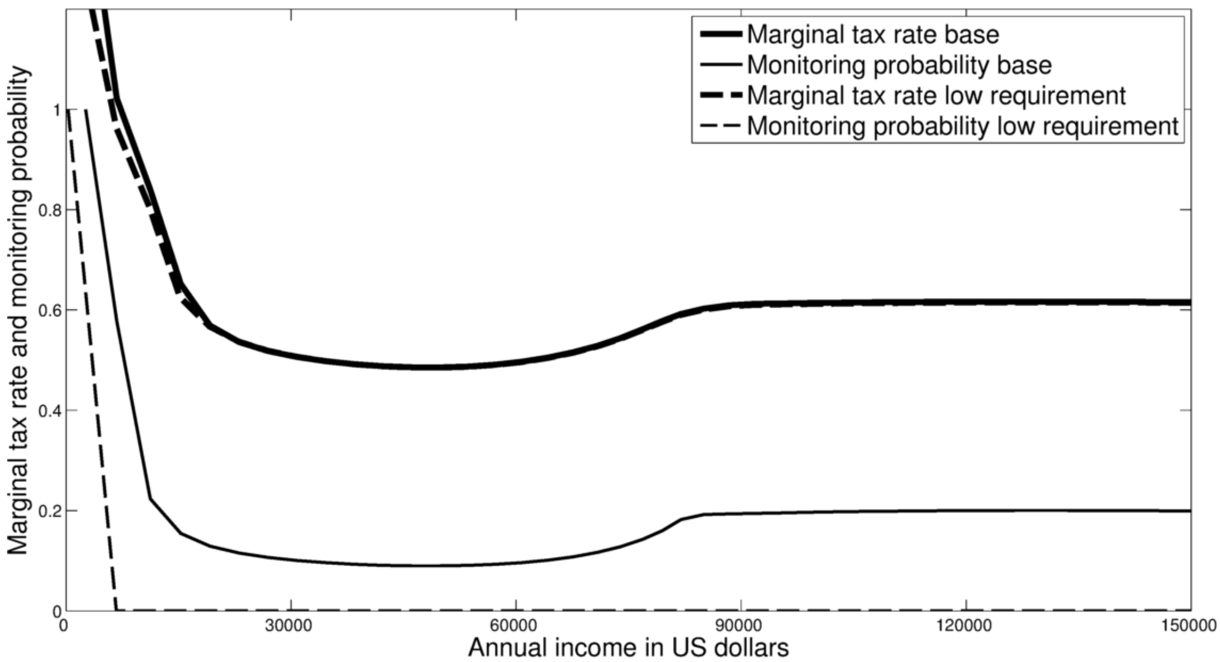


Figure 4.5: Optimal Tax and Monitoring Schedules for a Lower Reference Level of Work Effort ($l_n^* = 0.5$)
Note: All other parameters take baseline values, see Table 1.

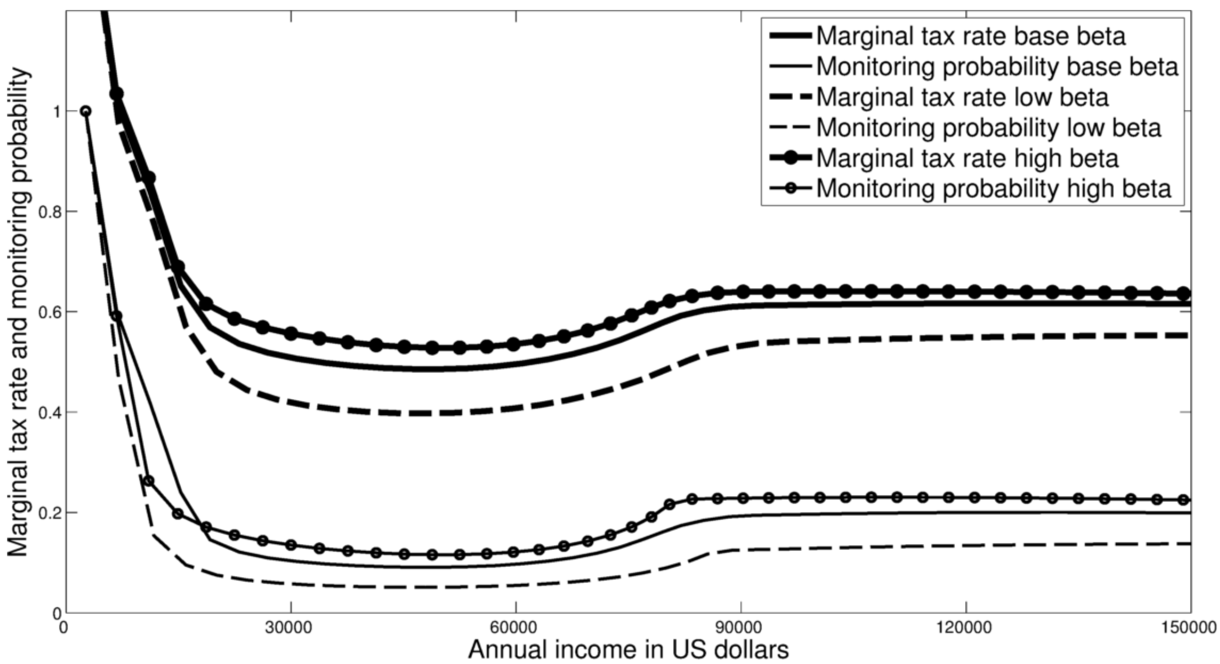


Figure 4.6: The Tax and Monitoring Schedule for Higher ($\beta = 1.5$) and a Lower ($\beta = 0.5$) Degree of of Inequality Aversion
Note: All other parameters take baseline values, see Table 1.

and the change in average earnings, $\Delta\bar{z}/\bar{z}$. All table entries are in percentages of average earnings.

From the first column we can infer that the average monitoring cost $\bar{k}/\bar{z}$ is a relatively small percentage of average labor earnings: about 0.5 percent of average earnings in the baseline. An increase in the marginal cost of monitoring raises total monitoring costs very little, since the increase in marginal cost is accompanied by a decrease in the optimal monitoring intensity at the optimum. A change in the reference level of labor effort reduces the total monitoring cost to almost zero, because monitoring is only used at the bottom if the labor requirement is low. In addition, the cost of monitoring is sensitive to the severity of the penalty as monitoring outlays rise (fall) with a stronger (weaker) penalty. A government having access to a stronger penalty technology will on average rely more heavily on monitoring to provide work incentives. Monitoring costs also increase (decrease) with inequality aversion, since a more inequality-averse government relies on average more heavily on monitoring to alleviate the equity-efficiency trade-off.

Table 4.2: Change in Allocation Due to Monitoring

	$\frac{\bar{k}}{\bar{z}}$	$\frac{\bar{P}}{\bar{z}}$	$\frac{P(\underline{n})}{\bar{z}}$	$\frac{-T(0)}{\bar{z}}$	$\frac{\Delta\bar{z}}{\bar{z}}$
No Monitoring	0.00	0.00	0.00	29.46	0
Base scenario	0.49	0.35	7.83	33.65	1.35
Low monitoring cost	0.40	0.34	7.55	34.20	1.50
High monitoring cost	0.61	0.36	8.08	33.13	1.18
Low reference effort	0.03	0.02	5.63	34.85	1.07
Low penalty	0.24	0.14	3.53	29.67	0.43
High penalty	0.63	0.49	10.02	37.01	2.04
Low inequality aversion	0.32	0.23	8.20	30.25	5.55
High inequality aversion	0.61	0.42	7.62	35.23	−0.90

Note: $\bar{z}$ is per capita labor income in the specified calibration, $\bar{k}$ is the per capita monitoring cost, $\bar{P}$ is the average penalty over the monitored population, $P(\underline{n})$ is the penalty at the lowest skill level, $-T(0)$ is the transfer and $\Delta\bar{z}$ is the change in average labor earnings as compared to the model without monitoring. All numbers are in percentages of average earnings.

The second column represents the average penalty given to monitored individuals as a percentage of average labor earnings $\bar{P}/\bar{z}$. As can be seen, penalties are relatively small. In the baseline, the average penalty equals 0.35 percent of average earnings. Penalties increase with the monitoring cost, because monitoring decreases with its marginal cost, and as a consequence, individuals work less and receive more severe penalties. The effects are very small, however. In addition, the average penalty falls strongly when the reference level of labor effort is lower. Similarly, the penalties increase (decrease) if the penalty parameter increases (decreases), as expected. The penalty also increases (decreases) with stronger (weaker) inequality aversion because a more (less) inequality-averse government sets higher (lower) wedges. The third column represents the average penalty at the bottom of the income distribution $P(\underline{n})/\bar{z}$. Penalties at the bottom are relatively large, because the wedge at the bottom is very high. Comparative-static effects of the penalty at the bottom are roughly similar to the comparative statics of the average penalty.

The fourth column represents the transfer as a fraction of earnings, $-T(0)/\bar{z}$, and the fifth column is the change in average labor earnings as compared to optimal taxation without monitoring, $\Delta\bar{z}/\bar{z}$. In almost all simulations, both the transfer and average labor earnings increase, indicating an improvement in both equity and efficiency of the tax-transfer system. This effect is surprisingly unsensitive to a change in the monitoring cost and to a change in the reference level of labor effort. These outcomes can be explained by the fact that monitoring is most effective at the bottom of the skill distribution. At this point in the earnings distribution, monitoring costs are relatively unimportant as the density of monitred individuals is low. A lower labor requirement is also unimportant, since individuals are working far less than any work requirement at the bottom end of the income scale. Results are more sensitive to the size of the penalty, because monitoring becomes less effective if the punishment technology is less effective. However, even if penalties are relatively low, the increase in both average labor earnings and the transfer is substantial. Finally, a change in the inequality aversion changes the emphasis given to either equity ($T(0)$) or efficiency (average labor earnings). In our scenario with low inequality aversion both increase. However, in the scenario with high inequality aversion average labor earnings decreases slightly.

Finally, Table 4.3 reports the welfare effects of monitoring. The first column represents the income-weighted average of the marginal deadweight loss of increasing the marginal tax rate by one percent. As can be seen, monitoring decreases the marginal deadweight loss by about 0.5 percent in our baseline simulation from 0.204 to 0.203. This result is robust in our sensitivity analyses.

The last column reports the monetized welfare gain of monitoring. We compute the compensating variation by calculating the amount of resources that have to be injected into an economy without monitoring in order to attain the same social welfare as the economy with optimal monitoring. In our base scenario, the welfare gain is about 1.4 percent of average labor earnings, i.e. 1.4 percent of total output. The welfare gain

Table 4.3: Welfare Effects of Monitoring.

	Marginal dead weight loss	Welfare gain
No Monitoring	0.204	–
Base	0.203	1.421
Low monitoring cost	0.203	1.592
High monitoring cost	0.204	1.073
Low reference effort	0.203	0.969
Low penalty	0.204	0.267
High penalty	0.203	1.835
Low inequality aversion	0.170	1.015
High inequality aversion	0.214	1.76

Note: The marginal deadweight loss refers to the income-weighted average of the marginal deadweight loss of all households as a consequence of increasing the labor wedge on labor with one percent. Welfare gains are obtained by calculating the compensating variation as a percentage of average earnings in the specified simulation.

increases if cost of monitoring are lower and if penalties are higher. Interestingly, the welfare gain is almost unaffected by a lower reference level of labor effort. The reason is that reference labor effort is still generating positive penalties at the bottom of the earnings distribution, where the benefits of monitoring are highest. Also, an increase in inequality aversion decreases the welfare gain of monitoring, because the within skill-group inequality created by monitoring is more costly for governments with a stronger inequality aversion. Nevertheless, in all cases we find quantitatively substantial social welfare gains.

4.6 Conclusions

In this paper we demonstrate that redistributive governments should optimally employ an effort-monitoring technology in order to redistribute income at the lowest efficiency cost. Monitoring of labor effort alleviates the equity-efficiency trade-off and raises equity, efficiency, or both. The reason is that distortions from redistribution derive from the informational problem that earning ability is private information. By using a monitoring technology this informational asymmetry is reduced. A first-best outcome cannot be reached, however, because monitoring is costly. Mirrlees (1971) is a special case of our model when monitoring is infinitely costly.

We demonstrated that monitoring works as an implicit subsidy on work effort, which partially off sets the explicit tax on work effort. We derived conditions on the desirability of monitoring and demonstrated that the optimal non-linear monitoring schedule generally follows the optimal labor wedge. Monitoring is more desirable when redistributive taxation creates larger distortions in labor supply. Moreover, optimal labor taxes can optimally be above 100 percent when monitoring is allowed for. At the endpoints of the earnings distribution labor wedges – including taxes and the implicit subsidy on work due to monitoring – are zero in the absence of bunching and with a finite skill level.

Simulations confirmed that the optimal monitoring intensity features a U-shaped pattern with income; very high at the lower end of the earnings distribution, declining towards the middle-income groups, increasing again towards the high-income groups, and becoming constant at the top-income groups. Our simulations demonstrated that marginal tax rates will be higher if the government monitors labor effort, while the labor wedges – including the explicit tax and implicit subsidy of monitoring – decreases. Indeed, monitoring is very effective to alleviate the equity efficiency trade-off.

In practice, monitoring is not infinitely costly as in Mirrlees (1971). By allowing for a monitoring technology we can explain our why work-dependent tax credits for low-income earners, that are employed in the UK, Ireland and New Zealand, are part of an optimal redistributive tax policy. Our findings also show that sanctions for welfare recipients, bonuses for low-income workers, and extensive monitoring of effort or working ability of low-earning individuals are especially desirable in more generous welfare states. Moreover, we can also explain why (large) penalties on low labor effort (high bonuses on high labor effort) are more desirable when the government desires to redistribute more income. Finally, we explain why marginal tax rates larger than 100 percent at the lower end of the earnings distribution, as commonly observed in many countries, can be optimal in the presence of monitoring of work effort.

Chapter 5

The Effect of Capital Taxes on Household's Portfolio Composition and Intertemporal Choice: Evidence from the Dutch 2001 Capital Income Tax Reform [1]

5.1 Introduction

Capital taxation is a contentious issue in public economics. Seminal papers by Atkinson and Stiglitz (1976), Judd (1985) and Chamley (1986) suggest capital should not be taxed at all. However, recent literature suggests that this result only holds in a very specific setting and the optimal tax rate on capital tax is generally non-zero (see for an overview Conesa *et al.*, 2009 and Diamond and Banks, 2010). In addition, many governments create tax incentives for households to hold specific assets, such as owner occupied housing and pension savings, but very little is known about the effectiveness of these subsidies.

In order to calculate the optimal capital tax rate on each asset it is of central concern to know if, and by how much, households respond to tax incentives when they choose their

[1]This chapter is based on Zoutman (2013). I would like to thank Eva Gavrilova, Aart Gerritsen, Bas Jacobs, Henrik Kleven, Sander Renes, Marcel Smeets and Hendrik Vrijburg for useful suggestions and comments. Furthermore, I would like to thank Statistics Netherlands for providing the data for this paper. All remaining errors are my own. The Stata programs used for the computations in this paper are available on request.

portfolio composition, and their level of savings. In this paper I answer this question by exploiting variation of the Dutch 2001 capital income tax reform.[2]

The reform was announced in 2000 and created enormous quasi-experimental variation in the after-tax return on assets. In particular, the reform drove a wedge between the taxation of owner-occupied housing, hereafter referred to as housing wealth, and the taxation of all other assets in household's portfolio, hereafter referred to as financial wealth.

Most households in the Netherlands owning both types of wealth have received a shock to their after-tax return on each of the two wealth types. At the household level, the two shocks are uncorrelated. In addition, the shock provides variation at each level of household income and all levels of (positive) wealth. This allows me to isolate the effect of the tax reform from other changes in the dependent variable that are correlated with income, wealth and many other control variables.

In order to estimate the effect of the reform, I use a specifically designed unique panel dataset provided by Statistics Netherlands over the period 1995-2004. The dataset is based on the Income Panel Investigation (IPO) which keeps track of administrative records of 0.61 percent of the Dutch population, as well as their household members. The original IPO contains individual tax records on capital and labor income collected from both employers and employees for each household member, as well as a large set of control variables collected at both the national and the municipal level. For the purpose of this study the dataset is extended at the household level with administrative data on household portfolios.

The use of this data is one of the main innovations in this paper. In his Presidential Address to the members of the American Finance Association Campbell (2006) points out that to estimate a portfolio choice model the ideal dataset should have the following five characteristics: i.) it should cover a representative sample of the population, ii.) it should contain wealth and break down wealth into categories, iii.) the categories should be sufficiently disaggregated, iv.) the reported data should be sufficiently accurate and v.) households should be followed over time. The IPO dataset exhibits all of these characteristics, and of top of that the 2001 tax reform offers quasi-experimental variation in the return on assets. Such data is not available in the US and Canada and, as such, most previous studies had to rely on cross-sectional survey data.[34] Therefore, unlike most other

[2]See Bovenberg and Cnossen (2001) for a comprehensive overview of the tax reform.

[3]See e.g. Hubbard (1985), King and Leape (1998) and Poterba and Samwick, 2003 and Alan *et al.* (2010)

[4]A notable exception is the working paper Alan and Leth-Petersen (2006) which uses administrative panel data around a capital-income tax reform in Denmark in the 80's.

studies in the literature, in this study I can control for unobserved household heterogeneity. This could be important, because unobserved heterogeneity, such as earnings ability, may be strongly correlated to the marginal tax rate of the household.

In addition, to my knowledge this is the first study to directly link portfolio choice to a tax-induced change in the after-tax return on assets. Unlike the Netherlands, most other tax systems in the world have some sort of capital gains tax. As a result, in other countries the capital tax affects both the expected return and the variance of the return, making it impossible to isolate the effect of taxation on the expected returns. Further, most other tax systems in the world tax all assets more or less synthetically. As such, it is impossible to separately identify asset-specific tax rates. In this respect the 2001 capital tax reform is an ideal experiment, because it drives a wedge between two asset types that were previously taxed synthetically. The estimates can therefore directly be used to predict the effect of the tax rate on a particular asset on the demand of the asset, and as such, they may be of large value to policy makers.

In order to aggregate the reform into economically meaningful statistics I impose some structure by developing a semi-structural model of the household's investment and savings decisions. In the spirit of the consumption-based capital asset pricing model (CCAPM)[5] I split the household decision in a first stage where the household chooses his level of savings and a second stage in which the household chooses its optimal portfolio composition. In the latter stage, I derive the optimal share invested in financial wealth and show that it should be a function of the gross expected after-tax return on financial and housing wealth, and the variance-covariance matrix of the returns. As such, the change in portfolio composition over the reform is a function of the change in each of these two components. I do not observe the after-tax return, since capital gains, and post-reform cash-returns are not recorded. In addition, I do not observe the variance-covariance matrix of the returns. However, by taking the assumption that the change in the expected before-tax returns and the change in the variance-covariance matrix are uncorrelated to the change in the capital-tax rate at the household level, it is still possible to identify the effect of a tax-induced change in the after-tax return on portfolio composition.

The validity of this assumption is discussed in detail in section 4 of this paper. However intuitively, the Netherlands is a small open economy. Therefore, the before-tax expected returns and variances in the capital market are unlikely to be correlated to the change in the tax rate. In addition, Domar and Musgrave (1944) already established that capital-income taxation may decrease the variance in after-tax returns, for given variance in the before-tax returns. However, the Dutch tax system only taxes cash returns. Since capital

[5] See e.g. Markowitz (1952) and Merton (1969, 1971, 1973).

gains are much more volatile than cash returns[6] the impact of the Dutch capital tax system on the variance of after-tax returns is likely negligible.

Since Hall (1978) the empirical literature on the trade-off between savings and consumption has focused on estimating the Euler-equation. In particular, the fundamental parameter of interest since Hansen and Singleton (1983) and Hall (1988) [7] has been the intertemporal elasticity of substitution. One of the difficulties in estimating a Euler equation is endogeneity. Proper instruments which are correlated to the rate of return but uncorrelated to the rate of consumption growth are difficult to find. However, the Dutch capital tax reform may provide just such an instrument. Unfortunately, the data do not allow me to uncover the consumption of households since I do not observe all objects in the budget constraint.[8] Instead I relate the change in total wealth accumulation to the change in the gross after-tax return on the portfolio. From this equation I retrieve the elasticity of the demand of total wealth with respect to a change in the return on total wealth. Although, the elasticity of intertemporal substitution cannot be retrieved from this equation, the sign of this elasticity equals the sign of the elasticity of intertemporal substitution. In addition, the estimated elasticity is interesting to policy makers in its own right, because it shows how capital taxation affects total accumulated wealth.

A particular concern in studies that use a tax-reform to identify the effect of taxation upon behavior is the endogeneity of the tax rate. In this case, the post-reform tax on housing and financial wealth may depend on the change in housing and financial wealth. For housing wealth this effect is likely small because housing is taxed together with labor income and labor income is orders of magnitude larger than housing income for most households. However, the marginal tax rate on financial wealth is crucially dependent on whether or not financial wealth exceeds a threshold. Therefore, I use the pre-reform data from 1995-1999 to construct a model to predict what financial wealth would have been without the reform. New tax rules are applied to predicted wealth levels to predict what the tax rate would have been without the reform. In the final regression I use the instrumented tax rates in order to determine the change in the after-tax return. This instrumentation strategy is standard in the estimation of the elasticity of taxable labor income (see e.g. Feldstein, 1995, Gruber and Saez, 2002 and Weber, 2013).

A second source of endogeneity may arise from the effect of wealth on portfolio composition. Empirical evidence shows a strong correlation between portfolio allocation and the level of wealth (see e.g. Mankiw and Zeldes, 1991, Poterba, 2002 and Campbell,

[6]See e.g. LeRoy and Porter (1981) and Shiller (1981) who show that capital gains in the stock market are much more volatile than may be expected by changes in the interest rate and dividends.

[7]See Attanasio and Weber (2010) for an overview of the literature.

[8]The most important missing variable are the capital gains.

2006). Hence, a change in wealth may lead to a change the optimal portfolio allocation. Since, the portfolio return depends on the the portfolio allocation the change in wealth may indirectly affect the change in the portfolio return. However, an instrument is readily available. I estimate the change in portfolio allocation using the change in the after-tax return on each asset. The instrument is valid because the after-tax return on each asset is unlikely to be correlated with the change in wealth, except through the change in the return on the entire portfolio. As a result, the portfolio-allocation stage of the household decision process can be used to instrument for the stage where the household chooses between savings and consumption.

In the estimations I use 1999 as a base year since decisions in 2000 may already have been affected by announcement effects. I look at long-run effects up to 2004 and short-run effects up to 2001. In the long-run I find that a tax-induced change in the after-tax return on financial and housing wealth has statistically significant but modest effects on portfolio composition. The central estimate is that a one-percent increase in the tax on financial wealth decreases the share invested in financial wealth by only 0.033 percent. The elasticity with respect to the after-tax return on housing has the expected negative sign, but the effect is economically negligible.

Furthermore, I find that accumulated wealth in the period 1999-2004 is positive and significantly correlated to the change in the after-tax return on the portfolio. However, again effects are rather modest. A 1 percent increase in a hypothetical tax that covers all wealth would decrease accumulated wealth by only 0.036 percent. The short-run elasticities are only slightly lower than long-run elasticities. This indicates that households respond to the change instantaneously.

In the sensitivity analysis I split the sample, and estimate the elasticity of single households and households that had high levels of wealth. The elasticity for these groups is significantly larger. However, this result has to be interpreted with some caution because the sample of single households is rather small and the measurement error may be smaller for the rich households than the poor households. Therefore, it is difficult to separate real heterogeneity in the behavioral response from possible attrition through the measurement error.

Two potential caveats should be discussed. First, unlike other studies on the effect of taxation on portfolio allocation, such as King and Leape (1998) and Poterba and Samwick (2003), I only study the intensive margin of the portfolio choice. Since almost all households own at least a little bit of financial wealth through a savings or demand deposit, I exclude all households that do not own a house. Arguably, the fact that I have aggregated the portfolio to only two assets alleviates the severity of this omission. In addition, the decision of buying a house is fundamentally different from other investment decisions due to the fundamental indivisibility of buying a house (see also Cocco, 2005), complicating the introduction of an extensive margin in this study.

A second caveat is the fact that I do not observe the wealth employees have in their pension fund. Total savings of the pension funds amount to 138% of GDP in 2013 and are as such a significant portion of total savings for Dutch employees. Unfortunately, for the studied period, pension funds did not keep any records on payments by individual employees and as such there is no way to reconstruct pension savings for households. However, households cannot alter their pension savings on the basis of the tax reform because the level of contributions is set in negotiations between employers and unions. In addition, the reform has had no impact on pension savings, since pensions were untaxed both before and after the reform.

Most classical studies, relying on (repeated) cross-sectional survey data, find a strong effect of taxation on portfolio composition (see e.g. Hubbard, 1985, King and Leape, 1998 and Poterba and Samwick, 2003). However, in a recent article Alan *et al.* (2010) partially control for unobserved household heterogeneity by cleverly exploiting intra-household variation in the capital income tax rate in the Canadian tax system. They find a significant but relatively modest effect of taxation on portfolio composition.

A direct comparison between previous studies and this study is complicated by the fact that previous studies could not directly link portfolio composition to the after-tax return on assets. However, findings in this paper broadly correspond with the modest behavioral response found in Alan *et al.* (2010), suggesting that not controlling for unobserved household heterogeneity leads to an overestimation of the impact of taxation.

A large literature has been devoted on estimating the Euler equation using a variety of datasets and instruments. Estimates of the intertemporal rate of substitution vary between 0.65 and 1 (see e.g. Attanasio and Weber, 1989, Attanasio and Weber, 1993, Blundell *et al.*, 1994, Banks *et al.*, 1994 , Attanasio and Weber, 1995 and Engelhardt and Kumar, 2009). This paper adds to this literature by estimating the effect of taxation on intertemporal choice behavior. In particular, the 2001 tax reform provides a strong instrument for the change in the after-tax return on assets. Unfortunately, there is no direct relationship between the elasticity of the after-tax return on wealth estimated in this paper and the intertemporal elasticity of substitution. However, Attanasio and Wakefield (2010) simulate a life-cycle model where the net-after tax return on assets in the UK is increased from 2 to 2.5 percent. Their simulations show that such an increase in the return significantly increases accumulated wealth if the elasticity of substitution equals one. A similar policy analysis using estimates from this study shows only a modest affect on accumulated wealth. The large difference between the estimated effect in this paper and the simulation tentatively suggest that the estimated elasticity of intertemporal substitution in this study is smaller than in the base-line simulations of Attanasio and Wakefield (2010).

This paper is organized as follows. In the next section explains the 2001 tax reform in detail. In the third section discusses the IPO data. The fourth section introduces the econometric specification. The main results are presented in the fifth section. The sixth section presents some sensitivity analysis and the final section concludes.

5.2 The 2001 Tax Reform

The 2001 tax reform was officially announced by the Dutch government in mid-2000. Rates, bracket thresholds, income definitions and tax deductions all changed. Also, the new system introduced a wedge between the taxation financial and housing wealth. The reform drastically changed incentives to for portfolio composition and savings. In this section I will highlight how the tax reform has affected incentives through the households' intertemporal budget constraint.

In the Netherlands, household wealth has four components, each of which are taxed according to a different tax-regime: i.) financial wealth, ii.) housing wealth, iii.) tax-deferred wealth, iv.) ownership of small firms and closely-held corporations. Financial wealth is the difference between financial assets such as bank accounts, stocks, bonds and real estate, and loans. Housing wealth is the difference between the value of the owner-occupied house and the mortgage on the house.

The largest part of tax-deferred wealth are so-called second pillar pension savings. Collective labor agreements between employers and employees require firms to set up pension funds or join in sectoral pension funds. Total savings of the pension funds amount to 138% of GDP in 2013 and are as such a significant portion of total savings for Dutch employees. Unfortunately, for the studied period, pension funds did not keep any records on payments by individual employees and there is no reliable way to reconstruct pension savings for households. As such, I have no choice but to ignore tax-deferred wealth in this study. Fortunately, the 2001 reform did not affect the taxation of these assets. Furthermore, behavioral responses in these savings at the household level are unlikely since the size of the contributions are set in negotiations between unions and employers. Up to 2003 there were no major changes in pension benefits, entitlements or contribution. In 2003 pension premiums did increase significantly due to the aftermath of burst of the dot-com bubble. However, the change in pension contributions was likely strongly correlated with household labor income and age, both of which I can control for in my estimation, and after controlling for those factors, only weakly correlated to the change in the capital tax rates.

In addition, this study ignores wealth stemming from small firms and closely-held corporations. The 2001 reform did change the taxation of wealth and income from closely-held corporation. However, households that own closely-held corporations likely have the possibility to shift income between various tax bases (see e.g. De Mooij and Nicodème, 2008). As such, I remedy this problem by simply excluding all households that owned close-held corporation, or small firms from my dataset. The focus of this study is therefore on housing and financial wealth.

In the remainder of this section I will explore the changes generated by the tax reform through the household's intertemporal budget constraint. The linearized intertemporal budget constraint of household i in period t is given by:

$$W_{i,t+1} + C_{it} = \left(1 - T_{it}^{L}\right) Y_{it} + V_{it} + R_{it}^{W} W_{it}, \tag{5.1}$$

where W_{it} denotes total household wealth of household i in time t, C_{it} consumption, T_{it} the marginal tax rate on labor income of the primary earner, Y_{it} gross labor income of the primary income earner, and R_{it}^{W} the gross-return rate on total wealth after capital taxes. The actual budget constraint is non-linear because the income is taxed progressively. However, linearizing the budget constraint simplifies the exposition considerably and is useful in deriving the relationship between behavior and the marginal tax rate (see also Saez, 2001, Gruber and Saez, 2002). The term V_{it} denotes virtual income, and contains a correction term for the fact that the actual budget constraint is non-linear, as well as net household income that does not pertain to labor of the primary earner.

The after-tax return on wealth is crucially dependent on the asset mix in the portfolio, since the different categories of wealth holdings face a different tax regime. Hence, it is useful to split up total wealth into financial and housing wealth:

$$W_{it} \equiv W_{it}^{F} + W_{it}^{H},$$

where, W_{it}^{F} is financial wealth, the difference between financial assets and loans, and W_{it}^{H} housing wealth, the difference between the value of the owner-occupied house and the mortgage resting on the house. Hence, the total gross after-tax return on wealth, R_{it}^{W} can be subdivided in the gross after-tax return on financial wealth and the gross after-tax return on housing wealth:

$$R_{it}^{W} = \alpha_{it}^{f} R_{it}^{F} + \left(1 - \alpha_{it}^{f}\right) R_{it}^{H},$$

where R_{it}^{j} is the after-tax return on wealth type j and α_{it}^{f} the share of financial wealth in the portfolio. The after-tax return on each asset can be characterized by the following equation:

$$R_{it}^{j} = 1 - \tau_{it}^{j} + \left(1 - T_{it}^{j}\right) R_{it}^{j} + R_{it}^{j*} \quad \forall \quad j \in \{F, H\},$$

where τ_{it}^{j} is the wealth-tax on wealth type j, R_{it}^{j} is the net taxable return on asset j and T_{it}^{Wj} the marginal tax rate over return j. Finally, R_{it}^{j*} is the untaxed return.

Before the reform the wealth tax was levied on the part of total household wealth that exceeded some threshold. The threshold value in turn depended on household characteristics X_{ib}.. The threshold was larger for couples than for singles, but independent of portfolio composition:

$$\tau_{ib}^{j} = \tau_{ib}\left(W_{ib}, X_{ib}\right) \quad \forall \quad j \in \{F, H\},$$

where the subscript b denotes the base year and X_{ib} is the status of the household.

Cash returns were taxed synthetically with labor income of the primary earner according to a non-linear, progressive tax-system.[9] In addition, for real-estate the government taxes an imputed rent. Capital gains were not taxed at all. Tax rates are age dependent since people over 65 do not have to pay the social premiums relating to the general pension. As such, their effective marginal tax rates are generally lower. In addition, there was a general tax deduction which depended on household status. Thus, the tax function could be expressed as follows :

$$T_{ib}^{j} = T_{ib}\left(Y_{ib} + \sum_{j \in \{A,H,M\}} r_{ib}^{Wj} W_{ib}^{j}, X_{ib}\right) \quad \forall \quad j \in \{F, H\}.$$

After the reform, the government introduced a tax-system based on imputed returns on financial wealth. Financial wealth above a threshold, which was again larger for couples than singles, are presumed to receive a return of 4 percent. The 4 percent in turn was subject to a tax rate of 30 percent. Effectively, the presumptive capital tax is equivalent to a wealth tax of $30\% \times 4\% = 1.2\%$. For future reference I will refer to this tax simply

[9]The marginal tax rate I use for this study consists of two parts: general social insurance premiums and taxes. However, I will treat both as taxes since there is no relationship between the payment of social insurance premiums and benefits.

as a wealth tax. The new wealth tax does not pertain to wealth from housing. As such, the after-reform wealth-tax on assets is given by:

$$\tau_{ir}^{A} = \tau_{ir}\left(W_{ir}^{F}, X_{ir}\right),$$

where subscript r stands for all post-reform years. Housing wealth is no longer subject to the wealth tax:

$$\tau_{ir}^{W^H} = 0.$$

For financial wealth the capital-income tax is abolished such that:

$$T_{ir}^{W^F} = 0.$$

Capital-income pertaining from housing wealth is still taxed synthetically with labor income from the primary earner. However, the general tax deduction is abolished and replaced with a tax credit which depends on household type and employment status. In addition, the rates in the income tax have changed. The post-reform income tax can therefore be expressed as:

$$T_{ir}^{W^j} = T_{ir}\left(Y_{ir} + \sum_{j \in \{H,M\}} r_{ir}^{W^j} j_{ir}, X_{ir}\right) \quad \forall \quad j \in \{H, M\},$$

where $T_{ir}(.)$ is the post-reform income tax rate.

Table 5.1 gives an overview of the changes in deductions, tax credits, threshold levels and tax rates for a single household. All amounts are expressed in 1999 euros. As can be seen, the wealth tax has increased from 0.7 to 1.2 percent and the tax exempt threshold has been lowered drastically. This increase in wealth taxes is offset by the fact that housing is now wealth-tax exempt. The marginal income tax rate has decreased for households at the bottom and the top. However, the tax rate has increased for some households that used to be in the third bracket and are now in the fourth bracket. In addition, the income definition has changed since actual return of financial wealth is no longer taxed. However, this is unlikely to affect the marginal tax rate much since cash returns from financial wealth are generally much smaller than labor income for most households.

The tax-reform created a wedge between the taxation of housing wealth, and the taxation of other financial assets, by excluding the former from the new wealth tax and the latter from the income tax. The shock has not affected all households symmetrically. In particular, in the market for financial assets the abolishment of the capital-income tax has

Table 5.1: Overview of the Tax System

	Pre-reform 1999			Post-reform 2001		
Wealth Tax						
Applies to	All Wealth			Financial Wealth		
General Tax Deduction	89,395			16,818		
Tax rate	0.70%			1.20%		
Income Tax						
Applies to	Full Synthetic Income			Labor and Housing Income		
General Tax Credit	0			3,284		
General Tax Deduction	3,993			0		
Tax Brackets	Starting	Up to	Percentage	Starting	Up to	Percentage
Bracket 1	0	6,807	35.75%	0	14,209	32.35%
Bracket 2	6,807	21,861	37.05%	14,209	25,808	37.60%
Bracket 3	21,861	48,080	50%	25,808	37,408	42%
Bracket 4	48,080	∞	60%	37,408	∞	52%

Note: The table gives an overview of the pre- and post-reform wealth and income tax. Deductions and credits apply to a single household without children. Tax rates apply to all income earners below 65. All monetary values are expressed in 1999 euros.

stronger effects for households with high synthetic income than for households with low synthetic income, due to tax progressivity. Additionally, because the tax rate on capital income is dependent on the earnings of the primary income earner, the size of the shock in the tax rate for given household income depends on the division of earnings within the household. Specifically, for a given household income, if the incomes of primary and secondary earners are relatively close, the tax rate on income earned by the primary earner is relatively low.[10] Also, the threshold of the wealth tax has shifted down affecting households that were previously below the threshold, but were not after the reform. Finally, the pre-reform tax on capital-income was only levied over cash returns. As such, households with relatively low cash returns and high capital gains paid a lower tax on their assets than households with high cash returns and low capital gains. This assymetric treatment of returns is now abolished since the post-reform tax rate is levied independent of the division of returns within the asset.

Furthermore, in the market for owner-occupied houseing the reform in the rates of the income tax have increased the tax rates for some households, but decreased the tax rates for other households, thereby providing a source of variation in the return on housing wealth. In addition, the abolishment of the wealth-tax on housing has affected those households above the wealth-tax threshold but has not affected those that were below the threshold.

As such, the reform offers a myriad of sources for identifying the effect of a change in the tax rate on portfolio composition and savings. The next section will present the data used to exploit the variation caused by the tax reform.

5.3 Data Description

The data used for the analysis is the Income Panel Investigation (IPO) provided by Statistics Netherlands. The IPO follows about 0.61 percent of individuals in the Dutch population in the period 1989-2010, and it follows all the household members of the original 0.61 percent. Individuals in the panel are unaware of their participation in the sample. In 1989 the dataset contained data on 210,000 individuals in 75,000 households. The size of the sample has steadily increased to correct for the increase in the population by adding newborns and immigrants such that the final sample size in 2010 consists of 270,000 individuals in 94,000 households. The sample is not entirely representative for the

[10]Note that this source of variation runs counter to the one exploited in Alan *et al.* (2010) in the Canadian tax system. In Canada, households can choose which partner pays capital income taxes. As such, households with a more unequal division of household income face a lower capital-income tax rate.

Dutch population because some groups were deliberately oversampled. However, sampling weights are provided.

For the purpose of this study, the IPO has been extended to contain administrative data on household wealth and portfolio composition. Data are collected at the household level through administrative tax records. The data contain financial wealth in three broad categories: financial wealth, housing wealth and closely-held corporations. All wealth is subdivided in assets and liabilities and for financial wealth they are further subdivided into saving accounts, stock, bonds, real estate and other assets. Loans, including mortgage loans, savings and checkings account are valued in their cash value. Stocks and bonds are valued at market prices. Dutch municipalities measure the value of all real estate in order to collect property taxes. These valuation have been used for real estate. Unfortunately, I have no data on the height of the property taxes themselves. However, there was no reform in the property taxes in the studied period. Therefore, this omission likely does not influence my results.

The dataset also contains some information on the taxable part of capital income. In particular, the data have some information on the cash returns. In the data, cash returns on financial wealth are measured as the sum of dividends, the difference between interest received and interest paid on all loans except the mortgage, and imputed returns on all real estate except the owner-occupied house divided by total financial wealth. Cash returns on housing wealth are defined as the difference between the imputed return on the owner-occupied house and interest paid over the mortage divided by total housing wealth. In the pre-reform period cash returns on both financial wealth and housing wealth are available. In the post-reform period cash returns on housing are observed but there are no accurate observations on returns on financial wealth since the government did not need to collect this information anymore after abolishing the capital-income tax on these assets. Note that even in the pre-reform actual returns on assets may be significantly larger than the cash returns due to the fact that capital gains are not reported at all.

The dataset also contains additional information on households such as primary income from labor-, transfer- and subsidy-income, gross income, taxable income after deductions, net income and disposable income at the individual level as well as many other income-related variables. Demographic variables such as age, region and country of origin are also included.

I study the data of households in the period 1995-2004. By exploiting the pre-reform period of 1995-1999, I can control for portfolio dynamics. In 2000 the reform was announced. Announcement effects are likely to bias the estimates and as such I do not use data on wealth or capital income from 2000. The main estimation period runs from 1999-

2004 and allows for estimation of the effects of the reform in both the short, 1999-2001, and long run, 1999-2004.

From the original data I select a balanced panel comprising the period 1999-2004. From these observations, I select the households whose structure has remained unchanged throughout the sample period. In particular, observations where households merged by marriage or cohabitation, or separated by divorce or death of one of the main partners were deleted from the sample. It is likely that the savings behavior of these households changed for reasons entirely unrelated to the tax reform. Observations where the size of the household increases through childbirth or decreases by one of the children leaving the house remain in the sample. In addition, I removed individuals that were in an institutional house during any of the years.

As mentioned in the previous section, I also filter out all households that own closely-held corporations, as well as self-employed individuals. In addition, I have filtered out all households that do not own positive financial and/or housing wealth. That is, I remove all non-home owners from the sample. Therefore, in this study I focus entirely on the intensive-margin portfolio choice. Finally, I remove outliers defined as households with reported cash returns lower than -20 percent or higher than 50 percent. The large returns for these households may stem from households that underreport their wealth. This is a particular concern for households with low wealth since they were not subject to the wealth tax and as such could not be penalized for underreporting their wealth. The summary statistics of the final sample for post- and pre-reform periods can be found in table 5.2. The appendix reports summary statistics for the unfiltered sample. All monetary values in the table are reported in 1999 euros. Pre-reform, net-return and after-tax returns on assets have been calculated by dividing cash and imputed income on wealth by wealth. In the post-reform period, returns are calculated under the assumption that before-tax returns were equal to returns in the pre-reform periods such that all the variation is driven by the change in the capital tax.

5.4 Methodology

It is clear that the tax reform provides many sources of variation in household's investment decisions. The upside of such a large reform is that the after-tax return on assets changed for almost all households. In addition, it has been shown in a large number of studies that portfolio choice and savings are strongly correlated to wealth and income (see e.g. Poterba, 2002, Campbell, 2006 and Attanasio and Weber, 2010). However, figures 5.1-

Table 5.2: Summary Statistics for Main Estimation Panel

Variable	Pre-reform	(1995-1999)	Postreform	(2001-2004)
	Mean	Mean Std	Mean	Mean Std
Single	0.082	0.272	0.063	0.242
Couple	0.376	0.484	0.391	0.488
Single with child	0.010	0.098	0.007	0.081
Couple with child	0.532	0.499	0.540	0.498
Nr Children¡18	1.002	1.089	1.101	1.177
Nr Household Members	3.072	1.206	3.350	1.248
Age	41.117	9.339	45.797	9.435
Wealth	118,965	118,343	219,544	244,821
Share Financial Wealth	0.279	0.261	0.220	0.209
Primary Household Labor Income	49,143	23,935	58,093	31,817
Effective Wealth Tax Rate	0.005	0.003	0.007	0.006
Marginal Income Tax Rate	0.438	0.077	0.423	0.052
Net After-Tax Return Financial Wealth	0.007	0.203	0.006	0.034
Net After-Tax Return Housing Wealth	-0.087	0.313	-0.023	0.144
Net After-Tax Return Total Wealth	-0.029	0.132	-0.013	0.035
Nr of observations	12,831		12,831	

Note: Summary statistics of the filtered sample. Pre-reform nr of observations were taken in 1999. All monetary values are expressed in 1999 euros. Post-reform returns are calculated under the assumption that before-tax returns remained equal, such that only the tax rate changes. Mean std denotes the mean standard deviation over all years.

5.6 show scatterplots of the relative change in the after-tax return on financial, housing and total wealth as a function of household pre-reform primary labor income and wealth under the assumption that the before-tax returns remain constant. As can be seen from the figures, there is very weak correlation between the change in the after-tax return and income and wealth. In addition, there is variation at all levels of wealth. It is therefore possible to control for these variables explicitly without soaking up any of the variation in the after-tax returns.

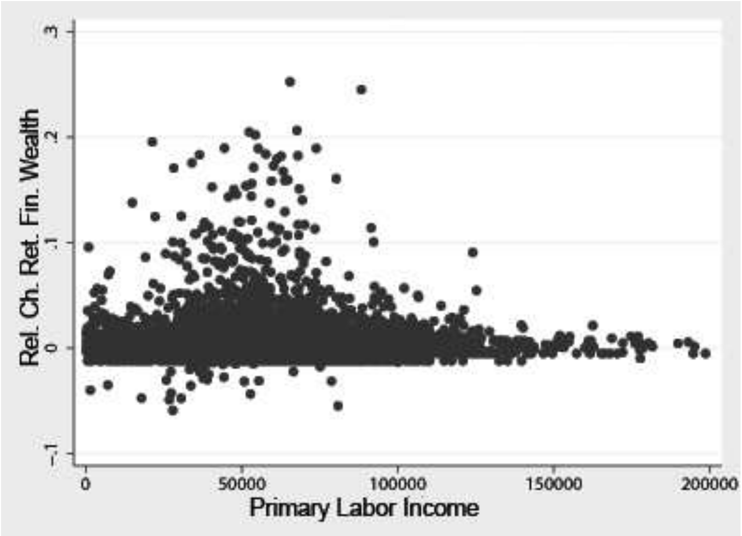
Note further that the variation in the return on housing and total wealth is very large for low levels of wealth. This may be due to the fact that homeowners with little housing wealth are highly leveraged. In that case, a small change in the tax rate induces a large change in the after-tax return on housing wealth. In my robustness analysis, I focus on a subsample with relatively high wealth-levels to see if excluding this group has a strong impact on my estimates.

In order to estimate the effect of the simultaneous change in the wealth and capital income tax on portfolio composition and savings it is necessary to take some structural assumptions in order to aggregate the reforms into a statistic. In addition, cash returns on wealth in the post-reform period are not accurately observed and capital gains are not observed at all. Finally, there may be multiple sources of endogeneity related to the non-linearity in the tax rate, and the relationship between wealth accumulation and portfolio composition. In this section I will first derive a semi-structural model for asset demand and a semi-structural model for wealth accumulation. Finally, a separate subsection explains the strategy to deal with potential endogeneity.

