Optimal Social Insurance and Redistribution: Incentives for Educational Investment, Work and Savings

Dissertation

zur Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften (Dr. rer. pol.)

vorgelegt von

Dominik Sachs

an der

Universität Konstanz

Sektion Politik – Recht – Wirtschaft

Fachbereich Wirtschaftswissenschaften

Konstanz, 2013

Tag der mündlichen Prüfung: 23.09.2013

1. Referent: Prof. Dr. Friedrich Breyer
2. Referent: Prof. Dr. Leo Kaas

Acknowledgement

I am deeply indebted to my advisor Friedrich Breyer. He has been a true mentor for me since the second year of my undergraduate studies and supported and advised me throughout all the years. He has crucially influenced my way of thinking about economic questions.

I also owe special thanks to my second advisor Leo Kaas. He has generously supported me and my work with a lot of valuable advice and feedback.

My coauthors Sebastian Findeisen and Normann Lorenz – who are both also friends – deserve special gratitude. Working and interacting with Sebastian over the last three years has been truly inspiring and I am glad that I got infected by his enthusiasm. Sharing and developing ideas with him is a real privilege. Normann had the role of a mentor for me throughout all my studies, who always urged me to think twice. Conducting research together with him has been a fruitful experience.

During my time in Konstanz, I have also interacted with many other faculty members and benefited from their comments: Tim Friehe, Bernd Genser, Volker Hahn, Matthias Hertweck, Zohal Hessami, Georgi Kocharkov, Niklas Potrafke, Almuth Scholl and Heinrich Ursprung. Dirk Schindler deserves special gratitude for having supported, motivated and advised me in pursuing research in public economics.

I would also like to thank all my PhD-colleagues in Konstanz for great discussions and, more importantly, for making the time so enjoyable. Especially Fabian Fink, Jan Häußler, Fabian Krüger, Verena Lauber, Maria Micevski and Ruben Seiberlich, who also became valuable friends.

In 2011-12 I had the privilege to spend several months at Stanford University, where I got generous academic support from Florian Scheuer. His advice has not only been enormously valuable for my thesis but also for my way of thinking about problems in public economics and how to tackle them. In Spring 2013 I spent five weeks at Yale University under the supervision of Aleh Tsyvinski. I am deeply indebted to him for his intense and enormously valuable supervision and guidance as well as his optimism and motivation.

Besides working on my thesis, my time as a PhD-student often involved writing research proposals and the like. All these proposals greatly benefited from comments of patient friends: Namely, Christian Adam, Miriam Benkißer, Susanne Benkißer, Sebastian Findeisen, Fabian Fink and Normann Lorenz. My mother Monika always reliably did the final proof reading.
I would like to thank my family for supporting me my entire life and for always believing in me. Thanks also to all my friends for making life so enjoyable. Finally, I would like to thank Susi for her unyielding support throughout the last years, always believing in me and her honest and valuable advice. Discussing my research ideas with her always helped me to shape them more precisely and without her the last years would have not been half as enjoyable.

Contents

List of Figures

Summary

This dissertation consists of four independent research papers. All four papers deal with the optimal design of tax, welfare and education policies, or generally speaking, the welfare state. The central underlying trade-off is the one between equity and efficiency, or redistribution and incentives. Before summarizing the four chapters in detail, I will briefly outline both sides of the trade-off and discuss how they are addressed in this dissertation.

Welfare state policies imply redistribution of income between ex-ante heterogeneous individuals. In addition, such policies provide insurance against certain risks over the life cycle, such as income, human capital and unemployment risk. While Chapters 2 and 4 focus on the classical redistributive role of the welfare state, in Chapters 1 and 3 life cycle models are considered, in which individuals are subject to income risk. Thus, these two chapters explicitly address the insurance motive of the welfare state in addition to the classical redistribution motive.

Redistribution within a welfare state always affects incentives of individuals. If one does not reap all the fruits of one's labor, working hard is usually less worthwhile. But also incentives for educational investment and wealth accumulation can be lowered because of redistribution. In all chapters of this dissertation, the labor supply margin of individuals is taken into account. In Chapters 1 and 4, it is also explicitly modeled that individuals' education decisions can be affected by the design of the welfare state. In Chapter 2, in addition to the decision of how much to work (intensive margin), the decision whether to participate in the labor force or not (extensive margin) is considered. Finally, in Chapter 3, the effect of redistribution on incentives for wealth accumulation is taken into account.

Chapter 1 is joint work with Sebastian Findeisen (University of Zurich and Centre for Equitable Growth, Berkeley). In this paper, we study the optimal integrated design of education finance and tax systems. From a policy perspective, we address the issue of income-contingent student loans, i.e. student loans that come with income-contingent repayment. Such loans have the advantage that they provide insurance

against human capital risk. At the same time they come with efficiency costs as they increase the effective marginal income tax rates. Whether the introduction of such income-contingent loans increases welfare therefore depends on the question whether the gains from insurance outweigh the efficiency costs. Our analysis suggests that the efficiency costs of income-contingent student loans are clearly outweighed by the insurance gains.

More formally, we study a life cycle model where individuals that are heterogeneous with respect to their innate ability make an educational investment early in their life. Consistent with empirical evidence, this educational investment is risky. Thus, having completed education, individuals enter the labor market and draw their wage from a distribution that depends on their innate ability as well as their educational investment. Within this formal environment we derive properties of second-best Pareto optimal allocations, where second-best refers to an informational asymmetry: the innate type of individuals as well as their wage are private information. Individuals reveal their innate type in a first step and their wage in a second step once uncertainty has materialized. To solve this screening problem, we use a dynamic first-order approach.

We find that an integrated education and tax system in which the government provides education loans to young individuals coupled with income-contingent repayment can always be designed in a Pareto optimal way. In other words: Pareto optimal allocations can be implemented with nonlinear income taxes and income-contingent students loans. We present a simple empirically driven application of the framework to US data in which individuals make a college entry decision. We find that the optimal repayment scheme for college loans is almost linearly increasing up to intermediate incomes and becomes flat afterwards. The welfare gains of making loan repayment contingent on income are significant.

Chapter 2 is joint work with Normann Lorenz (University of Trier). We study the optimal design of income transfer programs with a special focus on participation taxes and marginal tax rates in the income region where transfers are phased-out.

Our first contribution is technical: we study optimal nonlinear taxation in a model where individuals differ along two dimensions: their wage and their preference for leisure. Further, individuals can only participate in the labor market if they work a larger number of hours than some exogenous minimum hours constraint. This minimum hours constraint can be micro-founded by fixed costs on the side of the employer (e.g. for providing training or equipment). Focusing on a special kind of preferences, where a so called "type-aggregator" can be applied, we solve this non

standard two-dimensional screening problem, where the report set of the individuals is constrained endogenously. Then, we rewrite our results for the optimal nonlinear tax system in reduced form, i.e. in terms of elasticities and the income distribution and show how our results are related to previous results in the literature.
Based on this reduced form solution, our second contribution is to derive a novel formula for the optimal participation tax in a framework with both, intensive and extensive labor supply responses. We then generalize a famous result that within an optimal nonlinear tax system, a government always levies positive participation taxes if the social marginal utility of income of the lowest income worker is smaller than the marginal value of public funds.
Finally, as our most important contribution, we develop an empirical test for the Pareto efficiency of a given tax-transfer system. This test only requires knowledge about the income distribution as well as about intensive and extensive labor supply elasticities; it does not rest on the minimum hours assumption as a reason for extensive labor supply responses. The test can not only identify whether marginal tax rates are above their Laffer value but also whether the structure of marginal tax rates is second-best Pareto inefficient. We provide an empirical application by investigating the German tax-transfer system (for singles) and find that the structure of marginal tax rates is inefficient in the region where transfers (Hartz IV) are just phased-out. Our results suggest that a decrease of the transfer phase-out rates could yield a Pareto improvement.

Chapter 3 is joint work with Sebastian Findeisen. In this chapter, we consider optimal labor and capital income taxation over the life cycle. The main innovation is to make the problem of deriving optimal nonlinear labor income tax schedules and linear capital tax rates that do only condition on current (annual) income tractable. Importantly, we show that capital taxation is not superfluous in such a life cycle framework with realistic policy instruments and provide a novel formula for the optimal capital tax that arises in our life cycle framework. It follows a standard public finance trade-off: the optimal capital tax tends to increase in wealth inequality and decrease in the elasticity of savings with respect to taxes.
The key simplifying assumption to make the problem tractable is the absence of income effects on labor supply. Based on this assumptions, we develop a first-order approach which allows us to derive optimal nonlinear labor income tax schedules and linear capital taxes for a continuous type space. The formula that we derive for the optimal marginal labor income tax rates resembles the standard intuition of the static model. However, there is an additional dynamic term that captures the

impact of labor income taxes on individual savings behavior and the implied effect on public funds via the capital tax.

Finally, we simulate a calibrated version of the model. We find that labor and capital income taxes are increasing over the life cycle. Confirming our intuition from the theoretical section, we find that capital income taxes are significantly different from zero and yield significant welfare gains.

Chapter 4 is also joint work with Sebastian Findeisen. This chapter investigates the impact of the time-consistency problem of a government on education policies. We show that a lower commitment power of a government tends to make education subsidies more progressive.

We study a simple two-period model, where individuals can either be of high or low ability in the first period. In that first period, individuals make an educational investment which increases the wage that they earn in the second period. The government, which wants to redistribute from the high skilled to the low skilled, makes use of a linear labor income tax in the second period and subsidizes education non-linearly in the first period. Once educational investment is sunk, the government has the incentive to apply a different labor income tax rate as announced because then the impact on educational responses does not have to be considered. As rational individuals anticipate this temptation of the government, they decrease their educational investment.

As a benchmark, we first study the case of full commitment and derive results for the labor income tax and education subsidies. We then compare these results to the case where a government has no commitment power at all and always deviates from its tax announcements in the future and therefore always sets higher taxes. However, since the benevolent government anticipates its own temptation to increase taxes, it tries to mitigate this tax increase by keeping income inequality low. The lower the income inequality, the lower the increase in taxes. In order to keep income inequality low, the government keeps education inequality low. This can be done by subsidizing educational investment of the low (high) ability type more (less) heavily; thus by making education subsidies more progressive. Afterwards we show that this result generalizes to intermediate commitment cases. Finally, using data of the Worldbank and the UNESCO, we show that the correlation between the commitment power of a government and the progressivity of education subsidies is in line with our theoretical results.

Zusammenfassung

Die vorliegende Dissertation besteht aus vier eigenständigen Forschungsarbeiten. Alle vier Arbeiten befassen sich mit der optimalen Ausgestaltung des Sozialstaats bzw. der Steuer-, Sozial- und Bildungspolitik. Ein zentraler Zielkonflikt, welcher der Frage der optimalen Ausgestaltung stets zugrunde liegt, ist der zwischen (ökonomischer) Gleichheit und Effizienz, bzw. zwischen Umverteilung und Anreizen. Die vorliegende Arbeit untersucht folglich, wie die Abwägung zwischen Gleichheit und Effizienz getroffen werden sollte, so dass beide Ziele bestmöglich miteinander vereinbart werden. Bevor im Folgenden die einzelnen Kapitel näher vorgestellt werden, werden beide Seiten des Zielkonflikts kurz erläutert und es wird dargelegt, inwiefern sie in den einzelnen Kapiteln untersucht werden.

Sozialstaatliche Maßnahmen bewirken eine Umverteilung von Einkommen zwischen *ex-ante* ungleichen Individuen. Darüber hinaus stellt die Umverteilung von Reich nach Arm aber auch eine Versicherung gegenüber gewissen Lebensrisiken wie dem Einkommens-, Humankapital- oder Arbeitslosigkeitsrisiko dar. Während der Fokus der Arbeiten in Kapitel 2 und 4 auf dem klassischen Umverteilungsmotiv des Sozialstaats liegt, werden in den Kapiteln 1 und 3 Lebenszyklus-Modelle betrachtet, in denen die Individuen Einkommensrisiken ausgesetzt sind. In diesen Kapiteln wird daher neben dem klassischen Umverteilungsmotiv auch das Versicherungsmotiv betrachtet.

Umverteilung im Rahmen eines Sozialstaats mindert stets die Anreize der Menschen. So wird es typischerweise weniger rentabel, Arbeitseinkommen zu erwirtschaften, wenn die Erträge der Arbeit einem nicht vollständig selbst zukommen. Doch auch die Anreize, in Bildung zu investieren und Ersparnisse zu bilden, können aufgrund von Umverteilung verringert werden. In allen Kapiteln dieser Dissertation wird die Tatsache modelliert, dass Individuen ihre Arbeitsentscheidung in Abhängigkeit der Rahmenbedingungen des Sozialstaates treffen. In den Kapiteln 1 und 4 wird darüber hinaus auch explizit berücksichtigt, dass die Ausbildungsentscheidung der Individuen von der Ausgestaltung des Sozialstaates abhängen kann. In Kapitel 2 wird neben neben der Entscheidung bezüglich der Arbeitszeit auch die Entscheidung modelliert, ob überhaupt gearbeitet wird. In Kapitel 3 liegt der Fokus auf den Auswirkungen von Umverteilung auf die Sparentscheidung der Individuen.

Kapitel 1 ist eine Gemeinschaftsarbeit mit Sebastian Findeisen (Universität Zürich und Centre for Equitable Growth, Berkeley). In dieser Arbeit wird die optimale Ausgestaltung einer integrierten Bildungs- und Steuerpolitik betrachtet. Dabei wird die Idee von Studienkrediten mit einkommensabhängigen Rückzahlungen untersucht. Solche Kredite haben den Vorteil, dass sie Studenten gegen das Ausbildungsrisiko teilweise versichern. Auf der anderen Seite bringen sie Effizienz-Kosten mit sich, da sie die effektiven Grenzsteuersätze erhöhen. Ob solche Studienkredite mit einkommensabhängigen Rückzahlungen generell wohlfahrtserhöhend sind oder nicht, hängt davon ab, ob die Vorteile durch ein Mehr an Versicherung die Nachteile der Effizienzkosten überwiegen. Unsere Analyse zeigt, dass die Effizienzkosten von den Wohlfahrtsgewinnen durch das Mehr an Versicherung überwogen werden.
Dabei wird ein formales Modell betrachtet, in welchem Individuen, die sich bezüglich ihrer Fähigkeiten unterscheiden, eine Ausbildungsentscheidung treffen. Im Einklang mit empirischer Evidenz ist diese Ausbildungsentscheidung risikobehaftet. Mit dem Eintritt in den Arbeitsmarkt realisiert sich der Lohnsatz gemäß einer Lotterie, deren Wahrscheinlichkeiten sowohl von den Fähigkeiten des Individuums als auch von der Bildungsentscheidung abhängen. In diesem formalen Rahmen werden die Bedingungen von „second-best“ Pareto-optimalen Allokationen hergeleitet, wobei sich „second-best“ auf die Tatsache bezieht, dass der Staat weder die ursprüngliche Fähigkeit noch den Lohnsatz der Individuen beobachten kann. Um dieses „screening“-Problem zu lösen, wird von einem dynamischen „first-order“ Ansatz Gebrauch gemacht.
Anschließend wird gezeigt, dass die Pareto-optimalen Allokationen durch einen nichtlinearen Einkommensteuertarif sowie Studienkredite mit einkommensabhängigen Rückzahlungen implementiert werden können. In einer für die USA kalibrierten Version des Modells wird gezeigt, dass die Rückzahlungen von Studienkrediten im Optimum für geringe und mittlere Einkommen tatsächlich mit dem Einkommen ansteigen sollten. Die Wohlfahrtsgewinne durch die Einführung von Studienkrediten mit einkommensabhängigen Rückzahlungen sind dabei signifikant.

Kapitel 2 ist eine Gemeinschaftsarbeit mit Normann Lorenz (Universität Trier). In diesem Kapitel wird die optimale Ausgestaltung von Steuer-Transfer-Systemen mit einem speziellen Fokus auf Partizipationssteuern und Transferentzugsraten untersucht.

Der erste Beitrag zur Literatur ist technischer Natur: Es wird das nichtlineare Optimalsteuerproblem für den Fall, dass Individuen sich sowohl bezüglich ihres Lohnsatzes als auch bezüglich ihrer Freizeitpräferenz unterscheiden und außerdem bei der Arbeits-Freizeits-Entscheidung durch eine Mindestarbeitszeit eingeschränkt sind, gelöst. Eine Teilnahme am Arbeitsmarkt ist also nur möglich, wenn mehr als eine gewisse Mindestarbeitszeit gearbeitet wird, welche z.B. durch Fixkosten auf Seiten der Arbeitgeber (Bereitstellung von Arbeitsausstattung oder regelmäßige Weiterbildung) mikrofundiert werden kann. Durch die Annahme gewisser Präferenzen, die die Anwendung eines sogenannten „type aggregators“ erlaubt, wird dieses zweidimensionale „Screening“ -Problem lösbar gemacht. Anschließend werden die Ergebnisse in reduzierte Form umformuliert, also lediglich in Abhängigkeit von Elastizitäten und der Einkommensverteilung, und es wird gezeigt, wie die Resultate in die Literatur einzuordnen sind.
Darauf aufbauend wird eine Formel für die optimalen Partizipationssteuern hergeleitet, wenn Individuen ihr Arbeitsangebot sowohl marginal anpassen („intensive margin“) als auch Partizipationsentscheidungen treffen („extensive margin“). Darauf aufbauend wird ein bekanntes Resultat aus der Literatur verallgemeinert, welches besagt, dass Partizipationssteuern für jede Einkommenshöhe positiv sind, wenn der soziale Grenznutzen des Einkommens des Geringstverdieners kleiner ist als der Schattenwert der Staatseinnahmen.
Der letzte und wohl wichtigste Beitrag dieser Arbeit ist die Herleitung eines empirischen Tests bezüglich der Pareto-Effizienz eines Steuer-Transfer-Systems. Um diesen Test anwenden zu können, müssen nur die Arbeitsangebotselastizitäten und die Einkommensverteilung als Information vorliegen. Der Test kann nicht nur identifizieren, ob der Grenzsteuersatz für ein gewisses Einkommensniveau über dem sogenannten Laffer-Wert liegt, sondern kann auch eine („second-best“) ineffiziente Struktur von Grenzsteuersätzen identifizieren. Eine Anwendung dieses Test auf das deutsche Steuer-Transfer-Systems (für Alleinstehende) zeigt, dass dieses in der Tat eine Ineffizienz aufweist. So ist die Struktur der Grenzsteuersätze im Bereich des Einkommensniveaus, in welchem der letzte Euro Arbeitslosengeld II noch gezahlt wird, ineffizient. Unsere Ergebnisse legen nahe, dass eine Reduktion der Transferentzugsraten eine Pareto-Verbesserung darstellen würde.

Kapitel 3 ist eine Gemeinschaftsarbeit mit Sebastian Findeisen. In dieser Arbeit wird die optimale Besteuerung von Arbeits- und Kapitaleinkommen über den Lebenszyklus betrachtet. Die Hauptinnovation dieser Arbeit besteht darin, das Problem der optimalen nichtlinearen Besteuerung von Arbeitseinkommen sowie der linea-

ren Besteuerung von Kapitaleinkommen in einem dynamischen Lebenszyklus-Modell mit Unsicherheit lösbar zu machen. Der inhaltliche Hauptbeitrag besteht im Resultat, dass Kapitaleinkommensbesteuerung als zusätzliches Umverteilungsmittel nicht überflüssig ist. Insbesondere wird eine Formel für die optimale Höhe der Kapitalsteuer hergeleitet, welche einen klassischen Zielkonflikt in der Optimalsteuertheorie abbildet: Die Kapitalbesteuerung ist umso größer, je höher die Ungleichheit der Kapitaleinkommensverteilung und je kleiner die Elastizität der Ersparnisse bezüglich dieser Steuern ist.
Die zentrale Annahme, die das Poblem lösbar macht, ist die Abwesenheit von Einkommenseffekten auf die Arbeitsangebotsentscheidung der Individuen. Aufbauend auf dieser Annahme, wird für dieses Problem ein „first-order“-Ansatz entwickelt, der es ermöglicht, dieses dynamische Optimalsteuerproblem auch für den Fall einer kontinuierlichen Produktivitätsverteilung zu lösen. Die Formel für den optimalen Grenzsteuersatz auf das Arbeitseinkommen spiegelt die Intuition aus dem statischen Optimalsteuerproblem wider. Allerdings gibt es aufgrund der Dynamik einen zusätzlichen Term, welcher den Effekt der Arbeitseinkommensteuern auf die Ersparnisse und die dadurch implizierten Auswirkungen auf den Staatshaushalt (über die Kapitalsteuern) widerspiegelt. Letztlich wird eine kalibrierte Version dieses Modells simuliert, was das Ergebnis liefert, dass Arbeits- und Kapitaleinkommenssteuern über den Lebenszyklus steigen sollten. In Einklang mit der Intuition unserer theoretischen Resultate liefern die Simulationsergebnisse Kapitalsteuern, die deutlich positiv sind. Die Wohlfahrtsgewinne durch Kapitalbesteuerung sind signifikant.

Kapitel 4 ist ebenfalls eine Gemeinschaftsarbeit mit Sebastian Findeisen. In dieser Arbeit wird in einem theoretischen Modell der Zusammenhang zwischen der Progressivität der Bildungssubventionierung und der Fähigkeit einer Regierung, sich bezüglich zukünftiger Politikmaßnahmen zu verpflichten, untersucht. Es wird gezeigt, dass der Zusammenhang negativ ist.
Dafür wird ein einfaches zwei-Perioden-Modell betrachtet, in welchem Individuen entweder durch eine hohe oder eine niedrige Fähigkeit gekennzeichnet sind. In der ersten Periode investieren die Individuen in Bildung, was ihren Lohnsatz in der zweiten Periode erhöht. Die Regierung, welche von den Individuen mit hoher Fähigkeit zu denen mit niedriger Fähigkeit umverteilen möchte, wählt einen linearen Einkommensteuertarif in der zweiten Periode und subventioniert die Bildungsentscheidung in der ersten Periode. Die marginale Subventionierung der beiden Fähigkeitsniveaus kann dabei unterschiedlich sein, die Subventionierung kann beliebig nichtlinear sein. Sobald die Bildungsinvestition der Individuen getätigt ist, hat die Regierung den

Anreiz, die Einkommensteuer zu erhöhen, weil nun der negative Anreizeffekt auf die Bildungsentscheidung nicht mehr betrachtet werden muss. Rationale Individuen antizipieren diese Versuchung (also die Steuererhöhung) der Regierung jedoch und wählen eine niedrigere Bildungsinvestition.
Zunächst wird der Fall betrachtet, in der eine Regierung sich immer verpflichten kann, die Steuern in Zukunft nicht zu erhöhen. Diese Resultate werden dann mit dem Fall verglichen, dass eine Regierung sich überhaupt nicht dazu verpflichten kann, sich an Versprechen bezüglich zukünftiger Steuererhöhungen zu halten. Da die wohlwollende Regierung ihre eigene Versuchung vorhersieht, wird sie versuchen, diese Steuererhöhung gering zu halten. Dies lässt sich dadurch erreichen, dass die Einkommensungleichheit gering bleibt, da die Höhe der Steuern (in der zweiten Periode) mit der Einkommensungleichheit steigt. Um die Einkommensungleichheit zu verringern, muss die Regierung die Bildungsungleichheit verringern, was sie ceteris paribus dadurch erreicht, dass sie die Bildungsentscheidung des niedrigeren Fähigkeitsniveaus stärker und die des hohen Fähigkeitsniveaus schwächer subventioniert, also durch eine progressivere Bildungssubventionierung. Danach wird dieses Ergebnis auf die Fälle erweitert, in denen die Fähigkeit, sich zu verpflichten, zwischen den beiden Extremen liegt. Es wird gezeigt, dass Bildungssubventionen tendenziell umso progressiver sind, je schlechter eine Regierung sich verpflichten kann, sich an ihre Ankündigungen zu halten. Abschließend wird anhand von Daten der Weltbank und der UNESCO gezeigt, dass der empirische Zusammenhang zwischen der Progressivität der Bildungssubventionen und der Fähigkeit eines Staates, sich bezüglich der Ankündigungen zu verpflichten, in Einklang mit dem theoretischen Resultat dieses Kapitels steht.

Chapter 1

Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans

1.1 Introduction

How should governments design their higher education finance systems? There exist large differences across countries in the structure of higher education finance. In some countries, such as Denmark, Finland and Sweden, university and college students pay low or no tuition fees and in addition receive grants because of generous public subsidies for higher education. These countries have highly progressive tax systems, which allow to finance these education subsidies. By contrast, in the United Kingdom and the United States, e.g., the burden of educational costs mainly lies on the student and higher education is much less heavily subsidized by public finances. Instead, student loans offered by both the private and the public sector play a big part in financing higher education. From a policy perspective, the choice of an optimal education finance system is intimately linked to the tax system. Both underlie the same basic trade-offs, namely equity concerns in the form of redistribution and insurance against income risk versus efficiency concerns by distorting labor supply and education incentives.

In this paper, we address the optimal design of integrated education finance and tax systems. We build a novel optimal taxation framework in the spirit of Mirrlees (1971) and the vast literature following his footsteps, which allows to study the question from a new angle. In our framework, the distribution of wages is not exogenous but determined by the costly education decisions of individuals before labor market entry. Consistent with what is typically found in empirical studies, this human capital investment decision is risky. To solve the problem, we use an applied mechanism design approach. The benevolent government can observe total income and the education level of individuals, but it has to respect incentive compatibility – first, when individuals decide on education and second, when individuals decide on labor supply. The main novelty of our approach is that in our framework the government is not restricted to the use of predetermined instruments but is free to choose its own instruments, which can condition on education, income and savings. In addition, they are allowed to be fully nonlinear.

We find that an integrated education and tax system in which the government provides education loans to young individuals, coupled with income-contingent repayment rates of these loans after individuals enter the labor market, can effectively deal with all the major trade-offs underlying the education finance and tax problem. In other words, such systems can always be designed such that they are second-best Pareto efficient. This is because income-contingent repayment rates allow the government to *effectively differentiate tax distortions across education groups*, minimizing the efficiency cost of labor supply distortions. At the same time, it can subsidize education by varying the generosity of the loans.[1] Importantly, the government typically will find it optimal that some individuals partially default and never pay back the full value of their loans, while for some individuals the amount of repayment might exceed their loan values because this provides insurance.

We present a simple empirically driven application of the framework to US data in which individuals make a college entry decision. We simulate optimal income taxes and college student loans with income-contingent repayment. The optimal policy simulation provides three important insights. First, we find that the optimal repayment scheme for college loans can be well approximated by a schedule that is linearly increasing in income. So although the full optimum could lead to complicated nonlinear schedules in theory, very simple instruments can replicate it fairly well. Second, for our benchmark parameterization college graduates find it optimal to participate voluntarily in the loan schemes as compared to taking a risk-free loan on the private market. Third, we calculate the welfare gains of moving from a third-best scenario where the government optimally sets the income tax and offers a loan system with non-contingent repayment to the system with contingent repayments. We find welfare gains ranging from about 0.2% to 0.6% of lifetime consumption and we show how these gains vary with risk-aversion.

Several countries like the United Kingdom, Australia and New Zealand currently administer income-contingent college student loans, where repayment is proportional to income.[2] Our framework gives these policies a theoretical second-best foundation, based on an applied mechanism design approach to the education finance and taxation problem. Our theoretical considerations suggest that it might be optimal for the government to enforce that very rich individuals pay back more than the capitalized loan value or that repayment might actually be decreasing in income. In the mentioned countries, repayment never exceeds the loan value and repayment schedules are non-decreasing in income. To address these issues, we also consider policy experiments in which we restrict income-contingent repayment not to exceed the actual loan value and to be non-decreasing in income. We find that a large share of the welfare gains from the full optimum can be reaped with these simpler policies.

[1] We do not model credit market imperfection in the form of borrowing constraints. If these are relevant, as is still a debated question in the literature (Carneiro and Heckman, 2005), wide availability of student loans has the additional benefit of lifting these constraints.

[2] Chapman (2006) provides a survey for practices in those and other countries. To the best of our knowledge, the first economist to endorse the idea was Milton Friedman (1955). He envisioned repayment amounts to be proportional to income, i.e. a linearly increasing repayment schedule. Something we find as an optimal policy in our simulation for the most part of the income distribution.

Relation To Existing Literature. This chapter makes a contribution to the literature on optimal income taxation starting from Mirrlees (1971) (see the recent survey of Piketty and Saez (2013)). In Section 3 we discuss how the expression for optimal education-dependent marginal tax rates compares to the seminal optimal tax formulas from Diamond (1998) and Saez (2001) with exogenous human capital. In two papers, Bovenberg and Jacobs (2005) and Jacobs and Bovenberg (2011) analyze how endogenous education alters the optimal tax problem and discuss to what extent education should be subsidized. Bohacek and Kapicka (2008) study a dynamic model with certainty and obtain equivalent results regarding education subsidies. These articles work under certainty whereas we take idiosyncratic human capital risk into account. Using analytical results and numerical illustrations, we discuss in detail in Section 1.3 how our findings for optimal education subsidies in a general risky environment relate to their findings. Importantly, with idiosyncratic education risk, the necessity of education dependent labor wedges and income-contingent loans arises, as intuitively they can be understood as providing an additional source of insurance. As we discuss in Section 1.2, when we review some stylized empirical facts, there is strong evidence that uncertainty about college returns is important and matters for human capital investment decisions.[3]

Two recent papers, Best and Kleven (2013) and Kapicka and Neira (2013), study how human capital acquisition at the working age influence the optimal taxation problem. We focus on a different part of the human capital accumulation process, namely education before labor market entry. Importantly, both papers reasonably assume that tax policies cannot directly condition on human capital acquired while working. In contrast, we allow the government to use information about education before labor market entry in the tax code, as is done in the real world in some countries in the form of student loans with income-contingent repayment. In addition, our focus is on education finance instead of only tax policies.

Working with a two-type model, Gary-Bobo and Trannoy (2013) come to a similar conclusion concerning income-contingent loan repayment in a very recent paper. In

[3] One strand of literature has looked at first- versus second-best investment rules of human capital under risk with a representative agent. Da Costa and Maestri (2007) show that human capital should always be encouraged in the second-best optimum. Anderberg (2009) emphasizes that the risk properties of human capital are crucial for the question whether and how education should be distorted relative to a first-best rule. Focusing on linear policy instruments, Anderberg and Andersson (2003) as well as Jacobs et al. (2012) obtain similar results. An early treatment how taxes affect the risk properties of human capital investment is Eaton and Rosen (1980). Grochulski and Piskorski (2010) focus on the implications of unobservable human capital investment for capital taxation in an ex-ante homogeneous agent setting with uncertainty. Kapicka (2006) introduces non-observable endogenous human capital into a dynamic, non-stochastic Mirrlees model where taxes can only be conditioned on current income. He shows that marginal tax rates are lowered due to the education margin.

contrast to their work, we employ a continuous type approach in the tradition of the large literature on optimal income taxation going back to Mirrlees (1971). In particular, we are interested in determining the forces shaping the optimal design of student loan policies both theoretically and numerically, which requires a model with a continuum of types.

Concerning the implementation of history-dependent allocations, this chapter is related to Golosov and Tsyvinski (2006) who consider an environment with absorbing disability shocks and present an implementation in which disability insurance conditions on asset testing. Also in the context of optimal taxation, Scheuer (2012) considers differential taxation of profits and labor income; in our case a comparable logic applies for an endogenous education instead of an occupational choice.

Finally, taking a quantitative approach and working in the Ramsey tradition with simpler but given policy instruments, Krueger and Ludwig (2013) solve for the optimal income tax and education subsides in a rich macro model.

This chapter is organized as follows. Section 1.2 contains the basics of the model. In Section 1.3, we investigate dynamic incentive compatibility and describe the major properties of constrained efficient allocations. Decentralized implementations of constrained efficient allocations are provided in Section 1.4. We apply our model to the case of a binary education decision in Section 1.5 and Section 1.6 concludes.

1.2 The Model

1.2.1 Structure

We consider a stripped-down life-cycle model, in which individuals acquire formal education early in their life cycle and work afterwards. Individuals differ in innate ability θ, which can be interpreted as a one dimensional aggregate of (non-)cognitive skills, I.Q. and family background, and is distributed in the interval $[\underline{\theta}, \overline{\theta}]$ according to the cumulative density function (cdf) $F(\theta)$. After individuals learn their type θ, which is private information, they make a monetary educational investment e. Flow utility during education is denoted by $u^e(c_e)$ with $u^e_c > 0, u^e_{cc} < 0$. It takes T_e periods (years) until education is finished; the yearly education costs are denoted by e. To simplify exposition, we assume that all levels of education take the same amount of time, an assumption we relax later in our optimal policy simulations.

As individuals enter the labor market, they draw their labor market ability a from a continuous conditional cdf $G(a|e, \theta)$, which depends on *innate* ability θ and education e and has bounded support $[\underline{a}, \overline{a}]$, with $\underline{a} \geq 0$. We assume that preferences over consumption and leisure are given by the utility function $u^w(c_w, l)$, where labor

effort l is equal $\frac{y}{a}$, so that gross income is $y = a \times l$. We assume that $u^w(\cdot,\cdot)$ obeys the Spence-Mirrlees condition. The working life lasts for T_w periods.
Expected lifetime utility of an individual of type θ is hence given by

$$\sum_{t=1}^{T_e} \beta^{t-1} u^e(c_e(\theta)) + \int_{\underline{a}}^{\overline{a}} \sum_{t=T_e+1}^{T_e+T_w} \beta^{t-1} u^w \left(c_w(\theta,a), \frac{y(\theta,a)}{a} \right) dG(a|\theta, e(\theta)), \tag{1.1}$$

where we assume the allocation within the education and working period to be constant.[4] We will write $\beta^e = \sum_{t=1}^{T_e} \beta^{t-1}$ and $\beta^w = \sum_{t=T_e+1}^{T_e+T_w} \beta^{t-1}$. The yearly interest rate in the economy is given by $R = \frac{1}{\beta}$.
As equation (1.1) reveals, we abstract from further shocks to idiosyncratic labor productivity once individuals have entered the labor market. This simplifies and helps to focus the analysis on the education-taxation link. In the empirical literature, there is no ultimate consensus on the relative importance of heterogeneity before labor market entry (versus the role of shocks over the working life) for lifetime inequality, but different approaches have attributed a major role to it.[5]
Nevertheless, we capture many empirical regularities with this specification of the model. First, assuming $G(a|e,\theta)$ to be non-degenerate, our model captures the important fact of uncertainty in the labor market and risky educational investment. See e.g. Cunha and Heckman (2008) or Chen (2008) for recent contributions.
Second, we allow this cdf to be a function of innate ability θ and thereby capture the fact that inequality in earnings is – to a certain extent – also determined by innate ability. Taber (2001) and Hendricks and Schoellman (2012) suggest that much of the rise in the college premium may be attributed to a rise in the demand for unobserved skills, which are predetermined and independent of education. Indirect evidence for the importance of unobserved skills comes from the strong persistence of within-education-group inequality Acemoglu and Autor (2011).
Third, the cdf G being a function of e captures the returns to education. Importantly, for most of our results, we do not impose a certain assumption on the pattern of these returns.

[4]This is akin to the assumption that the first-order conditions of the second-best problem we solve are also sufficient.

[5]In recent work, Huggett et al. (2011) estimate a structural life-cycle model and find that differences realized at the age of 23 can account for more of the variation in lifetime outcomes than do shocks received over the working lifetime. A standard reference is Keane and Wolpin (1997) who attribute 90% to heterogeneity realized before labor market entry, while Storesletten et al. (2004) estimate a number of about 50%.

Fourth, as long as $\frac{\partial^2 G(a|e,\theta)}{\partial\theta\partial e} \neq 0$, returns to educational investment differ in innate ability θ. E.g., Carneiro and Heckman (2005) document that the returns can differ by as much as 19 percentage points across individuals for one year of college.[6]
To sharpen a few analytical results, it turns out helpful to place some structure on the behavior of $G(a|e,\theta)$:

Assumption 1.2.1 *$G(a|e',\theta) \geq_{FOSD} G(a|e,\theta) \Leftrightarrow G(a|e',\theta) \leq G(a|e,\theta)$, for all $e < e'$ and for all (θ, a).*

Assumption 1.2.2 *$G(a|e,\theta') \geq_{FOSD} G(a|e,\theta) \Leftrightarrow G(a|e,\theta') \leq G(a|e,\theta)$, for all $\theta < \theta'$ and for all (e, a).*

Assumption 1.2.3 *$\frac{\partial^2 G(a|e,\theta)}{\partial\theta\partial e} \leq 0$ for all (θ, e, a).*

These assumptions will not be needed to derive our main results, but help to illustrate important aspects of the model. Whenever an assumption is needed for a result, we refer to it. The first and the second one capture the notion that education and innate ability should both have a direct effect on labor market outcomes represented by a first-order stochastic dominance shift; a rather natural way of ordering distributions. The third one captures their interaction and respects the compelling evidence of complementarity between early ability and educational investment.

1.2.2 Definition of Wedges

For later purposes when we analyze optimal allocations and the respective tax and education finance systems that can implement such allocations, it is useful to define wedges. They are equal to implicit marginal tax rates. We are particularly interested in labor and education wedges. We use subscripts to indicate partial derivatives.
Labor Wedge: The labor wedge is positive (negative) if an individual works less (more) than it would at the intervention-free market price (which is her productivity level a). Formally the labor wedge reads as:

$$\tau_y(\theta, a) = 1 - \frac{u_l^w\left(c_w(\theta,a), \frac{y(\theta,a)}{a}\right)\frac{1}{a}}{u_c^w\left(c_w(\theta,a), \frac{y(\theta,a)}{a}\right)}.$$

Education Wedge: Here, a positive (negative) wedge corresponds to an upward (downward) distortion of the education decision. Formally the education wedge

[6]See also Lemieux (2006) for evidence on heterogeneity in returns.

reads as

$$\tau^e(\theta) = 1 - \frac{\beta^w \int_{\underline{a}}^{\overline{a}} u\left(c_w(\theta,a), \frac{y_w(\theta,a)}{a}\right) \frac{\partial g(a|e(\theta),\theta)}{\partial e(\theta)} da}{\beta^e u_c^e(c_e(\theta))}.$$

Finally, we will also look at optimal distortions of an individual's Euler equation between the education and the working period.

Savings Wedge:

$$\tau_s(\theta) = 1 - \frac{\beta^e u_c^e(c_e(\theta))}{\beta^w R \int_{\underline{a}}^{\overline{a}} u_c^w\left(c_w(\theta,a), \frac{y(\theta,a)}{a}\right) g(a|e,\theta) da}$$

where R is the gross return on savings between the education and the working life. $\tau_s(\theta) > (<)\, 0$ implies a downward (upward) distortion of savings.

1.3 Constrained Pareto Optimal Allocations

In this section, we characterize constrained Pareto efficient allocations, where "constrained" refers to the government being unable to observe agents' type θ at the education stage and a in the working stage. In Subsection 1.3.1, we show that the problem is tractable using a first-order approach. In addition, we provide necessary as well as sufficient conditions for this approach to be valid. In Subsection 1.3.2, we analyze optimality conditions and their consequences for optimal policies. In Subsection 1.3.3, we explore the model using numerical simulations.

1.3.1 Incentive Compatibility

We cast the problem as a sequential mechanism – agents report an initial type θ in the education period and, after uncertainty has materialized, report their productivity a in the working period. The planner assigns initial consumption levels $c_e(\theta)$ and education levels $e(\theta)$ to individuals with innate ability θ. Moreover, with each report there comes a sequence of utility promises for the next period $\{v^w(\theta,a)\}_{a\in[\underline{a},\overline{a}]}$. In the second period, the screening takes place over consumption levels $c_w(\theta,a)$ and labor supply $y(\theta,a)$. All these quantities define an allocation in the economy. Dynamic incentive compatibility is ensured backwards, so we start analyzing the problem from the second period.

1.3.1.1 Working Period Incentive Compatibility

By the revelation principle, we can restrict attention to direct mechanisms. Suppose that in the first period agents have made truthful reports $r_\theta(\theta) = \theta$, albeit this is not necessary and just simplifies the exposition.[7] Conditions for this to be true are given in the next subsection. Conditional on this report, the second period incentive constraint must be met for any history of types (θ, a) and reporting strategy $r_a(a)$:

$$u^w\left(c_w(\theta,a), \frac{y(\theta,a)}{a}\right) \geq u^w\left(c_w(\theta, r_a(a)), \frac{y(\theta, r_a(a))}{a}\right) \qquad \forall a, r_a(a), \theta.$$

Like in a standard Mirrleesian problem, preferences satisfy single-crossing for given first-period reports. For global incentive compatibility it is, hence, necessary and sufficient that all local envelope conditions hold:

$$\frac{\partial v^w(\theta,a)}{\partial a} = u_l^w\left(c_w(\theta,a), \frac{y(\theta,a)}{a}\right)\frac{y(\theta,a)}{a^2} \tag{1.2}$$

and the usual monotonicity condition, stating that $y(\theta,a)$ is non-decreasing in ability levels a, is satisfied:

$$\frac{\partial y(\theta,a)}{\partial a} \geq 0. \tag{1.3}$$

1.3.1.2 Education Period Incentive Compatibility

In the education period, an agent takes into account the effect of her report about θ on future utility. Education period incentive compatibility is ensured if and only if the following double continuum of weak inequalities holds:

$$\beta^e u^e\left(c_e(\theta)\right) + \beta^w \int_{\underline{a}}^{\overline{a}} v^w(\theta,a)dG(a|e(\theta),\theta) \geq$$
$$\beta^e u^e\left(c_e(r_\theta(\theta))\right) + \beta^w \int_{\underline{a}}^{\overline{a}} v^w(r_\theta(\theta),a)dG(a|e(r_\theta(\theta)),\theta)\ \forall\theta, r_\theta(\theta).$$

Let $V(\theta)$ be the associated value function. By using the FOC of an agent's reporting problem, one can easily derive the following envelope condition

$$\frac{dV(\theta)}{d\theta} = \beta^w \int_{\underline{a}}^{\overline{a}} v^w(\theta,a)\frac{\partial g(a|e(\theta),\theta)}{\partial\theta}da. \tag{1.4}$$

[7] The reason is that in the second period the utility is a function of $a, r_a(a)$ and $r_\theta(\theta)$ but not of θ.

As often done in screening problems, our strategy for solving the second-best problem is to work with a relaxed problem with only restrictions (1.2) and (1.4) being imposed and then check ex-post whether incentive compatibility is fulfilled. In the numerical explorations in Section 1.3.3 we find that incentive compatibility is always satisfied and therefore the first-order approach is valid for the primitives we consider.[8] Next, we present a set of sufficient conditions.

Lemma 1.3.1 *Suppose Assumptions 1.2.2 and 1.2.3 hold, conditions* (1.2), (1.3), (1.4) *are satisfied and we have:*
$(i) \frac{\partial y(\theta,a)}{\partial \theta} > 0$,
$(ii) \frac{\partial e(\theta)}{\partial \theta} > 0$,
then the considered allocation is incentive compatible.

Proof: See Appendix 1.A.1. ■
This lemma implies that instead of directly ex-post verifying whether period one incentive compatibility is satisfied in an allocation, one can alternatively check these two simple monotonicity conditions; if they are fulfilled, then the allocation is incentive compatible. Whereas condition (ii) is always fulfilled in our numerical examples, condition (i) was often violated for very low a; we will comment on the reasons in Section 1.3.3 when we present numerical illustrations of the model.

1.3.2 Properties of Constrained Pareto Optimal Allocations

The planner maximizes

$$\int_{\underline{\theta}}^{\overline{\theta}} \left(\beta^e u^e(c_e(\theta)) + \beta^w \int_{\underline{a}}^{\overline{a}} v^w(\theta, a) dG(a|e(\theta), \theta) \right) d\tilde{F}(\theta)$$

subject to (1.2), (1.4) and the resource constraint:

$$\int_{\underline{\theta}}^{\overline{\theta}} \left[\beta^e \left(c_e(\theta) + e(\theta) \right) + \beta^w \int_{\underline{a}}^{\overline{a}} (c_e(\theta, a) - y(\theta, a)) dG(a|e(\theta), \theta) \right] dF(\theta) = 0.$$

We let the planner assign Pareto weights $\tilde{f}(\theta)$ to individuals, depending on their initial skill level. Any distribution of these weights, which we normalize to satisfy $\int_{\underline{\theta}}^{\overline{\theta}} \tilde{f}(\theta) d\theta = 1$, corresponds to one point on the Pareto frontier. $\tilde{F}(\theta) = \int_{\underline{\theta}}^{\theta} \tilde{f}(\theta) d\theta$

[8] Our results of this section on dynamic incentive compatibility are related to previous work in the optimal nonlinear pricing literature by Courty and Li (2000). They study optimal pricing schemes of a monopolist facing consumers with stochastic tastes. In our case the distribution of types tomorrow is endogenous since education is a choice. In recent contributions, Kapicka (2013) as well as Pavan et al. (2012) investigate the robustness and validity of the Mirrleesian first-order approach in a large class of general dynamic environments.

denotes the cumulative Pareto weight. λ_R denotes the multiplier on the resource constraint and $\eta(\theta)$ the multiplier function of the first-period envelope conditions. The planner uses the same discount rate as all individuals. We now characterize the wedges of second-best Pareto optimal allocations.

1.3.2.1 Labor Distortions

The following proposition characterizes the optimal labor wedge.[9] For expositional reasons, we focus on the case where utility is separable in consumption and labor and show the formula for the general case in the appendix.

Proposition 1.3.1 *Suppose preferences are separable of the form $u(c) - \Psi(l)$ where $u'' < 0$ and $\Psi'' > 0$ and further that $u(\cdot) = u^e(\cdot)$. At any constrained Pareto optimum, labor wedges satisfy:*

$$\frac{\tau_y(\theta, a)}{1 - \tau_y(\theta, a)} = \frac{1 + \varepsilon^u(\theta, a)}{\varepsilon^c(\theta, a)} \frac{u'(c_w(\theta, a))}{ag(a|e(\theta), \theta)} \left[\mathcal{A}(\theta, a) + \mathcal{B}(\theta, a)\right],$$

where

$$\mathcal{A}(\theta, a) = G(a|e(\theta), \theta)\Bigg[\int_a^{\overline{a}} \frac{1}{u'(c_w(\theta, a^*))} dG(a^*|e(\theta), \theta) - \frac{1 - G(a|e(\theta)}{G(a|e(\theta), \theta)} \int_{\underline{a}}^{a} \frac{1}{u'(c_w(\theta, a^*))} dG(a^*|e(\theta), \theta)\Bigg]$$

$$\mathcal{B}(\theta, a) = \frac{1}{f(\theta)\lambda_R} \frac{\partial\left[1 - G(a|e(\theta), \theta)\right]}{\partial\theta} \eta(\theta),$$

where $\varepsilon^u(\theta, a)$ ($\varepsilon^c(\theta, a)$) is the uncompensated (compensated) labor supply elasticity of type (θ, a) and

$$\eta(\theta) = \tilde{F}(\theta) - \frac{\int_{\underline{\theta}}^{\theta} \frac{1}{u'(c_e(\theta))} f(\theta) d\theta}{\int_{\underline{\theta}}^{\overline{\theta}} \frac{1}{u'(c_e(\theta))} f(\theta) d\theta}.$$

Proof: See Appendix 1.B.1.1. ∎

To understand this analytical result and relate it to the literature, first assume that

[9] In a recent paper Golosov et al. (2011) provide formulas for dynamic optimal labor wedges with exogenous human capital, connecting them to empirical observables in the spirit of the contributions of Diamond (1998) and Saez (2001) for the static Mirrlees model.

there would be no incentive problem in period one, i.e. θ would be observable. In this case, the term $\mathcal{B}(\theta, a)$ would be zero everywhere because $\eta(\theta)$ would be zero everywhere. Then, the optimal labor wedge schedules for different values of θ would be the optimal insurance arrangement for the respective θ-type against income risk. In fact, we show that the formula resembles the standard formula of Saez (2001) for $\mathcal{B}(\theta, a) = 0$ in Appendix 1.B.1.1. If θ were observable, it would be an immutable *tag*. The planner would want to condition optimal insurance arrangements on θ in an Akerlof (1978) tagging manner.[10] The interpretation would be very standard that optimal effective marginal tax rates are decreasing in the compensated elasticity, typically larger for higher values of risk aversion and that the nonlinear shape of these effective marginal tax rates is to a large extent determined by the respective distribution function $G(a|\cdot, e(\cdot))$.

With θ being unobservable, the government has to take incentive compatibility in the first period into account. This is captured by term $\mathcal{B}(\theta, a)$. First, note that this term is proportional to the respective value of the Lagrangian-multiplier function $\eta(\theta)$. As long as the planner values the utility of low θ-types sufficiently high (i.e. such that $\tilde{F}(\theta)$ is not too low), $\eta(\theta)$ is positive. This is fulfilled for the Utilitarian case, but also for points on the Pareto frontier more in favor of higher θ-types. If, in addition, the rather natural Assumption 1.2.2 applies, term $\mathcal{B}(\theta, a)$ is unambiguously positive and thus is a force towards higher effective marginal labor income tax rates. To get an intuitive understanding for this term, it helps to think about a stylized example. Assume that $G(a|e(\theta^*), \theta^*) = 1$ for all $a > a^*$. Thus, given their choice e, individuals of type θ^* have a zero probability of having a larger labor market skill than a^*. In contrast, assume that $G(a|e(\theta^*), \theta^* + \varepsilon) < 1$ for $\varepsilon \to 0$. In that case $\frac{\partial[1-G(a|e(\theta^*),\theta^*)]}{\partial\theta} \to \infty$ for all $a > a^*$ and therefore $\tau_y(\theta^*, a) = 1$ for all $a > a^*$. Intuitively, effective marginal tax rates of 100% for individuals of type (θ^*, a) with $a > a^*$ have no costs as the mass of individuals whose behavior is distorted is equal to zero. At the same time, these high marginal tax rates make it less attractive for the type with ability $\theta^* + \varepsilon$ to mimic the θ^*-type.