5.4.1 A Model of Asset Demand

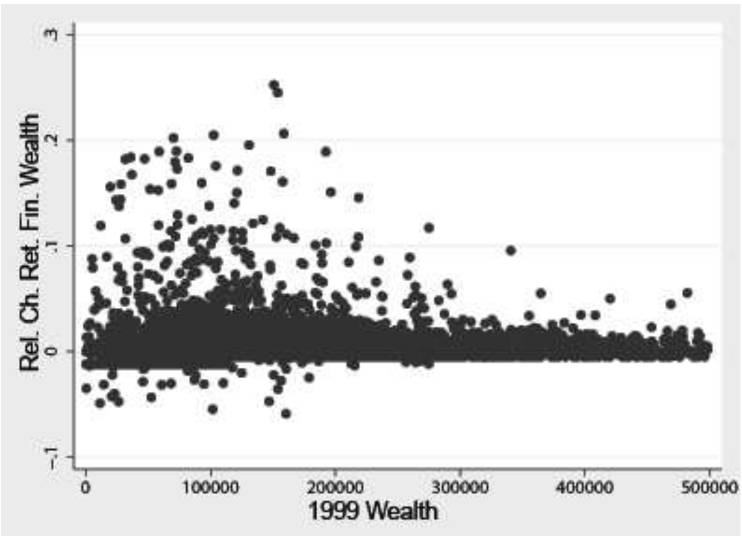
In this paper I study two household decisions. I study the trade-off between consumption and savings, and I study the trade-off between the different assets in the portfolio. In CCAPM it is shown that you can split up this decision into two stages. In the first stage households choose how much to save and how much to consume. In the second stage they choose their optimal portfolio-composition by maximizing a mean-variance utility function. I follow this approach and split up my estimation in a first stage where the dependent variable is accumulated wealth, and a second stage where the dependent variable is the share of financial wealth in total wealth. By backward induction, I will first derive the estimating equation in the second stage where the household chooses its optimal portfolio composition. Assume the household chooses its portfolio shares in each

Figure 5.1: The Change in Return on Financial Wealth Between 1999-2001 as Function of Base Year Income



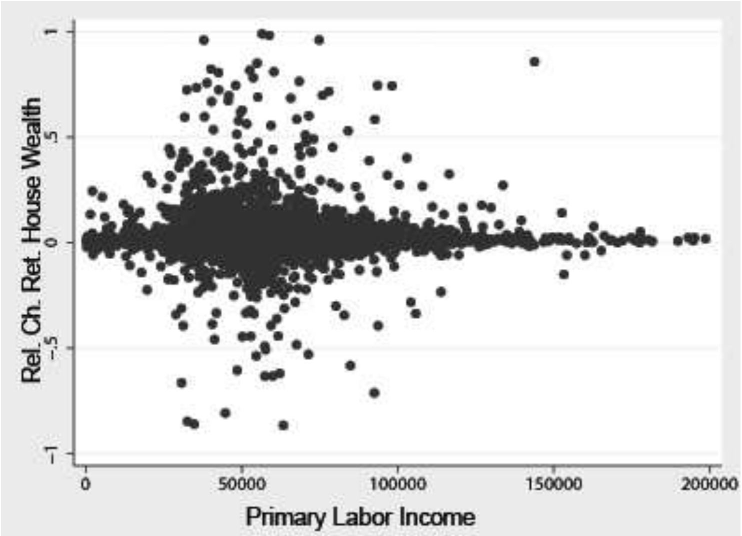
Note: Figure shows a scatterplot for all households of the relative change in the gross after-tax return on financial wealth between 1999-2001 against base year primary labor income.

Figure 5.2: The Change in Return on Financial Wealth Between 1999-2001 as Function of Base Year Wealth



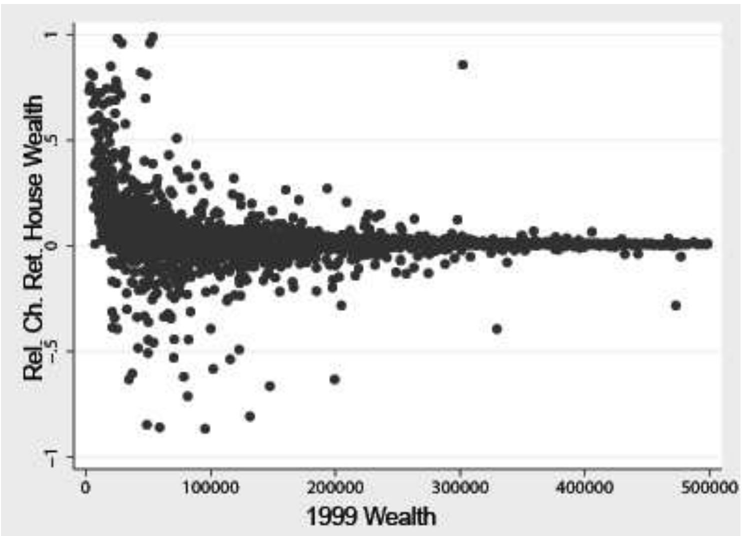
Note: Figure shows a scatterplot for all households of the relative change in the gross after-tax return on financial wealth between 1999-2001 against base year total wealth.

Figure 5.3: The Change in Return on Housing Wealth Between 1999-2001 as Function of Base Year Income



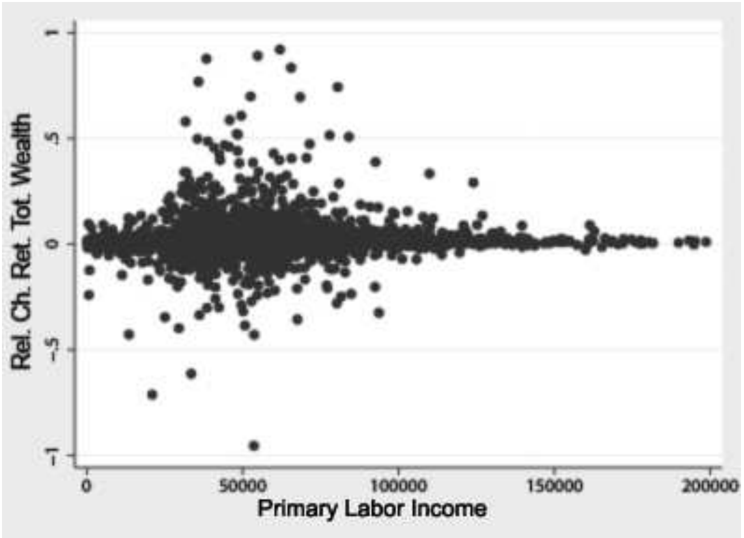
Note: Figure shows a scatterplot for all households of the relative change in the gross after-tax return on housing wealth between 1999-2001 against base year primary labor income.

Figure 5.4: The Change in Return on Housing Wealth Between 1999-2001 as Function of Base Year Wealth



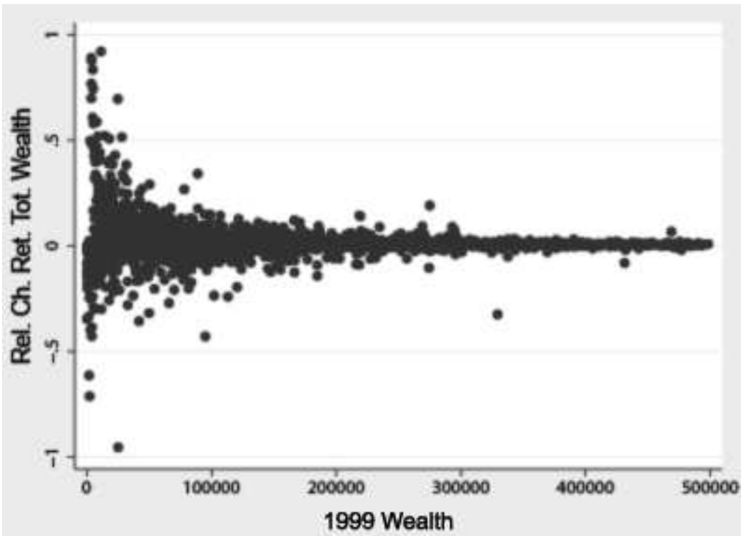
Note: Figure shows a scatterplot for all households of the relative change in the gross after-tax return on housing wealth between 1999-2001 against base year total wealth.

Figure 5.5: The Change in Return on Total Wealth Between 1999-2001 as Function of Base Year Income



Note: Figure shows a scatterplot for all households of the relative change in the gross after-tax return on total wealth between 1999-2001 against base year primary labor income.

Figure 5.6: The Change in Return on Total Wealth Between 1999-2001 as Function of Base Year Wealth



Note: Figure shows a scatterplot for all households of the relative change in the gross after-tax return on total wealth between 1999-2001 against base year total wealth.

period to maximize a mean-variance utility function of its portfolio returns:

$$f_i\left(E_t\left[R_{it}^W\right], E_t\left[\left(R_{it}^W - E_t\left(R_{it}^W\right)\right)^2\right]\right),$$

where E_t is the expectation operator in period t. It is well-known since Domar and Musgrave (1944) that capital-income taxation generally affects both the mean and the variance of the investment. A positive capital-income tax rate decreases the mean return on investment but at the same time it decreases the variance by letting the government share part of the losses. On the other hand, a wealth tax only affects the mean return since the size of the wealth tax is unrelated to the return obtained on the asset. Therefore, it might be expected that abolishing the capital-income taxes on the cash-returns of financial wealth and the abolishment of wealth taxes on housing wealth affected both the mean return and the variance of the assets. In this case, identification becomes difficult because it is unclear whether the behavioral changes of the reform were caused by a change in the mean or in the variance of the return. Fortunately for the econometrician, the Dutch government only taxes cash returns. Of these, both imputed returns on real estate and interest payments on loans, savings accounts and government bonds are generally known before the household makes an investment decision. Dividend pay-outs are arguably somewhat more volatile, but are still far less volatile than capital gains (see e.g. Shiller, 1981 and LeRoy and Porter, 1981). Therefore, I take the strong assumption that the taxable part of returns, R_{it}^j, is non-random at period t. I assume that untaxed returns are random variables and the vector of untaxed returns on financial and housing wealth follows a normal distribution:

$$r_{it}^* \tilde{} \mathcal{N}\left(\mu_{it}, \Sigma_{it}\right).$$

where μ_{it} is a vector, $\left[\mu_{it}^F, \mu_{it}^H\right]$ of the mean returns on financial and housing wealth, and Σ_{it} the variance-covariance matrix of returns. The expected portfolio return can be written as:

$$E_t\left[R_{it}^W\right] = E_t\left[\alpha_{it}^F R_{it}^F + \left(1-\alpha_{it}^F\right) R_{it}^H\right]$$

The variance of the return is given by:

$$E_t\left[\left(R_{it}^W - E_t\left(R_{it}^W\right)\right)^2\right] = \left(\alpha_{it}^F\right)^2 \sigma_{it}^{FF} + 2\alpha_{it}^F\left(1-\alpha_{it}^F\right)\sigma_{it}^{FH} + \left(1-\alpha_{it}^F\right)^2 \sigma_{it}^{HH},$$

where σ^{jk} denotes the covariance of assets j and k. As can be seen, the variance of the portfolio is independent of the tax rate. From the first-order condition of the household one can derive the demand for the share of financial wealth in the portfolio as a function of the after-tax returns on financial wealth and housing wealth. Asset pricing theory predicts that the share increases in the return on financial wealth and the variance in housing wealth and decreases in the return on housing wealth and the variance in financial wealth. Assume, as is standard in the literature (see e.g. King and Leape, 1998), that the log share of assets is log-linear in each of the returns, and separable in all returns and the variance. In that case the log share of financial wealth can be written as:

$$\ln \alpha_{it}^{F} = \zeta_i + \eta_t + \sum_{j \in \{F,H\}} \varepsilon^j \ln E_t \left[R_{it}^{J} \right] + g_i \left(\Sigma_i \right) + \nu_{it},$$

where ζ_i, is a household-specific intercept, η_t a period-specifc effect, ε^j is the elasticity of the share with respect to return wealth type j, $g_i(.)$ is a general function of the variance-covariance matrix and ν_{it} is the error term. In order to estimate this model, using variation induced by the reform, I take first differences over the reform:

$$\Delta \ln \alpha_{ir}^{F} = \gamma + \sum_{j \in \{F,H\}} \varepsilon^j \Delta \ln E_r \left[R_{ir}^{j} \right] + \Delta \nu_{ir},$$

where Δx_{ir} denotes the difference of variable x between the post-reform period r and the base year b and $\gamma^F = \Delta \eta_t^F$. Further simplify the equation by writing out the expectations:

$$\begin{aligned} \Delta \ln \alpha_{ir}^{F} &= \gamma + \sum_{j \in \{F,H\}} \varepsilon^j \left(\Delta r_{ir}^{j} - \Delta \left[T_{ir}^{j} r_{ir}^{j} \right] - \Delta \tau_{ir}^{j} + \Delta \mu_{ir}^{j} \right) + \Delta \nu_{ir}, \qquad (5.2) \\ &= \gamma + \varepsilon^F \left(\Delta r_{ir}^{F} + T_{ib}^{F} r_{ib}^{F} - \Delta \tau_{ir}^{F} + \Delta \mu_{ir}^{F} \right) + \varepsilon^H \left(\Delta r_{ir}^{H} - \Delta \left[T_{ir}^{H} r_{ir}^{H} \right] + \tau_{ib}^{H} + \Delta \mu_{ir}^{H} \right) + \Delta \nu_{ir}, \end{aligned}$$

where I have used the approximation $\ln(1+x) \approx x$ and the fact that $T_{ir}^{F} = \tau_{ir}^{H} = 0$. Clearly, direct estimation of (5.2) is problematic because we do not observe actual changes in the returns since returns are entirely unobserved after the reform. However, this omitted variable will not bias the final estimates as long as, after controlling for variables X_i, it is uncorrelated to the tax-induced change in the after-tax return on each asset. Control variables in X_i should obviously include variables that somehow influence household investment behavior and may be correlated to the change in the tax rate. I will first introduce the control variables that I will add to the model, before discussing the validity of this crucial assumption.

The first variable I include in X_i is the total sum of primary labor income the household earned during the reform period. Here primary labor income includes all taxes and employee and employer premiums. This variable is meant to capture the amount of disposable income a household had available during the reform period. Clearly, households that had a lot of income during the reform period might save more than households that received less income. In addition, households with more income might use different investment instruments. Finally, primary income may be seen as a good control variable measuring ability of the household.

Furthermore, X_i contains base-year wealth and base-year savings to control for portfolio dynamics such as mean reversion and persistence. In addition, I have added wealth splines to X_i. The wealth splines are dummy variables indicating whether a household was in a specific decile of the wealth distribution in 1999. These spline terms control for possible exogenous dispersion in the wealth distribution, as in Gruber and Saez (2002).

In addition, I control for age of the primary income earner using age dummies. It is likely that households with old primary income earners invest less, and less risky, than households with younger income earners due to the fact that the probability of death increases with age. Finally, I control for household type and household composition.

As mentioned before, the question is whether conditional on the X_i just mentioned, the omitted variables are uncorrelated to the tax-induced changes in the after-tax return. This condition is very likely satisfied in the market for financial wealth. The Netherlands is a small open economy and it is unlikely that the Dutch tax reform affected world market returns in any significant way. On the other hand, returns in the much less international housing market might be affected by the tax reform. However, estimation results remain valid as long as the tax reform affected the housing market symmetrically or its effect was assymetric, but strongly correlated with control variables in X_i. Especially the latter scenario seems likely. Although households with different wealth or income levels may have faced different shocks in their before-tax housing return, it seems unlikely that within these wealth and income classes the change in housing returns was somehow directly related to the tax rate. The assumption allows me to make the following substitution:

$$\varepsilon^F \left(\Delta r^F_{ir} + \Delta\mu^F_{ir}\right) + \varepsilon^H \left(\left(1 - T^H_{ir}\right)\Delta r^H_{ir} + \Delta\mu^H_{ir}\right) + \Delta\nu_{ir} = X_i\beta + \xi_{ir}, \tag{5.3}$$

where ξ_{ir} the new error-term. Through the substitution all variables relating to the change in before-tax return drop out. Note again that X_i should absorb all variation in the after-tax return that is unobserved but possibly correlated to the change in the capital taxes. Inserting (5.3) into equation (5.2) we arrive at an estimation equation with only observable variables:

$$\Delta \ln \alpha^F_{ir} = \gamma + \varepsilon^F \left(T^F_{ib} r^F_{ib} - \Delta\tau^F_{ir}\right) + \varepsilon^H \left(\tau^H_{ib} - \Delta\left[T^H_{ir}\right] r^H_{ib}\right) + X_i\beta + \xi_{ir}. \tag{5.4}$$

Note that a similar equation could be set up for the change in the log share of housing wealth. However, since both shares add up to one, such an equation would give no additional information with respect to the information contained in equation (5.4). If an increase in the return on housing decreases the share of financial wealth, than by definition it must also increase the share of housing wealth by the same percentage. Hence, in order to estimate portfolio allocation in a model with two assets, one only needs to estimate one equation.

The elasticities in equation (5.4) directly relate the change in portfolio share to a tax-induced change in the after-tax return on the asset. This contrasts sharply with estimates in the US and Canada in e.g. King and Leape (1998), Poterba and Samwick (2003) and Alan *et al.* (2010), where portfolio allocation is related to an overall measure of the marginal tax rate on capital income. The results in these studies can inform policy makers whether an increase in the marginal tax rate shifts asset demand from less to more tax-favored assets, but are unable to inform the policy makers about the effect of increasing the tax-favored status of a particular asset by one percent. Such inference can only be made if different assets are taxed according to entirely different rules and effective marginal tax rate can be calculated independently for each type of asset. An even stronger inference can be made when a country reforms its tax system from a system where assets were taxed according to the same rules, to a system where taxes differ along the different type of assets. It is in that respect that the Dutch 2001 capital tax reform gives the econometrician a close to perfect natural experiment.

In addition, the tax systems in the US and Canada tax both cash and capital gains. As such, the capital income tax lowers both the return and the risk of the asset. Hence, in the aforementioned studies it is impossible to directly relate the change in the tax-rate to a change in the after-tax return, without also making strong structural assumptions about the effect of taxation on the variance-covariance structure of asset returns. By contrast the Dutch tax system affects the return but gives close to zero insurance against asset price volatility. As such, the change in investment decisions can be related directly to the change in the after-tax return, allowing for a more fundamental unraveling of the asset-demand equation.

5.4.2 A Model of Household Wealth Accumulation

In the first stage of the household optimization problem a household decides how much to consume and how much to save. The typical approach is to assume that it maximizes life-time utility with respect to the intertemporal budget constraint, yielding a consumption-Euler equation. A large literature starting with Hall (1978) has been devoted to estimating the Euler equation. Of particular interest is the intertemporal elasticity of substitution, the relative increase in the rate of consumption growth as a result of a relative increase in the return on the portfolio. In many models of capital taxation the intertemporal elasticity is a sufficient statistic for the distortion induced by capital taxation. However, directly estimating the Euler equation may be difficult due to endogeneity. Many factors such as business cycle fluctuation likely affect both the rate of consumption growth and portfolio returns. In addition, it is difficult to come up with instruments which are correlated to the interest rate but not directly correlated to the rate of consumption growth. As a result estimates of the intertemporal elasticity of substitution are sensitive to the instrument used.

The 2001 tax reform in the Netherlands is a strong candidate for an instrument, since it affected after-tax returns without additionally affecting consumption directly. Unfortunately, the IPO data cannot be used to deduce household consumption levels, because in the intertemporal budget constraint (5.1) a large part of the returns are unobserved. However, wealth accumulation can be observed. Although, the effect of taxation on wealth accumulation does not give exact information on the elasticity of intertemporal substitution, unless you are willing to take strong structural assumptions, it does give policy makers an indication of the intertemporal distortion created by capital taxation. In addition, what is lacking in terms of estimating the relevant preference parameter, is made up for in precision by using a particulary strong instrument.

To estimate the effect of the 2001 reform on capital accumulation I assume that the demand for log wealth is linear in the expected log return on wealth:

$$\log W_{it} = \zeta_i + \eta_t + \varepsilon \log E_r \left[R_{it}^W \right] + \nu_{ir}., \tag{5.5}$$

where γ is a constant, ε is the elasticity of wealth with respect to the after-tax return on wealth and ξ_{ir} is the error-term. Note that the sign of ε is not a-priori determined by economic theory. An increase in the after-tax return leads to a substitution effect where people consume more in the future and less now. As is noted in Summers (1981) the substitution effect is reinforced by a negative wealth effect. The increase in the financial discount rate decreases the discounted value of future labor income. This wealth effects decreases consumption in each period, and hence, increases savings. However, households with positive wealth holdings also experience a positive wealth effect since the discounted value of their financial wealth increases with the interest rate. As such, households are induced to consume more in each period. In order to achieve this they have to consume part of their current wealth holdings. If the substitution effect dominates ε is greater than zero and vice versa if the wealth effect dominates. Note that the sign of ε corresponds directly to the sign of the intertemporal elasticity of substitution.

Taking first differences over equation (5.5) we arrive at:

$$\Delta \log W_{ir} = \gamma + \delta \Delta \log E_r \left[R_{ir} \right] + \Delta \nu_{ir},$$

where $\gamma = \Delta \eta_t$. The equation can be further simplified by writing out expected portfolio returns and using $\log(1+x) \approx x$:

$$\begin{aligned} \Delta \log W_{ir} &= \gamma + \varepsilon \sum_{j \in \{F,H\}} \Delta \left[\alpha^j_{ir} \left(r^j_{ir} - T^j_{ir} r^j_{ir} - \tau^j_{ir} + \mu^j_{ir} \right) \right] + \Delta \nu_{ir}, \\ &= \gamma + \varepsilon \sum_{j \in \{F,H\}} \Delta \alpha^j_{ir} \left(\left(1 - T^j_{ib} \right) r^j_{ib} - \tau^j_{ib} \right) + \\ & \quad \varepsilon \sum_{j \in \{F,H\}} \alpha^j_{ir} \left(\Delta r^j_{ir} - \Delta \left[T^j_{ir} r^j_{ir} \right] - \Delta \tau^j_{ir} \right) + \Delta \left[\alpha^j_{ir} \mu^j_{ir} \right] + \Delta \nu_{ir} \end{aligned} \quad (5.6)$$

where $\alpha^h_{it} = 1 - \alpha^f_{it}$ is the share of housing wealth in total wealth. Note that the change in before-tax returns is not observed. However, I have already assumed that, conditional on X_i, the change in the unobserved variables is independent to the change in the tax-rate. Hence, also here I can make the following substitution:

$$\sum_{j \in \{F,H\}} \alpha^j_{ir} \left(1 - T^j_{ir} \right) \Delta r^j_{ir} + \Delta \left[\alpha^j_{ir} \mu^j_{ir} \right] + \Delta \nu_{ir} = X_i \beta + \xi_{ir}. \quad (5.7)$$

Substitute equation (5.7) into (5.6) to arrive at the final relationship:

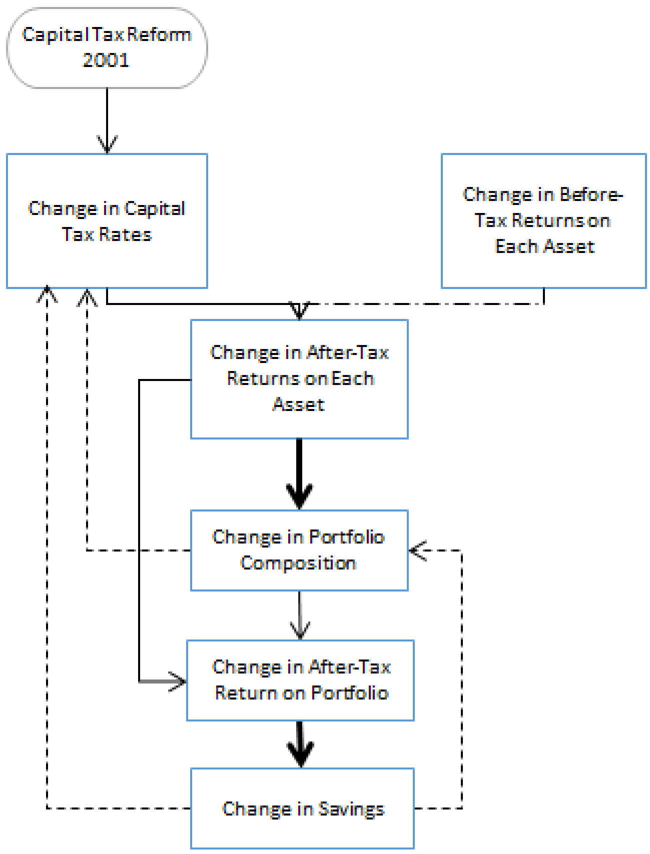
$$\Delta \log W_{it} = \gamma + \varepsilon \left[\sum_{j \in \{F,H\}} \Delta \alpha^j_{ir} \left(r^j_{ib} - T^j_{ib} r^j_{ib} - \tau^j_{ib} \right) - \alpha^j_{ir} \left(\Delta T^j_{ir} r^j_{ib} + \Delta \tau^j_{ir} \right) \right] + X_i \beta + \xi_{ir}. \quad (5.8)$$

Equation (5.8) gives the relationship between wealth accumulation and a tax-induced change in the after-tax return, expressed entirely in observables.

5.4.3 Endogeneity

The chain of causality is described in figure 5.7. In the flow chart solid arrows depict causal relations. The top fat solid arrow represents estimating equation (5.4) and the bottom arrow represents (5.8). Dashed arrows represent potential reversely causal links and the dashed-dotted arrow represents the omitted channel, the effect of the change in

Figure 5.7: A Flow Chart of the Chain of Causality



Note: Figure depicts the chain of causality. Solid lines depict the main causal relationship this study tries to investigate, the two fat solid lines indicate the two main estimation equations, dashed lines indicate potential reverse causality and the dashed-dotted line indicates the unobserved channel.

before-tax returns on after-tax returns. The flow chart starts at the 2001 reform which caused variation in the capital tax rates. In turn, the capital tax rate influenced the after-tax return on each of the two wealth types. Equation (5.4) has portfolio composition as a dependent variable and the after-tax return on each asset as the main independent variables. Both the change in portfolio composition and the change in the after-tax returns on each asset influence the return on the portfolio. Equation (5.4) has the change in wealth as a dependent variable and the change in the after-tax return on the portfolio as an independent variable.

As was argued in the previous subsection, the omitted change in before-tax returns might not be problematic as long as the change in those returns are, conditional on X_i, uncorrelated to the change in the tax-rate. However, there are also three potential reverse causal links. First, the capital tax rate may depend on the change in asset composition in the portfolio. For example, the reform might have incentivized some households to shift their financial wealth to housing wealth in order to bring financial wealth below the taxable threshold. In addition, households might decide to reduce their total wealth in order to bring their financial wealth below the threshold.

The standard approach in empirical tax reform studies to remove this reverse causal link consists of two steps. First, use all available information to predict what the dependent variable would have been had there not been a tax reform. Subsequently, use the new tax system to calculate what the tax rate would have been under the predicted value of the dependent variable (see e.g. Feldstein, 1995, Gruber and Saez, 2002 and Kleven and Schultz, forthcoming).[11]

Since non-capital income and income from housing wealth are taxed synthetically, and non-capital income is orders of maginitude higher than housing income, the tax rate on housing wealth is close to independent of the amount of housing wealth a household owns. As such, there should not be a large endogeneity issue with the after-tax return on housing. However, in the market for financial wealth, the endogeneity problem might be larger. Households can directly affect their tax rate by saving above or below their threshold. Therefore, I estimate a simple savings model on pre-reform data reaching back to 1995 in order to estimate what total wealth would be, had there been no reform. The dependent variable in this equation is household savings as measured by the relative increase in wealth, $\Delta \ln W_{it}$. I subsequently use this model to predict what total wealth would have been without a reform. I then predict financial wealth by assuming that the share of financial wealth in the post-reform year is equal to the share of financial wealth in the base-year 1999, α_{ib}^{f}. The underlying assumption is that the share invested in financial wealth α_{it}^{f} is a stationary variable such that the base year share is a good predictor for what the share would have been in the post-reform year, had there been no reform.

The model used to predict savings is given by:

$$\Delta \ln W_{it} = X_{it}\beta + \gamma_i + \eta_t + \nu_{it}.$$

Independent variables in X_{it} are the lagged value of wealth and savings to control for possible mean reversion in portfolio dynamics. In addition, it contains the log of primary income from labor, the number of children, the number of household members, the type of household and age dummies for the age of the primary earner in the household. Since

[11]See Weber (2013) for a discussion of the validity of this approach.

lagged savings can only be calculated from 1997 onwards the model is estimated in the period 1997-1999. Prediction takes place according to an iterative process. First, savings are predicted for 2000:

$$\Delta \ln \hat{W}_{i,2000} = X_{i,2000}\beta + \gamma_i + \eta_{2000}.$$

Obviously, η_{2000} does not follow directly from the model. However, I use the fact that in any model where η_{2000} would be estimated its value would be the difference between the cross-sectional mean of the dependent variable and the mean of its predicted value. That is,

$$\eta_{2000} = \overline{\Delta \ln W_{i,2000}} - \left(\overline{X_{i,2000}}\beta + \overline{\gamma_i}\right),$$

where a bar over a variable denotes its cross-sectional mean. Subsequently, I estimate wealth holdings in 2000 by using:

$$\ln \hat{W}_{i,2000} = \ln W_{i,1999} + \Delta \ln \hat{W}_{i,2000.} \tag{5.9}$$

I then update X_{i2001} by including predicted values of wealth and savings and again use the model to predict savings and wealth in 2001. The iterative process ends in the final period, 2004. In each post-reform year tax rules are applied to the predicted wealth level in order to estimate what the tax rate would have been if the tax reform had not affected household behavior. Finally, I use predicted instead of actual tax rates in the estimation of equation (5.4) and (5.8).

The second possible channel for reverse causality is the effect a change in wealth may have on the change in portfolio composition. Empirically portfolio composition is correlated with wealth (see e.g. Poterba, 2002, Campbell, 2006 and Attanasio and Weber, 2010). In addition, households may be limited in the amount of wealth they can invest in their house, in particular in the short run when they cannot change their mortgage. As such, a change in portfolio composition might be caused by a change of wealth. The econometric methodology in this paper allows for a straightforward approach to solve this issue. Equation (5.4) can be used to predict portfolio composition. Instruments are the change in the after-tax returns on each asset. Instrument validity requires that the change in the after-tax return is strongly correlated with the change in portfolio-composition, and hence, with the change in the after-tax return on the portfolio. In the results section it is shown that this is indeed the case.

Second, the exclusion restriction requires that the change in the after-tax return can only be related to the change in wealth through the change in the after-tax return. Note that standard tests for the exclusion restriction such as the Sargan-Hansen test (see Sargan, 1958 and Hansen, 1982) cannot be used here because instrumentation is non-linear. That is, the effect of the change in the after-tax return on each asset on the change in portfolio composition is estimated using linear regression. However, the after-tax return additionally has a direct effect on the after-tax return of the portfolio as can be seen in figure 5.7 by the arrow going from the after-tax return on each asset to the after-tax return on the portfolio. Hence, in equation (5.8), if α_{it}^H is instrumented using r_{it}^j, equation (5.8) becomes non-linear in r_{it}^j. However, there is no reason to assume that the tax-rate on each asset has a direct effect on wealth accumulation, other than through its effect on the after-tax return on the portfolio. Hence, I simply assume that the exclusion restriction is satisfied.

Under the assumption that the relative change in after-tax return on each asset is a valid instrument, equation (5.8) can be estimated using portfolio shares predicted in (5.4). This second instrumentation step is not standard in the literature. Therefore, in the sensitivity analysis I also study this equation with actual instead of instrumented shares to see if it has a strong effect on the results.

5.5 Results

5.5.1 First Stage

Table 5.3 represents the first-stage estimates. Regression tables in this and the following section only show the main covariates. Coefficients for the full set of covariates for each table can be found in appendix B. As can be seen, there is strong mean reversion in wealth accumulation. Households with higher levels of wealth save less in the next period. However, additionally, there is some persistence in savings since households with higher lagged growth in their wealth seem to save more in the next period. As expected, households with higher labor income save more. Lagged income does not seem to be correlated with current savings. Results from these first-stage estimates are used to instrument the post-reform tax rate.

5.5.2 Long-Run Effects on Portfolio Composition

Table 5.4 presents the first set of main results. It considers the long-run effects of the reform on portfolio allocation. As can be seen, the change in the after-tax return is positively correlated with the share of financial wealth in the portfolio. Elasticities range from 2.606-4.159 depending on the control variables included in the regression. The first

Table 5.3: First-Stage Results

Variables	Change in Log Wealth
Log Wealth t-1	-1.322***
	(0.00649)
Change in Log Wealth t-1	0.133***
	(0.00415)
Log Prim. Labor Inc.	0.0331***
	(0.00796)
Log prim. Labor Inc. t-1	-0.00644
	(0.00781)
Observations	89113
R-squared	0.604

Note: Dependent variable is change in log wealth. The regression equation is estimated using individual- and year fixed effects. In addition, age and household controls have been included in the estimation. Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.

Table 5.4: Long-Run Effects on Portfolio Composition

	(1) IV	(2) IV	(3) IV	(4) IV	(5) IV	(6) IV
Rel. Ch. R^F	2.715***	2.606***	3.645***	2.696***	3.774***	4.159***
	(0.856)	(0.884)	(0.751)	(0.883)	(0.749)	(0.741)
Rel. Ch. R^H	-0.00622***	-0.0113***	-0.0174**	-0.0112***	-0.0174**	-0.0178**
	(0.00210)	(0.00245)	(0.00717)	(0.00242)	(0.00717)	(0.00743)
Log Savings 1999	-0.0944***	-0.0845***		-0.0830***		
	(0.0311)	(0.0314)		(0.0313)		
Log Wealth 1999	-0.0329	0.134***		0.139***		
	(0.0525)	(0.0185)		(0.0185)		
Control for:						
Splines	YES	NO	NO	NO	NO	NO
Prim. Labor Income	YES	YES	YES	NO	NO	NO
Hh/Age Effects	YES	YES	YES	YES	YES	NO
Observations	12261	12261	13885	12261	13885	13885
R-squared	0.036	0.022	0.018	0.021	0.017	0.006

Note: Dependent variable is relative change in the share of financial wealth between 1999-2004. IV-estimates using instrumented tax rates. Splines are linear decile spline terms over the wealth distribution. Primary labor income is a term containing the log of the sum of primary labor income earned between 1999-2004. Household and age effects include household type dummies, number of children below 18 in the household, number of members in the household and age dummies for the primary income earner. Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.

column includes all control variables and it presents the preferred estimate of the elasticity at 2.715. This indicates that a one percent increase in the after-tax return increases the share of financial wealth in total wealth by 2.715 percent.

At first sight this effect seems rather large. However, recall that by the approximation $ln(1 + x) \approx x$, a one percentage point increase in the wealth tax on financial assets decreases the gross return on financial assets by one percent. Since the current wealth tax rate is 1.2 percent, a one-percentage point increase in this tax rate corresponds to an $1/1.2 = 83$ percentage increase in the tax rate. Therefore, these results imply that a one percent increase in the wealth tax on financial assets decreases the share of financial assets in the portfolio by $2.715/83 = 0.033$ percentage. Hence, the effect of taxation on asset allocation is relatively modest. A close to doubling of the wealth tax decreases financial assets by only 0.033 percent.

This finding corresponds with recent findings in Alan *et al.* (2010) but are in stark contrast to earlier findings using cross-sectional data in e.g. Feldstein (1976), King and

Leape (1998) and Poterba and Samwick (2003) who find strong effects of taxation on portfolio allocation. To my knowledge this study is the first to quantify the exact response of portfolio-allocation to a tax-induced change in the gross after-tax return on the asset.

The second row measures the effect of an increase in the return on housing wealth. As can be seen, an increase in the return on housing wealth slightly decreases the share invested in financial wealth, although the effect is much smaller. This asymmetric response could be explained by the fact that it is more costly for households to adjust their housing wealth. Households can increase their housing wealth by paying off their mortgage or by buying a new house. The former may be costly because households mortgage contracts usually fine households when they pay of more than the contracted amount. The latter is very costly due to moving cost and a 6 percent stamp duty that the government charges upon real estate transaction of owner-occupied houses.

Combining estimates from the first and second row of table 5.4 creates an interesting picture of household portfolio behavior. An increase in the return on financial wealth induces households to buy more financial wealth. Housing wealth remains remains unaffected due to high transaction costs, but since total wealth goes up, housing wealth as a percentage of total wealth goes down. On the other hand, an increase in the return on housing wealth does not induce households to buy more economically significantly more housing wealth, due to the transaction costs involved in buying housing wealth. Hence, there seems to be very little response to a change in the return on housing wealth.

With respect to the control variables, the change in the share invested in financial assets is decreasing in base-year savings, indicating that those individuals who have saved a lot in 1999 are less likely to save in financial assets in future periods. Wealth is uncorrelated to the change in the share invested in financial assets except when I do not control for wealth splines.

The size of both elasticities is sensitive with respect to the control variables used. In particular, the elasticities get larger when I do not control for household and age effects and when I do not control for portfolio dynamics by including base-year savings and wealth. In addition, the elasticity of the housing return is rather sensitive with respect to the use of splines. This indicates a slight positive correlation between the (instrumented) change in the after-tax return on the wealth types and each of these variables. However, the elasticities remain significant and of expected sign in each specification. In addition as has been argued above, even when the elasticity is at its highest, the effect of taxation on portfolio composition is rather modest.

5.5.3 Long-Run Effects on Wealth Accumulation

Table 5.5: Long-Run Effects on Wealth Accumulation

	(1) IV-IV	(2) IV-IV	(3) IV-IV	(4) IV-IV	(5) IV-IV	(6) IV-IV
Rel. Ch. R^W	3.031***	2.775***	4.256***	2.808***	4.257***	4.318***
	(0.431)	(0.353)	(0.429)	(0.355)	(0.429)	(0.432)
Savings 1999	-0.0959***	-0.0957***		-0.0941***		
	(0.0121)	(0.0120)		(0.0121)		
Wealth 1999	-0.210***	-0.263***		-0.257***		
	(0.0273)	(0.00977)		(0.00971)		
Control for:						
Splines	YES	NO	NO	NO	NO	NO
Prim. Labor Income	YES	YES	YES	NO	NO	NO
Hh/Age Effects	YES	YES	YES	YES	YES	NO
Observations	12261	12261	12261	12261	12261	12261
R-squared	0.356	0.343	0.230	0.340	0.230	0.218

Note: Dependent variable is relative change in wealth between 1999-2004. IV-estimates using instrumented tax rates and portfolio shares. Splines are linear decile spline terms over the wealth distribution. Primary labor income is a term containing the log of the sum of primary labor income earned between 1999-2004. Household and age effects include household type dummies, number of children below 18 in the household, number of members in the household and age dummies for the primary income earner. Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.

In table 5.5 I study the first stage of the household optimization problem, the trade-off between consumptions and savings. Results from the first column of table 5.4 were used to instrument for portfolio composition. My estimates show that a 1 percent increase in the after-tax return on the portfolio, reduces savings by somewhere between 2.775-4.318 percent. The preferred estimates are again shown in the first column where the elasticity equals 3.031. Under all specifications the elasticity is statistically significant at the one- percent level and positive. This indicates that the substitution effect dominates the income effect. This result indicates that the elasticity of intertemporal substitution is positive and this is in correspondence with all recent findings (see e.g. Attanasio and Weber, 1993, Blundell *et al.*, 1994, Attanasio and Browning, 1995 and Engelhardt and Kumar, 2009).

By the same argument as in the previous subsection, the estimated elasticity indicates that a one percent increase in a hypothetical wealth tax over all wealth of 1.2 percent, decreases savings by 0.036 percent. Hence, also here, the distortinary effect of taxation on wealth accumulation is relatively modest.

Controls for portfolio dynamics show a strong indication for mean reversion. Wealth accumulated during the reform period is decreasing in base-year wealth and in base-year savings.

Results are sensitive to the use of control variables. In particular, the elasticity becomes somewhat larger when I do not control for portfolio dynamics and become smaller when I do not control for wealth splines.

5.5.4 Short-Run Effects on Portfolio Composition

Table 5.6: Short-Run Effects on Portfolio Composition

	(1) IV	(2) IV	(3) IV	(4) IV	(5) IV	(6) IV
Rel. Ch. R^F	2.269***	3.368***	3.288***	2.235***	3.368***	3.882***
	(0.749)	(0.648)	(0.649)	(0.760)	(0.648)	(0.641)
Rel. Ch. R^H	-0.00366	-0.0146*	-0.0146*	-0.00837***	-0.0146*	-0.0147*
	(0.00241)	(0.00747)	(0.00746)	(0.00325)	(0.00747)	(0.00789)
Savings 1999	-0.0658**			-0.0521*		
	(0.0271)			(0.0273)		
Wealth 1999	-0.0146			0.121***		
	(0.0417)			(0.0157)		
Control for:						
Splines	YES	NO	NO	NO	NO	NO
Prim. Labor Income	YES	YES	YES	NO	NO	NO
Hh/Age Effects	YES	YES	YES	YES	YES	NO
Observations	15487	17570	17570	15487	17570	17570
R-squared	0.031	0.017	0.017	0.020	0.017	0.005

Note: Dependent variable is relative change in the share of financial wealth between 1999-2001. IV-estimates using instrumented tax rates. Splines are linear decile spline terms over the wealth distribution. Primary labor income is a term containing the log of the sum of primary labor income earned between 1999-2001. Household and age effects include household type dummies, number of children below 18 in the household, number of members in the household and age dummies for the primary income earner. Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.

Next, I study the short-run effect of the reform on portfolio composition in table 5.6 by taking the difference between the base year 1999 and the first post-reform year 2001. It is interesting to compare the short-run effects to the long-run effects since this may be indicative of how fast households can adjust their portfolio. In the preferred estimate, reported in the first column, the elasticity of the share invested in financial wealth with respect to the return on financial wealth equals 2.269. Surprisingly this elasticity is almost as high as the long-run elasticity, indicating that households reacted to the reform almost instantaneously.

The fast response may be explained by optimization frictions. The 2001 reform brought such a radical change in portfolio returns that households had to respond lest they would end up with highly inefficient portfolios. As a result, they responded right away. After the reform the returns did not fluctuate very much, and as a result, the households did not make anymore adjustments.

On the other hand, the elasticity with respect to the return on housing wealth is not significant. This again indicates that households react less strongly to changes in the return on housing wealth then to the return on financial assets.

Results are sensitive to the use of control variables where in some specifications the short-run elasticity exceeds the long-run elasticity. This could be seen as evidence for misspecification in those models since it seems unlikely that households overshoot their optimal portfolio allocation in the short run, especially if there are transaction costs.

Finally, note that there are slightly more observations in the short-run estimates than in the long-run estimates. This attrition is due to mortality and emigration, marriage and divorce, or missing variables somewhere in the period 2002-2004.

5.5.5 Short-Run Effects on Wealth Accumulation

In table 5.7 I present the short-run effects of the tax reform on wealth accumulation. One would expect that an increase in the after-tax return induces households to save more (or less if the income effect dominates) each period. Surprisingly, the preferred estimate the short-run elasticity of 2.739 is only slightly smaller than the long-run estimate. This indicates that households adjust their accumulated wealth immediately to the new after-tax return, but only slightly change their yearly savings in the periods afterward. This might again be the result of optimization frictions.

As mentioned before, it is not possible to directly compare estimates in this paper to estimates of the elasticity of intertemporal substitution. However, Attanasio and Wakefield (2010) perform a simulation for the UK where the after-tax net return is increased from 2 to 2.5 percent, using a life-cylce model. In their baseline simulation the elasticity of substitution equals 1. The simulations show that accumulated wealth may increase by as much as 18 percent at retirement age and by a very significant amount at all ages. On

Table 5.7: Short-Run Effects on Wealth Accumulation

	(1) IV-IV	(2) IV-IV	(3) IV-IV	(4) IV-IV	(5) IV-IV	(6) IV-IV
Rel. Ch. R^W	2.739***	3.491***	3.491***	2.365***	3.491***	3.548***
	(0.407)	(0.366)	(0.366)	(0.305)	(0.366)	(0.368)
Savings 1999	-0.0917***			-0.0895***		
	(0.00996)			(0.00989)		
Wealth 1999	-0.187***			-0.239***		
	(0.0194)			(0.00854)		
Control for:						
Splines	YES	NO	NO	NO	NO	NO
Prim. Labor Income	YES	YES	YES	NO	NO	NO
Hh/Age Effects	YES	YES	YES	YES	YES	NO
Observations	15487	15487	15487	15487	15487	15487
R-squared	0.346	0.222	0.222	0.326	0.222	0.211

Note: Dependent variable is relative change in wealth between 1999-2001. IV-estimates using instrumented tax rates and portfolio shares. Splines are linear decile spline terms over the wealth distribution. Primary labor income is a term containing the log of the sum of primary labor income earned between 1999-2001. Household and age effects include household type dummies, number of children below 18 in the household, number of members in the household and age dummies for the primary income earner. Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.

the other hand, estimates in this paper suggest that an increase in the net after-tax return of 0.5 percentage point increases average wealth by approximately $0.5 \times 3.031 = 1.5155$ percent in the long-run. This is far less than the simulation would indicate. Obviously, these two results are not directly comparable since Attanasio and Wakefield (2010) study the steady-state effects in a life-cycle model, where attaining the steady state takes the life-time of an entire generation. However, the small effect that I find in this study coupled with the fact that short-run estimates are very close to long-run estimates gives some tentative evidence that the elasticity of substitution in the Netherlands is smaller than the baseline value in Attanasio and Wakefield (2010).

5.6 Sensitivity Analysis

5.6.1 Portfolio Composition

Table 5.8 shows a sensitivity analysis of the effects of the reform on portfolio allocation. The first column considers only those households where the age of the primary earner is below 65 in 2004. As can be seen, the elasticities are virtually identical to the elasticities in the initial sample.

The second column considers only those households that were above the wealth tax threshold in the base year 1999. This sensitivity analysis serves two purposes. First, households above the wealth tax threshold are relatively wealthy. Therefore, the analysis may help uncover potential heterogeneity in the behavioral response between wealthy and less wealthy households. Such heterogeneity in responses has been found in the literature in e.g. Alan *et al.* (2010) and might exist because wealthy households may be different in unobservable characteristics such as transaction costs. Moreover, wealthy households might be less liquidity constrained and therefore better able to optimally adjust their portfolio. Second, the measurement error may be less severe for wealthy households. If a low-wealth household misreports its wealth there is no sanction for it, as long as wealth is below the threshold. However, if households above the threshold under report their wealth the tax authorities may sanction them severely.

As can be seen, the elasticity with respect to the return on financial wealth for this group is almost twice as large. In addition, the elasticity with respect to the return on housing wealth increases with a factor larger than 100. This result gives some evidence for heterogeneity in the response rate, although the implied heterogeneity needs to be interpreted with some caution, because the lower estimate for the initial sample may have also been driven by measurement error.

Table 5.8: Long-Run Effects on Portfolio Composition for Different Specifications and Subsamples

	(1) < 65	(2) Wealth Tax	(3) Singles	(4) Incl. outliers
Rel. Ch. R^F	2.786***	3.941***	6.438**	0.255***
	(0.865)	(1.102)	(3.168)	(0.0582)
Rel. Ch. R^H	-0.00621***	-7.343***	-2.843**	-0.00613***
	(0.00209)	(1.730)	(1.251)	(0.00157)
Log Savings 1999	-0.0964***	-0.187***	0.135	-0.0823**
	(0.0316)	(0.0452)	(0.116)	(0.0321)
Log Wealth 1999	-0.0304	-0.159***	0.0221	-0.0195
	(0.0584)	(0.0488)	(0.161)	(0.0538)
Control for:				
Splines	YES	YES	YES	YES
Prim. Labor Income	YES	YES	YES	YES
Hh/Age Effects	YES	YES	YES	YES
Observations	11903	8625	664	12510
R-squared	0.035	0.061	0.157	0.038

Note: Dependent variable is relative change in the share of financial wealth between 1999-2004. IV-estimates using instrumented tax rates. Splines are linear decile spline terms over the wealth distribution. Primary labor income is a term containing the log of the sum of primary labor income earned between 1999-2004. Household and age effects include household type dummies, number of children below 18 in the household, number of members in the household and age dummies for the primary income earner. Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.

The third group shows the result of single households without children. Their response with respect to the return on financial wealth is more than two times larger than it is for the initial sample. In addition, their response with respect to the return on housing wealth is more than 50 times larger. This again gives some indication of heterogeneity in the behavioral response where singles react much stronger to the change in the capital tax than other households. However, the results have to interpreted with some caution, since the number of observations is relatively small.

The final robustness analysis includes outliers that were filtered out for all other estimates. As can be seen, the number of observations increases with only 240 households. However, the response with respect to the after-tax return on financial wealth decreases by a factor ten. This indicates that these outliers have a very strong effect on the results.

5.6.2 Wealth Accumulation

Table 5.9: Long-Run Effects on Wealth Accumulation for Different Specifications and Subsamples

	(1) < 65	(2) Wealth Tax	(3) Singles	(4) Incl. outliers	(5) IV-OLS
Rel. Ch. R^W	3.008***	9.915***	7.822***	3.031***	1.091***
	(0.430)	(0.471)	(1.012)	(0.431)	(0.257)
Savings 1999	-0.0967***	-0.136***	-0.0676	-0.0959***	-0.115***
	(0.0122)	(0.0204)	(0.0417)	(0.0121)	(0.0122)
Wealth 1999	-0.215***	-0.191***	-0.225***	-0.210***	-0.418***
	(0.0306)	(0.0304)	(0.0812)	(0.0273)	(0.0302)
Control for:					
Splines	YES	YES	YES	YES	YES
Prim. Labor Income	YES	YES	YES	YES	YES
Hh/Age Effects	YES	YES	YES	YES	YES
Observations	11903	8625	664	12261	12831
R-squared	0.356	0.250	0.558	0.356	0.613

Note: Dependent variable is relative change in wealth between 1999-2004. IV-estimates using instrumented tax rates and portfolio shares except in final column which only uses instrumented tax rates. Splines are linear decile spline terms over the wealth distribution. Primary labor income is a term containing the log of the sum of primary labor income earned between 1999-2004. Household and age effects include household type dummies, number of children below 18 in the household, number of members in the household and age dummies for the primary income earner. Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.

Table 5.9 presents the sensitivity analysis of the effects of the reform on wealth accumulation. The first column presents the results on the subsample that was younger than 65 in 2004. As can be seen, the elasticity is not significantly different for this subsample.

The second column shows the elasticity for the group that was subject to the wealth tax. As can be seen, the elasticity of this group is about three times larger than for the entire sample, indicating that the wealthy may perhaps react stronger to the change in the after-tax return than the poor. A larger response of the the wealthy may be indicative of the fact that liquidity constraints are binding for the households with lower wealth. If liquidity constraints are binding an increase in the return on the portfolio may relax the liquidity constraint, inducing poor households to consume more and save less. Liquidity constraints play a similar role if the household is not currently at the liquidity constraint but may, due to uncertainty, end up at the liquidity constraint in some state of the world (see e.g. Attanasio and Weber, 2010). Clearly, the poor are more likely to be liquidity constrained than the rich and as such, they are less likely to accumulate more wealth if the after-tax return on their assets goes up. The resulting outcome would be that wealthy households have a higher elasticity which is exactly what I find.

The third column displays the results for singles. Singles also have a higher elasticity although the number of observations is rather limited. The fourth column shows the result when outliers are included. This does not seem to effect the elasticity at all. The final column shows the elasticity when I do not instrument for the change in asset composition. The elasticity in this specification is still significantly positive but almost three times smaller. This may indicate that it is indeed necessary to instrument for reverse causality running from wealth to portfolio composition.

5.7 Conclusion

In this paper I use the Dutch 2001 capital tax reform to estimate the effect of capital taxation on households' portfolio composition and intertemporal choice. To my knowledge this is the first study to directly link a tax-induced change on the after-tax return on assets to the portfolio and savings decisions of household. I find behavioral responses in the direction predicted by theory. However, in contrast to earlier findings in the literature, the estimated effect is modest. Therefore, the distortion caused by capital-income taxation is smaller than previously considered.

This finding is of direct impact to policy makers and researchers. A lower distortion of the capital-income tax on wealth accumulation implies a higher optimal capital tax rate. In addition, portfolio choice is not strongly affected by relative difference in the capital income tax rate on different assets. This indicates that nudges may perhaps be a more effective way to affect household's behavior (see e.g. Madrian and Shea, 2001).

In this study I investigate portfolio responses on the intensive margin. Future research should investigate whether Dutch households respond to tax incentives on the extensive margin, and what is the effect of pension savings on portfolio choice and wealth accumulation.

Chapter 6

As Easy as ABC? Multi-dimensional Screening in Public Finance[1]

6.1 Introduction

In two seminal articles Mirrlees (1971, 1976) characterizes the welfare-maximizing allocation in a model where individuals are heterogeneous in a single unobserved characteristic and make multiple decisions. This approach to the study of taxation has yielded many fruitful results since. In particular, Diamond (1998) and Saez (2001) rewrote the solution of the model into an multiplicative ABC-formula that describes the optimal wedge between the marginal rates of transformation and the marginal rates of substitution (further: wedge) as a function of measurable elasticities and the distribution of income. The ABC-formula allows for an intuitive explanation of the optimal wedge in the second-best, and serves as a convenient way to approach the data.

Up to now, the literature has almost exclusively focused on models where agents differ in only one dimension, namely earnings ability. So far, the technical complexity involved in introducing multi-dimensional heterogeneity has made the problem intractable. As a result, policy makers receive little guidance if they do not just want to redistribute from rich to poor, but also from, for instance, healthy to sick. In addition, the literature is

[1]This chapter is based on Renes and Zoutman (2013a). We would like to thank Felix Bierbrauer, Eva Gavrilova, Aart Gerritsen, Yasushi Iwamoto, Bas Jacobs, Laurence Jacquet, Etienne Lehmann, Dominik Sachs, Dirk Schindler, Bauke Visser, Casper de Vries and Hendrik Vrijburg for useful suggestions and comments on an earlier version of this paper. Furthermore, this paper benefited from comments and suggestions made by participants at the 2011 Nake Conference, Utrecht, the 2013 CESifo Area Conference on Public Economics, Munich, the 69th IIPF Conference, Taormina; ; seminar participants at the Erasmus School of Economics, the Norwegian University of Science and Technology, the Norwegian School of Economics, the University of Konstanz and the Centre for European Economic Research. All remaining errors are our own.

silent on how optimal policy should be adjusted when agents differ in preferences, for instance, the labor supply elasticity or the discount rate. Basic policy questions, such as, how should a government combine taxes on labor income with healthcare subsidies, what is the relation between capital and labor income taxes, and should housing subsidies depend on wealth or income, cannot be answered unless the extreme assumption is taken that the difference between agents can be expressed in a single parameter.

In this paper we make a first step towards answering these questions by using insights from the multi-dimensional screening literature (most notably McAfee and McMillan, 1988, Armstrong, 1996, and Rochet and Choné, 1998) to extend the Mirrleesian optimal tax model to a setting where individuals are heterogeneous in multiple unobservable characteristics.

In our model individuals differ in $p \geq 1$ hidden characteristics such as ability, health-status or patience. They choose k observable continuous choice variables such as, labor income, consumption of health care products and savings. Additionally, they choose how much to consume of a numeraire good, such as a general consumption good. We will often refer to the choice variables as goods, although they can be either inputs or outputs in the production process. In order to apply the revelation principle derived in Myerson (1979) we need to take two assumptions. First, we assume $k \geq p$, the number of choice variables is larger than the number of hidden characteristics. Second, we assume preferences allow the revelation of each characteristic given the proper mechanism.[2] The application of the revelation principle allows us to relate optimal policy to observable choice variables. Therefore, these assumptions are crucial to our analysis.

If all hidden characteristics of all agents were known to the central planner, the Second Welfare Theorem would imply the planner could select an efficient allocation and let the market reach it, only interfering through individualized lump-sum taxes. Since the planner cannot observe the type of each individual directly, he has to create the incentives for each individual to reveal his type. The central planner thus faces a multi-dimensional screening problem. The tax system is the tool used by the planner to gain information about the type of each individual so that he can redistribute from one type to another. Therefore, by reinterpreting the optimal tax problem as a screening or information problem we can combine insights from the screening literature with optimal taxation.

We compare our characterization of the second-best allocation to the second-best in the uni-dimensional Mirrleesian optimal tax model by establishing whether well-known

[2]In particular, we exclude characteristics that do not influence the preference of any choice variable, and multiple characteristics that influence the preference of only one choice variable. Such characteristics are fundamentally non-revealable in any mechanism.

results in the uni-dimensional setting hold in our more general model. First, we show the optimal wedge can be characterized by a generalized version of Diamond's (1998) and Saez' (2001) ABC-formula with an additive structure over the characteristics. The optimal wedge on each good increases in A, the quality of the signal obtained from observing the good: how much does a specific good reveal about an underlying characteristic. In the standard case with uni-dimensional heterogeneity in earnings ability it is well-known that the optimal tax rate decreases in the labor supply elasticity. We can easily explain this using our interpretation. As the labor supply elasticity increases the information gained from observing labor earnings decreases. It follows from our model that the tax on labor income decreases in the elasticity.

In addition, we find that the optimal wedge increases in B, the redistributive benefit of marginally distorting the price of the good, and decreases in C, the size of the tax base for which the good is distorted. The two latter properties have already been established in the uni-dimensional model.

A corollary to this result shows that an optimal wedge on a good is zero if a good does not reveal information about any of the hidden characteristic. This corollary implies the Atkinson-Stiglitz (A-S) theorem in case of uni-dimensional heterogeneity and weak separability of labor in the utility function. In our interpretation, if disutility of labor is the only aspect of utility that is not separable from type, the labor choice is the best signal of type. Indirect taxation yields no extra information and is thus not optimal.

The corollary also immediately implies the A-S theorem does not extend to multi-dimensional heterogeneity. With at least two types of heterogeneity, a single good can never extract all available information. If the planner wants to separate healthy and sick agents, and high-ability and low-ability agents, it will have to distort labor income, as well as a good that reveals the health status of the agent, such as consumption of health care products.

Mirrlees (1976) shows that if agents are heterogeneous in only one dimension the optimal wedge on each good can be written as a function of only that good. In case of multi-dimensional heterogeneity such separable wedges are impossible. We show that in general, in order to facilitate full revelation of the p underlying characteristics, the marginal wedge on each choice is a function of p choice variables[3]. Such interdependencies are very common in the stochastic dynamic models of the New Dynamic Public Finance (NDPF)[4]. This model extends the Mirrleesian framework to a setting where the type of the

[3] This result describes the general case. Special cases may exist where the wedge can be written in a simpler form.

[4] See Golosov *et al.* (2007) for an overview

agents evolves stochastically over time. In this model, the tax on labor income in period t may depend on all prior earned labor income in periods prior to t (see e.g. Kocherlakota, 2005). The history of play forms a natural extension to the type space, since it contains information about the preferences of the agents. We show that this result can be replicated in a deterministic Mirrleesian public finance model, provided earnings ability takes on a different value in each period. This suggests that the intertemporal interdependencies in the optimal tax-schedule in NDPF models stem from the multi-dimensionality of the type space, rather than from the stochastic process.

We also derive a generalization of the no-distortion at the top result (see e.g. Sadka, 1976 and Seade, 1977). As in the uni-dimensional case the optimal wedge at the extreme points of of the type-distribution are zero. If a type exists that has extreme values of all characteristics, his optimal marginal wedge on all choices equals zero.[5] Intuitively, since there are no "extremer" types, setting a wedge to separate out more extreme types yields no information to the planner. Hence, for any marginal distortion the efficiency cost of the distortion is higher than the welfare gain at the extreme-points of the distribution. Note though, that unlike in the uni-dimensional case, in the multi-dimensional case these types do not necessarily exist. For instance, the healthiest person in the economy may not be the richest person in the economy. In that case, the healthiest person may face a positive wedge on his labor income, whereas the richest person may face a positive wedge on his consumption of healthcare products. In the remainder of this paper we will refer to types located at extremes of the type-distribution as corner-types.

We overcome the technical complexities of deriving the second-best allocation under multi-dimensional heterogeneity by using a first-order approach. That is, we derive the optimal allocation in a relaxed problem that takes the first-order incentive constraints into account, while assuming the second-order incentive constraints are met in optimum.[6] This approach has become the standard in the optimal taxation literature with uni-dimensional heterogeneity. It is well-known that solutions obtained by the first-order approach consistently violate second-order incentive constraints at the bottom of the type space in screening models where both the incentive and the participation constraint is binding (see e.g. Armstrong, 1996, Rochet and Choné, 1998). Intuitively, if a principal tries to extract all of the rents of private information out of the bottom types, they will simply stop participating. Therefore in the second-best allocation, these types are bunched together. However, models of optimal taxation typically do not feature binding partici-

[5] This result is reminiscent of the theorem derived in Golosov *et al.* (2011) where the wedge at the bottom and top is zero if the stochastic process allows agents to be located at the extremes.

[6] An introduction to this technique can be found in Wilson (1996).

pation constraints because it is assumed to be too costly to leave the jurisdiction. Hence, there is no inherent conflict between participation constraints and incentive constraints. Although we cannot formally prove that bunching never occurs in our model we show in section 7 that if separation of types and bunching occur simultaneously, separation will occur in a single convex subset extending from "the top" of the type space. In this separating set, our solution obtained through the first-order approach still describes the second-best. Hence, even if the optimal allocation exhibits bunching of types at the bottom of the type space, our solution remains valid in the upper-interior part of the type space where full separation of types is optimal.

In this paper we focus on finding the characteristics of a second-best allocation in a direct mechanism without discussing the tax-mechanism used to implement this allocation in the market. As we show in the companion paper Renes and Zoutman (2013b) the design of the (tax-)mechanism that implements the second-best allocation in the market can be very complicated. However, as proposition 1 of that paper shows, if the government is welfarist and there are no externalities, the government can implement the second-best allocation by equating the marginal tax rate to the optimal wedge. In this paper we assume the objective of the planner is welfarist and there are no externalities. Therefore, the optimal wedges derived in this paper, also describe all the relevant characteristics of the optimal tax system.

The rest of this paper is organized as follows. The next section discusses related literature. Section three introduces the model. The fourth section derives the optimal allocation using the first-order approach. Section five discusses the ABC-formula. Section six compares our results to results obtained in the NDPF. Section seven discusses the validity of the first-order approach and the final section concludes.

6.2 Related Literature

In this paper we rely on the first-order approach to elicit the properties of the second-best allocation. Another approach in the literature on multi-dimensional screening is to discretize the type space (see e.g. Armstrong and Rochet, 1999 for a user's guide). In a model with discretely distributed types it is possible to (numerically) verify which incentive constraints are binding such that the optimal allocation can be derived without relying on the first-order approach. Cremer *et al.* (2001) apply this technique in an optimal tax model where agents are heterogeneous in earnings ability and wealth endowments and choose labor hours and savings. They show that the Atkinson-Stiglitz theorem fails to hold in this setting since the government optimally taxes savings. The downside of discretizing the distribution is that the optimal wedge can only be verified on a discrete number of points. Moreover, as the number of discretized types increases, the problem becomes less tractable. Because in our model types are continuously distributed it is possible to calculate the wedge for all levels of the choice variables, thereby deriving the entire shape of the optimal tax system.

In a setting with continuously distributed types Saez (2002a), derives the optimal income and commodity tax in a model where agents are heterogeneous in both earnings ability and preferences. He shows the Atkinson-Stiglitz theorem fails when preferences for a particular commodities are correlated with earnings ability, or the preference for leisure, since the government should optimally tax these commodities at a higher rate. Unfortunately, two strong assumptions make it difficult to use his approach to calculate the entire tax system. First, he assumes that welfare weights are correlated to ability, but uncorrelated to the other hidden characteristics. However, governments are also likely to give higher welfare weights to agents with lower health status and lower wealth endowments. Second, in his model all goods except labor income are taxed linearly. However, modern governments have access to a wider range of non-linear instruments such as the tax on capital income, health care subsidies and education subsidies. Our approach poses no such restrictions and can be used to calculate all optimal non-linear wedges.

Kleven *et al.* (2009) study the taxation of couples in a setting where both partners have different earnings ability. To maintain analytic tractability they assume the primary earner chooses labor supply on the intensive margin while the least-earning partner chooses on the extensive margin. In our model agents only make intensive-margin choices. We argue that many economic decisions such as savings and consumption choices are more accurately portrayed as choices on the intensive margin. We believe the best solution would be to combine both approaches by extending our model with extensive-margin decisions, as was done with uni-dimensional heterogeneity in Jacquet *et al.* (2010). However, we leave this for future research.

Lewis and Sappington (1988) and Pass (2012) study multi-dimensional screening in a setting where the number of goods is smaller than the number of characteristics, $k < p$. In such a setting a direct application of the revelation principle is not possible. However, it is shown that the revelation principle may be applied after ingeniously reducing the dimension of the type space to the dimension of the choice-space. This method has been successfully applied in Choné and Laroque (2010) in an optimal-tax model where agents choose labor supply and are heterogeneous in both opportunity cost of work and ability.

They show the income tax rate may be negative at the bottom of the income distribution if heterogeneity in the opportunity cost of work is relatively important. To limit the complexity of our problem we restrict our attention to the case where $k \geq p$ such that we can apply the revelation principle directly. However, it may be possible to extend our results to the case where $k < p$ by applying the method derived in Pass (2012).

6.3 The Model

6.3.1 Preferences

Before we start our analysis we formally introduce the preferences of the agents and the planner, and we set up the economies resource constraint.[7] The set-up of our model closely follows section 4 of Mirrlees (1976). The economy is populated by a unit mass of individuals that are characterized by a twice-differentiable utility-function:

$$u(\mathbf{x}, y, \mathbf{n}),$$

where $\mathbf{x} \in \mathbf{X} \subseteq \mathcal{R}^k$ denotes a vector of choice variables such as, effective labor supply, consumption of health care products and savings. $y \in Y \subseteq \mathcal{R}$ is the untaxed or numeraire commodity. In principle, the choice of the numeraire variable has no effect on the optimal allocation. However, we assume y is a normal good, such that $u_y > 0, u_{yy} \leq 0$ for any value of $(\mathbf{x}, y, \mathbf{n})$. Therefore, the utility function is non-satiated everywhere. These assumptions ease the interpretation of the optimal allocation in the remainder of this paper. Decision variables $\mathbf{x}$ and y are observable at the individual level, and the social planner can tax all choices in $\mathbf{x}$ non-linearly, but cannot tax y. Throughout the paper we will sometimes refer to the choice variables in $\{\mathbf{x}, y\}$ as goods, even though they can be both inputs and outputs to the production process.

$\mathbf{n} \in \mathbf{N} \subseteq \mathcal{R}^k$ denotes the type of an individual. Each element in type $\mathbf{n}$ is referred to as a characteristic. Characteristics in $\mathbf{n}$ may include earnings ability, health status and preference parameters. For technical convenience we assume $\mathbf{N}$ is a convex space. The space $\mathbf{N}$ will be referred to as the type space.

The distribution of $\mathbf{n}$ is given by the cumulative density function $F(\mathbf{n})$, with $F : \mathbf{N} \rightarrow [0, 1]$, and probability density function $f(\mathbf{n})$. Both are defined over the closure of $\mathbf{N}$. We assume each characteristic denotes some independent aspect of the individuals, such

[7] Note that the description of agent's preferences closely follows the description in Renes and Zoutman (2013b).

that no characteristic can be found as a deterministic function of the other characteristics. The type is private information to each individual and unobservable to the government.

Note that we do not restrict ourselves to static models: different choices can occur in different periods. However, we do assume that both the type and the direct mechanism are revealed to the individuals before they solve their maximization problem.[8][9].

We need two assumptions that allow us to apply the revelation principle in the subsequent analysis (see Myerson, 1979). First, we assume that $k \geq p \geq 1$, such that there are at least as many decision variables in $\mathbf{x}$ as characteristics in $\mathbf{n}$. Therefore, the choice space is large enough to contain all information of the type space. Second, let:

$$\mathbf{s}(\mathbf{x},y,\mathbf{n}) \equiv -\frac{u_{\mathbf{x}}(\mathbf{x},y,\mathbf{n})}{u_y(\mathbf{x},y,\mathbf{n})},$$

denote the vector of shadow prices, such that each element $\mathbf{s}_i$ denotes the marginal rate of substitution for decision variable x_i with respect to the numeraire y. We assume the Jacobian $\mathbf{s_n}$ is of full rank, p, for any combination $\{\mathbf{x}, y, \mathbf{n}\}$. This assumption excludes the possibility of having two characteristics that affect the preference of only one choice. The most famous example in the literature is the case where individuals differ in their degree of earnings ability and in their opportunity cost of work, and only choose effective labor supply. The utility cost of providing a unit of effective labor supply is decreasing in ability and increasing in the opportunity cost of work. If both characteristics act only on effective labor supply, it is fundamentally impossible to reveal them in the choice space. By assuming $\mathbf{s_n}$ is of full rank, we assume that there is always a second observable choice, which can be used to disentangle the effect of ability and the opportunity cost of work. A particular example in this case might be the amount spend on video games. Suppose the preference for video games increases in the opportunity cost of work. In that case, the planner can deduce both characteristics by observing both labor earnings and the amount spend on video games.

The social planner is assumed to maximize a concave sum of the individual's utility:

$$SW = \int_{\mathbf{N}} W(u(\mathbf{x},y,\mathbf{n}))\, dF(\mathbf{n}), \tag{6.1}$$

$$W' > 0, W'' \leq 0, \tag{6.2}$$

[8]The model with uni-dimensional heterogeneity has often been used to describe a dynamic economy. See Golosov *et al.* (2013) for a recent example.

[9]Note also that the conventional utility representation(See e.g. Mirrlees, 1971, Saez, 2001.) $\tilde{u}(y,l)$ with l denoting labor supply is a special case of our utility representation. If one takes the standard assumption that gross income equals $x_1 = n_1 l$ where n_1 is earnings ability, it can be seen that this utility function can be rewritten into our form: $\tilde{u}(y,l) = \tilde{u}\left(y, \frac{x_1}{n_1}\right) = u(x_1, y, n_1)$

where $W(\cdot)$ is a Bergson-Samuelson welfare function.[10] We assume the social planner commits to the allocation he offers such that he cannot alter the allocation after types are revealed.[11] Further, we assume that redistribution is welfare increasing because of (at least) one of two reasons. First, concavity in the utility functions of the individuals would imply that individuals with higher income have a lower marginal utility of income. Second, $W'' < 0$ would imply the social planner gives a higher welfare weight to individuals with lower utility. The social planner is bound by the economy's resource constraint:

$$\int_{\mathbf{N}} y(\mathbf{n})\, dF(\mathbf{n}) + R \leq \int_{\mathbf{N}} q(\mathbf{x}(\mathbf{n}))\, dF(\mathbf{n}), \tag{6.3}$$

where R denotes exogenous government expenditure and $q(\cdot)$ is the economy's production of y as a function of the decision variables in $\mathbf{x}$. A partial derivative q_{x_i} may be either positive or negative depending on whether choice variable x_i is an input, or an output variable of the production process. We assume the production technology exhibits diminishing marginal returns such that $q_{x_i x_i} \leq 0$ for all goods x_i to guarantee that an interior solution will be reached in laissez-faire.

For bookkeeping, the Jacobian of first-order derivatives $\phi'(\cdot)$ of any function $\phi(\cdot)$: $\mathcal{R}^a \to \mathcal{R}^b$, is of dimension $b \times a$, while the second-order derivatives $\phi''(\cdot)$ are of dimension $ab \times a$. For any multi-vector functions $\psi(\mathbf{z}^1, \mathbf{z}^2, \ldots) : \mathcal{R}^{a^1} \times \mathcal{R}^{a^2} \ldots \to \mathcal{R}$ the vector of first-order derivatives ψ_{z^i} are of dimension $a^i \times 1$ and the matrix of second-order derivatives $\psi_{z^i z^j}$ are of dimension $a^i \times a^j$ where the dimension of the matrix follows the order of the subscripts. Superscript T denotes the transpose operator. Vectors and multi-dimensional constructs are denoted in bold, scalars are in normal font.

6.3.2 Incentive Compatibility

Before we go to the problem faced by the social planner, we need to consider the problem of the individuals in our economy. In particular, we derive the conditions under which an allocation is incentive compatible. The incentive compatibility constraints will subsequently be used to solve for the optimal allocation. In a direct mechanism, the social planner offers bundles $\{\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m})\}$ for all $\mathbf{m} \in \mathbf{N}$. Each individual selects a bundle $\{\mathbf{x}(\mathbf{m}), y(\mathbf{m})\}$ by sending a message $\mathbf{m} \in \mathbf{N}$ to the social planner. Function $\mathbf{x}^*$ maps from the message space to the choice-variable space, $\mathbf{x}^* : \mathbf{N} \to \mathbf{X}$ and y^* maps from the message space to the numeraire commodity space, $y^* : \mathbf{N} \to Y$. An allocation $\{\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m})\}$ is

[10] As we show in proposition 1 of Renes and Zoutman (2013b) this assumption ensures that a price-based tax system exists that can implement the optimal allocation on the market.

[11] See Roberts (1984) for a discussion on the issue of commitment.

incentive compatible if each individual truthfully reveals all his unobserved characteristics and receives the bundle designed for him. Formally:

$$\mathbf{n} = \arg\max_{\mathbf{m}} u\left(\mathbf{x}^*\left(\mathbf{m}\right), y^*\left(\mathbf{m}\right), \mathbf{n}\right) \quad \forall \quad \mathbf{n} \in \mathbf{N} \tag{6.4}$$

Let:

$$V\left(\mathbf{n}\right) \equiv \max_{\mathbf{m}} u\left(\mathbf{x}^*\left(\mathbf{m}\right), y^*\left(\mathbf{m}\right), \mathbf{n}\right) \tag{6.5}$$

denote the indirect utility function as a function of the characteristics. In an incentive compatible allocation $V(\cdot)$ satisfies:

$$V\left(\mathbf{n}\right) = u\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right)$$

This equation simply states that maximized utility equals the utility function under optimal choices. Proposition 1 below largely follows Mirrlees (1976) and McAfee and McMillan (1988). It establishes the first and second-order conditions for incentive compatibility.

Proposition 5 *An allocation* $\{\mathbf{x} = \mathbf{x}^*\left(\mathbf{n}\right), y = y^*\left(\mathbf{n}\right)\}$ $\forall \mathbf{n} \in \mathbf{N}$ *is incentive compatible if:*

$$y^{*\prime}\left(\mathbf{n}\right) = \mathbf{s}\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right)^T \mathbf{x}^{*\prime}\left(\mathbf{n}\right), \tag{6.6}$$

$$\mathbf{x}^{*\prime}\left(\mathbf{n}\right)^T \mathbf{s}_{\mathbf{n}} \lessdot 0, \tag{6.7}$$

where the inequality sign, $\lessdot$*, signifies negative definiteness of the matrix.*

Through the envelope theorem a fully equivalent set of conditions can be derived:

$$V'\left(\mathbf{n}\right) = u_{\mathbf{n}}\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right)^T, \tag{6.8}$$

$$u_{\mathbf{nn}}\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right) - V''\left(\mathbf{n}\right) \lessdot 0. \tag{6.9}$$

Proof. *The proof can be found in the appendix.* ■

Equation (6.6) states that an individual should be indifferent between truth telling and mimicking at the margin for all characteristics. For each row j the left-hand side of the equation denotes the gain in y as a consequence of marginally changing the reported characteristic n_j. The right-hand side denotes the utility loss in $\mathbf{x}$ measured in units of y for the same change. Therefore, equation (6.6) states that in equilibrium the marginal cost of mimicking equals the marginal benefits for all characteristics. Equation (6.7) is the usual second-order condition as derived by Mirrlees (1976). If the marginal rate of substitution for decision variable x_i is increasing (decreasing) in characteristic n_j, $(s_i)_{n_j} >$

0 ($(s_i)_{n_j} < 0$), and the allocated amount of the good is also increasing (decreasing) in the characteristic, $(x_i^*)'_{n_j} > 0$ ($(x_i^*)'_{n_j} < 0$), the allocation induces self-selection. It implies higher (lower) quantities of the good are assigned to people with a stronger (weaker) preference for the good.

Equations (6.8,6.9) are fully equivalent formulations of the same incentive constraints. They are derived through the envelope theorem. Although their explanation is less intuitive, they are extremely convenient mathematical expressions in the derivations in subsequent sections.

Together equation (6.7), $k \geq p$, and the assumption that $\mathbf{s}_n$ is of rank p imply that the revelation principle is satisfied. This is shown in the next lemma:

Lemma 3 *If the allocation satisfies (6.7), $\mathbf{s}_n$ is of full rank and $k \geq p$ all characteristics are revealed through the bundles chosen by the agents.*

Proof. Note that (6.7) can only be satisfied if the product $\mathbf{x}^{*\prime}(\mathbf{n})^T \mathbf{s_n}$ is definite, and hence of full rank, p. Since in a matrix product $rank\,(AB) \leq \min\,(rank\,(A)\,, rank\,(B))$, it follows that (6.7) can only be satisfied if the Jacobian of the allocation, $\mathbf{x}^{*\prime}(\mathbf{n})^T$, is also of full rank p. Since $k \geq p$ it follows that the allocation is locally invertible around point $\mathbf{n}$ for all $\mathbf{n} \in \mathbf{N}$. Hence, at least one inverse mapping from the image of the allocation function, $\mathbf{x}^*(.)$ to the type space exists: $(x^*)^{\leftarrow} : \mathbf{X}^* \rightarrow \mathbf{N}$, where $\mathbf{X}^*$ denotes the image or range of the allocation function. It follows that by observing the bundle chosen by the agent, one can deduce all his characteristics. ■

By lemma 3 if the second order incentive constraints (6.7) are satisfied, it follows that the type of the agent can be deduced by observing all his choice variables. The application of the revelation principle is crucial to our analysis, since it allows us to relate optimal policy to observable choices in the remainder of the paper.

6.4 The Second Best Allocation: A First-Order Approach

Now that we have established the conditions for incentive-compatibility we can turn our attention to the social planner. We solve the social planner's problem using a direct mechanism. We will use the first-order approach, and assume that the second-order incentive compatibility conditions are met in the optimum. This can be verified ex-post by checking whether equation (6.7), or, equivalently, equation (6.9) is satisfied. We will return to the problem of violations of the second-order constraints in section 7. In the first-order approach the social planner maximizes social welfare subject to the first-order incentive compatibility constraint, (6.8), and the feasibility constraint, (6.3):

$$\max_{V(\mathbf{n}),\mathbf{x}^*(\mathbf{n}),y^*(\mathbf{n})} \int_{\mathbf{N}} W\left(V\left(\mathbf{n}\right)\right) dF\left(\mathbf{n}\right), \quad s.t. \tag{6.10}$$

$$0 \geq R + \int_{\mathbf{N}} \left(y^*(\mathbf{n}) - q(\mathbf{x}^*\left(\mathbf{n}\right))\right) dF\left(\mathbf{n}\right),$$

$$V'\left(\mathbf{n}\right) = u_{\mathbf{n}}\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right)^T, \tag{6.11}$$

$$V\left(\mathbf{n}\right) = u\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right), \tag{6.12}$$

where maximized utility $V(\mathbf{n})$ is explicitly modeled as a choice variable. The last constraint, (6.12), guarantees that maximized utility is equal to the value of the utility function on the allocation. The Lagrangian to this problem is given by:

$$\mathcal{L}= \int_{\mathbf{N}} \left[\left(W\left(V\right) - \lambda\left(R + y^* - q(\mathbf{x}^*)\right)\right) f + \theta^T \left(V'^T - u_{\mathbf{n}}\right) + \eta\left(u - V\right)\right] d\mathbf{n},$$

where λ is the Lagrangian multiplier associated with the resource constraint, $\theta(\mathbf{n})$ is a p-column vector of Lagrangian multipliers for the set of local incentive compatibility constraints, and $\eta(\mathbf{n})$ is the Lagrangian multiplier that ensures maximized utility equals the utility function for each type. Note that s, f, F, θ, u and their derivatives depend on $\mathbf{n}$, but for clarity of exposition this notation is suppressed. We let $\partial\mathbf{N}$ denote the boundary of $\mathbf{N}$ and $\mathbf{e}$ the outward unit surface normal vector to the boundary of $\mathbf{N}$. Through the divergence theorem (or multi-dimensional integration by parts) we can rewrite the Lagrangian as:

$$\begin{aligned}\mathcal{L} &= \int_{\mathbf{N}} \left[\left(W\left(V\right) - \lambda\left(R + y^* - q(\mathbf{x}^*)\right)\right) f - V \sum_{j=1}^{p} \frac{\partial \theta_j}{\partial n_j} - \theta^T u_{\mathbf{n}} + \eta\left(u - V\right)\right] d\mathbf{n} \\ &\quad + \int_{\partial\mathbf{N}} [V\theta^T \mathbf{e}] d\partial\mathbf{N}. \end{aligned} \tag{6.13}$$

Assuming the functions V and θ are smooth, this function can be maximized pointwise on the interior and boundary of the type space.

On the interior of the type space the first-order conditions with respect to variables $\mathbf{x}$, y and V are:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial y} &= 0 : -\lambda f - u_{y\mathbf{n}}\theta + \eta u_y = 0, && (6.14)\\ \frac{\partial\mathcal{L}}{\partial \mathbf{x}} &= \mathbf{0}_k : \lambda q'^T f - u_{\mathbf{xn}}\theta + \eta u_{\mathbf{x}} = \mathbf{0}_k, && (6.15)\\ \frac{\partial\mathcal{L}}{\partial V} &= 0 : W' f - \sum_{j=1}^{p}\frac{\partial\theta_j}{\partial n_j} - \eta = 0. && (6.16)\end{aligned}$$

The next proposition uses these first-order conditions to derive the ABC-formula for the optimal wedge in the spirit of Diamond (1998) and Saez (2001).

Proposition 6 *The optimal wedge on good i for type $\mathbf{n}$ can be described by the following formula:*

$$\begin{aligned}\frac{q_{x_i}\left(\mathbf{x}^*\left(\mathbf{n}\right)\right) - s_i\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right)}{s_i\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right)} &= \sum_{j=1}^{p} A_{ij}\left(\mathbf{n}\right) B_{ij}\left(\mathbf{n}\right) C_{ij}\left(\mathbf{n}\right) && (6.17)\\ \forall \quad i &= 1, \ldots, k; \mathbf{n} \in \mathbf{N},\end{aligned}$$

where:

$$\begin{aligned}A_{ij}\left(\mathbf{n}\right) &\equiv \varepsilon_{x_i n_j}\left(\mathbf{n}\right) = -\frac{\partial s_i\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right)}{\partial n_j}\frac{n_j}{s_i\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right)},\\ B_{ij}\left(\mathbf{n}\right) &= \theta_j\left(\mathbf{n}\right)\frac{u_y\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right)}{\lambda}, && (6.18)\\ C_{ij}\left(\mathbf{n}\right) &= \frac{1}{n_j f\left(\mathbf{n}\right)}.\end{aligned}$$

Proof. The proof can be found in the appendix. ■

Note that proposition 6 gives the equation for the optimal wedge but gives no information about the optimal tax rate. However, as we show in proposition 1 of Renes and Zoutman (2013b), if the allocation is optimal to a welfarist planner and there are no externalities, any tax system can implement the allocation as long as the marginal tax rate on each good is equated to its wedge. As such, the wedges derived above contain all relevant information for the tax system. In the next section we discuss the the ABC-formula and compare it to the optimal tax formula under uni-dimensional heterogeneity. In the next subsection we use proposition 6 to derive the optimal tax rate at the boundaries of the type space.

6.4.1 Boundary Conditions: No Distortion at the Corners

The boundary conditions can be found by differentiating equation (6.13) with respect to $V(\cdot)$ at the boundary of the type space ∂N. Only the final term of equation (6.13) depends on the boundary. Hence, we derive:

$$\theta_j\left(\underline{n}_j\right) = \theta_j\left(\overline{n}_j\right) = 0, \tag{6.19}$$

where $\underline{n}_j$ $(\overline{n}_j)$ represents the type that has the lowest (highest) value for characteristic j. Define corner types $\mathbf{n}_\lrcorner^\urcorner$ as agents that have either highest or lowest values for their characteristics. In a two-dimensional type space type $\mathbf{n} = (\underline{n}_1, \underline{n}_2)$ and $\mathbf{n} = (\overline{n}_1, \overline{n}_2)$ are obviously corner-types but so are the types that combine the lowest value of n_1 with the highest value of n_2 and vice versa: $\mathbf{n} = (\underline{n}_1, \overline{n}_2)$ and $\mathbf{n} = (\overline{n}_1, \underline{n}_2)$. There are at most 2^p corner types. However, there may be less, or none at all, depending on whether there is a positive density $f(.)$ at each corner type, and on whether the distribution is bounded. Corollary 7 establishes that for each existing corner-type the optimal wedge on all goods equals zero.

Corollary 7 *The optimal wedge for any type* $\mathbf{n}_\lrcorner^\urcorner$ *equals zero if the type exist.*

Proof. From the boundary conditions it follows that $\theta_j\left(\underline{n}_j\right) = \theta_j\left(\overline{n}_j\right) = 0$ for all $j = 1, \ldots, p$. The optimal wedge at the corner types can be found by taking the limit of equation (6.17) if $\mathbf{n}$ goes to a $\mathbf{n}_\lrcorner^\urcorner$.:

$$\lim_{\mathbf{n}\to\mathbf{n}_\lrcorner^\urcorner} \frac{q_{x_i} - s_i}{s_i} = \lim_{\mathbf{n}\to\mathbf{n}_\lrcorner^\urcorner} \sum_{j=1}^{p} \varepsilon_{x_i n_j} \frac{u_y \theta_j(\mathbf{n})/\lambda}{n_j f(\mathbf{n})}$$

which equals 0 provided $f\left(\mathbf{n}_\lrcorner^\urcorner\right)$ does not equal zero, that is provided the type exists in the economy. ■

Corollary (7) shows that the no-distortion at the top and the bottom result, as derived in Sadka (1976) and Seade (1977), remains valid in a multi-dimensional framework as long as the type-distribution is bounded.

Technically, the no-distortion at the corner result derived in corollary (7) follows from the transversality conditions of the optimization problem. If there are no individuals of extremer type, distorting their choices will not yield any extra information. In terms of our motivating examples, if an individual is the healthiest, most able person around, distorting his choices will not lead to a redistributive benefit, but will come at an efficiency loss. As such, the optimal wedge at the corners of the type distribution must equal zero.

Our results are similar to those of Golosov *et al.* (2011) who derive the optimal tax rate at the boundary in a model where earnings ability follows a stochastic progress. They show that the optimal tax for types that persistently have the highest or lowest skill-realization equals zero provided such types exist.

Mirrlees (1971), Diamond (1998) and Saez (2001) show that the optimal tax converges to a constant at the top in the uni-dimensional case provided the upper tail of ability follows a Pareto-distribution. We are not able to derive such a result in the multi-dimensional framework since this requires an explicit solution for all θ_j.

6.4.2 Finding the Optimal Allocation

The last step in the problem of the planner is to find the second-best allocation. Although, equation (6.17) gives a useful representation of the wedge, there is no closed-form solution for the optimal allocation. Depending on the specification, deriving the optimal allocation may be a computationally complex process and certainly goes beyond the scope of this paper. However, in this subsection we give a sketch of an algorithm that can solve for the optimal allocation. The solution method described here is largely based on Mirrlees (1976).

First, the set of equations (6.11, 6.12) can be used to solve for $\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\}$ as an (implicit) function of $V'(\mathbf{n}), V(\mathbf{n})$ and $\mathbf{n}$:

$$\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\} = \psi(V'(\mathbf{n}), V(\mathbf{n}), \mathbf{n}).$$

Second, by means of this equation and (6.14) and (6.15) we can solve for $\{\theta(\mathbf{n}), \eta(\mathbf{n})\}$ as an explicit function of $\mathbf{x}^*(\mathbf{n})$, $y^*(\mathbf{n})$ and $\mathbf{n}$, and hence as an (implicit) function of $V'(\mathbf{n}), V(\mathbf{n})$ and $\mathbf{n}$:

$$\{\theta(\mathbf{n}), \eta(\mathbf{n})\} = \hat{\phi}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})) = \phi(V'(\mathbf{n}), V(\mathbf{n}), \mathbf{n}).$$

Finally, if we substitute this result into the last first-order condition (6.16) it becomes a second-order partial differential equation, that can be integrated numerically under the boundary conditions (6.19) and (6.3). The solution provides us with $V'(\mathbf{n})$ and $V(\mathbf{n})$ which can subsequently be used to find the allocation $\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\}$. From here the optimal wedges can be found by substituting the solution $\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\}$ into the ABC-formula (6.17).

6.5 The ABC Formula for the Optimal Wedge

In this section we use the ABC-formula (6.17) to compare the optimal wedge under multi-dimensional heterogeneity to the outcome in the uni-dimensional case. In addition, we use the ABC-formula to revisit the Atkinson-Stiglitz theorem under multi-dimensional heterogeneity in the next subsection.

As in the uni-dimensional case, the left-hand side of equation (6.17) represents the optimal wedge on good i for type $\mathbf{n}$. This distortion is broken down into different factors of interest on the right-hand side.

The A-term is a measure of the informational value of good x_i. Intuitively, if the elasticity $\varepsilon_{x_i n_j}$ is large, it means that the preference for choice i strongly increases in characteristic j. Hence, x_i is a very strong signal of characteristic j, and therefore, the optimal wedge is large. Our A-term is more general than the one derived in, Diamond (1998), Saez (2001) and Jacquet *et al.* (2010) because we use a more general utility function. In Diamond (1998) and Saez (2001) the utility-function is of the form: $u(y,l) = u\left(y, \frac{x}{n}\right)$, where y is consumption, n is productivity and $x = nl$ is effective labor supply (or labor income). Their A-term is inversely related to the compensated labor supply elasticity. In Jacquet *et al.* (2010) the assumed utility function is $u^1(y) + u^2(x,n)$. Their A-term is also inversely related to the compensated labor supply elasticity. Their results can be explained using our interpretation. If the labor supply elasticity is large, this means that a small change in the net wage rate leads to a large change in labor income. Therefore, labor income is an imprecise signal of ability. It follows that the optimal tax rate on labor is decreasing in the labor supply elasticity, since the higher the elasticity the less information is gained from taxing labor income .

The B-term represents the redistributive benefits of distorting choice i for characteristic j. θ_j is the Lagrangian multiplier of incentive compatibility constraint j. Hence, it represents the welfare cost of separating type $\mathbf{n}$ in characteristic j. In equilibrium θ_j should equal the marginal welfare benefit of making the allocation marginally less incentive compatible in choice j (i.e. increase the distortion on choice j). By multiplying θ_j with u_y and dividing through λ the welfare gain for such a redistribution is expressed in units of the numeraire good. An increase in the marginal welfare benefit of distortion, higher θ_j, logically increases the optimal distortion.

The C-term is related to the size of the tax-base that is distorted by the wedge. The denominator represents the size of the tax base with respect to characteristic j. The larger this tax base is, the larger the incidence of the distortion and hence, the larger the efficiency cost associated with the distortion. Efficiency implies that the size of the distortion is inversely related to the incidence of the distortion, as the C-term clearly shows. In uni-dimensional cases, the C-term is often multiplied by $1 - F(n)$ to make it proportional to the (measurable) inverse hazard-rate of the ability distribution. This is corrected for by dividing the A or B-term through the same factor. In a single-dimensional distribution of types, such fractions have an intuitive interpretation as conditional means. Unfortunately, this interpretation is lost when the type-distribution is multi-dimensional.

The largest difference between the uni-dimensional and the multi-dimensional ABC-formula is the need to sum over all characteristics to get the optimal wedge for a good i in the latter case. This indicates that the optimal wedge on good i is the sum of the optimal wedge for each characteristic. For example, if earning labor income is easier for agents with higher ability and for agents with a better health state, the planner can calculate the optimal wedge on labor income by adding the optimal wedge on the basis of redistribution in ability to the optimal wedge of redistribution in health. This additive nature of the wedge is particularly useful for policy evaluation.

6.5.1 The Atkinson-Stiglitz Theorem

The Atkinson-Stiglitz theorem, Atkinson and Stiglitz (1976), from now on AS theorem, states that indirect taxation is superfluous in a setting where agents are heterogeneous in earnings ability if preferences are homogenous and the utility function is weakly separable in labor. The result has been generalized subsequently in Laroque (2005), Kaplow (2006), Gauthier and Laroque (2009) and Hellwig (2010) who show that in a setting with uni-dimensional heterogeneity the AS-theorem also holds with heterogeneous preferences as long as the preferences of all goods except labor income are weakly separable from ability. Therefore, if the conditions of the AS theorem are satisfied the government can reach the second-best allocation by only taxing labor income, or equivalently by taxing all commodities at an uniform rate. The main application of the theorem is perhaps that commodities should be taxed at the same rate over time. That is, the optimal capital-tax rate equals zero.

In the next corollary we use our ABC-formula to investigate when the optimal wedge on a good equals zero and use this to investigate under which conditions the AS-theorem holds under multi-dimensional heterogeneity.