In addition note that the education choice does not play a direct role for the results. If there were no education choices in the first period but individuals would just do nothing, Proposition 1.3.1 would be unchanged. Education only has an indirect effect on the value of optimal effective marginal tax rates through its impact on the distribution of skills. This does not imply that the planner does not take into account the adverse effects of labor supply distortions on the education margin. In

[10]More recently tagging is investigated by Cremer et al. (2010), Mankiw and Weinzierl (2010) as well as Weinzierl (2012).

fact, the social planner takes this into account by subsidizing the education margin as we discuss in the next subsection.
Finally, a no-distortion at the top and bottom result goes through since $\mathcal{B}(\theta,\overline{a}) = \mathcal{B}(\theta,\underline{a}) = \mathcal{A}(\theta,\overline{a}) = \mathcal{A}(\theta,\underline{a}) = 0$.

1.3.2.2 Education Distortions

The following proposition characterizes optimal education policies.

Proposition 1.3.2 *At any constrained Pareto optimum, the education wedge is given by:*

$$\begin{aligned}\tau^e(\theta) =& \frac{\beta^w}{\beta^e}\int_{\underline{a}}^{\overline{a}}(y(\theta,a)-c_w(\theta,a))\frac{\partial g(a|e(\theta),\theta)}{\partial e(\theta)}da\\ &+\frac{\beta^w}{\beta^e}\frac{\eta(\theta)}{\lambda_R f(\theta)}\int_{\underline{a}}^{\overline{a}}\frac{\partial v^w(\theta,a)}{\partial a}\frac{\partial^2 G(a|e(\theta),\theta)}{\partial e(\theta)\partial\theta}da.\end{aligned}$$

Proof: See Appendix 1.B.1.2. ■

The first term captures the expected marginal fiscal gain of an increase in education. One can show that it is always positive under Assumption 1.2.1 (FOSD shift of education) and positive labor wedges. Investing a dollar more into education increases the expected obligation of an agent. The first part of the education wedge exactly offsets this effect from the labor wedge. Bovenberg and Jacobs (2005) have discovered this effect for the static Mirrlees model, whereas we show this fiscal externality part of the wedge extends to the setting with uncertainty, holding in expectation.
We now turn to the second term. Under Assumption 1.2.3 the cross-derivative $\frac{\partial^2 G(a|e(\theta),\theta)}{\partial e(\theta)\partial\theta}$ is negative and $\frac{\partial v^w(\theta,a)}{\partial a}$ is positive everywhere by second-period incentive compatibility. Further, as discussed in the previous subsection, $\eta(\theta)$ is positive along a large part of the Pareto frontier. Then the second part of the education wedge is negative and acts as an implicit tax on education. By distorting education downward, the planner relaxes binding incentive constraints and can redistribute more effectively in line with her preferences. This is a consequence of the complementarity assumption, stating that agents endowed with higher innate skills gain more from education at the margin. The bundle of a lower type, hence, becomes less attractive from the perspective of an agent if education is downward distorted. Such an intuition is familiar from the standard static Mirrlees model concerning positive marginal income tax rates on the interior of the skill set. Relatedly, for this incentive term a zero at the top and at the bottom $(\underline{\theta},\overline{\theta})$ result holds.[11]

[11] Jacobs and Bovenberg (2011) discuss deviations from a first-best rule for the education subsidy for a general earnings function in the case without uncertainty. Our result is similar to their first

1.3.2.3 Savings Distortions

The characteristics of savings distortions depend on the properties of the utility function $u^w\left(c_w, \frac{y}{a}\right)$. In the case of separable preferences, the well explored inverse Euler equation holds (Diamond and Mirrlees, 1978; Rogerson, 1985; Golosov et al. 2003), making it optimal to tax savings at every initial skill level θ, improving the ability of the planner to provide labor supply incentives. In general, the sign of the wedge depends on the exact functional form assumption and especially on the interaction of labor effort and the marginal utility of consumption – see Golosov et al. (2011) for an elaborate discussion of the underlying forces in a dynamic Mirrleesian model.

1.3.3 Numerical Illustration

In this section we numerically explore our model in an illustrative manner. We consider two skill distributions as our primitives $G(a|e,\theta)$ that lead to very similar equilibrium wage distributions and educational expenses for actual given policies. In one of the cases, the distribution function is characterized by a strong complementarity between innate skills and education. In the other case, there is less complementarity and the direct effects of education and innate skills dominate. We solve for the Utilitarian optimum, so $\tilde{f}(\theta) = f(\theta)\ \forall\ \theta$. The utility function is:

$$U(c,l) = \frac{c^{1-\rho}}{1-\rho} - \frac{(y/a)^{\sigma}}{\sigma},$$

where we set $\sigma = 3$, implying a Frisch elasticity of 0.5 and the CRRA coefficient to $\rho = 2$.

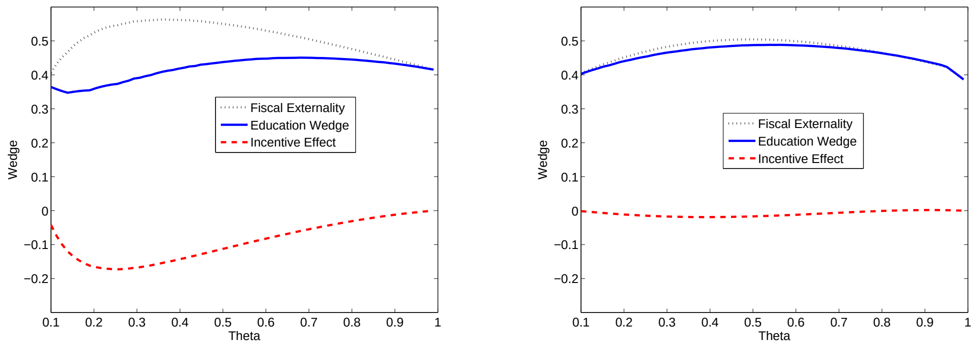
We assume that labor market abilities are distributed log-normally following common practice and impose the location parameter μ to be a function of θ and e. Concerning θ, we assume a uniform distribution within $[0.1, 1]$.

Case (a) - Strong Complementarity: The functional form of the location parameter is:

$$\mu(\theta, e) = 1.7 + 1.5\theta^{0.5}e^{0.15}.$$

result that a complementarity in education and ability leads to a tax on education. They also consider the degree of complementarity between labor supply and education which might call for an education subsidy in contrast. This second effect disappears in our environment since the returns to labor supply – once uncertainty has materialized – are independent from the education choice.

Figure 1.3.1: Optimal Education Wedges



(a) Strong complementarity (b) Strong direct effect

In this case, individuals are the same if they do not acquire any education at all. However, the more education they acquire the stronger are the differences in the location parameters. This inequality in μ is reinforced by the fact that agents have incentives to self-select into different education levels because of heterogeneous returns.

Case (b) - Strong Direct Effect: In the second case we assume,

$$\mu(\theta, e) = 1.5 + \theta + 0.75e^{0.25}.$$

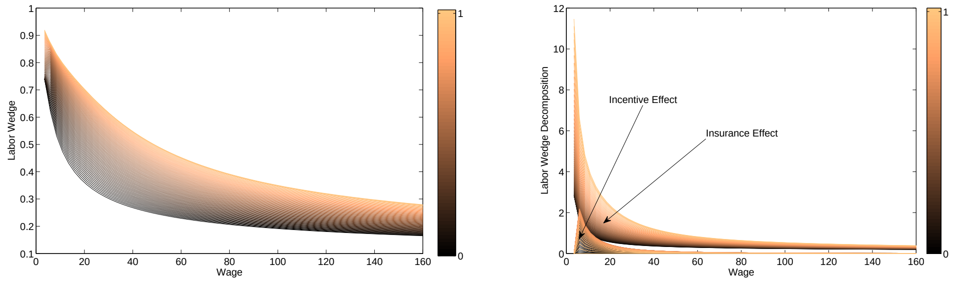
In this case, individuals are already very different from the outset, i.e. if nobody acquires any education. The difference in the location parameter then stays constant for a uniform increase in e across agents. Although $\frac{\partial^2 \mu(\theta,e)}{\partial\theta\partial e} = 0$ in this case, Assumption 1.2.3 is fulfilled for the relevant range. However, innate skills and education are weaker complements as compared to Case (a).

The respective parameters for the two cases as well as the respective constant marginal costs of education were chosen such that given an approximation of the current tax and college subsidy system in the US, the model roughly replicates per-capita expenditures on college education and the centers of the interval of the location parameters of the log-normal distributions is equal to the empirical value of the wage distribution.[12]

Figure 1.3.1 illustrates optimal education wedges for the two cases. In both cases,

[12]Following Gallipoli et al. (2011) we set the labor income tax to a flat rate of 27% and a lump sum transfer of one sixth of labor income per capita. We introduce a yearly education subsidy of 24%. In both cases, for these given policy instruments, average college education expenditures per year are roughly 30% of yearly median income; a long run average for the US (Gallipoli et al., 2011). The realized values of $\mu(\theta, e)$ are within the range $[2.02, 3.34]$, centered around 2.76, the value of the lognormal fit for the US wage distribution found by Mankiw et al. (2009); as them, we set the scale parameter equal to 0.565.

Figure 1.3.2: Optimal Labor Wedges



(a) Optimal Labor Wedges (b) Labor Wedge Decomposition

the optimal allocation features positive implicit education subsidies around 40%, which are relatively flat across innate types. The main difference between the two cases lies in the incentive effect. When innate skills and education are complements, the planner finds it optimal to tax education relative to a first best in line with Proposition 1.3.2. In Case (a) this incentive effect becomes as large as 17% whereas in Case (b) it hovers around zero.

Figure 1.3.2 illustrates the optimal labor wedges from Proposition 1.3.1. Panel (a) displays the optimal labor wedge as a function of income.[13] Darker regions refer to innate low types and lighter regions to innate high types. The picture shows that higher innate types face high labor wedges, whereas the shape of the wedges does not vary with θ.[14] In the next panel (b), we illustrate the decomposition from Proposition 1.3.1 into the insurance term and the incentive term by plotting $\mathcal{A}(\theta, a)$ and $\mathcal{B}(\theta, a)$. The set of insurance effects $\mathcal{A}(\theta, a)$ lies above the set of incentive effects $\mathcal{B}(\theta, a)$. Still, especially at the beginning of the income distribution incentive effects contribute to higher implicit tax rates. The graph also reveals that these incentive effects are of more importance for higher innate types on average.

1.4 Implementation

So far we only considered a direct mechanism, in which individuals make reports about their realized type and the planner assigns bundles of consumption, labor supply and education as functions of the reports. The focus in the characterization of the optimal allocation was on wedges or *implicit* price distortions of the allocation. In this section, we explore two decentralized implementations of constrained

[13] To economize on space we only show the figures for Case (b) here. The graphs for Case (a) turn out to look nearly identical.

[14] Since low incomes the distortions are strongly increasing in θ, condition (i) of Lemma 1.3.1 is typically not fulfilled for low a.

Pareto optima. We focus on utility functions $u\left(c_w, \frac{y}{a}\right)$ with no income effects, i.e. $u\left(c_w - \Psi\left(\frac{y}{a}\right)\right)$ as in Diamond (1998) or Greenwood et al. (1988). We do this for expositional purposes – in a working paper version of the chapter (Findeisen and Sachs, 2012) we discuss the implementation with income effects.[15] Additionally, in the main body, we focus on implementations where education $e(\theta)$ is monotone in type and discuss the case where $e(\theta)$ may be non-monotone in Appendix 1.C.1.1; in this case policies are very similar.

1.4.1 Implementation One: Education Dependent Taxes

The benevolent government offers a set of student grants to the agents. These grants $\mathcal{G}$ are conditional on education. In the working period, there is a tax function, which does not only condition on earnings but also on educational investment.

Proposition 1.4.1 *Suppose there are no income effects and education $e(\theta)$ is strictly monotone. Any constrained Pareto optimal allocation can be implemented by a grant schedule $\mathcal{G}(e)$, an education dependent income tax $T(y,e)$ and a savings tax $T^s(s)$, where*

- $\mathcal{G}(e(\theta)) = e(\theta) + c_e(\theta)$
- $T(y(\theta,a), e(\theta)) = y(\theta,a) - c_w(\theta,a)$
- $T^s(s)$ *is defined as in Appendix 1.C.1.*

Proof: See Appendix 1.C.1 ■

Implementation of Savings Wedges: The savings function $T^s(s)$ is prohibitively high such that all agents choose $s = 0$, hence in this implementation there are no private savings. However, as shown in Werning (2011) this comes without loss of generality: by a Ricardian equivalence argument, we can adjust $G(e(\theta))$ and $T(y(\theta,a), e(\theta))$ with lump-sum transfers and deductibles to arrive at a nonlinear savings tax schedule, which produces non-zero private savings for every agent and the same allocation with the same distortion of consumption across periods. The full argument can be found in Werning (2011).

Implementation of Labor Wedges: Agents enter the second period with no savings as argued above. Their budget constraint then is: $T(y(\theta,a), e(\theta)) = y(\theta,a) - c_w(\theta,a)$. From the agents' optimality conditions for y and c_w it follows that marginal

[15] In general, there is also the issue of needing history dependent taxes even with constant wages over the working life, see Werning (2007). We are grateful to Bas Jacobs for pointing out that this can be overcome by the assumption of no income effects.

tax rates $T_y(y(\theta,a),e(\theta))$ are equal to labor wedges $\tau_y(\theta,a)$ as characterized in Section 1.3.2.1.

Implementation of Education Wedges: In contrast to the optimal labor *wedge*, which equals the optimal labor *tax*, there is no single policy instrument for which the education wedge equals the marginal distortion of the policy. Instead, the government uses two instruments: i) the nonlinear grant schedule $\mathcal{G}(e)$, which depends on education chosen and ii) the labor tax code in the second period. Using the agents' optimality conditions in the proposed implementation one can show that the wedge equals:

$$\tau^e(\theta) = \mathcal{G}'(e) - \int_{\underline{a}}^{\overline{a}} \frac{u'(c_w(\theta,a))}{u'(c_e(\theta))} g(a|e(\theta),\theta) T_e(y(\theta,a),e(\theta))da.$$

A positive value of $\tau^e(\theta)$ encourages education at level θ. The incentive for agents to increase their educational attainment comes from: i) An increase in their grant measured by $\mathcal{G}'(e)$[16] and ii) an increase or decrease in their labor income tax burden for all states, i.e. $T_e(y(\theta,a),e(\theta))$.

1.4.2 Implementation Two: Income-Contingent Loans

The previous implementation required that people with the same income but different levels of education pay different taxes. In reality people might perceive this as a violation of horizontal equity concerns, which could hinder the political feasibility of such policies. In this light we now present a more appealing alternative implementation with only one labor income tax schedule and a repayment scheme of the education grant.[17] Technically, this can be seen as a simple reinterpretation of the previous implementation – we take the tax system of the $\underline{\theta}$-type as the *common* labor income tax schedule and introduce an income-contingent repayment schedule, which conditions on the size of the loan.[18] Together both instruments are sufficient

[16]Theoretically it could be that $\mathcal{G}$ is (partly) decreasing in e if $c_e(\theta)$ is sufficiently decreasing. However, this is rather unlikely and in all our numerical examples we have $c_e'(\theta) > 0$.

[17]Diamond and Saez (2011) argue that practical policy prescription from optimal tax models should not go against commonly held normative views (horizontal equity for example) and limit complexity to a reasonable degree. The second implementation seems in line with these recommendations.

[18]Related implementations are of course possible where the tax function of another θ-type can be the labor income tax schedule in place. The extreme case would just be that income taxes do not exist and all schedules that were interpreted as history-dependent labor income schedules in implementation 1 can now be interpreted as repayment schedules. Taking the labor income tax schedule of the $\underline{\theta}$-type, however, seems to be more natural in our view. Especially in our application of the theory in Section 1.5.

to replicate the optimal labor wedges. Formally we summarize this in the following proposition:

Corollary 1.4.1 *Suppose there are no income effects and education $e(\theta)$ is strictly monotone. Any constrained Pareto optimal allocation can be implemented by a (compulsory) loan schedule $L(e)$, a loan repayment schedule $\Gamma(y, L)$, an income tax $T(y)$ and a savings tax $T^s(s)$ where*

- $L(e(\theta)) = e(\theta) + c_e(\theta)$
- $\Gamma(y(\theta,a), L(e(\theta))) = c_w(\underline{\theta}, \tilde{a}(\underline{\theta}, y(\theta,a))) - c_w(\theta,a)$ *if* $y \in [y(\underline{\theta},\underline{a}), y(\underline{\theta},\overline{a})]$ *and* $\Gamma(y(\theta,a), L(e(\theta))) = y(\theta,a) - c_w(\theta,a)$ *otherwise.*
- $T(y) = y - c_w(\underline{\theta}, \tilde{a}(\underline{\theta}, y))\ \forall\ y\ \in\ [y(\underline{\theta},\underline{a}), y(\underline{\theta},\overline{a})]$ *and* $T = 0$ *otherwise*
- $T^s(s)$ *is defined as in Appendix 1.C.1,*

where $\tilde{a}(\theta, y)$ *is the inverse of* $y(\theta,\cdot)$ *for* a.

Proof: The budget constraint of an individual reads as:

$$\begin{aligned} c_e(\theta) + e(\theta) &\leq L(e(\theta)) \\ c_w(\theta,a) &\leq y(\theta,a) - T(y(\theta,a)) - \Gamma(y(\theta,a), L(e(\theta))), \end{aligned}$$

which is equivalent to the budget constraint in implementation 1 since by definition $\mathcal{G}(e) = L(e)\ \forall\ z$ and $T(y,z) = T(y) + \Gamma(y,z)\ \forall\ y,z$. Hence it is a direct consequence of Proposition 1.4.1. ∎

The similarity to the other implementation is apparent. Using the agents' optimality conditions, one can show that the education wedge equals

$$\tau^e(\theta) = L'(e) - \int_{\underline{a}}^{\overline{a}} \frac{u'(c_w(\theta,a))}{u'(c_e(\theta))} g(a|e(\theta),\theta)\Gamma_L(y(\theta,a), L(e(\theta))) \frac{dL(e(\theta)}{de} da,$$

and the labor wedge equals

$$\tau_y(\theta,a) = T'(y(\theta,a)) + \Gamma_y(y(\theta,a), L(e(\theta))).$$

Education wedges are implemented by the nonlinear loans schedule and how repayment varies with education level. The labor wedge is equal to the marginal tax rate and how loan repayment varies with income.

Note that in Proposition 1.4.1, we assume the loans to be mandatory. In the numerical simulation we check whether this is a restrictive assumption by allowing college graduates to opt out and instead take a loan with fixed repayment, i.e. with a fixed interest rate. For our baseline parameterization, we find that college students participate voluntarily in the government loan system. Finally, notice that we did not impose a cap on repayments so that in theory for some income and education levels, they might exceed the capitalized loan values. In our numerical simulations, we also consider income-contingent repayment polices, which are not allowed to exceed the loans value.

1.5 An Application of the Model: College vs. High-School

We now present an empirically driven application of our model. We limit education to be a binary instead of a continuous choice. Agents either enter the labor market directly after high-school graduation or go to college before working. Additionally, we restrict the analysis to two levels of innate ability levels, one that refers to high school and one that refers to college. These simplifications enable us to parameterize the model using factual and, importantly, estimated counterfactual income distributions from the empirical labor literature (Cunha and Heckman, 2007, 2008). Further, the simplification has the advantage that it is easy to incorporate foregone earnings as an implicit cost of education.

1.5.1 Parametrization

Individuals live for 47 years after they graduate from high-school (age 18-65). Afterwards they enter the labor market directly, or decide to go to college and graduate after four years. We label the two innate types θ_{HS} and θ_{CO}.[19] The incentive constraints read as:

$$\begin{aligned}
&\beta^e u(c_e) + \beta^{wCO} \int_{\underline{a}}^{\overline{a}} v_{CO}(a,\theta_{CO}) g(a|CO,\theta_{CO}) da \\
\geq &\beta^{wHS} \int_{\underline{a}}^{\overline{a}} v_{HS}(a,\theta_{HS}) g(a|HS,\theta_{CO}),
\end{aligned} \tag{1.5}$$

[19] We assume that it is *a priori* optimal that the low type θ_{HS} chooses the lower educational attainment (high school) and that θ_{CO} chooses the higher educational attainment (college).

and

$$\beta^{wHS} \int_{\underline{a}}^{\overline{a}} v_{HS}(a, \theta_{HS}) g(a|HS, \theta_{HS}) da$$
$$\geq \beta^{e} u(c_e) + \beta^{wCO} \int_{\underline{a}}^{\overline{a}} v_{CO}(a, \theta_{CO}) g(a|CO, \theta_{HS}) da, \tag{1.6}$$

where $g(a|CO, \theta_{CO})$ and $g(a|HS, \theta_{HS})$ are the probability density functions (pdfs) of the factual ability distributions and $g(a|HS, \theta_{CO})$ and $g(a|CO, \theta_{HS})$ are the pdfs of the counterfactual ability distributions. The discount factors take into account the different lengths of the periods, i.e. $\beta^{e} = \sum_{t=1}^{4} \beta^{t-1}$, $\beta^{wCO} = \sum_{t=5}^{47} \beta^{t-1}$ and $\beta^{wHS} = \sum_{t=1}^{47} \beta^{t-1}$. Note that college types now have to be compensated for their foregone labor earnings, the implicit cost of college education. To get the ability distributions, we take the factual and counterfactual earnings distributions for high-school graduates plotted in Cunha and Heckman (2007) in Figures 1 and 2.[20] After using a kernel smoother, we append Pareto tails at earnings of $88,000. Finally, we smooth the resulting distribution again to overcome the kink from the appended tail. Given a (linear) approximation of the real world tax system we calibrate the implied skill distributions as input for the model from the optimality conditions of the agents as pioneered by Saez (2001).[21] We assume there is an atom of workers equal to five percent for each distribution reflecting unemployment or disability as in Mankiw et al. (2009). The resulting calibrated skill distributions are illustrated in Figure 1.5.1. The share of high school and college types are set to 64.19% and 35.81%, respectively, as reported in Cunha and Heckman (2008). Following Gallipoli et al. (2011), we set the annual monetary cost of college education to $11,100, roughly a third of median income in our data. The yearly interest rate is set to 4% and the yearly discount factor β to 1/1.04. We work with a CRRA specification and focus on the case with no income effects so that:

$$U(c, y, a) = \frac{\left(c - \frac{(y/a)^{\sigma}}{\sigma}\right)^{1-\rho}}{1-\rho},$$

with $\sigma = 3$, implying a constant labor supply elasticity of 0.5[22] and set $\rho = 2$.[23]

[20]We used the software GetData Graph Digitizer to read out the data from the graphs. Since Cunha and Heckman (2007) consider the present value of lifetime earnings (18-65), we take a 47 years annuity with the same present value, i.e. we take something similar to average annual earnings.

[21]In Appendix 1.D.1, we describe this calibration method.

[22]In unreported simulations, we also varied σ up to 1.66 implying an elasticity of 1.5. The main conclusions of this sections do not change.

[23]Note that savings are not distorted in our application. As we assume an education period of length zero for the high school type, there is no transition from an education to a working period,

Figure 1.5.1: Skill Distributions

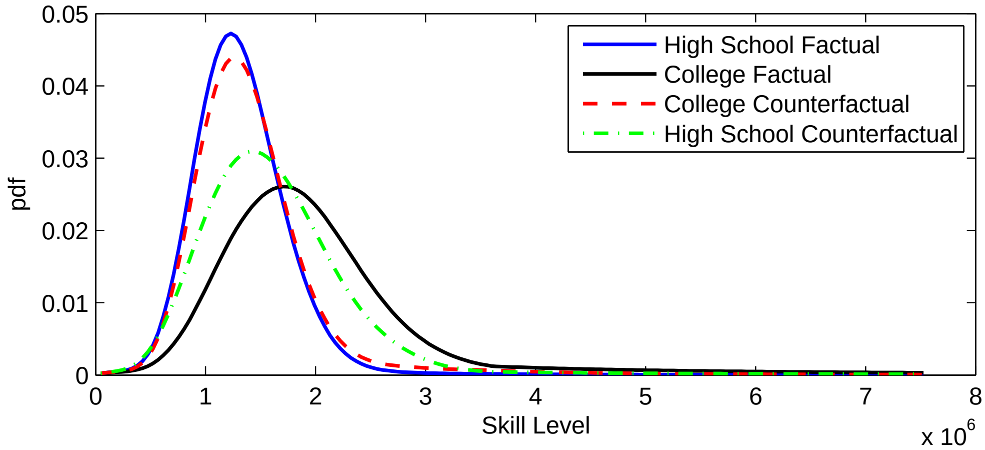
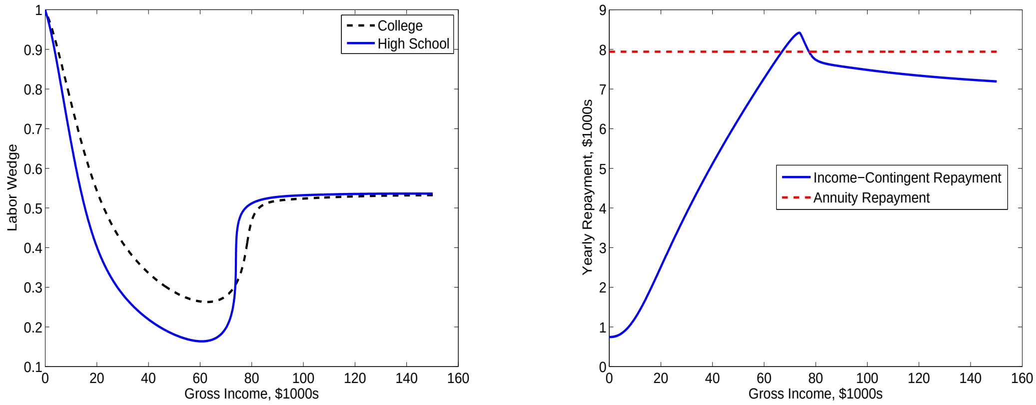


Figure 1.5.2: Utilitarian Optimum



(a) Labor Wedges (b) Repayment Schedule

1.5.2 Policies in Baseline Case

To the best of our knowledge, there exists no systematic evidence on the conditional distributions of top incomes for college graduates and non-graduates separately. In the baseline case, we conservatively assume identical tails for both groups, working with a Pareto parameter of 1.5 (Atkinson et al. 2011; Diamond and Saez, 2011).

1.5.2.1 Second-Best Optimal Policies

Optimal Labor Wedges: Figure 1.5.2(a) displays the optimal labor wedges as a function of yearly income up to $160,000. Both schedules follow a U-shaped pattern,

where the planner would find it optimal to distort savings for the high school type. For the college type, we get a no distortion at the top result for savings and therefore, for him the Euler equation holds between the education and the working period.

reflecting a result from the static Mirrlees problem (Diamond, 1998; Saez, 2001). The intuition for the pattern is simple: for very low incomes, marginal distortions are high for two reasons: first, distorting their labor supply is relatively harmless since they are rather unproductive. Second, the inverse hazard rate $\frac{1-G(a|\cdot,\cdot)}{g(a|\cdot,\cdot)}$ is rather high. Note that $1-G(a|\cdot,\cdot)$ is proportional to the additional revenue generated by the (implicit or explicit) marginal tax rate and $g(a|\cdot,\cdot)$ is the mass of individuals whose labor supply is distorted. For intermediate incomes the density $g(a|\cdot,\cdot)$ strongly increases making distortions more and more harmful, leading to a decrease in optimal distortions. Finally, due to the properties of the Pareto distribution, the ratio $\frac{1-G(a|\cdot,\cdot)}{ag(a|\cdot,\cdot)}$ converges to a constant and as a consequence the labor wedges start to converge.

Looking at Figure 1.5.1, one can see in which way tax distortions are tailored to the different income distributions. At every point of the skill support before the Pareto tail kicks in, college labor distortions generate much bigger mechanical revenue effects for the government. In the top income tails, the wedges converge to almost the same top tax rate (Saez, 2001), with a very small difference caused by the education incentive force $\mathcal{B}(\theta, a)$, which we discussed in the theoretical section of this chapter, that leads to slightly higher top tax rates for high school types to increase the attractiveness of going to college.[24]

Repayment Schedule: We now build on the implementation results from the previous section and illustrate optimal income-contingent repayment schedules. The (common) labor income tax schedule is determined by the high-school labor wedges. Figure 1.5.2(b) shows the yearly repayment of college debt as a function of income. The slope of the repayment schedule is given by the difference in the labor wedges as we outlined in the previous section. As the college wedge lies above the high-school wedge, repayment is increasing in income up to incomes of US-\$80,000. Repayments for college graduates start at about US-\$1,000.[25] Remarkably, the repayment schedule of loans is almost linear with a slope of roughly 0.1, because the difference in the labor wedge is almost constant. Afterwards, there is a very small decreasing range and the repayment schedule flattens out as the top labor wedges converge. In sum, optimal repayments can be very well approximated by an intercept of US-\$1,000, a US-\$1,000 increase in repayment for every US-\$10,000 earned up to earnings of US-\$70,000 and no additional repayments for incomes above that threshold. So al-

[24]Some of these results are related to the simulations of Luttmer and Zeckhauser (2008) who consider a static setting where going to college is purely a signal and not an investment; hence counterfactual and factual distributions are equal.

[25]As the common tax schedule has a generous lump sum element of US-\$18,898, individuals with zero income can afford to pay back this amount.

though we did not place any restrictions on the shape of the repayment schedule, linearity comes very close to the second-best optimum.

The red dotted horizontal line shows the yearly repayment that would occur if individuals chose a standard loan (with a yearly interest rate of 4%) where the repayment is not contingent on income and they repay the same amount every year. As can be seen, only some individuals pay back more than in the income-contingent case. This is sensitive to the interest rate, however. For 3%, e.g., more individuals would pay back more in the income-contingent case. For 5%, nobody would pay back more.

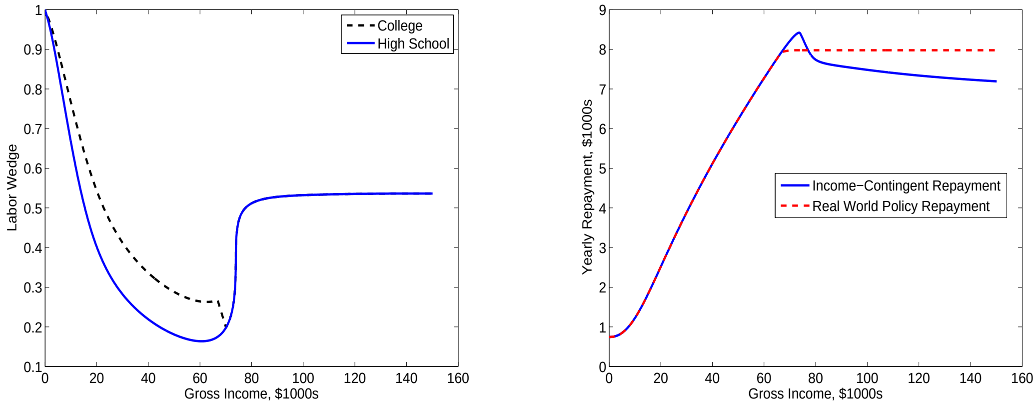
As discussed in the implementation section, we assume the college loan system to be mandatory. We check if this is a restrictive assumption by allowing college graduates to opt out and instead take a loan with a yearly interest rate of 4% to finance tuition and early consumption. We find that given the choice, individuals would opt into the loan system with income contingent repayment rates. This is also true for an interest rate of 3%. However, this is arguably a strict test of the assumption since it is not clear whether individuals would be able to borrow up to their desired amount and might face a substantial risk premium on their interest rate if they borrow in the private market.

1.5.2.2 Real World Policies: Cap on Repayment and Non-Decreasing Repayment

There might be two limitations to the full second-best optimum which could reduce its real world appeal. First, for some (small) range of the income distribution, repayment for college graduates actually exceeds the loan value, as becomes obvious from Figure 1.5.2(b). Second, for high earners the repayment schedule actually decreases in income. These properties are likely to go against commonly held normative views, when it comes to the actual implementation of an income-contingent loan system. Indeed, actual income-contingent repayment systems in the UK or Australia are never decreasing and cap repayment at the loans values. To deal with these concerns, we calculate an allocation which can be implemented with a repayment schedule respecting these constraints – i.e. it is never decreasing and capped at the loan value. In this scenario, effective marginal tax rates for college graduates are adjusted so that they are equal to the marginal tax rates for high school graduates as soon as repayment reaches the capitalized loan value. These modified polices still respect incentive compatibility and budget feasibility, of course.[26]

[26]More technically, we first adjusted the lump sum element of the common labor income tax schedule such that the government budget constraint holds. In case, the resulting allocation is

Figure 1.5.3: Real World Adjustment



(a) Labor Wedges

(b) Repayment Schedule

Figures 1.5.3(a) and 1.5.3(b) show the resulting labor wedges and the repayment schedule. By construction, this repayment system is, of course, inferior in welfare terms to the optimal repayment schedule. As we show in the next subsection, this welfare loss is small.

1.5.2.3 The Welfare Gains From Income-Contingent Repayment

We now aim at quantifying what the potential welfare effects of income-contingency might be and how much of these welfare gains can be obtained by the (ad-hoc) adjusted repayment schedules, which respect a "non-decreasing constraint" and put a cap on repayment.

The natural policy comparison is the case where repayment is not contingent on income. For this benchmark case, we allow the government to freely choose an income tax schedule and also optimize over education subsidies and savings taxes. Formally, the only additional restriction is that individuals with the *same income* should face the *same labor wedge*.

To be able to make such a welfare comparison, the crucial assumption is the absence of income effects. In this case, the restriction that the labor wedge is only a function

not incentive compatible, we adjusted the lump sum elements of the labor income tax and the repayment schedule such that the government budget constraint holds and the incentive constraint of the college type binds.

of current income is simply equivalent to:[27]

$$y(\theta, a) = y(a). \tag{1.7}$$

The following proposition states how a Pareto optimal allocation, subject to (1.7), can be implemented in the binary education model.[28]

Proposition 1.5.1 *Assume there are no income effects. Then any Pareto optimal allocation subject to private information and (1.7) can be implemented by a loan for college students L, a yearly loan repayment Γ and an income tax $T(y)$ that is constant over time, where these policy instruments satisfy*

- $T(y(a)) = y(a) - c(\theta_{HS}, a)$
- $\Gamma = c(\theta_{HS}, a) - c(\theta_{CO}, a)$
- $L = \beta^e(c_e + e)$.

Figure 1.5.4 shows the optimal education independent labor income tax in this case; the optimal marginal tax rates lie between their education dependent counterparts from the second best optimum.

We next calculate the welfare gains from income-contingent repayment schemes for both cases: the unrestricted repayment schedule from Section 1.5.2.1 and the constrained one from Section 1.5.2.2.

In Figure 1.5.5 we present the consumption equivalent welfare gains as the CRRA parameter ρ varies from 1 to 4. First, one can see that the "real world appeal"-repayment schedule is able to reap almost all the welfare gains from income-contingent loans. Second, the gains are increasing in risk-aversion which underscores the role of the loans as an insurance device. For a CRRA coefficient of two, the gains are about 0.32% in the unrestricted and about 0.25% in the restricted case. Thus, roughly 78%

[27] In contrast, in the case with income effects, education-independent marginal tax rates do not imply $y(\theta, a) = y(a)$. Concretely, as individuals with the same a but different θ typically differ concerning their optimal consumption, they choose different labor effort although they face the same labor wedge schedule. Imposing the same consumption and income level on all individuals with the same skill level a would overcome this problem, however, it would be a much stronger restriction on optimal policies. All these arguments do not necessarily imply that one cannot compute optimal history independent policies for the case with income effects. For this case, however, we would not be able to use a first-order approach but instead it would be necessary to check all possible incentive constraints. This would require us to significantly reduce the type space, severely limiting our ability to characterize nonlinear schedules and make welfare comparisons across scenarios.

[28] Naturally, another implementation of this optimum would involve a single labor tax schedule with education-dependent lump-sums and education grants offered by the government.

Figure 1.5.4: Optimal Education Independent Taxes – Baseline Case

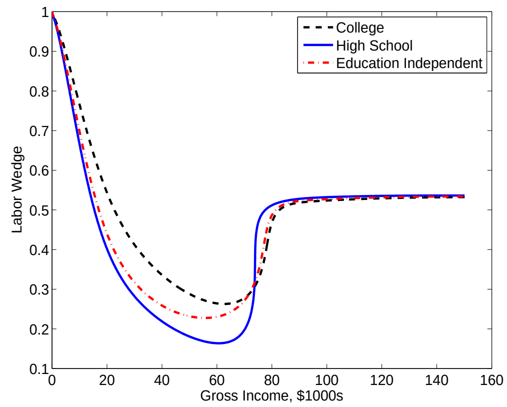
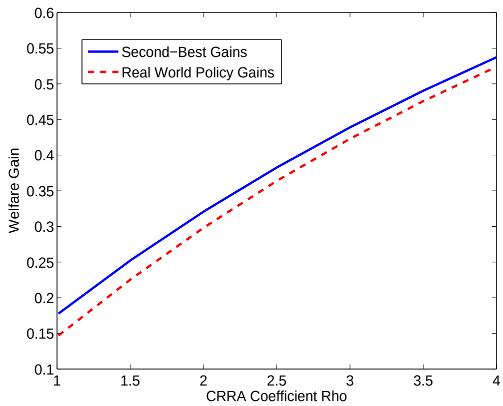


Figure 1.5.5: Welfare Gains – Baseline Case



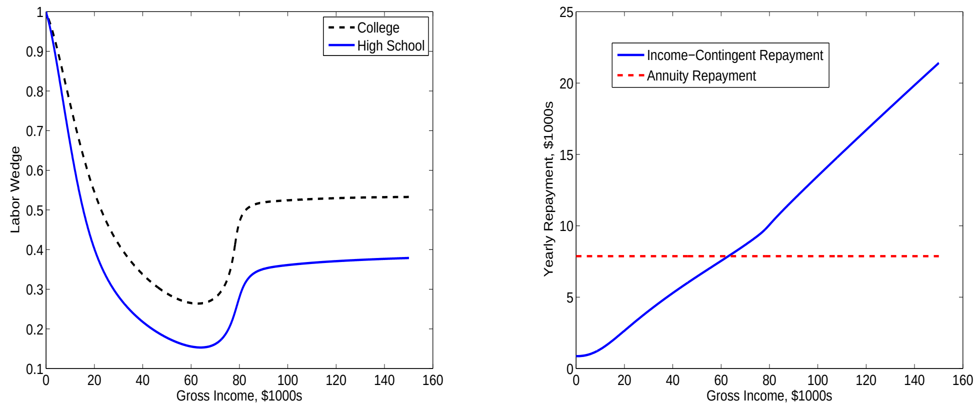
of the welfare gain from the second-best can be reaped with the restricted repayment. In case of an interest rate of 3%, 68% of the welfare gain can be reaped with simpler policies. For an interest rate of 5%, second-best optimal income-contingent repayment would actually never exceed the loan value.

Finally, the welfare gains are evenly distributed in the benchmark case ($\rho = 2$), implying that both the college and the high school type achieve a utility gain of 0.32% of consumptions equivalents. For lower values of ρ, a larger share of the gain is reaped by the high-school graduates, for higher values of ρ the result is reversed.

1.5.3 Policies in Case of Differing Top Income Tails

We now test if and how a different assumption on top incomes across income distributions changes the results. We focus on the case, where the college income distribution has a thicker tail than the high school income distribution. For college graduates, we choose a Pareto parameter of 1.28. For high-school graduates we

Figure 1.5.6: Utilitarian Optimum With Thick College Tails



(a) Labor Wedges (b) Repayment Schedule

choose a Pareto parameter of 3.[29] These values lie within the range of what has been typically found in empirical studies covering many countries and time periods (Atkinson et al, 2011). If we aggregate the two distributions to the aggregate income distribution, we find that the resulting tail for top incomes resembles a Pareto tail with a parameter not far away from 1.5.[30]

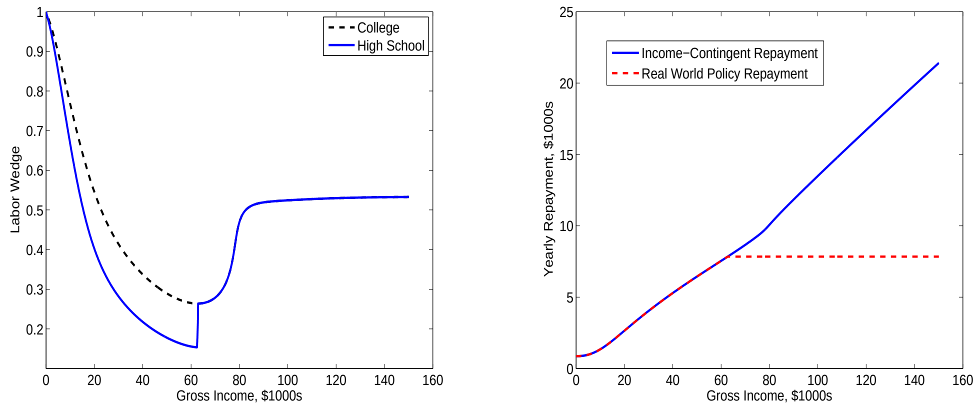
1.5.3.1 Second-Best Optimal Policies

Figures 1.5.6(a) and 1.5.6(b) display the corresponding schedules for labor wedges and the repayment schedule. The college labor wedge now lies above the high school labor wedge everywhere, leading to a strictly increasing repayment schedule. The implicit top tax rate for college graduates is higher than for high-school graduates, driven by the differences in the Pareto parameter. Interestingly, again a simple linear approximation of the repayment schedule with a linear slope of about 11% could almost perfectly implement the second-best optimum. Repayment of college graduates now exceeds the annuity loan value by a much more significant amount and for much bigger fraction of the population. We check again if a college type would prefer not to choose the income-contingent loan in this case and find that the loans indeed have to be compulsory. However, as we show next, one can again construct slightly different policies which respect a cap on repayment. These yield a large share of the welfare gain and do not require the loans to be compulsory.

[29] The top tails are not dependent on innate type θ but are just determined by the education level. In a working paper version of this chapter (Findeisen and Sachs, 2012), we also explore the case in which the tails are determined by innate type θ instead. The results are very similar.

[30] The sum of two Pareto distributions tends to behave like a Pareto distribution, where the heavier tail distribution seems to dominate (Ramsay, 2006). This implies that, in the tails, the resulting aggregate distribution is very close to the college distribution.

Figure 1.5.7: Real World Optimum With Thick College Tails



(a) Labor Wedges (b) Repayment Schedule

1.5.3.2 Real World Polices: Cap on Repayment

As in Section 1.5.2.2, we now adjust the second-best optimum towards policies that satisfy the same two mentioned real-world restrictions. The adjustment we make is slightly different this time. In Section 1.5.2.2, we lowered the labor wedges of the college types such that they equal the optimal ones for the high school types above all income levels, where the second-best repayment starts to exceed the loan value. Here, we do the opposite and increase the labor wedges of the high school types such that they are equal to the college labor wedges. The reason for this is that optimal history independent wedges (see Figure 1.5.8) are closer to the college wedges for high incomes, which is driven by the fatter college top income tail "dominating" the top income tail for the high school types, see footnote 30. The new adjusted policies respect incentive compatibility and budget feasibility. In order to avoid bunching because of a discrete upward jump in marginal tax rates, we smooth out the increase over an interval of roughly US-$5,000. The resulting labor wedges and repayment are illustrated in Figures 1.5.7(a) and 1.5.7(b).

1.5.3.3 The Welfare Gains From Income-Contingent Repayment

As in Section 1.5.2.3, we now calculate the welfare gains over student loans without income-contingent repayment. Due to the differing top income tails, the college and high school wedges are more distinct from each other (see Figure 1.5.8) than in the benchmark case. This yields to welfare gains (see Figure 1.5.9) that are slightly higher. They are 0.36% of lifetime consumption for a CRRA coefficient of 2. Again, the adjusted system respecting a cap can yield a large part of those gains: in fact,

Figure 1.5.8: Optimal Education Independent Taxes – Differing Tails

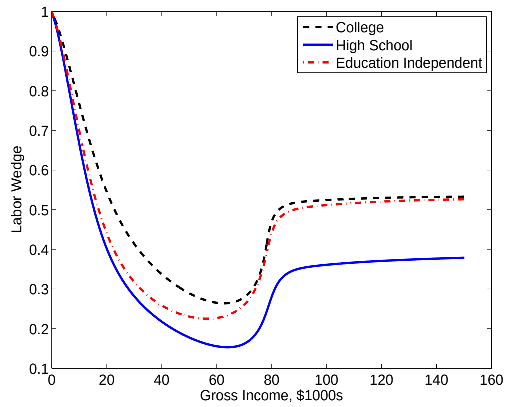
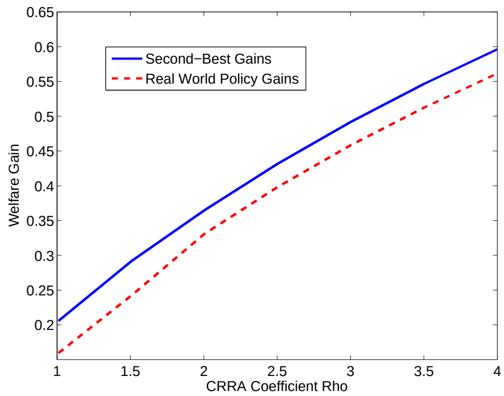


Figure 1.5.9: Welfare Gains – Differing Tails



they lead to a gain of 0.33%, which is almost 92% of the welfare gain. For an interest rate of 3% (5%) the latter value is 75% (95 %).

1.6 Conclusion

This chapter has studied the implications of endogenous education decisions before labor market entry on Pareto optimal tax policies in a dynamic environment with heterogeneous agents and uncertainty. An attractive way to decentralize Pareto optimal allocations is to have the government support students to finance consumption and tuition during education. During their working life students pay back these loans, contingent on income and loan size. We therefore make a second-best argument in favor of student loans with income-contingent repayment rates and, in addition, provide guidance for the optimal design of such repayment schedules.

We have abstracted from several aspects that can be tackled in future work. First, we have abstracted from initial wealth heterogeneity. In an environment where individuals differ concerning the income and wealth of their parents, typically the question arises to what extent optimal education policies should depend on parents' income and wealth. Closely related to this question, Gelber and Weinzierl (2012) have recently taken up the task of showing how the optimal history-independent tax system changes, when children's abilities depend on parents' financial resources. Second, due to our assumption that all labor market risk is realized directly after labor market entry, some aspects concerning the optimal timing of repayment were naturally disregarded. Relatedly, we did no consider human capital accumulation after labor market entry like on-the-job training.[31] Third, we assumed full commitment to all policies from the government side. In Chapter 4, lack of commitment is addressed in a more stylized model.

[31] In ongoing research, Stantcheva (2012) considers optimal taxation and human capital taxation in a life cycle economy, which encompasses on-the-job training.

Appendix 1.A Incentive Compatibility

1.A.1 Proof of Lemma 1.3.1

Denote by $U(\theta, r)$ the utility obtained by a type θ reporting r. Consider some admissible reporting strategy $r(\theta) = \theta'$. Then consider the following two derivatives

$$\begin{aligned}\frac{\partial U(\theta,\theta')}{\partial r(\theta)} &= u_c\left(c_e\left(\theta'\right)\right)\frac{\partial c_e(\theta')}{\partial r(\theta)} + \beta\int_{\underline{a}}^{\overline{a}}\frac{\partial v^w(\theta',a)}{\partial r(\theta)}g(a|e(\theta'),\theta)da\\ &+\frac{\partial e(\theta')}{\partial r(\theta)}\int_{\underline{a}}^{\overline{a}}v^w(\theta',a)\frac{\partial g(a|e(\theta'),\theta)}{\partial e(\theta')}da\end{aligned}$$

and

$$\begin{aligned}0 = \frac{\partial U(\theta',\theta')}{\partial r(\theta)} &= u_c\left(c_e\left(\theta'\right)\right)\frac{\partial c_e(\theta')}{\partial r(\theta)} + \beta\int_{\underline{a}}^{\overline{a}}\frac{\partial v^w(\theta',a)}{\partial r(\theta)}g(a|e(\theta'),\theta')da\\ &+\frac{\partial e(\theta')}{\partial r(\theta)}\int_{\underline{a}}^{\overline{a}}v^w(\theta',a)\frac{\partial g(a|e(\theta'),\theta')}{\partial e(\theta')}da\end{aligned}$$

which is equal to zero by first-order incentive compatibility. Subtracting from one another gives:

$$\begin{aligned}\frac{\partial U(\theta,\theta')}{\partial r(\theta)} = \beta\int_{\underline{a}}^{\overline{a}}&\left[\frac{\partial v^w(\theta',a)}{\partial r(\theta)}\left(g(a|e(\theta'),\theta) - g(a|e(\theta'),\theta')\right)\right.\\ &\left.+\frac{\partial e(\theta')}{\partial r(\theta)}v^w(\theta',a)\left(\frac{\partial g(a|e(\theta'),\theta)}{\partial e(\theta')} - \frac{\partial g(a|e(\theta'),\theta')}{\partial e(\theta')}\right)\right]da.\end{aligned}$$

We are now looking when this last expression always has the same sign as the difference $(\theta - \theta')$, which is sufficient for global incentive compatibility. For $(\theta - \theta') > 0$, using Assumption 2, the first line is positive if $\frac{\partial v^w(\theta',a)}{\partial r(\theta)}$ (or equivalently $\frac{\partial v^w(\theta,a)}{\partial \theta}$ in a truthful mechanism) is increasing in a. This can be shown to be equivalent to $\frac{\partial y(\theta,a)}{\partial \theta} > 0$ using the envelope theorem, which is part (i) of the lemma. Using assumption 3, one can show, that the second line is positive if $\frac{\partial e(\theta')}{\partial r(\theta)} > 0$ or equivalently $\frac{\partial e(\theta)}{\partial \theta} > 0$ in a truthful mechanism, which is part (ii) of the lemma. That (1.2) and (1.3) are sufficient for global incentive compatibility in the working period is a routine exercise and a proof can be found, for example in Salanié (2003).