Corollary 8 *The optimal wedge on good i is zero if $\varepsilon_{x_i n_j} = 0 \quad \forall \quad x_i, n_j,$, that is if the marginal rate of substitution for x_i does not depend on any characteristic n_j for all types.*

Proof. If the marginal rate of substitution, s_i, is independent of all characteristics n_j, then $\varepsilon_{x_i n_j} = 0 \quad \forall \quad x_i, n_j$, such that all A_{ij} are zero and the optimal wedge on x_i is zero by equation (6.17). ■

Intuitively, corollary 8 shows that the marginal wedge on a good equals zero if the preference for the good is not directly influenced by any characteristic. In that case the good does not provide any first order information and distorting it away from laissez-faire is not optimal. It follows immediately from corollary 8 that the optimal wedge on all goods except income equals zero if the marginal rate of substitution for all goods except income is independent of the type. Because we have assumed $\mathbf{s_n}$ has rank p, $\varepsilon_{x_i n_j} \neq 0$ for at least p choices in our model. Hence, if all characteristics are independent and revealable, the A-S theorem does not hold under multi-dimensional heterogeneity. Intuitively, a government that wants to redistribute in multiple dimensions, cannot do so by distorting the price of only one good.

In the literature many violations of the AS theorem have been recorded. In Erosa and Gervais (2002) preferences are not weakly separable over time since consumption at old age is a stronger complement to leisure than consumption at a younger age. Therefore, the distortion of the labor income tax is reduced by taxing capital income. In Golosov *et al.* (2013) capital is optimally taxed if households with higher ability also have higher patience. Boadway and Pestieau (2011) show that the theorem fails if households optimally choose a corner solution. Therefore, if some households are cash-constrained and do not buy all commodities, the tax rate on these commodities should be different. Farhi and Werning (2010) and Kopczuk (2013) show that the bequest motive may generate a negative externality which can be remedied through the taxation of capital. The argument that is closest to our is derived in Cremer *et al.* (2001) and Saez (2002a) who show that under bi-dimensional heterogeneity commodity taxation is not superfluous. However, the former result is derived in a setting with discrete types and the latter is derived under the assumption that welfare weights are only correlated with earnings ability and commodity taxes are linear.

Our result adds to this literature by showing generally that the AS theorem cannot hold under multi-dimensional heterogeneity. This has large implications for the evaluation of government policy. According to the A-S theorem we can obtain a second-best allocation if the only government intervention is the taxation of labor income. We show that if a government cares about redistribution in at least two dimensions, such as from healthy to sick, it needs to distort multiple choices in order to attain the second-best allocation. Therefore, government intervention in many markets, for example, the health care and the rental market may be optimal. In addition, the optimal wedge on capital income may be non-zero if households differ in their investment skills.

6.6 Interdependencies in the Tax System in Mirrleesian and New Dynamic Public Finance

The NDPF pioneered in Golosov *et al.* (2003) generalizes the Mirrlees model with unidimensional heterogeneity to a dynamic stochastic setting. For each agent, earnings ability takes a different value in each period. Each agent's earnings ability is revealed to him at the beginning of the period. Agents therefore do not know what their earnings ability will be in future periods, but they are aware of the stochastic process that generates it. The planner does not observe ability, but does observe all choice variables taken by the agents in the economy, and can keep records over time.

One of the most intriguing results in the NDPF is the complexity of the optimal tax system. Kocherlakota (2005) shows that the optimal tax on labor income in period t may depend on the entire history of labor income up to period t.[12] However, so far there is no clear explanation why interdependencies are present in the NDPF in the first place, or, alternatively, why they are typically absent in the Mirrleesian public finance. In this section we show that the interdependencies in the wedges generally do occur in Mirrleesian public finance models provided agents are heterogeneous in multiple dimensions. Subsequently, we give our model a dynamic interpretation, and provide a link between the result found in Kocherlakota (2005) and our model.

The next corollary of proposition 6 and lemma 3 shows that the optimal wedge on each choice depends on p choices.

Corollary 9 *The optimal wedge on good i depends on p choices.*

Proof. By lemma 3 there exists at least one inverted mapping of the allocation $(x^*)^{\leftarrow}$: $\mathbf{X}^* \to \mathbf{N}$ for all $\mathbf{n} \in \mathbf{N}$. This inverse mapping generally depends on p choice variables, since the allocation is of rank p everywhere on the type space. Therefore, we can write the optimal wedge as a function of p choices. By equation (6.17) we know that the wedge solves a set of partial differential equations, that are of rank p by assumption 1, hence each individual wedge can be written as a function of p choices. ■

[12] Subsequent papers have made some progress on limiting the number of interdependencies in special cases. In particular, Albanesi and Sleet (2006) show that the intertemporal interdependencies disappear when the stochastic process is iid. In that case, they show the optimal wedge in each period depends on current labor income and wealth, such that only an intratemporal interdependency remains between the optimal wedge on capital and labor income.

If all wedges are functions of p choices, then by construction so are all marginal taxes (see Mirrlees, 1976 and Renes and Zoutman, 2013b). In general, therefore a separable tax system where the wedge of each good depends only on consumption of that good does not exist if $p > 1$. Intuitively, the optimal wedge on each good generally depends on all characteristics. By the revelation principle in order to reveal all characteristics one needs at least p observable choices. For instance, if the optimal wedge on labor income depends on both earnings ability and health, the government needs to condition its tax rate on two observable choices that reveal both characteristics, for instance labor income and consumption of health care products.

There are a number of mathematical reasons why the optimal wedge for each good is almost always a function of all underlying characteristics. First, most probability density functions, $f(\mathbf{n})$, are a function of all characteristics such that equation (6.17) depends on all characteristics. In addition, the partial differential equations (6.16), which solve for the θ's, are a function of indirect utility. In equilibrium, indirect utility always has to be a function of all $\mathbf{n}$ in order to fulfill the incentive compatibility constraint (6.11). The solution to the set of partial differential equations in equation (6.17) therefore depend on all equations. Although, it may be possible to construct special cases where the optimal wedge on each good depends on only one characteristic, we generally expect any optimal tax system that redistributes in multiple dimensions to have interdependencies, equal to the dimensionality of the problem.

6.6.1 A Dynamic Interpretation

Consider a T-period economy where agents are heterogeneous in ability, and ability evolves over time. Denote ability in period t by n_t and assume the agents know in advance the entire history of ability levels, the vector $\mathbf{n} = [n_1, \ldots, n_T]$ at the beginning of the first period. Clearly, this is a strong assumption on the information available to the agents of the model. In contrast, in the NDPF agents do not know their ability before the beginning of each period, since ability evolves stochastically over time. Note however, that this model can be seen as a special case of a NDPF model where ability evolves according to a fully deterministic process. The government cannot observe the individual ability levels but is aware of its probability density function $f(\mathbf{n})$, and cumulative density function $F(\mathbf{n})$. In addition, assume for simplicity that each agent makes one independent choice each period, the amount of labor income he earns, x_t. Our model can be used to calculate the welfare-optimizing wedge on each period's labor income x_t. By proposition 6 the optimal wedge is given by equation (6.17). From corollary 9 it follows that the optimal wedge in each period can be written as a function of the entire history of labor income. Intuitively, the optimal wedge depend on terms A, how much information a choice reveals, B, the redistributive value of the wedge and C the density of the tax base. Each of these properties may depend on the entire vector $\mathbf{n}$. In turn, in order to reveal the entire vector the planner needs to observe all choices in $\mathbf{x}$, the history of labor income.

It may perhaps seem surprising that Kocherlakota's result can be replicated so easily in a deterministic setting. However, as was noted in Pavan *et al.* (2010), in a deterministic model agents can plan their entire choice vector $\mathbf{x}$ in the first period, with full knowledge about their type. In a stochastic setting information concerning the type is revealed over time. As such, a stochastic model may allow for less (profitable) deviations than a deterministic model. This implies the incentive constraints are more binding in our model than in the NDPF. Hence, if intertemporal interdependencies are optimal in a stochastic setting, they are likely also optimal in a deterministic setting. It also follows that interdependencies in the tax system might not be the result of any particular stochastic process, but instead follow from the multi-dimensionality in the type space.

The practical implications of this result are less clear. A literal reading of the model we described in this subsection would suggest the government should build intertemporal interdependencies in the tax system, in which for example, the labor income tax may be higher if a person has earned more income in previous periods. However, it is unclear how much welfare is gained by introducing these interdependencies. Simulations on a NDPF model in Farhi and Werning (2013) show that much of the welfare gain of optimal taxation can be obtained by tax systems that do not exhibit any interdependencies.

6.7 Bunching

In most multi-dimensional screening problems the second-best allocation contains bunching at the lower end of the type distribution. Bunching occurs if the first-order approach violates the second order conditions (6.7) on a part of the type space. In this case, certain types would prefer the bundle of another type over the one assigned to them. In our examples wealthy, highly able individuals might prefer the bundle the planner intents for low ability individual, and consume a lot of leisure. In this paper we relied on the first-order approach to derive the optimal allocation, and have therefore ignored possible violations of the second-order incentive constraints, (6.7).

Our analysis may still remain useful even if second-order incentive constraints are violated. Suppose, the first-order approach violates the second-order constraints (6.7) on some part of the type space with a non-zero measure, the bunching partition $\mathbf{N}_B$. In addition, suppose second-order conditions are not violated in another part of the space, the separating partition $\mathbf{N}_S$. The next proposition shows that the separating partition is convex and it extends to the upper boundary of the type space.

Proposition 10 *If $\mathbf{N}_S$ exists, it is a single convex set that extends from the upper boundary of the type space.*

Proof. The proof can be found in the appendix. ■

Intuitively, the proof of proposition 10 follows from the fact that incentive problems are more likely to occur at the bottom of the type space. The planner wants to extract as much informational rents from types at the top of the type space as possible, such that it can redistribute these rents to the bottom types. Therefore, bunching at the top of the type space is particularly costly to the planner. To facilitate full separation of the types at the top, it may be optimal in some problem to bunch together types at the bottom of the distribution. In screening models with binding participation constraints there is always bunching at the bottom of the type space as was proven in Zheng (2000) for auctions, Armstrong (1996) for non-linear multi-product monopoly pricing and Rochet and Choné (1998) for general screening problems. An attempt to extract all rents from agents at the bottom of the type space will inevitably lead to non-participation of some of these agents.

Although there is no participation constraint in our model, we cannot exclude the possibility that the first-order approach will violate second-order conditions on part of the type space. However, by proposition 10, if separation is optimal in a partition of the type space $\mathbf{N}_S$, this partition is convex. It follows, that the optimal wedge in the separating partition is described by proposition 6. Hence, even if the first-order approach does not yield the optimal wedge for every type, this does not necessarily invalidate our approach since it still describes the optimal wedge in the partition of the type space where full separation occurs.

6.8 Concluding Remarks

Although significant progress has been made in multi-dimensional mechanism design, the equilibrium in a multi-dimensional Mirrleesian optimal tax model had so far not been characterized. In this paper we characterize it and show some similarities and differences with the uni-dimensional Mirrleesian model. Furthermore we show how the equilibrium relates to the stochastic dynamic NDPF models and the wider class of multi-dimensional screening models.

Our model can be used to study the relationship between several tax-tools. Our characterization of the general equilibrium shows that the government should search for consumption patterns that provide as much information on the underlying types as possible. More importantly, the multi-dimensionality in the type space forces the government to make the redistributive taxes depend on several observable choices to separate out different aspects of the hidden types. It might not be optimal to separate out types everywhere in the type space, in which case some bunching occurs at the lower end of the type space. This prescription fits reasonably well with the tax-schedules observed in welfare states. The lowest earning individuals get welfare assistance, or income subsidies, creating a bunch at the lower end of the income distribution. Most assistance programs are conditioned on (the absence of) wealth, to make sure that no abuse occurs. This is the kind of interdependencies between underlying characteristics (wealth and ability) our model predicts. Many welfare states also subsidize medical expenses or housing for a large group of people. In theory the government could directly transfer the required money, rather than paying part of the price. A direct transfer, however, would make it impossible for the government to find out whether or not you are in need of health-care, i.e. the government could not determine your hidden type through a direct transfer, but can do so through the subsidy. We would therefore indeed predict the government can better use a subsidy that depends on income and expenses (or other observables) rather than direct transfers for differentiated assistance.

The equilibrium in our model depends on solving a set of partial differential equations for which no general solution exists, therefore we can only characterize equilibrium trough a set of necessary conditions. These conditions strongly limit the possible outcomes, but can never give a full description of the second-best. The next step in this line of research clearly is to find specific, realistic and relevant settings and simulate the model. This is, however, a difficult step. The multi-dimensional heterogeneity sets strong requirements on the optimization algorithms. In addition, the problem of implementation, which is discussed in the companion paper Renes and Zoutman (2013b), might add further difficulties. Implementation might proof especially difficult because implementation on the interior of the separating partition will likely require a different set of instruments than the implementation on the bunching partition. However, once these difficulties have been overcome the model presented in this paper can be used to provide a more precise insight in the optimal relation between the income tax system and the myriad of social schemes like healthcare subsidies, housing subsidies, and welfare that characterize modern welfare states.

Since this model contains the multi-dimensional type space that is also found in the NDPF, it could also provide a convenient middle ground between the complex stochastic dynamics in these models, and the known intuitions in the classical Mirrlees model. The problems of joint (or double) deviation and the interdependencies in the optimal wedges that plague NDPF models are, for instance, also prevalent in our setting, but can be traced much more conveniently. These results indicate at least part of the difficulties in the NDPF literature are due to the structure of hidden information, showing that we might be able to gain intuition for these tax-schedules from multi-dimensional screening models, in particular our model. In fact, the discussion in section six already suggests that our findings might be generalized to dynamic settings. This would allow an elasticity approach and a new focus on implementation in these models as well.

Chapter 7

When a Price is Enough: Implementation in Optimal Tax-Design[1]

7.1 Introduction

The tax system is one of the most important tools used by modern governments. In welfare states the tax system has evolved into a complex system, that is characterized by a myriad of non-linear instruments. Taxes on labor income, capital income and commodity consumption are combined with subsidies on healthcare, housing and education. Governments use these instrument in an effort to insure their constituents against adverse outcomes when insurance markets fail, and to redistribute from the fortunate to the less fortunate in society. In a game-theoretic framework, the tax system incentivizes the economic actors in society to make socially desirable choices, by influencing their budget-relevant choices.

The formal study of optimal non-linear redistributive tax systems was pioneered by Mirrlees (1971, 1976). In his model, agents are heterogeneous in their earnings ability. A social planner wants to redistribute from agents with high to agents with low earnings

[1]This chapter is based on Renes and Zoutman (2013b). We would like to thank Felix Bierbrauer, Eva Gavrilova, Aart Gerritsen, Yasushi Iwamoto, Bas Jacobs, Laurence Jacquet, Etienne Lehmann, Dominik Sachs, Dirk Schindler, Bauke Visser, Casper de Vries and Hendrik Vrijburg for useful suggestions and comments on an earlier version of this paper. Furthermore, this paper benefited from comments and suggestions made by participants at the 2011 Nake Conference, Utrecht, the 2013 CESifo Area Conference on Public Economics, Munich, the 69th IIPF Conference, Taormina; ; and seminar participants at the Erasmus School of Economics, the Norwegian University of Science and Technology, the Norwegian School of Economics, the University of Konstanz and the Centre for European Economic Research. All remaining errors are our own.

ability, but earnings ability is private information, and hence, the first-best is not generally attainable. Instead, in the second-best allocation the government optimally distorts agent's choices.

Following Mirrlees, the typical approach in optimal taxation consists of two steps. First, identify the second-best allocation. The second-best allocation maximizes the welfare function subject to a resource and an incentive-compatibility constraint in a direct mechanism. By the incentive-compatibility constraint, this allocation will incentivize agents to make socially desirable choices in a direct mechanism. However, this by no means guarantees that agents will make the same desirable choices in a market economy, since the market typically gives agents a much larger set of choice variables. Therefore, in the second step a tax system is designed that successfully implements the second-best allocation in a market economy.

Surprisingly, the second step of actually designing the optimal tax system has received very limited attention in the literature. Indeed, as we show in this paper, the canonical approach to tax design may in some cases incentivize agents to take choices that are entirely undesirable from a social perspective, and move the economy far away from the second-best allocation. Therefore, in this paper we derive conditions under which the canonical approach, and indeed any price mechanism where marginal tax rates are equated to the optimal wedge between the marginal rate of substitution and the marginal rate of transformation in the second-best allocation (further: taxes equal to wedges), successfully implements the second best in the market.

The canonical approach to optimal taxation was derived in Mirrlees (1976). His approach to implementing the second-best allocation is simply to equate the tax rate to the optimal wedge. This approach has subsequently throughout the literature on optimal taxation (see e.g. Atkinson and Stiglitz, 1976, Diamond, 1998, Saez, 2001, Bovenberg and Jacobs, 2005 and Golosov *et al.*, 2013). However, to the best of our knowledge, an attempt to derive the general conditions under which this canonical mechanism can implement the second-best allocation in a market economy have never been derived.

This concern has not been ignored entirely in the literature. Most notably the principle of taxation, derived by Hammond (1979), derives a different mechanism that can successfully implement every incentive compatible allocation.[2] This tax system combines a price mechanism with a potentially large set of rules. In particular, any bundle of choices that is not assigned to a type in the second-best allocation is prohibited or taxed at an infinite rate in the market. Therefore, the incentives created by this tax system

[2]A similar, more general result has been derived in Maskin (1999) for any countable number of agents.

are mathematically equivalent to the incentives in a direct mechanism. Hence, incentive compatibility in the direct mechanism implies implementability through this tax system.

However powerful, this result is of limited value to policy makers in market economies. Restricting the choices of economic agents to the choices that are available to them in a direct mechanism effectively removes all benefits of free choice and limited administrative costs associated with a market mechanism.[3] This may account for the fact that a large part of the literature ignores the design of these rules altogether.

This leaves a gap in our understanding. It is unclear whether the canonical tax implementation works, while the system that always works is of limited value to policy makers in market economies. We address this gap in the literature by the implementation of a second-best allocation in the market in a optimal tax model.

The analytic starting point of this paper is a second-best allocation of which we know only three things: i.) it satisfies the economy's resource constraint, ii.) it is incentive-compatible in a direct mechanism, iii.) it maximizes some welfare function under the constraints given by the other two conditions. The allocation may be multi-dimensional in both the number of decision variables and the number of unobserved characteristics in which agents differ. It is well-known in the literature that the derivation of the second-best allocation under multi-dimensional heterogeneity is technically complex. In the companion paper, Renes and Zoutman (2013a), we take up this issue and set the first steps towards fully characterizing the second-best under multi-dimensional heterogeneity of agents. However, in this paper we entirely ignore the first step of actually deriving the second-best allocation, and take it as a given. This allows us to study conditions for implementing a second-best in a wide variety of taxation models, and under a wide variety of welfare concepts. Starting from such an allocation, we study under which conditions a price mechanism without rules (further: a pure price mechanism) can implement the second-best allocation.

There is no clear distinction between a rule and a price. In a non-linear tax system, any legal rule can arbitrarily be replaced by an infinite, or arbitrarily high, tax rate, leading to an excessively high price. If we allow for such a broad definition of a pure price mechanism, the distinction between a pure price mechanism and a rule-based mechanism becomes meaningless. It is therefore necessary to tighten the definition of a pure price mechanism. A good starting point is the canonical approach to implementation which is

[3]The name 'principle of taxation' stems from later applications of the principle to taxation by e.g. Rochet (1985), Guesnerie (1995) and Bierbrauer (2009). However, the strength of the principle lies mostly in its application to the multi-product monopolist pricing problem and auction design (see e.g. Armstrong (1996)). In such a setting the application leads to perfectly realistic implementations, since the monopolist and the auctioneer can choose what (not) to produce and how to bundle their goods.

a unique pure price mechanism without any rules. However, the canonical approach is not defined under multi-dimensional heterogeneity of agents. Therefore, in order not to limit our attention to the case of the canonical approach of implementation under uni-dimensional heterogeneity, while at the same time not extending the definition of a pure price mechanism to tax rates that effectively act as rules, we define a pure price mechanism as a mechanism where the tax rate on a choice variable never exceeds (is below) the maximum (minimum) wedge on that choice variable in the second-best allocation. A nice property of this definition is that the canonical approach is a special case of a pure price mechanism. Therefore, if we prove any pure price mechanism implements the second-best allocation under a set of conditions, we implicitly also prove that the canonical approach is successful under the same set of conditions.

Our analysis starts with an example showing that a pure price mechanism may fail to implement the second-best allocation. We construct the simplest possible example, with uni-dimensional heterogeneity of agents, where the canonical approach steers agents on the market away from the second-best. The intuition of this result lies in the difference between incentive-compatibility in the direct mechanism and implementability in the market. Incentive compatibility in the direct mechanism requires that each agent prefers his bundle over the bundle of all other agents. In the market agents can create bundles themselves, rather than choosing one from the set of bundles designed by the planner. As a result, they can choose bundles that are not assigned to any type in the direct mechanism. A tax system can only implement the desired allocation if such non-marginal or joint deviation are not profitable or not possible for any agent. In a pure price mechanism it may be impossible to deter all agents from such joint deviations in equilibrium, hence a pure price mechanism is not always able to implement the allocation.

We proceed by deriving a lemma which describes the general conditions under which a pure price tax system implements the desired allocation. Implementation requires that marginal taxes are equal to optimal distortions and indifference curves are more convex than budget constraints in all linear combinations of the decision variables. Economists can use these implementability constraints to verify whether a proposed tax system implements the desired allocation. That is, after solving the maximization problem of the planner and formulating the entire tax system, it can be verified whether the tax system satisfies these constraints. or rules are required to implement the allocation. Unfortunately, the lemma is of limited use since most optimal allocations in public finance do not have a closed-form solution. Solutions may be obtained through numerical simulations, but this implies that verification of implementation can only be performed on the special cases that have been simulated. Verification of implementability is useful in such simulations, but does not provide insights in the general properties of optimal tax systems.

Therefore, our main contribution lies in identifying two classes of maximization problems in which a pure price mechanism can implement the allocation if taxes are equated to wedges on the allocation. First, the pure price mechanism successfully implements the second best if i.) the allocation is second-best for a welfarist social planner, ii.) the tax system does not have an internal maximum in tax revenue, and iii.) there are no externalities. This result holds independently of the preferences of the agents. Intuitively, if joint deviations are optimal to the agent, symmetry of second-order derivatives implies there exists a joint deviation which increases the utility of the agent and weakly increases tax revenue. Such a deviation entails a Pareto improvement over the original allocation. Hence, the initial allocation could not possibly have been second-best to a welfarist planner.

This result does not hold if the tax function has an internal maximum, because from such a maximum a deviation that weakly increases tax revenue does not exist. However, most optimal tax systems are either monotonic or convex and as such they usually do not exhibit internal maxima. More importantly, the result does not hold under non-welfarist governments or with externalities. In both cases an increase of utility for the agent does not automatically mean that the objective function of the government increases.

Second, we show a pure price mechanism succeeds if the allocation determines a one-to-one correspondence between the type space and the choice space. The prime example is the Mirrlees (1971) model where agents differ in earnings ability, and the only decision they make is how many hours they work. In this case, the agents problem on the market is identical to the agents problem in the direct mechanism. Therefore, incentive compatibility and implementability constraints coincide. Since we assumed the original allocation was incentive compatible, it must also be implementable in the market.

Note that we derive sufficient conditions, not necessary conditions. There are cases outside of these two classes for which a pure price mechanism suffices to implement the second-best allocation, but implementation cannot be guaranteed ex-ante through our propositions. However, the two identified classes are of enormous importance since they encompass virtually all models based on Mirrlees (1971, 1976) and Renes and Zoutman (2013a) and as a result validate almost all tax systems proposed in the literature.[4]

This paper provides a guide to the relatively understudied second step of optimal tax design. When an optimal tax problem fits one (or both) of the two classes identified a

[4]A notable exception is Jacobs and De Mooij (2011) which extends the Mirrlees model with externalities.

pure price mechanism can implement the second best. In addition, when the problem does not fit in these classes, our first lemma can be used to check whether a particular tax system yields the desired allocation. More generally our results show an imperfect link between direct and indirect mechanisms. A central planner that perfectly observes choices and can price/tax them non linearly, might still want to rely on quotas or legal prohibitions to reach the second-best allocation. This provides some intuition for the existence of (possibly optimal) complexities in the tax systems in modern welfare states. To prevent abuse of, for instance, social insurance a central planner may sometimes have to restrict the choices of (potential) beneficiaries, and force them to study, apply for jobs, or enroll in debt counseling for instance.

The rest of the paper is organized as follows. Section 2 discusses related literature. Section 3 introduces the model. Section 4 contains an example, which illustrates the failure of the pure price mechanism. Section 5 derives our main result and section 6 concludes.

7.2 Related Literature

In this paper we do not study the technically complex issue of deriving the optimal allocation under multi-dimensional heterogeneity. Instead we focus on implementing a given second-best allocation in the market through a tax system. However, a large and growing literature has been devoted to deriving the second-best allocation under multi-dimensional heterogeneity (see e.g Mirrlees, 1976, Cremer *et al.*, 2001, Saez, 2002a, Kleven *et al.*, 2009, Choné and Laroque, 2010 and Renes and Zoutman, 2013a for applications in an optimal taxation framework, and Armstrong, 1996, Rochet and Choné, 1998 and Armstrong and Rochet, 1999 for a more general treatment of screening problems). The mechanisms that implement the obtained second-best allocation in the market are often left implicit. Results obtained in this paper may be useful in designing mechanisms that implement these allocations in the market.

Implementation issues have received limited attention in the optimal tax literature but have received considerable attention in other fields of mechanism design. In auction and procurement theory joint deviations are known as unbalanced or skewed bidding. They have been studied extensively in the literature on procurement, auctions and operations research management.[5] The literature focuses on a principal that needs to procure several goods in a single contract, but is uncertain about the exact quantities required at the time

[5]See for an overview Cattell *et al.* (2007) and Renes (2011).

of the procurement procedure.[6] In the principal's first-best all goods are acquired from the cheapest firm at zero profit for the firm. Unfortunately, if the expected quantities are slightly misestimated by the principal, the firm can create a profitable joint deviation. By asking more for the goods that are under weighted in the score rule and less for the over weighted goods, the bidder can increase expected payment while keeping his score constant. For risk neutral bidders the optimal bid contains infinite prices, yielding unbounded profit and risk. Renes (2011) studies mechanisms to prevent skewed bidding but finds no general solution when the principal is committed to accepting the bid with the lowest score. He notes that legal rules in the US allow the government to reject unbalanced bids, creating a solution to the problem through prohibitions. Ewerhart and Fieseler (2003) study the optimal score rule under uni-dimensional firm heterogeneity. Using the restriction that unit prices have to be weakly positive, and thus prohibiting a large part of the choice space of bidders, they recoup a version of the revelation principle and are able to determine a second-best allocation. Both solutions are the logical equivalent to the principle of taxation applied to procurement. In both cases, a large set of joint deviations or skewed bids are punished by giving them zero expected profit, and are thus effectively prohibited.

In addition, the principle of taxation has been successfully applied in the literature on multi-product monopoly pricing (see e.g. Armstrong, 1996). In that literature, an application of the principle of taxation simply requires a monopolist to bundle its goods. In this way, the monopolist disallows joint deviations by not allowing costumers to buy more of one product, and less of another. Clearly, in this setting the principle of taxation elicits a very realistic mechanism that is likely available to monopolists. However, if bundling is not an option in a particular setting, the necessary and sufficient conditions for implementability we derive in this paper may be useful for this literature as well.

The New Dynamic Public Finance generalizes the Mirrlees model to a setting where a uni-dimensional hidden characteristic follows a stochastic dynamic process.[7] Kocherlakota (2005) shows that in this model the optimal tax system generally contains intertemporal interdependencies, where the tax rate on labor income depends on the entire history of labor income. Renes and Zoutman (2013a) show that one of the reasons behind these interdependencies may be the multi-dimensionality of the type space. However, an additional reason for interdependencies may be the prohibition of joint deviations. It is shown in for instance Albanesi and Sleet (2006) that excessive savings choices should be

[6]For simplicity we focus on procurement auctions in this review. However, these are simply reverse auctions and all issues encountered in procurement are also encountered in sale auctions. See Athey and Levin (2001) for an example of joint deviations in an sale auction.

[7]See Golosov *et al.* (2007) and Kocherlakota (2010) for an extensive overview of the literature.

prohibited by, for example, setting a borrowing limit. This ensures the joint deviation of first saving too much and then working too little is not optimal. Interdependencies in the tax system may achieve a similar goal by, for example, taxing individuals that save too much, and work too little at a prohibitive rate.

7.3 The Model

In this paper we study how a social planner can implement its second-best allocation through the tax system. In this section we lay down the formal structure of our model. Before we get to the design of the optimal tax system we need to set up the problem. First, we define the preferences of the agents in the economy. Second, we define the conditions to which a second-best allocation should adhere. Subsequently, we show the agent's maximization problem on the market. The final subsection discusses the two approaches to implementation that are currently used in the literature: the rule-based principle of taxation and the canonical approach due to Mirrlees (1976).

7.3.1 Preferences

The economy is populated by a unit mass of individuals that are characterized by a twice-differentiable utility function:[8]

$$u\left(\mathbf{x},y,\mathbf{n}\right) \tag{7.1}$$

Where $\mathbf{x} \in \mathbf{X} \subseteq \mathcal{R}^k$ denotes a vector of choice variables, $y \in Y \subseteq \mathcal{R}$ a numeraire choice variable, and $\mathbf{n} \in \mathbf{N} \subseteq \mathcal{R}^p$ denotes the type of an individual. Variables in $\mathbf{x}$ may include e.g. effective labor supply, consumption of housing or savings. Choice variables $\mathbf{x}$ and y are observable at the individual level, and the social planner can tax all choices in $\mathbf{x}$ non-linearly, but cannot tax y. In principle, the choice of the numeraire variable has no effect on the optimal allocation. However, we assume it is a normal good such that $u_y > 0, u_{yy} \leq 0$ for any value of $\{\mathbf{x}, y, \mathbf{n}\}$. This implies the utility function is non-satiated everywhere. The assumptions on the numeraire ease the interpretation in the remainder of the paper. Throughout the paper we will sometimes refer to the choice variables in $\{\mathbf{x}, y\}$ as goods, even though they can be both inputs and outputs to the production process.

[8]Note that the description of agent's preferences closely follows the description in Renes and Zoutman (2013a).

Each element in type $\mathbf{n}$ is referred to as a characteristic. Characteristics in $\mathbf{n}$ may include variables such as ability, health status and taste parameters. For technical convenience we assume $\mathbf{N}$ is an open convex set and the space $\mathbf{N}$ will be referred to as the type space. Let $\mathbf{n}$ follow a multi-dimensional differentiable cumulative distribution function $F(\mathbf{n})$, with $F : \mathbf{N} \to [0,1]$ and probability density $f(\mathbf{n})$ both defined over the closure of $\mathbf{N}$. We assume that each characteristic denotes some independent aspect of the individuals, such that no characteristic can be found as a deterministic function of the other characteristics. The type is private information to each individual and unobservable to the government. Note that we do not restrict ourselves to static models, different choices can occur in different periods. However, we do assume that both the type and the direct mechanism are revealed to the individuals in the first period.[9]

Preferences can be summarized by the marginal rate of substitution:

$$\mathbf{s}(\mathbf{x},y,\mathbf{n}) \equiv -\frac{u_{\mathbf{x}}(\mathbf{x},y,\mathbf{n})}{u_y(\mathbf{x},y,\mathbf{n})}.$$

Element s_i is the marginal rate of substitution for choice variable x_i with respect to the numeraire y. Therefore, s_i represents the marginal utility loss of receiving an extra unit of x_i, expressed in units of the numeraire variable y.

7.3.2 Incentive Compatibility and Feasibility

The second-best allocation is assumed to have been derived through a direct mechanism. Let the second-best allocation of goods be denoted by:

$$\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\} \quad \forall \quad \mathbf{n} \in \mathbf{N}.$$

Here, the function $\mathbf{x}^*$ maps from the type space to the good space, $\mathbf{x}^* : \mathbf{N} \to \mathbf{X}$ and y^* maps from the type space to the numeraire good space, $y^* : \mathbf{N} \to Y$. We assume $\mathbf{x}^*(\cdot)$ and $y^*(\cdot)$ are both twice differentiable in all their arguments. Further, let $\mathbf{X}^*$ denote the image or range of function $\mathbf{x}^*$, and Y^* the image of y^*, and $\{\mathbf{X}, Y\}^*$ be the image of $\{\mathbf{x}^*, y^*\}$. $\{\mathbf{X}, Y\}^*$ Contains all the bundles that are assigned to a type in the economy. By definition, the set of assigned bundles is a subset of the total goods space, $\mathbf{X}^* \subseteq \mathbf{X}$,

[9] Note that the conventional utility representation(see e.g. Mirrlees, 1971, Saez, 2001) $\tilde{u}(y,l)$ where l is labor supply is a special case of our utility representation. Assume, as is standard that gross income $x_1 = n_1 l$ where n_1 is earnings ability. It can readily be seen that this utility function can be rewritten into our form: $\tilde{u}(y,l) = \tilde{u}\left(y, \frac{x_1}{n_1}\right) = u(x_1, y, n_1)$.

$Y^* \subseteq Y$. The sets $\mathbf{X}^*$ and $\mathbf{X}$, (Y^* and Y) are equal if each possible combination of choice variables is assigned to a type in the economy.

We will use the direct mechanism to elicit some properties about the second-best allocation. First, the economy should be able to produce all goods in the economy. We assume the economy's resource constraint takes the form:

$$\int_{\mathbf{N}} y^*(\mathbf{n})\, dF(\mathbf{n}) + R = \int_{\mathbf{N}} q(\mathbf{x}^*(\mathbf{n})) dF(\mathbf{n}) \tag{7.2}$$

In this equation, R is the exogenous revenue requirement of the government and $q : \mathbf{X} \to Y$ is a function that describes the economy's production of y. The equation states that total production of y should equal the sum of consumption of the numeraire and exogenous government expenditure. Derivatives q_{x_i} may be positive or negative depending on whether x_i is an input or an output of the production process. We assume weakly decreasing returns to scale such that all $q_{x_i x_i}$ are non-positive. An allocation is feasible if it satisfies condition (7.2).

A second-best allocation also has to be incentive compatible. In the direct mechanism each agent can send a p-dimensional message about his type, $\mathbf{m} \in \mathbf{N}$, to the planner. On the basis of this message the planner assigns the agent the bundle $\{\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m})\}$. Individuals send a message $\mathbf{m} \in \mathbf{N}$, and will choose the message that maximizes their utility. An allocation $\{\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m})\}$ is incentive compatible if each individual truthfully reveals all his unobserved characteristics and receives the bundle designed for him. That is, if the agent maximizes his utility by sending the message $\mathbf{m} = \mathbf{n}$. Therefore, an incentive-compatible and feasible allocation can be defined as follows:

Definition 11 *An allocation* $\{\mathbf{x} = \mathbf{x}^*(\mathbf{n}), y = y^*(\mathbf{n})\}$ $\forall \mathbf{n} \in \mathbf{N}$ *is incentive compatible and feasible if each agent truthfully reveals his entire type in a direct mechanism:*

$$\mathbf{n} = \arg\max_{\mathbf{m}} u(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n}) \quad \forall \mathbf{n} \in \mathbf{N}. \tag{7.3}$$

and in addition it satisfies equation (7.2).

Characterizing conditions under which an allocation is incentive-compatible if agents are heterogeneous in multiple dimensions may be very complex. Rochet and Choné (1998) discuss general conditions of an incentive-compatible allocation and Mirrlees (1976), McAfee and McMillan (1988) and Renes and Zoutman (2013a) describe this problem in the context of a first-order approach where second-order conditions of maximization problem (7.3) have to be checked after the entire allocation is derived. However, in this paper we are not interested in deriving the second-best allocation. Instead we turn our attention to how a second-best allocation can be implemented in the market. As such, we can leave aside these complex issues and simply note that an incentive compatible and feasible allocation is one that satisfies definition 11.

Finally, we know that a second-best allocation must satisfy the objective function of the planner subject to the conditions given in definition 11. However, at this moment we do not make any assumption about the objective function of the planner. Therefore, all we know about the second-best allocation is that it is feasible and incentive-compatible.

The optimal distortion in the economy can be characterized by the wedges each agent faces on the allocation:

$$\mathcal{W}_i(\mathbf{n}) = q_{x_i}(\mathbf{x}^*(\mathbf{n})) - s_i(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}). \tag{7.4}$$

These wedges represent the difference between the marginal rate of substitution of the agent and the marginal rate of transformation (or production price) on the allocation. If the wedge is positive a good in the allocation is distorted below its Laissez-Faire value (i.e taxed). It is subsidized when the distortion is negative.

7.3.3 Market Implementation

We aim to find the properties of a tax system that implements the second-best allocation in the market. Therefore, we have to go beyond the direct mechanism that identifies the second-best, and study the choice problem of agents in a market. Agents maximize their utility function (7.1) with respect to their choice variables $\mathbf{x}$,y subject to their budget constraint in the market:

$$y \leq q(\mathbf{x}) - T(\mathbf{x}), \tag{7.5}$$

where the tax system or function T maps from the good space to the numeraire, $T : \mathbf{X} \rightarrow Y$. How much a consumer can spend on y depends on his choice of $\mathbf{x}$, the production function $q(\cdot)$ and the tax system, $T(\cdot)$.

A tax system implements an allocation if each agent weakly prefers his bundle over all other combinations of goods available to him in the market. This concept is formally defined in definition 12:

Definition 12 *A tax system implements an allocation* $\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\}$ *if each agent selects the bundle on the market that was assigned to him in the second-best allocation:*

$$\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\} = \arg\max_{\mathbf{x},y} \{u(\mathbf{x}, y, \mathbf{n}) : y = q(\mathbf{x}) - T(\mathbf{x}), \mathbf{x} \in \mathbf{X}, y \in Y\} \quad \forall \quad \mathbf{n} \in \mathbf{N} \tag{7.6}$$

Note that we do not need to check whether the planner's budget constraint holds as long as the tax system is successful in implementing the second-best allocation. By definition the second-best allocation is resource feasible, and on the market the agent's budget constraints must hold with equality given non-satiation in y. Then by Walras' law the government budget constraint must also be satisfied.

The difficulty of implementability can be understood by comparing definitions 11 and 12. In the direct mechanism the agent maximizes his utility by sending the optimal p-dimensional message, containing all p characteristics of his type, to the planner. In doing so, he can choose his optimal bundle of all bundles in the set $\{\mathbf{X}, Y\}^*$. That is, the agents can receive a bundle by mimicking another type, but they cannot receive a bundle that has not been assigned to any type.

However, in the market the agent can choose his optimal bundle out of all points in the choice space $\{\mathbf{X}, Y\}$ within his budget constraint. That is, the market allows the agents to create new bundles that were not assigned to any type in the direct mechanism. Such a strategy is called a joint deviation, since in order to create a new bundle that satisfies the budget constraint an agent has to deviate in at least two goods. If a tax system allows for profitable joint deviations, it cannot satisfy definition 12, even if the allocation satisfies definition 11. Incentive-compatibility does not imply market implementability under all tax systems.

7.3.4 Pure Price Mechanisms

Conceptually, the difference between the problem of the agents in the direct mechanism, and the market is largest when $k > p$, the number of choice variables is larger than the number of hidden characteristics. Suppose for example, that the agents are couples that maximize joint utility by choosing labor income of both spouses x_1 and x_2 and let y denote their consumption. Suppose in addition, that earnings ability differs between, but not within couples. In the direct mechanism each couple is awarded a bundle according to their shared ability: $\{x_1^*(n), x_2^*(n), y^*(n)\}$. The direct mechanism allows each couple to mimic a different type by sending a message $m \neq n$ to the planner. However, it does not allow differentiation between spouses. Since the planner knows couples differ in only one dimension, it only accepts one message from each couple. As such, it is impossible to jointly deviate by sending a message n for the husband, and a message $m \neq n$ for the wife. On the other hand, the market does allow joint deviations and, depending on the tax function $T(x_1, x_2)$, may give them positive consumption levels if they make this choice.

This naturally brings us to the principle of taxation derived in Hammond (1979). The principle of taxation says that at least one tax system can implement the second-best allocation. This tax system has two properties. First, if the agent chooses the bundle $\mathbf{x}^*(\mathbf{n})$ designed for him, he will receive the corresponding value of $y^*(\mathbf{n})$. That is, the tax function satisfies:

$$T(\mathbf{x}^*(\mathbf{n})) = q(\mathbf{x}^*(\mathbf{n})) - y^*(\mathbf{n}) \qquad \forall \quad \mathbf{n} \in \mathbf{N}.$$

Second, it restricts agents to making a choice within $\{\mathbf{X}, Y\}^*$. This implementation effectively disallows all joint deviations. It follows immediately that problems 11 and 12 are isomorphic and hence the outcome is identical.

The mechanism with which the tax system disallows joint deviations is not explicitly described. It is easy to think of a way in which the planner can deter agents from making joint deviations. The planner could tax joint deviations at an infinite rate, or forbid joint deviations explicitly, giving severe punishment to trespassers. It is clear that such a tax system can deter agents from making joint deviations. However, this mechanism can have very undesirable properties. It is for example sensitive to a trembling hand. If in the example, one spouse works slightly more than he should whereas his partner works exactly the right amount of time, this would be considered a joint deviation. This 'mistake' causes the planner to give the couple an infinite amount of taxes or receive a severe penalty.

Mirrlees (1976) proposes a market-based approach. He derives the optimal second-best allocation with multiple goods under uni-dimensional heterogeneity in earnings ability of the agents. The outcome of this maximization defines the wedges, $\mathcal{W}_i(n)$, for each level of earnings ability, n, and each good x_i. Under certain regularity conditions[10] each good can be used to infer n. That is, there exists an inverse function of the allocation $\mathbf{x}^*(n)$, such that $(\mathbf{x}^*)_i^{-1}(x_i) = n$ for each good x_i. Intuitively, in the example of the couples, one can derive the earnings ability of each couple by observing either the husband's labor

[10] Non-satiation of the utility function, the Spence-Mirrlees condition on preferences and a monotonicity condition on the allocation.

income, or the wife's labor income, in an incentive compatible allocation. Mirrlees (1976) then proposes to implement the second-best allocation through a tax system that has the following properties. First, each agent should be able to afford his bundle:

$$T\left(\mathbf{x}^*\left(n\right)\right) = q\left(\mathbf{x}^*\left(n\right)\right) - y^*\left(n\right) \qquad \forall \quad n \in N,$$

as in the principle of taxation However, unlike in the principle of taxation, it does not limit the choice set to $\{\mathbf{X}, Y\}^*$. Instead, it sets the marginal tax on each good x_i equal to the wedge:

$$T_i'\left(x_i\right) = \mathcal{W}_i((\mathbf{x}^*)^{-1}\left(x_i\right)).$$

In the literature this approach is known simply as equating the tax to the wedge (see e.g. Kocherlakota, 2005). Note that the tax system described does not limit the choice of the agent in any way. That is, under this tax system the agent can choose joint deviations if he so desires. Also note that the tax system is unique if all $x_i \in \mathbf{X}$ are awarded to a type in the direct mechanism, because in that case a marginal tax rate is assigned to each level of the choice variable. Finally, unlike the tax system prescribed by the principle of taxation, there is no easy proof to show that this tax system can implement all allocations. This, of course, is the object of our study.

The biggest difference between the principle of taxation type of implementation and the price mechanism of Mirrlees is the off-allocation behavior of the tax system.[11] The fundamental property of the canonical implementation by Mirrlees is that the government only influences behavior through relative prices. In contrast, the principle of taxation disallows all choice that are made off the allocation. In the remainder of this paper we draw a distinction between a pure price mechanism, such as the one designed by Mirrlees, and a rule-based mechanism such as the principle of taxation, and look for allocations that can be implemented through a price mechanism alone. However, rules can be arbitrarily replaced by prohibitive prices. Therefore, we need to limit the definition of a pure price mechanism, such that such excessive taxes are not allowed. This is done in definition 13. The definition limits the marginal tax rate in a pure price mechanism such that it never exceeds the maximum wedge and is never smaller than the minimum wedge found on the allocation. This definition excludes the possibility of using rule-equivalent taxes, but at

[11] The problem of how to define the tax rate off the allocation has been asked elsewhere. Mirrlees (1976) suggested making the tax system separable. This leads to a unique implementation as in figure 1. Similarly, Hellwig (2007) focuses on "canonical" tax systems, which he defines as being continuous at end-points of the type-space. Although not necessarily unique, this does narrow down the class of tax systems.

the same incorporates most price mechanisms proposed in the literature, among them the one proposed by Mirrlees.

Definition 13 *A tax system $T : \mathbf{X} \to Y$ is a pure price mechanism if it satisfies the following condition:*

If a choice $\mathbf{x} \in \mathbf{X}$ is outside the image of the allocation $\mathbf{X}^*$, the marginal tax rate on x_i should satisfy:

$$\min \{\mathcal{W}_i(\mathbf{n}) : \mathbf{n} \in \mathbf{N}\} \leq T_i'(\mathbf{x}) \leq \max \{\mathcal{W}_i(\mathbf{n}) : \mathbf{n} \in \mathbf{N}\}$$

. .

7.4 Failure of the Price Mechanism: A Simple Example

In this section we show graphically how the pure price mechanism may fail.[12] The purpose of this example is to show in the simplest possible setting that a pure price mechanism cannot always achieve the second-best allocation. For that purpose the model is highly stylized and very much simplified. The idea is specifically not to give a realistic example, but to show the reader that even in very simple settings a pure price mechanism may fail. The example as depicted 7.1 and 7.2 give an example where a pure price mechanism does not achieve the desired result. The figures describe a situation with two goods in $\mathbf{x}$ and one exogenous characteristic, $k = 2$ and $p = 1$. In this example we again fall back on the couples which we now assume maximize a perfectly symmetric joint utility function. For simplicity we take the case with uni-dimensional heterogeneity such that each spouse in the couple has exactly the same earnings ability level. The spouses have to decide how much time each of the partners works and how much each of them tends to the household and children. The optimal allocation specifies how much labor income is generated by each spouse, $x_1^*(n)$ and $x_2^*(n)$, and how much the couple consumes, $y^*(n)$, as a function of ability. The production function that we have used in this example is $x_i = nl_i$, where l_i represents the labor effort of spouse i. Clearly, highly able couples have to provide less labor to reach a certain income level than lowly able couples. Since there is only one hidden characteristic, the bundles assigned to the types by this allocation form a line in $X_1 \times X_2 \times Y$ space. This line is represented by the black line in figures 7.1. The line is sloping upward, indicating that the government wants couples with higher ability to work

[12]The mathematics behind this example can be found in the appendix.

more. In addition, an increase in gross income $x_1 + x_2$ leads to a less than one-to-one increase in consumption for the couple indicating that the government is redistributing from higher to lower ability couples. Finally, on any point of the line $x_1 = x_2$ indicating that the government wants each spouse in the couple to supply the same amount of labor effort. The dots represent the bundle of one particular couple.

The hyper-plane shows the budget constraint of individuals in the unique Mirrleesian implementation where taxes are equated to wedges and marginal tax rates are separable. Each point on the surface represents a combination of labor income of the husband x_1, labor income of the wife, x_2 and the amount of consumption goods they can buy after taxes, y.

In figure 7.2 the vertical axis shows the utility level at each point of the budget constraint for one particular couple that is about in the middle of the ability distribution. The surface represents the utility function of the couple, with the assigned bundle at the dot, for all combinations $\{x_1, x_2, y\}$ that satisfy their budget constraint with equality given the price mechanism imposed. In figure 7.2 we can see that the assigned bundle (dot) marks the highest utility level on the allocation (line), such that the couple prefers their bundle over any of the other bundles in the allocation. The allocation is therefore incentive compatible for this couple in the direct mechanism, since mimicking another couples labor income would decrease their utility level. However, in the market, the couple are also allowed joint deviations where the husband works more hours than the wife or vice versa. In figure 7.2 there are such joint deviations that give this couple more utility than their assigned amount of labor effort. Therefore, the pure price mechanism fails to implement the allocation.

In this particular example, the failure results from the fact that our fictitious couple prefers specialization, where one of the partners does all the work and the other partner stays at home, over an allocation of tasks where both partners share the burden of working on the labor market equally. Such a preference may be the result of increasing returns or increasing utility to specialization. One may therefore conclude that the price mechanism could fail due to specific agents preferences. However, note another particular feature of the optimal allocation. Even though, the couple likes to specialize, the government prefers each partner to work the same number of hours. There is thus a misalignment between the preferences of the government and the preferences of the agents. Such a misalignment may have several causes. First, the government may be paternalistic, simply forcing households to equally divide the task between both partners for no reason other than the fact that government thinks this is right. Second, the couples in the economy might have children. The children in turn may be better off with attention of both spouses than with attention

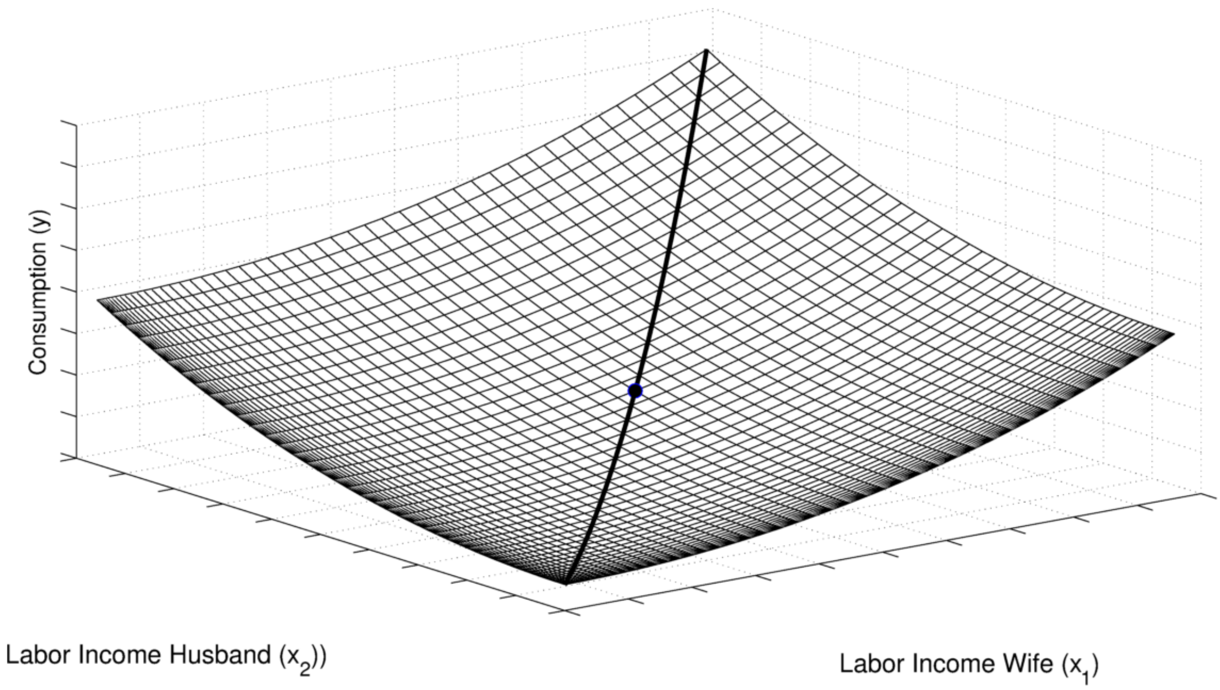


Figure 7.1: An Optimal Allocation and a Budget Constraint.

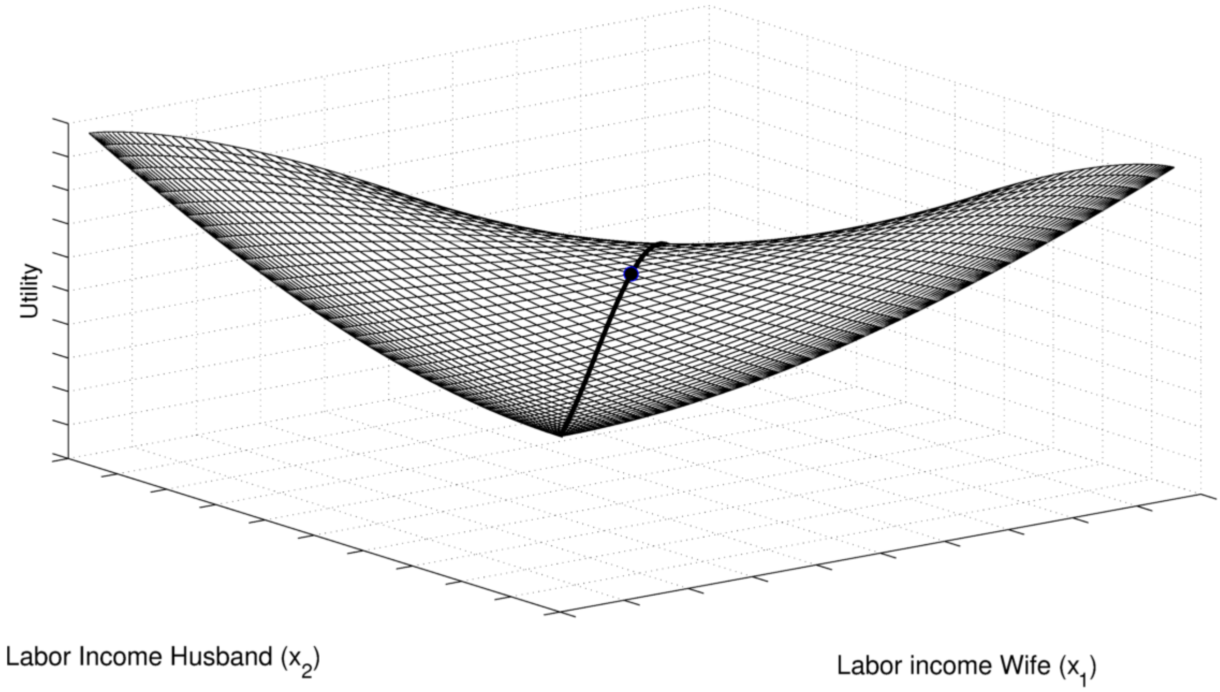


Figure 7.2: Utility of the Couple

of only the stay-at-home spouse. If the utility of the child is not sufficiently weighted in the utility function of the couple, the government may want to correct the externality and make both spouses stay at home part of the time. Note that both of these arguments can lead to exactly the same first-order optimality conditions for the second-best allocation. This example indicates that non-welfarist motives such as paternalism or correction of externalities may lead to a failure of the price mechanism, independent of the preferences of the agents. We proof rigorously that a price mechanism always suffices provided that no such misalignment exists in proposition 14.

For future reference note that the utility function of the agents could only take on one of two forms around the bundle that was allocated to them. Because taxes are equal to wedges, the bundle must be a stationary point in the allocation. Further, because the allocation is assumed to be incentive-compatible it must be a maximum in at least one direction, namely along the allocation. It then follows that it must either be a maximum or a saddle point. This point is crucial to understanding the proof for proposition 14.

The fact that the pure price mechanism cannot implement the allocation does not mean that the allocation cannot be implemented at all. Theoretically, the principle of taxation applies, so the government can implement its second-best by simply disallowing the couples to specialize. However, such an approach might not be desirable for reasons already discussed.f

7.5 Implementation through Pure Price Mechanisms

In this section we first derive the necessary and sufficient conditions for a pure price mechanism to implement the second-best allocation. These conditions can be a useful test to verify whether a specific pure price mechanism implements a specific allocation. However, in order to perform this test, one first needs to derive the entire allocation and the tax system. In many cases an explicit solution for the second-best allocation does not exist. Numerical solutions are available, but these describe only special cases by definition. The implementation can then only be checked on the specific tax system and parametrization studied. Therefore, these explicit solutions cannot be used to say anything about tax systems in general. To overcome this problem we therefore continue by identifying two classes of problems where the pure price mechanism is always effective in implementing the second-best. We describe the characteristics of these classes in the second and third subsection.

7.5.1 Conditions for Implementation

Lemma 4 derives the general conditions under which a pure price mechanism implements an allocation, by formally solving the problem of definition 12.

Lemma 4 *An incentive compatible and feasible allocation can be implemented through a twice differentiable tax system $T(\mathbf{x})$ iff a.e.:*
i.)

$$y^*(\mathbf{n}) = q(\mathbf{x}^*(\mathbf{n})) - T(\mathbf{x}^*(\mathbf{n})), \tag{7.7}$$

ii.)

$$T_i'(\mathbf{x}^*(\mathbf{n})) = \mathcal{W}_i(\mathbf{n}), \tag{7.8}$$

iii.)

$$-\frac{\partial \mathbf{s}(\mathbf{x}, y(u, \mathbf{x}, \mathbf{n}), \mathbf{n})}{\partial \mathbf{x}} + q''(\mathbf{x}^*(\mathbf{n})) - T''(\mathbf{x}^*(\mathbf{n})) \lessdot 0. \tag{7.9}$$

Proof. The proof can be found in the appendix. ■

Equation (7.7) ensures that the amount of taxes paid for any bundle of $\mathbf{x}^*(\mathbf{n})$ within the allocation is uniquely determined. If the total tax level $T(\mathbf{x}^*(\mathbf{n}))$ is too high, the tax schedule cannot implement the allocation because people receive too little $y^*(\mathbf{n})$ if they choose their assigned quantities $\mathbf{x}$, and vice versa. Equation (7.8) is the first-order condition for a market implementation. It states that marginal taxes are equated to marginal wedges. There are always as many marginal tax rates in T' as there are goods in $\mathbf{X}$, for all $\mathbf{n} \in \mathbf{N}$. Such that there is always a unique vector of marginal tax rates $T'(\mathbf{x}^*(\mathbf{n}))$ that satisfies (7.8) within any possible incentive compatible allocation. In effect, this means that the first order conditions of this problem can always be met and that the solution is unique on the allocation, but undefined off the allocation. In our example in figures 1 and 2, this translates to a tax system that is fully defined on the line, but undefined everywhere else.

Equation (7.9) states that the indifference curve of any linear combination of $\mathbf{x}$'s with respect to y should be more convex than the budget constraint for the same linear combination of $\mathbf{x}$. This condition is different from the standard second-order condition of utility maximization with two goods (see e.g. Mas-Collel *et al.*, 1995) in two ways. First, in standard micro-economic theory the budget constraint is linear and hence if the indifference curve is convex, it is automatically more convex than the budget constraint. Second, since there are multiple choices, sufficiency requires that the indifference curve of all linear combinations of $\mathbf{x}$ with respect to y are more convex than the budget constraint.

Since the conditions derived are both necessary and sufficient, they can be used to verify whether or not a specific tax system implements an allocation, after both the allocation and the tax schedule have been determined. However, the result is only applicable to allocations that have a closed-form solution. Therefore, in the next two sections we provide our main results, and characterize two situations in which we can guarantee that a pure price mechanism can implement the allocation.

7.5.2 Pure Price Mechanisms: Second-Best of a Welfarist Planner

As we have seen in figure 7.2 a pure price mechanism may sometimes place agents on utility saddle points, causing them to deviate from the desired allocation if given the possibility. In the next proposition we show that if there are no externalities an allocation that places individuals in saddle points, allows Pareto improvements. Therefore, they cannot be part of a second-best allocation for a welfarist planner.

Proposition 14 *If an allocation maximizes a welfare function* $SW = \int W(u)\, dF(\mathbf{n})$*, subject to the incentive compatibility constraints and the resource constraint, and* $W' \geq 0$ *, then any pure price mechanism can implement the allocation, provided taxes are equated to wedges on the allocation,* $T_i'(\mathbf{x}^*(\mathbf{n})) = \mathcal{W}_i(\mathbf{n})$.

Proof. The proof can be found in the appendix. ■

Intuitively, any tax schedule that does not satisfy (7.9) allows for at least one deviation that increases the utility of at least one agent. In addition, since the first-order conditions (7.8) combined with the violation of (7.9) imply that the agent is located in a saddle point, the exact opposite deviation must increase his utility by approximately the same amount. This can easily be seen in figure 7.2. The agents' utility, by approximation, increases as much if he moves to the right off the allocation as when he moves to the left. Provided tax revenue is not maximized in the allocation, tax revenue must weakly increase either for the deviation to the right, or for the opposite deviation to the left. In figure 7.1 the tax schedule is monotonic and hence such a deviation exists. Therefore, there is a deviation which increases the utility of the agent, and hence increases welfare, and weakly increases the tax revenue of the planner. Such a deviation must be a Pareto improvement. Since such a deviation is possible in any allocation that places individuals in a saddle-point, any allocation containing saddle-points (like the allocation in figure 7.1) can not be second-best for a welfarist social planner. Since incentive compatibility guarantees that the allocation cannot be minimum in individual utility and a Pareto argument rules out saddle-points, the second-best of a welfarist social planner has to form a maximum in the agent's problem. This second-best must therefore be implementable by a tax system satisfying equations (7.7) and (7.8). Since in this case the individuals do not wish to double deviate, the planner can set off allocation taxes that satisfy definition 13 and a pure price tax system can implement the second-best.

This proof breaks down in the presence of externalities, internalities or a non-welfarist planner. With externalities and internalities the deviation of any agent can influence the utility of other agents, such that it is unclear when a deviation from the saddle-point entails a Pareto- mprovement. This implies that implementability has to be checked through lemma 4 in this case, since the goals of the individuals and the planner might differ too much for implementation through prices.

In practice, the restriction that a tax schedule does not contain an internal maximum in revenue is rather weak. The sign of the marginal tax rate is equal to the sign of the wedge. In most models of optimal taxation the optimal wedge does not change sign such that the resulting tax system must be monotonic. A monotonic tax system does not have an internal maximum. Even in models such as Saez (2002b) and Choné and Laroque (2010) where the optimal wedge does change sign, it changes signs from negative to positive. As such, the resulting tax system has an internal minimum, but not an internal maximum. We are not aware of any articles where the optimal tax system does have an internal maximum in revenue.

Mirrleesian Implementation

A simple corollary shows that the Mirrleesian implementation can implement the second-best under uni-dimensional heterogeneity, provided there are no externalities and the second-best allocation is optimal to a welfarist planner.

Corollary 15 *If $p = 1$ and the conditions of proposition 14 are met, the Mirrleesian implementation can implement the second-best.*

Proof. The proof follows from the fact that the Mirrleesian planner is welfarist and equates off-allocation wedges to on allocation wedges, such that definition 13 and proposition 14 are satisfied. ■

It follows that the use of the Mirrleesian implementation in e.g. Mirrlees (1976), Atkinson and Stiglitz (1976), Bovenberg and Jacobs (2005) and Golosov *et al.* (2013) was indeed correct. Despite the fact that none of these papers provide a formal proof for the fact that their proposed tax system can implement the second-best allocation they have derived, our result shows ex-post that it can because each of this articles assumes the planner is welfarist and there are no externalities.

7.5.3 Pure Price Mechanisms: A Bijective Allocation

The combination of (7.7) and (7.8) defines the tax schedule on the allocation. If the allocation perfectly covers the choice space this pure price tax-schedule must implement the allocation. Proposition 16 provides a sufficient condition for such a unique tax implementation to exist.

Proposition 16 *If the mapping* $\mathbf{x}^*(\mathbf{n})$ *is bijective, then the tax implementation described by equations (7.7) and (7.8) is the unique differentiable tax-schedule that implements the second-best allocation.*

Proof. proof in appendix ■

Note that bijectiveness of the mapping $\mathbf{x}^*(\mathbf{n})$ is a rather strict requirement. It requires $\mathbf{x}$ and $\mathbf{n}$ to be of the same dimension and both be similarly (un)bounded, such that incentive compatibility and implementability coincide. In this situation every choice in the market corresponds to the choice of a unique type in the direct mechanism, and every type in the direct mechanism to a unique bundle on the market. Since all types prefer their own bundle over the bundles assigned to other types and all bundles are assigned to a type, it follows that all types prefer their bundle above any other bundle in the budget set.

The allocation derived in Mirrlees (1971) is an example of a bijective allocation, provided the ability distribution is unbounded. In the direct mechanism all ability types are assigned a specific gross income level x. Mirrlees shows that if ability is continuously distributed in $\mathcal{R}_+$, the second-best allocation assigns all gross income levels to a specific ability type without bunching. Hence, the function $x^*(n)$, mapping ability to gross income, is bijective. Then by definition incentive compatibility and implementability coincide. Thus, incentive compatibility is enough to ensure double deviations are not profitable even in case of multiple dimensions of heterogeneity.

7.6 Concluding Remarks

The results presented in this paper cover the understudied second step of tax design for a general class of models. Propositions 14 and 16 show that a relatively simple tax system, a (separable) pure price tax system, can implement the allocation found in most problems studied in the existing literature. Lemma 4 gives the conditions that need to be checked if the problem does not fit one of the two classes. The results are, however, not restricted to the uni-dimensional heterogeneity that has been the main focus of the literature. All of our results are directly applicable to the multi-dimensional problem studied in the companion paper (Renes and Zoutman, 2013a).

Proposition 14 highlights a unique feature of the Mirrleesian optimal tax model. Unlike the design problem of auctioneers and monopolists, the maximization of the central planner is quite closely aligned with that of the agents he faces. In fields such as monopoly pricing and auction theory the objectives of the principal and the agents are opposed. An increase in a monopolist's profits (at fixed quantities) automatically comes at the expense of the consumers. As such, implementation will generally not be possible through a pure price mechanism, consequently bunching, prohibitions or more generally restrictions in the choice space will be quite prevalent (see also Armstrong, 1996, Renes, 2011, Rochet and Choné, 1998).

The alignment between agents and planner means a relatively broad class of tax systems implements the second-best of a welfarist planner. This implies the planner can let the agents maximize their utility with relatively few restrictions, irrespective of the actual utility function of individuals. This alignment also has interesting effects on the restrictions that are required for implementation. Necessary restrictions will often require interdependence in tax rates, increasing the complexity of the tax system. These interdependencies, like wealth tests on income assistance in welfare states, are necessary to prevent rational people from taking advantage of subsidies that are not targeted at them.

Future work could focus on the possibility to extend this work to dynamically stochastic settings and on finding more tight descriptions of the classes of models where violations of the second order implementation conditions do not occur.

Chapter 8

Conclusion

Trends such as globalization, and skill-biased technological growth affect the global economy in a very positive way. The average employee has seen stellar growth in his productivity. Products and services such as information sharing through the internet, connectivity through smartphones, access to media content all around the world at very low prices, access to a wide supply of products from different continents, and regular out-of-continent holidays are taken for granted by a large part of the population. All of these possibilities have become available to us during my life-time.

The positive effects of these developments were perhaps felt even stronger in some developing nations. In 1990 43 percent of the people in developing nations lived in extreme poverty. In 2010 this proportion was down to 21 percent. Most of the reduction in poverty occurred in East-Asia. However, in recent years even countries in Sub-Saharan Africa have found their way to growth. Characteristic for our current day and age, a lot of the growth in Africa is due to the introduction of the mobile phone, which allowed for connectivity between people in countries where the quality of the traditional infrastructure is weak at best.

The most important downside of these developments is the massive increase in inequality in developed countries. The contrast between rich and poor has taken stellar proportions, with the middle-class almost disappearing in some countries. Although the Great Recession has affected the income of both the rich and the poor, it served to highlight the weak position of those at the bottom of the labor market. It is this group which has faced higher probabilities of unemployment, loss in purchasing power, and more restrictive access to credit.

Looking towards their leaders, political parties in Europe present their constituents with two choices. Either, we accept the increase in inequality with all of its negative effects, and accept that global developments have made it impossible to create a society with economic justice. This view is presented in Europe by the political center. Or, as proposed by parties at the left and right fringe of the political spectrum, xenophobic measures aimed at closing off the economy from all global developments should be taken. The goal of these measures is to recreate a society that faced neither the costs, nor the benefits created by globalization and technological progress.

Are these the only options available to us? Do we have to accept either increasing economic inequality, or stark reductions in our economic potential? The answer given in this dissertation is a resounding 'no'. We have not explored all of our options. Many instruments, simply available to politicians in developed economies can reduce inequality, while simultaneously boosting economic growth.

We have seen that the Dutch government does not make efficient use of their labor-income tax in chapter two and three of this thesis. It could do better by redistributing less to middle incomes, and more to the poor. In the fourth chapter it is shown that monitoring of labor effort, and punishing those that shirk, in particular at the bottom of the income distribution, raises equity and lowers redistributive costs. The fifth chapter, provides evidence that capital income taxation is less distortive than previously assumed, suggesting that countries should intensify their reliance on this redistributive instrument. The final two chapters have made a first step towards analyzing redistribution in multiple dimensions. This analysis can help governments to properly align the full orchestra of their redistributive instruments. All the evidence presented in this dissertation shows that current governments can do much more to reduce inequality in their own countries, without closing it off to the rest of the world.

Clearly, future research is necessary. For instance, we do not know exactly how costly monitoring of labor effort would be. Governments could start pilot projects to measure the costs of this instrument. The analysis in chapter five shows that the cost of capital income taxation may be relatively low, but quantifying the benefits of taxing capital income is an arduous task, since the literature have provided us with so many different reasons to tax capital income. Finally, chapter six and seven set the first steps towards a better alignment of different redistributive instruments, but future research is necessary before this knowledge can be applied to the real world.

The new global developments require us to change the structure of our society. Perhaps, we will have to live with slightly more economic inequality after we have exhausted all redistributive instruments within our welfare state. But this moment has not arrived yet. We have not pushed our welfare state to its redistributive limits. Governments around the world can, and should do more, to create the redistributive symphony their constituents deserve.

Appendix A

Appendix to Chapter 2

A.1 Optimal Income Taxation with Intensive Margin Only

We will solve the optimal income tax using Lagrangian methods. Multiply the incentive constraint with θ_n and apply integration by parts to $\theta_n \frac{\mathrm{d}u_n}{\mathrm{d}n}$ so as to find:

$$\int_{\mathcal{N}} \left(-\theta_n \frac{z_n h'(z_n/n)}{n^2} - u_n \frac{\mathrm{d}\theta_n}{\mathrm{d}n} \right) \mathrm{d}n + \theta_{\overline{n}} u_{\overline{n}} - \theta_{\underline{n}} u_{\underline{n}} = 0. \tag{A.1}$$

Now, set up the optimal-tax problem as a Lagrangian with c_n, z_n, and u_n as control variables. We furthermore introduce λ as the Lagrange multiplier of the economy's resource constraint. $\eta_n f(n)$ denotes the composite Lagrange multiplier of the utility constraint at n (we have harmlessly pre-multiplied each multiplier η_n with $f(n)$ to avoid some additional notation). θ_n is the Lagrange multiplier of the incentive-compatibility constraint at n:[1]

$$\mathcal{L} \equiv \int_{\mathcal{N}} \left(W(u_n) + \lambda\left(z_n - c_n - R\right)\right) f(n)\mathrm{d}n + \int_{\mathcal{N}} \eta_n \left(v(c_n) - h(z_n/n) - u_n\right) f(n)\mathrm{d}n \tag{A.2}$$
$$- \int_{\mathcal{N}} \left(\theta_n \frac{z_n h'(z_n/n)}{n^2} + u_n \frac{\mathrm{d}\theta_n}{\mathrm{d}n} \right) \mathrm{d}n + \theta_{\overline{n}} u_{\overline{n}} - \theta_{\underline{n}} u_{\underline{n}}.$$

[1] We need the latter constraint because all variables in the utility function c and z as well as utility itself u are considered choice variables for the government in this optimization procedure. Alternatively, one may invert the utility function and write consumption as a function of the allocation, $c(z, u)$, which is usually done in the literature.

The first-order and transversality conditions for this control problem are given by:

$$\frac{\partial \mathcal{L}}{\partial c_n} = 0 \quad : \quad -\lambda f(n) + \eta_n f(n) v'(c_n) = 0, \quad \forall n, \tag{A.3}$$

$$\frac{\partial \mathcal{L}}{\partial z_n} = 0 \quad : \quad \lambda f(n) - \eta_n \frac{h'(l_n)}{n} - \theta_n \frac{h'(l_n) + l_n h''(l_n)}{n^2} = 0, \quad \forall n, \tag{A.4}$$

$$\frac{\partial \mathcal{L}}{\partial u_n} = 0 \quad : \quad W'(u_n) f(n) - \eta_n - \frac{\mathrm{d}\theta_n}{\mathrm{d}n} = 0, \quad \forall n \neq \underline{n}, \overline{n}, \tag{A.5}$$

$$\lim_{n \to \underline{n}} \theta_n = 0, \quad \lim_{n \to \overline{n}} \theta_n = 0. \tag{A.6}$$

We omitted restating the incentive-compatibility and resource constraints. We now derive the optimal tax formula as reported in Saez (2001).

First, solve (A.3) for η_n to find $\eta_n = \frac{\lambda}{v'(c_n)}$, and substitute this into (A.4) and simplify:

$$1 - \frac{h'(l_n)}{n v'(c_n)} = \frac{\theta_n \left(h'(l_n) + l_n h''(l_n) \right)}{\lambda f(n) n^2}. \tag{A.7}$$

Substitute the individuals' FOC (2.4) into (A.7) and simplify the resulting equation:

$$\frac{T'(z_n)}{1 - T'(z_n)} = \left(1 + \frac{l_n h''(l_n)}{h'(l_n)} \right) \frac{\theta_n v'(c_n)/\lambda}{(1 - F(n))} \frac{1 - F(n)}{f(n) n}. \tag{A.8}$$

For the utility function we used, the compensated and uncompensated labor supply elasticities are given by (see a later Appendix):

$$\varepsilon_n^c \equiv -\frac{\partial l_n}{\partial T'} \frac{1 - T'}{l_n} = \frac{v'}{\frac{l h'' v'}{h'} - \frac{l h' v''}{v'} + v' n l \frac{T''}{1 - T'}}, \tag{A.9}$$

$$\varepsilon_n^u \equiv \frac{\partial l_n}{\partial n} \frac{n}{l_n} = \frac{v' + \frac{l h' v''}{v'} - v' n l \frac{T''}{1 - T'}}{\frac{l h'' v'}{h'} - \frac{l h' v''}{v'} + v' n l \frac{T''}{1 - T'}}. \tag{A.10}$$

Therefore, we find

$$\frac{1 + \varepsilon_n^u}{\varepsilon_n^c} = 1 + \frac{l_n h''(l_n)}{h'(l_n)}. \tag{A.11}$$

In addition, integrating equation (A.5), and using a transversality condition, a solution for θ_n can be derived:

$$\theta_n = \int_n^{\overline{n}} \left(\frac{\lambda}{v'(c_m)} - W'(u_m) \right) f(m) \, \mathrm{d}m. \tag{A.12}$$

By introducing normalized social welfare weight $g_n \equiv \frac{W'(u_n) v'(c_n)}{\lambda}$, which denotes the monetized welfare gain of providing one euro to individual n, this expression can be simplified:

$$\theta_n = \lambda \int_n^{\overline{n}} \frac{(1-g_m)}{v'(c_m)} f(m)\,\mathrm{d}m. \tag{A.13}$$

Substituting results (A.11) and (A.13) into (A.8). The final constraints on θ are the transversality conditions. Note that the transversality conditions imply that the distortion on labor supply at the top and the bottom should equal zero.

A.2 Optimal Income Taxation with Both Intensive and Extensive Margins

The incentive-compatibility constraint is unaffected by introducing the extensive margin. We will solve the optimal income tax again using a Lagrangian, which uses c_n, z_n, and u_n as control variables. λ is the Lagrange multiplier of the economy's resource constraint, $\eta_n f(n)$ denotes the composite Lagrange multipliers on the utility constraint for each n, and θ_n is the Lagrange multiplier of the incentive-compatibility constraint at n. The Lagrangian can be written as:

$$\begin{aligned}\mathcal{L} \;=\; & \int_{\mathcal{N}} \left(\int_{\underline{\varphi}^n}^{u_n - v(b)} W(u_n-\varphi)k(\varphi|n)\mathrm{d}\varphi f(n) + W\left(v(b)\right)\left(f(n) - K(u_n - v(b)|n)f(n)\right) \right) \mathrm{d}n \\ & + \int_{\mathcal{N}} \lambda \left((z_n - c_n)K(u_n - v(b)|n)f(n) - (f(n) - K(u_n-v(b)|n)f(n))b) - Rf(n) \right) \mathrm{d}n \\ & + \int_{\mathcal{N}} \eta_n \left(v(c_n) - h(z_n/n) - u_n \right) f(n)\mathrm{d}n \\ & - \int_{\mathcal{N}} \left(\theta_n \frac{z_n h'(z_n/n)}{n^2} + u_n \frac{\mathrm{d}\theta_n}{\mathrm{d}n} \right) \mathrm{d}n + \theta_{\overline{n}} u_{\overline{n}} - \theta_{\underline{n}} u_{\underline{n}}. \end{aligned} \tag{A.14}$$

where we substituted the definition for $\tilde{k}(n) \equiv K(u_n - v(b)|n)f(n)$. The first-order and transversality conditions for this control problem are given by:

$$\frac{\partial \mathcal{L}}{\partial c_n} = 0 \ : \ -\lambda \tilde{k}(n) + \eta_n v'(c_n) f(n) = 0, \quad \forall n, \tag{A.15}$$

$$\frac{\partial \mathcal{L}}{\partial z_n} = 0 \ : \ \lambda \tilde{k}(n) - \eta_n \frac{h'(l_n)}{n} - \theta_n \frac{h'(l_n) + l_n h''(l_n)}{n^2} = 0, \quad \forall n, \tag{A.16}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial u_n} = 0 \ : \ & \int_{\underline{\varphi}^n}^{u_n - v(b)} W'(u_n - \varphi) k(\varphi|n) \mathrm{d}\varphi f(n) + \\ & \lambda \kappa_n \left(T(z_n) + b\right) \tilde{k}(n) - \eta_n - \frac{\mathrm{d}\theta_n}{\mathrm{d}n} = 0, \quad \forall n \neq \underline{n}, \overline{n}, \end{aligned} \tag{A.17}$$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial b} = 0 \ : \ & \int_{\mathcal{N}} \Big(W'(v(b)) v'(b) (f(n) - \tilde{k}(n)) - \lambda (f(n) - \tilde{k}(n)) \Big) \mathrm{d}n, \\ & -\lambda \int_{\mathcal{N}} \kappa_n v'(b) \left(T(z_n) + b\right) \tilde{k}(n) \mathrm{d}n = 0, \end{aligned} \tag{A.18}$$

$$\lim_{n \to \underline{n}} \theta_n = 0, \quad \lim_{n \to \overline{n}} \theta_n = 0, \tag{A.19}$$

where we used $\kappa_n \equiv \frac{K'(u_n - v(b)|n) f(n)}{\tilde{k}(n)}$, which denotes the semi-elasticity of participation with respect to a utility increase for employed. We employed Leibniz' rule in the first-order conditions for u_n and b to find the derivatives of the Lagrangian with respect to the the bound $(u_n - v(b))$ of the integrals. We also used $T(z_n) = z_n - c_n$ to simplify the first-order conditions for u_n and b. We omitted restating the incentive-compatibility and resource constraints.

We can solve for the modified ABC-formula using the same procedure as for the intensive margin. First, solve for η_n using equation (A.15): $\eta_n = \frac{\lambda \tilde{k}(n)}{v'(c_n) f(n)}$ and substitute this into (A.16) and simplify:

$$1 - \frac{h'(l_n)}{n v'(c_n)} = \frac{\theta_n \left(h'(l_n) + l_n h''(l_n)\right)}{\lambda \tilde{k}(n) n^2}. \tag{A.20}$$

Substitute (A.20) in first-order condition (2.4) and rewrite:

$$\frac{T'(z_n)}{1 - T'(z_n)} = \left(1 + \frac{l_n h''(l_n)}{h'(l_n)}\right) \frac{\theta_n v'(c_n)/\lambda}{\tilde{K}(\overline{n}) - \tilde{K}(n)} \frac{\tilde{K}(\overline{n}) - \tilde{K}(n)}{n \tilde{k}(n)}, \tag{A.21}$$

where $\tilde{K}(n) = \int_{\underline{n}}^{n} \tilde{k}(m) \, \mathrm{d}m$ is the fraction of employed workers in the population with skill level n or less.

Integrate equation (A.17), and use the transversality condition, to find the solution for θ_n:

$$\theta_n = \int_n^{\overline{n}} \lambda \left(\frac{1}{v'(c_m)} - \kappa_m (T(z_m) + b) \right) \tilde{k}\left(m\right) \mathrm{d}m - \int_n^{\overline{n}} \int_{\underline{\varphi}^n}^{u_m - v(b)} W'\left(u_m - \varphi\right) k(\varphi|n) \mathrm{d}\varphi f\left(m\right) \mathrm{d}m. \tag{A.22}$$

Use the expected, conditional welfare weight of an individual with ability n $g_n^P \equiv \int_{\underline{\varphi}^n}^{u_n - v(b)} \frac{W'(u_n - \varphi) v'(c_n)}{\lambda} k(\varphi|n) \mathrm{d}\varphi / K\left(u_n - b\right)$ and simplify (A.22):

$$\theta_n = \int_n^{\overline{n}} \lambda \left(\frac{1 - g_m^P}{v'(c_m)} - \kappa_m (T(z_m) + b) \right) \tilde{k}\left(m\right) \mathrm{d}m, \tag{A.23}$$

Finally, combine expressions (A.23), (A.21), and (A.11) to obtain the adjusted ABC-formula (2.18).

In addition, equation (A.18) describes an optimality condition for unemployment benefits b. It can be simplified by solving the integrals and introducing the marginal social welfare weight g_0 of unemployed individuals: $g_0 \equiv W'(v(b))v'(b)/\lambda$. Use g_0 to simplify (A.18):

$$(g_0 - 1)(1 - \tilde{K}(\overline{n})) = v'(b) \int_{\mathcal{N}} \kappa_m (T\left(z_m\right) + b) \tilde{k}(m) \mathrm{d}m. \tag{A.24}$$

Simplify the right-hand side by imposing the transversality condition at the top:

$$\theta_{\overline{n}} = \int_{\mathcal{N}} \lambda \left(\frac{g_m^P - 1}{v'(c_m)} + \kappa_m (T\left(z_m\right) + b) \right) \tilde{k}(m) \mathrm{d}m = 0. \tag{A.25}$$

From (A.25) then follows the expression for the optimal participation tax:

$$v'(b) \int_{\mathcal{N}} \kappa_n (T(z_n) + b) \tilde{k}(m) \mathrm{d}m = v'(b) \int_{\mathcal{N}} \frac{(1 - g_m^P) \tilde{k}(m) \mathrm{d}m}{v'(c_m)}. \tag{A.26}$$

Use (A.26) to simplify (A.24):

$$\frac{(g_0 - 1)(1 - \tilde{K}(\overline{n}))}{v'(b)} = \int_{\mathcal{N}} \frac{(1 - g_m^P) \tilde{k}(m) \mathrm{d}m}{v'(c_n)}. \tag{A.27}$$

A.3 Deriving Behavioral Elasticities

This appendix derives the exact elasticities – taking into account the non-linearity of the tax system, as in Jacquet *et al.* (2010). Individuals maximize utility $u\left(c, l\right)$ subject to their budget constraint $c = nl - T\left(nl\right)$. The first-order condition (FOC) is given by

$n(1-T')\,u_c(.)+u_l(.)=0$. Define the following function

$$\begin{aligned} Y(l,n,\tau,\rho) \;\equiv\;& n\left(1-T'(nl)+\tau\right)u_c\left(nl-T(nl)+\tau\left(nl-nl_n\right)+\rho,l\right) \\ &+u_l\left(nl-T(nl)+\tau(nl-n\hat{l})+\rho,l\right). \end{aligned} \tag{A.28}$$

$Y(z,n,\tau,\rho)$ measures the shift of the first-order condition of the household when the marginal tax rate exogenously increases with τ (i.e., for any level of earnings) or when the household receives an exogenous amount of income ρ, irrespective of the amount of work effort. The first-order condition of the household is equivalent to $Y(z_n,n,0,0)=0$. Introducing the second term, $\tau(nl-n\hat{l})$, has the following intuition. Suppose we raise the marginal tax rate – irrespective of income level nl – and we evaluate the impact at $n\hat{l}$ (the optimum choice for $\hat{l}$ of household n), then this marginal tax increase does not change income, only the marginal incentives to supply labor. ρ represents the income effect: suppose that we give the household a marginal increase in income of ρ, starting from $\rho=0$, what will happen to labor supply?

We find the following derivatives, using the first-order condition $-u_l=n(1-T')\,u_c$:

$$Y_l(l_n,n,0,0) = u_{ll}+\left(\frac{u_l}{u_c}\right)^2 u_{cc}-2\frac{u_l}{u_c}u_{cl}+nu_l\frac{T''}{1-T'}, \tag{A.29}$$

$$Y_n(l_n,n,0,0) = \left(-u_l/l+nu_l\frac{T''}{1-T'}+\left(\frac{u_l}{u_c}\right)^2 u_{cc}-\left(\frac{u_l}{u_c}\right)u_{lc}\right)\frac{l}{n}, \tag{A.30}$$

$$Y_\tau(l_n,n,0,0) = nu_c, \tag{A.31}$$

$$Y_\rho(l_n,n,0,0) = n(1-T')\,u_{cc}+u_{lc}=\frac{u_{lc}u_c-u_lu_{cc}}{u_c}. \tag{A.32}$$

Now, by applying the envelope theorem we find

$$\frac{\partial l}{\partial x}=-\frac{Y_x}{Y_l},\quad x=n,\tau,\rho \tag{A.33}$$

Hence, the uncompensated *wage* elasticity of labor supply ε^u is equal to:

$$\varepsilon^u\equiv\frac{\partial l}{\partial n}\frac{n}{l}=\frac{u_l/l+\left(\frac{u_l}{u_c}\right)u_{lc}-\left(\frac{u_l}{u_c}\right)^2u_{cc}-nu_l\frac{T''}{1-T'}}{u_{ll}+\left(\frac{u_l}{u_c}\right)^2u_{cc}-2\frac{u_l}{u_c}u_{cl}+nu_l\frac{T''}{1-T'}}. \tag{A.34}$$

The compensated *wage* elasticity of labor supply ζ^c is given by:

$$\zeta^c \equiv \frac{\partial l}{\partial n}\frac{n}{l} = \frac{u_l/l - nu_l\frac{T''}{1-T'}}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\frac{u_l}{u_c}u_{cl} - nu_l\frac{T''}{1-T'}}. \tag{A.35}$$

And, the compensated *tax* elasticity ε^c is:

$$\varepsilon^c \equiv -\frac{\partial l}{\partial \tau}\frac{1-T'}{l} = \frac{u_l/l}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\frac{u_l}{u_c}u_{cl} + nu_l\frac{T''}{1-T'}}. \tag{A.36}$$

Note that the compensated *wage* elasticity of labor supply and the compensated *tax* elasticity of labor supply are not identical due to the non-linearity in the tax system. Increasing the marginal tax rate τ amounts to increasing the marginal tax, irrespective of the income level, whereas increasing the wage rate also changes the marginal tax rates as a result of the non-linearities in the tax system.

The income elasticity of labor supply is defined by the Slutsky equation ($\eta \equiv \varepsilon^u - \zeta^c$):

$$\eta = (1-T')\,n\frac{\partial l}{\partial \rho} = \frac{\frac{-u_l}{u_c}\left(\frac{u_l}{u_c}u_{cc} - u_{lc}\right)}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\frac{u_l}{u_c}u_{cl} + nu_l\frac{T''}{1-T'}}. \tag{A.37}$$

All the elasticities depend on the second derivatives of the tax function. Hence, in contrast to Saez (2001), the second-derivatives cannot be ignored in the expressions of the elasticities if tax systems are non-linear. We thus confirm Blomquist and Simula (2010). If $T'' > 0$ distortions of taxes are lower – ceteris paribus. However, if $T'' < 0$ the reverse is true. The reason is that if marginal tax rates are increasing ($T'' > 0$) the labor-supply response dampens out, but if the marginal tax rates are decreasing ($T'' < 0$) the labor-supply response is magnified by the non-linearity in the tax schedule.

Note that we can derive that

$$1 + \frac{lu_{ll}}{u_l} - \frac{lu_{lc}}{u_c} = \frac{1+\varepsilon^u}{\varepsilon^c} \tag{A.38}$$

Thus, the term $1 + \frac{lu_{ll}}{u_l} - \frac{lu_{lc}}{u_c}$ equals one plus the uncompensated *wage* elasticity of labor supply, divided by the compensated *tax* elasticity of labor supply. The former does include the impact of the non-linear tax schedule, whereas the latter does not. Only when the tax system is linear, this expression reduces to $\frac{1+\varepsilon^u}{\zeta^c}$ as in Saez (2001).

For the specific utility function $u(c,l) \equiv v(c) - h(l)$ we obtain the following elasticities:

$$\varepsilon^u = \frac{v' + \frac{lh'v''}{v'} - v'nl\frac{T''}{1-T'}}{\frac{lh''v'}{h'} - \frac{lh'v''}{v'} + v'nl\frac{T''}{1-T'}}, \tag{A.39}$$

$$\varepsilon^c = \frac{v'}{\frac{lh''v'}{h'} - \frac{lh'v''}{v'} + v'nl\frac{T''}{1-T'}}. \tag{A.40}$$

A.4 Marginal Dead Weight Loss Non-linear Income Tax

To determine the deadweight loss of a non-linear tax schedule $T(z_n)$ with $T' \equiv \mathrm{d}T(\cdot)/\mathrm{d}z_n$, suppose that we increase the marginal tax rates at each and every point of the tax schedule with $\mathrm{d}T'$, how large is the marginal deadweight loss of that tax increase? To answer this question, we conduct the following hypothetical thought experiment. Each household n gets perfectly compensated via a household-specific lump-sum transfer T_n so that its utility remains unaffected.[2] This implies that our deadweight loss measure is based on the compensating variation. The marginal deadweight loss then equals the net loss in public revenue so as to keep everyone's utility constant. Note that this hypothetical tax reform does not affect the participation margin, since the benefit given to non-working individuals b remains constant and the utility of all working individuals u_n does not change.

Indirect utility of all working individuals can be written as a function $v((1-T'(nl_n))n, \tilde{T}_n)$ of the net marginal wage rate $(1-T')n$, and so-called virtual income $\tilde{T}_n$. Virtual income is defined as

$$\tilde{T}_n \equiv nl_n - T(nl_n) - (1 - T'(nl_n))nl_n. \tag{A.41}$$

So that the household budget constraint can be written as:

$$c_n = nl_n - T(nl_n) = (1 - T'(nl_n))nl_n + \tilde{T}_n. \tag{A.42}$$

Note that virtual income works like the intercept of the tax function if the marginal tax rate T' had been constant. From applying Roy's identity we find that

$$\frac{\partial v_n}{\partial \tilde{T}_n} = \lambda_n, \quad \frac{\partial v_n}{\partial T'} = -\lambda_n nl_n, \tag{A.43}$$

[2]Of course this instrument does not exist, since it boils down to an individualized lump-sum tax. However, this thought-exercise allows us to calculate the excess burden of the tax.

The change in taxes $\mathrm{d}T'$ and lump-sum income $\mathrm{d}T_n$ for each household n, which leaves private utility unaffected satisfies:

$$\mathrm{d}v_n = \lambda_n \mathrm{d}T_n - \lambda_n n l_n \mathrm{d}T' = 0. \tag{A.44}$$

where we used the derivatives of indirect utility here. The transfers T_n play the same role as the virtual income $\tilde{T}_n$. Hence, the derivatives of indirect utility with respect to $\tilde{T}_n$ or T_n are identical. Consequently, when each individual gets a perfect compensation for the tax change, we have

$$\mathrm{d}T_n = n l_n \mathrm{d}T'. \tag{A.45}$$

What is the effect of this tax policy on the public budget? There are three effects. i) For each working individual n, the government loses revenue $\mathrm{d}T_n$. ii) When the tax rate increases, the government also gains revenue $n l_n \mathrm{d}T'$. iii) The individual will change its (compensated) labor supply in response to higher taxation. This results in a decline of total tax revenue for the government with $T' n \frac{\partial l_n^c}{\partial t} \mathrm{d}T'$.