Appendix 1.B Pareto Optimal Allocations

1.B.1 Optimal Labor and Education wedges

We start by stating the objective for the case of separable preferences of the form $u(c) - \Psi(l)$, where Ψ are the convex utility costs of labor. Further, we assume that $u(\cdot) = u^e(\cdot)$. After integrating by parts and using the transversality conditions $\eta(\underline{\theta}) = \eta(\overline{\theta}) = 0$ as well as $\mu(\theta, \underline{a}) = \mu(\theta, \overline{a}) = 0 \ \ \forall\ \theta$, the Lagrangian for the social planner's problem reads as[32]

[32] With more general preferences, the fourth line would be $\beta^w \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} \gamma\left(v^w(\theta,a), \frac{y(\theta,a)}{a}\right) dG(a|e(\theta),\theta)dF(\theta)$ with $\gamma(v,l)$ being the inverse func-

$$\begin{aligned}\max_{c_e(\theta),v^w(\theta,a),e(\theta),y(\theta,a)} \mathcal{L} =& \beta^e \int_{\underline{\theta}}^{\overline{\theta}} u(c_e(\theta))d\tilde{F}(\theta) \\
&+ \beta^w \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} v^w(\theta,a)dG(a|e(\theta),\theta)d\tilde{F}(\theta) \\
&+ \beta^w \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} y(\theta,a)dG(a|e(\theta),\theta)dF(\theta) \\
&- \beta^w \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} u^{-1}\left[v^w(\theta,a) + \Psi\left(y(\theta,a)/a\right)\right] dG(a|e(\theta),\theta)dF(\theta) \\
&- \beta^e \lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \left(c_e(\theta) + e(\theta)\right) dF(\theta) \\
&- \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} \left(\mu'(\theta,a)v^w(\theta,a) + \mu(\theta,a)\Psi'\left(\frac{y(\theta,a)}{a}\right)\frac{y(\theta,a)}{a^2}\right) dad\theta \\
&- \int_{\underline{\theta}}^{\overline{\theta}} \eta'(\theta)\left[\beta^e u(c_e(\theta)) + \beta^w \int_{\underline{a}}^{\overline{a}} v^w(\theta,a)dG(a|e(\theta))da\right] d\theta \\
&- \beta^w \int_{\underline{\theta}}^{\overline{\theta}} \eta(\theta) \int_{\underline{a}}^{\overline{a}} v^w(\theta,a)\frac{\partial g(a|e(\theta),\theta)}{\partial\theta} dad\theta.\end{aligned}$$

With first-order conditions:

$$u'(c_e(\theta))(\tilde{f}(\theta) - \eta'(\theta)) - \lambda_R f(\theta) = 0 \tag{1.8}$$

$$\begin{aligned}&\left(\tilde{f}(\theta) - \eta'(\theta)\right) g(a|e(\theta),\theta) - \lambda_R \frac{1}{u'(c_w(\theta,a))} g(a|e(\theta),\theta)f(\theta) - \frac{\mu'(\theta,a)}{\beta^w} \\
&- \frac{\partial g(a|z(\theta),\theta)}{\partial\theta}\eta(\theta) = 0.\end{aligned} \tag{1.9}$$

$$\begin{aligned}&\lambda_R g(a|e(\theta),\theta)f(\theta) - \frac{\mu(\theta,a)}{\beta^w}\left[\Psi''\left(\frac{y(\theta,a)}{a}\right)\frac{y(\theta,a)}{a^3} + \frac{1}{a^2}\Psi'\left(\frac{y(\theta,a)}{a}\right)\right] \\
&- \lambda_R g(a|e(\theta),\theta)f(\theta)\frac{\Psi'\left(\frac{y(\theta,a)}{a}\right)}{au'(c_w(\theta,a))} = 0,\end{aligned} \tag{1.10}$$

$$\begin{aligned}&\tilde{f}(\theta)\beta^w \int_{\underline{a}}^{\overline{a}} v^w(\theta,a)\frac{\partial g(a|e(\theta),\theta)}{\partial e(\theta)}da + \beta^w\lambda_R f(\theta) \int_{\underline{a}}^{\overline{a}} \frac{\partial g(a|e(\theta),\theta)}{\partial e(\theta)}\left(y(\theta,a) - c_w(\theta,a)\right) da \\
&-\eta'(\theta)\beta^w \int_{\underline{a}}^{\overline{a}} v^w(\theta,a)\frac{\partial g(a|e(\theta),\theta)}{\partial e(\theta)}da - \beta^w\eta(\theta) \int_{\underline{a}}^{\overline{a}} v^w(\theta,a)\frac{\partial^2 g(a|e(\theta),\theta)}{\partial e(\theta)\partial\theta}da - \beta^e\lambda_R f(\theta) = 0\end{aligned} \tag{1.11}$$

Combining equations (1.8) and (1.9) and integrating directly gives the inverse Euler equation.

---

tion of u over c. The sixth line would be $-\beta^e\lambda_R \int_{\underline{\theta}}^{\overline{\theta}} \left(c_e(\theta) + e(\theta)\right) dF(\theta) - \int_{\underline{\theta}}^{\overline{\theta}} \int_{\underline{a}}^{\overline{a}} \left(\mu'(\theta,a)v^w(\theta,a) + \mu(\theta,a)u_l\left(\gamma\left(v^w(\theta,a), \frac{y(\theta,a)}{a}\right), \frac{y(\theta,a)}{a}\right)\frac{y(\theta,a)}{a^2}\right) dad\theta$ instead.

1.B.1.1 Proposition 1.3.1

Rewriting (1.10):

$$\lambda_R g(a|e(\theta),\theta)f(\theta)\left[1-\frac{\Psi'\left(\frac{y(\theta,a)}{a}\right)}{au'(c_w(\theta,a))}\right]$$

$$-\frac{1}{\beta^w}\mu(\theta,a)\left[\Psi''\left(\frac{y(\theta,a)}{a}\right)\frac{y(\theta,a)}{a^3}+\frac{1}{a^2}\Psi'\left(\frac{y(\theta,a)}{a}\right)\right]=0.$$

Dividing by $\frac{\Psi'}{au'}$ and $\lambda_R g(a|e,\theta)f(\theta)$ and using the definition of the labor wedge, i.e. $u'(1-\tau_y)=\Psi'\frac{1}{a}$ yields

$$\frac{\tau_y(\theta,a)}{1-\tau(\theta,a)}=\frac{1}{\beta^w}\frac{\mu(\theta,a)}{\lambda_R g(a|e(\theta),\theta)f(\theta)a}\left[\frac{\Psi''\frac{y}{a^2}+\Psi'\frac{1}{a}}{\frac{\Psi'}{au'}}\right],$$

which can be written as

$$\frac{\tau_y(\theta,a)}{1-\tau(\theta,a)}=\frac{1}{\beta^w}\cdot\frac{\mu(\theta,a)}{\lambda_R g(a|e(\theta),\theta)f(\theta)a}\frac{1+\varepsilon_u(\theta,a)}{\varepsilon_c(\theta,a)},$$

where $\frac{\Psi'\frac{1}{a}}{\Psi''\frac{y}{a^2}+\Psi'\frac{1}{a}}=\frac{1+\varepsilon_u(\theta,a)}{\varepsilon_c(\theta,a)}$ can be shown by simple algebra, see Saez (2001, p.227). In particular, with the isoelastic specification used in the computations $\frac{(y/a)^\sigma}{\sigma}$ one can verify that this term is equal to $\frac{1}{\sigma}$.

The multiplier $\mu(\theta,a)$ can be obtained using (1.9) and (1.8):

$$\frac{\mu(\theta,a)}{\beta^w}=\frac{\lambda_R f(\theta)}{u'(c_e(\theta))}G(a|e(\theta),\theta)-\lambda_R f(\theta)\int_{\underline{a}}^{a}\frac{1}{u'(c_w(\theta,a^*))}dG(a^*|e(\theta),\theta)$$

$$-\frac{\partial G(a|e(\theta),\theta)}{\partial\theta}\eta(\theta^*),$$

yielding:

$$\frac{\tau_y(\theta,a)}{1-\tau(\theta,a)}=\frac{1+\varepsilon_u(\theta,a)}{\varepsilon_c(\theta,a)}\frac{u'(c_w(\theta,a))}{ag(a|e(\theta),\theta)}\left[\mathcal{A}(\theta,a)+\mathcal{B}(\theta,a)\right]$$

where

$$\mathcal{A}(\theta,a)=\frac{G(a|e(\theta),\theta)}{u'(c_e(\theta))}-\int_{\underline{a}}^{a}\frac{1}{u'(c_w(\theta,a^*))}dG(a^*|e(\theta),\theta)$$

$$\mathcal{B}(\theta,a)=-\frac{1}{f(\theta)\lambda_R}\frac{\partial G(a|e(\theta),\theta)}{\partial\theta}\eta(\theta).$$

Using the inverse Euler equation, the term $\mathcal{A}(\theta, a)$ can be written as in the proposition:

$$\begin{aligned}
&\frac{G(a|e(\theta),\theta)}{u'(c_e(\theta))} - \int_{\underline{a}}^{a} \frac{1}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta) \\
&= G(a|e(\theta),\theta) \int_{\underline{a}}^{\overline{a}} \frac{1}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta) - \int_{\underline{a}}^{a} \frac{1}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta) \\
&= G(a|e(\theta),\theta) \int_{\underline{a}}^{a} \frac{1}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta) + G(a|e(\theta),\theta) \int_{a}^{\overline{a}} \frac{1}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta) \\
&- \int_{\underline{a}}^{a} \frac{1}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta) \\
&= G(a|e(\theta),\theta) \int_{a}^{\overline{a}} \frac{1}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta) - (1 - G(a|e(\theta),\theta)) \int_{\underline{a}}^{a} \frac{1}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta).
\end{aligned}$$

From (1.8), $\eta(\theta)$ is given by:

$$\eta(\theta) = \tilde{F}(\theta) - \lambda_R \int_{\underline{\theta}}^{\theta} \frac{1}{u'(c_e(\theta))} f(\theta) d\theta.$$

The direct benefit of raising utils for agents with skills lower than θ is $\tilde{F}(\theta)$. The monetary cost is $\int_{\underline{\theta}}^{\theta} \frac{1}{u'(c_e(\theta))} f(\theta) d\theta$, transformed into welfare units by λ_R.

Relation to the formula of Saez (2001):
The insurance part of the labor wedge can be expressed as in Saez (2001) for our case with separable preferences. This relation applies if agents do not differ ex-ante. By some abuse of notation, then $\mathcal{B}(\theta, a) = 0$ and for $\mathcal{A}(\theta, a)$, using the inverse Euler equation, we obtain

$$\begin{aligned}
\mathcal{A}(\theta, a) &= \int_{\underline{a}}^{\overline{a}} \frac{G(a|e(\theta),\theta)}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta) - \int_{\underline{a}}^{a} \frac{1}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta) \\
&= \int_{\underline{a}}^{\overline{a}} \frac{G(a|e(\theta),\theta)}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta) - \int_{\underline{a}}^{\overline{a}} \frac{1}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta) \\
&+ \int_{a}^{\overline{a}} \frac{1}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta) \\
&= \int_{a}^{\overline{a}} \frac{1}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta) - \int_{\underline{a}}^{\overline{a}} \frac{1 - G(a|e(\theta),\theta)}{u'(c_w(\theta,a^*))} dG(a^*|e(\theta),\theta),
\end{aligned}$$

where the second equality follows from the transversality condition. This term can be expressed as in Saez (2001) as shown by Mankiw et al. (2009) in their online appendix.
General Preferences: Carrying out the analogous steps with a general utility function $u(c, l)$ we get:

$$\frac{\tau_y(\theta,a)}{1-\tau(\theta,a)} = \frac{1+\varepsilon_u(\theta,a)}{\varepsilon_c(\theta,a)} \frac{u_c(\theta,a)}{a g(a|e(\theta),\theta)} \int_{a}^{\overline{a}} \exp\left(- \int_{a}^{x} \frac{u_{c,l}(\theta,s)}{u_c(\theta,s)} \frac{l(\theta,a)}{a} ds \right) \times [\mathcal{A}(\theta,x) + \mathcal{B}(\theta,x)]\, dx,$$

where $\mathcal{A}(\theta, x) = g(x|e(\theta),\theta) \left(\frac{1}{u_c(c_w(\theta,x))} - \frac{1}{u_c(c_e(\theta))} \right)$ and $\mathcal{B}(\theta, x) = \frac{\eta(\theta)}{\lambda f(\theta)} \frac{\partial g(x|e(\theta),\theta)}{\partial \theta}$.

1.B.1.2 Proposition 1.3.2

Plugging (1.8) into (1.11) gives:

$$\frac{\lambda_R f(\theta)}{u'(c_e(\theta))}\beta^w \int_{\underline{a}}^{\overline{a}} v^w(\theta,a)\frac{\partial g(a|e(\theta),\theta)}{\partial e(\theta)}da + \beta^w \lambda_R f(\theta)\int_{\underline{a}}^{\overline{a}} \frac{\partial g(a|e(\theta),\theta)}{\partial e(\theta)}\left(y(\theta,a) - c_w(\theta,a)\right)da$$
$$- \beta^w \eta(\theta)\int_{\underline{a}}^{\overline{a}} v^w(\theta,a)\frac{\partial^2 g(a|e(\theta),\theta)}{\partial e(\theta)\partial\theta}da - \beta^e \lambda_R f(\theta) = 0.$$

Proposition 1.3.2 directly follows. Note that the relevant first-order conditions are identical for general utility function, so that the formula for the optimal wedge in the proposition also applies.

Appendix 1.C Implementation

1.C.1 Proof of Proposition 1.4.1

Starting from a direct mechanism we show in four steps that optimal allocations can indeed be implemented with the policy instruments as defined in Proposition 1.4.1. The idea to work with a history independent savings tax builds upon the work of Werning (2011).

Step 1: Introduce Savings

Imagine the constrained efficient allocation is implemented by a direct mechanism. From that point on, assume that individuals could freely save s at rate R. Let r_1 denote the report about θ. With savings tax $T^s(s, r_1)$, the budget constraints read as

$$\tilde{c}_e(r_1) + s = c_e(r_1) \tag{1.12}$$

$$\tilde{c}_w(r_1, r_2) = c_w(r_1, r_2) + Rs - T^s(s, r_1). \tag{1.13}$$

Define the optimal report r_2 about a, for a given report r_1 about θ, a given savings tax schedule $T^s(s, r_1)$ and a given level of savings s:

$$r_2^*(a, r^1, s, T^s) = \arg\max_{r_2} u\left[c_w(r_1, r_2) + Rs - T^s(s, r_1) - \psi\left(\frac{y(r_1, r_2)}{a}\right)\right].$$

Then the optimal report in period one, for a given level of savings and a given savings tax schedule $T^s(s, r_1)$, is defined by

$$r_1^*(\theta, s, T^s(r^1, s)) = \arg\max_{r_1} u(c_e(r_1) - s)$$
$$+ \beta\int_{\underline{a}}^{\overline{a}} u\left[c_w(r_1, r_2^*) + Rs - T^s(s, r_1) - \psi\left(\frac{y(r_1, r_2^*)}{a}\right)\right]dG(a|e(r^1), \theta). \tag{1.14}$$

Then define a hypothetical tax schedule $T^*(r_1, s, \theta)$ for each θ implicitly by[33]

$$V(\theta) = V(\theta, s, r_1^*, T^*(r_1, s, \theta))\ \forall\ s.$$

[33] Recall that $V(\theta)$ is the value function of a truth teller of type θ.

This hypothetical tax schedule would make individuals of type θ indifferent between truth telling and the optimal joint deviation for any s. It is hypothetical since it does not only depend on the report r_1, which is observable but also on the unobservable type θ. However, we know that for each θ such a tax schedule exists. Therefore, taking the upper envelope over these functions yields a savings tax function $\hat{T}(s, r_1)$ that also implements zero savings and is feasible since it does not condition on θ:

$$\hat{T}(s, r_1) = \sup_{\theta} T^*(r_1, s, \theta). \tag{1.15}$$

Lemma 1.C.1 *A constrained efficient allocation can be implemented by a direct mechanism extended by a savings choice and history-dependent savings tax schedules* $\hat{T}^s(s, r_1)$.

Step 2: Make the savings tax history-independent
A simple way to make the savings tax history-independent is to take the upper envelope of all functions $T^s(s, r_1)$, i.e.

$$T^s(s) = \sup_{r_1} \hat{T}^s(s, r_1). \tag{1.16}$$

Lemma 1.C.2 *A constrained efficient allocation can be implemented by a direct mechanism extended by a savings choice and a history-independent savings tax schedule* $T^s(s)$.

Note that this savings tax function $T^s(s)$ is not differentiable and implies zero savings. As Werning (2011) shows one can, using Ricardian equivalence arguments, also construct a history-independent savings tax function that is differentiable and yields non-zero savings choices.
Step 3: Allow for labor-leisure decisions
To get closer to a decentralized implementation now assume the following extended direct mechanism.

1. Individuals report $r(\theta)$
2. They get assigned 'income to consume' $c_e(\theta)$
3. They face the savings tax schedule $T^s(s)$ and save $s(\theta) = 0$
4. In period two, instead of directly revealing their type, individuals of type θ face an income tax schedule that is defined by

$$T(y(\theta, a), e(\theta)) = y(\theta, a) - c_w(\theta, a) \;\forall\; a.$$

By the same arguments as in the standard Mirrlees model it follows that this extended direct mechanism can also implement the constrained efficient allocations. We can summarize this in the following lemma.

Lemma 1.C.3 *A constrained efficient allocation can be implemented by a direct mechanism in the first period extended by a savings choice and a history-independent differentiable savings tax schedule* $\tilde{T}^s(s)$ *and a history-dependent labor income tax schedule* $T(y, e)$ *in period two.*

Step 4: Complete Decentralization – allow for educational investment

1. Individuals buy (or tell the government that they want to buy) $e(\theta)$ units of education
2. They get assigned a student loan $\mathcal{G}(e(\theta)) = c_e(\theta) + e(\theta)$ (and are obliged to actually buy $e(\theta)$ units of education)
3. They face the savings tax schedule $T^s(s)$ and save $s(\theta) = 0$
4. in period two, instead of directly revealing their type, individuals of type θ face an income tax schedule that is defined by

$$T(y(\theta,a), e(\theta)) = y(\theta,a) - c_w(\theta,a) \ \forall \ a$$

Since the mechanism in step 4 is just a reformulation of the mechanism in step 3 this directly leads us to Proposition 1.4.1.

1.C.1.1 Discussion of Implementations with Non-Monotone Education

If education is not strictly monotone, it is not enough to condition tax and grant schedules on education for education levels which are assigned to more than one type. In this case for those respective education levels, the planner can augment the system of education grants, such that there is no more a unique grant per education level but a set of grants, which contains the respective correct grant. More formally, if the education level e^* is the optimal education level for all individuals within a certain set $\Theta(e^*)$, then the set of grants assigned to education level e^* must contain $\mathcal{G}(e^*, \theta^*) = c_e(\theta^*) + e^*$, for each $\theta^* \in \Theta(e^*)$. Every education dependent tax function associated with a level e^*, can then in addition be conditioned on consumption during education $c_e(\theta^*)$. Analogously, the planner can offer multiple loans sizes per education level, and repayment schedules which condition on the loan size, income and early consumption.

Appendix 1.D Application

1.D.1 Calibration of Skill Distribution

From Cunha and Heckman (2007, 2008), we only have the distribution of labor incomes not the distribution skills. Saez (2001) proposed to uncover the underlying skill distribution from the first-order conditions of the agents. To uncover the skill a of an individuals that earns y when facing the marginal tax rate T' one only has to solve the first-order condition for a. As the utility function is of the form

$$U(c,y,a) = \frac{\left(c - \frac{(y/a)^\sigma}{\sigma}\right)^{1-\rho}}{1-\rho},$$

and $c = y - T(y)$, the first-order condition reads as

$$(1 - T') = (y/a)^{\sigma-1}\frac{1}{a},$$

which can then be solved for a. To get the respective numbers for T', we consider the same approximation of the US tax code as in Section 1.3.3, i.e. we assume T' to be 27% for each income level. Finally, note that this also implies that changing the assumed elasticity of labor supply implies a different distribution of skills for a given distribution of incomes. As Saez (2001, p.223) notes "...changing the elasticity parameter without changing the skill distribution, as usually done in numerical simulations, might be misleading because changing the elasticities modifies the resulting income distribution and thus might affect optimal rates also through this indirect effect."

Bibliography of Chapter 1

ACEMOGLU, D. AND D. AUTOR (2011): "Skills, Tasks and Technologies: Implications for Employment and Earnings," in *Handbook of Labor Economics Vol. 4*, ed. by O. Ashenfelter and D. Card, Elsevier, 1043–1171.

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

ANDERBERG, D. (2009): "Optimal Policy and the Risk Properties of Human Capital Reconsidered," *Journal of Public Economics*, 93, 1017–1026.

ANDERBERG, D. AND F. ANDERSSON (2003): "Investments in Human Capital, Wage Uncertainty, and Public Policy," *Journal of Public Economics*, 87, 1521 – 1537.

ATKINSON, A. B., T. PIKETTY, AND E. SAEZ (2011): "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 49, 3–71.

BEST, M. AND H. KLEVEN (2013): "Optimal Income Taxation with Career Effects of Work Effort," *Working Paper, London School of Economics.*

BOHACEK, R. AND M. KAPICKA (2008): "Optimal Human Capital Policies," *Journal of Monetary Economics*, 55, 1–16.

BOVENBERG, L. AND B. JACOBS (2005): "Redistribution and Education Subsidies are Siamese Twins," *Journal of Public Economics*, 89, 2005–2035.

CARNEIRO, P. AND J. HECKMAN (2005): " Human Capital Policies," in *Inequality in America: What Role for Human Capital Policies*, ed. by J. Heckman and A. Krueger, The MIT Press, 77–240.

CHAPMAN, B. (2006): "Income contingent loans for higher education: International reforms," *Handbook of the Economics of Education*, 2, 1435–1503.

Chen, S. (2008): "Estimating the Variance of Wages in the Presence of Selection and Unobserved Heterogeneity," *Review of Economics and Statistics*, 90, 275–289.

Courty, P. and H. Li (2000): "Sequential Screening," *The Review of Economic Studies*, 67, 697–717.

Cremer, H., F. Gahvari, and J. Lozachmeur (2010): "Tagging and Income Taxation: Theory and an Application," *American Economic Journal: Economic Policy*, 2, 31–50.

Cunha, F. and J. Heckman (2007): "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling," *Labour Economics*, 14, 870–893.

——— (2008): "The Evolution of Labor Earnings Risk in the US Economy," *Working Paper, University of Chicago.*

Da Costa, C. E. and L. J. Maestri (2007): "The risk Properties of Human Capital and the Design of Government Policies," *European Economic Review*, 51, 695–713.

Diamond, P. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88, 83–95.

Diamond, P. A. and J. A. Mirrlees (1978): "A model of Social Insurance with Variable Retirement," *Journal of Public Economics*, 10, 295–336.

Diamond, P. A. and E. Saez (2011): "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25, 165–90.

Eaton, J. and H. S. Rosen (1980): "Taxation, human capital, and uncertainty," *The American Economic Review*, 70, 705–715.

Findeisen, S. and D. Sachs (2012): "Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans," *Working Paper No. 40, Department of Economics, University of Zurich.*

Friedman, M. (1955): "The role of government in education," *Rutgers University Press.*

Gallipoli, G., C. Meghir, and G. Violante (2011): "Equilibrium effects of education policies: A quantitative evaluation," *Working Paper.*

Gary-Bobo, R. and A. Trannoy (2013): "Optimal Student Loans and Graduate Tax under Moral Hazard and Adverse Selection," *Working Paper.*

Gelber, A. and M. Weinzierl (2012): "Equalizing Outcomes vs. Equalizing Opportunities: Optimal Taxation when Children's Abilities Depend on Parents' Resources," *NBER Working Paper 18332.*

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. 17642.*

Golosov, M. and A. Tsyvinski (2006): "Designing optimal disability insurance: A case for asset testing," *Journal of Political Economy*, 114, 257–279.

Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988): "Investment, capacity utilization, and the real business cycle," *The American Economic Review*, 402–417.

Grochulski, B. and T. Piskorski (2010): "Risky Human Capital and Deferred Capital Income Taxation," *Journal of Economic Theory*, 145, 908–943.

Hendricks, L. and T. Schoellman (2012): "Student Abilities During the Expansion of US Education," *Working Paper.*

Huggett, M., G. Ventura, and A. Yaron (2011): "Sources of Lifetime Inequality," *American Economic Review*, 101(7), 2923–2954.

Jacobs, B. and A. L. Bovenberg (2011): "Optimal Taxation of Human Capital and the Earnings Function," *Journal of Public Economic Theory*, 13, 957–971.

Jacobs, B., D. Schindler, and H. Yang (2012): "Optimal Taxation of Risky Human Capital," *Scandinavian Journal of Economics*, 114, 908–931.

Kapicka, M. (2006): "Optimal Income Taxation with Human Capital Accumulation and Limited Record Keeping," *Review of Economic Dynamics*, 9, 612–639.

——— (2013): "Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach," *The Review of Economic Studies (forthcoming).*

Kapicka, M. and J. Neira (2013): "Optimal taxation in a life-cycle economy with endogenous human capital formation," *Working paper, UC Santa Barbara.*

KEANE, M. AND K. WOLPIN (1997): "The career decisions of young men," *Journal of Political Economy*, 105, 473–522.

KRUEGER, D. AND A. LUDWIG (2013): "Optimal Progressive Labor Income Taxation and Education Subsidies when Education Decisions and Intergenerational Transfers are Endogenous," *American Economic Review Papers & Proceedings*, 103, 496–501.

LEMIEUX, T. (2006): "Postsecondary Education and Increasing Wage Inequality," *American Economic Review*, 96, 195–199.

LUTTMER, E. AND R. ZECKHAUSER (2008): "Schedule selection by agents: from price plans to tax tables," *NBER Working Paper No. 13808.*

MANKIW, N. AND M. WEINZIERL (2010): "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution," *American Economic Journal: Economic Policy*, 2, 155–176.

MANKIW, N., M. WEINZIERL, AND D. YAGAN (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23, 147–174.

MIRRLEES, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38, 175–208.

PAVAN, A., I. SEGAL, AND J. TOIKKA (2012): "Dynamic Mechanism Design," *Working Paper.*

PIKETTY, T. AND E. SAEZ (2013): "Optimal labor income taxation," *Handbook of Public Economics, Vol 5.*

RAMSAY, C. M. (2006): "The distribution of sums of certain iid Pareto variates," *Communications in Statistics- Theory and Methods*, 35, 395–405.

ROGERSON, W. P. (1985): "Repeated Moral Hazard," *Econometrica*, 53, 69–76.

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205–229.

SALANIÉ, B. (2003): *The Economics of Taxation*, The MIT Press.

SCHEUER, F. (2012): "Entrepreneurial Taxation with Endogenous Entry," *Working Paper, Stanford University.*

STANTCHEVA, S. (2012): "Optimal Taxation and Human Capital Policies Over the Lifecycle," *Graduate Student Research Paper 12-02 MIT.*

STORESLETTEN, K., C. TELMER, AND A. YARON (2004): "Consumption and risk sharing over the life cycle," *Journal of Monetary Economics*, 51, 609–633.

TABER, C. (2001): "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?" *Review of Economic Studies*, 68, 665–691.

WEINZIERL, M. (2012): "Why Do We Redistribute So Much But Tag So Little? Normative Diversity and Optimal Taxation," *Working Paper, Harvard Business School.*

WERNING, I. (2007): "Optimal Fiscal Policy with Redistribution," *The Quarterly Journal of Economics*, 122, 925–967.

——— (2011): "Nonlinear Capital Taxation," *Working Paper, MIT.*

Chapter 2

Optimal Participation Taxes and Efficient Transfer Phase-Out

2.1 Introduction

Redistribution schemes that support the unemployed and individuals with low income exist in all developed countries. There is, however, a public debate on the appropriate design of such schemes. One issue in this debate is whether it is the unemployed or individuals with low income who should receive the largest benefits. Under a Negative Income Tax (NIT), transfers are highest for the unemployed. Individuals with positive income receive lower transfers and thus pay a participation tax when entering the labor market. This contrasts to an Earned Income Tax Credit (EITC): Here individuals with low income receive the highest transfers. Because these transfers exceed those for the unemployed, low income individuals receive a participation subsidy (a negative participation tax) for entering the labor market.

A second issue in this debate concerns the marginal tax rates for those income levels where transfers are phased out. In most real world tax-transfer systems – regardless of whether NIT or EITC – these phase-out rates are very high. On the one hand, one may argue that this is unavoidable if society wants to grant large transfers. On the other hand, high marginal tax rates heavily distort labor supply. As these phase-out rates are close to 100% in many countries, one may suspect that they leave room for Laffer reforms, i.e. tax cuts that are self-financing because of strong labor supply effects.

Using methods of optimal nonlinear taxation, we address these two issues and ask (i) whether a tax-transfer system should levy participation taxes on or provide participation subsidies for individuals with low income and (ii) under what conditions a tax-transfer system is beyond the top of the Laffer curve, so that there is room for tax cuts which increase tax revenue. We derive the following main results: (i) We generalize a well-known theoretical result from the pure extensive margin model going back to Diamond (1980) to a framework with intensive *and* extensive labor supply responses: participation subsidies are never part of the optimal tax schedule if the social marginal utility of the lowest income workers is smaller than the marginal value of public funds.[1] (ii) We develop a test – based only on intensive and extensive labor supply elasticities and the income distribution – that can uncover whether a nonlinear tax schedule is beyond the top of the Laffer curve. Applying this test to Germany, we find that the marginal tax rates in the phase-out region may or may not be inefficiently high (depending on labor supply elasticities), but that they certainly exhibit an inefficient structure so that there is room for Pareto improving reforms.

[1] Throughout this chapter we will use the term tax schedule to describe the effective schedule of the entire tax-transfer system incorporating income taxes and all benefit programs.

As a formal starting point, we solve the optimal nonlinear income tax problem in a model with both intensive *and* extensive labor supply responses. As pointed out by numerous empirical studies, extensive responses are large, in particular for individuals with low income.[2] Addressing them when analyzing the optimal design of income transfer programs is therefore crucial.

One reason for an extensive margin to exist is a minimum hours constraint, as first proposed by Moffitt (1982); see also Stewart and Swaffield (1997), Gustman and Steinmeier (2008) and Gielen (2009) for empirical evidence. Such a constraint can be due to several causes: For example, some tasks require the worker to be present for a certain amount of time, or there may be fixed costs on the side of the firm (e.g., for training or for providing equipment) on which the firm wants to economize.

We incorporate such a minimum hours constraint in a model without income effects where individuals differ in two dimensions, productivity and preferences for leisure. To keep this two-dimensional screening problem tractable, we focus on a special kind of separable preferences which allows to apply a type-aggregator.[3]

Our first contribution is of methodological nature in that we solve this two-dimensional screening problem which is non-standard because the report set of the agents is constrained endogenously. We then show how to express the optimality conditions in reduced form. The reduced form solution for the marginal tax rates shows a tight connection to Saez (2002) and Jacquet et al. (2012). Saez (2002) considers a model, where each individual can choose among two different occupations and unemployment. Jacquet et al. (2012) analyze a Mirrlees model, in which the extensive margin arises due to disutility of participation.

Based on our reduced form solution, we contribute to the EITC versus NIT debate by deriving a formula for the optimal participation taxes in the presence of intensive and extensive labor supply responses. This formula allows to state the condition that participation subsidies are never part of the optimal tax-transfer system if the social marginal utility of the lowest income workers is smaller than the marginal value of public funds. This is a generalization of the result from the pure extensive model, see Diamond (1980), Saez (2002) and Choné and Laroque (2011b). Importantly, this result does not depend on our specific setting with a minimum hours constraint, but holds in general, i.e. also for other frameworks with intensive and extensive labor supply responses.

[2]See Heckman (1993), Eissa and Liebman (1996), Meyer and Rosenbaum (2001) and Meghir and Phillips (2010).

[3]The concept of a type-aggregator has also been applied by Brett and Weymark (2003) and Choné and Laroque (2011a).

Concerning the issue of high marginal tax rates in the phase-out region, we ask under what conditions a given nonlinear tax schedule is beyond the top of the Laffer curve. Whereas the concept of the Laffer curve is well understood for a linear tax, no one explicitly derived the conditions for a nonlinear income tax schedule to be efficient prior to Laroque (2005) (extensive margin) and Werning (2007) (intensive margin). For a setting with both intensive and extensive labor supply responses, we first propose a simple test whether the marginal tax rate at a certain income level is above its Laffer value, i.e. whether a decrease of the marginal tax rate at that income level would increase tax revenue. We then show that a tax schedule may be inefficient for more subtle reasons: even if each marginal tax rate itself is below its Laffer value, the structure of marginal tax rates may be inefficient. We therefore develop a stronger version of the test that can also identify such inefficient structures of marginal tax rates. We express both versions in reduced form, so that they only require knowledge of the income distribution and elasticities. Thus, no assumptions concerning the underlying reason for extensive labor supply responses are necessary when applying the test to a certain tax-transfer system.

Finally, we apply this test to the tax-transfer system in Germany (for singles). Whether marginal tax rates in the phase-out region are beyond their Laffer values crucially depends on the values of the intensive elasticities. However, with the stronger version of the test, we identify an inefficiency irrespective of the values of the elasticities. This inefficiency is caused by the structure of the marginal tax rates which heavily decrease at the income level where the transfer is (just) phased out.[4] Our analysis suggests that a reform that decreases the marginal tax rate below this threshold income level, and increases it above, constitutes a Pareto improvement: As the absolute level of taxes does not increase for any income level, no individual is made worse off, but tax revenue increases due to the induced labor supply responses along the intensive and extensive margin.

Besides the above mentioned papers, this chapter is also related to the following studies. Kleven and Kreiner (2006) find that the marginal cost of public funds is higher if in addition to an intensive margin also extensive labor supply responses are taken into account. Boone and Bovenberg (2004) also consider the optimal tax problem in the presence of both margins: individuals have to search for a job and can either be unemployed voluntarily or involuntarily. They elaborate how the government should optimally balance distortions on search incentives with those on work effort incentives.

[4] The inefficiency does not hinge on the discontinuity in the marginal tax rates but exists as well if marginal tax rates decrease smoothly.

Kanbur et al. (1994) consider the optimal design of tax-transfer systems for the case that the government's goal is the alleviation of poverty. They show that if it is optimal that all individuals work, the marginal tax rate for the lowest income is negative. Pirttilä and Tuomala (2004) extended this result for the case that the government can also levy linear commodity taxes.

Our Pareto efficiency test is related to the test derived by Scheuer (2012), who considers differential taxation of entrepreneurs and workers. In his model the extensive margin captures the decision of being a worker or entrepreneur. Da Costa and Pereira (2011) derive the properties of tax schedules that satisfy a minimum equal sacrifice rule and use the Pareto efficiency test of Werning (2007) to test whether these schedules are Pareto efficient.

The application of our Pareto efficiency test to Germany is related to a study by Blundell et al. (2009). For Great Britain and Germany, they calculate the welfare weights that would render the given tax-transfer systems for lone mothers optimal. In a similar vein, Bargain et al. (2011) pursue this approach for 17 EU countries and the US focusing on singles without children.[5] Both of these studies estimate the relevant elasticities and – for a discretized income distribution with a small number of intervals – then invert the optimal tax formula to calculate the respective welfare weights. We instead apply a more continuous approach and thereby examine the structure of marginal tax rates in greater detail. This makes our approach more powerful in identifying inefficiencies.

The remainder of this chapter is organized as follows: In Section 2.2 we present our model of labor supply (Section 2.2.1) and the government's problem (Section 2.2.2). We reformulate the government's problem as a direct mechanism in Section 2.2.3. We derive the solution and express it in reduced form in Section 2.2.4. We discuss optimal participation taxes in Section 2.3. In Section 2.4 we develop the Pareto efficiency test (Section 2.4.1) and apply it to the German tax-transfer system (Section 2.4.2). Section 2.5 concludes.

[5] Relatedly, Immervoll et al. (2007) consider two kinds of marginal reforms for several European countries: increasing the welfare benefit and increasing in-work benefits. They find that the latter would increase welfare in many EU countries for a large set of redistributive preferences whereas the former can only be justified by very strong redistributive preferences.

2.2 The Model

2.2.1 Individuals' Labor Supply

Individuals' preferences over consumption C and hours of work L are characterized by the quasi-linear utility function

$$U(C, L; \alpha) = C - v(\alpha L),$$

with $v(0) = 0$, $v' \geq 0$, $v'' > 0$. Individuals differ in their productivity w and in α, which measures preferences for leisure. α is assumed to enter the utility function in this way to render the two-dimensional screening problem tractable. w and α are distributed within $[w_0, w_1]$ and $[\alpha_0, \alpha_1]$ according to a joint density function $d(w, \alpha)$, which we represent by the marginal density $f(w)$ and the conditional density $g(\alpha|w)$: $d(w, \alpha) = f(w)\, g(\alpha|w)$. The mass of individuals is normalized to one.

Individuals have to pay (possibly negative) taxes T, and because α, w and L are unobservable for the government, these taxes only depend on income $Y = wL$. All individuals with the same income therefore receive the same consumption level $C = Y - T(Y)$.

Without the minimum hours constraint, income and utility depend on w and α only via the one-dimensional aggregate $\beta = \frac{w}{\alpha}$, which can easily be seen by expressing preferences in terms of C and Y:[6]

$$U = C - v(\alpha L) = C - v\left(\frac{\alpha Y}{w}\right) = C - v\left(\frac{Y}{\beta}\right) = U(C, Y; \beta).$$

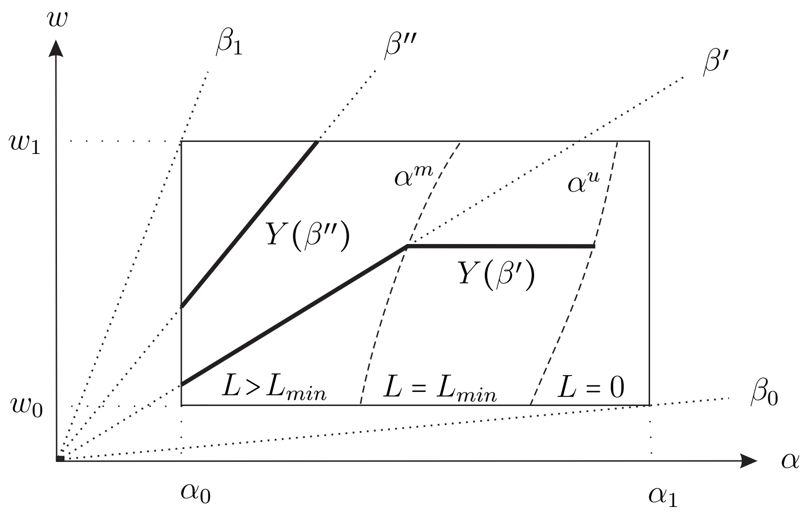
Note that the smallest and largest value of β are $\beta_0 = w_0/\alpha_1$ and $\beta_1 = w_1/\alpha_0$ respectively (see Figure 2.2.1). Let $K(\beta)$ be the distribution function of β with corresponding density $k(\beta)$ that has support $[\beta_0, \beta_1]$. Also, let the conditional density of α in terms of β be $\widetilde{g}(\alpha|\beta)$, with corresponding distribution function $\widetilde{G}(\alpha|\beta)$.

Along each β-line individuals are identical concerning income and consumption, but – in moving away from the origin – the number of hours an individual works decreases: Because of their higher preference for leisure α, these individuals work less, but because of their higher productivity, they earn the same income and receive the same utility.

With the minimum hours constraint $L \geq L_{min}$, this no longer holds. As labor supply decreases along each β-line, it will at some point equal L_{min}, and would then fall below it. This, however, is not possible, so these individuals have to work the

[6] Choné and Laroque (2011a) consider a similar model, but allow for a more general aggregation function than $\beta = w/\alpha$. Also, Brett and Weymark (2003) use such a type-aggregator in a model with endogenous education.

Figure 2.2.1: Partition of the Type-Space by α^m and α^u and Iso-Income Curves



minimum number of hours. We denote the critical values of α that separate those that are not restricted by the minimum hours constraint ($L > L_{min}$) from those that are ($L = L_{min}$) by $\alpha^m(w)$; we provide a formal definition of this α^m-curve below. Since all individuals to the right of α^m work L_{min}, income in this area is constant along a horizontal line, see Figure 2.2.1. Along this horizontal line, α is increasing and for sufficiently large values of it, individuals prefer not to work at all. We denote this second threshold by $\alpha^u(w)$; again, we provide a formal definition below. In the following we present the model and the derivation of the results for the case depicted in Figure 2.2.1, i.e. that both curves are interior and do not cross. This assumption is not necessary to derive the results but greatly simplifies the notation. Also, we could let the minimum hours constraint depend on w, i.e. define a function $L_{min}(w)$. Again, to simplify the notation we refrain from doing so.

For income levels like $Y(\beta')$, the iso-income curve is a kinked line. The group of individuals earning this income consists of two subgroups: Those on the increasing part of the curve can adjust their labor supply freely and those on the horizontal part cannot. This is important for the average elasticity of income with respect to a change in the marginal tax rate: for each income level, this elasticity will depend on the share of these two subgroups.

Finally, let the density of Y be $h(Y)$, with corresponding distribution function $H(Y)$; the formal definition is provided in Section 2.2.4.

2.2.2 The Government's Problem

The government's aim is to choose the nonlinear tax schedule $T(Y)$ that maximizes social welfare

$$W = \int_{w_0}^{w_1} \int_{\alpha_0}^{\alpha_1} \Psi(V(w,\alpha))dG(\alpha|w)dF(w), \tag{2.1}$$

subject to a budget constraint (where R denotes exogenous government expenditure)

$$\int_{\underline{w}}^{\overline{w}} \int_{\underline{\alpha}}^{\overline{\alpha}} T(Y(w,\alpha))dG(\alpha|w)dF(w) = R$$

and

$$V(w,\alpha) = \max\ C - v(\alpha L) \quad \text{s.t.} \quad C \leq wL - T(wL) \quad \text{and} \quad L \geq L_{min} \vee L = 0. \tag{2.2}$$

We denote the welfare benefit by $b = -T(0)$. Participation taxes then are

$$T_{part}(Y) = T(Y) - T(0) = T(Y) + b.$$

Note that the government may find it optimal to have a discontinuity in the tax schedule at the bottom, i.e. $T(Y_{min}) \neq T(0)$, so that $T_{part}(Y_{min}) \neq 0$. This can well be the case even for $Y_{min} \to 0$.

$\Psi(\cdot)$ is increasing and concave and may either represent redistributive preferences of the government or a concave transformation of individual utilities that does not change preferences over leisure and consumption. With $\Psi(V(w,\alpha))$, all individuals with the same utility have the same impact on welfare. This implies that (e.g. along the increasing part of the iso-income curve) the utility of a "lazy and able person" is valued the same as that of an "unable and hardworking" individual. One could, however, easily generalize all our results to the case $\Psi(V(w,\alpha),\alpha)$.

2.2.3 The Government's Problem as a Direct Mechanism

As is well known, a way to make the problem of optimally choosing a nonlinear tax function tractable is to formulate it as a direct mechanism, where the government chooses the optimal income-consumption bundle $(C(w,\alpha), Y(w,\alpha))$ for each type (w,α).[7] The government then maximizes social welfare (2.1) subject to the resource

[7] The reader not interested in the detailed derivation of the representation as a mechanism may skip this section and move immediately to Section 2.2.4, where we present the solution to the government's problem in terms of the tax schedule.

constraint

$$\int_{w_0}^{w_1}\int_{\alpha_0}^{\alpha_1} C(w,\alpha)dG(\alpha|w)dF(w)+R \leq \int_{w_0}^{w_1}\int_{\alpha_0}^{\alpha_1} Y(w,\alpha)dG(\alpha|w)dF(w), \quad \text{(RC)}$$

the minimum hours constraint

$$Y(w,\alpha) \geq wL_{min} \lor Y(w,\alpha) = 0, \quad \text{(MHC)}$$

and the incentive compatibility constraints[8]

$$C(w,\alpha) - v\left(\frac{\alpha Y(w,\alpha)}{w}\right) \geq C(w',\alpha') - v\left(\frac{\alpha Y(w',\alpha')}{w}\right) \quad \text{(IC)}$$

$$\forall \ w,\alpha,w',\alpha' \text{ with } \frac{Y(w',\alpha')}{w} \geq L_{min} \lor Y(w',\alpha') = 0.$$

Denote (by some abuse of notation) the associated indirect utility function also by $V(w,\alpha)$; it is the "direct-mechanism equivalent" to (2.2) and defined by:

$$V(w,\alpha) = \max_{w',\alpha'} C(w',\alpha') - v\left(\frac{\alpha Y(w',\alpha')}{w}\right) \text{ s.t. } \frac{Y(w',\alpha')}{w} \geq L_{min} \lor Y(w',\alpha') = 0.$$

This problem is not straightforward to solve. For example, the set of incentive compatibility constraints is diminished by the fact that for most individuals it is impossible to mimic some of the other types because the income-consumption bundle designated to these other types would require the individual to work less than L_{min}. In other words, the reporting set of the agents is constrained endogenously. In the following we show how to rewrite this problem in a tractable manner. For this purpose, we state three lemmas.

Lemma 2.2.1 *In any incentive compatible allocation, $Y(w,\alpha)$ must be non-increasing in α.*

Proof: If for $\tilde{\alpha} > \alpha$ we have $Y(w,\tilde{\alpha}) > Y(w,\alpha)$, we must have $C(w,\tilde{\alpha}) > C(w,\alpha)$. But if, of two bundles, $(C(w,\tilde{\alpha}), Y(w,\tilde{\alpha})) \gg (C(w,\alpha), Y(w,\alpha))$, type $(w,\tilde{\alpha})$ prefers the first one, so must type (w,α) because of the lower disutility of labor. ■

Based on this lemma, we can define the threshold functions α^u and α^m; based on α^m we can then define the value of β that is associated with the lowest income $Y_{min} = w_0 L_{min}$:

[8] One should also consider a non-negativity constraint on consumption. In the following we assume that $\Psi' \to \infty$ as $V \to 0$, which ensures that negative consumption will never be optimal.

Definition 2.2.1 *For any incentive compatible allocation, define the threshold function* $\alpha^u(w)$ *such that* $\frac{Y(w,\alpha)}{w} \geq L_{min}$ *for* $\alpha \leq \alpha^u(w)$ *and* $\frac{Y(w,\alpha)}{w} = 0$ *for* $\alpha > \alpha^u(w)$.

Definition 2.2.2 *For any incentive compatible allocation, define the threshold function* $\alpha^m(w)$ *such that* $\frac{Y(w,\alpha)}{w} > L_{min}$ *for* $\alpha < \alpha^m(w)$ *and* $\frac{Y(w,\alpha)}{w} \leq L_{min}$ *for* $\alpha \geq \alpha^m(w)$. *This threshold can also be expressed in terms of* β, *which will sometimes simplify the notation. In this case we denote it by* $\alpha^m_\beta(\beta)$; *formally it is given by* $\alpha^m(w) = \alpha^m_\beta(\frac{w}{\alpha^m(w)})$.

Definition 2.2.3 *For any incentive compatible allocation, define* $\underline{\beta} = \frac{w_0}{\alpha^m(w_0)}$, *or implicitly by* $Y(\underline{\beta}) = w_0 L_{min}$.

In Figure 2.2.1, $\underline{\beta}$ would correspond to the β-line through the intersection of the α^m-curve and the lower border w_0 of the type space. Note that $\underline{\beta}$ constitutes the lower bound for unconstrained workers, i.e. for $\beta > \underline{\beta}$ there is a positive mass of individuals working more than L_{min}.
The next lemma is based on the type aggregator β and allows to define income, consumption and utility for all workers that work more than L_{min} solely as a function of β.

Lemma 2.2.2 *If it is optimal for type* (w,α) *to choose income-consumption bundle* $(\widetilde{C},\widetilde{Y})$ *with* $\frac{\widetilde{Y}}{w} > L_{min}$, *then it is also optimal for any other type* (w',α') *with* $\frac{w'}{\alpha'} = \frac{w}{\alpha}$ *and* $\frac{\widetilde{Y}}{w'} \geq L_{min}$.

The next lemma simply follows from the fact that the government can only observe income. It will be used to link workers who work L_{min} with those who work more than L_{min} but earn the same income.

Lemma 2.2.3 *In any incentive compatible allocation, whenever* $Y(w,\alpha) = Y(w',\alpha')$ *for some types* (w,α) *and* (w',α'), *then* $C(w,\alpha) = C(w',\alpha')$.

Based on these three lemmas, we show in Appendix 2.A.2 that the government's problem can be rewritten in the following way, which can then be solved using standard Lagrangian techniques:

Proposition 2.2.4 *Instead of choosing* $\{C(w,\alpha), Y(w,\alpha)\}$ *in order to maximize (2.1) subject to (MHC), (RC) and (IC), the planner can also choose* $\{C(\beta), Y(\beta)\}$ *for all unconstrained workers, consumption levels for the constrained workers* $\{C(w,\alpha)\}$ *and consumption levels* b *for all inactive workers subject to* *(i) a no discrimination constraint*

$$C(\beta) = C\left(\frac{Y(\beta)}{L_{min}}, \alpha\right) \quad \forall\, \beta \text{ and } \alpha \in \left[\alpha^m\left(\frac{Y(\beta)}{L_{min}}\right), \alpha^u\left(\frac{Y(\beta)}{L_{min}}\right)\right], \qquad (NDC)$$

(ii) an envelope condition

$$V'(\beta) = v'\left(\frac{Y(\beta)}{\beta}\right)\frac{Y(\beta)}{\beta^2} \quad \forall\, \beta \geq \underline{\beta}, \qquad (EC)$$

(iii) a monotonicity constraint

$$Y'(\beta) \geq 0 \;\forall\; \beta \geq \underline{\beta}, \qquad (MC)$$

(iv) and the government budget constraint

$$\int_{w_0}^{w_1}\int_{\alpha^u(w)}^{\alpha_1} b\, dG(\alpha|w)dF(w) + R = \int_{\underline{\beta}}^{\beta_1} (Y(\beta) - C(\beta))\, \widetilde{G}(\alpha^m_\beta(\beta)|\beta)dK(\beta) \quad (GBC)$$

$$+ \int_{w_0}^{w_1}\int_{\alpha^m(w)}^{\alpha^u(w)} (wL_{min} - C(w,\alpha))\, dG(\alpha|w)dF(w),$$

where $\underline{\beta} = \frac{w_0}{\alpha^m(w_0)}$ *and the threshold functions* $\alpha^u(w)$ *and* $\alpha^m_\beta(\beta)$ *satisfy*

$$C(w,\alpha) - v(\alpha^u(w), L_{min}) = b \quad \forall\, w$$

and

$$Y(\beta) = \beta\alpha^m_\beta(\beta)L_{min} \;\forall\; \beta.$$

Proof: See Appendix 2.A.2 ■

The intuition for the restatement of the government's problem is as follows: For all individuals to the left of the α^m-line, incentive constraints can be replaced by an envelope condition and a monotonicity constraint as in a standard Mirrlees problem. Any further incentive compatibility is guaranteed by the definition of the thresholds and the no discrimination constraint, which formalizes the fact that individuals with the same income must be assigned the same consumption as well.