The change in total public revenue $\mathrm{d}R_n$ per individual is the sum of these three effects:

$$\mathrm{d}R_n = -\mathrm{d}T_n + n l_n \mathrm{d}T' + T' n \frac{\partial l_n^c}{\partial T'} \mathrm{d}T' = T' n \frac{\partial l_n^c}{\partial T'} \mathrm{d}T'. \tag{A.46}$$

Note that the first two terms sum to zero, since each household gets perfectly compensated: $\mathrm{d}T_n = n l_n \mathrm{d}T'$, see above. Therefore, the total revenue loss for the government on individual n is

$$\frac{\mathrm{d}R_n}{\mathrm{d}T'} = n T' \frac{\partial l_n^c}{\partial T'} = \frac{T'}{1-T'} n l_n \frac{\partial l_n^c}{\partial T'} \frac{1-T'}{l_n} = -\frac{T'}{1-T'} n l_n \varepsilon_n^c. \tag{A.47}$$

Finally, summing the revenue losses $\frac{\mathrm{d}R_n}{\mathrm{d}T'}$ over all working households and dividing this sum by total taxable income $\int_{\mathcal{N}} n l_n \tilde{k}(n)\,\mathrm{d}n$ yields the total marginal excess burden as a fraction of taxed income:

$$MEB \equiv \frac{\int_{\mathcal{N}} -\frac{\mathrm{d}G_n}{\mathrm{d}T'} \tilde{k}(n)\,\mathrm{d}n}{\int_{\mathcal{N}} n l_n \tilde{k}(n)\,\mathrm{d}n} = \frac{\int_{\mathcal{N}} \frac{T'(z_n)}{1-T'(z_n)} n l_n \varepsilon^c \tilde{k}(n)\,\mathrm{d}n}{\int_{\mathcal{N}} n l_n \tilde{k}(n)\,\mathrm{d}n}, \tag{A.48}$$

Note that this deadweight loss formula is applicable to any tax schedule, including the optimal one. If the compensated tax elasticity of labor supply (ε_n^c) is constant across skills, as we assume, then we find that the marginal deadweight loss is a function of the income-weighted marginal tax rates $\frac{T'(z_n)}{1-T'(z_n)}$:

$$MEB \equiv \varepsilon_n^c \frac{\int_{\mathcal{N}} \frac{T'(z_n)}{1-T'(z_n)} n l_n \tilde{k}(n)\,\mathrm{d}n}{\int_{\mathcal{N}} n l_n \tilde{k}(n)\,\mathrm{d}n}. \tag{A.49}$$

Appendix B

Appendix to Chapter 3

B.1 Derivations

B.1.1 Optimal Income Taxation with Intensive Margin Only

We will solve the optimal income tax using Lagrangian methods. Multiply the incentive constraint with θ_n and apply integration by parts to $\theta_n \frac{\mathrm{d}u_n}{\mathrm{d}n}$ so as to find:

$$\int_{\mathcal{N}} \left(-\theta_n \frac{z_n h'(z_n/n)}{n^2} - u_n \frac{\mathrm{d}\theta_n}{\mathrm{d}n} \right) \mathrm{d}n + \theta_{\overline{n}} u_{\overline{n}} - \theta_{\underline{n}} u_{\underline{n}} = 0. \tag{B.1}$$

Now, set up the optimal-tax problem as a Lagrangian with c_n, z_n, and u_n as control variables. We furthermore introduce λ as the Lagrange multiplier of the economy's resource constraint. $\eta_n f(n)$ denotes the composite Lagrange multiplier of the utility constraint at n (we have harmlessly pre-multiplied each multiplier η_n with $f(n)$ to avoid some additional notation). θ_n is the Lagrange multiplier of the incentive-compatibility constraint at n:[1]

$$\begin{aligned}\mathcal{L} \equiv & \int_{\mathcal{N}} \left(W(u_n) + \lambda\left(z_n - c_n - R\right)\right) f(n)\mathrm{d}n + \int_{\mathcal{N}} \eta_n \left(v(c_n) - h(z_n/n) - u_n\right) f(n)\mathrm{d}n \\ & - \int_{\mathcal{N}} \left(\theta_n \frac{z_n h'\left(z_n/n\right)}{n^2} + u_n \frac{\mathrm{d}\theta_n}{\mathrm{d}n} \right) \mathrm{d}n + \theta_{\overline{n}} u_{\overline{n}} - \theta_{\underline{n}} u_{\underline{n}}.\end{aligned} \tag{B.2}$$

[1] We need the latter constraint because all variables in the utility function c and z as well as utility itself u are considered choice variables for the government in this optimization procedure. Alternatively, one may invert the utility function and write consumption as a function of the allocation: $c(z, u)$, which is usually done in the literature.

The first-order and transversality conditions for this control problem are given by:

$$\frac{\partial \mathcal{L}}{\partial c_n} = 0 \quad : \quad -\lambda f(n) + \eta_n f(n) v'(c_n) = 0, \quad \forall n, \tag{B.3}$$

$$\frac{\partial \mathcal{L}}{\partial z_n} = 0 \quad : \quad \lambda f(n) - \eta_n \frac{h'(l_n)}{n} - \theta_n \frac{h'(l_n) + l_n h''(l_n)}{n^2} = 0, \quad \forall n, \tag{B.4}$$

$$\frac{\partial \mathcal{L}}{\partial u_n} = 0 \quad : \quad W'(u_n) f(n) - \eta_n - \frac{\mathrm{d}\theta_n}{\mathrm{d}n} = 0, \quad \forall n \neq \underline{n}, \overline{n}, \tag{B.5}$$

$$\lim_{n \to \underline{n}} \theta_n = 0, \quad \lim_{n \to \overline{n}} \theta_n = 0. \tag{B.6}$$

We omitted restating the incentive-compatibility and resource constraints. We now derive the optimal tax formula as reported in Saez (2001).

First, solve (B.3) for η_n to find $\eta_n = \frac{\lambda}{v'(c_n)}$, and substitute this into (B.4) and simplify:

$$1 - \frac{h'(l_n)}{nv'(c_n)} = \frac{\theta_n \left(h'(l_n) + l_n h''(l_n)\right)}{\lambda f(n) n^2}. \tag{B.7}$$

Substitute the individuals' FOC into (B.7) and simplify the resulting equation:

$$\frac{T'(z_n)}{1 - T'(z_n)} = \left(1 + \frac{l_n h''(l_n)}{h'(l_n)}\right) \frac{\theta_n v'(c_n)/\lambda}{(1 - F(n))} \frac{1 - F(n)}{f(n) n}. \tag{B.8}$$

For the utility function we used, the compensated and uncompensated labor supply elasticities are given by (see last Appendix):

$$\varepsilon_n^c \equiv -\frac{\partial l_n}{\partial \tau} \frac{1 - T'}{l_n} = \frac{v'}{\frac{lh''v'}{h'} - \frac{lh'v''}{v'} + v'nl\frac{T''}{1-T'}}, \tag{B.9}$$

$$\varepsilon_n^u \equiv \frac{\partial l_n}{\partial n} \frac{n}{l_n} = \frac{v' + \frac{lh'v''}{v'} - v'nl\frac{T''}{1-T'}}{\frac{lh''v'}{h'} - \frac{lh'v''}{v'} + v'nl\frac{T''}{1-T'}}. \tag{B.10}$$

Therefore, we find

$$\frac{1 + \varepsilon_n^u}{\varepsilon_n^c} = 1 + \frac{l_n h''(l_n)}{h'(l_n)}. \tag{B.11}$$

In addition, integrating equation (B.5), and using a transversality condition, a solution for θ_n can be derived:

$$\theta_n = \int_n^{\overline{n}} \left(\frac{\lambda}{v'(c_m)} - W'(u_m)\right) f(m)\,\mathrm{d}m. \tag{B.12}$$

By introducing normalized social welfare weight $g_n \equiv \frac{W'(u_n)v'(c_n)}{\lambda}$, which denotes the monetized welfare gain of providing one euro to individual n, this expression can be simplified:

$$\theta_n = \lambda \int_n^{\overline{n}} \frac{(1-g_m)}{v'(c_m)} f(m)\,\mathrm{d}m. \tag{B.13}$$

Substituting results (B.11) and (B.13) into (B.8). The final constraints on θ are the transversality conditions. Note that the transversality conditions imply that the distortion on labor supply at the top and the bottom should equal zero.

B.1.2 Optimal Income Taxation with Both Intensive and Extensive Margins

The incentive-compatibility constraint is unaffected by introducing the extensive margin. We will solve the optimal income tax again using a Lagrangian, which uses c_n, z_n, and u_n as control variables. λ is the Lagrange multiplier of the economy's resource constraint, $\eta_n f(n)$ denotes the composite Lagrange multipliers on the utility constraint for each n, and θ_n is the Lagrange multiplier of the incentive-compatibility constraint at n. The Lagrangian can be written as:

$$\begin{aligned}\mathcal{L} \;=\; & \int_{\mathcal{N}} \left(\int_{\underline{\varphi}^n}^{u_n - v(b)} W(u_n-\varphi)k(\varphi|n)\mathrm{d}\varphi f(n) + W\left(v(b)\right)\left(f(n) - K(u_n - v(b)|n)f(n)\right) \right) \mathrm{d}n \\ & + \int_{\mathcal{N}} \lambda \left((z_n - c_n)K(u_n - v(b)|n)f(n) - (f(n) - K(u_n - v(b)|n)f(n))b) - Rf(n) \right) \mathrm{d}n \\ & + \int_{\mathcal{N}} \eta_n \left(v(c_n) - h(z_n/n) - u_n \right) f(n)\mathrm{d}n \\ & - \int_{\mathcal{N}} \left(\theta_n \frac{z_n h'(z_n/n)}{n^2} + u_n \frac{\mathrm{d}\theta_n}{\mathrm{d}n} \right) \mathrm{d}n + \theta_{\overline{n}} u_{\overline{n}} - \theta_{\underline{n}} u_{\underline{n}}. \end{aligned} \tag{B.14}$$

where we substituted the definition for $\tilde{k}(n) \equiv K(u_n - v(b)|n)f(n)$. The first-order and transversality conditions for this control problem are given by:

$$\frac{\partial \mathcal{L}}{\partial c_n} = 0 \quad : \quad -\lambda\tilde{k}(n) + \eta_n v'(c_n)f(n) = 0, \quad \forall n, \tag{B.15}$$

$$\frac{\partial \mathcal{L}}{\partial z_n} = 0 \quad : \quad \lambda\tilde{k}(n) - \eta_n \frac{h'(l_n)}{n} - \theta_n \frac{h'(l_n) + l_n h''(l_n)}{n^2} = 0, \quad \forall n, \tag{B.16}$$

$$\frac{\partial \mathcal{L}}{\partial u_n} = 0 \quad : \quad \int_{\underline{\varphi}^n}^{u_n - v(b)} W'(u_n - \varphi)k(\varphi|n)\mathrm{d}\varphi f(n) + \tag{B.17}$$

$$\lambda\kappa_n \left(T(z_n) + b\right)\tilde{k}(n) - \eta_n - \frac{\mathrm{d}\theta_n}{\mathrm{d}n} = 0, \quad \forall n \neq \underline{n}, \overline{n},$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \quad : \quad \int_{\mathcal{N}} \Big(W'(v(b))v'(b)(f(n) - \tilde{k}(n)) - \lambda(f(n) - \tilde{k}(n))\Big)\,\mathrm{d}n, \tag{B.18}$$

$$-\lambda \int_{\mathcal{N}} \kappa_n v'(b)\left(T(z_n) + b\right)\tilde{k}(n)\mathrm{d}n = 0,$$

$$\lim_{n \to \underline{n}} \theta_n = 0, \quad \lim_{n \to \overline{n}} \theta_n = 0, \tag{B.19}$$

where we used $\kappa_n \equiv \frac{K'(u_n - v(b)|n)f(n)}{\tilde{k}(n)}$, which denotes the semi-elasticity of participation with respect to a utility increase for employed. We employed Leibniz' rule in the first-order conditions for u_n and b to find the derivatives of the Lagrangian with respect to the the bound $(u_n - v(b))$ of the integrals. We also used $T(z_n) = z_n - c_n$ to simplify the first-order conditions for u_n and b. We omitted restating the incentive-compatibility and resource constraints.

We can solve for the modified ABC-formula using the same procedure as for the intensive margin. First, solve for η_n using equation (B.15): $\eta_n = \frac{\lambda\tilde{k}(n)}{v'(c_n)f(n)}$ and substitute this into (B.16) and simplify:

$$1 - \frac{h'(l_n)}{nv'(c_n)} = \frac{\theta_n\left(h'(l_n) + l_n h''(l_n)\right)}{\lambda\tilde{k}(n)n^2}. \tag{B.20}$$

Substitute (B.20) in first-order condition the household's first order condition and rewrite:

$$\frac{T'(z_n)}{1 - T'(z_n)} = \left(1 + \frac{l_n h''(l_n)}{h'(l_n)}\right)\frac{\theta_n v'(c_n)/\lambda}{\tilde{K}(\overline{n}) - \tilde{K}(n)}\frac{\tilde{K}(\overline{n}) - \tilde{K}(n)}{n\tilde{k}(n)}, \tag{B.21}$$

where $\tilde{K}(n) = \int_{\underline{n}}^{n} \tilde{k}(m)\,\mathrm{d}m$ is the fraction of employed workers in the population with skill level n or less.

Integrate equation (B.17), and use the transversality condition, to find the solution for θ_n:

$$\theta_n = \int_n^{\overline{n}} \lambda \left(\frac{1}{v'(c_m)} - \kappa_m (T(z_m) + b) \right) \tilde{k}\,(m)\,\mathrm{d}m - \int_n^{\overline{n}} \int_{\underline{\varphi}^n}^{u_m - v(b)} W'\,(u_m - \varphi)\, k(\varphi|n) \mathrm{d}\varphi f\,(m)\,\mathrm{d}m. \tag{B.22}$$

Use the expected, conditional welfare weight of an individual with ability n $g_n^P \equiv \int_{\underline{\varphi}^n}^{u_n - v(b)} \frac{W'(u_n - \varphi) v'(c_n)}{\lambda} k(\varphi|n) \mathrm{d}\varphi / K\,(u_n - b)$ and simplify (B.22):

$$\theta_n = \int_n^{\overline{n}} \lambda \left(\frac{1 - g_m^P}{v'(c_m)} - \kappa_m (T(z_m) + b) \right) \tilde{k}\,(m)\,\mathrm{d}m, \tag{B.23}$$

Finally, combine expressions (B.23), (B.21), and (B.11) to obtain the adjusted ABC-formula (3.9).

In addition, equation (B.18) describes an optimality condition for unemployment benefits b. It can be simplified by solving the integrals and introducing the marginal social welfare weight g_0 of unemployed individuals: $g_0 \equiv W'(v(b))v'(b)/\lambda$. Use g_0 to simplify (B.18):

$$(g_0 - 1)(1 - \tilde{K}(\overline{n})) = v'(b) \int_{\mathcal{N}} \kappa_m (T\,(z_m) + b) \tilde{k}(m) \mathrm{d}m. \tag{B.24}$$

Simplify the right-hand side by imposing the transversality condition at the top:

$$\theta_{\overline{n}} = \int_{\mathcal{N}} \lambda \left(\frac{g_m^P - 1}{v'(c_m)} + \kappa_m (T\,(z_m) + b) \right) \tilde{k}(m) \mathrm{d}m = 0. \tag{B.25}$$

From (B.25) then follows the expression for the optimal participation tax:

$$v'(b) \int_{\mathcal{N}} \kappa_n (T(z_n) + b) \tilde{k}(m) \mathrm{d}m = v'(b) \int_{\mathcal{N}} \frac{(1 - g_m^P)\tilde{k}(m)\mathrm{d}m}{v'(c_m)}. \tag{B.26}$$

Use (B.26) to simplify (B.24):

$$\frac{(g_0 - 1)(1 - \tilde{K}(\overline{n}))}{v'(b)} = \int_{\mathcal{N}} \frac{(1 - g_m^P)\tilde{k}(m)\mathrm{d}m}{v'(c_n)}. \tag{B.27}$$

B.1.3 Deriving Behavioral Elasticities

This appendix derives the exact elasticities – taking into account the non-linearity of the tax system, as in Jacquet *et al.* (2010). Individuals maximize utility $u\,(c, l)$ subject to their budget constraint $c = nl - T\,(nl)$. The first-order condition (FOC) is given by

$n(1-T')u_c(.)+u_l(.)=0$. Define the following function

$$\begin{aligned} Y(l,n,\tau,\rho) &\equiv n(1-T'(nl)+\tau)u_c(nl-T(nl)+\tau(nl-nl_n)+\rho,l) \\ &\quad +u_l\left(nl-T(nl)+\tau(nl-n\hat{l})+\rho,l\right). \end{aligned} \tag{B.28}$$

$Y(z,n,\tau,\rho)$ measures the shift of the first-order condition of the household when the marginal tax rate exogenously increases with τ (i.e., for any level of earnings) or when the household receives an exogenous amount of income ρ, irrespective of the amount of work effort. The first-order condition of the household is equivalent to $Y(z_n,n,0,0)=0$. Introducing the second term, $\tau(nl-n\hat{l})$, has the following intuition. Suppose we raise the marginal tax rate – irrespective of income level nl – and we evaluate the impact at $n\hat{l}$ (the optimum choice for $\hat{l}$ of household n), then this marginal tax increase does not change income, only the marginal incentives to supply labor. ρ represents the income effect: suppose that we give the household a marginal increase in income of ρ, starting from $\rho=0$, what will happen to labor supply?

We find the following derivatives, using the first-order condition $-u_l=n(1-T')u_c$:

$$Y_l(l_n,n,0,0) = u_{ll}+\left(\frac{u_l}{u_c}\right)^2 u_{cc}-2\frac{u_l}{u_c}u_{cl}+nu_l\frac{T''}{1-T'}, \tag{B.29}$$

$$Y_n(l_n,n,0,0) = \left(-u_l/l+nu_l\frac{T''}{1-T'}+\left(\frac{u_l}{u_c}\right)^2 u_{cc}-\left(\frac{u_l}{u_c}\right)u_{lc}\right)\frac{l}{n}, \tag{B.30}$$

$$Y_\tau(l_n,n,0,0) = nu_c, \tag{B.31}$$

$$Y_\rho(l_n,n,0,0) = n(1-T')u_{cc}+u_{lc}=\frac{u_{lc}u_c-u_lu_{cc}}{u_c}. \tag{B.32}$$

Now, by applying the envelope theorem we find

$$\frac{\partial l}{\partial x}=-\frac{Y_x}{Y_l}, \quad x=n,\tau,\rho \tag{B.33}$$

Hence, the uncompensated *wage* elasticity of labor supply ε^u is equal to:

$$\varepsilon^u \equiv \frac{\partial l}{\partial n}\frac{n}{l}=\frac{u_l/l+\left(\frac{u_l}{u_c}\right)u_{lc}-\left(\frac{u_l}{u_c}\right)^2 u_{cc}-nu_l\frac{T''}{1-T'}}{u_{ll}+\left(\frac{u_l}{u_c}\right)^2 u_{cc}-2\frac{u_l}{u_c}u_{cl}+nu_l\frac{T''}{1-T'}}. \tag{B.34}$$

The compensated *wage* elasticity of labor supply ζ^c is given by:

$$\zeta^c \equiv \frac{\partial l}{\partial n}\frac{n}{l} = \frac{u_l/l - nu_l\frac{T''}{1-T'}}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\frac{u_l}{u_c}u_{cl} - nu_l\frac{T''}{1-T'}}. \tag{B.35}$$

And, the compensated *tax* elasticity ε^c is:

$$\varepsilon^c \equiv -\frac{\partial l}{\partial \tau}\frac{1-T'}{l} = \frac{u_l/l}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\frac{u_l}{u_c}u_{cl} + nu_l\frac{T''}{1-T'}}. \tag{B.36}$$

Note that the compensated *wage* elasticity of labor supply and the compensated *tax* elasticity of labor supply are not identical due to the non-linearity in the tax system. Increasing the marginal tax rate τ amounts to increasing the marginal tax, irrespective of the income level, whereas increasing the wage rate also changes the marginal tax rates as a result of the non-linearities in the tax system.

The income elasticity of labor supply is defined by the Slutsky equation ($\eta \equiv \varepsilon^u - \zeta^c$):

$$\eta = (1-T')\,n\frac{\partial l}{\partial \rho} = \frac{\frac{-u_l}{u_c}\left(\frac{u_l}{u_c}u_{cc} - u_{lc}\right)}{u_{ll} + \left(\frac{u_l}{u_c}\right)^2 u_{cc} - 2\frac{u_l}{u_c}u_{cl} + nu_l\frac{T''}{1-T'}}. \tag{B.37}$$

All the elasticities depend on the second derivatives of the tax function. Hence, in contrast to Saez (2001), the second-derivatives cannot be ignored in the expressions of the elasticities if tax systems are non-linear. We thus confirm Blomquist and Simula (2010). If $T'' > 0$ distortions of taxes are lower – ceteris paribus. However, if $T'' < 0$ the reverse is true. The reason is that if marginal tax rates are increasing ($T'' > 0$) the labor supply response dampens out, but if the marginal tax rates are decreasing ($T'' < 0$) the labor supply response is magnified by the non-linearity in the tax schedule.

Note that we can derive that

$$1 + \frac{lu_{ll}}{u_l} - \frac{lu_{lc}}{u_c} = \frac{1+\varepsilon^u}{\varepsilon^c} \tag{B.38}$$

Thus, the term $1 + \frac{lu_{ll}}{u_l} - \frac{lu_{lc}}{u_c}$ equals one plus the uncompensated *wage* elasticity of labor supply, divided by the compensated *tax* elasticity of labor supply. The former does include the impact of the non-linear tax schedule, whereas the latter does not. Only when the tax system is linear, this expression reduces to $\frac{1+\varepsilon^u}{\zeta^c}$ as in Saez (2001).

For the specific utility function $u(c,l) \equiv v(c) - h(l)$ we obtain the following elasticities:

$$\varepsilon^u = \frac{v' + \frac{lh'v''}{v'} - v'nl\frac{T''}{1-T'}}{\frac{lh''v'}{h'} - \frac{lh'v''}{v'} + v'nl\frac{T''}{1-T'}}, \tag{B.39}$$

$$\varepsilon^c = \frac{v'}{\frac{lh''v'}{h'} - \frac{lh'v''}{v'} + v'nl\frac{T''}{1-T'}}. \tag{B.40}$$

B.2 Scatter Plots and Kernels of Marginal Tax Rates

Figures B.1 to B.5 give the scatter plots of effective marginal tax rates in the baseline and for the different political parties, respectively. The figures show that there is considerable variation in marginal tax rates at most income levels, in particular for low incomes.

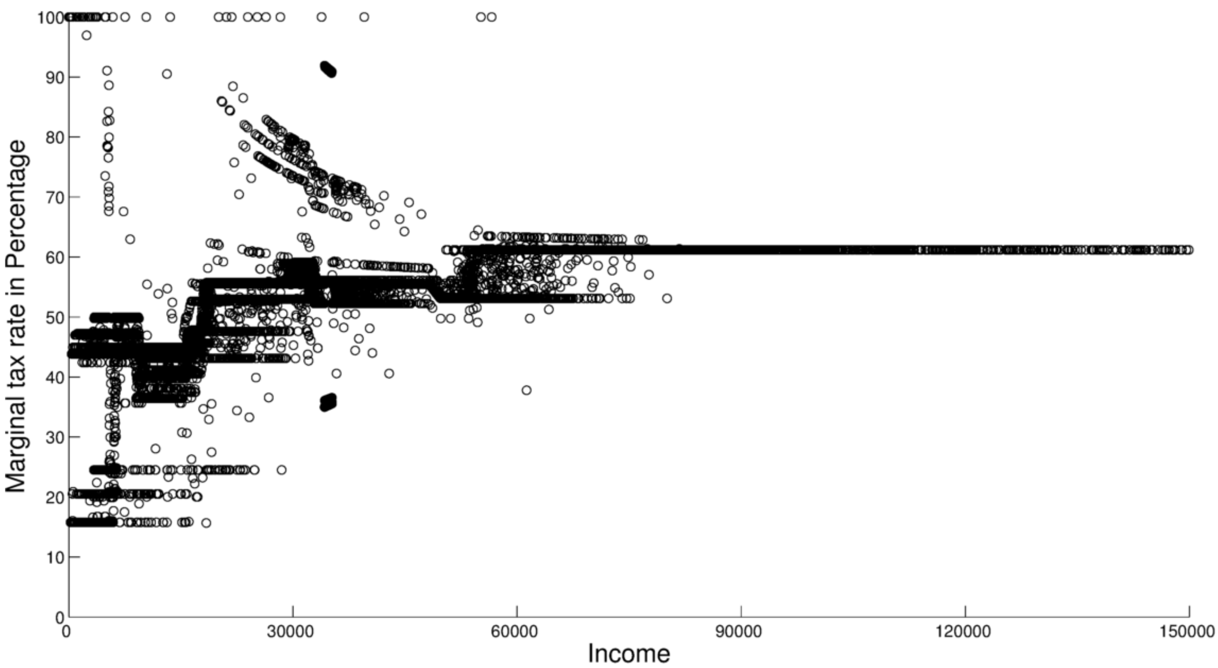


Figure B.1: Scatter Plot and Kernel Estimate of Effective Marginal Tax Rates: Baseline

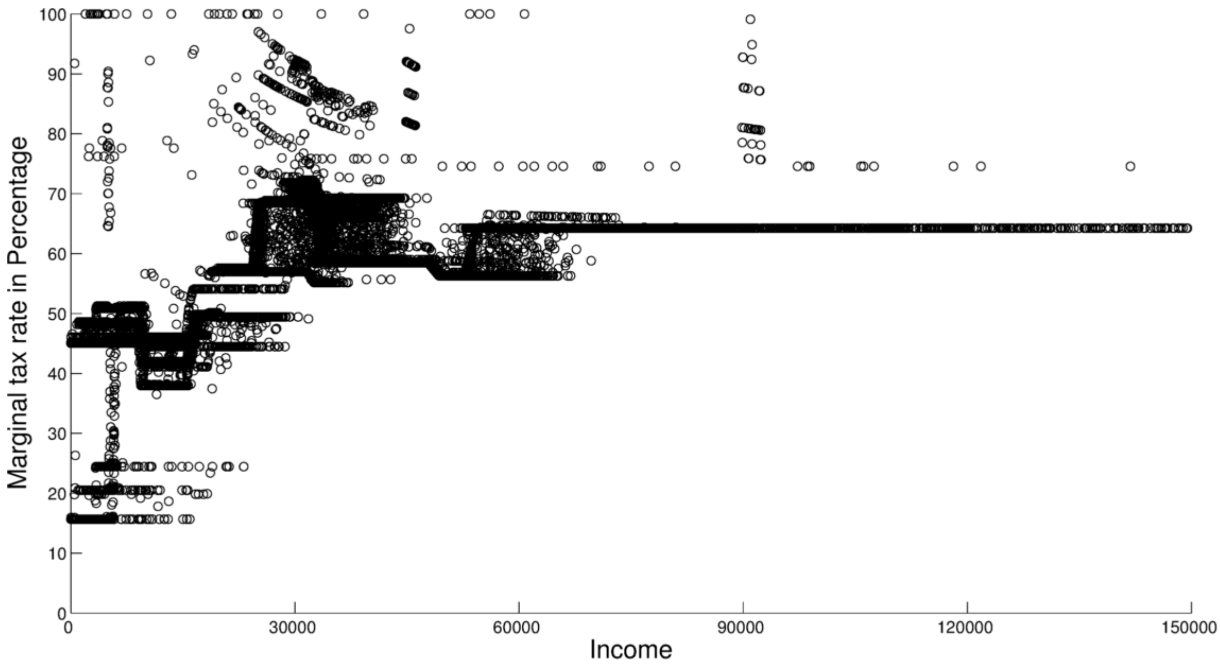


Figure B.2: Scatter Plot and Kernel Estimate of Effective Marginal Tax Rates: SP

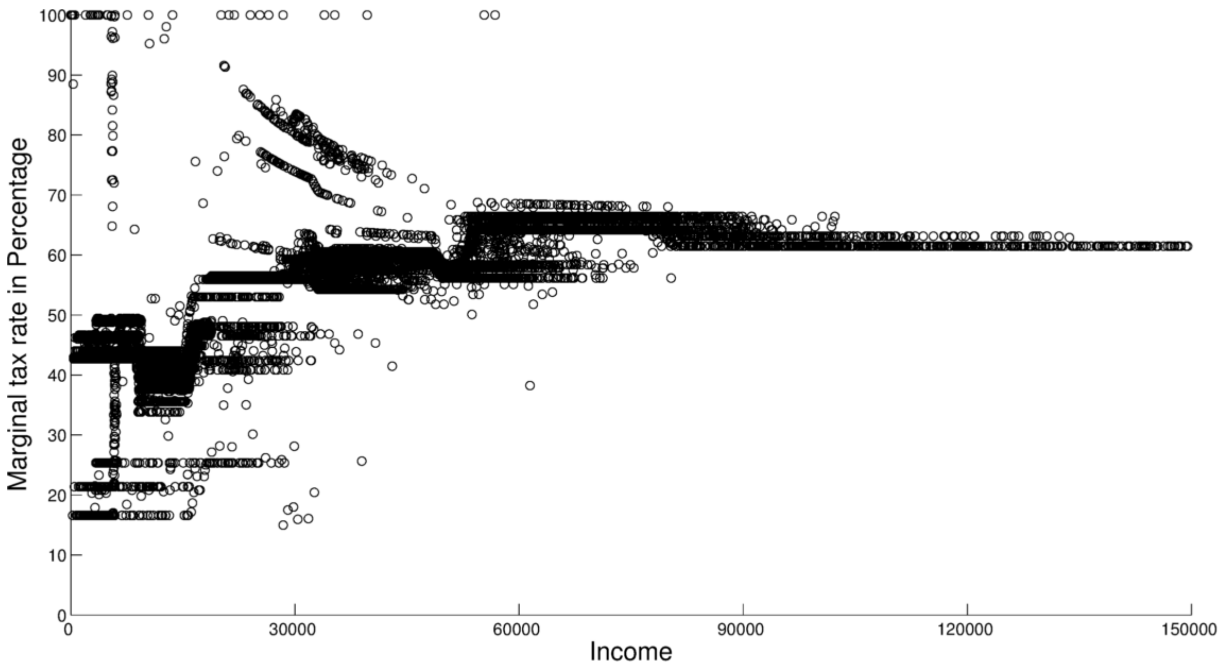


Figure B.3: Scatter Plot and Kernel Estimate of Effective Marginal Tax Rates: $PvdA$

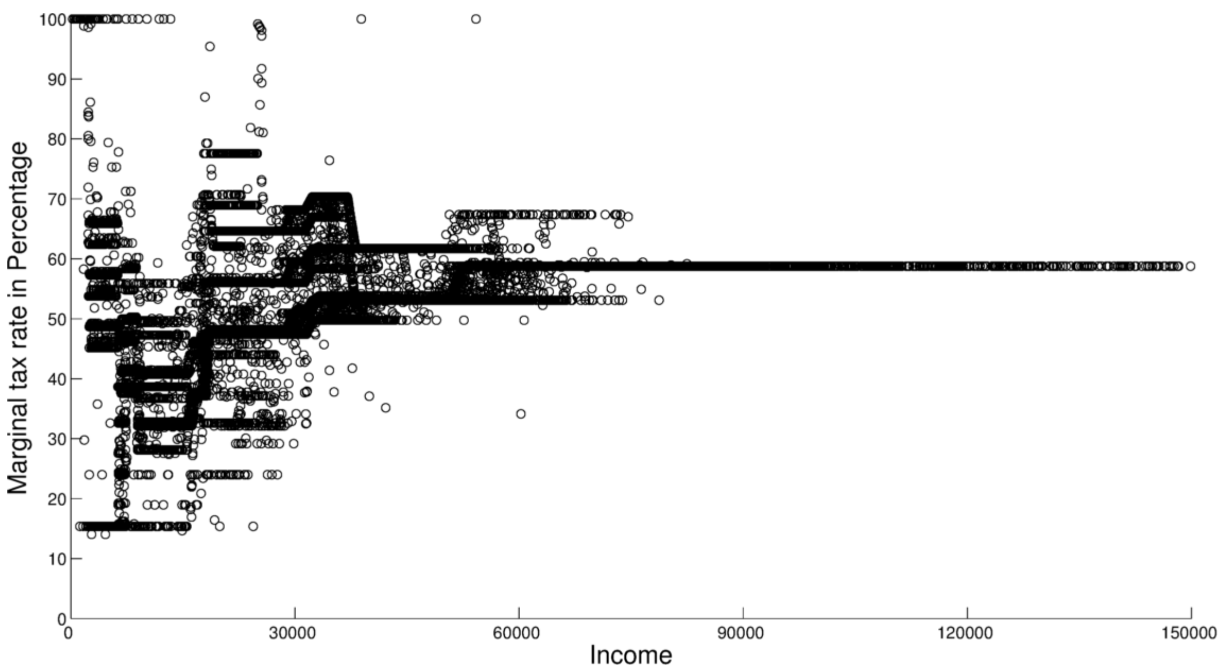


Figure B.4: Scatter Plot and Kernel Estimate of Effective Marginal Tax Rates: CDA

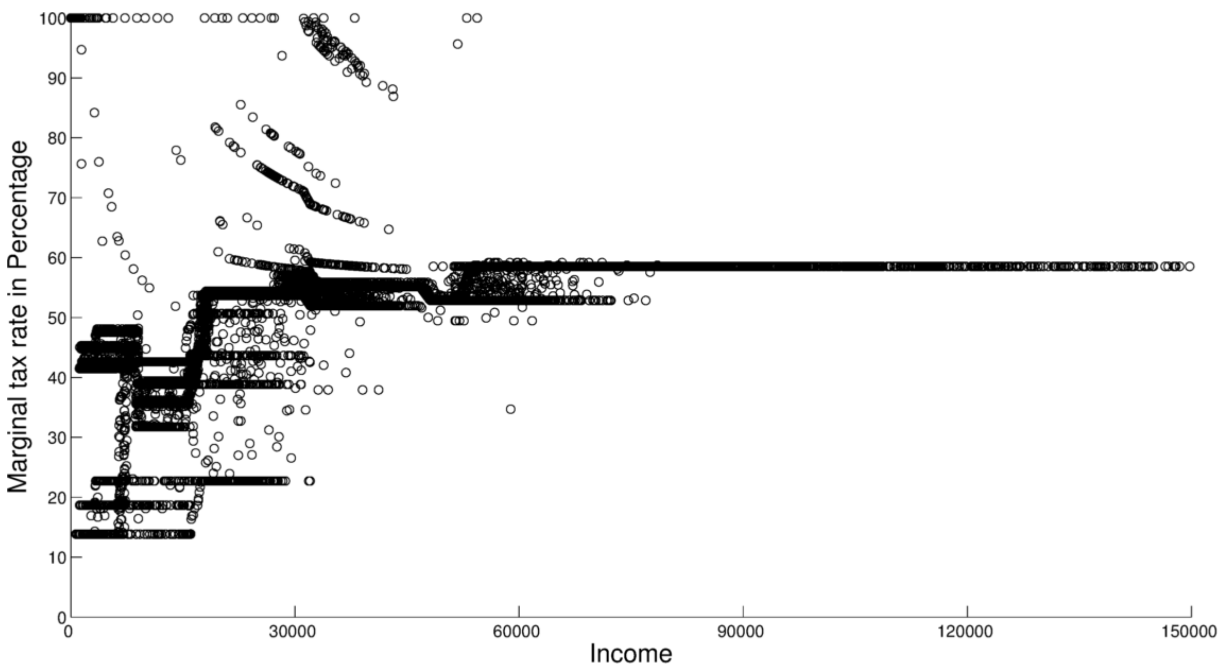


Figure B.5: Scatter Plot and Kernel Estimate of Effective Marginal Tax Rates: VVD

Appendix C

Appendix to Chapter 4

C.1 Simulation Algorithm

The algorithm we use to solve for the optimal allocation consists of two steps. First, we find the optimal allocation using a shooting method. Second, we calculate the implied wedge, tax, and monitoring schedule.[1]

C.1.1 Finding the Optimal Allocation

We find the optimal allocation through four nested loops:

1. The outer loop solves the resource constraint (4.13) for λ. A higher value of λ implies a higher shadow value of resources, and thus, a lower resource deficit, and vice versa. Therefore, we can satisfy the resource constraint arbitrarily by altering the value of λ.

2. The second loop solves the transversality condition at the top (4.32) for a given utility level at the bottom $u_{\underline{n}}$, and λ. The most important determinant in $u_{\underline{n}}$ is the transfer implied by $T(0)$. Therefore, one can think of this procedure as finding the intercept of the tax function $T(0)$. If the intercept is too low, the distortion at the top has to be positive to finance the transfer, and vice versa if the intercept is set too high. As a consequence, by varying the transfer $T(0)$ we can satisfy the transversality condition arbitrarily closely.

3. The third loop solves the differential equations (4.14) and (4.30) for given $u_{\underline{n}}$, λ, and $\theta_{\underline{n}}$ using a Runge-Kutta method to integrate over n.

[1]All Matlab programs used in the computations are available from the authors upon request.

4. The inner loop maximizes the Hamiltonian (4.26) with respect to π_n and z_n for a given state u_n and costate variable θ_n at each n.

The above algorithm is known as a shooting method because it shoots for the initial values of the differential equations that satisfy the boundary condition.

C.1.2 Finding the Optimal Wedge, Tax, and Monitoring Schedules

The above algorithm gives us a numerical approximation of the allocation $\{u_n, \theta_n, z_n, \pi_n\}$ at each n. $\pi'(z_n)$ can be approximated by taking the first difference:

$$\pi'(z_n) \approx \frac{\Delta \pi_n}{\Delta z_n}. \tag{C.1}$$

With $\pi'(z_n)$ we have all the information we need to find the optimal tax schedule using eq. (4.38).

Appendix D

Appendix to Chapter 5

D.1 Summary Statistics for the Unfiltered Sample

Table D.1: Summary Statistics for the Unfiltered Sample

Variable	Pre-reform	(1995-1999)	Postreform	(2001-2004)
	Mean	Mean Std	Mean	Mean Std
Single	0.228	0.419	0.147	0.354
Couple	0.372	0.483	0.341	0.242
Single with child	0.041	0.198	0.040	0.488
Couple with child	0.359	0.479	0.472	0.081
Nr Children<18	0.743	1.066	1.016	1.190
Nr Household Members	2.567	1.336	3.142	1.423
Age	40.980	12.913	42.294	11.299
Wealth	81043.710	135244.035	136714.048	273311.630
Share Financial Wealth	0.566	0.422	0.421	0.362
Primary Household Labor Income	35518.320	27199.737	47129.161	35455.063
Effective Wealth Tax Rate	0.003	0.004	0.004	0.005
Marginal Income Tax Rate	0.416	0.112	0.361	0.064
Nr of observations	42,595		57,558	

Note: Summary statistics of the unfiltered sample. All monetary values are expressed in 1999 euros. Post-reform returns are calculated under the assumption that before-tax returns remained equal, such that only the tax rate changes. Pre-reform number of observations are taken in 1999, post-reform in 2004.

D.2 Regression Tables with All Covariates

Table D.2: Long-Run Effects on Portfolio Composition

	(1) IV	(2) IV	(3) IV	(4) IV	(5) IV	(6) IV
Rel. Ch. R^F	2.715***	2.606***	3.645***	2.696***	3.774***	4.159***
	(0.856)	(0.884)	(0.751)	(0.883)	(0.749)	(0.741)
Rel. Ch. R^H	-0.00622***	-0.0113***	-0.0174**	-0.0112***	-0.0174**	-0.0178**
	(0.00210)	(0.00245)	(0.00717)	(0.00242)	(0.00717)	(0.00743)
Log Savings 1999	-0.0944***	-0.0845***		-0.0830***		
	(0.0311)	(0.0314)		(0.0313)		
Log Wealth 1999	-0.0329	0.134***		0.139***		
	(0.0525)	(0.0185)		(0.0185)		
Primary Labor Income	0.0131***	0.0120***	0.0121***			
	(0.00312)	(0.00315)	(0.00314)			
2nd Decile Wealth	-2.498***					
	(0.469)					
3rd Decile Wealth	-1.229***					
	(0.239)					
4th Decile Wealth	-0.490***					
	(0.152)					
5th Decile Wealth	-0.185*					
	(0.110)					
6th Decile Wealth	0.0365					
	(0.0838)					
7th Decile Wealth	0.0639					
	(0.0683)					
8th Decile Wealth	0.117**					
	(0.0548)					
9th Decile Wealth	0.122***					
	(0.0425)					
Couple	-0.0680	-0.0388	-0.0734	-0.00315	-0.0377	
	(0.0508)	(0.0511)	(0.0498)	(0.0503)	(0.0491)	
Single with Child	-0.0372	-0.0491	-0.147	-0.0750	-0.171	
	(0.219)	(0.215)	(0.216)	(0.214)	(0.215)	
Couple with Child	-0.156**	-0.131**	-0.193***	-0.0961	-0.159***	
	(0.0607)	(0.0610)	(0.0587)	(0.0605)	(0.0582)	
Nr Children< 18	-0.0855***	-0.0793***	-0.0810***	-0.0840***	-0.0854***	
	(0.0242)	(0.0244)	(0.0236)	(0.0243)	(0.0236)	
Nr Household Members	0.103***	0.0999***	0.111***	0.103***	0.115***	
	(0.0176)	(0.0178)	(0.0174)	(0.0177)	(0.0173)	
Constant	-0.854	-2.692***	-1.464***	-2.028***	-0.737*	-0.0559***
	(0.785)	(0.464)	(0.456)	(0.434)	(0.418)	(0.0133)
Observations	12261	12261	13885	12261	13885	13885
R-squared	0.036	0.022	0.018	0.021	0.017	0.006

Note: Dependent variable is relative change in the share of financial wealth between 1999-2004. IV-estimates using instrumented tax rates. Splines are linear decile spline terms over the wealth distribution. Primary labor income is a term containing the log of the sum of primary labor income earned between 1999-2004. Age dummies for the primary earner were included in the regression $***p < 0.01, **p < 0.05, *p < 0.1$.

Table D.3: Long-Run Effects on Wealth Accumulation

	(1) IV-IV	(2) IV-IV	(3) IV-IV	(4) IV-IV	(5) IV-IV	(6) IV-IV
Rel. Ch. R	3.031***	2.775***	4.256***	2.808***	4.257***	4.318***
	(0.431)	(0.353)	(0.429)	(0.355)	(0.429)	(0.432)
Savings 1999	-0.0959***	-0.0957***		-0.0941***		
	(0.0121)	(0.0120)		(0.0121)		
Wealth 1999	-0.210***	-0.263***		-0.257***		
	(0.0273)	(0.00977)		(0.00971)		
Primary Labor Income	0.00845***	0.00898***	0.00240*			
	(0.00118)	(0.00120)	(0.00125)			
2nd Decile Wealth	-1.252**					
	(0.590)					
3rd Decile Wealth	0.122					
	(0.127)					
4th Decile Wealth	0.240***					
	(0.0739)					
5th Decile Wealth	0.0846*					
	(0.0505)					
6th Decile Wealth	0.0113					
	(0.0386)					
7th Decile Wealth	-0.0181					
	(0.0306)					
8th Decile Wealth	-0.0207					
	(0.0237)					
9th Decile Wealth	-0.00274					
	(0.0169)					
Couple	-0.0241	-0.0329*	0.00254	-0.00631	0.00956	
	(0.0181)	(0.0184)	(0.0200)	(0.0182)	(0.0199)	
Single with Child	0.0657	0.0676	0.0954*	0.0482	0.0900*	
	(0.0514)	(0.0521)	(0.0546)	(0.0512)	(0.0545)	
Couple with Child	-0.0362*	-0.0427*	0.0228	-0.0166	0.0295	
	(0.0218)	(0.0220)	(0.0241)	(0.0219)	(0.0240)	
Nr Children< 18	-0.0362***	-0.0389***	-0.0388***	-0.0424***	-0.0398***	
	(0.00842)	(0.00852)	(0.00898)	(0.00848)	(0.00893)	
Nr Household Members	0.0415***	0.0421***	0.0255***	0.0446***	0.0263***	
	(0.00621)	(0.00628)	(0.00662)	(0.00622)	(0.00657)	
Constant	2.255***	2.838***	0.271*	3.320***	0.416***	0.491***
	(0.388)	(0.203)	(0.158)	(0.200)	(0.142)	(0.00852)
Observations	12261	12261	12261	12261	12261	12261
R-squared	0.356	0.343	0.230	0.340	0.230	0.218

Note: Dependent variable is relative change in wealth between 1999-2004. IV-estimates using instrumented tax rates and portfolio shares. Splines are linear decile spline terms over the wealth distribution. Primary labor income is a term containing the log of the sum of primary labor income earned between 1999-2004. Age dummies for the primary earner were included in the estimation. $***p < 0.01, **p < 0.05, *p < 0.1$.

Table D.4: Short-Run Effects on Portfolio Composition

	(1)	(2)	(3)	(4)
	¡65	wtax	Singles	Incl. outliers
Rel. Ch. R	2.722***	11.51***	9.826***	2.739***
	(0.406)	(0.456)	(1.029)	(0.407)
Savings 1999	-0.0922***	-0.143***	-0.0723**	-0.0917***
	(0.0100)	(0.0161)	(0.0289)	(0.00996)
Wealth 1999	-0.199***	-0.161***	-0.224***	-0.187***
	(0.0216)	(0.0187)	(0.0713)	(0.0194)
Primary Labour Income	0.00941***	0.00416**	0.0133**	0.00901***
	(0.00218)	(0.00192)	(0.00636)	(0.00194)
1st Decile Wealth 1999	-3.806***		-19.24***	-3.776***
	(0.877)		(2.263)	(0.882)
2nd Decile Wealth	-1.135**		-4.360***	-1.098**
	(0.483)		(0.760)	(0.485)
3rd Decile Wealth	0.0498		-0.741**	0.0860
	(0.115)		(0.368)	(0.112)
4th Decile Wealth	0.193***		-0.00663	0.222***
	(0.0584)		(0.189)	(0.0547)
5th Decile Wealth	0.0770*		0.143	0.0979***
	(0.0398)		(0.142)	(0.0364)
6th Decile Wealth	-0.0186	-0.0744**	0.0246	-0.00346
	(0.0307)	(0.0305)	(0.107)	(0.0281)
7th Decile Wealth	-0.0541**	-0.0583***	-0.0269	-0.0426*
	(0.0243)	(0.0217)	(0.0874)	(0.0224)
8th Decile Wealth	-0.0399**	-0.0254	0.0167	-0.0299*
	(0.0193)	(0.0173)	(0.0727)	(0.0179)
9th Decile Wealth	-0.0201	-0.00846	-0.0148	-0.0171
	(0.0138)	(0.0124)	(0.0541)	(0.0131)
Couple	0.0183	0.0140		0.0147
	(0.0161)	(0.0157)		(0.0157)
Single with Child	0.0449	0.0332		0.0380
	(0.0392)	(0.0476)		(0.0389)
Couple with Child	-0.0336*	-0.0290		-0.0364*
	(0.0190)	(0.0185)		(0.0188)
Nr Children<18	-0.0321***	-0.0283***		-0.0318***
	(0.00704)	(0.00698)		(0.00701)
Nr Household Members	0.0433***	0.0406***	0.169***	0.0431***
	(0.00544)	(0.00534)	(0.0237)	(0.00539)
Constant	2.250***	1.919***	1.990*	2.114***
	(0.309)	(0.270)	(1.020)	(0.283)
Observations	15144	10729	1001	15487
R-squared	0.347	0.270	0.593	0.346

Note: Dependent variable is relative change in the share of financial wealth between 1999-2001. IV-estimates using instrumented tax rates. Splines are linear decile spline terms over the wealth distribution. Primary labor income is a term containing the log of the sum of primary labor income earned between 1999-2001. Age dummies for the primary earner were included in the regression $***p < 0.01, **p < 0.05, *p < 0.1$.