2.2.4 Solution to the Government's Problem

In Appendix 2.A.3, we state the first order conditions of the government's problem the way it was expressed as a direct mechanism in Proposition 2.2.4. Using these first order conditions, in Appendix 2.A.4, we then derive the solution in terms of the optimal tax schedule which can be summarized as follows:

$$\frac{T'(Y(\beta))}{1-T'(Y(\beta))}\lambda\beta\frac{\varepsilon_{Y,1-T'}}{\varepsilon_{Y,1-T'}+1}\widetilde{G}(\alpha^m(\beta)|\beta)k(\beta)-\mathcal{A}(Y(\beta))=0\ \ \forall\beta, \tag{2.3}$$

where $\varepsilon_{Y,1-T'}$ denotes the elasticity of the unconstrained workers, and with

$$\begin{aligned}\mathcal{A}(Y(\beta)) \ &= \ \int_{\beta}^{\beta_1}\int_{\underline{\alpha}(\beta')}^{\alpha^m(\beta')}\left(\lambda-\Psi'(V(\beta'))\right)d\widetilde{G}(\alpha|\beta')dK(\beta') \\ &+\int_{\frac{Y(\beta)}{L_{min}}}^{w_1}\left[\int_{\alpha^m(w)}^{\alpha^u(w)}\left(\lambda-\Psi'(V(w,\alpha))\right)dG(\alpha|w)\right. \\ &\left.+\lambda g(\alpha^u(w)|w)\frac{\partial\alpha^u(w)}{\partial T(wL_{min})}(T(wL_{min})+b)\right]dF(w).\end{aligned} \tag{2.4}$$

The Lagrangian multiplier λ, associated with the government's budget constraint (GBC), is equal to the average social marginal utility of income, i.e.

$$\lambda=\int_{w_0}^{w_1}\int_{\alpha_0}^{\alpha_1}\Psi'(V(w,\alpha))dG(\alpha|w)dF(w). \tag{2.5}$$

Further, we have

$$\mathcal{A}(Y(\underline{\beta}))=\mathcal{A}(Y(\beta_1))=0. \tag{2.6}$$

Deriving the conditions for the optimal marginal tax rates and the optimal participation taxes from equations (2.3) and (2.4) is rather cumbersome. Also, we want to express the test for Pareto inefficiency in terms of observable labor supply elasticities and the income distribution. We therefore show how to rewrite the above solution in reduced form. To do so, denote by $\overline{\Psi}'(Y)$ the average social marginal utility of income for all individuals earning Y. The cumulative distribution function of income Y is given by

$$H(Y(\beta))=\int_{\underline{\beta}}^{\beta}\widetilde{G}(\alpha_\beta^m(\beta')|\beta')dK(\beta')+\int_{w_0}^{\frac{Y(\beta)}{L_{min}}}\left[G(\alpha^u(w)|w)-G(\alpha^m(w)|w)\right]dF(w)$$

with corresponding density

$$h(Y(\beta))=\widetilde{G}(\alpha_\beta^m(\beta)|\beta)k(\beta)\frac{d\beta}{dY}+\left(G\left(\alpha^u(w_\beta)|w_\beta\right)-G\left(\alpha^m(w_\beta)|w_\beta\right)\right)\frac{f(w_\beta)}{L_{min}},$$

where $w_\beta = \frac{Y(\beta)}{L_{min}}$. Note that the income distribution is endogenous with respect to the tax-transfer system.

Also, we have to express the labor supply responses in terms of elasticities. Let $\xi(Y)$ be the semi-elasticity of participation, i.e. the increase in the number of unemployed[9] relative to the number of individuals earning income level Y due to an absolute increase in $T(Y)$ (or b); formally it is given by

$$\xi(Y(\beta)) = \frac{-\frac{\partial \alpha^u(w_\beta)}{\partial T} g(\alpha^u(w_\beta)|w_\beta)\frac{f(w_\beta)}{L_{min}}}{h(Y(\beta))}.$$

Likewise, let $\overline{\varepsilon}(Y)$ be the average elasticity of income with respect to $1 - T'$ of all individuals earning income Y:

$$\overline{\varepsilon}(Y(\beta)) = \frac{\widetilde{G}(\alpha^m(\beta)|\beta)k(\beta)}{h(Y)\frac{\partial Y}{\partial \beta}}\varepsilon_{Y,1-T'}.$$

For each income level Y, it is the share of individuals with $L > L_{min}$ multiplied by their intensive margin elasticity $\varepsilon_{Y,1-T'}$.

Finally, denote the maximum income by $Y_{max} = Y(\beta_1)$ and recall that $Y_{min} = w_0 L_{min}$. Based on these definitions, we state the following proposition:

Proposition 2.2.5 *The optimality conditions (2.3) - (2.6) can be expressed in reduced form as*

$$\frac{T'(Y)}{1 - T'(Y)}\lambda Y \overline{\varepsilon}(Y)h(Y) - \mathcal{A}(Y) = 0 \;\; \forall\, Y \tag{2.7}$$

with

$$\mathcal{A}(Y) = \int_Y^{Y_{\max}} \left[(\lambda - \overline{\Psi}'(\widetilde{Y})) - \lambda\xi(\widetilde{Y})T_{part}(\widetilde{Y})\right] dH(\widetilde{Y}), \tag{2.8}$$

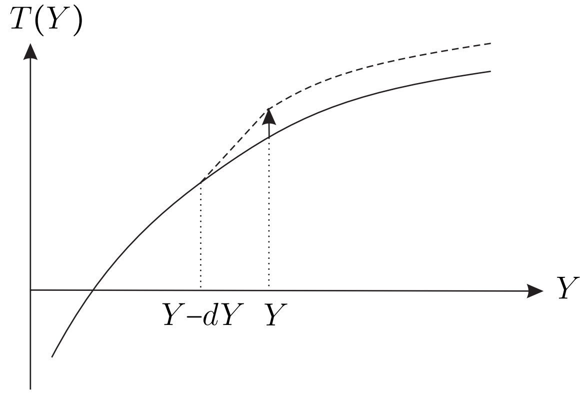
$$\lambda = \overline{\Psi}'(0)H(0) + \int_{Y_{min}}^{Y_{max}} \overline{\Psi}'(Y)dH(Y), \tag{2.9}$$

$$\mathcal{A}(Y_{min}) = \mathcal{A}(Y_{max}) = 0. \tag{2.10}$$

Proof: See Appendix 2.A.5. ■

[9] As in our model there is only voluntary unemployed, the notion “unemployed” always refers to inactive individuals.

Figure 2.2.2: Tax Perturbation



Simply rewriting condition (2.7) then yields the expression for the optimal marginal tax rates:

$$\frac{T'(Y)}{1-T'(Y)} = \frac{\int_{Y}^{Y_{\max}} \left[(\lambda - \overline{\Psi}'(\widetilde{Y})) - \lambda \xi(\widetilde{Y}) T_{part}(\widetilde{Y}) \right] dH(\widetilde{Y})}{\lambda Y \overline{\varepsilon}(Y) h(Y)}. \tag{2.11}$$

This condition resembles the results of Saez (2002) and Jacquet et al. (2012). Once our results for the optimal marginal tax rates are rewritten in reduced form, they are equivalent to theirs.[10]

Although our theoretical contribution is not on this formula for the optimal marginal tax rates, we now provide an intuitive derivation of (2.11) using a tax perturbation method.[11] This derivation will sharpen the economic intuition about the trade-offs the government faces and will facilitate the understanding of our results for the optimal participation taxes and the test for Pareto-efficiency.

Consider an infinitesimal increase dT' of the marginal tax rate in an income interval of infinitesimal length dY around income Y, see Figure 2.2.2. This will have the following three effects on welfare:

Substitution Effect: Individuals within the interval adjust their labor supply along the intensive margin. By the envelope theorem, these labor supply responses only change welfare by their impact on public funds. The mass of individuals affected is $h(Y)dY$. The average increase in income is given by $\frac{\partial Y}{\partial T'}dT' = -\frac{Y\overline{\varepsilon}(Y)}{1-T'(Y)}dT'$, which multiplied by $T'(Y)$ yields the effect on public funds. The substitution effect in

[10] A result that has been obtained by Jacquet et al. (2012) is that marginal tax rates are positive for interior income levels if $\overline{\Psi}'(Y)$ is decreasing in income and $\frac{\partial}{\partial Y}\left[\frac{\lambda-\overline{\Psi}'(Y)}{\xi(Y)}\right] > 0$. For completeness, we give a simple proof of this result in our reduced form notation in Appendix 2.A.6.

[11] See Piketty (1997) and Saez (2001).

terms of welfare then is

$$dW^S(Y) = -\lambda \frac{T'(Y)}{1 - T'(Y)} Y \overline{\varepsilon}(Y) h(Y) dT' dY. \tag{2.12}$$

Redistribution Effect: The increase of the marginal tax rate results in a higher overall tax of $dT'dY$ for all individuals earning more than Y and thereby redistributes money from these individuals (valued by $\overline{\Psi}'$) to the government (valued by λ). This redistribution effect on welfare therefore reads as

$$dW^R(Y) = dT'dY \int_Y^{Y_{max}} (\lambda - \overline{\Psi}'(\widetilde{Y})) dH(\widetilde{Y}). \tag{2.13}$$

Participation Effect: Some of the individuals earning more than Y will stop working due to the higher participation tax. For each income level $\widetilde{Y} \geq Y$, their mass is captured by $\frac{\partial h(\widetilde{Y})}{\partial T_{part}} dT'dY = \xi(\widetilde{Y}) h(\widetilde{Y}) dT'dY$. By choosing unemployment over employment, the governments tax revenue is decreased by the participation tax of these individuals $T_{part}(\widetilde{Y})$. As these individuals are indifferent between working and not working, these labor supply responses only change welfare by their impact on public funds. This effect on welfare equals

$$dW^P(Y) = -dT'dY \int_Y^{Y_{max}} \lambda \xi(\widetilde{Y}) T_{part}(\widetilde{Y}) dH(\widetilde{Y}). \tag{2.14}$$

For the tax function to be optimal, we have to have

$$dW^S(Y) + dW^R(Y) + dW^P(Y) = 0 \ \ \forall \ Y,$$

yielding (2.7) and (2.11). Condition (2.11) allows an intuitive interpretation of the economic forces that determine the value of the optimal marginal tax rates: The denominator on the RHS captures the substitution effect. The higher the mass of individuals $h(Y)$ whose marginal incentives are distorted and the higher their average elasticity $\overline{\varepsilon}(Y)$ and their productivity, the larger the excess burden and therefore the lower the marginal tax rate.

The first term in the numerator represents the redistribution effect as defined in (2.13). The greater the aggregated difference between λ and $\overline{\Psi}'$, the larger the welfare gain from additional redistribution and the higher marginal tax rates should be.

The participation effect as defined in (2.14) is captured by the second term in the numerator. It is increasing in the mass of individuals responding along the extensive margin $\xi(Y)h(Y)$ and the participation tax. It counteracts the redistribution effect and leads to lower marginal tax rates.

We finally comment on the optimality conditions (2.9) and (2.10). Equation (2.9) states the well known result that the average social marginal utility of income is equal to the marginal value of public funds if there are no income effects. A marginal increase in income for all individuals increases welfare by the aggregate social marginal utility of income, and decreases government revenue by one (as the mass of individuals is one), which is valued by λ. Without income effects, such a reform has no impact on labor supply and therefore, for the tax schedule to be optimal, the redistribution effect on welfare must be zero, yielding (2.9).

$\mathcal{A}(Y_{min}) = 0$ could also have been derived by a small perturbation of the tax schedule such that all employed individuals pay marginally higher taxes, leaving $T(0) = -b$ constant. This marginal and identical increase in the participation tax for all employed individuals would only cause a redistribution and a participation effect as defined in (2.13) and (2.14), both integrated over all income levels greater than or equal to Y_{min}. The condition $\mathcal{A}(Y_{min}) = 0$ then follows from $dW^R(Y_{min}) + dW^P(Y_{min}) = 0$. Because this reform increases $T_{part}(Y_{min})$ while $T(0)$ stays constant, this condition implicitly determines the optimal "size" of the discontinuity in the tax schedule. As argued by Choné and Laroque (2011b) and Jacquet et al. (2012), it may well be that the government finds it optimal to have negative participation taxes for low income workers induced by such a discontinuity and at the same time have strictly positive marginal tax rates.

Note that $\mathcal{A}(Y_{min}) = \mathcal{A}(Y_{max}) = 0$ does not necessarily imply that marginal tax rates are zero at the bottom and at the top. According to (2.10), the numerator of condition (2.11) equals zero for Y_{min} and Y_{max}. For these two values of Y we then have to have

$$\frac{T'(Y)}{1 - T'(Y)} \lambda Y \bar{\varepsilon}(Y) h(Y) = 0.$$

For $\bar{\varepsilon}(Y)h(Y) > 0$ we get the standard result of no distortion, but if $\bar{\varepsilon}(Y)h(Y) = 0$ we may have $T'(Y) \neq 0$. This case applies in our minimum hours model for both Y_{min} and Y_{max}: Because the increasing part of the iso-income curve is infinitesimally small for Y_{min}, we have $\bar{\varepsilon}(Y_{min}) = 0$. For Y_{max} we have $h(Y_{max}) = 0$ because the length of the respective iso-income curve is infinitesimally small.

2.3 Optimal Participation Taxes

We now turn to the question of whether the optimal tax-transfer system should levy participation taxes or provide participation subsidies. To do so, we first state the optimality condition for the participation taxes:[12]

Corollary 2.3.1 *For income levels with $\xi(Y) > 0$, optimal participation taxes are given by*

$$T_{part}(Y) = \frac{\left(\lambda - \overline{\Psi}'(Y)\right)h(Y) + \frac{\partial}{\partial Y}\left[\frac{T'(Y)}{1-T'(Y)}h(Y)Y\lambda\overline{\varepsilon}(Y)\right]}{\lambda\xi(Y)h(Y)}. \tag{2.15}$$

Proof: Because equation (2.7) holds for all values of Y, we can take its derivative with respect to Y; rearranging terms then yields (2.15). ■

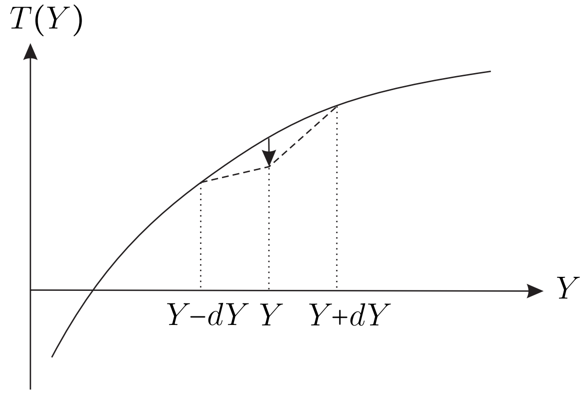
To gain an intuitive understanding of this expression, first consider equation (2.15) without the term $\frac{\partial}{\partial Y}[\cdot]$. We then have the standard interpretation of a model with only extensive labor supply responses going back to Diamond (1980): The sign of the optimal participation tax only depends on the social marginal utility of income compared to the marginal value of public funds: For income levels with $\overline{\Psi}' < \lambda$, participation taxes are positive, for those with $\overline{\Psi}' > \lambda$, they are negative (Diamond, 1980; Saez, 2002; Choné and Laroque, 2011b).[13] This result can most easily be understood by considering an (infinitesimally) small perturbation of the tax schedule as shown in Figure 2.3.1, so that the tax at income Y is reduced by dT due to a small decrease of the marginal tax rates in the interval $[Y - dY, Y]$ and a small increase of the marginal tax rates in the interval $[Y, Y + dY]$.[14] Without intensive labor supply responses, this only has two effects: a redistribution effect of $(\overline{\Psi}'(Y) - \lambda)h(Y)$ as individuals with income Y pay lower taxes, and a participation effect of $T_{part}(Y)\lambda\xi(Y)h(Y)$ as some of the unemployed start working if the participation tax is reduced. For the optimal tax schedule, these two effects on welfare have to add up to zero and therefore the sign of the participation tax is equal to the sign of $\lambda - \overline{\Psi}'$.

[12] In our minimum hours model, the extensive margin may be missing for very high income levels, so that $\xi(Y) = 0$. For these income levels, T_{part} cannot be inferred from (2.15), but is implicitly defined by the set of optimality conditions stated in Proposition 2.2.5. However, condition (2.15), when multiplied by the denominator, holds for all Y.

[13] Christiansen (2012) also discusses the question of negative participation taxes in an extensive margin model and refers to the important role of labor supply responses of high-skilled for this condition to hold. He also generalizes the result to a general equilibrium framework.

[14] Werning (2007) considers such a reform in a model with only intensive labor supply responses to test whether a given income tax schedule is Pareto inefficient.

Figure 2.3.1: Tax Perturbation



With labor supply responses along the intensive margin such a perturbation also has a substitution effect as defined in (2.12) because of the change in marginal tax rates. Individuals with income in $[Y - dY, Y]$ will increase their labor supply, and those with income in $[Y, Y + dY]$ will reduce it. Whether this increases or decreases government revenue depends on the difference of these two effects, which in the limit, as $dT \to 0$, is captured by the derivative of the substitution effect, i.e. $\frac{\partial}{\partial Y}[\cdot]$. This derivative can be smaller than zero if for example the density $h(Y)$ is decreasing quickly. However, for a constant density, a constant elasticity and a constant marginal tax rate, the substitution effect is increasing (so that $\frac{\partial}{\partial Y}[\cdot]$ is positive), which then makes negative participation taxes less likely compared to the pure extensive model. This shows that we can have $\overline{\Psi}' > \lambda$ and still $T_{part} > 0$, so that the result of the pure extensive model does not carry over to a setting with both margins.

As we have shown at the end of Section 2.2.4, the substitution effect at Y_{min} equals zero since either $\overline{\varepsilon}(Y_{min})h(Y_{min}) = 0$ or $T'(Y_{min}) = 0$. This raises the question if at least at the bottom of the income distribution the result of the pure extensive model holds. Note that the term $\frac{\partial}{\partial Y}[\cdot]$ can be decomposed as

$$\frac{T'(Y)}{1 - T'(Y)} \lambda \frac{\partial}{\partial Y}\left[h(Y) Y \overline{\varepsilon}(Y)\right] + h(Y) Y \lambda \overline{\varepsilon}(Y) \frac{\partial}{\partial Y}\left[\frac{T'(Y)}{1 - T'(Y)}\right]. \tag{2.16}$$

For $\overline{\varepsilon}(Y_{min})h(Y_{min}) = 0$, the second term vanishes and the derivative in the first term is unambiguously positive so that the first term as a whole is positive (negative) if $\frac{T'(Y_{min})}{1-T'(Y_{min})} > (<)\, 0$. For $T'(Y_{min}) = 0$, the first term vanishes and the second term is positive (negative) if T' is increasing (decreasing) in Y at Y_{min}. Thus, although the substitution effect equals zero at the bottom of the income distribution, the result from the binary model that the sign of the participation tax depends solely on $\overline{\Psi}'$

relative to λ does not carry over. However, using (2.16) one can show that the reverse holds:

Proposition 2.3.1 *If the social marginal utility at the bottom of the income distribution is smaller than the social marginal value of public funds, i.e.* $\overline{\Psi}'(Y_{min}) < \lambda$, *and* $\overline{\Psi}'(Y)$ *decreases in income, then* $T_{part}(Y)$ *is positive for all* $Y \geq Y_{min}$.

Proof: If $\overline{\Psi}'(Y_{min}) < \lambda$, then $T_{part}(Y_{min})$ can only be negative if (2.16) is negative. Again we have to distinguish two cases: For $\overline{\varepsilon}(Y_{min})h(Y_{min}) = 0$, the second term of (2.16) vanishes, so that (2.16) can only be negative if $T'(Y_{min}) < 0$, because $\frac{\partial}{\partial Y}\left[h(Y)Y\overline{\varepsilon}(Y)\right] \geq 0$ for Y_{min}. For $T'(Y_{min}) = 0$, the first term of (2.16) vanishes, so that (2.16) can only be negative if $\frac{\partial}{\partial Y}\left[\frac{T'(Y_{min})}{1-T'(Y_{min})}\right] < 0$, which implies $T'(Y_{min}+\epsilon) < 0$ for some small ϵ. In both cases, for $T_{part}(Y_{min})$ to be negative, T' has to be negative for Y equal or close to Y_{min}.
However, because $T(0) \leq 0$ by definition (since individuals without income cannot pay taxes), T_{part} has to be positive for some Y so that the government budget constraint is satisfied. This implies that T' has to turn positive for some value of Y, say $\widetilde{Y}$, where T_{part} is still negative. At $\widetilde{Y}$, $T'(\widetilde{Y}) = 0$, $\frac{\partial}{\partial Y}\left[\frac{T'(\widetilde{Y})}{1-T'(\widetilde{Y})}\right] > 0$, and $\overline{\Psi}'(\widetilde{Y}) < \lambda$, so the right hand side of (2.15) is unambiguously positive, a contradiction to $T_{part}(\widetilde{Y})$ still being negative at that point. (For $\xi(\widetilde{Y}) = 0$, the numerator of the right hand side of (2.15) would be positive, while $T_{part}(\widetilde{Y})\lambda\xi(\widetilde{Y})h(\widetilde{Y}) = 0$, again a contradiction.)
If $T_{part}(Y_{min}) \geq 0$ and $T_{part}(\widehat{Y}) < 0$ for some $\widehat{Y} > Y_{min}$, there must be a $\overline{Y} < \widehat{Y}$ such that $T'(\overline{Y}) < 0$ and $T_{part}(\overline{Y}) = 0$. If then T_{part} becomes positive for some $Y > \overline{Y}$, the same reasoning of the previous paragraph applies again. If T_{part} did not become positive, we would have $T_{part}(Y) < 0 \; \forall \; Y > \overline{Y}$. But then the right hand side of (2.11) would be positive for $Y = \overline{Y}$, a contradiction to $T'(\overline{Y}) < 0$. ■

This proposition generalizes a well-known result from the optimal tax model with only participation decisions (Diamond, 1980) to a framework with both intensive and extensive labor supply responses. Whether the condition that the social marginal utility at the bottom of the income distribution is smaller than the social marginal value of public funds is fulfilled depends on the welfare function and, e.g., on the number of inactive workers. The higher this number, the stronger the impact of their social marginal utility on the marginal value of public funds and the more likely this condition is fulfilled. Also, the more concave the social welfare function, the more likely it is fulfilled. In the extreme case of a Rawlsian welfare function, the condition always holds.

2.4 A Test for Pareto Inefficiency

In the previous sections, we analyzed the properties of the optimal tax-transfer system for some given model primitives and a particular social welfare function Ψ. In the following, we will invert the optimal tax approach and derive a simple empirical test to answer the following question: Given the tax-transfer system and what is known about the primitives in this country: Does there exist a Bergson-Samuelson welfare function that would render the current system as optimal?[15] If this is not the case, the observed tax-transer system is (second-best) Pareto inefficient. This in turn implies that reforms of the tax-transfer system can be undertaken that yield Pareto improvements or – equivalently – there is room for self-financing tax cuts. Thus, our test is also a test for a nonlinear tax-transfer system to be beyond the top of the Laffer curve. In Section 2.4.1 we derive the test. We apply this test to Germany in Section 2.4.2.

2.4.1 Theoretical Considerations

2.4.1.1 Inefficiently High Marginal Tax Rates

We first ask whether the marginal tax rate at a certain income level (given the marginal tax rates for the other income levels) is so high that it is beyond its Laffer value. To determine this value, it is helpful to rewrite the optimality condition (2.7) in the following way:

$$\frac{T'(Y)}{1-T'(Y)}\bar{\varepsilon}(Y)h(Y)Y-(1-H(Y))+\int_{Y}^{Y_{max}}\xi(\widetilde{Y})T_{part}(\widetilde{Y})dH(\widetilde{Y}) \tag{2.17}$$

$$=-\frac{1}{\lambda}\int_{Y}^{Y_{max}}\overline{\Psi}'(\widetilde{Y})dH(\widetilde{Y}).$$

The Laffer value, i.e. the revenue maximizing marginal tax rate is found by setting all welfare weights $\overline{\Psi}'$ to zero. It then immediately follows that $T'(Y)$ is too high if the left hand side of (2.17) is greater than zero. This yields a first test for inefficiency, which can be applied if the tax schedule, the income distribution and the labor supply elasticities are known:

Proposition 2.4.1 *For given intensive elasticities $\bar{\varepsilon}(Y)$, extensive semi-elasticities $\xi(Y)$, an income distribution $H(Y)$ and quasi-linear preferences, whenever a tax*

[15] Saez (2001) first suggested this idea. Werning (2007) elaborates this approach for the Mirrlees model with intensive labor supply responses. We extend this approach to the case of intensive and extensive labor supply responses.

schedule satisfies

$$\frac{T'(Y)}{1-T'(Y)}\overline{\varepsilon}(Y)h(Y)Y-(1-H(Y))+\int_{Y}^{Y_{max}}\xi(\widetilde{Y})T_{part}(\widetilde{Y})dH(\widetilde{Y})>0 \quad (2.18)$$

for at least some Y, then the tax schedule is second-best Pareto inefficient.

This proposition can be considered as the natural extension of the Laffer argument to nonlinear taxation: With a linear tax schedule, it is the constant marginal tax rate that is too high over the entire schedule; here, it is the marginal tax rate $T'(Y)$ at a specific income level Y. Lowering $T'(Y)$ will increase tax revenue; it will also reduce T for all income levels Y and above, which will make these individuals better off. A small reduction of the marginal tax rate $T'(Y)$ therefore constitutes a Pareto improvement.

This test will identify some of the inefficient tax schedules, but we will now argue that a stronger test exists: Even if each marginal tax rate itself is below the Laffer value, the tax schedule can be inefficient because the structure of marginal tax rates is not efficient. In this case, a different reform will be needed to achieve a Pareto improvement.

2.4.1.2 Inefficient Structure of Marginal Tax Rates

To derive the stronger version of the test we use the fact that for each Pareto efficient tax schedule, there exists a set of nonnegative welfare weights so that the tax schedule is the solution to the welfare maximization problem for these weights. If one of these weights has to be negative, the tax schedule cannot be efficient. Taking the derivative of condition (2.17) yields an expression for these weights:

$$\frac{\partial}{\partial Y}\left[\frac{T'(Y)}{1-T'(Y)}\overline{\varepsilon}(Y)h(Y)Y\right]+h(Y)-\xi(Y)T_{part}(Y)h(Y)=\frac{\overline{\Psi}'(Y)}{\lambda}h(Y). \quad (2.19)$$

A negative welfare weight $\overline{\Psi}'(Y)$ and thus a Pareto inefficiency exists if the left hand side of (2.19) is negative, i.e., if the left hand side of (2.17) is decreasing in income. This defines the stronger version of the test:[16]

Proposition 2.4.2 *Given intensive elasticities $\overline{\varepsilon}(Y)$, extensive semi-elasticities $\xi(Y)$, an income distribution $H(Y)$ and quasi-linear preferences, a tax schedule $T(Y)$ is*

[16]Note that the condition in this proposition is not necessary for an inefficiency. A tax-transfer system could alternatively also be inefficient at a possible discontinuity at $Y=0$. Such an inefficiency cannot be identified by this test. However, as such discontinuities rather seem to be theoretical idea, we refrain from discussing such inefficient discontinuities in the following.

second-best Pareto inefficient, if

$$\frac{T'(Y)}{1-T'(Y)}\bar{\varepsilon}(Y)h(Y)Y-(1-H(Y))+\int_{Y}^{Y_{max}}\xi(\widetilde{Y})T_{part}(\widetilde{Y})dH(\widetilde{Y}) \tag{2.20}$$

is decreasing in Y for at least one Y.

Again, this test can be applied if the tax schedule, the income distribution and the labor supply elasticities are known. Note that it nests the condition for Pareto inefficiency of Proposition 2.4.1, i.e., whenever a tax schedule is inefficient according to (2.18), it is also inefficient according to (2.20): If the cumulative welfare weights are smaller than zero (so that the right hand side of (2.17) is positive), then at least one of the welfare weights has to be negative. On the other hand, the weighted sum might still be positive although some of the weights are negative.

If the test indicates that a tax schedule is inefficient, then a reform as depicted in Figure 2.3.1 in Section 2.3, conducted at income level Y, will yield a Pareto improvement.[17] Such a reform will be self-financing or even increase tax revenue. Without labor supply responses, this tax cut would decrease tax revenue, but the labor supply responses will outweigh this loss. Using equation (2.19) instead of (2.20) makes it easier to see when that will be the case.

The mechanical loss in tax revenue is given by $h(Y)$, the mass of individuals affected by the tax cut. The participation effect on public funds induced by the tax reform is captured by the third term on the left hand side of (2.19): The larger $T_{part}(Y)$ and the larger the participation semi-elasticity $\xi(Y)$, the larger is this participation effect. The argument for the substitution effect is more subtle as the tax reform on the one hand increases marginal tax rates for incomes slightly higher than Y and on the other hand decreases marginal tax rates for incomes slightly lower than Y. In the limit, the overall sign of these intensive labor supply responses is captured by the derivative of the substitution effect captured by the first term on the left hand side of (2.19). That is, the positive effect on public funds induced by the labor supply increase of those with slightly lower income is more likely to outweigh the other effect if the marginal tax rate, the density $h(Y)$ or the elasticity is decreasing in income.

2.4.1.3 Overcoming Inefficiencies

A favorable property of the proposed tests is that to apply them only the income distribution and elasticities are required. In the terminology of Chetty (2009), the

[17] For the case without extensive labor supply responses, such a reform has already been proposed by Werning (2007).

income distribution and the elasticities are sufficient statistics to uncover inefficiencies. However, for overcoming an inefficiency, one has to know how individuals react to large tax reforms and therefore has to make structural assumptions about their labor supply decisions. Nevertheless our analysis provides theory-based guidance for such reforms. Whenever a tax schedule is characterized by inefficiently high marginal tax rates as discussed in Section 2.4.1.1, we know that a small decrease in these marginal tax rates yields a Pareto improvement. In order to know how strong these decreases have to be to not only yield a Pareto improvement but to completely eliminate the inefficiency, one has to make structural assumptions. Similarly, if the structure of marginal tax rates is inefficient as discussed in Section 2.4.1.2, we know that a small reform as depicted in Figure 2.3.1 yields a Pareto improvement. But again, structural assumptions are required to determine how to eliminate the inefficiency.

Of course there will always exist not only one, but a whole set of Pareto improving reforms. Each of these reforms would yield a different allocation on the Pareto frontier. That is, when deciding how to overcome the inefficiency, one has to abandon the sole "efficiency consideration" and make a choice of how to value the utility of different individuals.

2.4.2 An Application to Germany

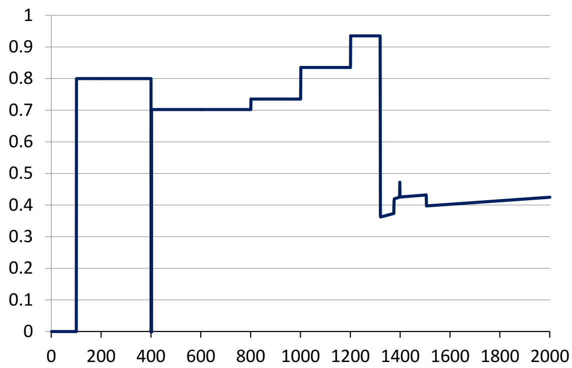
In order to apply the Pareto inefficiency test, the tax-transfer schedule has to be known. For Germany (and likely for other countries as well), it is not immediately apparent what this schedule looks like because it is the result of the interplay of three different systems. We discuss how to construct this schedule and how we estimate the income distribution in the following Section 2.4.2.1. The results are presented in Section 2.4.2.2. Policy implications are discussed in Section 2.4.2.3.

2.4.2.1 Income Distribution, Tax-Transfer System and Elasticities

As in most countries, the tax-transfer system conditions on marital status as well as on the number of children. As the taxation of families raises a number of additional issues, we focus on singles without children. In addition, eligibility for welfare benefits depends on assets. Therefore, we only consider individuals with sufficiently low assets such that eligibility for welfare benefits is ensured.

The tax-transfer system results from the interplay of three different systems: the income tax schedule, the welfare benefit system including the phase-out region and social insurance contributions. The main steps to derive the effective tax schedule are as follows: Gross income determines payments to the social insurance system according to the Social Security Code. Gross income and social insurance contributions then determine the tax liability according to the Personal Income Tax Code. Transfers then depend on gross income, taxes and social insurance contributions. A more detailed derivation and the link for the Excel file which contains the exact calculation of the schedule can be found in Appendix 2.A.1. Integrating the three systems, we arrive at the schedule of effective marginal tax rates for the year 2011 as shown in Figure 2.4.1.[18] Marginal tax rates are very high for low incomes. As soon as transfers are phased out, marginal tax rates decrease drastically.[19]

Figure 2.4.1: Marginal Tax Rates as Function of Monthly Income for the Year 2011



In contrast to other studies (like Sinn et al. (2006)), the highest phase-out rate is below 100%. This is because we consider contributions to the pension system not purely as a tax, as there is a Bismarckian pension system in place in Germany, see OECD (2011). Although the rate of return in the pension system is likely to be very low, it seems reasonable to assume that individuals will (on average) receive at least half of their (marginal) contributions as (higher) pensions; this reduces the effective marginal tax rate by about five percentage points.[20]

To estimate the income distribution we use data for the year 2011 of the German Socio-Economic Panel (SOEP), which is a representative sample of German households that are interviewed annually, see Wagner et al. (2007). Our sample (of singles, aged 18 to 65, out of education, and with sufficiently low assets) consists of 627 ob-

[18] In 2011 there was a reform of the welfare benefit system (Hartz-IV-Reform 2011) that was explicitly targeted to the phase out region, where marginal tax rates were reduced. We present the results for 2011, i.e. after the reform.

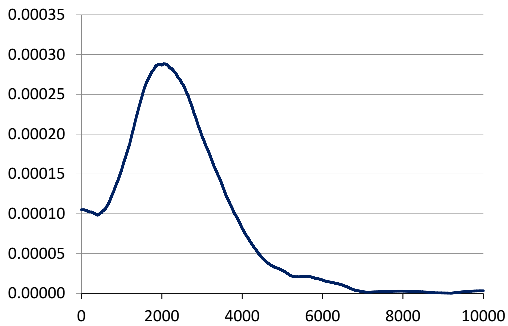
[19] There is a small downward jump in the tax schedule at 400 Euro, which is why T' tends to $-\infty$ at this income level. As this inefficiency is of second-order importance, we do not further comment on it. Also, there is a small spike at 1,398 Euro, which is due to the way the tax formula is stated in the tax code. As it arises due to rounding, it can be ignored. Note, that this small spike is also visible in Figure 2.4.3. Marginal tax rates are shifted upwards for income levels between 1,375 and 1,504 Euro because of the solidarity surcharge which is phased in at a higher rate in this interval.

[20] The result of an inefficient structure of the marginal tax rates is robust with regard to how contributions to the pension system are taken into account.

servations. The minimum and maximum value of gross monthly income are 0 and 13,921 Euro. The mean income is 1,977 Euro (2,375 Euro if restricted to positive incomes).

We estimate the density of the income distribution nonparametrically (using the standard SOEP weights), employing an Epanechnikov kernel and Silverman's rule of thumb to determine the bandwidth, see Fan and Gijbels (2003). Results for the Pareto efficiency test are, however, basically identical for different values of the bandwidth, so we refrain from applying any cross-validation procedure to determine an optimal bandwidth. The distribution of monthly gross incomes is shown (up to 10.000 Euro) in Figure 2.4.2.

Figure 2.4.2: Density of the Income Distribution for the Year 2011



We do not estimate elasticities ourselves but instead apply a range of values of the empirical literature. For the benchmark case we use 0.25 for the extensive elasticities (which we denote by ν), and 0.33 for the intensive elasticities, see Chetty et al. (2011), but our main result holds for a large set of values (see below).

2.4.2.2 Results

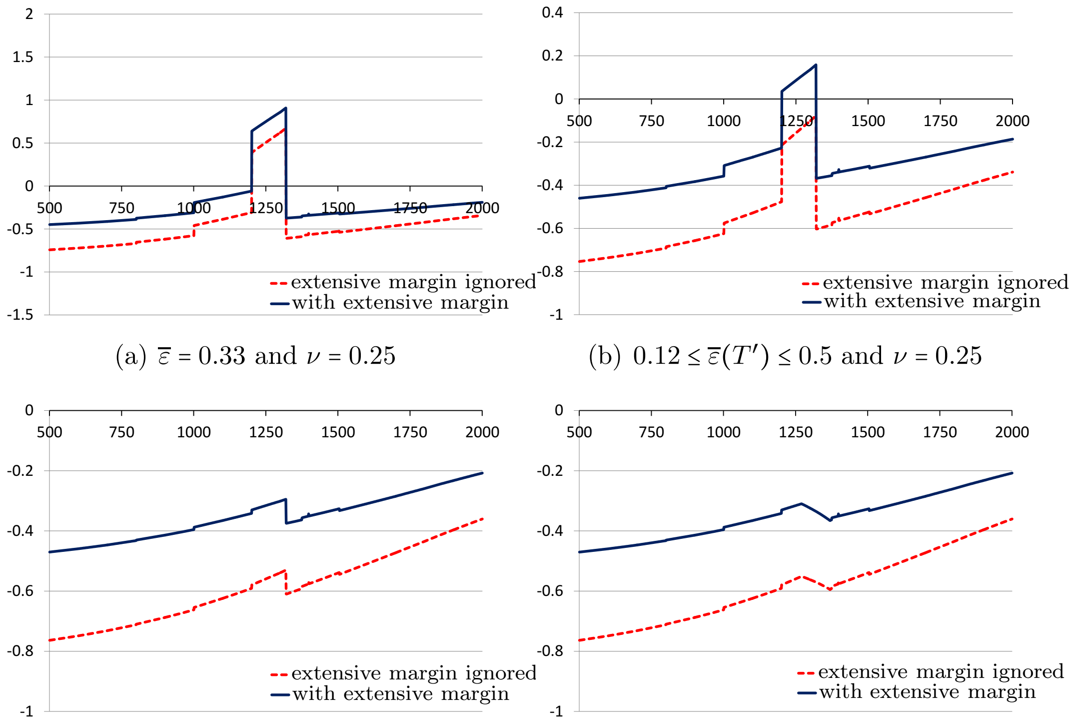
As marginal tax rates are very high in the phase-out region, one might suspect that they are beyond their Laffer value as defined in Section 2.4.1.1.

Figure 2.4.3(a) shows our test function (2.20) for the benchmark case with intensive elasticities $\overline{\varepsilon} = 0.33$ and extensive elasticities $\nu = 0.25$. For the interval where marginal tax rates are about 95%, they are indeed above their Laffer value, since the test function is larger than zero.[21]

This could be considered a strong result, but it may need the following qualification: Assuming an intensive elasticity that does not depend on the value of the marginal

[21] As in most data sets, top incomes are underrepresented in the SOEP data we use. Taking this into account would slightly weaken the case for the marginal tax rates being above their Laffer values. However, our main result, that the structure of marginal tax rates is inefficient, is independent of any underrepresentation of high incomes.

Figure 2.4.3: Graph of the Test Function (Left Hand Side of (2.17)) for Different Intensive Elasticities $\bar{\varepsilon}$ and Extensive Elasticity $\nu = 0.25$; (a)-(c) Original Tax Schedule, (d) Smoothed Tax Schedule.



(a) $\bar{\varepsilon} = 0.33$ and $\nu = 0.25$

(b) $0.12 \le \bar{\varepsilon}(T') \le 0.5$ and $\nu = 0.25$

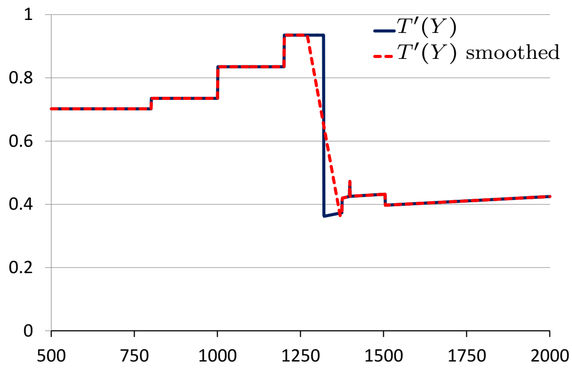
(c) $0 \le \bar{\varepsilon}(T') \le 0.5$ and $\nu = 0.25$

(d) $0 \le \bar{\varepsilon}(T') \le 0.5$ and $\nu = 0.25$, T' smoothed

tax rate may not be appropriate. With a constant elasticity, the percentage increase in income due to a 1 percentage point increase in T' strongly increases in T'. For example, a decrease in T' from 95% to 94% induces a relative increase in income that is 10 times as high as for a decrease from 50% to 49%. This would imply a 10 times as large semi-elasticity. Such huge differences in the semi-elasticities might be considered too large. We therefore also apply our test for the case that the semi-elasticity is constant, and for an intermediate case. To keep the semi-elasticity constant, we let the elasticity decrease linearly in T' from 0.5 to 0, i.e., we assume $\bar{\varepsilon}(T') = 0.5 - 0.5T'$.[22] In this case, the inefficiency according to Proposition 2.4.1 vanishes, see Figure 2.4.3(c), as our test function is now below zero. For the intermediate case, we let the semi-elasticity decrease less heavily in T' than is the case with a constant elasticity. When we assume $\bar{\varepsilon}(T') = 0.5 - 0.38T'$ (so that the lower bound for $\bar{\varepsilon}$ is not 0, but 0.12), our test identifies an inefficiency when extensive effects are incorporated, but fails to do so, if they are ignored, see Figure 2.4.3(b).

[22] For the intermediate value of $T' = 0.33$ we get the intensive elasticity $= 0.33$ that we assumed before.

Figure 2.4.4: Marginal Tax Rates and Smoothed Tax Rates as Function of Monthly Income for the Year 2011



Whether the German tax-transfer system passes the test of Proposition 2.4.1 is therefore very sensitive with respect to the elasticities. However, in all three cases (Figure 2.4.3(a)-(c)) the test curve is falling, so the test shows an inefficiency according to Proposition 2.4.2. Since marginal tax rates drop discontinuously (see Figure 2.4.1), one might argue that this is actually trivial. Whenever there is a discontinuity in the marginal tax rates, the test function (2.20) will be characterized by a discontinuous downward jump if elasticities and the income distribution are smooth. We therefore test the following: We smooth the decrease in marginal tax rates and determine how large the smoothing interval has to be (leaving everything else equal), so that the inefficiency disappears; see Figure 2.4.4 for a smoothing interval of 100 Euro (50 Euro below and above the discontinuity in T'). If the interval where the inefficiency just disappears is small, the inefficiency is of second-order importance. However, this is not the case. Figure 2.4.3(d) shows our test function for a smoothing interval of 100 Euro and for the same elasticities as in Figure 2.4.3(c): The inefficiency clearly stays present. Indeed the inefficiency does not disappear for any smoothing interval smaller than 326 Euro.[23]

The decrease of marginal tax rates after transfers are phased out, as it is observed in many countries, has already been criticized by (Kaplow, 2007, p. 304). Referring to results from numerical simulations based on Utilitarian welfare functions, he argues that marginal tax rates in the phase-out region are too high, and too low afterwards. For Germany we show this to be correct but also make the argument even stronger

[23]Considering in addition income effects would of course yield different numbers. With income effects, a tax reform as depicted in Figure 2.3.1 would slightly reduce labor supply of individuals earning Y, which would decrease tax revenue and therefore make such a reform less likely to be feasible. As the tax schedule remains inefficient even for a large smoothing interval, taking into account income effects would not change our main result but only slightly decrease the extent of the inefficiency, especially because the literature has found income effects to be rather small, see Meghir and Phillips (2010) and the references therein.

since the tax-transfer system is (second-best) Pareto inefficient and can therefore not be justified by any welfare function.

2.4.2.3 Possible Reforms and Policy Implications

The German tax-transfer system has often been criticized for its disincentives to work for individuals with low incomes. One proposal has been to lower marginal tax rates in the phase-out region, financed by a decrease in the welfare benefit (Sinn et al. (2006)). For individuals that cannot find a job this proposal also included a guaranteed job offer in the public sector; if accepted, transfers would then be as high as before the reform. Such a reform would increase employment, but its welfare consequences are ambiguous because at least some of the welfare recipients are worse off.

Our analysis clearly suggests that the high marginal tax rates in the phase-out region are indeed hard to justify. If intensive elasticities in this region are above 0.1, they might even be above their Laffer value, which implies that slightly lowering them would increase tax revenue. As already mentioned, this result may have to be qualified because high incomes are underrepresented in the SOEP data. <

In contrast, the inefficiency identified by the second test is independent of how accurately the density of high income is estimated (see equation (2.19)). Also, the test indicates an inefficiency for a very wide range of elasticities. We therefore conclude that there is room for a Pareto improving reform, where phase-out rates of the welfare benefits ("Hartz IV") are decreased and therefore the income region where individuals receive transfers is increased; as these new transfer recipients would then face transfer phase-out rates as well, their effective marginal tax rate would be increased. Since the absolute level of taxes does not increase for any income level, no individual is made worse off, but tax revenue increases due to the induced labor supply responses along the intensive and extensive margin. To obtain concrete numbers for such a reform so that not only a Pareto improvement, but also a Pareto optimal allocation is achieved one has to make structural assumptions about the elasticities in order to predict labor supply responses to reforms that come with substantial changes in marginal tax rates. One should then also take consumption taxes into account. Whereas this would have no effect on the general pattern of such a Pareto improving reform, it may well influence concrete numbers.

2.5 Conclusion

We analyzed the optimal design of tax-transfer systems in the presence of intensive and extensive labor supply responses. We derived optimality conditions for the entire tax schedule, but our interest was mainly on that part of the schedule where individuals receive transfers. More specifically, we asked whether participation subsidies and high marginal tax rates in the transfer phase-out region can be grounded in optimal tax theory.

Concerning participation subsidies, we derived a condition for negative participation taxes to be never part of an optimal tax schedule: the social marginal utility of the lowest income worker being smaller than the marginal value of public funds. We thereby extended the result of Diamond (1980), Saez (2002) and Choné and Laroque (2011b) to the case where in addition to extensive labor supply responses also intensive labor supply responses are present.

Regarding the issue of high marginal tax rates in the phase-out region, we developed a test for the Pareto inefficiency of a given tax-transfer system. This test is expressed in reduced form and is an extension of Werning (2007) to the case of intensive and extensive labor supply responses. When applied to the German tax-transfer system the results suggest an inefficient structure of marginal tax rates: a decrease of marginal tax rates in the phase-out region combined with an increase of marginal tax rates for slightly higher incomes could yield a Pareto improvement. Using these insights as a starting point for analyzing Pareto improving tax reforms in a structural labor supply model for Germany in the spirit of Blundell and Shephard (2012) would be an interesting task for future research.

Applying this test to other countries would also be worthwhile. Constructing the schedule of effective marginal tax rates, however, requires detailed knowledge of the interplay of the tax code and all elements of the welfare benefit program (at the federal and the state level in some countries), but once the schedule is known, the test can easily be applied. The extension of such a test to tax-transfer systems for couples and families should also be pursued.[24] This would add additional interesting aspects because the marginal tax rate of the primary earner often depends on the earnings of the secondary earner and vice versa.

[24]See Cremer et al. (2012), Kleven et al. (2009), Immervoll et al. (2011) and Bach et al. (2012) for recent contributions to the optimal tax treatment of couples.

Appendix 2.A Proofs for Section 2.2

2.A.1 Derivation of the Effective Tax Schedule

In the following, the capital letters in parenthesis refer to columns in the Excel file containing the calculation of the effective tax schedule; the file can be downloaded at
https://www.dropbox.com/s/ra405spwaegbhji/German-tax-transfer-system.xlsx.
An individual with gross income y (column A) may have to pay contributions to the pension fund (C to F), the sickness fund (G to J), the long term care fund (K to N) and to the unemployment insurance (O to R).[25]
The individual may also have to pay income taxes. The income tax schedule including the solidarity surcharge (AF to AP) is applied to the taxable income (AE) which is derived by subtracting several tax deductibles (S to AB, including part of the contributions to the social insurance system) from gross income (A).
An individual without income receives the welfare benefit of 675 Euro (AQ). If the individual earns a positive gross income (A), this welfare benefit is reduced by a fictitious "net"-income (AZ): this "net"-income equals gross income (A) minus a deductible (AW, which itself depends on gross income), the income tax (AR) and the contributions to the social insurance system (AS).[26]
The actual net income of the individual (including the welfare benefit) can be found in column BC. It is given by the income dependent welfare benefit (BA) – i.e. the constant benefit of 675 Euro minus the fictitious "net"-income – plus gross income, minus contributions to the social security system, minus income taxes. Column BD then contains the marginal tax rate that results from this actual income.
Finally, one has to decide how to take contributions to the pension fund into account: Since there is a Bismarckian pension system in place in Germany, see OECD (2011), higher contributions imply higher pension benefits (with an approximately linear relationship between contributions and benefits). However, due to the demographic situation in Germany this Bismarckian pension system may not be sustainable, in which case higher contributions may not increase benefits. Column BG contains the effective marginal tax rate if contributions are considered purely as a tax, and column BO if contributions increase benefits by an equal amount; column BK contains the intermediate case where only half of the contributions are considered to be a tax. This is the effective marginal tax schedule that we use for our Pareto-efficiency test.

2.A.2 Proof of Proposition 2.2.4

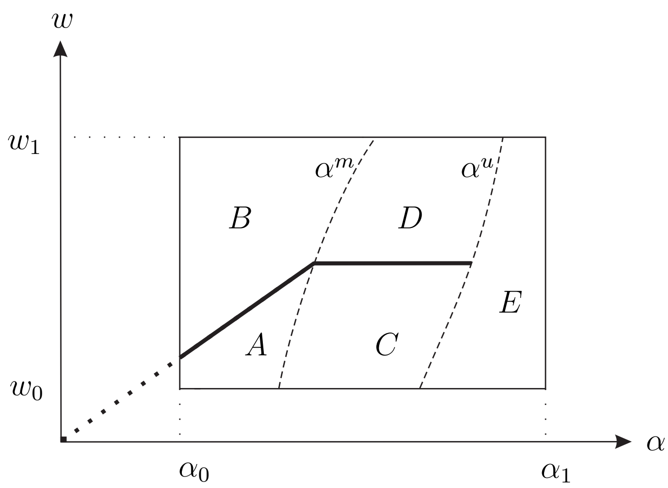
To prove that (NDC), (EC), (MC) and the definition of the thresholds imply incentive compatibility, we use Figure 2.A.1, where a representative iso-income curve is illustrated. We first argue that an iso-income curve indeed has a shape as illustrated in Figure 2.A.1: Note that by Lemma 2.2.2 income and consumption are constant along the increasing part of that line because $\frac{w}{\alpha}$ is constant. By the definition of the $\alpha^m(w)$-curve, income is the same on the flat part of this kinked line; thus this curve is indeed an iso-income curve. Because of (NDC), we know that consumption on the

[25] In each case, the first column contains the contribution if income is below 400 Euro, the second if it is above 400 but below 800, and the third if it is above 800 Euro.

[26] To be more precise, this "net"-income is the maximum of zero and the gross income (A) minus the deductible (AW), the income tax (AR) and the contributions to the social insurance system (AS), i.e., if this "net"-income is below zero, it is set to zero.

flat part must also be equal to consumption on the increasing part, so the iso-income curve is also an iso-consumption-curve. Finally note that by the definition of the $\alpha^u(w)$-curve, income is zero to the right of it and therefore consumption must be the same for all those types.

Figure 2.A.1: Incentive Compatibility Constraints



We now show that (NDC), (EC), (MC) and the definition of the thresholds imply incentive compatibility for all types on such a kinked line:

Incentive constraints for the increasing part: For the increasing part of the iso-income curve, (EC) and (MC) guarantee that no income-consumption bundle to the left of the $\alpha^m(w)$-curve is preferred; individuals on the increasing part prefer their income-consumption bundle to any income-consumption bundle in A or B. By the no-discrimination constraint we know that for each income-consumption bundle in D or C there exists an equivalent income-consumption bundle in A or B; individuals on the increasing part thus prefer their income-consumption bundle to any income-consumption bundle in C or D. Since along the flat part of the iso-income curve, utility is decreasing in α, we know that utility at the kink is larger than at the point where the iso-income curve intersects the $\alpha^u(w)$-curve; thus individuals on the increasing part prefer their income-consumption bundle to the income-consumption bundle in E.

Incentive constraints for the flat part: By the same argument as above, it follows that individuals on the flat part prefer their income-consumption bundle to that in E. By the minimum hours constraint, they cannot choose income-consumption bundles in A and C. Each consumption bundle in B is not preferred by the type on the kink of the curve as argued above. Since income in B is higher and disutility of work is increasing in α along the flat part, no individuals in the flat part prefer any income-consumption bundle in B (and therefore also none in D) to their own income-consumption bundle.

Finally, by the definition of the $\alpha^u(w)$-line and the same arguments as above, no individuals in E prefer any income-consumption bundle to the left of the $\alpha^u(w)$-line.

2.A.3 The First-Order Conditions of the Government's Problem

As a first step, rewrite the government's objective (2.1) in terms of $V(\beta)$, $(\alpha^u(w))$, $\alpha^m(w)$, $\alpha^m_\beta(\beta)$, b and $V(w,\alpha)$ as

$$
\begin{aligned}
W &= \int_{\underline{\beta}}^{\beta_1} \int_{\underline{\alpha}(\beta)}^{\alpha^m_\beta(\beta)} \Psi(V(\beta))\, d\widetilde{G}(\alpha|\beta)\, dK(\beta) \\
&+ \int_{w_0}^{w_1} \left[\int_{\alpha^m(w)}^{\alpha^u(w)} \Psi(V(w,\alpha))\, dG(\alpha|w) + \int_{\alpha^u(w)}^{\alpha_1} \Psi(b)\, dG(\alpha|w) \right] dF(w),
\end{aligned}
$$

where $\underline{\alpha}(\beta)$ is the lowest value of α associated with a certain value of beta, i.e. $\underline{\alpha}(\beta) = \frac{w_0}{\beta}$. Using the definition of the indirect utility function, replace $C(\beta)$ by $V(\beta) + v\left(\frac{Y(\beta)}{\beta}\right)$ and $C(w,\alpha)$ by $V(w,\alpha) + v(\alpha L_{min})$. The Lagrangian for the problem then reads as:

$$
\begin{aligned}
\mathcal{L} &= \int_{\underline{\beta}}^{\beta_1} \int_{\underline{\alpha}(\beta)}^{\alpha^m_\beta(\beta)} \Psi(V(\beta))\, d\widetilde{G}(\alpha|\beta)\, dK(\beta) \\
&+ \int_{w_0}^{w_1} \int_{\alpha^m(w)}^{\alpha^u(w)} \Psi(V(w,\alpha))\, dG(\alpha|w) dF(w) + \int_{w_0}^{w_1} \int_{\alpha^u(w)}^{\alpha_1} \Psi(b)\, dG(\alpha|w) dF(w) \\
&+\lambda \left\{ \int_{\underline{\beta}}^{\beta_1} \left[Y(\beta) - \left(V(\beta) + v\left(\frac{Y(\beta)}{\beta}\right)\right)\right] \widetilde{G}(\alpha^m_\beta(\beta)|\beta) dK(\beta) \right. \\
&\qquad + \int_{w_0}^{w_1} \int_{\alpha^m(w)}^{\alpha^u(w)} \left(wL_{min} - \left(V(w,a) + v\left(\alpha L_{min}\right)\right)\right) dG(\alpha|w) dF(w) \\
&\qquad \left. - \int_{w_0}^{w_1} b(1 - G(\alpha^u(w)|w)) dF(w) \right\} \\
&+ \int_{\underline{\beta}}^{\beta_1} \int_{\alpha^m(Y(\beta)/L_{min})}^{\alpha^u(Y(\beta)/L_{min})} \eta(\beta,\alpha) \left[V(\beta) + v\left(\frac{Y(\beta)}{\beta}\right) - V(w,\alpha) - v\left(\alpha L_{min}\right)\right] d\alpha d\beta \\
&+ \int_{\underline{\beta}}^{\beta_1} \left[\mu(\beta) V'(\beta) - \mu(\beta) v'\left(\frac{Y(\beta)}{\beta}\right) \frac{Y(\beta)}{\beta^2}\right] d\beta,
\end{aligned}
$$

where λ is the multiplier of the resource constraint, $\eta(\beta,\alpha)$ is the multiplier function of the no discrimination constraint and $\mu(\beta)$ is the multiplier function of the envelope condition. Partially integrating the term $\int_{\underline{\beta}}^{\beta_1} \mu(\beta) V'(\beta) d\beta$ yields $-\int_{\underline{\beta}}^{\beta_1} \mu'(\beta) V(\beta) + \mu(\beta_1) V(\beta_1) - \mu(\underline{\beta}) V(\underline{\beta})$, so that the last line of the Lagrangian can be replaced by

$$
+ \int_{\underline{\beta}}^{\beta_1} \left[-\mu'(\beta) V(\beta) - \mu(\beta) v'\left(\frac{Y(\beta)}{\beta}\right) \frac{Y(\beta)}{\beta^2}\right] d\beta + \mu(\beta_1) V(\beta_1) - \mu(\underline{\beta}) V(\underline{\beta}).
$$

The first order conditions are:

$$\frac{\partial \mathcal{L}}{\partial V(\beta)} = \int_{\underline{\alpha}(\beta)}^{\alpha_\beta^m(\beta)} \left(\Psi'(V(\beta)) - \lambda\right) d\widetilde{G}(\alpha|\beta)\, k(\beta) - \mu'(\beta) + \int_{\alpha^m(Y(\beta)/L_{min})}^{\alpha^u(Y(\beta)/L_{min})} \eta(\beta,\alpha) d\alpha = 0 \tag{2.21}$$

$$\left.\frac{\partial \mathcal{L}}{\partial V(w,\alpha)}\right|_{\alpha<\alpha^u} = \left(\Psi'(V(w,\alpha)) - \lambda\right) g(\alpha|w) f(w) - \eta(Y^{-1}((wL_{min}),\alpha) = 0 \tag{2.22}$$

$$\frac{\partial \mathcal{L}}{\partial V(w,\alpha^u(w))} = \left(\Psi'(V(w,\alpha)) - \lambda\right) g(\alpha|w) f(w) - \eta(Y^{-1}((wL_{min}),\alpha) + \lambda g(\alpha^u(w)|w)\frac{\partial \alpha^u(w)}{\partial V}\left(b + wL_{min} - \left(V(w,a) + v\left(\alpha^u(w)L_{min}\right)\right)\right) = 0 \tag{2.23}$$

$$\frac{\partial \mathcal{L}}{\partial Y(\beta)} = \lambda\left(1 - v'\left(\frac{Y(\beta)}{\beta}\right)\frac{1}{\beta}\right)\widetilde{G}(\alpha_\beta^m(\beta)|\beta)k(\beta) - \mu(\beta)\frac{v'\left(\frac{Y(\beta)}{\beta}\right) + v''\left(\frac{Y(\beta)}{\beta}\right)\frac{Y(\beta)}{\beta}}{\beta^2} + \int_{\alpha^m(Y(\beta)/L_{min})}^{\alpha^u(Y(\beta)/L_{min})} \eta(\beta,\alpha)\left[v'\left(\frac{Y(\beta)}{\beta}\right)\frac{1}{\beta} - \frac{\partial V}{\partial w}\frac{1}{L_{min}}\right] d\alpha = 0 \tag{2.24}$$

$$\frac{\partial \mathcal{L}}{\partial b} = \int_{w_0}^{w_1}\int_{\alpha^u(w)}^{\alpha_1} \Psi'(b) dG(\alpha|w) dF(w) - \lambda \int_{w_0}^{w_1}\int_{\alpha^u(w)}^{\alpha_1} dG(\alpha|w) dF(w) - \lambda \int_{w_0}^{w_1} \frac{\partial \alpha^u(w)}{\partial b} g(\alpha^u(w)|w)(b + wL_{min} - V(w,\alpha) - v(\alpha L_{min}))dF(w). \tag{2.25}$$

Finally the derivatives with respect to the endpoint conditions are

$$\frac{\partial \mathcal{L}}{\partial V(\beta_1)} = \mu(\beta_1) = 0 \tag{2.26}$$

and

$$\frac{\partial \mathcal{L}}{\partial V(\underline{\beta})} = \mu(\underline{\beta}) = 0. \tag{2.27}$$

2.A.4 Solution to the Government's Problem

First integrating (2.22) over α^m to α^u and adding (2.23), then integrating this expression over w_0 to w_1, finally adding (2.25) as well as (2.21) integrated over $\underline{\beta}$ to β_1 yields

$$\lambda = \int_{w_0}^{w_1}\int_{\alpha_0}^{\alpha_1} \Psi'(V(w,\alpha)) dG(\alpha|w) dF(w),$$

i.e. equation (2.5). Integrating (2.21) yields

$$\mu(\beta) = \int_{\beta}^{\beta_1}\int_{\underline{\alpha}(\beta')}^{\alpha^m_{\beta'}(\beta)} \left(\lambda - \Psi'(V(\beta'))\right) d\widetilde{G}(\alpha|\beta')dK(\beta') - \int_{\beta}^{\beta_1}\int_{\alpha^u(Y(\beta')/L_{min})}^{\alpha^m(Y(\beta')/L_{min})} \eta(\beta,\alpha)d\alpha d\beta'. \tag{2.28}$$

Inserting (2.22) and (2.23) into (2.28) then results in

$$\begin{aligned}\mu(\beta) = & \int_{\beta}^{\beta_1}\int_{\underline{\alpha}(\beta')}^{\alpha^m_{\beta}(\beta')} \left[\lambda - \Psi'(V(\beta'))\right] d\widetilde{G}(\alpha|\beta')dK(\beta') \\ & + \int_{\beta}^{\beta_1}\left[\int_{\alpha^m(Y(\beta')/L_{min})}^{\alpha^u(Y(\beta')/L_{min})}\left[\lambda - \Psi'\left(V\left(\frac{Y(\beta')}{L_{min}},\alpha\right)\right)\right]d\alpha \right. \\ & \left. +\lambda g(\alpha^u(w)|w)\frac{\partial\alpha^u(w)}{\partial V(w,\alpha^u(w))}\left(T(wL_{min})+b\right)\right]dK(\beta').\end{aligned}$$

Using $\frac{\partial V(w,\alpha)}{\partial w} = (1-T'(Y(\beta)))L_{min}$ and $v'\left(\frac{Y(\beta)}{\beta}\right)\frac{1}{\beta} = 1-T'(Y(\beta))$ to simplify (2.24) yields

$$\lambda\left(1-v'\left(\frac{Y(\beta)}{\beta}\right)\frac{1}{\beta}\right)\widetilde{G}(\alpha^m(\beta)|\beta)k(\beta) - \mu(\beta)\frac{v'\left(\frac{Y(\beta)}{\beta}\right)+v''\left(\frac{Y(\beta)}{\beta}\right)\frac{Y(\beta)}{\beta}}{\beta^2} = 0. \tag{2.29}$$

Inserting (2.28) into (2.29) and using $\varepsilon_{Y,1-T'} = \frac{\beta^2}{v''}\frac{1-T'}{Y(\beta)}$, (where $\frac{\partial Y}{\partial(1-T')} = \frac{\beta^2}{v''}$ can be derived by implicitly differentiating the FOC of the unconstrained individuals), we have

$$\begin{aligned}&\frac{T'(Y(\beta))}{1-T'(Y(\beta))}\lambda\beta\left(\frac{\varepsilon_{Y,1-T'}}{\varepsilon_{Y,1-T'}+1}\right)\widetilde{G}(\alpha^m(\beta)|\beta)k(\beta) \\ = & \int_{\beta}^{\beta_1}\int_{\underline{\alpha}(\beta)}^{\alpha^m_{\beta}(\beta)}\left(\lambda-\Psi'(V(\beta'))\right)d\widetilde{G}(\alpha|\beta')dK(\beta') \\ & + \int_{\frac{Y(\beta)}{L_{min}}}^{w_1}\left[\int_{\alpha^m(w)}^{\alpha^u(w)}\left(\lambda-\Psi'(V(w,\alpha))\right)dG(\alpha|w)\right. . \\ & \left. +\lambda g(\alpha^u(w)|w)\frac{\partial\alpha^u(w)}{\partial T(wL_{min})}\left(T(wL_{min})+b\right)\right]dF(w).\end{aligned}$$

Together with the endpoint conditions (2.26) and (2.27) this constitutes the solution.

2.A.5 Proof of Proposition 2.2.5

Let $\overline{\Psi}'(Y)$ be the average marginal utility of income of all individuals earning income Y; it is given by

$$\overline{\Psi}'(Y(\beta)) = \frac{\int_{\underline{\alpha}(\beta)}^{\alpha^m(\beta)}\Psi'(V(\beta))dG(\alpha|\beta)h(\beta)\frac{\partial\beta}{\partial Y} + \int_{\alpha^m(w_\beta)}^{\alpha^u(w_\beta)}\Psi'(V(w_\beta,\alpha))dG(\alpha|w_\beta)\frac{f(w_\beta)}{L_{min}}}{\tilde{h}(Y(\beta))}.$$

Using $\overline{\Psi}'(Y)$, the first and second line of (2.4) can be rewritten as $\int_{Y}^{Y_{\max}}\left[(\lambda-\overline{\Psi}'(\widetilde{Y}))\right]dH(\widetilde{Y})$. Let $\xi(Y)$ be the semi-elasticity of participation, i.e. the increase in the number of unemployed relative to the number of individuals earning income level Y, $h(Y)$, due to an absolute increase in

$T(Y)$ (or b); it is given by

$$\xi(Y(\beta)) = \frac{-\frac{\partial \alpha^u(w_\beta)}{\partial T} g(\alpha^u(w_\beta)|w_\beta)\frac{f(w_\beta)}{L_{min}}}{h(Y(\beta))}.$$

Applying integration by substitution, the term

$$\int_{\frac{Y(\beta)}{L_{min}}}^{w_1} g(\alpha^u(w)|w)\frac{\partial \alpha^u(w)}{\partial T(wL_{min})}(T(wL_{min})+b)dF(w)$$

can be rewritten as

$$\int_{Y(\beta)}^{w_1 L_{min}} \underbrace{g\left(\alpha^u\left(\frac{Y}{L_{min}}\right)\middle|\frac{Y}{L_{min}}\right)\frac{\partial \alpha^u\left(\frac{Y}{L_{min}}\right)}{\partial T(Y)}\frac{1}{h(Y)}\frac{f\left(\frac{Y}{L_{min}}\right)}{L_{min}}}_{\xi(Y)}(T(Y)+b)dH(Y).$$

Using this, the third line of (2.4) can be rewritten as

$$\int_{Y}^{Y_{\max}} -\lambda\xi(\widetilde{Y})T_{part}(\widetilde{Y})\, dH(\widetilde{Y}).$$

Using the definition of $\varepsilon_{Y,\beta}$ and the first order condition $(1-T')\beta = v'$, we have

$$\varepsilon_{Y,\beta} = -\frac{v'\left(\frac{Y(\beta)}{\beta}\right)\frac{1}{\beta} + v''\left(\frac{Y(\beta)}{\beta}\right)\frac{Y(\beta)}{\beta^2}}{-\frac{1}{\beta^2}v''\left(\frac{Y(\beta)}{\beta}\right)} = \frac{(1-T')\beta^2}{v''\left(\frac{Y(\beta)}{\beta}\right)Y(\beta)} + 1 = \varepsilon_{Y,1-T'} + 1.$$

We can therefore rewrite the first term of (2.3) as

$$\frac{T'(Y(\beta))}{1-T'(Y(\beta))}\lambda\beta\frac{\varepsilon_{Y,1-T'}}{\frac{\partial Y}{\partial \beta}\frac{\beta}{Y}}\widetilde{G}(\alpha^m(\beta)|\beta)k(\beta). \tag{2.30}$$

Since the average elasticity $\overline{\varepsilon}(Y)$ for income Y and the elasticity $\varepsilon_{Y,1-T'}$ are linked by

$$\frac{\overline{\varepsilon}(Y)}{\varepsilon_{Y,1-T'}} = \frac{\widetilde{G}(\alpha^m(\beta)|\beta)k(\beta)}{h(Y)\frac{\partial Y}{\partial \beta}},$$

we can rewrite (2.30) as $\frac{T'(Y)}{1-T'(Y)}\lambda Y\overline{\varepsilon}(Y)h(Y)$.

2.A.6 Proof of Conditions for Nonnegative Marginal Tax Rates

In our minimum hours model we have $\xi(Y) = 0\ \forall\, Y \geq \overline{Y}$ for some $\overline{Y} \in]Y_{min}, Y_{max}[$. In the following we show that – if $\frac{\partial}{\partial Y}\left[\frac{\lambda - \overline{\Psi}'(Y)}{\xi(Y)}\right] > 0$ and $\overline{\Psi}'(Y)$ is decreasing in income – there can be no interval $]Y_1, Y_2[$ with negative marginal tax rates for

- Case 1: $Y_1 < Y_2 < \overline{Y}$,
- Case 2: $Y_1 < \overline{Y} \leq Y_2$,
- Case 3: $\overline{Y} \leq Y_1 < Y_2$.

In a model where the extensive margin is always present, only Case 1 applies.

In all three cases we would have $\mathcal{A}(Y_1) = 0$ and $\mathcal{A}(Y_2) = 0$ because $T'(Y_1) = 0$ and $T'(Y_2) = 0$.[27] We would also have $\mathcal{A}'(Y_1) \leq 0$ and $\mathcal{A}'(Y_2) \geq 0$, because for T' to be negative in $]Y_1, Y_2[$, $\mathcal{A}(Y)$ has to be negative within this interval.

Case 1: In this case we have

$$\mathcal{A}'(Y) = [\overline{\Psi}'(Y) - \lambda + \lambda\xi(Y)T_{part}(Y)]h(Y) \tag{2.31}$$

for $Y = Y_1$ and $Y = Y_2$. Solving $\mathcal{A}'(Y_1) \leq 0$ for $T_{part}(Y_1)$ and $\mathcal{A}'(Y_2) \geq 0$ for $T_{part}(Y_2)$, and using $T_{part}(Y_1) > T_{part}(Y_2)$ since marginal tax rates are negative, we have

$$\frac{\lambda - \overline{\Psi}'(Y_1)}{\lambda\xi(Y_1)} \geq \frac{\lambda - \overline{\Psi}'(Y_2)}{\lambda\xi(Y_2)},$$

which cannot hold if $\frac{\partial}{\partial Y}\left[\frac{\lambda - \overline{\Psi}'(Y)}{\xi(Y)}\right] > 0$.

Case 2: Using (2.31), in this case we have

$$\begin{aligned} \overline{\Psi}'(Y_1) - \lambda + \lambda\xi(Y_1)T_{part}(Y_1) &\leq 0 \\ \overline{\Psi}'(Y_2) - \lambda &\geq 0, \end{aligned}$$

which implies $\lambda\xi(Y_1)T_{part}(Y_1) \leq \overline{\Psi}'(Y_2) - \overline{\Psi}'(Y_1)$. Because the right hand side is negative, this requires $T_{part}(Y_1) < 0$. We now have to distinguish two cases:
If $\overline{\Psi}'(Y_1) < \lambda$, we would then have $T_{part}(Y_1 + \epsilon) < 0$, $\overline{\Psi}'(Y_1 + \epsilon) < \lambda$ and $T'(Y_1 + \epsilon) < 0$, which cannot hold as we show in the proof of Proposition 2.3.1.
If $\overline{\Psi}'(Y_1) > \lambda$, we must have $\overline{\Psi}'(\widetilde{Y}) = \lambda$ for some $\widetilde{Y} \leq \overline{Y}$; if not, $\frac{\partial}{\partial Y}\left[\frac{\lambda - \overline{\Psi}'(Y)}{\xi(Y)}\right] > 0$ would be violated (close to $\overline{Y}$). Again we have to distinguish two cases:
If $\widetilde{Y} < \overline{Y}$ or $\widetilde{Y} = \overline{Y} < Y_2$, we would have $T_{part}(\widetilde{Y} + \epsilon) < 0$, $\overline{\Psi}'(\widetilde{Y} + \epsilon) < \lambda$ and $T'(\widetilde{Y} + \epsilon) < 0$, which again cannot hold as we show in the proof of Proposition 2.3.1.
If $\widetilde{Y} = \overline{Y} = Y_2$, we would have $\xi(\widetilde{Y}) = 0$ and $\overline{\Psi}'(\widetilde{Y}) = \lambda$. But then, since $\overline{\Psi}'(Y) < \lambda \ \forall Y > \overline{Y}$, we would have $\mathcal{A}(Y_2) > 0$, see (2.8).

Case 3: In this case we have

$$\mathcal{A}'(Y) = (\overline{\Psi}'(Y) - \lambda)h(Y)$$

for $Y = Y_1$ and $Y = Y_2$. $\mathcal{A}'(Y_1) \leq 0$ and $\mathcal{A}'(Y_2) \geq 0$ would imply $\overline{\Psi}'(Y_2) \geq \overline{\Psi}'(Y_1)$, a contradiction to decreasing social marginal utility of income.

[27] If $Y_1 = Y_{min}$, we may have $T'(Y_{min}) \neq 0$, but nevertheless $\mathcal{A}(Y_{min}) = 0$, see (2.10). The same applies for Y_{max}, see (2.10) again.

Bibliography of Chapter 2

BACH, S., G. CORNEO, AND V. STEINER (2012): "Optimal top marginal tax rates under income splitting for couples," *European Economic Review*, 56, 1055 – 1069.

BARGAIN, O., M. DOLLS, D. NEUMANN, A. PEICHL, AND S. SIEGLOCH (2011): "Tax-benefit systems in Europe and the US: between equity and efficiency," *CESifo Working Paper 3534*.

BLUNDELL, R., M. BREWER, P. HAAN, AND A. SHEPHARD (2009): "Optimal Income Taxation of Lone Mothers: An Empirical Comparison Of The UK and Germany," *The Economic Journal*, 119, 101–121.

BLUNDELL, R. AND A. SHEPHARD (2012): "Employment, Hours of Work and the Optimal Taxation of Low-Income Families," *The Review of Economic Studies*, 79, 481–510.

BOONE, J. AND L. BOVENBERG (2004): "The optimal taxation of unskilled labor with job search and social assistance," *Journal of Public Economics*, 88, 2227–2258.

BRETT, C. AND J. A. WEYMARK (2003): "Financing education using optimal redistributive taxation," *Journal of Public Economics*, 87, 2549–2569.

CHETTY, R. (2009): "Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods," *Annual Review of Economics*, 1, 451–488.

CHETTY, R., A. GUREN, D. MANOLI, AND A. WEBER (2011): "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins," *American Economic Review Papers and Proceedings*, 101, 471–475.

CHONÉ, P. AND G. LAROQUE (2011a): "Negative marginal tax rates and heterogeneity," *American Economic Review*, 100, 2532–47.

——— (2011b): "Optimal Taxation in the Extensive Model," *Journal of Economic Theory*, 146, 425–453.

CHRISTIANSEN, V. (2012): "Optimal Participation Taxes," *CESifo Working Paper No. 3941*.

CREMER, H., J.-M. LOZACHMEUR, AND P. PESTIEAU (2012): "Income taxation of couples and the tax unit choice," *Journal of Population Economics*, 25, 763–778.

DA COSTA, C. AND T. PEREIRA (2011): "Sacrifice and Efficiency of the Income Tax Schedule," *Escola de Pós-Graduação em Economia da FGV.*

DIAMOND, P. (1980): "Income Taxation with Fixed Hours of Work," *Journal of Public Economics*, 13, 101–110.

EISSA, N. AND J. B. LIEBMAN (1996): "Labor Supply Response to the Earned Income Tax Credit," *The Quarterly Journal of Economics*, 111, 605–637.

FAN, J. AND I. GIJBELS (2003): *Local Polynomial Modelling and Its Applications*, Boca Raton: CRC Press.

GIELEN, A. C. (2009): "Working hours flexibility and older workers' labor supply," *Oxford Economic Papers*, 61, 240–274.

GUSTMAN, A. L. AND T. L. STEINMEIER (2008): "Projecting Behavioral Responses to the Next Generation of Retirement Policies," *Research in Labor Economics*, 28, 141–195.

HECKMAN, J. J. (1993): "What Has Been Learned about Labor Supply in the Past Twenty Years?" *American Economic Review*, 83, 116–121.

IMMERVOLL, H., H. KLEVEN, C. KREINER, AND E. SAEZ (2007): "Welfare Reform in European Countries: A Microsimulation Analysis," *The Economic Journal*, 117, 1–44.

IMMERVOLL, H., H. KLEVEN, C. KREINER, AND N. VERDELIN (2011): "Optimal Tax and Transfer Programs for Couples with Extensive Labor Supply Responses," *Journal of Public Economics*, 95, 1485–1500.

JACQUET, L., E. LEHMANN, AND B. VAN DER LINDEN (2012): "Signing distortions in optimal tax and other adverse selection problems with random participation," *CESifo Working Paper 3766.*

KANBUR, R., M. KEEN, AND M. TUOMALA (1994): "Optimal non-linear income taxation for the alleviation of income-poverty," *European Economic Review*, 38, 1613 – 1632.

KAPLOW, L. (2007): "Optimal Income Transfers," *International Tax and Public Finance*, 14, 295–325.

Kleven, H., C. Kreiner, and E. Saez (2009): "The Optimal Income Taxation of Couples," *Econometrica*, 77, 537–560.

Kleven, H. J. and C. T. Kreiner (2006): "The marginal cost of public funds: Hours of work versus labor force participation," *Journal of Public Economics*, 90, 1955–1973.

Laroque, G. (2005): "Income Maintenance and Labor Force Participation," *Econometrica*, 73, 341–376.

Meghir, C. and D. Phillips (2010): "Labour Supply and Taxes," in *Dimensions of Tax Design: The Mirrlees Review*, ed. by Institute for Fiscal Studies, Oxford University Press.

Meyer, B. D. and D. T. Rosenbaum (2001): "Welfare, The Earned Income Tax Credit, and The Labor Supply of Single Mothers," *Quarterly Journal of Economics*, 116, 1063–1114.

Moffitt, R. (1982): "The Tobit Model, Hours of Work and Institutional Constraints," *The Review of Economics and Statistics*, 64, 510–515.

OECD (2011): "Pensions at a Glance 2011: Retirement-Income Systems in OECD and G20 Countries," URL: www.oecd.org/els/social/pensions/PAG.

Piketty, T. (1997): "La redistribution fiscale face au chomage," *Revue française d'économie*, 12, 157–201.

Pirttilä, J. and M. Tuomala (2004): "Poverty alleviation and tax policy," *European Economic Review*, 48, 1075–1090.

Saez, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205–229.

——— (2002): "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses," *Quarterly Journal of Economics*, 117, 1039–1073.

Scheuer, F. (2012): "Entrepreneurial Taxation with Endogenous Entry," *Working Paper, Stanford University.*

Sinn, H.-W., C. Holzner, W. Meister, W. Ochel, and M. Werding (2006): "Aktivierende Sozialhilfe 2006: Das Kombilohn-Modell des ifo Instituts," *ifo Schnelldienst*, 59, 6–27.

STEWART, M. B. AND J. K. SWAFFIELD (1997): "Constraints on the Desired Hours of Work of British Men," *The Economic Journal*, 107, 520–535.

WAGNER, G., J. FRICK, AND J. SCHUPP (2007): "The German Socio-Economic Panel Study – Scope, Evolution and Enhancements," *Schmoller's Jahrbuch*, 127, 626–654.

WERNING, I. (2007): "Optimal Fiscal Policy with Redistribution," *The Quarterly Journal of Economics*, 122, 925–967.

Chapter 3

Efficient Labor and Capital Income Taxation over the Life Cycle

3.1 Introduction

How can a government efficiently redistribute income among individuals? Should the government solely rely on the taxation of labor income or should capital be taxed as well? What role does the evolution of inequality over the life cycle play for the design of optimal policies?

This chapter addresses these questions in a life cycle framework. Consistent with a large empirical literature, inequality changes over the life cycle, both because of forecastable heterogeneity across individuals and because of idiosyncratic risk. Consistent with current tax practices, we characterize taxes that condition on current (annual) earnings. The labor income tax is allowed to be fully non-linear in the tradition of the seminal approach to optimal taxation pioneered by Mirrlees (1971). The tax on wealth (or equivalently capital income) is linear for tractability.

We show that a novel and simple formula for the optimal taxation of capital arises in our simple life cycle model. It follows a standard public finance trade-off: the tax rate tends to increase in wealth inequality and decrease in the elasticity of savings with respect to this tax rate. For commonly used social welfare functions, the tax on wealth is likely to be positive, as higher income households tend to hold more wealth. Wealth accumulation is a consequence of precautionary savings in our framework.

Optimal labor taxes in our life cycle setting are determined by the same forces as in the static models of labor income taxation (Diamond, 1998; Saez, 2001). An additional force arises, however, in the presence of capital income taxes. In particular, the labor income tax formulas are adjusted by how much labor taxes influence savings decisions.

We then turn to numerical simulations of optimal policies, using estimates from the recent literature on earnings dynamics over the life cycle (Karahan and Ozkan, 2013). We can confirm the intuition from the theoretical analysis of the model that the government taxes capital income at positive rates. If capital taxes are allowed to be age-dependent, they increase over the life cycle as wealth concentration increases. If labor taxes are allowed to be age-dependent, they also increase over the life cycle as labor income inequality increases.

The normative theory of capital taxation is a controversial subject in the literature. An important and influential benchmark in the academic and popular debate on capital taxation is the simple life cycle model by Atkinson and Stiglitz (1976). In their seminal framework, individuals differ in labor abilities, preferences are (weakly) separable between consumption and leisure and there is a retirement period individuals want to save for. The main result is that a nonlinear labor income tax is the more efficient tool for redistribution and the optimal capital tax rate is zero.[1] One intuition, e.g. recently stated by Piketty and Saez (2013), for this strong normative prescription is that heterogeneity is one-dimensional in the Atkinson and Stiglitz (1976) framework. This makes it sufficient to use only one instrument (the labor tax) applied to the source of heterogeneity directly without distorting other margins of behavior (savings). Many empirical papers have documented that the distribution of wages and labor income within a cohort changes dramatically over their life cycle (among many, see, e.g., Heathcote et al. (2010)). With changes in inequality over the life cycle, heterogeneity becomes multidimensional. We argue and show that with a realistic life cycle structure with changing inequality, a natural role for the capital tax arises on top of nonlinear labor income taxes.[2] In our model, wealth inequality arises as inequality in labor productivities changes both because of forecastable reasons and because of idiosyncratic risk. Individuals engage in precautionary savings and a wealth distribution arises – as is well understood from the incomplete-markets literature (Aiyagari, 1994). The optimal capital tax rate tends to increase with wealth inequality as the wealthy also consume more and consequently have a lower social marginal welfare weight.

A relatively recent literature, often called the *New Dynamic Public Finance* (NDPF), has explicitly taken into account how inequality evolves over the life cycle because of idiosyncratic risk, as is done in this chapter. Our current work is particularly related to the papers by Golosov et al. (2011) and Farhi and Werning (2013), who characterize history-dependent optimal labor and savings distortions in such dynamic environments. In contrast to their work we limit our attention to simple tax structures which are only allowed to condition on current earnings (and potentially age). We view these two approaches as clearly complementary. The NDPF approach has the advantage that history-dependent tax systems are more powerful to raise welfare. Our approach has the advantage of being within the realm of current tax practices.

There is an increasing interest in tax reforms which would move current tax policies towards conditioning on the age of the taxpayer. Weinzierl (2011) and Bastani et al. (2011) study optimal age-dependent labor income taxation in a discrete type model with a small number of types. They find large welfare gains from age-dependent labor income taxes and find these taxes to be increasing with age. In contrast, we

[1] This result has been generalized by Kaplow (2006) to the case where the labor income tax is not chosen optimally by the government.

[2] Some papers have emphasized that time preference heterogeneity may justify positive capital taxation in models similar to the Atkinson-Stiglitz framework; see Saez (2002), Diamond and Spinnewijn (2011) and Golosov et al. (2013a).

develop a first-order approach, which allows to study a much richer type space, in line with the continuous version of the Mirrlees model (Saez, 2001; Golosov et al., 2011). Thus, we are able to optimize over a fully nonlinear labor income tax schedule and characterize it theoretically, connecting it precisely to the static literature. In addition, our focus also lies on age-dependent capital income taxation, which we find to increase over the life cycle.[3] Conesa et al. (2009), in tradition with the Ramsey approach to optimal taxation, study optimal labor and capital income taxes in a computational life cycle framework. While our approach shares some features with a Ramsey type of exercise, we allow labor income taxes to be an arbitrarily non-linear function in the Mirrlees tradition and theoretically highlight the forces driving labor and capital taxation.

Finally our paper is related to Golosov et al. (2013b), who study general dynamic tax reforms and elaborate the welfare gains from the sophistication of the tax code such as age-dependence, history-dependence or joint taxation of labor and capital income. They also show that the presence of savings responses favors increasing labor income taxes over the life cycle in the presence of age-dependent taxation. Whereas our formulas – when written as functions of behavioral responses – are somehow encompassed by their results, our study differs in that we are interested in characterizing the optimum for our given tax instruments.

This chapter is structured as follows. In Section 3.2, we state our formal framework and show how we make the optimal tax problem tractable. In Section 3.3, we derive and discuss our results on optimal taxes. We first discuss the simpler two periods case in Section 3.3.1 and then show how our results generalize to the T-periods case in Section 3.3.2. Section 3.4 contains the calibration and simulation of optimal policies and in Section 3.5 we conclude.

3.2 The Formal Framework

3.2.1 The Model

We consider a life cycle framework with T periods where individuals at a certain point in time t are characterized by their productivity θ_t. Further, we define the history of shocks as $\theta^t = (\theta_1, \theta_2, \cdots, \theta_t)$. In each period, individuals make a savings

[3] Referring to the study of Weinzierl (2011), Banks and Diamond (2011) emphasize that further research in this area "seems to have a good probability of leading to significant policy improvements".

and a labor supply decision. Flow utility is given by

$$U(c_t, y_t, \theta_t) = U\left(c_t - \Psi\left(\frac{y_t}{\theta_t}\right)\right),$$

where $U'' < 0$, $\Psi'' > 0$, c_t is consumption in period t and y is gross income in period t. $\frac{y_t}{\theta_t}$ captures labor effort. For brevity, we sometimes write $R_t = c_t - \Psi\left(\frac{y_t}{\theta_t}\right)$. Abusing notation, we sometimes write the utility function or marginal utility as function of the history of shocks, i.e. $U(\theta^t)$ and $U'(\theta^t)$.

Importantly, the functional form of U eliminates income effects on labor supply, while allowing for risk-aversion. As we explain in more detail below, this assumption is crucial for the tractability of the dynamic optimal tax problem. The empirical literature using detailed micro data sets has typically not rejected a zero income elasticity on labor supply or found very small effects (see Gruber and Saez (2002) for the US or a recent paper by Kleven and Schultz (2012) using the universe of Danish tax records).[4] Eliminating income effects has also proven to be a key simplification in making progress on the theoretical and computational side in public finance models and especially in optimal tax problems (Diamond, 1998; Golosov et al., 2011).

We assume that agents already differ in the first period. The conditional density function (*cdf*) of the initial distribution of productivities is denoted by $F(\theta)$ and captures the ex-ante heterogeneity of agents. The reader should think about this heterogeneity as the level of heterogeneity of individuals at the age of roughly 25.

In the following periods, productivities evolve stochastically over time according to a Markov process. The respective *cdf* is $F_t(\theta_t|\theta_{t-1})$. Further, let $H_t(\theta^t)$ be the measure over the history θ^t.[5] We consider an economy with exogenous prices, so the interest on savings r is fixed. Further, we assume incomplete markets in a sense that individuals only have access to risk-free one period bonds.[6]

In the absence of any taxes, the value function of an individual with history θ^t reads as

[4] In macroeconomics, this class of preferences has shown to be very useful in matching business cycle moments (Greenwood et al., 1988; Mendoza and Yue, 2012).

[5] Sometimes we also write $h_t(\theta^t)$ in order to express the density of history θ^t, i.e. $f_t(\theta_t|\theta_{t-1})f_{t-1}(\theta_{t-1}|\theta_{t-2}).....f_1(\theta_1)$.

[6] We allow agents to borrow up to the natural debt limit (see Aiyagari (1994)). There are two differences to Aiyagari (1994): First, labor supply is endogenous; the minimal amount of future earnings in period s is $\sum_{t=s}^{T} y_t(\underline{\theta})$. Secondly, individuals can actually not borrow that much as repaying everything would yield a negative number in $U(\cdot)$ because of the disutility of labor. Therefore, the maximal amount of debt is $\sum_{t=s}^{T}\left(y_t(\underline{\theta}) - \Psi\left(\frac{y_t(\underline{\theta})}{\underline{\theta}}\right)\right)$. For CRRA preferences, e.g., this constraint would never be binding as the marginal utility of consumption would be ∞ in the worst case scenario.

$$V_t(\theta_t, a_t(\theta^{t-1})) = \max_{y_t, a_{t+1}} U\left(c_t - \Psi\left(\frac{y_t}{\theta_t}\right)\right) + \beta \int_{\theta_{t+1}} V_t(\tilde{\theta}_{t+1}, a_{t+1}) dF_{t+1}(\tilde{\theta}_{t+1}|\theta_t)$$
$$\text{subject to } c_t + a_{t+1} = y_t + (1+r)a_t$$
$$\text{and } a_{T+1} \geq 0 \tag{3.1}$$

where $a_t(\theta^{t-1})$ are period t assets of an individual with history θ^{t-1} and β is the discount factor. In the following, we assume $\beta(1+r)=1$.

3.2.2 The Social Planner's Problem

We are interested in the Pareto efficient set of nonlinear labor income tax schedule and linear capital income tax rates that only condition on current income. Thus, we are not solving for a second-best Pareto problem, where the government could condition policy instruments on all public information (typically the history of income and savings), but rather restrict the set of policy instruments in a Ramsey manner. However, our approach shares the feature of the Mirrlees approach that labor income taxes can be an arbitrarily nonlinear function of current income. One could call it a third-best Pareto problem, where third-best refers to the restriction on policy instruments. In the remainder of this chapter, solely the phrase Pareto optimal will be used.

We consider two scenarios. In the first, the government can condition labor income tax schedules $\mathcal{T}(\cdot)$ and capital tax rates τ on age t, so $\mathcal{T} = \{\mathcal{T}_t(\cdot)\}_{t=1,\ldots,T}$ and $\tau = \{\tau_t\}_{t=1,\cdots,T}$. In the second scenario, we study income tax functions which are independent of age.

The preferences of the social planner are described by the set of Pareto weights $\{\tilde{f}(\theta)\}_{\theta\in[\underline{\theta},\overline{\theta}]}$. The cumulative Pareto weights are defined by $\tilde{F}(\theta) = \int_{\underline{\theta}}^{\theta} \tilde{f}(\tilde{\theta})d\tilde{\theta}$. The set of weights are restricted such that $\tilde{F}(\overline{\theta}) = 1$. Different sets of Pareto weights refer to different points on the Pareto frontier. The set of weights where $\tilde{f}(\theta) = f(\theta)\ \forall\ \theta$, e.g., refers to the Utilitarian planner.[7]

Before writing down the problem of the planner, we refer to a special property that any equilibrium, with taxes as defined above, has:

Lemma 3.2.1 *The optimal labor supply of an individual, given taxes* $\mathcal{T} = \{\mathcal{T}_t(\cdot)\}_{t=1,\ldots,T}$ *and* $\tau = \{\tau_t\}_{t=1,\cdots,T}$, *is only a function of the current shock, i.e.* y_t *is only a function of* θ_t *and* $\mathcal{T}_t$.

[7] Similar as $H_t(\theta^t)$ and $h_t(\theta^t)$, we sometimes use $\tilde{H}_t(\theta^t)$ and $\tilde{h}_t(\theta^t)$ to express the Pareto weights for individuals with certain histories.

This lemma is a direct consequence of our preference assumption and will render the problems for age-independent and age-dependent taxes tractable.[8] In the remainder of this subsection, we always write $\mathcal{T}_t$ and τ_t, implying age-dependent taxes. For the age-independent tax problem we would only have to drop the subscript t. The age-dependent tax problem reads as:

$$\max_{\mathcal{T},\tau} \int_{\theta_1} V(\theta_1) d\tilde{F}(\theta_1) \tag{3.2}$$

with

$$V(\theta_1) = U\left(c_1(\theta_1) - \Psi\left(\frac{y_1(\theta_1)}{\theta_1}\right)\right) + \sum_{t=2}^{T} \beta^{t-1} \int_{\theta^t \in \mathcal{B}_2(\theta_1)} U\left(c_t(\theta^t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right) dH_t(\theta^t), \tag{3.3}$$

where $\mathcal{B}_2(\theta_1)$ is the set of all θ^t that contain θ_1 as its first element. The government has to balance the budget intertemporally

$$\sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\theta^t} \mathcal{T}_t(y_t(\theta_t)) dH_t(\theta^t) + \sum_{t=2}^{T} \frac{1}{(1+r)^{t-1}} \int_{\theta^{t-1}} \tau_t (1+r) a_t(\theta^{t-1}) dH_{t-1}(\theta^{t-1}) \geq \mathcal{R} \tag{3.4}$$

where $\mathcal{R}$ is some exogenous revenue requirement of the government. Finally, the government has to take into account individual behavior, i.e. $\forall t$ and $\forall \theta^t : \{c_t(\theta^t), y_t(\theta_t)\}$ solves

$$\begin{aligned} V_t(\theta_t, a_t(\theta^{t-1})) = \arg\max_{c_t, y_t} \; & U\left(c_t - \Psi\left(\frac{y_t}{\theta_t}\right)\right) + E_t\left[V_{t+1}(\theta^t, a_{t+1})\right] \\ & \text{where: } c_t + a_{t+1}(\theta^t) = y_t - \mathcal{T}_t(y_t) + (1+r)(1-\tau_t) a_t(\theta^{t-1}) \\ & \text{and: } a_{T+1} \geq 0. \end{aligned} \tag{3.5}$$

[8] With income effects, labor supply would also be a function of the assets. This would complicate the analysis heavily in two manners. First, there would be no obvious ordering of the income vector. Second, without income effects, the Hessian matrix of the individual choice problem (choosing labor supply and savings jointly) has an empty minor diagonal; taking into account optimal behavior of individuals is thus much simpler.

Constraint (3.5) makes the solution of the problem with Lagrangian methods non-trivial. In the following subsection, we argue that (3.5) can be replaced by a set of first-order conditions for a_t and y_t.
In the remainder of this chapter, we use the notions wealth, savings or capital for a_t interchangeably. Also note that the way we define τ_t, it is a stock tax not a flow tax. However, there is always a one to one mapping between such a stock tax and a tax on capital income, i.e. ra_t. Thus, there is no loss of generality in the way we defined τ_t. In the following, we use the notions capital taxes, wealth taxes and capital income taxes interchangeably. Further note that the way we define capital taxes implies that borrowing is subsidized at the same rate as saving is taxed.

3.2.3 First-Order Approach

The set of first-order conditions for the individual problem (3.5) are standard. For the labor supply decision, we have $\forall t$ and $\forall \theta_t$:

$$1 - \mathcal{T}_t'(y_t(\theta_t)) = \Psi'\left(\frac{y_t(\theta_t)}{\theta_t}\right)\frac{1}{\theta_t}. \tag{3.6}$$

For the intertemporal consumption decision we have $\forall t = 1, ..., T-1$ and $\forall \theta^t$:

$$U'\left(c_t(\theta^t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right) = \int_{\theta_{t+1}} U'\left(c_t(\theta^t, \theta_{t+1}) - \Psi\left(\frac{y_{t+1}(\theta_{t+1})}{\theta_{t+1}}\right)\right) dF_{t+1}(\theta_{t+1}|\theta_t). \tag{3.7}$$

with

$$c_t(\theta^t) = y_t(\theta_t) - \mathcal{T}_t(y_t(\theta_t)) + (1+r)(1-\tau_t)a_t(\theta^{t-1}) - a_{t+1}(\theta^t).$$

These conditions are only necessary and not sufficient for the agents' choices to be optimal. Due to the assumption about preferences, the second order conditions are of particularly simple form. The derivative of the first-order condition of labor supply with respect to consumption, i.e. the cross derivative of the value function, is zero. By symmetry of the Hessian, the same holds for the derivative of the Euler equation with respect to labor supply. Thus, the minor diagonal of the Hessian matrix contains only zeros. For (3.6) and (3.7) to represent a maximum, only the second derivatives of the value function with respect to labor supply and consumption have to be ≤ 0. For labor supply, a familiar argument from the standard Mirrlees model implies that this holds if and only if[9]

$$y_t'(\theta_t) \geq 0. \tag{3.8}$$

The second order condition for savings is always fulfilled due to concavity of the utility function. Hence, (3.6) and (3.7) represent a maximum whenever $y_t'(\theta_t) \geq 0$. As $y_t'(\theta_t) \geq 0$ even implies global concavity, (3.6) and (3.7) even represent a global maximum if $y_t'(\theta_t) \geq 0$ holds.

Lemma 3.2.2 *When choosing $\mathcal{T} = \{\mathcal{T}_t(\cdot)\}_{t=1,...,T}$ and $\tau = \{\tau_t\}_{t=1,\cdots,T}$ to maximize (3.2) subject to (3.4) and (3.5), the last constraint can be replaced by (3.6), (3.7) and (3.8).*

Incorporating (3.6) into a Lagrangian, however, is still problematic as it contains $\mathcal{T}'$, i.e. the derivative of the function with respect to which we want to maximize. To tackle this problem, we make use of the following derivative

$$\frac{\partial\left(y_t(\theta_t) - \mathcal{T}_t(y_t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)}{\partial\theta_t} = y_t'(\theta_t)(1 - \mathcal{T}_t'(y_t)) - \Psi'\left(\frac{y_t(\theta_t)}{\theta_t}\right)\left[\frac{y_t'(\theta_t)}{\theta_t} - \frac{y_t(\theta_t)}{\theta_t^2}\right].$$

Inserting (3.6) into this derivative yields:

$$\frac{\partial\left(y_t(\theta_t) - \mathcal{T}_t(y_t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)}{\partial\theta_t} = \Psi'\left(\frac{y_t(\theta_t)}{\theta_t}\right)\frac{y_t(\theta_t)}{\theta_t^2}. \tag{3.9}$$

Thus, (3.9) implies (3.6). As we show in the Appendix 3.A.1, (3.9) can easily be incorporated into the Lagrangian. Further, when solving for optimal policies, we do not incorporate the monotonicity constraint (3.8) into the Lagrangian as is standard practice in the optimal tax literature. In the numerical simulations we will ex-post check whether the condition is fulfilled or not. The Lagrangian and all first-order conditions for the age-dependent and age-independent case are stated in Appendices 3.A.1 and 3.B.1 respectively.

[9] See, e.g., Salanié (2003).

3.3 Pareto Optimal Taxes

We start by deriving and discussing optimal taxes in a two period model, so $T = 2$. The reason is that optimal taxes in the two periods case are much easier to derive and to interpret. Building on our results we then briefly discuss formulas for the general case.