Table D.5: Short-Run Effects on Wealth Accumulation

	(1) IV-IV	(2) IV-IV	(3) IV-IV	(4) IV-IV	(5) IV-IV	(6) IV-IV
Rel. Ch. R	2.739***	3.491***	3.491***	2.365***	3.491***	3.548***
	(0.407)	(0.366)	(0.366)	(0.305)	(0.366)	(0.368)
Savings 1999	-0.0917***			-0.0895***		
	(0.00996)			(0.00989)		
Wealth 1999	-0.187***			-0.239***		
	(0.0194)			(0.00854)		
Primary Labour Income	0.00901***		0.000427			
	(0.00194)		(0.00202)			
1st Decile Wealth 1999	-3.776***					
	(0.882)					
2nd Decile Wealth	-1.098**					
	(0.485)					
3rd Decile Wealth	0.0860					
	(0.112)					
4th Decile Wealth	0.222***					
	(0.0547)					
5th Decile Wealth	0.0979***					
	(0.0364)					
6th Decile Wealth	-0.00346					
	(0.0281)					
7th Decile Wealth	-0.0426*					
	(0.0224)					
8th Decile Wealth	-0.0299*					
	(0.0179)					
9th Decile Wealth	-0.0171					
	(0.0131)					
Couple	0.0147	0.0428**	0.0423**	0.0251	0.0428**	
	(0.0157)	(0.0178)	(0.0180)	(0.0162)	(0.0178)	
Single with Child	0.0380	0.0868**	0.0872**	0.0450	0.0868**	
	(0.0389)	(0.0431)	(0.0431)	(0.0396)	(0.0431)	
Couple with Child	-0.0364*	0.0126	0.0120	-0.0282	0.0126	
	(0.0188)	(0.0214)	(0.0215)	(0.0194)	(0.0214)	
Nr Children< 18	-0.0318***	-0.0266***	-0.0264***	-0.0372***	-0.0266***	
	(0.00701)	(0.00753)	(0.00757)	(0.00710)	(0.00753)	
Nr Household Members	0.0431***	0.0235***	0.0234***	0.0466***	0.0235***	
	(0.00539)	(0.00566)	(0.00573)	(0.00539)	(0.00566)	
Constant	2.114***	0.228*	0.215	2.980***	0.228*	0.326***
	(0.283)	(0.135)	(0.148)	(0.167)	(0.135)	(0.00705)
Observations	15487	15487	15487	15487	15487	15487
R-squared	0.346	0.222	0.222	0.326	0.222	0.211

Note: Dependent variable is relative change in wealth between 1999-2004. IV-estimates using instrumented tax rates and portfolio shares. Splines are linear decile spline terms over the wealth distribution. Primary labor income is a term containing the log of the sum of primary labor income earned between 1999-2001. Age dummies for the age of the primary earner were included in the estimation $***p < 0.01, **p < 0.05, *p < 0.1$.

Table D.6: Long-Run Effects on Portfolio Composition for Different Specifications and Subsamples

	(1) ¡65	(2) wtax	(3) Singles	(4) Incl. outliers
Rel. Ch. R^F	2.786***	3.941***	6.438**	0.255***
	(0.865)	(1.102)	(3.168)	(0.0582)
Rel. Ch. R^H	-0.00621***	-7.343***	-2.843**	-0.00613***
	(0.00209)	(1.730)	(1.251)	(0.00157)
Log Savings 1999	-0.0964***	-0.187***	0.135	-0.0823**
	(0.0316)	(0.0452)	(0.116)	(0.0321)
Log Wealth 1999	-0.0304	-0.159***	0.0221	-0.0195
	(0.0584)	(0.0488)	(0.161)	(0.0538)
Primary Labour Income	0.0124***	0.0172***	0.00885	0.0139***
	(0.00365)	(0.00325)	(0.00950)	(0.00315)
1st Decile Wealth 1999				
0				
2nd Decile Wealth	-2.494***			-2.157***
	(0.482)			(0.465)
3rd Decile Wealth	-1.228***		0.354	-1.192***
	(0.253)		(0.694)	(0.241)
4th Decile Wealth	-0.492***		0.428	-0.408***
	(0.165)		(0.476)	(0.155)
5th Decile Wealth	-0.188		0.274	-0.124
	(0.119)		(0.372)	(0.112)
6th Decile Wealth	0.0341	-0.0563	0.497*	0.0625
	(0.0911)	(0.0981)	(0.296)	(0.0852)
7th Decile Wealth	0.0623	-0.0338	0.335	0.0920
	(0.0739)	(0.0664)	(0.236)	(0.0694)
8th Decile Wealth	0.112*	0.0306	0.287	0.125**
	(0.0588)	(0.0532)	(0.200)	(0.0556)
9th Decile Wealth	0.117***	0.0561	0.415***	0.134***
	(0.0450)	(0.0418)	(0.158)	(0.0429)
Couple	-0.0857*	0.0137		-0.0640
	(0.0519)	(0.0527)		(0.0518)
Single with Child	-0.0527	-0.151		-0.0692
	(0.219)	(0.310)		(0.215)
Couple with Child	-0.170***	-0.0544		-0.146**
	(0.0614)	(0.0637)		(0.0620)
Nr Children< 18	-0.0881***	-0.0647**		-0.0813***
	(0.0244)	(0.0253)		(0.0245)
Nr Household Members	0.105***	0.0884***	0.0622	0.0984***
	(0.0178)	(0.0175)	(0.0773)	(0.0177)
Constant	-0.829	0.343	-3.569	-0.930
	(0.858)	(0.795)	(2.421)	(0.789)
Observations	11903	8625	664	12510
R-squared	0.035	0.061	0.157	0.038

Note: Dependent variable is relative change in the share of financial wealth between 1999-2004. IV-estimates using instrumented tax rates. Splines are linear decile spline terms over the wealth distribution. Primary labor income is a term containing the log of the sum of primary labor income earned between 1999-2004. Age dummies for age of the primary earner were included in the estimation. Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.

Table D.7: Long-Run Effects on Wealth Accumulation for Different Specifications and Subsamples

	(1) ¡65	(2) wtax	(3) Singles	(4) Incl. outliers	(5) IV-OLS
Rel. Ch. R	3.008***	9.915***	7.822***	3.031***	1.091***
	(0.430)	(0.471)	(1.012)	(0.431)	(0.257)
Savings 1999	-0.0967***	-0.136***	-0.0676	-0.0959***	-0.115***
	(0.0122)	(0.0204)	(0.0417)	(0.0121)	(0.0122)
Wealth 1999	-0.215***	-0.191***	-0.225***	-0.210***	-0.418***
	(0.0306)	(0.0304)	(0.0812)	(0.0273)	(0.0302)
Primary Labour Income	0.00954***	0.00470***	0.0114***	0.00845***	0.0108***
	(0.00139)	(0.00119)	(0.00398)	(0.00118)	(0.00129)
1st Decile Wealth 1999					2.099***
					(0.227)
2nd Decile Wealth	-1.247**			-1.252**	1.299***
	(0.588)			(0.590)	(0.171)
3rd Decile Wealth	0.117		-0.282	0.122	0.435***
	(0.134)		(0.353)	(0.127)	(0.117)
4th Decile Wealth	0.235***		0.0739	0.240***	-0.0131
	(0.0803)		(0.211)	(0.0739)	(0.0800)
5th Decile Wealth	0.0806		0.0640	0.0846*	-0.238***
	(0.0559)		(0.153)	(0.0505)	(0.0566)
6th Decile Wealth	0.00800	-0.0899**	0.00524	0.0113	-0.262***
	(0.0427)	(0.0439)	(0.122)	(0.0386)	(0.0431)
7th Decile Wealth	-0.0214	-0.0451	-0.0149	-0.0181	-0.235***
	(0.0337)	(0.0337)	(0.0978)	(0.0306)	(0.0340)
8th Decile Wealth	-0.0224	-0.0234	0.0655	-0.0207	-0.186***
	(0.0259)	(0.0257)	(0.0813)	(0.0237)	(0.0263)
9th Decile Wealth	-0.00187	-0.00300	0.0153	-0.00274	-0.107***
	(0.0181)	(0.0177)	(0.0644)	(0.0169)	(0.0184)
Couple	-0.0256	-0.0167		-0.0241	-0.0695***
	(0.0188)	(0.0197)		(0.0181)	(0.0206)
Single with Child	0.0691	0.0611		0.0657	0.0224
	(0.0516)	(0.0659)		(0.0514)	(0.0614)
Couple with Child	-0.0390*	-0.0523**		-0.0362*	-0.0886***
	(0.0222)	(0.0238)		(0.0218)	(0.0242)
Nr Children< 18	-0.0359***	-0.0223**		-0.0362***	-0.0460***
	(0.00849)	(0.00895)		(0.00842)	(0.00925)
Nr Household Members	0.0416***	0.0410***	0.166***	0.0415***	0.0536***
	(0.00631)	(0.00652)	(0.0290)	(0.00621)	(0.00684)
Constant	2.241***	2.511***	1.700	2.255***	4.683***
	(0.428)	(0.423)	(1.076)	(0.388)	(0.437)
Observations	11903	8625	664	12261	12831
R-squared	0.356	0.250	0.558	0.356	0.613

Note: Dependent variable is relative change in wealth between 1999-2004. IV-estimates using instrumented tax rates and portfolio shares except in final column which only uses instrumented tax rates. Splines are linear decile spline terms over the wealth distribution. Primary labor income is a term containing the log of the sum of primary labor income earned between 1999-2004. Age dummies for the age of the primary earner were included in the estimation. Robust standard errors in parentheses. $***p < 0.01, **p < 0.05, *p < 0.1$.

Appendix E

Appendix to Chapter 6

E.1 Proof of Proposition 5

Proof. The first order condition for incentive compatibility is given by:

$$\begin{aligned} \mathbf{0}_p &= \frac{\partial u\left(\mathbf{x}^*\left(\mathbf{m}\right), y^*\left(\mathbf{m}\right), \mathbf{n}\right)}{\partial \mathbf{m}}|_{\mathbf{m}=\mathbf{n}}, \\ &= \mathbf{x}^{*\prime}\left(\mathbf{n}\right)^T u_{\mathbf{x}}\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right) + u_y\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right) y^{*\prime}\left(\mathbf{n}\right)^T, \end{aligned} \tag{E.1}$$

where $\mathbf{0}_p$ denotes a p-column vector of zeros. This can be rewritten to:

$$y^{*\prime}\left(\mathbf{n}\right) = s\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right)^T \mathbf{x}^{*\prime}\left(\mathbf{n}\right),$$

proving equation (6.6).

We can derive (6.8) from (6.5) using the envelope theorem:

$$\begin{aligned} V'\left(\mathbf{n}\right) &= \mathbf{x}^{*\prime}\left(\mathbf{n}\right)^T u_{\mathbf{x}} + u_y y^*\left(\mathbf{n}\right)^T + u_{\mathbf{n}}^T, \\ V'\left(\mathbf{n}\right) &= u_{\mathbf{n}}\left(\mathbf{x}^*\left(\mathbf{n}\right), y^*\left(\mathbf{n}\right), \mathbf{n}\right)^T, \end{aligned} \tag{E.2}$$

where the latter equality follows from the first-order conditions.

The second-order conditions of a maximum are :

$$\frac{\partial^2 u\left(\mathbf{x}^*\left(\mathbf{m}\right), y^*\left(\mathbf{m}\right), \mathbf{n}\right)}{\partial \mathbf{m}^2}|_{\mathbf{m}=\mathbf{n}} \lessdot 0, \tag{E.3}$$

Where $\prec 0$ denotes the negative definiteness of the matrix.
Taking the derivative of (E.2) with respect to $\mathbf{m}$ gives:

$$\begin{aligned}\frac{\partial^2 u\left(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n}\right)}{\partial \mathbf{m}^2} &= \left(u_{\mathbf{x}}\left(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n}\right)^T \otimes I_p\right) \mathbf{x}^{*\prime\prime}(\mathbf{m}) \\ &\quad + \mathbf{x}^{*\prime}(\mathbf{m})^T u_{\mathbf{xx}}\left(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n}\right) \mathbf{x}^{*\prime}(\mathbf{m}) \\ &\quad + u_{yy}\left(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n}\right) y^{*\prime}(\mathbf{m}) y^{*\prime}(\mathbf{m})^T \\ &\quad + u_y\left(\mathbf{x}^*(\mathbf{m}), y^*(\mathbf{m}), \mathbf{n}\right) y^{*\prime\prime}(\mathbf{m}). \end{aligned} \tag{E.4}$$

where $\otimes$ denotes the Kronecker product.

To simplify this expression we take the total derivative of the first order condition (E.1):

$$D_{\mathbf{n}} \mathbf{0}_p = D_{\mathbf{n}}\left[\mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{x}}\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right) + u_y\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right) y^{*\prime}(\mathbf{n})^T\right] \tag{E.5}$$

$$\begin{aligned}\mathbf{0}_{p\times p} &= \left(u_{\mathbf{x}}\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right)^T \otimes I_p\right) \mathbf{x}^{*\prime\prime}(\mathbf{n})^T + \mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{xx}}\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right) \mathbf{x}^{*\prime}(\mathbf{n}) + \\ &\quad u_{yy}\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right) y^{*\prime}(\mathbf{n}) y^{*\prime}(\mathbf{n})^T + u_y\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right) y^{*\prime\prime}(\mathbf{n}) + \\ &\quad \mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{xn}}\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right) + y^{*\prime}(\mathbf{n})^T u_{y\mathbf{n}}\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right), \end{aligned} \tag{E.6}$$

Now combine equations (E.5,E.3,E.4) to get the following expression:

$$\begin{aligned} 0 &\succ \left(u_{\mathbf{x}}\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right)^T \otimes I_p\right) \mathbf{x}^{*\prime\prime}(\mathbf{n})^T + \mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{xx}}\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right) \mathbf{x}^{*\prime}(\mathbf{n}) + \\ &\quad u_{yy}\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right) y^{*\prime}(\mathbf{n}) y^{*\prime}(\mathbf{n})^T + u_y\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right) y^{*\prime\prime}(\mathbf{n}) - \mathbf{0}_{p\times p} \\ 0 &\succ \mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{xn}}\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right) + y^{*\prime}(\mathbf{n})^T u_{y\mathbf{n}}\left(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}\right). \end{aligned} \tag{E.7}$$

Then partially differentiate the vector of shadow prices with respect to $\mathbf{n}$ to get:

$$\begin{aligned} s_{\mathbf{n}} &= \frac{-u_{\mathbf{xn}} u_y + u_{\mathbf{x}} u_{y\mathbf{n}}}{\left(u_y\right)^2} \\ u_{\mathbf{xn}} &= -s_{\mathbf{n}} u_y - s u_{y\mathbf{n}}, \end{aligned}$$

and substitute this result and (6.6) into (E.7) to yield

$$\begin{aligned} 0 &\succ \mathbf{x}^{*\prime}(\mathbf{n})^T \left(-s_{\mathbf{n}} u_y - s u_{y\mathbf{n}}\right) + y^{*\prime}(\mathbf{n})^T u_{y\mathbf{n}} \\ 0 &\prec \mathbf{x}^{*\prime}(\mathbf{n})^T s_{\mathbf{n}} u_y \Leftrightarrow \\ 0 &\prec \mathbf{x}^{*\prime}(\mathbf{n})^T s_{\mathbf{n}}, \end{aligned}$$

where the final inequality, equation (6.7), follows from the fact that $u_y > 0$.

An equivalent expression can be derived by totally differentiating, equation (E.2) with respect to $\mathbf{n}$:

$$\begin{aligned} V''(\mathbf{n}) &= Du_{\mathbf{n}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) \\ &= \mathbf{x}^{*\prime}(\mathbf{n})^T u_{\mathbf{xn}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) + y^{*\prime}(\mathbf{n})^T u_{y\mathbf{n}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) \\ &\quad + u_{\mathbf{nn}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) \end{aligned}$$

Now combine this last expression with (E.7) to get the final equation:

$$V''(\mathbf{n}) - u_{\mathbf{nn}}(\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n}), \mathbf{n}) \gg 0.$$

■

E.2 Proof of Proposition 6

Proof. Starting from the first order conditions (6.15), (6.14) and (6.16). First solving (6.14) for yields η:

$$\eta = \frac{\lambda f + u_{y\mathbf{n}}\theta}{u_y}.$$

Now substitute this expression into (6.15) and simplify to get the desired equation:

$$\begin{aligned} \lambda q'^T f - u_{\mathbf{xn}}\theta + \frac{\lambda f + u_{y\mathbf{n}}\theta}{u_y} u_{\mathbf{x}} &= \mathbf{0}_k, \\ \lambda q_{x_i} f - u_{x_i\mathbf{n}}\theta + \frac{\lambda f + u_{y\mathbf{n}}\theta}{u_y} u_{x_i} &= \mathbf{0}_k, \\ \lambda f (q_{x_i} - s_i) &= u_{x_i\mathbf{n}}\theta + u_{y\mathbf{n}}\theta\mathbf{s}, \\ q_{x_i} - s_i &= (u_{x_i\mathbf{n}} + u_{y\mathbf{n}} s_i)\frac{\theta}{\lambda f}, \\ q_{x_i} - s_i &= \sum_{j=1}^{p} -\frac{\partial s_i}{\partial n_j}\theta_j \frac{u_y}{\lambda f}, \\ \frac{q_{x_i} - s_i}{s_i} &= \sum_{j=1}^{p} \varepsilon_{x_i n_j} \times \theta_j \frac{u_y}{\lambda} \times \frac{1}{n_j f}. \end{aligned}$$

■

E.3 Proof of Proposition 10

Proof. Proposition 10 consists of 2 statements. First, we prove that the separating partition $\mathbf{N}_S$ is either convex or empty.

Define the types α, β, γ s.t. $\alpha, \gamma \in \mathbf{N}_S$ and $\exists k \in (0,1)$ s.t. $(1-k)\,\alpha + k\gamma = \beta$. Denote by $\{\widetilde{x}, \widetilde{y}\}$ the solution to the full problem, and by $u\left(\{\overline{x}, \overline{y}\}, \mathbf{n}\right) = \max_{\{x,y\}} u(x, y, \mathbf{n} | \{\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}\}^{\leftarrow} \subseteq \mathbf{N}_B)$ the bunching choice that delivers type $\mathbf{n}$ the highest utility from the set of bunching allocations.
Then define the function $L(\mathbf{n}) = u\left(\{\mathbf{x}^*(\mathbf{n}), y^*(\mathbf{n})\}, \mathbf{n}\right) - u\left(\{\overline{x}, \overline{y}\}, \mathbf{n}\right)$. By individual rationality we know that $0 < L(\alpha), L(\gamma)$. Equation (6.9) implies that L is convex and monotone, and continuity of u implies L is piece-wise continuous. For any $0 < k < 1$ it must be then be the case that $0 < L(\ \beta)$, and β must also be part of the separating set.

Second we prove that bunching occurs below the boundary $\partial N_S \cap \partial N_B$, and only one such boundary exists.

Since $V(\mathbf{n})$ is continuous, individual rationality requires any type at $\mathbf{b} \in \partial N_S \cap \partial N_B$, $u\left(\{\mathbf{x}^*(\mathbf{b}), y^*(\mathbf{b})\}, \mathbf{b}\right) = u\left(\{\overline{x}, \overline{y}\}, \mathbf{b}\right)$. While on the separating set equation (6.8) guarantees that the first derivative of u has to be equal to the first-order derivative of v. By equation (6.9), however, the second-order derivative on the optimal allocation has to be higher than the second derivative of the utility function on $\{\overline{x}, \overline{y}\}$, such that the utility profile of the separating and the bunching partition cross only once for each type, and they cross at $\{\mathbf{x}^*(\mathbf{b}), y^*(\mathbf{b})\}$ for type $\mathbf{b}$. The single-crossing property (6.7) then implies that this is the unique border and that separation occurs on side of the border where types have relatively high utility.
This can easily be seen if we assume $\frac{\partial u}{\partial n_i} > 0 \quad \forall i$. In that case, for any type $\mathbf{g}$, with $b_j \leq g_j \; j \in \{1, .., p\}$ and at least one inequality strict, the utility profile of the optimal allocation, $V(\mathbf{n})$, has to be higher than $u\left(\{\overline{x}, \overline{y}\}, \mathbf{g}\right)$. For any type $\mathbf{g}$, therefore, (6.8) holds and the allocation found above the boundary $\partial N_S \cap \partial N_B$ induces separation. Simultaneously, below the boundary equation (6.9) cannot hold, which (together with (6.8) and continuity of V) implies that bunching yields a higher utility, such that bunching occurs there. ■

Appendix F

Appendix to Chapter 7

F.1 Proofs

For bookkeeping, the Jacobian of first derivatives $\phi'(\cdot)$ of any function $\phi(\cdot) : \mathcal{R}^a \to \mathcal{R}^b$, is of dimension $b \times a$, while the second derivatives $\phi''(\cdot)$ are of dimension $ab \times a$. For any multi-vector functions $\psi(\mathbf{z}^1, \mathbf{z}^2, \ldots) : \mathcal{R}^{a^1} \times \mathcal{R}^{a^2} \ldots \to \mathcal{R}$ the vector of first derivatives ψ_{z^i} are of dimension $a^i \times 1$ and the matrix of second derivatives $\psi_{z^i z^j}$ are of dimension $a^i \times a^j$ where the dimension of the matrix follows the order of the subscripts. In addition, let superscript T be the transpose operator. Vectors and multi-dimensional constructs are denoted in bold, scalars are in normal font.

F.1.1 Proof of Lemma 4

Proof. Due to non-satiation of the utility function we know that the budget constraint will hold with equality such that we know that:

$$y^*(\mathbf{n}) = q(\mathbf{x}^*(\mathbf{n})) - T(\mathbf{x}^*(\mathbf{n}))$$

Direct substitution of the budget constraint into the utility function allows us to write the first-order conditions to problem (12) as:

$$\mathbf{0} = u_{\mathbf{x}} + (q' - T')^T u_y \tag{F.1}$$

which directly implies equations (7.7) and (7.8).

Now take the second-order derivative of the utility function with respect to $\mathbf{x}$ to get the second-order conditions:

$$u_{\mathbf{xx}}+\left(2u_{\mathbf{x}y}+u_{yy}\left(q'\left(\mathbf{x}^*\right)-T'\left(\mathbf{x}^*\right)\right)^T\right)\left(q'\left(\mathbf{x}^*\right)-T'\left(\mathbf{x}^*\right)\right)+u_y\left(q''\left(\mathbf{x}^*\right)-T''\left(\mathbf{x}^*\right)\right)\lessdot 0 \quad \text{(F.2)}$$

Differentiate the marginal rate of substitution, $\mathbf{s}$, to $\mathbf{x}$ using the definition of $\mathbf{s}$ and using the implicit function theorem to define $y\left(u,\mathbf{x},\mathbf{n}\right)$:

$$\frac{\partial \mathbf{s}\left(\mathbf{x},y\left(u,\mathbf{x},\mathbf{n}\right),\mathbf{n}\right)}{\partial \mathbf{x}}=-\frac{\left(u_{\mathbf{xx}}+2u_{\mathbf{x}y}\mathbf{s}^T\right)-u_{yy}\mathbf{s}\mathbf{s}^T}{u_y} \quad \text{(F.3)}$$

Now combining (7.8) with (F.3) allows us to simplify (F.2) and obtain the final condition:

$$\begin{aligned}-\left(\frac{\partial \mathbf{s}\left(\mathbf{x},y\left(u,\mathbf{x},\mathbf{n}\right),\mathbf{n}\right)}{\partial \mathbf{x}}+q''\left(\mathbf{x}^*\right)-T''\left(\mathbf{x}^*\right)\right)u_y &\lessdot 0\Leftrightarrow\\ -\frac{\partial \mathbf{s}\left(\mathbf{x},y\left(u,\mathbf{x},\mathbf{n}\right),\mathbf{n}\right)}{\partial \mathbf{x}}+q''\left(\mathbf{x}^*\right)-T''\left(\mathbf{x}^*\right) &\lessdot 0\end{aligned}$$

where the final step follows from the assumption that $u_y>0$. ■

F.1.2 Proof of Proposition 14

Suppose on the contrary that (7.9) is not satisfied for some agent of type $\mathbf{n}$. Consider a deviation from the second-best allocation $\alpha\Delta\mathbf{x}$ where $\alpha>0$ and $\Delta\mathbf{x}$ is a $k\times 1$ vector with length one. The utility gain of such a deviation can be approximated by a second-order Taylor expansion:

$$\begin{aligned}u\left(\mathbf{x}^*\left(\mathbf{n}\right)+\alpha\boldsymbol{\Delta}\mathbf{x},q\left(\mathbf{x}^*\left(\mathbf{n}\right)+\alpha\Delta\mathbf{x}\right)-T\left(\mathbf{x}^*\left(\mathbf{n}\right)+\alpha\Delta\mathbf{x}\right),\mathbf{n}\right)-u^* &=\\ \left(u_{\mathbf{x}}^T+u_y\left(q'-T'\right)\right)\alpha\boldsymbol{\Delta}\mathbf{x} &+\\ \frac{1}{2}\alpha^2\boldsymbol{\Delta}\mathbf{x}^T\left(u_{\mathbf{xx}}+\left(2u_{\mathbf{x}y}+u_{yy}\left(q'-T'\right)^T\right)\left(q'-T'\right)+u_y\left(q''-T''\right)\right)\boldsymbol{\Delta}\mathbf{x} &=\\ \frac{1}{2}\alpha^2\boldsymbol{\Delta}\mathbf{x}^T\left(u_{\mathbf{xx}}+\left(2u_{\mathbf{x}y}+u_{yy}\left(q'-T'\right)^T\right)\left(q'-T'\right)+u_y\left(q''-T''\right)\right)\boldsymbol{\Delta}\mathbf{x} &=\\ \frac{1}{2}\alpha^2u_y\boldsymbol{\Delta}\mathbf{x}^T\left(-\frac{\partial \mathbf{s}\left(\mathbf{x},y\left(u,\mathbf{x},\mathbf{n}\right),\mathbf{n}\right)}{\partial \mathbf{x}}+q''\left(\mathbf{x}^*\right)-T''\left(\mathbf{x}^*\right)\right)\boldsymbol{\Delta}\mathbf{x}&\end{aligned}$$

where the first order terms equal zero by the assumption that taxes are equated to wedges, $T_i'\left(\mathbf{x}^*\left(\mathbf{n}\right)\right)=\mathcal{W}_i(\mathbf{n})$. Due to symmetry of the matrix of second order conditions for sufficiently small α the deviation strategy $\alpha\Delta\mathbf{x}$ and $-\alpha\Delta\mathbf{x}$ yield the same utility. In addition, if:

$$-\frac{\partial \mathbf{s}\left(\mathbf{x},y\left(u,\mathbf{x},\mathbf{n}\right),\mathbf{n}\right)}{\partial \mathbf{x}}+q''\left(\mathbf{x}^*\right)-T''\left(\mathbf{x}^*\right)$$

is not negative semi-definite there is at least one deviation strategy $\Delta\hat{\mathbf{x}}$ which yields a positive utility gain. The change in tax revenue due to such a deviation can also be found by means of a second-order Taylor expansion:

$$T\left(\mathbf{x}^*\left(\mathbf{n}\right)+\alpha\Delta\hat{\mathbf{x}}\right)-T\left(\mathbf{x}^*\left(\mathbf{n}\right)\right)\approx \alpha T'\Delta\hat{\mathbf{x}}+\frac{1}{2}\alpha^2\Delta\hat{\mathbf{x}}^T T''\Delta\hat{\mathbf{x}}.$$

The first-order term will always be non-negative for either strategy $-\Delta\hat{\mathbf{x}}$ or $\Delta\hat{\mathbf{x}}$. If for either choice it is positive, the first-order term dominates the second-order term for sufficiently small α and hence, the deviation results in higher tax revenue. If the first-order term is zero, we need to consider the second-order term. If it is negative apparently the tax schedule contains an internal maximum on the allocation in $\Delta\hat{\mathbf{x}}$ which violates our assumption. Therefore, if the first-order term is zero the second term must be non-negative. Hence, tax revenue always weakly increases in either $-\Delta\hat{\mathbf{x}}$ or $\Delta\hat{\mathbf{x}}$. Therefore, one of these deviations must be a Pareto improvement and we run into a contradiction. If a Pareto improvement over the allocation can be found within a particular implementation, then the original allocation could not have been second-best.

F.1.3 Proof of Proposition 16

Equations (7.7) and (7.8) uniquely define the tax schedule for $\mathbf{x}^*\left(\mathbf{n}\right)$ on its domain $\mathbf{N}$. If $\mathbf{x}^*\left(\mathbf{n}\right)$ is bijective there is an unique inverse mapping $\mathbf{n}^*\left(\mathbf{x}\right)$ for all $\mathbf{x}\in\mathbf{X}$. Therefore, equations (7.7) and (7.8) define the tax schedule for $\mathbf{x}^*\left(\mathbf{n}^*\left(\mathbf{x}\right)\right)$ on its domain $\mathbf{x}\in\mathbf{X}$. Hence, the tax schedule is defined on the entire choice space. Note that we do not need to check for second-order conditions (7.9) in this case, because we have assumed that the allocation $\mathbf{x}^*\left(\mathbf{n}\right)$ is (second-order) incentive compatible for all $\mathbf{n}\in\mathbf{N}$. Therefore, the unique tax schedule that implements this allocation must also be implementable.

F.2 Example

In figure 1 and 2, welfare function and resource constraint are equal to:

$$
\begin{aligned}
u &= log(y) - \frac{1}{1.5}\left(\frac{x_1}{n}\right)^{1.5} - \frac{1}{1.5}\left(\frac{x_2}{n}\right)^{1.5} \\
W &= \int \left[u(n) + E(x_1.x_2)\right] dF(n) \\
E(x_1, x_2) &= \frac{1}{1.5}\left(\left(\frac{x_1(n)}{n}\right)^{1.5} + \left(\frac{x_2(n)}{n}\right)^{1.5} - \left(\frac{x_1(n)}{n}\right)^{1.5} * \left(\frac{x_2(n)}{n}\right)^{1.5}\right) \\
\int_N y(n)dF(n) &= \int_N x_1(n) + x_2(n)dF(n)
\end{aligned}
$$

We assume that the type space is uni-dimensional and the types are uniformly distributed over a closed interval on the real line. The first-order approach to this problem yields the allocation shown in the figures 7.1 7.2. This second-best allocation can only be implemented by the central planner if he uses interdependencies to map out the off-allocation consumption choices/coordinates. The planner can than determine what off-allocation points have to be taxed prohibitively to ensure that each individual prefers his own bundle over any other choice.

Nederlandse Samenvatting (Summary in Dutch)

Een van de belangrijkste vraagstukken in de politiek is hoe de overheid moet herverdelen van rijk naar arm. De identiteit van politieke partijen wordt traditioneel zelfs in grote mate bepaald door de vraag hoe, en in hoeverre de overheid inkomensverschillen moet nivelleren. Enerzijds willen rechtse partijen slechts met mate herverdelen. Zij pleiten daarom voor lage belastingen en een kleine overheid. Anderzijds, willen linkse partijen juist zoveel mogelijk doen om het lot van de armsten in de maatschappij verbeteren. Financiering en organisatie van deze herverdeling vereist vanzelfsprekend een hogere belastingdruk en een grotere overheid.

In dit proefschrift wordt in zes afzonderlijke hoofdstukken het herverdelingsvraagstuk nader bestudeerd. De vraag hoeveel de overheid zou moeten herverdelen wordt hierbij niet beantwoord. De mate van herverdeling is een politieke keuze die uiteindelijk door de kiezer moet worden genomen. De kiezer kan echter pas een geïnformeerde keuze maken nadat de kosten en baten van herverdeling in kaart zijn gebracht. Is onze huidige welvaartstaat wel effectief, zorgt de progressiviteit van ons belastingstelsel ervoor dat de sterkste schouders de zwaarste lasten dragen en komen subsidies en toelages terecht bij huishoudens die dit het hardste nodig hebben? De centrale vraag in dit proefschrift is dan ook hoe de overheid zijn herverdelingsinstrumenten zo effectief mogelijk kan gebruiken, zodat herverdeling tot stand komt tegen de laagst mogelijke economische kosten.

De overheid heeft vele instrumenten voor herverdeling tot zijn beschikking. Het belangrijkste herverdelingsinstrument is natuurlijk de belasting op inkomsten uit arbeid. De progressiviteit van de inkomstenbelasting neemt traditioneel een centrale plaats in op alle politieke partijprogramma's. De overheid heeft daarnaast echter de beschikking over talloze andere instrumenten die voor herverdeling gebruikt kunnen worden zoals de belasting op kapitaalinkomsten, toelages zoals de huurtoeslag en de zorgtoeslag, de bijstand en de BTW. Helaas spelen deze instrumenten in vele landen in kakofonie. De overheid gebruikt zijn instrumenten verkeerd, en de samenklank tussen verschillende instrumenten bevat dissonanten. Zo wordt de herverdelende werking van een progressief inkomstenbelastingstelsel vaak gedeeltelijk tenietgedaan door vele aftrekposten die slechts ten goede komen aan rijke belastingbetalers, subsidies en toelages bedoeldt voor de armsten in onze maatschappij komen vaak terecht bij mensen met een gemiddeld of bovengemiddeld inkomen en voorgestelde belastingverhogingen die tot doel hebben de belastingopbrengst te verhogen werken soms averechts doordat de belastingbasis krimpt wanneer de belasting verhoogd wordt. In dit proefschrift vinden overheden dan ook belangrijke adviezen om de klank en samenklank van de verschillende instrumenten te versterken. Hierdoor kan de huidige kakofonie worden omgevormd tot een symfonie van herverdelingsinstrumenten.

Het belasting-subsidie stelsel op arbeidsinkomen in Nederland

In de eerste twee hoofdstukken van dit proefschrift wordt gekeken hoe de Nederlandse overheid met het huidige “belasting-subsidie stelsel op arbeidsinkomen” beter kan herverdelen van huishoudens met een hoog arbeidsinkomen naar huishoudens met een laag inkomen. Het belasting-subsidie stelsel op arbeidsinkomen is het stelsel van herverdelingsinstrumenten die allen een eigenschap delen: naarmate de verdiensten uit arbeid omhoog gaan, stijgt de netto betaling aan de overheid. Het belangrijkste instrument dat deze eigenschap heeft is de belasting op arbeidsinkomsten. Naarmate een huishouden meer verdient betaaldt het huishouden meer inkomstenbelasting aan de overheid. Maar daarnaast zijn ook toeslagen en subsidies vaak afhankelijk van het verdiende inkomen. Zo ontvangt een huishouden steeds minder zorgtoeslag naarmate het meer verdient. Bovendien is de betaling van indirecte belastingen zoals de BTW en accijns sterk afhankelijk van het verdiende inkomen, omdat huishoudens met hogere inkomens meer consumeren en daardoor via de indirecte belastingen meer afdragen aan de overheid. Hoeveel een huishouden met een gegeven inkomen netto meer moet betalen aan de overheid wanneer de inkomsten met een euro stijgen, wordt het tarief op arbeidsinkomen genoemd. Dit tarief is niet gelijk voor elk huishouden. Het tarief voor rijke huishoudens is relatief hoog omdat het marginale belastingpercentage dat zij betalen over arbeidsinkomen relatief hoog is. Maar ook huishoudens met lage inkomens hebben een relatief hoog tarief, omdat zij relatief veel subsidie inkomen verliezen als hun verdiensten omhoog gaan.

In het eerste hoofdstuk worden de optimale belasting-subsidie tarieven op arbeidsinkomen berekend voor Nederland. De berekening wordt gemaakt met behulp van de optimale belasting theorie. Deze theorie is in 1971 ontwikkeld door James Mirrlees en in 1997 beloond met de Nobelprijs. Centraal in deze theorie staat de afweging tussen gelijkheid en doelmatigheid. De overheid wil herverdelen van rijk naar arm, maar met deze herverdeling gaan kosten gepaard. Als tarieven te hoog zijn verliezen werknemers hun prikkel om te werken. Daardoor bieden zij minder arbeid aan of stoppen ze volledig met werken. Bovendien zorgen hoge tarieven op het verdiende inkomen ervoor dat vooral hoogopgeleide werknemers eerder naar het buitenland vertrekken of via ingewikkelde fiscale constructies de belasting op arbeidsinkomsten ontwijken. De belasting gaat daarom ten koste van de doelmatigheid in de economie. De optimale tarieven in het belasting-subsidie stelsel creëren de grootste herverdeling winst tegen de laagste doelmatigheidskosten.

De berekening van de optimale tarieven in het belasting-subsidie stelsel voor Nederland is van bijzonder belang omdat een dergelijke berekening nog nooit is gemaakt voor een Europees continentaal land. Optimale tarieven voor de Verenigde Staten en het Verenigd Koninkrijk zijn wel bekend, maar deze landen zijn nauwelijks te vergelijken met Nederland, omdat inkomensongelijkheid in de VS en het VK ontzettend hoog zijn, terwijl Nederland een van de ontwikkelde landen is met de laagste inkomensongelijkheid.

In dit hoofdstuk wordt het optimale belastingmodel bijzonder nauwkeurig gekalibreerd naar de stand van de Nederlandse economie met behulp van representatieve data over het verdiende inkomen van Nederlandse huishoudens en de beste empirische schattingen met betrekking tot het gedragseffect van een verhoging van de tarieven op het arbeidsaanbod van huishoudens. Het blijkt dat de lage inkomensongelijkheid in Nederland ervoor zorgt dat het optimale marginale tarief in de hoogste schijf van de inkomstenbelasting veel lager is dan in de VS en het VK. Er wordt aangetoond dat een verhoging van de belasting voor veelverdieners zelfs zal leiden tot een verlaging van de belastingopbrengst. Voorstellen om het belastingpercentage in de hoogste belastingschijf te verhogen hebben dan ook een averechts effect.

Aan de andere kant is het huidige marginale tarief aan de onderkant relatief laag. Dit betekent dat inkomensafhankelijke regelingen zoals de zorgtoeslag erg langzaam worden uit gefaseerd. Dit is wellicht optimaal onder de meest rechtse preferenties voor herverdeling, maar linksere partijen zullen ervoor willen zorgdragen dat er meer wordt herverdeeld aan de allerarmsten en minder aan de middeninkomens.

De grote winnaars van het huidige stelsel zijn de middeninkomens. Deze groep wordt disproportioneel bevoordeeld. De overheid zou de gemiddelde belastingdruk van deze groep dan ook moeten verhogen zodat de druk op de armste inkomens verlaagd kan worden. In dit hoofdstuk wordt ook onderzocht of de introductie van een flattax de kosten van herverdeling kan drukken. Bij een flattax wordt het tarief gelijk gesteld voor alle huishoudens. Wanneer het huidige stelsel vervangen wordt door een flattax blijkt dit inderdaad dezelfde herverdeling tot stand te kunnen brengen tegen lagere kosten. Echter, een verbetering van het huidige stelsel waarbij de tarieven gedifferentieerd zijn naar inkomen kan leiden tot een veel grotere welvaartswinst. Als zodanig is de introductie van een flattax een slecht idee, want verbeteringen waarbij de tarieven, net als in het huidige stelsel, verschillen per inkomensniveau leiden tot meer herverdeling tegen lagere kosten.

Het tweede hoofdstuk bestudeerd de herverdelingspreferenties van politieke partijen in Nederland. In de "Keuzes In Kaart" analyse van het Centraal Planbureau worden voor elke nationale verkiezing de partijprogramma's van politieke partijen doorberekend. Daardoor kan precies bepaald worden hoe politieke partijen het belasting-subsidie stelsel willen veranderen. In dit hoofdstuk worden de voorgestelde belasting-subsidie systemen van de SP, PVDA, CDA en VVD voor de nationale verkiezien in 2002 nader onderzocht met behulp van de optimale belastingtheorie. Het uitgangspunt is dat de wijzigingen die worden voorgesteld door de politieke partijen inzicht geven in de preferenties van herverdeling voor de politieke partijen. Zo zou het partijprogramma van de SP de armen in Nederland moeten helpen, terwijl het programma van de VVD vooral voorstellen zou moeten bevatten die de rijke Nederlanders bevoordelen. Gebruik makend van de optimale belastingtheorie kan worden achterhaald of de voorstellen van de politieke partijen inderdaad overeenkomen met hun ideologische kleur.

De ideologie die wordt achterhaald blijkt in schril contrast te staan tot de ideologie die door de politieke partijen zelf wordt uitgedragen. Zoals reeds uitgevonden in het vorige hoofdstuk bevoordeeldt het huidige stelsel de middeninkomens disproportioneel. In de voorstellen van alle politieke partijen, van de SP tot en met de VVD, blijft deze bevoordeelde positie van middeninkomens overeind en in sommige voorstellen wordt deze positie zelfs versterkt. Deze status quo kan verklaard worden met behulp van politieke economie. Wellicht ruilen politieke partijen hun ideologische belangen uit in de hoop om meer stemmen te winnen van de grote groep middeninkomens tijdens de verkiezingen.

Monitoring van arbeidsaanbod

In het derde hoofdstuk wordt bestudeerd of het belasting-subsidie stelsel op arbeidsinkomen beter kan functioneren wanneer een nieuw instrument wordt geïntroduceerd: monitoring van arbeidsaanbod. In de huidige optimale belastingtheorie wordt aangenomen dat de overheid wel kan bepalen hoeveel iemand verdient, maar niet hoe hard deze persoon moet werken om dit inkomen te verdienen. Het ligt voor de hand dat laagopgeleide werknemers die werken tegen het minimumloon veel meer uren moeten maken om hetzelfde te verdienen als een hoogopgeleide parttimemedewerker. In theorie kan de overheid veel efficiënter herverdelen indien de laagopgeleide persoon die veel uren maakt wordt beloond voor zijn harde werk, of indien de hoogopgeleide parttimemedewerker wordt gestraft door bijvoorbeeld minder steun te ontvangen van de staat. Dit geeft werknemers namelijk een sterke prikkel om harder te werken, zodat de overheid tegen lagere kosten meer herverdelen van rijk naar arm. Door de belasting op arbeidsinkomen verliezen werknemers hun prikkel om te werken, maar een monitoring instrument dat hard werkende burgers beloond kan deze prikkel gedeeltelijk herstellen.

Natuurlijk gaat de introductie van een monitoringinstrument ook gepaard met kosten. Indien de overheid wil achterhalen hoeveel uren werknemers werken zal het moeten investeren in ambtenaren die deze controle mogelijk maken. De kosten van deze controles moeten dan ook grondig worden afgewogen tegen de baten.

In dit hoofdstuk wordt een optimale formule voor de monitoringsintensiteit als functie van deze kosten en baten afgeleid. De monitoringsintensiteit moet het hoogste zijn bij die werknemers wiens arbeidsaanbod keuze het sterkst verstoord wordt door het belasting-subsidie stelsel op arbeidsinkomsten. Daarnaast kan de overheid de tarieven verhogen bij inkomensgroepen die sterk gemonitord worden, omdat monitoring deze groepen een sterke prikkel geeft om te werken, zodat deze werknemers ook actief zijn in de arbeidsmarkt wanneer hun tarief hoog is.

Simulaties op Amerikaanse data laten zien dat monitoring het belasting-subsidie stelsel aanzienlijk progressiever kan maken. Monitoring is vooral belangrijk bij de onderste inkomensgroepen, omdat daar inkomensafhankelijke regelingen de prikkel om te werken aanzienlijk vermindert. Het effect van monitoring is dat de overheid veel meer kan herverdelen naar de onderste inkomensgroepen. Het gevolg is dat de totale herverdeelde som stijgt, terwijl de kosten van herverdeling dalen. Monitoring blijkt een potentieel belangrijk instrument voor elke politieke kleur en onder allerlei gevoeligheidsanalyses. Dit hoofdstuk adviseert landen dan ook om hun belasting-subsidie systeem te complementeren met een instrument dat mensen beloond voor hun arbeidsinzet.

De belasting op kapitaalinkomen

In hoofdstuk vier wordt onderzocht hoe de belasting op kapitaalinkomsten de beleggingsbeslissingen van huishoudens beïnvloed. De belasting op kapitaalinkomen is potentieel een krachtig herverdelingsinstrument voor de overheid omdat deze herverdeelt van huishoudens met veel vermogen naar huishoudens met minder vermogen. In de economisch theoretische modellen is de verstoring die gepaard gaat met deze belasting echter ook zeer groot. Dat komt doordat de belasting op kapitaalinkomsten gespaard vermogen meerdere keren belast. De eerste belasting vindt plaats wanneer het vermogen verdiend wordt via de belasting op arbeidsinkomsten. Maar wanneer dit zelfde inkomen op bijvoorbeeld een spaarrekening wordt geplaatst, wordt het volgende jaar de verdiende rente nog een keer belast via de belasting op kapitaalinkomsten. Op deze manier wordt gespaard vermogen elk jaar opnieuw belast met als resultaat een zeer grote verstoring van de economie. Het klassieke resultaat uit de economische literatuur is dan ook dat inkomen eenmaal belast moet worden bij uitkering. Gespaard vermogen moet geheel vrijgesteld worden van belasting.