3.3.1 The Two Period Model

3.3.1.1 Labor Taxes

We derive optimal tax formulas for our *dynamic problem* following an intuitive tax reform approach as in the *static* Mirrlees literature, pioneered by Piketty (1997) and further developed by Saez (2001). In Appendix 3.A.3.2 we document how to derive optimal labor income taxes by maximizing the Lagrangian-functional associated with Lemma 3.2.2 using calculus of variation.

We start with the age-dependent tax in period two. Consider an infinitesimal increase in the marginal tax rates $d\mathcal{T}_2'$ within an income interval of infinitesimal length dy_2 around some income level $y_2(\theta_2)$. First, such a change in marginal tax rates triggers a *mechanical* increase in tax revenue by raising the tax obligation for all individuals at age two with higher income in that period by $d\mathcal{T}_2' dy_2$. Since the mass of these individuals is $\int_{\theta_1}(1 - F_2(\theta_2|\theta_1))dF(\theta_1)$, the overall effect on the government budget in present value terms reads as

$$\frac{1}{1+r} d\mathcal{T}_2' dy_2 \int_{\theta_1}(1 - F_2(\theta_2|\theta_1))dF(\theta_1). \tag{3.10}$$

A benevolent government also takes the utility loss of these individuals into account. Denote by λ the marginal value of public funds, i.e. the Lagrangian-multiplier on the government budget constraint. The utility loss on welfare, denoted in (period one) monetary units, then reads as

$$d\mathcal{T}_2' dy_2 \int_{\theta_1}\int_{\theta_2}^{\overline{\theta}} \beta \frac{U'(\theta_1, \tilde{\theta}_2)}{\lambda} dF_2(\tilde{\theta}_2|\theta_1) d\tilde{F}(\theta_1). \tag{3.11}$$

The tax increase also triggers behavioral responses. First, it affects the labor supply margin. Individuals with income $y_2(\theta_2)$ work less: $\frac{\partial y_2(\theta_2)}{\partial \mathcal{T}'} d\mathcal{T}'$. For each dollar they earn less, the government loses the fraction $\mathcal{T}'$ as this is the money that is taxed away marginally. The effect on the government budget in period two thus reads as[10]

$$-\frac{1}{1+r}\mathcal{T}_2'\frac{\partial y_2(\theta_2)}{\partial(1-\mathcal{T}_2')}d\mathcal{T}_2' \times \mathcal{H}_2(y_2(\theta_2))dy_2. \tag{3.12}$$

where $\mathcal{H}_2(\cdot)$ denotes the probability density function of second period income. It can be rewritten as $\frac{\partial\theta_2}{\partial y_2(\theta_2)}\int_{\theta_1} f_2(\theta_2|\tilde{\theta}_1)dF(\theta_1)$. We also use $\frac{\partial y_2(\theta_2)}{\partial\theta_2} = \varepsilon_{y_2,\theta_2}\frac{y_2(\theta_2)}{\theta_2}$. Using $\varepsilon_{y_2,\theta_2} = 1 + \varepsilon_{y_2,1-\mathcal{T}'}$ in addition, (3.12) can be rewritten as:

$$-\frac{1}{1+r}\frac{\mathcal{T}_2'}{1-\mathcal{T}_2'}\varepsilon_{y_2,1-\mathcal{T}'}d\mathcal{T}_2' \times \frac{\theta_2}{(1+\varepsilon_{y_2,1-\mathcal{T}'})}\int_{\theta_1} f_2(\theta_2|\tilde{\theta}_1)dF(\theta_1)dy_2. \tag{3.13}$$

Finally, the change in tax liability makes individuals adjust their savings decisions.[11] Intuitively, an increase in taxes at the old age induces the young to save more. If capital taxes τ are different from zero, this has the following first-order effect on the government's budget in present value terms:

$$\tau\int_{\theta_1}\frac{\partial a_2(\theta_1)}{\partial\mathcal{T}_2'(y_2(\theta_2))}dF(\theta_1). \tag{3.14}$$

At the optimum, we have to have (3.10)+(3.13)+(3.14) = 0, which gives the following optimal tax formula.

Proposition 3.3.1 *Pareto optimal marginal labor income tax rates for the old satisfy:*

$$\frac{\mathcal{T}_2'(y_2(\theta_2))}{1-\mathcal{T}_2'(y_2(\theta_2))} = \left(1+\frac{1}{\varepsilon_{y_2,1-\mathcal{T}'}(\theta_2)}\right)\frac{1}{\theta_2\int_{\theta_1} f_2(\theta_2|\theta_1)dF(\theta_1)} \times \left[\mathcal{M}_2(\theta_2)+\mathcal{S}_2(\theta_2)\right],$$

where the total "mechanical effect" equals

$$\mathcal{M}_2(\theta_2) = \int_{\theta_1}\int_{\theta_2}^{\overline{\theta}}\left(1-\frac{U'(\theta_1,\tilde{\theta}_2)}{\lambda}\frac{\tilde{f}(\theta_1)}{f(\theta_1)}\right)dF_2(\tilde{\theta}_2|\theta_1)d\tilde{F}(\theta_1)$$

and the "savings effect" equals

$$\mathcal{S}_2(\theta_2) = \tau(1+r)\int_{\theta_1}\frac{\partial a_2(\theta_1)}{\partial\mathcal{T}_2'(y_2(\theta_2))}dF(\theta_1).$$

The optimal tax formula follows the same logic as in the static literature. It is decreasing in the elasticity of labor supply and trades-off how an additional dollar

[10] This behavioral adjustment has no first-order effect on the utility of the individuals by the envelope theorem.

[11] Again, this behavioral adjustment has no first-order effect on the utility of the individuals by the envelope theorem.

in the hand of the government is weighted against the welfare loss of individuals who face higher taxes which is captured by $\mathcal{M}_2(\theta_2)$. The novel behavioral response in a dynamic model is the savings effect $\mathcal{S}_2(\theta_2)$. Notice that $\mathcal{S}_2(\theta_2)$ has the same sign as τ, because $\frac{\partial a_2(\theta_1)}{\partial \mathcal{T}_2'(y_2(\theta_2))} > 0 \ \forall \ \theta_1$. We argue in the next section and confirm in our numerical simulations that a positive $\tau > 0$ is likely in a realistic life cycle environment, so that $\mathcal{S}_2(\theta_2)$ is likely to be positive as well. The additional force in our life cycle model thus leads to higher labor income taxes for the old.

Golosov et al. (2013b) also show that this savings effect leads to larger welfare gains from tax increases for older individuals as compared to welfare gains from tax increases for younger individuals and as a consequence to higher (lower) optimal taxes for older (younger) individuals as compared to the static case.

The following proposition states the optimal marginal tax rates for the young.

Proposition 3.3.2 *Pareto optimal marginal labor income tax rates for the young satisfy:*

$$\frac{\mathcal{T}_1'(y_1(\theta_1))}{1 - \mathcal{T}_1'(y_1(\theta_1))} = \left(1 + \frac{1}{\varepsilon_{y_1, 1-\mathcal{T}'}(\theta_1)}\right) \frac{1}{\theta_1 f(\theta_1)} \times \left[\mathcal{M}_1(\theta_1) + \mathcal{S}_1(\theta_1)\right],$$

where the total mechanical effect equals

$$\mathcal{M}_1(\theta_1) = \int_{\theta_1}^{\overline{\theta}} \left(1 - \frac{U'(\tilde{\theta}_1)}{\lambda} \frac{\tilde{f}(\tilde{\theta}_1)}{f(\tilde{\theta}_1)}\right) d\tilde{F}(\tilde{\theta}_1)$$

and the savings response equals

$$\mathcal{S}_1(\theta_1) = \tau \int_{\tilde{\theta}_1} \frac{\partial a_2(\tilde{\theta}_1)}{\partial \mathcal{T}_1'(y_1(\theta_1))} dF(\tilde{\theta}_1).$$

This result for period one is even closer to the results from the static model without income effects (Diamond, 1998) because in period one individuals only differ in their current skill. In case of zero capital taxation $\tau = 0$, the formula is indeed the same as in Diamond (1998). As capital taxes are generally non-zero as we argue in the next subsection, the formula is adjusted by a savings effect $\mathcal{S}_1(\theta_1)$ similar as in Proposition 3.3.1. The savings effect is of opposite sign on the young as compared to the old. Because of $\frac{\partial a_2(\tilde{\theta}_1)}{\partial \mathcal{T}_1'(y_1(\theta_1))} < 0 \ \forall \ \tilde{\theta}_1$, $\mathcal{S}_1(\theta_1)$ is negative for a positive capital tax. Compared to the static arguments, labor income taxes for the young should be decreased since this increases savings of the young which in turn has a positive fiscal effect on the government budget. Taken together, the presence of positive capital taxation tends to increase labor taxes over the life cycle.

How do age-independent taxes change in comparison to the static case? In fact, if labor taxes are restricted to be independent of age, a mixture of the formulas in Propositions 3.3.1 and 3.3.2 applies as the following proposition summarizes.[12]

Proposition 3.3.3 *Pareto optimal marginal labor income tax rates for the young satisfy:*

$$\frac{\mathcal{T}'(y(\theta))}{1-\mathcal{T}'(y(\theta))} = \left(1+\frac{1}{\varepsilon_{y,1-\mathcal{T}'}(\theta)}\right)\frac{1}{\theta f(\theta)\int_{\theta_1} f_2(\theta|\theta_1)dF(\theta_1)} \times \left[\sum_{i=1}^{2}\mathcal{M}_i(\theta)+\mathcal{S}_i(\theta)\right].$$

Intuitively, the underlying trade-offs are the same for the government. The difference is that the mentioned effects have to be considered for period one and period two. The effect of labor taxes on savings will in general be ambiguous – as our discussion on age-dependent labor taxes highlighted, higher taxes on the young reduce savings by an income effect but higher taxes on the old also increase savings as individuals anticipate higher taxes later in life. Determining the sign of $\mathcal{S}_1(\theta)+\mathcal{S}_2(\theta)$ is thus not possible because it is unclear which effect dominates in general. We now turn to the optimal capital tax rate.

3.3.1.2 Capital Taxes

In the two period model there is no difference between age-dependent and age-independent capital taxation because young agents start with zero wealth and capital is only taxed in period two. We will again use a perturbation argument, looking at a small change in the capital tax rate $d\tau$.[13]
This will increase government's revenue in present value terms by

$$\int_{\theta_1} a_2(\theta_1)dF(\theta_1) \tag{3.15}$$

and decrease utility of individuals, which is valued (in terms of public funds) by

$$-\int_{\theta_1} a_2(\theta_1)\beta(1+r)\int_{\theta_2}^{\overline{\theta}}\frac{U'(\theta_1,\tilde{\theta}_2)}{\lambda}dF_2(\tilde{\theta}_2|\theta_1)d\tilde{F}(\theta_1). \tag{3.16}$$

It also discourages savings and thereby decreases tax revenue from the capital income tax. This effect on public funds (in present value terms) is given by:

$$\tau\int_{\theta_1}\frac{\partial a_2(\theta_1)}{\partial\tau}dF(\theta_1). \tag{3.17}$$

[12] For a derivation of this formula with Lagrangian methods, see Appendix 3.B.3.2.
[13] In Appendix 3.A.4.2 a derivation of the formula with Lagrangian methods is provided.

Note that $\frac{\partial a_2(\theta_1)}{\partial \tau}$ is in general of ambiguous sign. The substitution effect calls for lower savings due to an increase in τ. The income effect is of opposite sign: the higher τ, the less after tax wealth one has tomorrow (for a given amount of savings), which makes one save more. For expositional reasons, we now rewrite (3.17) as

$$-\tau \int_{\theta_1} \zeta_{a_2,1-\tau}(\theta_1) \frac{a_2(\theta_1)}{1-\tau} dF(\theta_1), \tag{3.18}$$

where $\zeta_{a_2,1-\tau}(\theta_1)$ is the elasticity of savings with respect to the net of tax rate $1-\tau$. If $\frac{\partial a_2(\theta_1)}{\partial(1-\tau)} > (<) 0$, $\zeta_{a_2,1-\tau}(\theta_1)$ is positive if $a_2(\theta_1) > (<)\ 0$.
The absence of income effects on labor implies that labor supply does not change in response to a capital tax increase. Optimality of τ then requires (3.15) + (3.16) + (3.18) = 0, which yields the following result.

Proposition 3.3.4 *The optimal linear capital tax rate τ satisfies*

$$\frac{\tau}{1-\tau} = \frac{\int_{\theta_1} a_2(\theta_1)\left[f(\theta_1) - \int_{\theta_2} \frac{U'(\theta_1,\tilde{\theta}_2)\tilde{f}_1(\theta_1)}{\lambda} dF_2(\tilde{\theta}_2|\theta_1)\right] d\theta_1}{\int_{\theta_1} a_2(\theta_1)\zeta_{a_2,1-\tau}(\theta_1) dF(\theta_1)}. \tag{3.19}$$

The optimal taxation of capital follows a very simple and intuitive equity-efficiency trade-off as is standard in the public finance literature. Capturing efficiency arguments, it is decreasing in the weighted elasticity of savings with respect to the net of tax rate $1-\tau$.[14] Capturing equity arguments, the tax rate is higher, the more the government values redistribution from savers to non-savers. The numerator is similar to the mechanical effects as defined in Propositions 3.3.1 and 3.3.2. Without the term $a_2(\theta_1)$ the numerator in (3.19) would say by how much welfare (in terms of public funds) increases if one dollar is redistributed in period two from all individuals to the government. With the term $a_2(\theta_1)$, however, this effect is weighted for each θ_1 by the amount of savings. For reasonable conditional distribution functions $F(\theta_2|\theta_1)$, one can expect $a_2'(\theta_1) > 0$ implying that individuals with higher innate ability θ_1 save more and therefore are affected stronger by the increase in the capital income tax rate τ.[15] If in addition Pareto weights are such that the government wants to redistribute from high innate types to low innate types, the numerator of (3.19) is positive, yielding positive capital taxes. For a given set of Pareto weights, capital taxes are then increasing in wealth inequality. Our numerical results in Section 3.4 will confirm this intuition.

[14] In the following argumentation, we are only considering the case where the denominator of (3.19) is positive, i.e. we assume that on average the substitution effect dominates the income effect. In all our numerical results, this is actually the case.

[15] In our numerical simulations that are based on a realistic calibration, this is always the case.

Note that our result for the distortion of the savings margin are different from other recent arguments in favor of capital taxation. In the NDPF-literature, savings are distorted in order to relax incentive constraints via an income effect on labor supply. They are not used to redistribute, however, see Kocherlakota (2005). The reason is that in these papers, capital taxes are superfluous to redistribute since history dependent labor taxes can condition on the whole history of shocks θ^t.[16] As we show, with more realistic policy instruments like standard income taxes or age-dependent income taxes, capital taxes can provide additional redistribution.
In a recent paper, Jacobs and Schindler (2012) show that in a two-period model with linear labor taxes, a similar role for the capital tax as in the NDPF-literature arises as capital taxes have the positive effect of boosting labor supply in the second period. In their framework, a positive capital tax also provides insurance against idiosyncratic risk.
Using a dynamic tax reform approach, Golosov et al. (2013b) look at the welfare effects of an increase of a linear capital tax rate and obtain a formula which is very similar to (3.19).

3.3.2 The T-Period Case

We now briefly present the optimal tax formulas in the general T-period case. In the main text, we only refer to age-dependent taxes. The respective results for age-independent taxes are discussed in Appendices 3.B.3.1 and 3.B.4.

3.3.2.1 Labor Taxes

We first present the formula for a marginal tax rate on labor income if age-dependent taxes are available for the government.

Proposition 3.3.5 *Age-dependent Pareto optimal marginal labor income tax rates in period t satisfy*

$$\frac{\mathcal{T}_t'(y_t(\theta_t))}{1-\mathcal{T}_t'(y_t(\theta_t))} = \left(1+\frac{1}{\varepsilon(\theta_t)}\right)\frac{1}{\theta_t\int_{\theta^{t-1}} f_t(\theta_t|\theta_{t-1})dH(\theta^{t-1})} \times \left[\mathcal{M}_t(\theta_t)+\mathcal{S}_t(\theta_t)\right],$$

where the total mechanical effect equals

$$\begin{aligned}\mathcal{M}_t(\theta_t) = &\int_{\theta^{t-1}}\int_{\theta_t}^{\overline{\theta}} dF_t(\tilde{\theta}_t|\theta_{t-1})dH_{t-1}(\theta^{t-1})\\ &-\int_{\theta^{t-1}}\int_{\theta_t}^{\overline{\theta}}\frac{U'(R_t(\theta^{t-1},\tilde{\theta}_t))}{\lambda}dF_t(\tilde{\theta}_t|\theta_{t-1})d\tilde{H}_{t-1}(\theta^{t-1})\end{aligned}$$

[16] Indeed, in his model, the expected tax payment of each agent is always zero.

and the savings response equals

$$\mathcal{S}_t(\theta_t) = -\int_{\theta^{t-1}} \int_{\theta_t}^{\overline{\theta}} \mu_t(\theta^{t-1}, \theta_t) U''(R_t(\theta^{t-1}, \theta_t)) d\theta_t d\theta^{t-1}$$
$$+ (1-\tau_t) \int_{\theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\theta_t}^{\overline{\theta}} U''(R_t(\theta^{t-1}, \theta_t)) dF_t(\theta_t|\theta^{t-1}) d\theta^{t-1},$$

where $\mu_t(\theta^t)$ *is the multiplier on the Euler equation of an agent with history* θ^t*.*

Proof: See Appendix 3.A.3.1 ■

The main insight is that the same basic forces underlying optimal taxes in the two period model also determine tax rates in the T period case. Higher mechanical revenue effects tend to increase tax rates, the labor supply elasticity tends to decrease tax rates. Labor taxes will also influence savings behavior as captured by $\mathcal{S}_t(\theta_t)$. In the first period $t = 1$ this leads to lower tax rates whereas in the last period $t = T$ it leads to higher tax rates. In intermediate periods the effects is not as clear. In general, the difficulty is that savings in *all* periods change once the labor income tax is changed in one period. For future research, we are planning to express the savings responses $\mathcal{S}_t(\theta_t)$ as a direct function of the behavioral responses, thus as a function of $\frac{\partial a_i}{\partial T_j'(y_j(\theta_j))}$ for all $i, j = 1, \ldots, T$. Finally, note that if capital taxation is unavailable to the government, it follows that $\mathcal{S}_t(\theta_t) = 0$ in all periods because all multipliers $\mu_t(\theta^t) = 0$ for each history θ^t. Then, our formula again collapses to the static version as in Diamond (1998).

3.3.2.2 Capital Taxes

A small increase in the age-dependent capital tax at age t again triggers mechanical revenue effects for the government and reduces the welfare of wealth holders at age t. As in the case of the age-dependent labor tax, the behavioral response of savings will not be limited to wealth holdings a_t. Instead the whole sequence of capital holdings $a_2, \ldots, a_T$ will respond. The following formula can be derived:

Proposition 3.3.6 *The optimal linear capital tax τ_t is implicitly defined by*

$$(1-\tau_t) = \frac{1}{\beta(1+r)^2\mu_{t-1}(\theta^{t-1})\int_{\theta_t} U''(R_t(\theta^{t-1},\theta_t))a_t(\theta^{t-1})dF_t(\theta_t|\theta_{t-1})d\theta^{t-1}}$$
$$\Bigg[-\frac{\lambda}{(1+r)^{t-2}}\int_{\theta^{t-1}} a_t(\theta^{t-1})dH_{t-1}(\theta^{t-1})$$
$$+\beta^{t-1}(1+r)\int_{\theta^{t-1}} a_t(\theta^{t-1})\int_{\theta_t} U'(R_t(\theta^{t-1},\theta_t))dF_t(\theta_t|\theta_{t-1})d\tilde{H}_{t-1}(\theta^{t-1})$$
$$+(1+r)\int_{\theta^{t-1}} a_t(\theta^{t-1})\int_{\theta_t}\mu_t(\theta^{t-1},\theta_t)U''(R_t(\theta^{t-1},\theta_t))d\theta_t d\theta^{t-1}$$
$$-\int_{\theta^{t-1}}\mu_{t-1}(\theta^{t-1})\int_{\theta_t} U'(R_t(\theta^{t-1},\theta_t))dF_t(\theta_t|\theta_{t-1})d\theta^{t-1}\Bigg],$$

where $\mu_t(\theta^t)$ is the multiplier on the Euler equation of an agent with history θ^t.

Proof: See Appendix 3.A.4.1 ■

The formula is significantly more complex than the formula in Proposition 3.3.4. Lines two and three again capture the mechanical redistributive effect. The other lines capture all savings responses over the life cycle: increasing the capital income tax in a certain period induces changes in savings behavior over the whole life cycle.

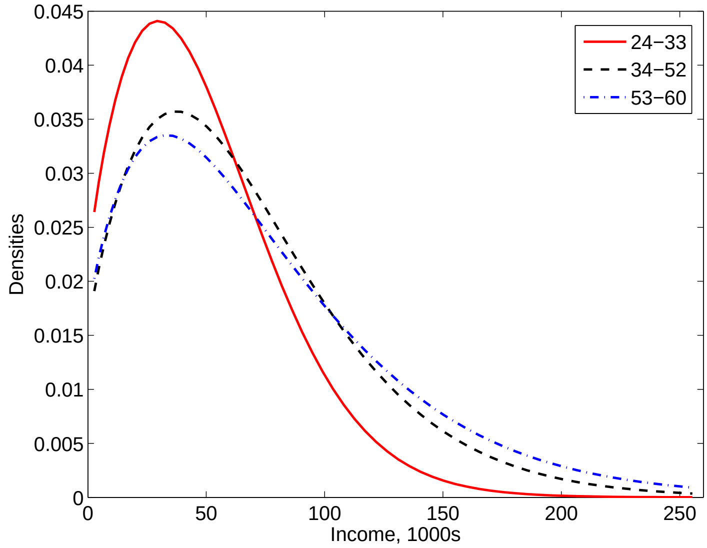
3.4 A Numerical Exploration

We now simulate optimal policies for a $T = 3$ period economy. The formulas derived in Section 3.3.2 will be the basis for our numerical simulation. In Section 3.4.1 we explain our parameterization. In the following subsections, we present the results for optimal taxes.

3.4.1 Inequality over the Life Cycle and Parameters

There is large literature on the estimation of earnings dynamics over the life cycle – see Meghir and Pistaferri (2011) and Jappelli and Pistaferri (2010) for recent surveys. For the parameterization of our model, we use the recent empirical approach taken by Karahan and Ozkan (2013). In their analysis, they estimate the persistence of permanent shocks as well as the variance of permanent and transitory income shocks for US workers. Innovatively and in contrast to most previous work in this strand of the literature, they allow these parameters to be age-dependent. They find two structural breaks in how the key parameters change over the life cycle,

Figure 3.4.1: Income Distribution for Each Age Group



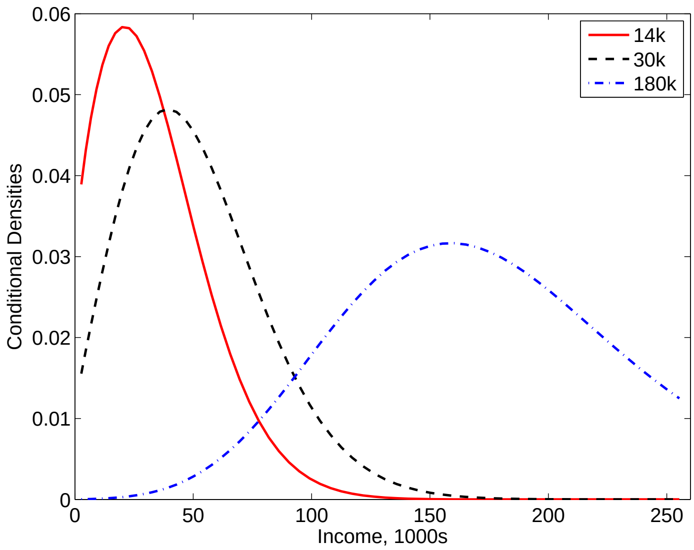
yielding three age groups, in which income dynamics are governed by the same risk parameters. The age groups are 24-33, 34-52 and 53-60 years.

We base our parameterization on their results. Thus, we interpret our period one as age 24-33 and periods two and three as 34-52 and 53-60 respectively. Given the estimates of Karahan and Ozkan (2013) for the evolution of income over the life cycle[17], we simulate millions of labor income histories. After having simulated those earnings histories, i.e. an economy with millions of individuals, we calibrate the cross sectional income distributions for each age group and respective transition probabilities. Figure 3.4.1 shows the three cross-sectional income distributions for each age group. It becomes clear how inequality evolves over the life cycle. In the older age groups there are more people with top incomes than in the young age group. Figure 3.4.2 shows three conditional income distributions for the middle age-group, conditioning on earnings of $14,000, $30,000 and $180,000 in the previous period respectively. The role of both persistence and risk for earnings becomes clear from this picture. As a last step to complete the calibration of the model, we calibrate all conditional skill distributions from their income counterparts, as suggested by Saez (2001).[18] We describe all these steps in more detail in Appendix 3.C.

[17] We gratefully acknowledge that they shared some estimates with us that are not in the main body of their paper.

[18] We back out the skill from the first-order condition of individual labor supply given a rough approximation of the current US-tax system. We do it analogously as in the first chapter of this dissertation, see Section 1.5.1 and Appendix 1.D.1.

Figure 3.4.2: Conditional Income Distributions, Middle Aged Workers



We assume the utility function to be of the form

$$U = \frac{\left(C - \frac{L^{1+\varepsilon}}{1+\varepsilon}\right)^{1-\gamma}}{1-\gamma}. \tag{3.20}$$

For the benchmark, we set $\varepsilon = 3$, implying a labor supply elasticity of 0.33 (Chetty, 2012) and set $\gamma = 1.5$ (Chetty, 2006).

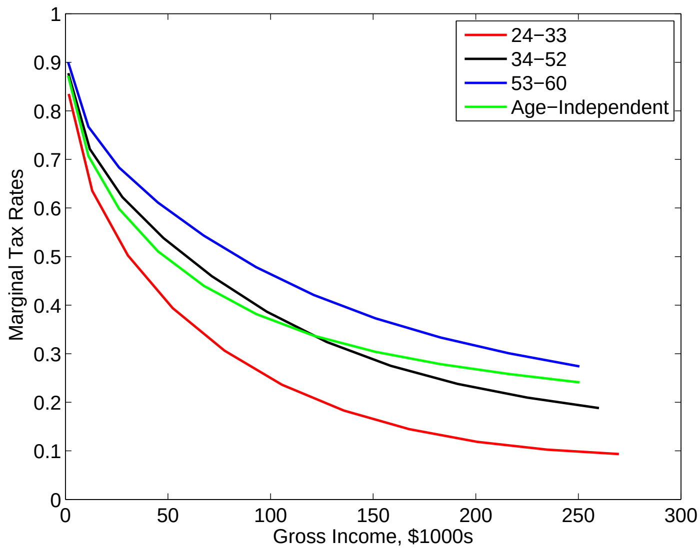
The yearly interest rate is 4% and we set $\beta = \frac{1}{1+r}$. In this version of our numerical simulations, we look at three periods with the same length; i.e. we do not take into account that the three considered age intervals in the calibration are of different length. Given that the three periods are of the same length, each period consists of roughly 12 years implying an interest rate of approximately 30%.[19]

3.4.2 Results in the Benchmark Case

We present results for a Utilitarian social welfare function. We calculate optimal policies for four cases: age-dependent taxes are available or not and wealth/capital income taxes are available or not.

[19]Note that we did not simply take an overall interest rate of $1.04^{12} - 1$ as this would imply that all the savings get compounded over the whole 12 years. Instead, 30% is the average interest rate one gets for the overall savings if in each period the same amount of money is added to the stock of savings; whereas money that is added to the savings stock in the first period is indeed compounded with $1.04^{12} - 1$, money that is added in the last period is only compounded with 4%.

Figure 3.4.3: Optimal Utilitarian Marginal Tax Rates

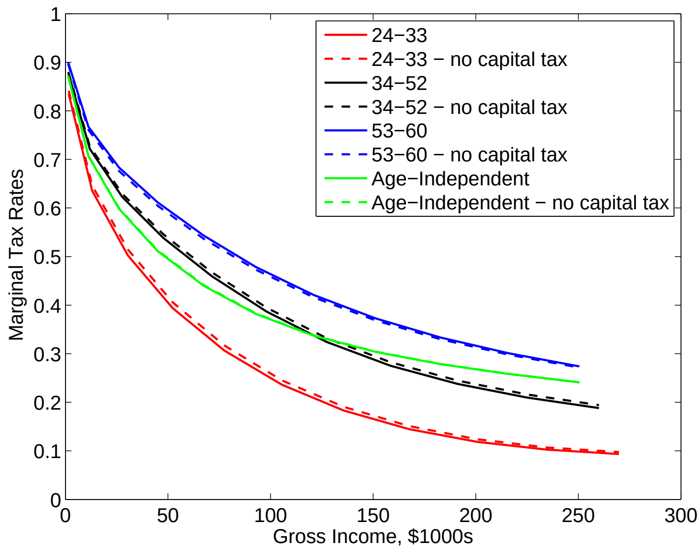


In Figure 3.4.3 we illustrate optimal marginal labor income tax rates, both age-dependent and age-independent, for the case when wealth taxation is available to the government. First, all marginal tax rates are decreasing over the income distribution, reflecting that the income distributions have a log-normal shape. This marginal tax rate regressivity is well-understood from the static literature on optimal income taxation (Diamond, 1998).[20] Second, labor income taxes are increasing over the life cycle. The intuition is that labor income inequality is much higher at later points in the life cycle. This leads to bigger mechanical revenue effects for the government when raising tax rates. Third, in the case in which the government is restricted to set age-independent taxes, optimal marginal tax rates are roughly an average of the age-dependent taxes.

Concerning the optimal taxation of capital, we obtain a wealth tax of 3.11% in period two and of 7.18% in period three. These numbers are significant and one can see that they are increasing over the life cycle. The reason is that wealth concentration is increasing. The optimal age-independent wealth tax is 2.84%. The age-independent tax is not an average of its age-dependent counterparts in this case. The reason is that individuals' savings behavior is different in the case with age-dependent taxes. The mean and the variance of the wealth distribution tend to be higher with age-dependent labor taxes, as younger individuals anticipate higher labor tax rates in

[20]Notice that our income distributions have no Pareto tails. From figure 3.4.1 it becomes obvious that tails tend to become much thicker over the life cycle, but never thick enough to resemble a Pareto tail as in Saez (2001).

Figure 3.4.4: Optimal Marginal Tax Rates: Savings Effect



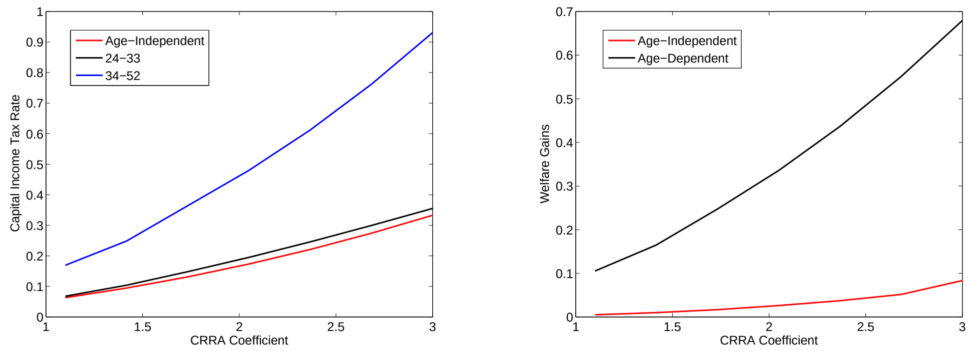
the future. Thus, due to higher wealth inequality, capital income taxes are more effective as a redistribution tool in the age-dependent case.

We also rephrase our results in terms of a yearly capital income tax. We therefore make the assumption that the amount of savings that are added to one's stock of savings are constant during the 12 years of one period and then calculate the yearly capital income tax rate that would yield the same tax revenue in present value terms as the wealth tax. We obtain a yearly capital income tax of 11.41% for the young and 27.10% for the old individuals. The optimal age-independent capital income tax rate is 10.38%.

Interestingly, varying the yearly interest rate has only very small effects on the optimal wealth tax as long as we adjust the discount factor such that $\beta(1 + r) = 1$. This implies in turn that capital income taxes are higher (lower) if interest rates are lower (higher).

An interesting question is to what extent the presence of capital taxes influences the optimal labor taxes. We therefore illustrate in Figure 3.4.4 not only the optimal Utilitarian taxes in the presence of a capital tax but also for the case where we constrained capital taxes to be zero. Age-independent taxes are almost equivalent in both cases; the solid green line and the dotted green line are almost indistinguishable. For age-dependent taxes, the effect is small but visible. The presence of positive capital taxes leads to higher labor income taxes in period three and lower labor income taxes in periods one and two. This is in line with our interpretations of the theoretical formulas in Section 3.3.

Figure 3.4.5: Capital Income Taxes



(a) Risk-Aversion and Capital Taxes (b) Welfare Gains of Capital Taxation

3.4.3 Welfare Gains and Comparative Statics

We next test the sensitivity of optimal policies to γ, which controls both risk-aversion and the intertemporal elasticity of substitution (IES). Figure 3.4.5(a) illustrates that the value of the capital income tax rate is quite sensitive to the value of γ. The optimal age-independent rate gets close to 40% for values of γ around three. The same is true for age-dependent capital tax rates for the young. For the old, however, they are almost 100%.

Higher risk-aversion tends to increase capital taxes, as the value placed on redistribution and insurance increases. Simultaneously, the IES decreases, decreasing the elasticity of savings with respect to the tax rate. Both effects explain the upward sloping profile.

Figure 3.4.5(b) shows the consumption equivalent welfare gains in percentage points of being able to tax wealth. These gains are also increasing in risk-aversion and are higher in the age-dependent case, underscoring the complementarity of being able to condition both labor and wealth taxes on age. The gains range from 0.11% to 0.68% points as γ goes from 1.1 to 3 in the age-dependent case and from 0.001% to 0.08% points in the age-independent case.

3.5 Conclusion

We have developed a formal framework to study Pareto optimal nonlinear taxation of annual labor income as well as linear taxation of capital in a framework with heterogeneous agents whose skills evolve stochastically over time. This method can be used to study age-dependent and age-independent taxes. By focusing on preferences without income effects on labor supply, we developed a first-order approach to make this problem tractable also for a continuous type space.

In this dynamic environment where inequality evolves over the life cycle, we derive a novel and simple formula for optimal capital income taxes. It follows a standard public finance intuition, trading redistributive benefits versus efficiency costs resulting from behavioral responses. In our realistically calibrated numerical simulations capital income taxation plays an important role as a redistribution device. Both, savings and labor income taxes tend to increase over the lifecycle, as inequality in both dimensions increases.

Appendix 3.A Age-Dependent Taxes

3.A.1 The Lagrangian and the First-Order Conditions

Define net income by $M_t(\theta_t) = y_t(\theta_t) - \mathcal{T}_t(y_t(\theta_t))$. Then, the Lagrangian can be written as

$$\begin{aligned}
\mathcal{L} = \int_{\theta_1} \Bigg(\sum_{t=1}^{T} \beta^{t-1} \int_{\theta^t} U \Bigg(& M_t(\theta_t) - a_{t+1}(\theta^t) \\
& + (1-\tau_t)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\Bigg) dH(\theta^t) \Bigg) d\tilde{F}_1(\theta_1) \\
& + \lambda \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\theta^{t-1}} \int_{\theta_t} y_t(\theta_t) - M(\theta_t) + \tau_t(1+r)a_t(\theta^{t-1}) dF(\theta_t|\theta_{t-1}) dH(\theta^{t-1}) \\
& + \sum_{t=1}^{T-1} \int_{\theta^t} \mu_t(\theta^t) \Bigg[U'\left(M_t(\theta_t) - a_{t+1}(\theta^t) + (1-\tau_t)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right) \\
& - \beta(1+r)(1-\tau_{t+1}) \int_{\theta_{t+1}} U' \Bigg(M_{t+1}(\theta_{t+1}) - a_{t+2}(\theta^t, \theta_{t+1}) \\
& + (1-\tau_{t+1})(1+r)a_{t+1}(\theta^t) - \Psi\left(\frac{y_{t+1}(\theta_{t+1})}{\theta_{t+1}}\right)\Bigg) dF(\theta_{t+1}|\theta^t) \Bigg] d\theta^t \\
& + \sum_{t=1}^{T} \int_{\theta_t} \eta_t(\theta_t) \frac{\partial \left(M_t(\theta_t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)}{\partial \theta_t} d\theta_t - \sum_{t=1}^{T} \int_{\theta_t} \eta_t(\theta_t) \Psi'\left(\frac{y_t(\theta_t)}{\theta_t}\right) \frac{y_t(\theta_t)}{\theta_t^2} d\theta_t.
\end{aligned}$$

Partially integrating $\sum_{t=1}^{T} \int_{\theta_t} \eta_t(\theta_t) \frac{\partial \left(M_t(\theta_t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right)}{\partial \theta_t} d\theta_t$ yields

$$\eta_t(\overline{\theta}) \left(M_t(\overline{\theta}) - \Psi\left(\frac{y_t(\overline{\theta})}{\overline{\theta}}\right)\right) - \eta_t(\underline{\theta}) \left(M_t(\underline{\theta}) - \Psi\left(\frac{y_t(\underline{\theta})}{\underline{\theta}}\right)\right) - \int_{\theta_t} \eta_t'(\theta_t) \left(M_t(\theta_t) - \Psi\left(\frac{y_t(\theta_t)}{\theta_t}\right)\right) d\theta_t.$$

The derivatives with respect to the endpoint conditions yield $\forall t: \ \eta_t(\overline{\theta}) = \eta_t(\underline{\theta}) = 0$. The first-order conditions read as

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial M_s(\theta_s)} = & - \frac{\lambda}{(1+r)^{s-1}} \int_{\theta^{s-1}} f_s(\theta_s|\theta_{s-1}) dH(\theta^{s-1}) \\
& + \beta^{s-1} \int_{\theta^{s-1}} U'(R_s(\theta^{s-1}, \theta_s)) f_s(\theta_s|\theta_{s-1}) d\tilde{H}(\theta^{s-1}) \\
& + \int_{\theta^{s-1}} \mu_s(\theta^{s-1}, \theta_s) U''(R_s(\theta^{s-1}, \theta_s)) d\theta^{s-1} \\
& - \beta(1+r)(1-\tau_s) \int_{\theta^{s-1}} \mu_{s-1}(\theta^{s-1}) U''(R_s(\theta^{s-1}, \theta_s)) f_s(\theta_s|\theta^{s-1}) d\theta^{s-1} \\
& - \eta_s'(\theta_s) = 0 \qquad (3.21)
\end{aligned}$$

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial y_s(\theta_s)} =& \frac{\lambda}{(1+r)^{s-1}}\int_{\theta^{s-1}} f_s(\theta_s|\theta_{s-1})dH(\theta^{s-1})\\ &-\beta^{s-1}\int_{\theta^{s-1}} U'(R_s(\theta^{s-1},\theta_s))\Psi'\left(\frac{y_s(\theta_s)}{\theta_s}\right)\frac{1}{\theta_s}f_s(\theta_s|\theta_{s-1})d\tilde{H}(\theta^{s-1})\\ &+\int_{\theta^{s-1}}\mu_s(\theta^{s-1},\theta_s)U''(R_s(\theta^{s-1},\theta_s))\Psi'\left(\frac{y_s(\theta_s)}{\theta_s}\right)\frac{1}{\theta_s}d\theta^{s-1}\\ &-\beta(1+r)(1-\tau_s)\int_{\theta^{s-1}}\mu_{s-1}(\theta^{s-1})U''(R_s(\theta^{s-1},\theta_s))\Psi'\left(\frac{y_s(\theta_s)}{\theta_s}\right)\frac{1}{\theta_s}f_s(\theta_s|\theta^{s-1})d\theta^{s-1}\\ &-\eta_s'(\theta_s)\Psi'\left(\frac{y_s(\theta_s)}{\theta_s}\right)\frac{1}{\theta_s}-\eta_s(\theta_s)\left(\Psi'\left(\frac{y_s(\theta_s)}{\theta_s}\right)\frac{1}{\theta_s^2}+\Psi''\left(\frac{y_s(\theta_s)}{\theta_s}\right)\frac{y_s(\theta_s)}{\theta_s^2}\right)=0\end{aligned} \tag{3.22}$$

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial a_{s+1}(\theta^s)} =& \frac{\lambda}{(1+r)^{s-1}}\tau_{s+1}h(\theta^{s-1})-\mu_s(\theta^s)U''\left(R_s(\theta^s)\right)\\ &-(1-\tau_{s+1})^2\beta(1+r)^2\mu_s(\theta^s)\int_{\theta_{s+1}}U''(R_s(\theta^s,\theta_{s+1}))dF_{s+1}(\theta_{s+1}|\theta^s)\\ &+(1-\tau_s)\beta(1+r)\mu_{s-1}(\theta^{s-1})U''\left(R_s(\theta^s)\right)f_s(\theta_s|\theta^{s-1})\\ &+(1-\tau_{s+1})(1+r)\int_{\theta_{s+1}}\mu_{s+1}(\theta^s,\theta_{s+1})U''(R_{s+1}(\theta^s,\theta_{s+1})d\theta_{s+1}=0\end{aligned} \tag{3.23}$$

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial \tau_s} =& \frac{\lambda}{(1+r)^{s-2}}\int_{\theta^{s-1}} a_t(\theta^{s-1})dH(\theta^{s-1})\\ &-\beta^{s-1}(1+r)\int_{\theta^{s-1}}a_s(\theta^{s-1})\int_{\theta_s}U'(R_s(\theta^{s-1},\theta_s))dF_s(\theta_s|\theta_{s-1})d\tilde{H}(\theta^{s-1})\\ &-(1+r)\int_{\theta^{s-1}}a_s(\theta^{s-1})\int_{\theta_s}\mu_s(\theta^{s-1},\theta_s)U''(R_s(\theta^{s-1},\theta_s))d\theta_s d\theta^{s-1}\\ &+\beta(1+r)^2(1-\tau_s)\int_{\theta^{s-1}}\mu_{s-1}(\theta^{s-1})\int_{\theta_s}U''(R_s(\theta^{s-1},\theta_s))a_s(\theta^{s-1})dF_s(\theta_s|\theta_{s-1})d\theta^{s-1}\\ &+\beta(1+r)\int_{\theta^{s-1}}\mu_{s-1}(\theta^{s-1})\int_{\theta_s}U'(R_s(\theta^{s-1},\theta_s))dF_s(\theta_s|\theta_{s-1})d\theta^{s-1}=0\end{aligned} \tag{3.24}$$

3.A.2 Multiplier Functions η_t and μ_t

We now derive the multiplier functions $\eta_t(\theta_t)$ and $\mu_t(\theta^t)$, which are needed in order to obtain the results for optimal taxes. Use (3.21) to obtain

$$\begin{aligned}\eta_s(\theta_s) =& \frac{\lambda}{(1+r)^{s-1}}\int_{\theta^{s-1}}\int_{\theta_s}^{\overline{\theta}} dF_s(\tilde{\theta}_s|\theta^{s-1})dH(\theta^{s-1})\\ &-\beta^{s-1}\int_{\theta^{s-1}}\int_{\theta_s}^{\overline{\theta}}U'(R_s(\theta^{s-1},\theta_s))dF_s(\tilde{\theta}_s|\theta^{s-1})d\tilde{H}(\theta^{s-1})\\ &-\int_{\theta^{s-1}}\int_{\theta_s}^{\overline{\theta}}\mu_s(\theta^{s-1},\theta_s)U''(R_s(\theta^{s-1},\theta_s))d\tilde{\theta}_s d\theta^{s-1}\\ &+\beta(1+r)(1-\tau_s)\int_{\theta^{s-1}}\mu_{s-1}(\theta^{s-1})\int_{\theta_s}^{\overline{\theta}}U''(R_s(\theta^{s-1},\theta_s))dF_s(\tilde{\theta}_s|\theta^{s-1})d\theta^{s-1}.\end{aligned} \tag{3.25}$$

Next, we derive μ_t. Use (3.23) to obtain, with $SOC_s(\theta^s)$ basically being the second-order condition for savings from the individuals problem:

$$\mu_s(\theta^s) = \frac{\frac{\lambda}{(1+r)^{s-1}}\tau_{s+1}h(\theta^{s-1}) + (1-\tau_s)\beta(1+r)\mu_{s-1}(\theta^{s-1})U''\left(R_s(\theta^s)\right) f_s(\theta_s|\theta^{s-1})}{SOC_s(\theta^s)} + \frac{(1-\tau_{s+1})(1+r)\int_{\theta_{s+1}} \mu_{s+1}(\theta^s,\theta_{s+1})U''(R_{s+1}(\theta^s,\theta_{s+1})d\theta_{s+1}}{SOC_s(\theta^s)}. \tag{3.26}$$

Therefore, we define some terms that make notation less burdensome:

$$A_s(\theta^s) = \frac{\frac{\lambda}{(1+r)^{s-1}}\tau_{s+1}h(\theta^{s-1})}{SOC_s}$$

$$B_s(\theta^s) = \frac{(1-\tau_s)\beta(1+r)U''\left(R_s(\theta^s)\right) f_s(\theta_s|\theta^{s-1})}{SOC_s}$$

$$C_s(\theta^s,\theta_{s+1}) = \frac{(1-\tau_{s+1})(1+r)U''(R_{s+1}(\theta^s,\theta_{s+1}))}{SOC_s},$$

then, we can rewrite (3.26) as

$$\mu_s(\theta^s) = A_s(\theta^s) + B_s(\theta^s)\mu_{s-1}(\theta^{s-1}) + \int_{\theta_{s+1}} C_s(\theta^s,\theta_{s+1})\mu_{s+1}(\theta^s,\theta_{s+1})d\theta_{s+1}.$$

Or, more concretely for $s = T-2$:

$$\mu_{T-2}(\theta^{T-2}) = A_{T-2}(\theta^{T-2}) + B_{T-2}(\theta^{T-2})\mu_{T-3}(\theta^{T-3}) + \int_{\theta_{T-1}} C_{T-2}(\theta^{T-2},\theta_{T-1})\mu_{T-1}(\theta^{T-2},\theta_{T-1})d\theta_{T-1}. \tag{3.27}$$

For $s = T-1$, we get:

$$\mu_{T-1}(\theta^{T-1}) = A_{T-1}(\theta^{T-1}) + B_{T-1}(\theta^{T-1})\mu_{T-2}(\theta^{T-2}). \tag{3.28}$$

Now insert (3.28) into (3.27). Omitting arguments, this yields:

$$\mu_{T-2} = \frac{A_{T-2} + B_{T-2}\mu_{T-3} + \int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1 - \int_{\theta_{T-1}} C_{T-2}(\theta^{T-2},\theta_{T-1})B_{T-1}(\theta^{T-1})d\theta_{T-1}}.$$

Now insert this into μ_{T-3}

$$\mu_{T-3} = A_{T-3} + B_{T-3}\mu_{T-4} + \int_{\theta_{T-2}} C_{T-3}\frac{A_{T-2} + B_{T-2}\mu_{T-3} + \int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1 - \int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}d\theta_{T-2}, \tag{3.29}$$

yielding

$$\mu_{T-3} = \frac{A_{T-3} + B_{T-3}\mu_{T-4} + \int_{\theta_{T-2}} C_{T-3} \frac{A_{T-2}+\int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}} d\theta_{T-2}}{1 - \int_{\theta_{T-2}} \frac{C_{T-3}B_{T-2}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}}. \tag{3.30}$$

Now insert this into μ_{T-4}

$$\begin{aligned}\mu_{T-4} =& A_{T-4} + B_{T-4}\mu_{T-5} \\ &+ \int_{\theta_{T-3}} C_{T-4} \frac{A_{T-3} + B_{T-3}\mu_{T-4} + \int_{\theta_{T-2}} C_{T-3} \frac{A_{T-2}+\int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}} d\theta_{T-2}}{1 - \int_{\theta_{T-2}} \frac{C_{T-3}B_{T-2}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}} d\theta_{T-3}.\end{aligned} \tag{3.31}$$

Rewrite to obtain

$$\begin{aligned}\mu_{T-4} =& \left[1 - \int_{\theta_{T-3}} C_{T-4}B_{T-3}\left[1 - \int_{\theta_{T-2}} C_{T-3}B_{T-2}\left[1 - \int_{\theta_{T-1}} C_2 B_{T-1}d\theta_{T-1}\right]^{-1}\right]^{-1}\right]^{-1} \\ &\Bigg(A_{T-4} + B_{T-4}\mu_{T-5} \\ &+ \int_{\theta_{T-3}} C_{T-4} \frac{A_{T-3} + \int_{\theta_{T-2}} C_{T-3} \frac{A_{T-2}+\int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}} d\theta_{T-2}}{1 - \int_{\theta_{T-2}} \frac{C_{T-3}B_{T-2}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}} d\theta_{T-3}\Bigg).\end{aligned} \tag{3.32}$$

Now we also calculate μ_{T-5} to make sure how the pattern looks like.

$$\begin{aligned}\mu_{T-5} =& \left[1 - \int_{\theta_{T-4}} C_{T-5}B_{T-4}\left[...\left[1 - \int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}\right]^{-1}....\right]^{-1}\right]^{-1} \\ &(A_{T-5} + B_{T-5}\mu_{T-6} \\ &+ \int_{\theta_{T-4}} C_{T-5} \frac{A_{T-4} + \int_{\theta_{T-3}} C_{T-4} \frac{A_{T-3}+\int_{\theta_{T-2}} C_{T-3} \frac{A_{T-2}+\int_{\theta_{T-1}} C_{T-2}A_{T-1}d\theta_{T-1}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}} d\theta_{T-2}}{1-\int_{\theta_{T-2}} \frac{C_{T-3}B_{T-2}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}}}{1 - \int_{\theta_{T-3}} \frac{C_{T-4}B_{T-3}}{1-\int_{\theta_{T-2}} \frac{C_{T-3}B_{T-2}}{1-\int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}}}}\Bigg).\end{aligned} \tag{3.33}$$

Now define

$$D_s = \left[1 - \int_{\theta_{s+1}} C_s B_{s+1}\left[1 - \int_{\theta_{s+2}} C_{s+1}B_{s+2}\left[...\left[1 - \int_{\theta_{T-1}} C_{T-2}B_{T-1}d\theta_{T-1}\right]^{-1}....\right]^{-1} d\theta_{s+2}\right]^{-1} d\theta_{s+1}\right]^{-1}.$$

Using this definition, we can write μ_{T-5} as

$$\mu_{T-5} = \frac{A_{T-5} + B_{T-5}\mu_{T-6} + \int_{\theta_{T-4}} C_{T-5} \frac{A_{T-4} + \int_{\theta_{T-3}} C_{T-4} \frac{A_{T-3} + \int_{\theta_{T-2}} C_{T-3} \frac{A_{T-2} + \int_{\theta_{T-1}} C_{T-2} A_{T-1}}{D_{T-2}}}{D_{T-3}}}{D_{T-4}}}{D_{T-5}}. \tag{3.34}$$

It now turns out helpful to make another definition:

$$E_s = \int_{\theta_{s+1}} C_s \frac{A_{s+1} \int_{\theta_{s+2}} C_{s+1} \frac{A_{s+2} + \int_{\theta_{s+3}} C_{s+2} \frac{A_{s+3} + \int_{\theta_{s+4}} C_{s+3} \frac{A_{s+4} \cdots\cdots}{D_{s+4}}}{D_{s+3}}}{D_{s+2}}}{D_{s+1}}.$$

Then we can write μ_{T-5} as

$$\mu_{T-5} = \frac{A_{T-5} + B_{T-5}\mu_{T-6} + E_{T-5}}{D_{T-5}}.$$

In genereal, we thus obtain:

$$\mu_s = \frac{A_s + B_s \mu_{s-1} + E_s}{D_s}.$$

For the second period, we thus obtain

$$\mu_2 = \frac{A_2 + B_2 \mu_1 + E_2}{D_2}. \tag{3.35}$$

then we should get

$$\mu_1 = \frac{A_1 + E_1}{D_1}. \tag{3.36}$$

Now we can recursively calculate all other μ_t for $t = 2, ..., T$.
In equation (3.36) one can see that the $\mu_1(\theta_1) = 0$ if savings taxes are zero. Recursive calculation reveals that all μ_t are equal to zero.

3.A.3 Labor Income Taxes

3.A.3.1 T periods

Dividing (3.22) by $\Psi' \frac{1}{\theta_s}$ and adding (3.21) yields

$$\frac{\mathcal{T}_s'(y_s(\theta_s))}{1 - \mathcal{T}_s'(y_s(\theta_s))} = \left(1 + \frac{1}{\varepsilon_{y_s, 1-\mathcal{T}'}(\theta_s)}\right) \frac{\eta_s(\theta^s)}{\lambda \frac{1}{(1+r)^{s-1}} \theta_s \int_{\theta_s} f(\theta_s|\theta_{s-1}) dH(\theta^{s-1})}. \tag{3.37}$$

Inserting (3.25) in (3.37) yields Proposition 3.3.5.

3.A.3.2 Two Periods

For the multiplier function $\mu_1(\theta_1)$ we now have:

$$\mu_1(\theta_1) = \frac{\lambda\tau}{SOC(\theta_1)}$$

where $SOC(\theta_1)$ are the second-order conditions of the savings decision for an individual of type θ_1. Thus we obtain from (3.25):

$$\eta_1(\theta_1) = \lambda(1 - F(\theta_1)) - \int_{\theta_1}^{\overline{\theta}} U'(\tilde{\theta}_1)d\tilde{F}(\tilde{\theta}_1) - \lambda\tau \int_{\theta_1}^{\overline{\theta}} \frac{U''(\tilde{\theta}_1)}{SOC(\theta_1)} dF(\tilde{\theta}_1) \tag{3.38}$$

and

$$\eta_2(\theta_1, \theta_2) = \frac{1}{1+r} \int_{\theta_1} \int_{\theta_2}^{\overline{\theta}} dF(\tilde{\theta}_2|\theta_1)dF(\theta_1) - \beta \int_{\theta_1} \int_{\theta_2}^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2)dF(\tilde{\theta}_2|\theta_1)d\tilde{F}(\tilde{\theta}_1) \tag{3.39}$$

$$+ \lambda\tau \int_{\theta_1} \frac{\int_{\theta_2}^{\overline{\theta}} U''(\theta_1, \tilde{\theta}_2)dF(\tilde{\theta}_2|\theta_1)}{SOC(\theta_1)} dF(\theta_1). \tag{3.40}$$

By implicitly differentiating the Euler equation of an individual of type θ_1, one can show that the last terms in (3.38) and (3.40) capture the impact of savings responses on the government budget and therefore one arrives at the results in Propositions 3.3.1 and 3.3.2.

3.A.4 Capital Income Taxes

3.A.4.1 T periods

Simply rearranging (3.24) yields the formula as in Proposition 3.3.6.

3.A.4.2 Two periods

In case of two periods, (3.24) reads as (after inserting $\mu_1(\theta_1)$)

$$\begin{aligned} 0 =& \lambda \int_{\theta_1} a_2(\theta_1)dF(\theta_1) \\ &- \beta(1+r) \int_{\theta_1} a_2(\theta_1) \int_{\theta_2} U'(R_2(\theta_1, \theta_2))dF(\theta_2|\theta_1) \\ &+ \lambda\tau\beta(1+r)^2(1-\tau) \int_{\theta_1} \frac{\int_{\theta_2} U''(R_2(\theta_1, \theta_2))a_2(\theta_1)dF(\theta_2|\theta_1)}{SOC(\theta_1)} d\theta_1 \\ &+ \lambda\tau\beta(1+r) \int_{\theta_1} \frac{\int_{\theta_2} U'(R_2(\theta_1, \theta_2))dF(\theta_2|\theta_1)}{SOC(\theta_1)} d\theta_1 = 0. \end{aligned} \tag{3.41}$$

Again by implicitly differentiating the Euler equation, one can show that line three and four of (3.41) capture the impact of savings responses on the government budget, where line three captures the income effect on savings and line four captures the price effect. Simple rearranging then yields the formula as in Proposition 3.3.4.

Appendix 3.B Age-Independent Taxes

3.B.1 The Lagrangian and the First-Order Conditions

Here we have $y_t(\theta_t) = y(\theta_t)$ and $M_t(\theta_t) = M(\theta_t)$. The Lagrangian then reads as

$$\begin{aligned}
\mathcal{L} = &\int_{\theta_1}\left(\sum_{t=1}^{T}\beta^{t-1}\int_{\theta^t} U\Bigg(M(\theta_t) - a_{t+1}(\theta^t)\right.\\
&\left. + (1-\tau)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y(\theta_t)}{\theta_t}\right)\Bigg)dH(\theta^t)\right)d\tilde{F}_1(\theta_1)\\
&+ \lambda\sum_{t=1}^{T}\frac{1}{(1+r)^{t-1}}\int_{\theta^{t-1}}\int_{\theta_t} y(\theta_t) - M(\theta_t) + \tau(1+r)a_t(\theta^{t-1})dF(\theta_t|\theta_{t-1})dH(\theta^{t-1})\\
&+ \sum_{t=1}^{T-1}\int_{\theta^t}\mu_t(\theta^t)\Bigg[U'\left(M(\theta_t) - a_{t+1}(\theta^t) + (1-\tau)(1+r)a_t(\theta^{t-1}) - \Psi\left(\frac{y(\theta_t)}{\theta_t}\right)\right)\\
&- \beta(1+r)(1-\tau_{t+1})\int_{\theta_{t+1}} U'\Bigg(M(\theta_{t+1}) - a_{t+2}(\theta^t,\theta_{t+1})\\
&+ (1-\tau)(1+r)a_{t+1}(\theta^t) - \Psi\left(\frac{y(\theta_{t+1})}{\theta_{t+1}}\right)\Bigg)dF(\theta_{t+1}|\theta^t)\Bigg]d\theta^t\\
&+ \int_\theta \eta(\theta)\frac{\partial\left(M(\theta) - \Psi\left(\frac{y(\theta)}{\theta}\right)\right)}{\partial\theta}d\theta - \int_\theta \eta(\theta)\Psi'\left(\frac{y(\theta)}{\theta}\right)\frac{y(\theta)}{\theta^2}d\theta.
\end{aligned}$$

Partially integrating $\int_\theta \eta(\theta)\frac{\partial\left(M(\theta)-\Psi\left(\frac{y(\theta)}{\theta}\right)\right)}{\partial\theta}d\theta$ yields

$$\eta(\overline{\theta})\left(M(\overline{\theta}) - \Psi\left(\frac{y(\overline{\theta})}{\overline{\theta}}\right)\right) - \eta(\underline{\theta})\left(M(\underline{\theta}) - \Psi\left(\frac{y(\underline{\theta})}{\underline{\theta}}\right)\right) - \int_\theta \eta'(\theta)\left(M(\theta) - \Psi\left(\frac{y(\theta)}{\theta}\right)\right)d\theta.$$

The derivatives with respect to the endpoint conditions yield $\forall t:\ \eta_t(\overline{\theta}) = \eta_t(\underline{\theta}) = 0$. The first-order conditions read as

$$\begin{aligned}
\frac{\partial\mathcal{L}}{\partial M(\theta)} = &-\sum_{t=1}^{T}\frac{\lambda}{(1+r)^{t-1}}\int_{\theta^{t-1}} f(\theta|\theta_{t-1})dH(\theta^{t-1})\\
&+ \sum_{t=1}^{T}\beta^{t-1}\int_{\theta^{t-1}} U'(R_t(\theta^{t-1},\theta))f(\theta|\theta_{t-1})d\tilde{H}(\theta^{t-1})\\
&+ \sum_{t=1}^{T-1}\int_{\theta^{t-1}}\mu_t(\theta^{t-1},\theta)U''(R_t(\theta^{t-1},\theta))d\theta^{t-1}\\
&- \sum_{t=2}^{T}\beta(1+r)(1-\tau)\int_{\theta^{t-1}}\mu_{t-1}(\theta^{t-1})U''(R_t(\theta^{t-1},\theta))f(\theta|\theta^{t-1})d\theta^{t-1}\\
&- \eta'(\theta) = 0
\end{aligned} \tag{3.42}$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial y(\theta)} =& \sum_{t=1}^{T} \frac{\lambda}{(1+r)^{t-1}} \int_{\theta^{t-1}} f(\theta|\theta_{t-1}) dH(\theta^{t-1}) \\
&- \sum_{t=1}^{T} \beta^{t-1} \int_{\theta^{t-1}} U'(R_t(\theta^{t-1},\theta)) \Psi'\left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta} f(\theta|\theta_{t-1}) d\tilde{H}(\theta^{t-1}) \\
&+ \sum_{t=1}^{T-1} \int_{\theta^{t-1}} \mu_t(\theta^{t-1},\theta) U''(R_t(\theta^{t-1},\theta)) \Psi'\left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta} d\theta^{t-1} \\
&- \sum_{t=2}^{T} \beta(1+r)(1-\tau) \int_{\theta^{t-1}} \mu_{t-1}(\theta^{t-1}) U''(R_t(\theta^{t-1},\theta)) \Psi'\left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta} f(\theta|\theta^{t-1}) d\theta^{t-1} \\
&- \eta'(\theta) \Psi'\left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta} - \eta(\theta) \left(\Psi'\left(\frac{y(\theta)}{\theta}\right) \frac{1}{\theta^2} + \Psi''\left(\frac{y(\theta)}{\theta}\right) \frac{y(\theta)}{\theta^2} \right) = 0 \qquad (3.43)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial a_{s+1}(\theta^s)} =& \frac{\lambda}{(1+r)^{s-1}} \tau h(\theta^{s-1}) - \mu_s(\theta^s) U''\left(R_s(\theta^s)\right) \\
&- (1-\tau)^2 \beta (1+r)^2 \mu_s(\theta^s) \int_{\theta_{s+1}} U''(R_s(\theta^s,\theta_{s+1})) dF(\theta_{s+1}|\theta^s) \\
&+ (1-\tau)\beta(1+r)\mu_{s-1}(\theta^{s-1}) U''\left(R_s(\theta^s)\right) f(\theta_s|\theta^{s-1}) \\
&+ (1-\tau)(1+r) \int_{\theta_{s+1}} \mu_{s+1}(\theta^s,\theta_{s+1}) U''(R_{s+1}(\theta^s,\theta_{s+1}) d\theta_{s+1} = 0 \qquad (3.44)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \tau} =& \sum_{t=2}^{T} \frac{\lambda}{(1+r)^{t-2}} \int_{\theta^{t-1}} a_t(\theta^{t-1}) dH(\theta^{t-1}) \\
&- \sum_{t=2}^{T} \beta^{t-1}(1+r) \int_{\theta^{t-1}} a_t(\theta^{t-1}) \int_{\theta_t} U'(R_t(\theta^{t-1},\theta_t)) dF(\theta_t|\theta_{t-1}) d\tilde{H}(\theta^{t-1}) \\
&- \sum_{t=1}^{T-1} (1+r) \int_{\theta^{t-1}} a_t(\theta^{t-1}) \int_{\theta_t} \mu_t(\theta^{t-1},\theta_t) U''(R_t(\theta^{t-1},\theta_t)) d\theta_t d\theta^{t-1} \\
&+ \sum_{t=2}^{T} \beta(1+r)^2(1-\tau) \int_{\theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\theta_t} U''(R_t(\theta^{t-1},\theta_t)) a_t(\theta^{t-1}) dF(\theta_t|\theta_{t-1}) d\theta^{t-1} \\
&+ \sum_{t=2}^{T} \beta(1+r) \int_{\theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\theta_t} U'(R_t(\theta^{t-1},\theta_t)) dF(\theta_t|\theta_{t-1}) d\theta^{t-1} = 0. \qquad (3.45)
\end{aligned}
$$

3.B.2 Multiplier Functions η and μ_t

We now derive the multiplier functions $\eta(\theta_t)$, which is needed i order to obtain the results for optimal taxes. Obtaining $\mu_t(\theta^t)$ is equivalent to the age-dependent case in Appendix 3.A.2. Use (3.42) to obtain

$$\begin{aligned}\eta(\theta) =& \sum_{t=1}^{T} \frac{\lambda}{(1+r)^{t-1}} \int_{\theta^{t-1}} \int_{\theta}^{\overline{\theta}} dF(\tilde{\theta}|\theta^{t-1}) dH(\theta^{t-1}) \\ &- \sum_{t=1}^{T} \beta^{t-1} \int_{\theta^{t-1}} \int_{\theta}^{\overline{\theta}} U'(R_t(\theta^{t-1}, \theta_t)) dF(\tilde{\theta}|\theta^{t-1}) d\tilde{H}(\theta^{t-1}) \\ &- \sum_{t=1}^{T} \int_{\theta^{t-1}} \int_{\theta}^{\overline{\theta}} \mu_t(\theta^{t-1}, \tilde{\theta}) U''(R_t(\theta^{t-1}, \tilde{\theta})) d\tilde{\theta} d\theta^{t-1} \\ &+ \sum_{t=1}^{T} \beta(1+r)(1-\tau) \int_{\theta^{t-1}} \mu_{t-1}(\theta^{t-1}) \int_{\theta}^{\overline{\theta}} U''(R_t(\theta^{t-1}, \tilde{\theta})) dF(\tilde{\theta}|\theta^{t-1}) d\theta^{t-1}. \end{aligned} \tag{3.46}$$

3.B.3 Labor Income Taxes

3.B.3.1 T Periods

Dividing (3.43) by $\Psi' \frac{1}{\theta}$ and adding (3.42) yields

$$\frac{\mathcal{T}'(y(\theta))}{1-\mathcal{T}'(y(\theta))} = \left(1 + \frac{1}{\varepsilon_{y_t, 1-\mathcal{T}'}(\theta)}\right) \frac{\eta(\theta)}{\lambda\theta \sum_{t=1}^{T} \frac{1}{(1+r)^{t-1}} \int_{\theta_{t-1}} \int_{\theta_t} f(\theta_t|\theta_{t-1} dH(\theta^{t-1})}. \tag{3.47}$$

Inserting (3.46) into (3.47) yields the formula for optimal labor tax rates.

3.B.3.2 Two Periods

As in Appendix 3.A.3.2, for the multiplier function $\mu_1(\theta_1)$ we now have:

$$\mu_1(\theta_1) = \frac{\lambda\tau}{SOC(\theta_1)}.$$

For $\eta(\theta)$ we then have

$$\begin{aligned}\eta(\theta) =& \lambda(1-F(\theta)) - \int_{\theta}^{\overline{\theta}} U'(\tilde{\theta}_1) d\tilde{F}(\tilde{\theta}_1) - \lambda\tau \int_{\theta}^{\overline{\theta}} \frac{U''(\tilde{\theta}_1)}{SOC(\theta_1)} dF(\tilde{\theta}_1) \\ & \frac{1}{1+r} \int_{\theta_1} \int_{\theta}^{\overline{\theta}} dF(\tilde{\theta}_2|\theta_1) dF(\theta_1) - \beta \int_{\theta_1} \int_{\theta}^{\overline{\theta}} U'(\theta_1, \tilde{\theta}_2) dF(\tilde{\theta}_2|\theta_1) d\tilde{F}(\tilde{\theta}_1) \qquad (3.48) \\ &+ \lambda\tau \int_{\theta_1} \frac{\int_{\theta}^{\overline{\theta}} U''(\theta_1, \tilde{\theta}_2) dF(\tilde{\theta}_2|\theta_1)}{SOC(\theta_1)} dF(\theta_1). \qquad (3.49)\end{aligned}$$

By similar reasoning as in Proposition 3.A.3.2, we arrive at Proposition 3.3.3.

3.B.4 Capital Income Taxes

Rearranging (3.45) would yield the age-independent equivalent to the age-dependent formula in Proposition 3.3.6.

Appendix 3.C Details on Numerical Simulations

We use the empirical model from Karahan and Ozkan (2013), who estimate their model using PSID-data. $y^i_{h,t}$ denotes log income of individual i at age h in period t. To obtain residual log incomes $\tilde{y}^i_{h,t}$, the authors regress log earnings on some observables (age and education):

$$y^i_{h,t} = f(X^i_a; \theta_t) + \tilde{y}^i_{h,t},$$

where $f(X^i_a)$ is a function of the observable characteristics. Residual income is then decomposed into a fixed effect (α^i), an AR(1) component ($z^i_{h,t}$) and a transitory component ($\phi_t \epsilon^i_h$):

$$\tilde{y}^i_{h,t} = \alpha^i + z^i_{h,t} + \phi_t \epsilon^i_{h,t},$$

where the AR(1) process is given by

$$z^i_{h,t} = \rho_{h-1} z^i_{h-1,t-1} + \pi_t \eta^i_h,$$

and where the error term η^i_h captures persistent shocks, π_t is a time dependent loading factor and ρ_{h-1} measures the persistence of these shocks.
Based on non-parametric estimates, Karahan and Ozkan (2013) divide individuals into three age groups: 24-33 (young), 34-52 (middle age) and 53-60 (old). In the following, we list the values they obtain for the different parameters, where the indices Y, M, O correspond to the three age groups.
Age-dependent parameters:

- Persistence parameters: $\rho_Y = 0.88$, $\rho_M = 0.97$ and $\rho_O = 0.96$,
- Variances of the persistent error terms: $\sigma^2_{\eta,Y} = 0.027$, $\sigma^2_{\eta,M} = 0.013$ and $\sigma^2_{\eta,O} = 0.026$
- Variances of the transitory shock: $\sigma^2_{\epsilon,Y} = 0.056$, $\sigma^2_{\epsilon,M} = 0.059$ and $\sigma^2_{\epsilon,O} = 0.068$

Age-independent parameters:

- Variance of individual fixed effect: $\sigma^2_\alpha = 0.0707$
- Variance of z_1 (i.e. the starting value of the persistence term): $\sigma^2_z = 0.0767$

Time-dependent parameters:

- As we consider only one cohort, we assume the time dependent loading factors π_t and ϕ_t to be constant. Indeed, we set them to $\pi = 1.1253$ and $\phi = 1.1115$ which corresponds to the values from 1996 as they lie roughly in the middle of all estimates for the years from 1968-1997.

Finally, we had to make some assumptions about the observable characteristics X, which should capture education and age. In the spirit of a numerical example, we chose values that are within the range of numbers one can find in different studies and data bases. In unreported simulations, we also conducted robustness checks by varying these numbers; the main results from our numerical exercise were unaffected. The assumptions that we made are the following. First, we had to make an assumption about the education composition at the beginning of their lives. We chose 38% college graduates, 37% high school graduates and 25% individuals without high school degree. We then had to choose the average income of each education group for the age of 24: We assumed

$18,000 for the individuals without high school degree, $32,000 for the high school graduates and $48,000 for the college graduates. For the deterministic age trend, we assumed average income of the middle age group is 20% higher than that of the young age group, and average income of the old age group is 10% higher than that of the young age group.
Based on all these parameters, one can now simulate the evolution of the earnings distribution. We simulated millions of lives such that a law of large numbers applies. For each simulated life, we then have the income for each year, which allows us to calculate the average income of one individual for all three parts of his life (24-33, 34-52 and 53-60). For each simulated life, we then have three levels of earnings.
We next discretized the earnings distribution. Thus for each simulated life, we then have 3 grid points; one for each period. With a standard kernel smoother (bandwidth of $2,500), we then smoothed the unconditional earnings distributions over this grid space as well as the conditional earnings distributions and therefore the transition probabilities. The final step was then to calibrate the skill distributions from the earnings distributions which we did in the same way as in Chapter 1, see Appendix 1.D.1 for details.

Bibliography of Chapter 3

AIYAGARI, S. R. (1994): "Uninsured idiosyncratic risk and aggregate saving," *The Quarterly Journal of Economics*, 109, 659–684.

ATKINSON, A. AND J. STIGLITZ (1976): "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics*, 6, 55–75.

BANKS, J. AND P. DIAMOND (2011): "The Base for Direct Taxation," in *Dimensions of Tax Design. The Mirrlees Review*, Oxford University Press, chap. 6, 548–648.

BASTANI, S., S. BLOMQUIST, AND L. MICHELETTO (2011): "The welfare gains of age related optimal income taxation," *International Economic Review (forthcoming).*

CHETTY, R. (2006): "A new method of estimating risk aversion," *The American Economic Review*, 96, 1821–1834.

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

CONESA, J. C., S. KITAO, AND D. KRUEGER (2009): "Taxing Capital? Not a Bad Idea after All!" *The American Economic Review*, 25–48.

DIAMOND, P. (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88, 83–95.

Diamond, P. and J. Spinnewijn (2011): "Capital income taxes with heterogeneous discount rates," *American economic journal: economic policy*, 3, 52–76.

Farhi, E. and I. Werning (2013): "Insurance and Taxation over the Life Cycle," *The Review of Economic Studies*.

Golosov, M., M. Troshkin, A. Tsyvinski, and M. Weinzierl (2013a): "Preference heterogeneity and optimal capital income taxation," *Journal of Public Economics*, 97, 160–175.

Golosov, M., A. Tsyvinski, and N. Werquin (2013b): "Dynamic Tax Reforms," *Working Paper*.

Greenwood, J., Z. Hercowitz, and G. W. Huffman (1988): "Investment, capacity utilization, and the real business cycle," *The American Economic Review*, 402–417.

Gruber, J. and E. Saez (2002): "The elasticity of taxable income: evidence and implications," *Journal of Public Economics*, 84, 1–32.

Heathcote, J., F. Perri, and G. L. Violante (2010): "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006," *Review of Economic Dynamics*, 13, 15–51.

Jacobs, B. and D. Schindler (2012): "On the desirability of taxing capital income in optimal social insurance," *Journal of Public Economics*, 96, 853–868.

Jappelli, T. and L. Pistaferri (2010): "The Consumption Response to Income Changes," *Annual Review of Economics*, 2, 479–506.

Kaplow, L. (2006): "On the undesirability of commodity taxation even when income taxation is not optimal," *Journal of Public Economics*, 90, 1235–1250.

Karahan, F. and S. Ozkan (2013): "On the persistence of income shocks over the life cycle: Evidence, theory, and implications," *Review of Economic Dynamics*, 26, 452–476.

Kleven, H. J. and E. A. Schultz (2012): "Estimating Taxable Income Responses using Danish Tax Reforms," .

Kocherlakota, N. R. (2005): "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 73, 1587–1621.

Meghir, C. and L. Pistaferri (2011): "Earnings, consumption and life cycle choices," *Handbook of Labor Economics*, 4, 773–854.

Mendoza, E. G. and V. Z. Yue (2012): "A general equilibrium model of sovereign default and business cycles," *The Quarterly Journal of Economics*, 127, 889–946.

Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38, 175–208.

Piketty, T. (1997): "La redistribution fiscale face au chomage," *Revue française d'économie*, 12, 157–201.

Piketty, T. and E. Saez (2013): "A Theory of Optimal Inheritance Taxation," *Econometrica*, forthcoming.

Saez, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205–229.

——— (2002): "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes," *Journal of Public Economics*, 83, 217–230.

Salanié, B. (2003): *The Economics of Taxation*, The MIT Press.

Weinzierl, M. (2011): "The Surprising Power of Age-Dependent Taxes," *The Review of Economic Studies*, 78, 1490–1518.

Chapter 4

Education Policies and Taxation without Commitment

4.1 Introduction

Public finance economists have long recognized that the challenges involved in the design of optimal education policies and income tax systems are intimately related. Income taxation influences the incentives to invest in education.[1] Education subsidies and policies, in turn, influence the choice of an optimal income tax system as they have a direct effect on both the level and the distribution of wages. Many papers have studied the design of education and tax policies jointly from a normative perspective – see, for example, Bovenberg and Jacobs (2005) for a state-of-the-art treatment in a heterogeneous agent model.[2] This strand of literature assumes that individuals rationally make human capital investment decisions, reacting to incentives set by the tax code and education subsidies. Importantly, the government fully commits to the income tax schedule that it announces before education decisions are made.

Boadway et al. (1996) have drawn attention to the issue of time-consistency, in the spirit of Kydland and Prescott (1977), inherent in the design of optimal tax and education policies. If the government lacks a device to credibly commit to tax policies at the time individuals make education decisions, this can dramatically depress the incentives of young individuals to invest into human capital. In their framework, they show that this underinvestment arises and make a case for mandatory education as a second-best policy in the presence of commitment problems.

This chapter looks at the implications of limited commitment and policy credibility on education and tax policies from a new perspective. Consistent with real world practices, the government can decide to subsidize different levels of education at different rates. The idea here is that governments typically intervene at primary, secondary and tertiary education levels. However, as we will also exploit in our empirical section, the rate at which these different education levels are subsidized is very different. We formalize this by allowing the government to set a nonlinear schedule of education subsidies. The income tax is linear and the revenue is redistributed lump-sum and used to finance education subsidies. We derive our results in a transparent and simple heterogeneous agent model with two types (Stiglitz, 1982) and two periods: an education and a working period. Consistent with empirical evidence, individual wages are determined by both innate abilities and education levels.

[1] See Abramitzky and Lavy (2012) for recent quasi-experimental evidence on the negative effect of redistributive taxation on education investment. More structural and model based approaches as the classic work by Trostel (1993) also have found big effects of income taxation on human capital investment.

[2] See Richter (2009) for a recent treatment in a Ramsey setting with a representative agent.

We first consider the two polar cases where the government has full commitment to stick to tax promises and no commitment at all. Under full commitment the optimal income tax rate takes into account education incentives. The tax rate is smaller, when the effect of education on wages is large relative to the effect of innate abilities on wages for the initially high skilled. Intuitively, the more important the role of education for wages the more important taxes become to incentivize high-skilled agents to self-select into high education level. Education subsidies for the high types are set such that a first-best rule for education is fulfilled: the subsidy corrects for the fiscal externality as in Bovenberg and Jacobs (2005). For the low type, in addition to this correction of the fiscal externality, education is downwards distorted at the margin to relax the incentive constraint of the high type.

Without any commitment, no tax promise of the government is credible and individuals rationally anticipate that the government re-optimizes after education is sunk. In line with previous results, this leads to excessive taxation and depresses human capital investment. An important result we find here concerns the design of education subsidies. We show that they tend to become more progressive when there is a commitment problem. The intuition is that a higher subsidy for low types and a lower subsidy for high types will compress the distribution of education. As education inequality is reduced, also the wage distribution in the next period is more compressed. And as wage inequality decreases, the redistributive government sets a lower tax rate in the second period. This lower tax rate will help to boost education incentives and helps to alleviate the commitment problem. This is consistent with the recent results from Farhi et al. (2012) who first detected a similar channel for nonlinear capital taxation.

We move on to study intermediate scenarios, where the government has some form of limited commitment, nesting the two polar cases. More concretely, the government can deviate from its announcements but this induces some output costs capturing the idea of a reputational loss.[3] The forward looking government wants to avoid costly deviation and announces policies respectively. Labor income taxes are still designed to take into account their effect on education incentives. However, the strength of the effect is decreasing the more severe the commitment problem is because too low tax promises lack credibility. Education policies become more progressive the more severe the commitment problem if the government is sufficiently redistributive towards low types. Another way to look at this result is to interpret the design of education and tax policies as a choice to engage in redistribution *ex-ante* through

[3]Farhi et al. (2012) show how to microfound such an output loss in a dynamic repeated game, where a deviation today brings a reputational cost borne in the future, because of depressed investment of future generations.

more progressive education subsidies – which decreases wage inequality – as opposed to engaging in redistribution *ex-post* through income taxation. With a lack of commitment, the government tries to weaken its own temptation to engage in costly ex-post redistribution by increasing the amount of ex-ante redistribution.

We conclude by providing suggestive evidence for the described mechanisms in the form of cross-country correlations. We proxy for commitment power using data from the World Bank's Worldwide Governance Indicators. Specifically, we use the variable Government Effectiveness capturing "...the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies." (Kaufmann et al., 2010). Controlling for income, geographical variables and the overall share of government involvement in education, we find a robust correlation, indicating that countries with higher policy credibility employ more regressive subsidies.

As already mentioned, this chapter is related to Farhi et al. (2012), who consider capital taxation without commitment. An important difference is that, in their framework, deviation could imply full redistribution, i.e. 100% marginal tax rates on capital. In our case, deviation is constrained to be less extreme as the government still has to respect labor supply responses and therefore cannot fully tax away human capital returns. Despite this article by Farhi et al. (2012) and the mentioned work of Boadway, Marceau, and Marchand (1996), this chapter is related to the work on time inconsistency and education policies by Konrad (2001) and Andersson and Konrad (2003).[4] Konrad (2001) shows how the time inconsistency problem is alleviated by the presence of private information in an optimal taxation framework. In particular, he shows that the strong no-education result obtained in Boadway et al. (1996) no longer applies, as with private information some rents of education are still captured by individuals, preserving some incentives to invest in education.[5] In our framework, a similar logic applies as the government uses a linear tax income rate together with lump-sums, in the spirit of a simple negative income tax system. This also preserves some incentives to invest in education, even in the complete absence of credible policy promises, as full equalization of incomes is not feasible. In contrast to Konrad (2001), we consider ex-ante heterogeneous individuals and thereby address the progressivity of education subsidies. Lastly, Andersson and Konrad (2003) investigate education policies chosen by extortionary

[4]In a related paper, Pereira (2009) studies linear education subsidies and shows that this subsidy offsets some of the excessive redistribution from income taxes, when the government lacks commitment.

[5]Poutvaara (2003) shows that redistribution without commitment may still involve more education than in the laissez-faire if the insurance effect of taxes is important.

governments lacking commitment and how migration and tax competition affect policies.[6] We depart from these papers by placing our focus on nonlinear education subsidies as used in the real world.

In Section 4.2, we introduce the formal model and look at optimal income taxation with exogenous education as a benchmark case. In Section 4.3, we derive optimal policies for the full-commitment government before we look at the other extreme case where the government cannot commit at all in Section 4.4. We then look at the intermediate case of partial commitment in Section 4.5 before we present some suggestive empirical evidence in Section 4.6 and conclude in Section 4.7.

4.2 The Model Basics

In this section, we present the model basics and characterize optimal taxes with exogenous education, an important benchmark which helps to understand the further results with endogenous education.

4.2.1 Environment

We consider a two-period model, where ex-ante heterogeneous agents make an educational investment in the first period. In the second period, they make a labor leisure decision. More formally, there are two types of ex-ante heterogeneous agents. The θ_1-type and the θ_2-type with $\theta_2 > \theta_1$. Their masses are $f(\theta_1)$ and $f(\theta_2)$ with $f(\theta_1) + f(\theta_2) = 1$. In Period 1, they make a monetary educational investment e. The wage w they earn in period two is a function of innate type and education, i.e. $w(\theta, e)$.

We impose three intuitive assumptions on the wage function $w(\theta, e)$. First, education is productive and raises wages $\frac{\partial w(\theta,e)}{\partial e} > 0$. Second education and innate ability are complements implying higher marginal returns to education for the higher innate type: $\frac{\partial w(\theta_2,e)}{\partial e} - \frac{\partial w(\theta_1,e)}{\partial e} > 0$. Finally, innate abilities positively influence wages for a given level of education: $w(\theta_2, e) - w(\theta_1, e) > 0$. None of these assumption are needed for most of the results we derive in the sense that all formulas are valid if we deviate from those assumptions. These assumptions ease the understanding of the model, however, and have strong empirical support.[7]

[6] In a median voter framework, Poutvaara (2011) shows that generous subsidies for high education may make the median voter of the future a college graduate, leading to lower taxes compared to a world with lower subsidies for high education.

[7] Card (1999) summarizes a long strand of literature estimating the *causal* effect of education on earnings. Carneiro and Heckman (2005) and Lemieux (2006), among others, document complementarity between innate skills and formal education. Taber (2001) and Hendricks and Schoellman

We assume quasi-linear preferences. To minimize the notational burden we often write all the variables not as a function of θ but with subscript instead. E.g. e_1 instead of $e_1(\theta_1)$ or w_2 instead of $w(e_2(\theta_2), \theta_2)$. The utility functions are $U^1 = c^1$ in period one and $U^2 = c^2 - \Psi(h)$ in period two, where h are hours worked. For simplicity, we assume that $\Psi(h) = \frac{h^{1+\frac{1}{\varepsilon}}}{1+\frac{1}{\varepsilon}}$, i.e. that Ψ exhibits a constant elasticity of labor supply ε. Before tax income is denoted by $y_i = w_i h_i$. Further we assume no discounting and a zero interest rate for notational convenience.
We are considering redistributive linear taxation. That is, we are interested in the policies of a government that is interested in redistributing from the high type θ_2 to the low type θ_1 via linear taxes used to finance a lump-sum rebate such as in a negative income tax system. To capture this redistributive concern, we set the Pareto weights $\tilde{f}(\theta_1)$ and $\tilde{f}(\theta_2)$ such that $\frac{\tilde{f}(\theta_1)}{f(\theta_1)} > \frac{\tilde{f}(\theta_2)}{f(\theta_2)}$. When deciding about the optimal degree of redistribution, the government has to take into account that taxes will (i) lower incentives to work and also (ii) lower incentives to invest in education. The education margin, however, can also be influenced by nonlinear education subsidies. Before looking at optimal policies in the different commitment scenarios, we look at the simple benchmark case of exogenous education where commitment issues do not arise.

4.2.2 Optimal Policies with Exogenous Education

Assume a one period setting where education levels e_1 and e_2 are exogenous. In that case, the only relevant margin for the government when choosing taxes is the labor-leisure margin. The problem of the government then simply is

$$\begin{aligned} \max_t \quad & \tilde{f}(\theta_1)\Bigg((1-t)w_1 h(t,w_1) + T - \Psi[h(t,w_1)]\Bigg) \\ & + \tilde{f}(\theta_2)\Bigg((1-t)w_2 h(t,w_2) + T - \Psi[h(t,w_2)]\Bigg) \end{aligned} \tag{4.1}$$

subject to a government budget constraint

$$T = t\Bigg(f(\theta_1)w_1 h(t,w_1) + (1 - f(\theta_1))w_2 h(t,w_2)\Bigg) \tag{4.2}$$

(2012) suggest that much of the rise in the college premium may be attributed to a rise in the demand for unobserved skills, which are predetermined and independent of education.

and optimal labor supply of the individuals

$$h(t, w_i) = \arg\max_h (1 - t)hw_i - \Psi(h).$$

The government thus only has to choose t optimally and thereby take into account how the transfer T is determined by the government budget constraint (4.2) and how individuals' hours worked h respond.[8] It is then easy to show that the optimal linear tax rate t^{ex}, in this case with exogenous human capital, satisfies

$$\frac{t^{ex}}{1 - t^{ex}} = \frac{\left(\tilde{f}(\theta_1) - f(\theta_1)\right)\left(\frac{y_2 - y_1}{\bar{y}}\right)}{\varepsilon}, \tag{4.3}$$

where $\bar{y}$ is average income $f(\theta_1)y_1 + f(\theta_2)y_2$. The optimal tax rate is increasing in redistributive preferences $\left(\tilde{f}(\theta_1) - f(\theta_1)\right)$, increasing in inequality measured by $\frac{y_2 - y_1}{\bar{y}}$ and decreasing in the elasticity of labor supply. The formula (4.3) is a variation for the optimal linear tax rate of Sheshinski (1972).[9] We refrain from providing a formal proof for this simple case as it is nested in the following formulas with endogenous educational attainment.

4.3 Optimal Policies with Full Commitment

4.3.1 The Government's Problem

We now consider the case where the educational decision is endogenous and the government can influence the decision of the agents by setting a nonlinear subsidy schedule. Thus, the government chooses a (nonlinear) subsidy function $S(e)$ and an income tax rate t subject to a government budget constraint and subject to behavioral responses of the individuals.[10] Thus, formally we have:

[8] In addition, one must also respect a non-negativity constraint on consumption that might be binding for some "extreme" Pareto weights. In the following we assume that the Pareto weights which we are considering will never be that "extreme".

[9] See Stantcheva (2013) for a similar formula in a discrete type setting.

[10] As individuals already reveal their type with their education decision, the government could actually also levy individualized lump sum taxes in period 2. In the spirit of the *Ramsey approach*, we constrain the labor income taxes to be linear, however.

$$\max_{t,S(\cdot)} \tilde{f}(\theta_1)\Bigg((1-t)w_1h(t,w_1)+T-\Psi[h(t,w_1)]-e_1+S(e_1)\Bigg)$$
$$+\tilde{f}(\theta_2)\Bigg((1-t)w_2h(t,w_2)+T-\Psi[h(t,w_2)]-e_1+S(e_2)\Bigg) \quad (4.4)$$

subject to a government budget constraint

$$T=t\Bigg(f_1w(e_1,\theta_1)h(t,w_1)+f_2w_2h(t,w_2)-f_1S(e_1)-f_2S(e_2)\Bigg)$$

and optimal individual behaviour

$$\forall\ i=1,2\ :\ (e_i,h_i)=\arg\max_{e,h}(1-t)w(e,\theta_i)h+T-\Psi(h)-e+S(e). \quad (4.5)$$

This problem has some similarities to the problem in Stiglitz (1982), where a non-linear tax schedule is chosen in an economy with two groups of individuals. By the revelation principle we can formulate the part of choosing $S(\cdot)$ as choosing e_1, e_2, c_1^1, c_2^1 directly, where c_1^1 and c_2^1 denote first period consumption.[11] In that case, we can replace (4.5) by

$$h(t,w_i)=\arg\max_h(1-t)hw_i-\Psi(h) \quad (4.6)$$

and an incentive compatibility constraint[12]

$$\begin{aligned}&c_2^1+(1-t)w_2h(t,w_2)-\Psi(h(t,w_2))\\ \geq &c_1^1+(1-t)w(e_1,\theta_2)h(t,w(e_1,\theta_2))-\Psi(h(t,w(e_1,\theta_2))).\end{aligned} \quad (4.7)$$

Notice that in the incentive constraint (4.7) the deviation utility on the right-hand-side, the terms $w(e_1,\theta_2)$ and $h(t,w(e_1,\theta_2))$, show up. A deviating high-skilled agent receives the education level of the low skilled agent e_1. The wage she receives differs from the wage of the low skilled agent because of the effect of innate abilities

[11] With linear utility, the timing of consumption across the first and the second period is not determined. This implies that there are alternative ways to state the problem. E.g., we dropped first period consumption in problem (4.4) without loss of generality.

[12] Since we assume $\tilde{f}(\theta_1) > f(\theta_1)$, we focus on downward redistributive taxation where only the incentive constraint of the θ_2-type is binding.

on wages. To keep notation simple we will call this $w(e_1, \theta_2) = w^c$, with a c for counterfactual as in equilibrium by incentive compatibility the wage will never be observed. We call the associated hour choice $h(t, w(e_1\theta_2)) = h^c$ and associated income $y^c = h^c w^c$.

The government's problem now reads as:

$$\max_{c_1^1, c_2^1, t, e_1, e_2} \tilde{f}(\theta_1)\Bigg(c_1^1 + (1-t)w_1 h(t, w_1) + T - \Psi[h(t, w_1)]\Bigg) + \tilde{f}(\theta_2)\Bigg(c_2^1 + (1-t)w_2 h(t, w_2) + T - \Psi[h(t, w_2)]\Bigg) \tag{4.8}$$

subject to a government budget constraint

$$T = t\Bigg(f_1 w_1 h(t, w_1) + f_2 w_2 h(t, w_2) - f_1(c_1^1 + e_1) - f_2(c_2^1 + e_2)\Bigg), \tag{4.9}$$

and subject to (4.6) and (4.7), where we denote as η the Lagrangian multiplier of the incentive compatibility constraint. The Lagrangian and the first-order conditions are stated in Appendix 4.A.

The solution (4.8) can then be implemented with a nonlinear subsidy function $S(\cdot)$ that has to yield the desired consumption levels, i.e. $S(e_1) = c_1^1 + e_1$ and $S(e_2) = c_2^1 + e_2$. In addition, we have to make sure that incentives for the level of education and labor supply are jointly optimal for the individual. This implies that – given the subsidy function – (4.5) has to hold. Naturally, infinitely many nonlinear subsidy schedules can implement the desired allocation, as in the nonlinear tax problem with two types of Stiglitz (1982). We will in the following be interested in those subsidy functions that are differentiable at e_1 and e_2. In these cases, we know that the first-order condition for education of an individual can be rearranged as:

$$(1 - S'(e(\theta_i))) = (1-t)\frac{\partial w_i}{\partial e} h_i \;\; \forall\, i = 1, 2.$$

In the following, we will therefore be interested in

$$s(\theta_i) \equiv 1 - (1-t)\frac{\partial w_i}{\partial e} h_i \;\; \forall\, i = 1, 2.$$

Having computed an optimal allocation, we can therefore infer the implicit marginal education subsidies $s(\theta_1)$ and $s(\theta_2)$ for this allocation. For simplicity, we will call $s(\theta_1)$ and $s(\theta_2)$ education subsidies in the remainder of this chapter.[13] Note also that throughout this chapter, we only characterize *marginal* subsidies and not *average* subsidies.

4.3.2 Optimal Tax and Education Policies

We start by characterizing the optimal linear income tax rate. As the following proposition shows, the optimal linear tax rate is corrected by the endogeneity of education as compared to the optimal tax rate with exogenous education in equation (4.3).

Proposition 4.3.1 *In a full-commitment economy, the optimal linear tax rate satisfies*

$$\frac{t^f}{1-t^f} = \frac{(\tilde{f}(\theta_1) - f(\theta_1))\left(\frac{y_2-y_1}{\bar{y}}\right) - \eta\left(\frac{y_2-y_2^c}{\bar{y}}\right)}{\epsilon},$$

where the multiplier satisfies $\eta = \tilde{f}(\theta_1) - f(\theta_1)$.

Proof: See Appendix 4.A.1 ■

The tax rate with endogenous education decisions is still increasing in income inequality and decreasing in the labor supply elasticity. As can be seen, there is an additional force given by $\eta\left(\frac{y_2-y_2^c}{\bar{y}}\right)$ in the numerator as compared to the case where education is taken as exogenous. It decreases the optimal tax rate, and the effect is stronger the bigger the difference $y_2 - y_2^c$. y_2^c is the income level that the high type θ_2 would attain when only taking the education level of the low type e_1. The difference, hence, captures the effect of a higher education level for the high type on her earnings. The more important the effect of education on earnings, the smaller the tax rate tends to be. Consider the one extreme case, where additional education does not change wages at all for the high-type, so $y_2 = y_2^c$. In this case, there is no need for the optimal tax rate to take into account education incentives, and the formula collapses to the case with exogenous human capital. In the other extreme case, we would have $y_1 = y_2^c$, so with the same education level both agents would

[13] As is in the optimal taxation problem with discrete types, we can always pick a nonlinear subsidy schedule such that the first-order conditions of an individuals are also sufficient and her problem is concave. In order to ensure that locally linear subsidy schedules implement the desired allocation, further assumptions on $w(\theta, e)$ have to e made, see, e.g., Bovenberg and Jacobs (2005, p.2010) for a discussion of that in a similar framework.

receive the same wage. This would essentially eliminate agent heterogeneity and the optimal tax rate would be zero in a model without risk. The following corollary summarizes the above reasoning.

Corollary 4.3.1 *Let e_1^* and e_2^* be the solution to the problem (4.8). Then the respective optimal linear tax rate is smaller than the linear tax rate as defined by (4.3) for $e_1 = e_1^*$ and $e_2 = e_2^*$, i.e. $t^f(e_1^*, e_2^*) < t^{ex}(e_1^*, e_2^*)$.*

Income taxes are not the only instrument of the government. Governments do rely on education subsidies to increase the incentives to invest into education.
We now characterize optimal education subsidies.

Proposition 4.3.2 *In a full-commitment economy, education subsidies satisfy*

$$s^f(\theta_1) = t^f \frac{\partial w_1}{\partial e_1} h_1(1+\epsilon) - \frac{\eta}{f(\theta_1)}(1-t^f)\left[h_2^c \frac{\partial w_2^c}{\partial e_1} - h_1 \frac{\partial w_1}{\partial e_1}\right]$$

and

$$s^f(\theta_2) = t^f \frac{\partial w_2}{\partial e_2} h_2(1+\epsilon).$$

Proof: See Appendix 4.A.2 ■
First, looking at the education subsidy for the low type one can see that there are two parts. The first term reflects the fiscal externality effect of private education decisions (Bovenberg and Jacobs, 2005): the education decision of individuals imposes an externality on the government budget as individuals with higher education pay higher taxes. The government internalizes this fiscal externality by subsidizing education in a Pigouvian way. As the formula reveals, the larger the labor supply elasticity is, the larger the subsidy. Intuitively, the stronger individuals' working hours react to wage increases, the larger is the fiscal externality on the government budget. Relatedly, the subsidy increases in the marginal return of education $\frac{\partial w_1}{\partial e_1}$ and in the income tax rate.
The second term captures the fact that innate abilities and education are complements (Jacobs and Bovenberg, 2011). The marginal return to education is increasing in innate ability. As the government is redistributive, there is a force towards lowering education subsidies, as they tend to profit more the initially high types. For the high type θ_2 only the fiscal externality part is present because a standard "no-distortion-at-the-top" result applies for the second part.

4.4 Optimal Policies without Commitment

We now characterize policies when the government has no commitment power and contrast them to the full-commitment results. We start backwards, looking at optimal tax policies, once education decisions are sunk.

4.4.1 The Problem in Period Two

The problem of the planner is basically equivalent to that of the planner in Section 4.2.2, as the distribution of wages is taken as exogenous. In particular, the same tax formula applies:

$$\frac{t^{nc}}{1-t^{nc}} = \frac{\left(\tilde{f}(\theta_1) - f(\theta_1)\right)\left(\frac{y_2-y_1}{\bar{y}}\right)}{\varepsilon}. \tag{4.10}$$

In the following, we write the optimal tax rate for the second period planner as a function of both education levels, so $t^{nc}(e_1, e_2)$.

4.4.2 The Problem in Period One

In the first period, the planner anticipates that he will set taxes according to (4.10). Therefore, in the first period, the problem reads as:

$$\begin{aligned}\max_{c_1,c_2,e_1,e_2} = \tilde{f}(\theta_1)\Bigg(& c_1^1 + (1-t^{nc}(e_1,e_2))w_1 h_1(t^{nc}(e_1,e_2),w_1) \\ &+ T - \Psi[h_1(t^{nc}(e_1,e_2),w_1)]\Bigg) \\ + \tilde{f}(\theta_2)\Bigg(& c_2^1 + (1-t^{nc}(e_1,e_2))w_2 h_2(t^{nc}(e_1,e_2),w_2) \\ &+ T - \Psi[h_2(t^{nc}(e_1,e_2),w_2)]\Bigg)\end{aligned}$$

subject to the government budget constraint

$$T = t^{nc}(e_1, e_2)\Bigg(f(\theta_1) w_1 h_1(t^{nc}(e_1, e_2), w_1) + (1 - f(\theta_1)) w_2 h_2(t^{nc}(e_1, e_2), w_2) - f(\theta_1) c_1^1 - f(\theta_2) c_2^1 - f(\theta_2) e_2 - f(\theta_1) e_1 \Bigg) \quad (4.11)$$

and the incentive constraint

$$c_2^1 + (1 - t^{nc}(e_1, e_2)) w_2 h_2 - \Psi(h_2) \geq c_1^1 + (1 - t^{nc}(e_1, e_2)) w_2^c h_2^c - \Psi(h_2^c).$$

This problem is very similar to the problem in Section 4.3. The difference is that the government cannot choose t but instead takes into account how it will choose t in the future once education decisions are sunk. The Lagrangian and first-order conditions are stated in Appendix 4.B. The following proposition shows how optimal optimal education subsidies are designed in a no-commitment economy.

Proposition 4.4.1 *In a no-commitment economy, education subsidies satisfy*

$$s^{nc}(\theta_1) = t^{nc} \frac{\partial w_1}{\partial e_1} h_1(1 + \epsilon) - \frac{\eta}{f(\theta_1)} \frac{\partial t^{nc}}{\partial e_1} (y_2 - y_2^c)$$

and

$$s^{nc}(\theta_2) = t^{nc} \frac{\partial w_2}{\partial e_2} h_2(1 + \epsilon) - \frac{\eta}{f(\theta_1)} \frac{\partial t^{nc}}{\partial e_2} (y_2 - y_2^c)$$

where $\frac{\partial t^{nc}}{\partial e_1} < 0$ *and* $\frac{\partial t^{nc}}{\partial e_2} > 0$ *and the multiplier satisfies* $\eta = \tilde{f}(\theta_1) - f(\theta_1)$.