Nieuwe resultaten in de literatuur laten echter zien dat kapitaalinkomen waarschijnlijk toch belast moeten worden. Verschillende huishoudens hebben soms compleet verschillende beleggingsresultaten. Herverdeling tussen gelukkige huishoudens met goede beleggingsresultaten en de rest van de maatschappij kan in dat geval leiden tot een grote herverdelingswinst en deze herverdeling kan alleen plaats vinden via een belasting op kapitaalinkomen. Bovendien, blijkt een belasting op kapitaalinkomsten soms de herverdelende werking van het belasting-subsidie stelsel op arbeidsinkomen te versterken. Daarom complementeren veel landen hun belasting-subsidie stelsel op arbeidsinkomsten met een belasting op kapitaalinkomen.

Om de hoogte van de optimale belasting op kapitaalinkomsten te bepalen is het belangrijk om te weten in hoeverre de belasting de beleggingsbeslissing van huishoudens beïnvloed. In dit hoofdstuk wordt getest in hoeverre de belasting op kapitaalinkomen de keuze om te sparen beïnvloed en of de belasting invloed heeft op de portfoliocompositie van huishoudens. Helaas is hierover heel weinig bekend omdat in de meeste landen de portfolio van huishoudens niet wordt geregistreerd. Het Centraal Bureau voor Statistiek heeft echter speciaal voor deze studie data beschikbaar gesteld over de portfolio's van Nederlandse huishoudens in de periode 1995-2004. Deze periode is bijzonder interessant omdat Nederland in 2001 de belasting op kapitaalinkomsten grondig heeft hervormd. Deze hervorming wordt gebruikt om te meten wat het effect van belastingen is op het beleggingsgedrag van huishoudens.

De 2001-hervorming veranderde de belasting op inkomsten uit kapitaal voor vrijwel alle Nederlandse huishoudens. Daarnaast veranderde de fiscale behandeling van alle vormen van kapitaalinkomsten, behalve de kapitaalinkomsten op het eigen huis. De effecten van de hervorming zijn zeer verschillend voor de verschillende huishoudens. Sommige huishoudens moesten meer belasting betalen, terwijl andere huishoudens juist hun belasting op inkomen uit kapitaal zagen dalen. Daardoor is het mogelijk om de effecten van de verandering in de belasting te scheiden van andere schokken in de Nederlandse economie. Deze hervorming vormt dan ook een uitgelezen mogelijkheid om de invloed van de belasting op het beleggingsgedrag van huishoudens te bestuderen.

Eerdere studies met data van veel lagere kwaliteit vonden een sterk effect van de belasting op kapitaalinkomen op het beleggingsgedrag van huishoudens. De resultaten in dit hoofdstuk laten zien dat de belasting inderdaad statistisch significante effecten heeft, maar de effecten zijn economisch beperkt. Een verhoging van de belasting op kapitaalinkomen zorgt ervoor dat de besparingen van een huishouden naar beneden gaan, maar minder dan voorheen werd aangenomen. Daarnaast zorgt een verhoging van de belasting het eigen huis ervoor dat mensen minder sparen in hun eigen huis en meer in andere financiële middelen, maar ook deze reactie is zeer beperkt. Dit betekent dat de belasting op kapitaalinkomen een veel beter instrument is voor herverdeling dan eerder aangenomen. De kapitaalinkomstenbelasting kan een belangrijke begeleidende rol spelen bij herverdeling met behulp het belasting-subsidie stelsel op arbeidsinkomsten en overheden zouden hier meer gebruik van kunnen maken.

Herverdeling in meerdere dimensies

Individuen in de economie verschillen in veel verschillende dimensies, zoals hun loon, hun gezondheid, hun vermogen en hun preferenties. De taak van de overheid is dan ook om niet alleen te herverdelen van rijk naar arm, maar ook van vermogend naar minder-vermogend en van gezond naar ziek. De optimale belastingtheorie kon deze verschillende dimensies van herverdeling tot nu toe slechts in isolatie bestuderen. In de eerste drie hoofdstukken van dit proefschrift werd aangenomen dat huishoudens slechts verschilde in hun loonvoet, terwijl in het vierde hoofdstuk gekeken werd naar herverdeling tussen huishoudens met verschillend gespaard vermogen. Er zijn daarnaast vele studies gewijd aan de herverdeling tussen gezonde en zieke individuen. Het tegelijkertijd bestuderen van herverdeling over verschillende dimensies bleek tot nu toe wiskundig te ingewikkeld.

De resultaten in dit hoofdstuk laten echter zien dat analyse van het herverdeling in verschillende dimensies wel degelijk mogelijk is. De simplificerende aanname die analyse van het optimale belasting systeem onder meerdere dimensies mogelijk maakt is de zogenaamde "first-order" analyse. Deze vorm van analyse, die gemeengoed is bij de analyse van herverdeling onder een dimensie, maakt het mogelijk om het probleem van herverdeling "lokaal" te analyseren. De overheid bepaalt voor elk individu de optimale herverdeling tussen dat individu en een individu die bijzonder veel op het eerste individu lijkt, maar in één dimensie iets afwijkt. Door dit voor alle individuen en dimensies te doen kan de overheid de optimale herverdeling in de gehele economie bepalen. Vervolgens kan het deze herverdeling bewerkstelligen met het gehele orkest aan herverdelingsinstrumenten.

De formule voor de optimale tarieven blijkt te kunnen worden gevangen in een inzichtelijke formule die herverdelingswinsten afweegt tegen de doelmatigheid kosten voor elk herverdelingsinstrument en elke dimensie waarover de overheid wil herverdelen. Met behulp van de formule wordt laten zien dan de overheid altijd meerdere instrumenten moet gebruiken indien het wil herverdelen in verschillende dimensies. Het beroemde Atkinson-Stiglitz theorema dat zegt dat onder bepaalde aannames de overheid alleen herverdeelt via het belasting-subsidie stelsel op arbeidsinkomen vervalt volledig wanneer de overheid in meerder dimensies wil herverdelen. Dit stelsel kan onmogelijk zowel van rijk naar arm als van gezond naar ziek herverdelen. De overheid heeft daarom aparte instrumenten nodig die herverdelen van vermogend naar onvermogend en van gezond naar ziek.

Als laatste wordt laten zien dat de optimale marginale belasting op alle keuzevariabelen voor de onderste individuen, bijvoorbeeld de individuen die zowel het minst vermogend, het ongezondst als het armst zijn, nul moeten zijn. Datzelfde geldt voor de bovenste individuen al is dit tweede inzicht meer theoretisch van aard, omdat de vaststelling van het bovenste type niet triviaal is terwijl types die zich daar net onder bevinden juist zeer hoge marginale tarieven moeten betalen.

Het laatste hoofdstuk behandelt een belangrijk hiaat in de optimale belastingliteratuur. In de literatuur wordt de optimale herverdeling bepaald voor een centraal geplande economie, het zogenaamde directe mechanisme. Vervolgens wordt aangenomen dat deze optimale herverdeling ook in een markteconomie tot stand gebracht kan worden via een relatief eenvoudig belasting-subsidie systeem. Deze koppeling tussen twee compleet verschillende problemen lijkt in eerste instantie lastig te begrijpen. Een vrije markteconomie geeft burgers aanzienlijk meer vrijheid dan een centraal geplande economie. Deze keuzevrijheid heeft vele economische voordelen omdat burgers bijvoorbeeld kunnen kiezen om hun eigen bedrijf op te richten wanneer ze een goed idee hebben, en de markteconomie bedrijven in staat stelt in te spelen op nieuwe technologische trends. Maar tegelijkertijd bemoeilijkt deze keuzevrijheid de herverdelende taak van de overheid, omdat het veel lastiger is om te bepalen naar wie precies herverdeelt moet worden en via welk instrument dat het beste kan gebeuren.

Dit hoofdstuk laat zien dat de koppeling tussen het probleem in de centraal geplande economie en de markteconomie niet altijd bestaat. Wanneer de overheid paternalistische voorkeuren aan de burgers wil opdringen of wanneer het marktmechanisme sterk faalt, moet de overheid ingrijpen door de keuzevrijheid van burgers te beperken. In de meeste optimale belasting modellen zijn deze problemen echter niet aanwezig. In dat geval laten we zien dat de overheid de keuzevrijheid van burgers niet hoeft te beperken en dat relatief eenvoudige belastingsysteem inderdaad afdoende kan herverdelen in een markteconomie. Daarmee laat dit hoofdstuk ex-post zien dat de optimale herverdeling berekend in het grootste gedeelte van de economische literatuur tot stand kan worden gebracht met een relatief simpel belastingstelsel in een vrije markt economie. Ook de voorgestelde belastingsystemen in hoofdstuk 1, 3 en 5 van dit proefschrift worden hiermee ex-post gevalideerd.

Bibliography

Ahmad, E. and N. Stern (1984), The Theory of Reform and Indian Indirect Taxes, *Journal of Public Economics*, 25, 259–298.

Akerlof, G. A. (1976), The Economics of the Caste and of the Rat Race and Other Woeful Tales, *Quarterly Journal of Economics*, 90, 599–617.

Alan, S., K. Atalay, T. F. Crossley, and S.-H. Jeon (2010), New Evidence on Taxes and Portfolio Choice, *Journal of Public Economics*, 94, 813–823.

Alan, S. and S. Leth-Petersen (2006), Tax Incentives and Household Portfolios: A Panel Data Analysis, CAM 2006-13, Copenhagen.

Albanesi, S. and C. Sleet (2006), Dynamic Optimal Taxation with Private Information, *Review of Economic Studies*, 73, 1–30.

Alesina, A., E. Glaeser, and B. Sacerdote (2005), Work and Leisure in the United States and Europe: Why So Different?, *NBER Macroeconomics Annual*, 20, 1–64.

Alesina, A., A. Ichino, and L. Karabarbounis (2011), Gender-Based Taxation and the Division of Family Chores, *American Economic Journal: Economic Policy*, 3, 1–40.

Allingham, M. G. and A. Sandmo (1972), Income Tax Evasion: A Theoretical Analysis, *Journal of Public Economics*, 1, 323–338.

Apps, P. and R. Rees (1998), Taxation and the Household, *Journal of Public Economics*, 35, 355–369.

Armenter, R. and T. M. Mertens (2013), Fraud deterrence in dynamic Mirrleesian economies, *Journal of Monetary Economics*, 60, 139–151.

Armstrong, M. (1996), Multiproduct Nonlinear Pricing, *Econometrica*, 64, 51–75.

Armstrong, M. and J.-C. Rochet (1999), Multi-dimensional Screening: A User's Guide, *European Economic Review*, 43, 959–979.

Athey, S. and J. Levin (2001), Information and Competition in US Forest Service Timber Auctions, *Journal of Political Economy*, 109, 375–417.

Atkinson, A. B., T. Piketty, and E. Saez (2011), Top Incomes over a Century or more, *Journal of Economic Literature*, 49.

Atkinson, A. B. and W. Salverda (2005), Top Incomes in the Netherlands and the United Kingdom over the Twentieth Century, *Journal of the European Economic Association*, 3, 883–913.

Atkinson, A. B. and J. E. Stiglitz (1976), The Design of Tax Structure: Direct Versus Indirect Taxation, *Journal of Public Economics*, 6, 55–75.

Attanasio, O. P. and M. Browning (1995), Consumption over the Life Cycle and the Business Cycle, *American Economic Review*, 85, 1118–1137.

Attanasio, O. P. and M. Wakefield (2010), The Effects on Consumption and Saving of Taxing Asset Returns, in J. A. Mirrlees, S. Adam, T. J. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. D. Myles, and J. M. Poterba (eds.), *The Mirrlees Review – Dimensions of Tax Design*, chap. 7, Oxford University Press, Oxford, pp. 202–274.

Attanasio, O. P. and G. Weber (1989), Intertemporal Substitution, Risk Aversion and the Euler Equation for Consumption, *Economic Journal*, 99, 59–73.

Attanasio, O. P. and G. Weber (1993), Consumption Growth, the Interest Rate and Aggregation, *Review of Economic Studies*, 60, 631–649.

Attanasio, O. P. and G. Weber (1995), Is Consumption Growth Consistent with Intertemporal Optimization? Evidence from the Consumer Expenditure Survey, *Journal of Political Economy*, 103, 1121–1157.

Attanasio, O. P. and G. Weber (2010), Consumption and Saving: Models of Intertemporal Allocation and Their Implications for Public Policy, *Journal of Economic Literature*, 48, 693–751.

Banks, J., R. Blundell, and I. Preston (1994), Life-Cycle Expenditure Allocations and the Consumption Costs of Children, *European Economic Review*, 38, 1391–1410.

Bargain, O., M. Dolls, D. Neumann, A. Peichl, and S. Siegloch (forthcoming), Comparing Inequality Aversion Across Countries when Labor Supply Responses Differ, *International Tax and Public Finance*.

Bargain, O., M. Dolls, D. Nuemann, A. Peichl, and S. Siegloch (2011), Tax-Benefit Systems in Europe and the US: Between Equity and Efficiency, IZA Discussion Paper No. 5440, Bonn.

Bargain, O. and C. Keane (2010), Tax-benefit-revealed Redistributive Preferences over Time: Ireland 1987-2005, *Labour*, 24, 141–167.

Bargain, O., K. Orsini, and A. Peichl (forthcoming), Labor Supply Elasticities in Europe and the US, *Journal of Human Resources*.

Bassetto, M. and C. Phelan (2008), Tax Riots, *Review of Economic Studies*, 75, 649–669.

Bettendorf, L., S. Cnossen, and C. v. Ewijk (2012), BTW-Verhoging Treft Hoge en Lage Inkomens Even Sterk, *Me Judice*, 25 April.

Bierbrauer, F. (2009), A Note on Optimal Income Taxation, Public Goods Provision and Robust Mechanism Design, *Journal of Public Economics*, 93, 667–670.

Bloemen, H. G. (2009), An Empirical Model of Collective Household Labour Supply with Non-participation, *Economic Journal*, 120, 183–214.

Bloemen, H. G. (2010), Income Taxation in an Empirical Collective Household Labour Supply Model with Discrete Hours, Tinbergen Institute Discussion Paper 2010-010/3, Amsterdam.

Blomquist, S. and L. Simula (2010), Marginal Deadweight Loss when the Income Tax is Nonlinear, mimeo, Uppsala University.

Blundell, R., A. Bozio, and G. Laroque (2011), Labor Supply and the Extensive Margin, *American Economic Review*, 101, 482–486.

Blundell, R., M. Brewer, P. Haan, and A. Shephard (2009), Optimal Income Taxation of Lone mothers: An Empirical Comparison of the UK and Germany, *Economic Journal*, 119, 101–121.

Blundell, R., M. Browning, and C. Meghir (1994), Consumer Demand and the Life-cycle Allocation of Household Expenditures, *Review of Economic Studies*, 61, 57–80.

Boadway, R. and K. Cuff (1999), Monitoring Job Search as an Instrument for Targeting Transfers, *International Tax and Public Finance*, 6, 317–337.

Boadway, R. and P. Pestieau (2011), Indirect Taxes for Redistribution: Should Necessity Goods be Favored?, CESifo Working Paper Series, Munich.

Bonenkamp, J. (2009), Measuring Lifetime Redistribution in Dutch Occupational Pensions, *De Economist*, 157, 49–77.

Boone, J., P. Fredriksson, B. Holmlund, and J. C. Van Ours (2007), Optimal Unemployment Insurance with Monitoring and Sanctions, *Economic Journal*, 117, 399–421.

Boone, J. and J. C. Van Ours (2006), Modeling Financial Incentives to Get the Unemployed Back to Work, *Journal of Institutional and Theoretical Economics*, 162, 227–252.

Boskin, M. J. and E. Sheshinski (1983), Optimal Tax Treatment of the Family: Married Couples, *Journal of Public Economics*, 20, 281–297.

Bourguignon, F. and Amedeo Spadaro (2010), Tax-benefit Revealed Social Preferences, *Journal of Economic Inequality*.

Bourguignon, F. and A. Spadaro (2000), Social Preferences Revealed through Effective Marginal Tax Rates, DELTA Working Paper 29, Paris.

Bourguignon, F. and A. Spadaro (2012), Tax–benefit Revealed Social Preferences, *Journal of Economic Inequality*, 10, 75–108.

Bovenberg, A. L. and B. Jacobs (2005), Redistribution and Education Subsidies are Siamese Twins, *Journal of Public Economics*, 89, 2005–2035.

Bovenberg, L. and S. Cnossen (2001), Fundamental Tax Reform in the Netherlands, *International Tax and Public Finance*, 8, 471–484.

Bovenberg, L. and C. Teulings (2005), De Vlaktaks in Concreto, *Economisch Statistische Berichten*, 90, 112–114.

Brewer, M., E. Saez, and A. Shephard (2010), Means-testing and Tax Rates on Earnings, in J. A. Mirrlees, S. Adam, T. J. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. D. Myles, and J. M. Poterba (eds.), *The Mirrlees Review – Dimensions of Tax Design*, chap. 3, Oxford University Press, Oxford, pp. 202–274.

Campbell, J. Y. (2006), Household Finance, *Journal of Finance*, 61, 1553–1604.

Cattell, D., P. Bowen, and A. Kaka (2007), Review of Unbalanced Bidding Models in Construction, *Journal of Construction engineering and Management*, 133, 562–573.

Chamley, C. (1986), Optimal Taxation of Capital Income in General Equilibrium with Infinite Lives, *Econometrica*, 54, 607–22.

Chander, P. and L. L. Wilde (1998), A General Characterization of Optimal Income Tax Enforcement, *Review of Economic Studies*, 65, 165–183.

Chetty, R. (2006), A New Method of Estimating Risk Aversion, *American Economic Review*, 96, 1821–1834.

Chetty, R. (2012), Bounds on Elasticities with Optimization Frictions: a Synthesis of Micro and Macro Evidence on Labor Supply, *Econometrica*, 80(3), 969–1018.

Choné, P. and G. Laroque (2010), Negative Marginal Tax Rates and Heterogeneity, *American Economic Review*, 100, 2532–2547.

Christiansen, V. and E. S. Jansen (1978), Implicit Social Preferences in the Norwegian System of Social Preferences, *Journal of Public Economics*, 10, 217–245.

Clauset, A., C. R. Shalizi, and M. E. Newman (2009), Power-Law Distributions in Empirical Data, *SIAM Review*, 51, 661–703.

Clementi, F. and M. Gallegati (2005a), Pareto's Law of Income Distribution: Evidence for Germany, the United Kingdom, and the United States, in A. Chatterjee, S. Yarlagadda, and B. K. Chakrabarti (eds.), *Econophysics of Wealth Distributions*, chap. 1, Springer, Milan, pp. 3–14.

Clementi, F. and M. Gallegati (2005b), Power Law Tails in the Italian Personal Income Distribution, *Physica A: Statistical Mechanics and its Applications*, 350, 427–438.

Cocco, J. F. (2005), Portfolio Choice in the Presence of Housing, *Review of Financial Studies*, 18, 535–567.

Conesa, J. C., S. Kitao, and D. Krueger (2009), Taxing Capital? Not a Bad Idea After All!, *American Economic Review*, 99, 25–48.

CPB (2002a), Charting Choices 2003-2006: Economic Effects of Eight Election Platforms, CPB Netherlands Bureau for Economic Policy Analysis, The Hague.

CPB (2002b), *Keuzes in Kaart: 2003-2006*, CPB Netherlands Bureau for Economic Policy Analysis, The Hague.

CPB (2010a), *Centraal Economisch Plan 2010*, CPB Netherlands Bureau for Economic Policy Analysis, The Hague.

CPB (2010b), Hervorming van Het Nederlandse Woonbeleid, The Hague.

CPB and PBL (2012), *Charted Choices 2013-2017: Effects of Nine Election Platforms on the Economy and the Environment*, no. 5 in CPB Book, CPB Netherlands Bureau for Economic Policy Analysis, The Hague.

Cremer, H. and F. Gahvari (1994), Tax Evasion, Concealment and the Optimal Linear Income Tax, *Scandinavian Journal of Economics*, 2, 219–239.

Cremer, H. and F. Gahvari (1996), Tax Evasion and the Optimum General Income Tax, *Journal of Public Economics*, 60, 235–249.

Cremer, H., P. Pestieau, and J.-C. Rochet (2001), Direct Versus Indirect Taxation: the Design of the Tax Structure Revisited, *International Economic Review*, 42, 781–800.

De Koning, J., H. Kroes, and A. van der Steen (2006), Patronen van Werk en Gebruik van Sociale Regelingen, SEOR working paper, Rotterdam.

De Mooij, R. A. and G. Nicodème (2008), Corporate Tax Policy and Incorporation in the EU, *International Tax and Public Finance*, 15, 478–498.

Decoster, A. and E. Schokkaert (1989), Equity and Efficiency of a Reform of Belgian Indirect Taxes, *Recherches Economiques De Louvain*, 55, 155–176.

Dekker, P. and S. Ederveen (2003), *Europese Verkenning 1: Sociaal Europa 1*, SCP Netherlands Institute for Social Research, The Hague.

Diamond, P. A. (1980), Income Taxation with Fixed Hours of Work, *Journal of Public Economics*, 13, 101–110.

Diamond, P. A. (1998), Optimal Income Taxation: An Example with a U-shaped Pattern of Optimal Marginal Tax Rates, *American Economic Review*, 88, 83–95.

Diamond, P. A. and J. Banks (2010), The Base for Direct Taxation, in J. A. Mirrlees, S. Adam, T. J. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. D. Myles, and J. M. Poterba (eds.), *The Mirrlees Review – Dimensions of Tax Design*, chap. 6, Oxford University Press, Oxford, pp. 548–648.

Domar, E. D. and R. A. Musgrave (1944), Proportional Income Taxation and Risk-taking, *Quarterly Journal of Economics*, 58, 388–422.

Ebert, U. (1992), A Reexamination of the Optimal Non-linear Income Tax, *Journal of Public Economics*, 49, 47–73.

Engelhardt, G. V. and A. Kumar (2009), The Elasticity of Intertemporal Substitution: New Evidence from 401 (k) Participation, *Economics Letters*, 103, 15–17.

Erosa, A. and M. Gervais (2002), Optimal Taxation in Life-Cycle Economies, *Journal of Economic Theory*, 105, 338–369.

Ewerhart, C. and K. Fieseler (2003), Procurement Auctions and Unit-price Contracts, *RAND Journal of Economics*, 569–581.

Farhi, E. and I. Werning (2010), Progressive Estate Taxation, *Quarterly Journal of Economics*, 125, 635–673.

Farhi, E. and I. Werning (2013), Insurance and Taxation over the Life Cycle, *Review of Economic Studies*, 1, 596–635.

Feldstein, M. (1995), The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act, *Journal of Political Economy*, 551–572.

Feldstein, M. (1999), Tax Avoidance and the Deadweight Loss of the Income Tax, *Review of Economics and Statistics*, 81, 674–680.

Feldstein, M. S. (1976), Personal Taxation and Portfolio Composition: An Econometric Analysis, *Econometrica*, 44, 631–50.

Fredriksson, P. and B. Holmlund (2006), Improving Incentives in Unemployment Insurance: A Review of Recent Research, *Journal of Economic Surveys*, 20, 357–386.

Gauthier, S. and G. Laroque (2009), Separability and Public Finance, *Journal of Public Economics*, 93, 1168–1174.

Gerritsen, A. (2013), Optimal Taxation when People Do Not Maximize Well-Being, mimeo, Erasmus University Rotterdam.

Gielen, M., J. Goes, M. Lever, and R. Van Opstal (2009), Ontwikkeling en Verdeling van de Marginale Druk 2001-2011 (Development and Distribution of Marginal Tariffs 2001-2011, in Dutch), CPB Document 195, The Hague.

Golosov, M., N. Kocherlakota, and A. Tsyvinski (2003), Optimal Indirect and Capital Taxation, *Review of Economic Studies*, 70, 569–587.

Golosov, M., M. Troshkin, and A. Tsyvinski (2011), Optimal Dynamic Taxes, NBER Working Paper No. w17642, Cambridge, MA.

Golosov, M., A. Tsyvinski, and M. Weinzierl (2013), Preference Heterogeneity and Optimal Commodity Taxation, *Journal of Public Economics*, 97, 160–175.

Golosov, M., A. Tsyvinski, and I. Werning (2007), New Dynamic Public Finance: A User's Guide, in D. Acemoglu, K. Rogoff, and M. Woodford (eds.), *NBER Macroeconomics Annual 2006, Volume 21*, MIT Press, Cambridge-MA, pp. 317–388.

Graafland, J. J. and A. P. Ros (eds.) (2003), *Economic Assesment of Election Programmes: Does it Make Sense?*, Kluwer Academic Publishers, Dordrecht.

Gruber, J. and E. Saez (2002), The Elasticity of Taxable Income: Evidence and Implications, *Journal of Public Economics*, 84, 1–32.

Guesnerie, R. (1995), *A Contribution to the Pure Theory of Taxation*, vol. 25, Cambridge University Press, Cambridge.

Hall, R. E. (1978), Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence, *Journal of Political Economy*, 96, 339–357.

Hall, R. E. (1988), Intertemporal Substitution in Consumption, *Journal of Political Economy*, 86, 971–87.

Hammond, P. J. (1979), Straightforward Individual Incentive Compatibility in Large Economies, *Review of Economic Studies*, 46, 263–282.

Hansen, L. P. (1982), Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica*, 50, 1029–54.

Hansen, L. P. and K. J. Singleton (1983), Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns, *Journal of Political Economy*, 91, 249–265.

Harris, M. and R. M. Townsend (1981), Resource Allocation under Asymmetric Information, *Econometrica*, 49, 33–64.

Heady, C. (2010), Trends in Top Incomes and Income Inequality and Their Implications for Tax Policy, OECD Working Paper (concept) WP2(2010)6, Paris.

Heckman, J. J. (1979), Sample Selection Bias as a Specification Error, *Econometrica*, 47, 153–161.

Hellwig, M. F. (2007), A Contribution to the Theory of Optimal Utilitarian Income Taxation, *Journal of Public Economics*, 91, 1449–1477.

Hellwig, M. F. (2010), A Generalization of the Atkinson–Stiglitz (1976) Theorem on the Undesirability of Nonuniform Excise Taxation, *Economics Letters*, 108, 156–158.

Holmstrom, B. (1979), Moral Hazard and Observability, *Bell Journal of Economics*, 10, 74–91.

Hubbard, R. (1985), Personal Taxation, Pension Wealth, and Portfolio Composition, *Review of Economics and Statistics*, 67, 53–60.

Immervoll, H. (2004), Average and Marginal Effective Tax Rates Facing Workers in the EU: A Micro-level Analysis of Levels, Distributions and Driving Factors, OECD Unclassified DELSA/ELSA/WD/SEM(2004)4, Paris.

Jacobs, B. (2008), *De Prijs van Gelijkheid (The Price of Equality, in Dutch)*, Bert Bakker, Amsterdam.

Jacobs, B. (2013), The Marginal Cost of Public Funds is One at the Optimal Tax System, mimeo Erasmus University Rotterdam.

Jacobs, B. and R. A. De Mooij (2011), Pigou Meets Mirrlees: On the Irrelevance of Tax Distortions for the Second-best Pigouvian Tax, CESifo Working Paper No. 3250, Munich.

Jacquet, L., E. Lehmann, and B. Van Der Linden (2010), The Optimal Marginal Tax Rates with both Extensive and Intensive responses, IRES Working Paper 2010-33, Paris.

Jacquet, L., E. Lehmann, and B. Van Der Linden (2013), Optimal Redistributive Taxation with Both Extensive and Intensive Responses, *Journal of Economic Theory*, 148, 1770–1805.

Jongen, E. L. and M. Stoel (2013a), Estimating the Elasticity of Taxable Labour Income in the Netherlands, CPB Background Document, The Hague.

Jongen, E. L. and M. Stoel (2013b), Estimating the Elasticity of Taxable Labour Income in the Netherlands, CPB Background Document, CPB Netherlands Bureau for Economic Research, may 29.

Judd, K. L. (1985), Redistributive Taxation in a Simple Perfect Foresight Model, *Journal of Public Economics*, 28, 59–83.

Kanbur, R., M. Tuomala, and J. Pirttilae (2006), Non-Welfarist Optimal Taxation and Behavioral Economics, *Journal of Economic Surveys*, 20, 1467–6419.

Kaplow, L. (2006), On the Undesirability of Commodity Taxation Even when Income Taxation is Not Optimal, *Journal of Public Economics*, 90, 1235–1250.

King, M. A. and J. I. Leape (1998), Wealth and Portfolio Composition: Theory and Evidence, *Journal of Public Economics*, 69, 155–193.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011), Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark, *Econometrica*, 79, 651–592.

Kleven, H. J. and C. T. Kreiner (2006), *Skat, Arbejde og Lighed*, chap. Beskatning af Arbejdsindkomst i Danmark (Taxation of Labor Income in Denmark, in Danish), Gyldendal, Copenhagen, ch. 7.

Kleven, H. J., C. T. Kreiner, and E. Saez (2009), The Optimal Income Taxation of Couples, *Econometrica*, 77, 537–560.

Kleven, H. J. and E. A. Schultz (forthcoming), Estimating Taxable Income Responses using Danish Tax Reforms, *American Economic Journal: Economic Policy*.

Kocherlakota, N. R. (2005), Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation, *Econometrica*, 73, 1587–1621.

Kocherlakota, N. R. (2006), Advances in Dynamic Optimal Taxation, in R. Blundell, W. K. Newey, and T. Persson (eds.), *Econometric Society Monographs*, chap. 7, Cambridge University Press, Cambridge, pp. 269–297.

Kocherlakota, N. R. (2010), *The New Dynamic Public Finance*, Princeton University Press, Princeton-NJ.

Kopczuk, W. (2013), Incentive Effects of Inheritances and Optimal Estate Taxation, *American Economic Review*, 103, 472–77.

Laroque, G. R. (2005), Indirect Taxation is Superfluous under Separability and Taste Homogeneity: A Simple Proof, *Economics Letters*, 87, 141–144.

Layard, R. (1980), Human Satisfactions and Public Policy, *Economic Journal*, 90, 737–750.

LeRoy, S. F. and R. D. Porter (1981), The Present-Value Relation: Tests Based on Implied Variance Bounds, *Econometrica*, 49, 555–74.

Lewis, T. R. and D. E. Sappington (1988), Regulating a Monopolist with Unknown Demand and Cost Functions, *RAND Journal of Economics*, 438–457.

Ljungqvist, L. and T. J. Sargent (1995a), The Swedish Unemployment Experience, *European Economic Review*, 39, 1043–1070.

Ljungqvist, L. and T. J. Sargent (1995b), Welfare States and Unemployment, *Economic Theory*, 6, 143–160.

Madrian, B. C. and D. F. Shea (2001), The Power of Suggestion: Inertia in 401 (k) Participation and Savings Behavior, *Quarterly Journal of Economics*, 116, 1149–1187.

Mankiw, N., M. Weinzierl, and D. Yagan (2009), Optimal Taxation in Theory and Practice, *Journal of Economic Perspectives*, 23, 147–174.

Mankiw, N. and S. P. Zeldes (1991), The Consumption of Stockholders and Nonstockholders, *Journal of Financial Economics*, 29, 97–112.

Markowitz, H. M. (1952), Portfolio Selection, *Journal of Finance*, 7, 77–91.

Mas-Collel, A., M. D. Whinston, and J. R. Green (1995), *Microeconomics*, Oxford University Press, Oxford.

Maskin, E. (1999), Nash Equilibrium and Welfare Optimality, *Review of Economic Studies*, 66, 23–38.

Mastrogiacomo, M., N. M. Bosch, M. D. Gielen, and E. L. Jongen (2013), A Structural Analysis of Labour Supply Elasticities in the Netherlands, CPB Discussion Paper 235, The Hague.

McAfee, R. P. and J. McMillan (1988), Multidimensional Incentive Compatibility and Mechanism Design, *Journal of Economic Theory*, 46, 335–354.

Meltzer, A. H. and S. F. Richard (1981), A Rational Theory of the Size of Government, *Journal of Political Economy*, 914–927.

Merton, R. (1969), Life-time Portfolio Selection under Uncertainty, *Review of Economics and Statistics*, 51, 247–259.

Merton, R. C. (1971), Optimum Consumption and Portfolio Rules in a Continuous-time Model, *Journal of Economic Theory*, 3, 373–413.

Merton, R. C. (1973), An Intertemporal Capital Asset Pricing Model, *Econometrica*, 41, 867–887.

Mirrlees, J. A. (1971), An Exploration in the Theory of Optimum Income Taxation, *Review of Economic Studies*, 38, 175–208.

Mirrlees, J. A. (1976), Optimal Tax Theory: A Synthesis, *Journal of Public Economics*, 6, 327–358.

Mirrlees, J. A. (1997), Information and Incentives: The Economics of Carrots and Sticks, *Economic Journal*, 107, 1311–1329.

Mirrlees, J. A. (1999), The Theory of Moral Hazard and Unobservable Behaviour: Part I, *Review of Economic Studies*, 66, 3–21.

Mookherjee, D. and I. Png (1989), Optimal Auditing, Insurance, and Redistribution, *Quarterly Journal of Economics*, 104, 399–415.

Myerson, R. B. (1979), Incentive Compatibility and the Bargaining Problem, *Econometrica*, 47, 61–73.

OECD (2011), Taxation and Employment, OECD Tax Policy Studies, No.21, Paris.

Okun, A. M. (1975), *Equality and Efficiency: The Big Tradeoff*, Brookings Institution Press.

Olson, M. (1982), *The Rise and Decline of Nations: Economic Growth, Stagflation, And Social Rigidities*, Yale University Press, New Haven and London.

Pass, B. (2012), Convexity and Multi-Dimensional Screening for Spaces with Different Dimensions, *Journal of Economic Theory*, 147, 2399–2418.

Pavan, A., I. Segal, and J. Toikka (2010), Dynamic Mechanism Design: Incentive Compatibility, Profit Maximization and Information Disclosure, SSRN 1620662.

Persson, T. and G. Tabellini (2000), *Political Economics: Explaining Economic Policy*, MIT Press.

Piketty, T. and E. Saez (2003), Income Inequality in The United States, 1913-1998, *Quarterly Journal of Economics*, 118, 1–39.

Poterba, J. M. (2002), Taxation, Risk-taking, and Household Portfolio Behavior, in M. Feldstein and A. J. Auerbach (eds.), *Handbook of Public Economics*, vol. 3, Elsevier, Amsterdam, pp. 1109–1171.

Poterba, J. M. and A. A. Samwick (2003), Taxation and Household Portfolio Composition: US Evidence from the 1980s and 1990s, *Journal of Public Economics*, 87, 5–38.

Renes, S. (2011), Balancing the Bids, Solutions for Unit Price Auctions, Tinbergen Institute Discussion Paper 11-047/1, Amsterdam.

Renes, S. and F. T. Zoutman (2013a), As Easy as ABC? Multi-dimensional Screening in Public Finance, Rotterdam, mimeo: Erasmus University Rotterdam.

Renes, S. and F. T. Zoutman (2013b), When a Price is Enough, A Toolbox for Optimal Tax-Design, Rotterdam, mimeo: Erasmus University Rotterdam.

Roberts, K. W. S. (1977), Voting over Income Tax Schedules, *Journal of Public Economics*, 8, 329–340.

Roberts, K. W. S. (1984), The Theoretical Limits to Redistribution, *Review of Economic Studies*, 51, 177–195.

Rochet, J.-C. (1985), The Taxation Principle and Multi-Time Hamilton-Jacobi Equations, *Journal of Mathematical Economics*, 14, 113–128.

Rochet, J.-C. and P. Choné (1998), Ironing, Sweeping, and Multidimensional Screening, *Econometrica*, 66, 783–826.

Roemer, J. E. (1998), Why the Poor Do Not Expropriate the Rich: An Old argument in New Garb, *Journal of Public Economics*, 70, 399 – 424.

Romer, T. (1975), Individual Welfare, Majority Voting, and the Properties of a Linear Income Tax, *Journal of Public Economics*, 4, 163–185.

Rothschild, C. and F. Scheuer (2013), Redistributive Taxation in the Roy Model, *Quarterly Journal of Economics*, 128, 623–668.

Sadka, E. (1976), On Income Distribution, Incentive Effects and Optimal Income Taxation, *Review of Economic Studies*, 43, 261–267.

Saez, E. (2001), Using Elasticities to Derive Optimal Income Tax Rates, *Review of Economic Studies*, 68, 205–229.

Saez, E. (2002a), The Desirability of Commodity Taxation under Non-linear Income Taxation and Heterogeneous Tastes, *Journal of Public Economics*, 83, 217–230.

Saez, E. (2002b), Optimal Income transfer Programs: Intensive Versus Extensive Labor Supply Responses, *Quarterly Journal of Economics*, 117, 1039–1073.

Saez, E., J. B. Slemrod, and S. H. Giertz (2012), The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review, *Journal of Economic Literature*, 50, 3–50.

Sandmo, A. (1981), Income Tax Evasion, Labour Supply, and the Equity–efficiency Tradeoff, *Journal of Public Economics*, 16, 265–288.

Sargan, J. D. (1958), The Estimation of Economic Relationships using Instrumental Variables, *Econometrica*, 26, 393–415.

Schroyen, F. (1997), Pareto Efficient Income Taxation under Costly Monitoring, *Journal of Public Economics*, 65, 343–366.

Schroyen, F. (2003), Redistributive Taxation and the Household: The Case of Individual Filings, *Journal of Public Economics*, 87, 2527–2547.

Seade, J. K. (1977), On the Shape of Optimal Tax Schedules, *Journal of Public Economics*, 7, 203–235.

Seade, J. K. (1982), On the Sign of the Optimum Marginal Income Tax, *The Review of Economic Studies*, 49, 637–643.

Shiller, R. J. (1981), Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?, *American Economic Review*, 71, 421–436.

Slemrod, J. (1994), Fixing the Leak in Okun's Bucket: Optimal Tax Progressivity when Avoidance Can be Controlled, *Journal of Public Economics*, 55, 41–51.

Slemrod, J. and W. Kopczuk (2002), The Optimal Elasticity of Taxable Income, *Journal of Public Economics*, 84, 91–112.

Slemrod, J. and S. Yitzhaki (2002), Tax Avoidance, Evasion, and Administration, in A. J. Auerbach and M. Feldstein (eds.), *Handbook of Public Economics Volume 3*, chap. 22, Elsevier, Amsterdam, pp. 1423–1470.

Spadaro, A. (2005), Micro-simulation and Normative Policy Evaluation: An Application to Some EU Tax Benefits Systems, *Journal of Public Economic Theory*, 7, 593–622.

Statistics Netherlands (2013), Input-Output Tables The Netherlands, www.cbs.nl.

Stern, N. (1977), Welfare Weights and the Elasticity of the Marginal Valuation of Income, in M. Artis and R. Nobay (eds.), *Studies in Modern Economic Analysis*, Basil Blackwell Publishers, Oxford.

Stigler, G. J. (1970), Director's Law of Public Income Redistribution, *Journal of Law and Economics*, 13, 1–10.

Stiglitz, J. E. (1982), Self-selection and Pareto Efficient Taxation, *Journal of Public Economics*, 17, 213–240.

Summers, L. H. (1981), Capital Taxation and Accumulation in a Life Cycle Growth Model, *American Economic Review*, 71, 533–544.

Townsend, R. M. (1979), Optimal Contracts and Competitive Markets with Costly State Verification, *Journal of Economic Theory*, 21, 265–293.

Tuomala, M. (1984), On the Optimal Income Taxation: Some Further Numerical Results, *Journal of Public Economics*, 23, 351–366.

Tuomala, M. (2010), On Optimal Non-linear Income Taxation: Numerical Results Revisited, *International Tax and Public Finance*, 17, 259–270.

Van Ewijk, C., M. Koning, M. Lever, and R. A. De Mooij (2006), Economische Effecten van Aanpassing Fiscale Behandeling Eigen Woning (Economic Effects of Fiscal Treatment of Owner-Occupied Housing, in Dutch, CPB bijzondere Publicaties 62, The Hague.

Vermeulen, F. (2005), And the Winner is ... an Empirical Evaluation of Unitary and Collective Labour Supply Models, *Empirical Economics*, 30, 711–734.

Vickrey, W. S. (1947), *Agenda for Progressive Taxation*, Ronald Press, New York.

Weber, C. (2013), Obtaining a Consistent Estimate of the Elasticity of Taxable Income using Difference-in-Differences, mimeo, University of Oregon.

Wetenschappelijk Instituut voor Het CDA (2009), Een Sociale Vlaktaks, naar Werkbare en Begrijpelijke Inkomstenbelastingen, Wetenschappelijk Instituut voor Het CDA, The Hague.

Wilson, R. (1996), Nonlinear Pricing and Mechanism Design, *Handbook of Computational Economics*, 1, 253–293.

Zheng, C. Z. (2000), *Optimal Auction in a Multidimensional World*, vol. 1282, Center for Mathematical Studies in Economics and Management Science, Northwestern University.

Zoutman, F. T. (2013), The Effect of Capital Taxes on Household's Portfolio Composition and Intertemporal Choice: Evidence from the Dutch 2001 Capital Income Tax Reform, Rotterdam, mimeo: Erasmus University Rotterdam.

Zoutman, F. T. and B. Jacobs (2013), Optimal Redistribution and Monitoring of Labor Effort, Rotterdam, mimeo: Erasmus University Rotterdam.

Zoutman, F. T., B. Jacobs, and E. L. Jongen (2013a), Optimal Redistributive Taxes and Redistributive Preferences in the Netherlands, Rotterdam, mimeo: Erasmus University Rotterdam/CPB Netherlands Bureau for Economic Policy Analysis.

Zoutman, F. T., B. Jacobs, and E. L. Jongen (2013b), Revealed Social Preferences of Dutch Political Parties, Rotterdam, mimeo: Erasmus University Rotterdam/CPB Netherlands Bureau for Economic Policy Analysis.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

521. Y.ZU, *Essays of nonparametric econometrics of stochastic volatility.*
522. B.KAYNAR, *Rare Event Simulation Techniques for Stochastic Design Problems in Markovian Setting.*
523. P. JANUS, *Developments in Measuring and Modeling Financial Volatility.*
524. F.P.W. SCHILDER, *Essays on the Economics of Housing Subsidies.*
525. S. M MOGHAYER, *Bifurcations of Indifference Points in Discrete Time Optimal Control Problems.*
526. C. ÇAKMAKLI, *Exploiting Common Features in Macroeconomic and Financial Data.*
527. J. LINDE, *Experimenting with new combinations of old ideas.*
528. D. MASSARO, *Bounded rationality and heterogeneous expectations in macroeconomics.*
529. J. GILLET, *Groups in Economics.*
530. R. LEGERSTEE, *Evaluating Econometric Models and Expert Intuition.*
531. M.R.C. BERSEM, *Essays on the Political Economy of Finance.*
532. T. WILLEMS, *Essays on Optimal Experimentation.*
533. Z. GAO, *Essays on Empirical Likelihood in Economics.*
534. J. SWART, *Natural Resources and the Environment: Implications for Economic Development and International Relations.*
535. A. KOTHIYAL, *Subjective Probability and Ambiguity.*
536. B. VOOGT, *Essays on Consumer Search and Dynamic Committees.*
537. T. DE HAAN, *Strategic Communication: Theory and Experiment.*
538. T. BUSER, *Essays in Behavioural Economics.*
539. J.A. ROSERO MONCAYO, *On the importance of families and public policies for child development outcomes.*
540. E. ERDOGAN CIFTCI, *Health Perceptions and Labor Force Participation of Older Workers.*
541. T.WANG, *Essays on Empirical Market Microstructure.*

542. T. BAO, *Experiments on Heterogeneous Expectations and Switching Behavior.*
543. S.D. LANSDORP, *On Risks and Opportunities in Financial Markets.*
544. N. MOES, *Cooperative decision making in river water allocation problems.*
545. P. STAKENAS, *Fractional integration and cointegration in financial time series.*
546. M. SCHARTH, *Essays on Monte Carlo Methods for State Space Models.*
547. J. ZENHORST, *Macroeconomic Perspectives on the Equity Premium Puzzle.*
548. B. PELLOUX, *the Role of Emotions and Social Ties in Public On Good Games: Behavioral and Neuroeconomic Studies.*
549. N. YANG, *Markov-Perfect Industry Dynamics: Theory, Computation, and Applications.*
550. R.R. VAN VELDHUIZEN, *Essays in Experimental Economics.*
551. X. ZHANG, *Modeling Time Variation in Systemic Risk.*
552. H.R.A. KOSTER, *The internal structure of cities: the economics of agglomeration, amenities and accessibility.*
553. S.P.T. GROOT, *Agglomeration, globalization and regional labor markets: Micro evidence for the Netherlands.*
554. J.L. MÖHLMANN, *Globalization and Productivity Micro-Evidence on Heterogeneous Firms, Workers and Products.*
555. S.M. HOOGENDOORN, *Diversity and Team Performance: A Series of Field Experiments.*
556. C.L. BEHRENS, *Product differentiation in aviation passenger markets: The impact of demand heterogeneity on competition.*
557. G. SMRKOLJ, *Dynamic Models of Research and Development.*
558. S. PEER, *The economics of trip scheduling, travel time variability and traffic information.*
559. V. SPINU, *Nonadditive Beliefs: From Measurement to Extensions.*
560. S.P. KASTORYANO, *Essays in Applied Dynamic Microeconometrics.*
561. M. VAN DUIJN, *Location, choice, cultural heritage and house prices.*
562. T. SALIMANS, *Essays in Likelihood-Based Computational Econometrics.*
563. P. SUN, *Tail Risk of Equidity Returns.*
564. C.G.J. KARSTEN, *The Law and Finance of M&A Contracts.*
565. C. OZGEN, *Impacts of Immigration and Cultural Diversity on Innovation and Economic Growth.*
566. R.S. SCHOLTE, *The interplay between early-life conditions, major events and health later in life.*

567. B.N.KRAMER, *Why don't they take a card? Essays on the demand for micro health insurance.*

568. M. KILIÇ, *Fundamental Insights in Power Futures Prices.*

569. A.G.B. DE VRIES, *Venture Capital: Relations with the Economy and Intellectual Property.*

570. E.M.F. VAN DEN BROEK, *Keeping up Appearances.*

571. K.T. MOORE, *A Tale of Risk: Essays on Financial Extremes.*