Proof: See Appendix 4.B.1 ■

In comparison to the full-commitment case in Proposition 4.3.2, there is now an additional term in both formulas for the optimal education subsidy. In fact, for the low type, this additional term favors higher subsidies and for the high type, this additional term favors lower subsidies. Together this tends to make education policies more progressive.

The intuition behind this result is clear and simple. In the case of the low type the additional term is given by:

$$-\frac{\eta}{f(\theta_1)} \frac{\partial t^{nc}}{\partial e_1} (y_2 - y_2^c) > 0.$$

A higher education subsidy and a higher level of education for the low type will decrease the optimal tax rate chosen by the government in period two $\frac{\partial t^{nc}}{\partial e_1} < 0$. This will strengthen education incentives for both types. In other words, the government anticipates its temptation to set too high taxes in the second period. By compressing the distribution of eduction across the two agents, it can avoid some of the harmful spillover from too high taxes on the education margin. Consistent with that argument, there is a downward adjustment in the optimal subsidy for the high type

$$-\frac{\eta}{f(\theta_1)}\frac{\partial t^{nc}}{\partial e_2}(y_2 - y_2^c) < 0,$$

as a higher education level for the high type will tend to increase taxes because of higher income inequality.

4.5 Varying the Degree of Commitment

In the previous section we studied two polar cases. We now look at economies, where the degree of commitment power of the government is allowed to differ, nesting the two cases from the previous sections. This allows us to show that smoother versions of our previous results hold.

4.5.1 Costs of Deviating and the Commitment Technology

Following Farhi, Sleet, Werning, and Yeltekin (2012), we introduce output costs of deviation. This implies that the government lacks commitment and can always deviate from its announced tax rate. However, deviation will incur some output loss κ, which can be considered as a reduced form for a reputational loss. Farhi et al. (2012) show how to microfound such an output loss in a dynamic repeated game, where a deviation today brings a reputational cost borne in the future because of depressed investment of future generations.

Formally, this implies an additional credibility constraint on the government problem. It takes the form:

$$\mathcal{W}^2_{PC}(e_1, e_2, t) \geq \mathcal{W}^2_{Dev}(e_1, e_2) - \kappa, \tag{4.12}$$

where $\mathcal{W}^2_{PC}(e_1, e_2, t)$ is second period welfare as a function of education levels for both types and the promised tax rate t, under the assumption that the government sticks to its promise. $\mathcal{W}^2_{Dev}(e_1, e_2)$ on the other is the second period welfare obtained if the government reneges on its tax promise and effectively takes the education levels as exogenous as in Section 4.2.2.

This form of deviation costs allows to flexibly capture different levels of *limited commitment*. At the one extreme end, when κ is zero, there is no way for the government to credibly commit and we arrive at the case from Section 4.4. At the other extreme end, when κ is above some positive threshold $\bar{\kappa} > 0$, all tax promises are fully credible and we arrive at the full-commitment solution of Section 4.3, which naturally achieves the highest welfare. In this section we focus on the intermediate cases where κ lies between zero and $\bar{\kappa}$.

4.5.2 Optimal Policies and Discussion

In comparison to the full-commitment problem in Section 4.3, the government has to respect the credibility constraint (4.12) in addition to all other constraints. We denote the Lagrangian multiplier on this credibility constraint as ζ. The Lagrangian function and the first-order conditions are stated in Appendix 4.C. The following proposition shows the optimal income tax rate for this case.

Proposition 4.5.1 *In a partial-commitment economy, the optimal linear tax rate satisfies:*

$$\frac{t^{pc}}{1-t^{pc}} = \frac{\left(\tilde{f}(\theta_1) - f(\theta_1)\right)\left(\frac{y_2-y_1}{\bar{y}}\right) - \frac{\eta}{1+\zeta}\left(\frac{y_2-y_2^c}{\bar{y}}\right)}{\epsilon},$$

Proof: See Appendix 4.C.1 ∎

One can see how this case nests the full-commitment case, i.e. the optimal income tax rate from Proposition 4.3.1. If the credibility constraint is not binding for sufficiently high κ (hence $\kappa > \bar{\kappa}$), ζ is equal to zero and the government is able to implement the full-commitment tax rate. As discussed above, the second term in the numerator reflects how labor taxes are adjusted to provide education incentives and complement education subsidies. This effect is now scaled down by $\frac{1}{1+\zeta}$. The more severe the commitment problem, the bigger ζ tends to be. This will make any tax promises less credible and, anticipating this, the government will set a higher, more credible tax rate. Next, we characterize the resulting education subsidies.

Proposition 4.5.2 *In a partial-commitment economy, education subsidies satisfy:*

$$s^{pc}(\theta_1) = t^{pc}\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - \frac{\eta}{f(\theta_1)}(1-t^{pc})\left[h_2^c\frac{\partial w_2^c}{\partial e_1} - h_1\frac{\partial w_1}{\partial e_1}\right] + \frac{\zeta}{f(\theta_1)}\left(\frac{\partial \mathcal{W}_{PC}}{\partial e_1} - \frac{\partial \mathcal{W}_{Dev}}{\partial e_1}\right)$$

where $\frac{\partial \mathcal{W}_{PC}}{\partial e_1} - \frac{\partial \mathcal{W}_{Dev}}{\partial e_1} > 0$ *and*

$$s^{pc}(\theta_2) = t^{pc}\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) + \frac{\zeta}{f(\theta_2)}\left(\frac{\partial \mathcal{W}_{PC}}{\partial e_2} - \frac{\partial \mathcal{W}_{Dev}}{\partial e_2}\right).$$

where $\frac{\partial \mathcal{W}_{PC}}{\partial e_2} - \frac{\partial \mathcal{W}_{Dev}}{\partial e_2} < 0$.

Proof: See Appendix 4.C.2 ■

Whenever the credibility constraint is binding, the subsidies get adjusted by

$$\frac{\zeta}{f(\theta_1)}\left(\frac{\partial \mathcal{W}^2_{PC}}{\partial e_1} - \frac{\partial \mathcal{W}^2_{Dev}}{\partial e_1}\right) > 0$$

and

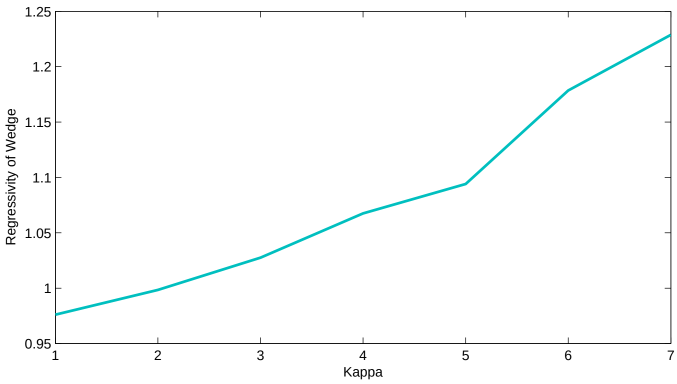
$$\frac{\zeta}{f(\theta_2)}\left(\frac{\partial \mathcal{W}^2_{PC}}{\partial e_2} - \frac{\partial \mathcal{W}^2_{Dev}}{\partial e_2}\right) < 0$$

respectively. This implies that whenever there is a commitment problem, the marginal value of low level education goes up as it strengthens the credibility of tax promises. Relatedly, the marginal value of high level education goes down as it increases the temptation to renege on tax promises and increase the tax rate to redistribute. Taken together, a more compressed education distribution leads to a more compressed wage distribution, decreasing the value of an ex-ante harmful deviation of the government. With limited commitment, the government wants to avoid excessive ex-post redistribution, by engaging already in ex-ante redistribution through the use of education policies. The larger the commitment problem, the larger ζ and the stronger this effect on the progressivity of education subsidies. In the next subsection, we illustrate with a numerical example how a decrease in κ leads to more progressive education subsidies.

4.5.3 Numerical Illustration

We assume an equal mass of high and low types $f(\theta_1) = f(\theta_2) = 0.5$ and set the welfare weights to $\tilde{f}(\theta_1) = 0.9$ and $\tilde{f}(\theta_2) = 0.1$. We set the labor supply elasticity to 0.25. For wages, we assume that they are determined by a simple Cobb Douglas production function $w_i = \theta_i^{0.5}e_i^{0.5}$ with equal weights. There is a constant marginal cost of education. We start by assuming that the government has limited commitment power and the cost of reneging on tax promises (κ) is set at 5% of output calculated from the full-commitment economy.

Figure 4.5.1: Regressivity of Education Subsidies



The equilibrium tax rate in the partial-commitment case is $t^{pc} = 35.56\%$. For comparison it is $t^{fc} = 19.56\%$ in the full-commitment case. This illustrates the workings of the formula in Proposition 4.5.1, as the human capital effect on taxes is scaled down relative to the full-commitment benchmark, because the government lacks full credibility. A deviating government which would take the wage distribution as exogenously given (Section 4.2.2) would set a tax rate of 65.95%. Thus, although the government lacks commitment power, it still takes human capital investment incentives into account as it sets a significantly smaller tax rate (about 30 percentage points).

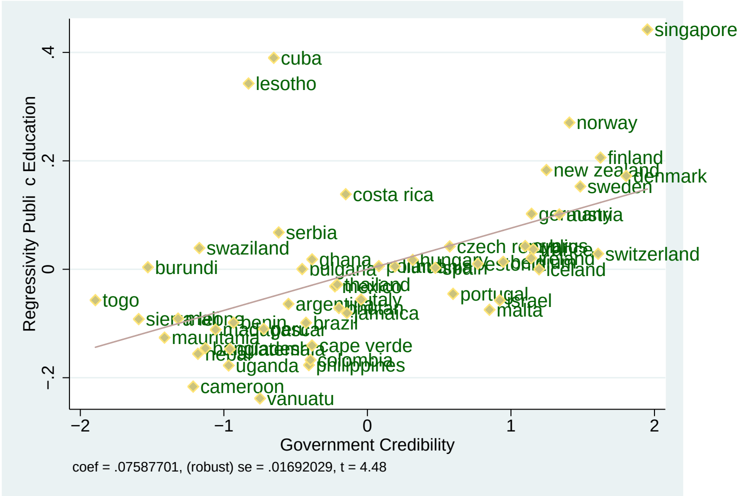
The main predictions from our analysis of education subsides concern the degree of the progressivity of education subsidies, see Proposition 4.5.2. In line with these predictions we find that the ratio of the subsidies $\frac{s^{pc}(\theta_2)}{s^{pc}(\theta_1)}$ of the high relative to the low type is 0.99, as compared to 2.38 in the full-commitment case. A higher $\frac{s^{pc}(\theta_2)}{s^{pc}(\theta_1)}$ ratio implies a more regressive incidence of subsidies.

Finally, we illustrate how the regressivity of education policies varies with the commitment technology. Figure 4.5.1 plots $\frac{s^{pc}(\theta_2)}{s^{pc}(\theta_1)}$ against κ as it varies, measured in percentage points of output lost when reneging on tax promises. Moving from left to right, the commitment power of the government gradually increases. In line with the mechanism outlined above, a government with more commitment power can afford to set less progressive subsidies as its credibility increases.

4.6 Empirical Implications

The model predicts a more regressive incidence of subsidies when the ability of a government to commit is high. We now provide suggestive cross-country evidence

Figure 4.6.1: Government Education Expenditures and Policy Credibility



for this. The estimates of this section should be interpreted as correlations only.

Data. As a measure for commitment power and the credibility of policy announcements we use data from the World Bank's Worldwide Governance Indicators database. We use the variable Government Effectiveness capturing "...the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies." (Kaufmann et al., 2010). To proxy for the regressivity/progressivity of public education subsidies, we use the share of public education expenditures at each educational level relative to total public education expenditures. We then take the share spent on tertiary education relative to the total spending share on all lower levels of education (primary and secondary). The bigger the value of this variable, the more regressive is the incidence of public education expenditure, in the sense that more is spent on tertiary education relative to lower tier education. To construct the measure, we take data from the UNESCO on public educational expenditures across education levels. We also use GDP data, which we take from the World Bank database. We use the year 2008 as the most recent year with a reasonable number of observations. We are left with a sample of 54 countries, for which all the relevant data is available.

Results. For Figure 4.6.1, we regress our measure of the regressivity of public education on the government credibility index. The correlation is positive and

Table 4.6.1: Credibility of Government and Education Regressivity

Dependent variable: Regressivity of Education Expenditure				
Policy Credibility	0.759***	1.226***	1.366***	1.252**
	(0.169)	(0.330)	(0.550)	(0.550)
Log GDP			- 0.020	- 0.021
			(0.030)	(0.032)
Total Education				0.031**
Expenditure Share				(0.013)
Continent Dummies	No	Yes	Yes	Yes
R-squared	0.290	0.352	0.356	0.490

Observations: 54. Year: 2008. List of countries see Appendix. Policy Credibility coefficient multiplied by 10. Robust errors. Last column based on 52 observation, since data on Bhutan and Uganda is missing.

highly significant.[14] The coefficient in column one of Table 1 implies that a one standard deviation increase in policy credibility increases the regressivity of public education expenditures by 0.53 standard deviations. Next, we include continent dummies. Only exploiting the variation within continents increases the credibility coefficient. Adding the log of per capita GDP does not affect the conclusion. Maybe surprisingly, income per capita seems not to be correlated with a more regressive incidence of public expenditure, as is seen in column three. The raw correlation between GDP per capita and our regressivity index is, however, positive and significant (0.49). But as the estimates indicate, this effect vanishes with continent dummies and controlling for government credibility. Finally, we control for the overall share of public education expenditures aggregated across all levels as a fraction of GDP. This approximates for the overall importance of the public sector in providing and paying for education. The main correlation concerning the effect of governmental policy credibility remains unaffected. As column four shows, countries in which the government has a relatively larger stake in education, tend to spend more on higher education.

[14]It is even stronger when excluding Lesotho and Cuba, which are two outliers with high regressivity but weak institutional commitment. On the other side, Singapore is an outlier with very strong policy credibility and a high incidence of regressivity.

4.7 Conclusion

Optimal income tax and education policies depend on the degree of commitment power or policy credibility the government has. We build a transparent and simple heterogeneous agent model to understand the economic mechanisms involved. Individual wages are determined by both innate abilities and education levels. Without any commitment, the labor income tax does not take into account the incentives to acquire education. When some or full commitment is available, income tax rates are adjusted to incentivize education. The tax rate is smaller, when the effect of education on wages is large relative to the effect of innate abilities on wages for the initially high skilled. We allow the government to subsidize different levels of education at different rates. The main implication of limited commitment is that education policies become more progressive relative to the full-commitment benchmark: the government takes into account that a more compressed wage distribution limits its own temptation to tax excessively. By adjusting the distribution of education, the government effectively creates its own commitment device. This mirrors previous findings from Farhi et al. (2012) concerning the design of capital taxes. Using data on the credibility of policy announcements from the World Bank database, we find a positive and significant correlation between the degree of commitment power and how regressive education expenditures are across countries, consistent with the mechanism highlighted in this chapter. This correlation is conditional on income and geographical controls.

Appendix 4.A The Full-Commitment Planner

We first substitute the government budget constraint into the problem. The Lagrangian function then reads as

$$\begin{aligned}\mathcal{L} =&\tilde{f}(\theta_1)\Bigg(c_1^1 + (1-t)w(e_1,\theta_1)h(t,w_1) - \Psi[h(t,w_1)]\Bigg)\\ &+ \tilde{f}(\theta_2)\Bigg(c_2^1 + (1-t)w_2(e_2,\theta_2)h(t,w_2) - \Psi[h(t,w_2)]\Bigg)\\ &+ t\Bigg(f_1 w(e_1,\theta_1)h(t,w_1) + (1-f_1)w(e_2,\theta_2)h(t,w_2) - f_1(c_1^1+e_1) - f_2(c_2^1+e_2)\Bigg)\\ &+ \eta\left(c_2^1 + (1-t)w_2 h(t,w_2) - \Psi(h(t,w_2)) - c_1^1 + (1-t)y^c - \Psi(h^c)\right)\end{aligned}$$

The first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial c_1^1} = \tilde{f}_1 - f_1 - \eta = 0$$

$$\frac{\partial \mathcal{L}}{\partial c_2^1} = \tilde{f}_2 - f_2 + \eta = 0$$

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial t} =& -\tilde{f}(\theta_1)y_1 - (1-\tilde{f}(\theta_1))y_2 + f(\theta_1)y_1 + (1-f(\theta_1))y_2 + tf(\theta_1)w_1\frac{\partial h_1}{\partial t}\\ &+ t(1-f(\theta_1))w_2\frac{\partial h_2}{\partial t} - \eta(w_2h_2 - w_2^c h_2^c) = 0\end{aligned} \tag{4.13}$$

$$\frac{\partial \mathcal{L}}{\partial e_1} = \tilde{f}(\theta_1)(1-t)\frac{\partial w_1}{\partial e}h_1 + tf(\theta_1)\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - f(\theta_1) + \eta\left[-(1-t)h_2^c\frac{\partial w_2^c}{\partial e_1}\right] = 0,$$

$$\frac{\partial \mathcal{L}}{\partial e_2} = \tilde{f}(\theta_2)(1-t)\frac{\partial w_2}{\partial e}h_2 + tf(\theta_2)\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - f(\theta_2) + \eta(1-t)\frac{\partial w_2}{\partial e}h_2 = 0.$$

From the FOC for c_1^1, one directly obtains $\eta = \tilde{f}_1(\theta) - f_1(\theta)$.

4.A.1 Proof of Proposition 4.3.1

Two manipulations of (4.13) yield

$$\left(\tilde{f}_1 - f_1\right)(y_2 - y_1) - \frac{t}{1-t}\left(f_1 w_1 h_1 \frac{\partial h_1}{\partial 1-t}\frac{1-t}{h_1} + f_2 w_2 h_2 \frac{\partial h_2}{\partial 1-t}\frac{1-t}{h_2}\right) - \eta(y^c - y_1).$$

Now use $\varepsilon = \frac{\partial h_2}{\partial 1-t}\frac{1-t}{h_2} = \frac{\partial h_1}{\partial 1-t}\frac{1-t}{h_1}$ and $\bar{y} = f_1 y_1 + f_2 y_2$ and solve for $\frac{t}{1-t}$ to obtain the result.

4.A.2 Proposition 4.3.2

We start with the θ_2 type. Rewriting the FOC for e_2 yields

$$\tilde{f}(\theta_2)(1-t)\frac{\partial w_2}{\partial e}h_2 + tf(\theta_2)\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - f(\theta_2) + \left(f_2 - \tilde{f}_2\right)(1-t)\frac{\partial w_2}{\partial e}h_2 = 0,$$

which yields

$$tf(\theta_2)\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - f(\theta_2) + f_2(1-t)\frac{\partial w_2}{\partial e}h_2 = 0.$$

This can be rewritten as

$$f_2(1-t)\frac{\partial w_2}{\partial e}h_2 = 1 - t\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon),$$

where the RHS is the definition of the implicit education subsidy for the θ_2-type. Now we look at the θ_1-type. Rewriting the FOC for e_1 yields:

$$\tilde{f}(\theta_1)(1-t)\frac{\partial w_1}{\partial e}h_1 + tf(\theta_1)\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - f(\theta_1) + \eta\left[-(1-t)h_2^c\frac{\partial w_2^c}{\partial e_1}\right]$$
$$+ \eta(1-t)h_1\frac{\partial w_1}{\partial e_1} - \eta(1-t)h_1\frac{\partial w_1}{\partial e_1} = 0.$$

Now use $\eta = \tilde{f}_1 - f_1$ and obtain

$$tf(\theta_1)\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - f(\theta_1) + \eta(1-t)\left[h_1\frac{\partial w_1}{\partial e_1} - h_2^c\frac{\partial w_2^c}{\partial e_1}\right] - f_1(1-t)h_1\frac{\partial w_1}{\partial e_1} = 0.$$

Rearranging and again using the definition of the implicit subsidy yields the result.

Appendix 4.B The No-Commitment Planner

We first substitute the government budget constraint into the problem. The Lagrangian function then reads as

$$\begin{aligned}\mathcal{L} =& \tilde{f}(\theta_1)\left(c_1^1 + (1-t^{nc}(e_1,e_2))w_1(e_1,\theta_1)h_1(t^{nc}(e_1,e_2),w_1) - \Psi[h_1(t^{nc}(e_1,e_2),w_1)]\right)\\ &+ \tilde{f}(\theta_2)\left(c_2^1 + (1-t^{nc}(e_1,e_2))w_2(e_2,\theta_2)h_2(t^{nc}(e_1,e_2),w_2) - \Psi[h_2(t^{nc}(e_1,e_2),w_2)]\right)\\ &+ t^{nc}(e_1,e_2)\Big(f(\theta_1)w_1(e_1,\theta_1)h_1(t^{nc}(e_1,e_2),w_1)\\ &+ (1-f(\theta_1))w_2(e_2,\theta_2)h_2(t^{nc}(e_1,e_2),w_2) - f(\theta_1)c_1^1 - f(\theta_2)c_2^1 - f(\theta_2)e_2 - f(\theta_1)e_1\Big)\\ &+ \eta\left(c_2^1 + (1-t^{nc}(e_1,e_2))w_2h_2 - \Psi(h_2) - c_1^1 + (1-t^{nc}(e_1,e_2))w_2^c h_2^c - \Psi(h_2^c)\right)\end{aligned}$$

For the first-order condition for e_1 and e_2, we know that their impact on Period 2 welfare via t is zero due to the envelope theorem. Thus the first-order conditions read as:

$$\frac{\partial \mathcal{L}}{\partial e_1} = \tilde{f}(\theta_1)(1-t)\frac{\partial w_1}{\partial e}h_1 + tf(\theta_1)\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - f(\theta_1) + \eta\left[-(1-t^F)h_2^c\frac{\partial w_2^c}{\partial e_1}\right] - \eta\frac{\partial t}{\partial e_1}\left[w_2h_2 - w_c^c h_2^c\right] = 0,$$

$$\frac{\partial \mathcal{L}}{\partial e_2} = \tilde{f}(\theta_2)(1-t)\frac{\partial w_2}{\partial e}h_2 + tf(\theta_2)\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - f(\theta_2) + \eta(1-t)\frac{\partial w_2}{\partial e}h_2 - \eta\frac{\partial t}{\partial e_2}\left[w_2h_2 - w_c^c h_2^c\right] = 0.$$

4.B.1 Proof of Proposition 4.4.1

The optimal education subsidies can be obtained by some analytical manipulations almost equivalently as in Appendix 4.A.2.

We now look at the derivatives of t with respect to e_1 and e_2. We know that:

$$\frac{t}{1-t} = \frac{(\tilde{f}(\theta_1) - f(\theta_1))\left[\frac{w_2h_2 - w_1h_1}{\overline{\overline{wh}}}\right]}{\epsilon}.$$

We now show that t is increasing in e_2 and decreasing in e_1. Define the implicit function:

$$F(e_1,e_2,t(e_1,e_2)) = \frac{t}{1-t} - \frac{(\tilde{f}(\theta_1) - f(\theta_1))\left[\frac{w_2h_2 - w_1h_1}{\overline{\overline{wh}}}\right]}{\epsilon} = 0.$$

As $F = 0$ for any (e_1,e_2), the derivatives of F w.r.t to e_1 and e_2 have to be zero as well. In general these derivates are characterized by

$$\frac{\partial F}{\partial t}\frac{\partial t}{\partial e_i} + \frac{\partial F}{\partial e_i} = 0$$

and therefore can reveal the sign of $\frac{\partial t}{\partial e_i}$. Spelling this out for e_1 yields

$$\frac{\partial t}{\partial e_1}\left[\frac{1}{(1-t)^2}-\frac{(\tilde{f}(\theta_1)-f(\theta_1))}{\epsilon\overline{y}^2}\left(\left(\frac{\partial y_2}{\partial t}-\frac{\partial y_1}{\partial t}\right)\overline{y}+\left(f(\theta_1)\frac{\partial y_1}{\partial t}+f(\theta_2)\frac{\partial y_2}{\partial t}\right)(y_2-y_1)\right)\right] + \frac{(\tilde{f}(\theta_1)-f(\theta_1))}{\epsilon\overline{y}^2}\frac{\partial y_1}{\partial e_1}\left((y_2-y_1)f(\theta_1)+\bar{y}\right)=0$$

and hence

$$\frac{\partial t}{\partial e_1}\left[\frac{1}{(1-t)^2}-\frac{(\tilde{f}(\theta_1)-f(\theta_1))}{\epsilon\overline{y}^2}\underbrace{\left(\left(\frac{\partial y_2}{\partial t}-\frac{\partial y_1}{\partial t}\right)\overline{y}+\left(f(\theta_1)\frac{\partial y_1}{\partial t}+f(\theta_2)\frac{\partial y_2}{\partial t}\right)(y_2-y_1)\right)}_{\equiv X}\right] + \frac{(\tilde{f}(\theta_1)-f(\theta_1))}{\epsilon\overline{y}^2}\frac{\partial y_1}{\partial e_1}y_2=0.$$

For X we obtain:

$$\begin{aligned}&\left(\frac{\partial y_2}{\partial t}-\frac{\partial y_1}{\partial t}\right)\overline{y}-\left(f(\theta_1)\frac{\partial y_1}{\partial t}+f(\theta_2)\frac{\partial y_2}{\partial t}\right)(y_2-y_1)\\&=\frac{\partial y_1}{\partial t}(-f(\theta_1)y_1-(1-f(\theta_1))y_2-f(\theta_1)y_2+f(\theta_1)y_1)+\frac{\partial y_2}{\partial t}(f(\theta_1)y_1\\&+(1-f(\theta_1))y_2-(1-f(\theta_1))y_2+(1-f(\theta_1))y_1))\\&=\frac{\partial y_1}{\partial t}(-y_2)+\frac{\partial y_2}{\partial t}y_1=\frac{\partial y_1}{\partial 1-t}y_2-\frac{\partial y_2}{\partial 1-t}y_1=\varepsilon\frac{y_1y_2}{1-t}-\varepsilon\frac{y_1y_2}{1-t}=0,\end{aligned}$$

so it follows $\frac{\partial t}{\partial e_1}<0$. Similar reasoning shows $\frac{\partial t}{\partial e_2}>0$.

Appendix 4.C The Partial-Commitment Planner

We first substitute the government budget constraint into the problem. The Lagrangian function then reads as

$$\begin{aligned}\mathcal{L}=&\tilde{f}(\theta_1)\left(c_1^1+(1-t)w(e_1,\theta_1)h(t,w_1)-\Psi[h(t,w_1)]\right)\\&+\tilde{f}(\theta_2)\left(c_2^1+(1-t)w_2(e_2,\theta_2)h(t,w_2)-\Psi[h(t,w_2)]\right)\\&+t\left(f_1w(e_1,\theta_1)h(t,w_1)+(1-f_1)w(e_2,\theta_2)h(t,w_2)-f_1(c_1^1+e_1)-f_2(c_2^1+e_2)\right)\\&+\eta\left(c_2^1+(1-t)w_2h(t,w_2)-\Psi(h(t,w_2))-c_1^1+(1-t)y^c-\Psi(h^c)\right)\\&+\zeta\left(\mathcal{W}_{PC}^2(e_1,e_2,t)-\mathcal{W}_{Dev}^2(e_1,e_2)+\kappa,\right)\end{aligned}$$

where

$$\tilde{f}_1(1-t)w_1h_1 - \Psi(h_1) + \tilde{f}_2(1-t)w_2h_2 -$$
$$\left(\tilde{f}_1(1-t^d)w_1h_1(w_1,t^d) - \Psi(h_1(w_1,t^d)) + \tilde{f}_2(1-t)w_2h_2(w_2,t^d)\right)$$
$$+ t\,(w_1h_1f_1 + w_2h_2f_2) - t^d\left(w_1h_1(w_1,t^d)f_1 + w_2h_2(w_2,t^d)f_2\right).$$

For c_1 and c_2 we get the same FOC as in the full-commitment case. For t we get:

$$\frac{\partial \mathcal{L}}{\partial t} = (1+\zeta)\left(-\tilde{f}(\theta_1)y_1 - (1-\tilde{f}(\theta_1))y_2 + f(\theta_1)y_1 + (1-f(\theta_1))y_2 + tf(\theta_1)w_1\frac{\partial h_1}{\partial t} + t(1-f(\theta_1))w_2\frac{\partial h_2}{\partial t}\right) - \eta(w_2h_2 - w_2^c h_2^c) = 0$$

$$\frac{\partial \mathcal{L}}{\partial e_1} = \tilde{f}(\theta_1)(1-t)\frac{\partial w_1}{\partial e}h_1 + tf(\theta_1)\frac{\partial w_1}{\partial e_1}h_1(1+\epsilon) - f(\theta_1) + \eta\left[-(1-t^F)h_2^c\frac{\partial w_2^c}{\partial e_1}\right]$$
$$+\zeta\frac{\partial w_1}{\partial e_1}\Bigg[\tilde{f}_1\left((1-t^{PC})h_1(e_1,t^{PC}) - (1-t^{Dev})h_1(e_1,t^{Dev})\right)$$
$$+ f_1\left(t^{PC}h_i(e_1,t^{PC}) - t^{Dev}h_1(e_1,t^{Dev})\right)(1+\varepsilon_{h,w})\Bigg] = 0$$

$$\frac{\partial \mathcal{L}}{\partial e_2} = \tilde{f}(\theta_2)(1-t)\frac{\partial w_2}{\partial e}h_2 + tf(\theta_2)\frac{\partial w_2}{\partial e_2}h_2(1+\epsilon) - f(\theta_2) + \eta(1-t)\frac{\partial w_2}{\partial e}h_2$$
$$+\zeta\frac{\partial w_2}{\partial e_2}\Bigg[\tilde{f}_2\left((1-t^{PC})h_2(e_2,t^{PC}) - (1-t^{Dev})h_2(e_2,t^{Dev})\right)$$
$$+ f_2\left(t^{PC}h_2(e_2,t^{PC}) - t^{Dev}h_1(e_2,t^{Dev})\right)(1+\varepsilon_{h,w})\Bigg] = 0.$$

4.C.1 Proof of Proposition 4.5.1

Dividing the FOC for T by $1+\zeta$ directly shows that the FOC is equivalent to the one in Appendix 4.A; the only difference is that η is now replaced by $\frac{\eta}{1+\zeta}$. The proof is then equivalent to the proof in Appendix 4.A.1

4.C.2 Proof of Proposition 4.5.2

The FOC for e_1 and e_2 are equivalent to those in Appendix 4.A.2 apart from the additional terms multiplied by ζ. The steps are, however, the same as in 4.A.2 and the additional terms multiplied with ζ then appear in the education subsidy formula as well.

These additional terms in the formula for the education subsidy read as:

$$\zeta\frac{\partial w_i}{\partial e_i}\Bigg[\frac{\tilde{f}_i}{f_i}\left((1-t^{PC})h_i(e_i,t^{PC})-(1-t^{Dev})h_i(e_i,t^{Dev})\right)$$
$$+\left(t^{PC}h_i(e_i,t^{PC})-t^{Dev}h_i(e_i,t^{Dev})\right)(1+\varepsilon_{h,w})\Bigg]. \tag{4.14}$$

By assumption we have $\frac{\tilde{f}(\theta_1)}{f(\theta_1)}>1$ and $\frac{\tilde{f}(\theta_2)}{f(\theta_2)}<1$. In what follows we will write RF_i for $\frac{\tilde{f}_i}{f_i}$ to denote the relative Pareto weight and save on notation. We also simplify the notation for h and write $h_i(e_i,t^{Dev})=h_i^{dev}$ and similarly for the other expressions. Then (4.14) can be rearranged as:

$$\zeta\frac{\partial w_i}{\partial e_i}\left(\left(h_i^{pc}-h_i^{dev}\right)-\left[t^{dev}h_i^{dev}-t^{pc}h_i^{pc}\right]\left[\frac{1+\varepsilon_{h,w}}{RF_i}-1\right]\right). \tag{4.15}$$

The sign of this term is equivalent to the sign of:

$$\frac{h_i^{pc}-h_i^{dev}}{t^{dev}h_i^{dev}-t^{pc}h_i^{pc}}-\left[\frac{1+\varepsilon_{h,w}}{RF_i}-1\right] \tag{4.16}$$

if $h(t,w)t$ is increasing in t (which implies $t^{dev}h_i^{dev}-t^{pc}h_i^{pc}>0$). The latter is the case if $\varepsilon_{h,t}>-1$. Note that $\varepsilon_{h,t}=-\frac{t}{1-t}\varepsilon$. One can show that for the Laffer tax rate, we have $\frac{t}{1-t}=\frac{1}{\varepsilon}$. As we are below the Laffer rate, we get $\varepsilon_{h,t}>-\frac{\varepsilon}{\varepsilon}=-1$. Thus, $h(t,w)t$ is increasing in t in the cases we consider. Since $h_i=(w_i(1-t))^{\varepsilon}$, (4.16) is $>(<)0$ if:

$$(1-t)^{\varepsilon}-\left(1-t^d\right)^{\varepsilon}>(<)\left(\frac{1+\varepsilon}{RF_i}-1\right)\left(t^d\left(1-t^d\right)^{\varepsilon}-t\left(1-t\right)^{\varepsilon}\right)$$

which is $>(<)0$ whenever

$$H(t)\equiv(1-t)^{\varepsilon}\left(1+t\frac{1+\varepsilon}{RF_i}-t\right)>(<)\left(1-t^d\right)^{\varepsilon}\left(1+t^d\frac{1+\varepsilon}{RF_i}-t^d\right)=H(t^d).$$

We now have to show that $H(t)>H(t^d)$ for the low type and $H(t)<H(t^d)$ for the high type. We therefore take the derivative:

$$H'(t)=-\varepsilon(1-t)^{\varepsilon-1}\left(1+t\frac{1+\varepsilon}{RF_i}-t\right)+(1-t)^{\varepsilon}\left(\frac{1+\varepsilon}{RF_i}-1\right)=(1-t)^{\varepsilon}\left(\frac{1+\varepsilon}{RF_i}-1-\varepsilon\frac{1+t\frac{1+\varepsilon}{RF_i}-t}{1-t}\right).$$

We need to show that it is <0 for the low type and >0 for the high type, which is equivalent to

$$(1-t)\frac{1+\varepsilon}{RF_i}-(1-t)-\varepsilon\left(1+t\frac{1+\varepsilon}{RF_i}-t\right)<(>)0$$

respectively, which is equivalent to

$$(1-t)(1+\varepsilon)-(1-t)RF_i-RF_i\varepsilon-t(1+\varepsilon)\varepsilon+tRF_i\varepsilon<(>)0$$

and therefore

$$(1-t)(1+\varepsilon)-t(1+\varepsilon)\varepsilon < (>) RF_i\left((1-t)+\varepsilon-t\varepsilon\right)$$

which yields

$$RF_i > (<) \frac{(1-t)(1+\varepsilon)-t(1+\varepsilon)\varepsilon}{(1-t)(\varepsilon+1)} = 1-\frac{t}{1-t}\varepsilon.$$

As $RF_1 > 1$, we directly see that this condition is always fullfilled for the θ_1-type. Importantly, it is fulfilled for any $t > 0$ and therefore we know that $H(t^d) < H(t)$ for the low type.
How about our result for the θ_2-type? We need

$$RF_2 < \frac{(1-t)(1+\varepsilon)-t(1+\varepsilon)\varepsilon}{(1-t)(\varepsilon+1)} = 1-\frac{t}{1-t}\varepsilon$$

and

$$RF_2 < \frac{(1-t)(1+\varepsilon)-t(1+\varepsilon)\varepsilon}{(1-t)(\varepsilon+1)} = 1-\frac{t^d}{1-t^d}\varepsilon.$$

If both of these inequalities are fulfilled we can be sure that $H(t^d) > H(t)$ for the high type. Since $t^d > t$, the second is the stricter requirement. Inserting the formula for $\frac{t^d}{1-t^d}$ yields:

$$\frac{\tilde{f}_2}{f_2} < 1-(f_2-\tilde{f}_2)\underbrace{\frac{y_2-y_1}{\bar{y}}}_{\equiv A}.$$

This is equivalent to

$$\tilde{f}_2(1-f_2A) < f_2(1-f_2A).$$

Thus, whenever $1-f_2A > 0$, we have our result. Term A can be written as

$$A = \frac{w_2^{1+\varepsilon}(1-t)^{\varepsilon}-w_1^{1+\varepsilon}(1-t)^{\varepsilon}}{f_1w_1^{1+\varepsilon}(1-t)^{\varepsilon}+f_2w_2^{1+\varepsilon}(1-t)^{\varepsilon}} = \frac{w_2^{1+\varepsilon}-w_1^{1+\varepsilon}}{f_1w_1^{1+\varepsilon}+f_2w_2^{1+\varepsilon}}.$$

$f_2A < 1$ therefore implies

$$f_2w_2^{1+\varepsilon}-f_2w_1^{1+\varepsilon} < f_1w_1^{1+\varepsilon}+f_2w_2^{1+\varepsilon}$$

and hence

$$-f_2w_1^{1+\varepsilon} < f_1w_1^{1+\varepsilon}$$

which is always fulfilled.

4.C.3 List of Countries in the Empirical Part

Argentina, Austria, Bangladesh, Belgium, Benin, Bhutan, Brazil, Bulgaria, Burundi, Cameroon, Cape Verde, Colombia, Costa Rica, Cuba, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ghana, Guatemala, Hungary, Iceland, Ireland, Israel, Italy, Jamaica, Lesotho, Lithuania, Madagascar, Mali, Malta, Mauritania, Mexico, Nepal, New Zealand, Norway, Peru, Philippines, Poland, Portugal, Serbia, Sierra Leone, Singapore, Spain, Swaziland, Sweden, Switzerland, Thailand, Togo, Uganda.

Bibliography of Chapter 4

Abramitzky, R. and V. Lavy (2012): "How Responsive is Investment in Schooling to Changes in Redistributive Policies and in Returns?" *Working Paper, Stanford University.*

Andersson, F. and K. A. Konrad (2003): "Human capital investment and globalization in extortionary states," *Journal of Public Economics*, 87, 1539–1555.

Boadway, R., N. Marceau, and M. Marchand (1996): "Investment in education and the time inconsistency of redistributive tax policy," *Economica*, 171–189.

Bovenberg, L. and B. Jacobs (2005): "Redistribution and Education Subsidies are Siamese Twins," *Journal of Public Economics*, 89, 2005–2035.

Card, D. (1999): "The causal effect of education on earnings," *Handbook of labor economics*, 3, 1801–1863.

Carneiro, P. and J. Heckman (2005): " Human Capital Policies," in *Inequality in America: What Role for Human Capital Policies*, ed. by J. Heckman and A. Krueger, The MIT Press, 77–240.

Farhi, E., C. Sleet, I. Werning, and S. Yeltekin (2012): "Non-linear Capital Taxation Without Commitment," *The Review of Economic Studies*, 79, 1469–1493.

Hendricks, L. and T. Schoellman (2012): "Student Abilities During the Expansion of US Education," *Working Paper.*

Jacobs, B. and A. Bovenberg (2011): "Optimal taxation of human capital and the earnings function," *Journal of Public Economic Theory*, 13, 957–971.

KAUFMANN, D., A. KRAAY, AND M. MASTRUZZI (2010): "The worldwide governance indicators: methodology and analytical issues," *World Bank Policy Research Working Paper.*

KONRAD, K. A. (2001): "Privacy and time-consistent optimal labor income taxation," *Journal of Public Economics*, 79, 503–519.

KYDLAND, F. E. AND E. C. PRESCOTT (1977): "Rules rather than discretion: The inconsistency of optimal plans," *The Journal of Political Economy*, 473–491.

LEMIEUX, T. (2006): "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?" *The American Economic Review*, 461–498.

PEREIRA, J. (2009): "Essays on Time-Consistent Fiscal Policy," Ph.D. thesis, European University Institute.

POUTVAARA, P. (2003): "Investment in education and redistributive taxation without precommitment," *FinanzArchiv/Public Finance Analysis*, 177–188.

——— (2011): "The expansion of higher education and time-consistent taxation," *European Journal of Political Economy*, 27, 257–267.

RICHTER, W. F. (2009): "Taxing education in Ramsey's tradition," *Journal of Public Economics*, 93, 1254–1260.

SHESHINSKI, E. (1972): "The optimal linear income-tax," *The Review of Economic Studies*, 39, 297–302.

STANTCHEVA, S. (2013): "Optimal Income Taxation with Adverse Selection in the Labor Market," *MIT Department of Economics Graduate Student Research Paper.*

STIGLITZ, J. E. (1982): "Self-selection and Pareto efficient taxation," *Journal of Public Economics*, 17, 213–240.

TABER, C. (2001): "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?" *Review of Economic Studies*, 68, 665–691.

TROSTEL, P. A. (1993): "The effect of taxation on human capital," *Journal of Political Economy*, 327–350.

Complete Bibliography

ABRAMITZKY, R. AND V. LAVY (2012): "How Responsive is Investment in Schooling to Changes in Redistributive Policies and in Returns?" *Working Paper, Stanford University.*

ACEMOGLU, D. AND D. AUTOR (2011): "Skills, Tasks and Technologies: Implications for Employment and Earnings," in *Handbook of Labor Economics Vol. 4*, ed. by O. Ashenfelter and D. Card, Elsevier, 1043–1171.

AIYAGARI, S. R. (1994): "Uninsured idiosyncratic risk and aggregate saving," *The Quarterly Journal of Economics*, 109, 659–684.

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

ANDERBERG, D. (2009): "Optimal Policy and the Risk Properties of Human Capital Reconsidered," *Journal of Public Economics*, 93, 1017–1026.

ANDERBERG, D. AND F. ANDERSSON (2003): "Investments in Human Capital, Wage Uncertainty, and Public Policy," *Journal of Public Economics*, 87, 1521 – 1537.

ANDERSSON, F. AND K. A. KONRAD (2003): "Human capital investment and globalization in extortionary states," *Journal of Public Economics*, 87, 1539–1555.

ATKINSON, A. AND J. STIGLITZ (1976): "The Design of Tax Structure: Direct versus Indirect Taxation," *Journal of Public Economics*, 6, 55–75.

ATKINSON, A. B., T. PIKETTY, AND E. SAEZ (2011): "Top Incomes in the Long Run of History," *Journal of Economic Literature*, 49, 3–71.

BACH, S., G. CORNEO, AND V. STEINER (2012): "Optimal top marginal tax rates under income splitting for couples," *European Economic Review*, 56, 1055 – 1069.

BANKS, J. AND P. DIAMOND (2011): "The Base for Direct Taxation," in *Dimensions of Tax Design. The Mirrlees Review*, Oxford University Press, chap. 6, 548–648.

BARGAIN, O., M. DOLLS, D. NEUMANN, A. PEICHL, AND S. SIEGLOCH (2011): "Tax-benefit systems in Europe and the US: between equity and efficiency," *CESifo Working Paper 3534*.

BASTANI, S., S. BLOMQUIST, AND L. MICHELETTO (2011): "The welfare gains of age related optimal income taxation," *International Economic Review (forthcoming)*.

BEST, M. AND H. KLEVEN (2013): "Optimal Income Taxation with Career Effects of Work Effort," *Working Paper, London School of Economics*.

BLUNDELL, R., M. BREWER, P. HAAN, AND A. SHEPHARD (2009): "Optimal Income Taxation of Lone Mothers: An Empirical Comparison Of The UK and Germany," *The Economic Journal*, 119, 101–121.

BLUNDELL, R. AND A. SHEPHARD (2012): "Employment, Hours of Work and the Optimal Taxation of Low-Income Families," *The Review of Economic Studies*, 79, 481–510.

BOADWAY, R., N. MARCEAU, AND M. MARCHAND (1996): "Investment in education and the time inconsistency of redistributive tax policy," *Economica*, 171–189.

BOHACEK, R. AND M. KAPICKA (2008): "Optimal Human Capital Policies," *Journal of Monetary Economics*, 55, 1–16.

BOONE, J. AND L. BOVENBERG (2004): "The optimal taxation of unskilled labor with job search and social assistance," *Journal of Public Economics*, 88, 2227–2258.

BOVENBERG, L. AND B. JACOBS (2005): "Redistribution and Education Subsidies are Siamese Twins," *Journal of Public Economics*, 89, 2005–2035.

BRETT, C. AND J. A. WEYMARK (2003): "Financing education using optimal redistributive taxation," *Journal of Public Economics*, 87, 2549–2569.

CARD, D. (1999): "The causal effect of education on earnings," *Handbook of labor economics*, 3, 1801–1863.

Carneiro, P. and J. Heckman (2005): “ Human Capital Policies,” in *Inequality in America: What Role for Human Capital Policies*, ed. by J. Heckman and A. Krueger, The MIT Press, 77–240.

Chapman, B. (2006): “Income contingent loans for higher education: International reforms,” *Handbook of the Economics of Education*, 2, 1435–1503.

Chen, S. (2008): “Estimating the Variance of Wages in the Presence of Selection and Unobserved Heterogeneity,” *Review of Economics and Statistics*, 90, 275–289.

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

——— (2009): “Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods,” *Annual Review of Economics*, 1, 451–488.

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

Chetty, R., A. Guren, D. Manoli, and A. Weber (2011): “Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins,” *American Economic Review Papers and Proceedings*, 101, 471–475.

Choné, P. and G. Laroque (2011a): “Negative marginal tax rates and heterogeneity,” *American Economic Review*, 100, 2532–47.

——— (2011b): “Optimal Taxation in the Extensive Model,” *Journal of Economic Theory*, 146, 425–453.

Christiansen, V. (2012): “Optimal Participation Taxes,” *CESifo Working Paper No. 3941.*

Conesa, J. C., S. Kitao, and D. Krueger (2009): “Taxing Capital? Not a Bad Idea after All!” *The American Economic Review*, 25–48.

Courty, P. and H. Li (2000): “Sequential Screening,” *The Review of Economic Studies*, 67, 697–717.

Cremer, H., F. Gahvari, and J. Lozachmeur (2010): “Tagging and Income Taxation: Theory and an Application,” *American Economic Journal: Economic Policy*, 2, 31–50.

CREMER, H., J.-M. LOZACHMEUR, AND P. PESTIEAU (2012): "Income taxation of couples and the tax unit choice," *Journal of Population Economics*, 25, 763–778.

CUNHA, F. AND J. HECKMAN (2007): "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling," *Labour Economics*, 14, 870–893.

——— (2008): "The Evolution of Labor Earnings Risk in the US Economy," *Working Paper, University of Chicago.*

DA COSTA, C. AND T. PEREIRA (2011): "Sacrifice and Efficiency of the Income Tax Schedule," *Escola de Pós-Graduação em Economia da FGV.*

DA COSTA, C. E. AND L. J. MAESTRI (2007): "The risk Properties of Human Capital and the Design of Government Policies," *European Economic Review*, 51, 695–713.

DIAMOND, P. (1980): "Income Taxation with Fixed Hours of Work," *Journal of Public Economics*, 13, 101–110.

——— (1998): "Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 88, 83–95.

DIAMOND, P. AND J. SPINNEWIJN (2011): "Capital income taxes with heterogeneous discount rates," *American economic journal: economic policy*, 3, 52–76.

DIAMOND, P. A. AND J. A. MIRRLEES (1978): "A model of Social Insurance with Variable Retirement," *Journal of Public Economics*, 10, 295–336.

DIAMOND, P. A. AND E. SAEZ (2011): "The Case for a Progressive Tax: From Basic Research to Policy Recommendations," *Journal of Economic Perspectives*, 25, 165–90.

EATON, J. AND H. S. ROSEN (1980): "Taxation, human capital, and uncertainty," *The American Economic Review*, 70, 705–715.

EISSA, N. AND J. B. LIEBMAN (1996): "Labor Supply Response to the Earned Income Tax Credit," *The Quarterly Journal of Economics*, 111, 605–637.

FAN, J. AND I. GIJBELS (2003): *Local Polynomial Modelling and Its Applications*, Boca Raton: CRC Press.

FARHI, E., C. SLEET, I. WERNING, AND S. YELTEKIN (2012): "Non-linear Capital Taxation Without Commitment," *The Review of Economic Studies*, 79, 1469–1493.

FARHI, E. AND I. WERNING (2013): "Insurance and Taxation over the Life Cycle," *The Review of Economic Studies.*

FINDEISEN, S. AND D. SACHS (2012): "Education and Optimal Dynamic Taxation: The Role of Income-Contingent Student Loans," *Working Paper No. 40, Department of Economics, University of Zurich.*

FRIEDMAN, M. (1955): "The role of government in education," *Rutgers University Press.*

GALLIPOLI, G., C. MEGHIR, AND G. VIOLANTE (2011): "Equilibrium effects of education policies: A quantitative evaluation," *Working Paper.*

GARY-BOBO, R. AND A. TRANNOY (2013): "Optimal Student Loans and Graduate Tax under Moral Hazard and Adverse Selection," *Working Paper.*

GELBER, A. AND M. WEINZIERL (2012): "Equalizing Outcomes vs. Equalizing Opportunities: Optimal Taxation when Children's Abilities Depend on Parents' Resources," *NBER Working Paper 18332.*

GIELEN, A. C. (2009): "Working hours flexibility and older workers' labor supply," *Oxford Economic Papers*, 61, 240–274.

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. 17642.*

GOLOSOV, M., M. TROSHKIN, A. TSYVINSKI, AND M. WEINZIERL (2013a): "Preference heterogeneity and optimal capital income taxation," *Journal of Public Economics*, 97, 160–175.

GOLOSOV, M. AND A. TSYVINSKI (2006): "Designing optimal disability insurance: A case for asset testing," *Journal of Political Economy*, 114, 257–279.

GOLOSOV, M., A. TSYVINSKI, AND N. WERQUIN (2013b): "Dynamic Tax Reforms," *Working Paper.*

GREENWOOD, J., Z. HERCOWITZ, AND G. W. HUFFMAN (1988): "Investment, capacity utilization, and the real business cycle," *The American Economic Review*, 402–417.

Grochulski, B. and T. Piskorski (2010): "Risky Human Capital and Deferred Capital Income Taxation," *Journal of Economic Theory*, 145, 908–943.

Gruber, J. and E. Saez (2002): "The elasticity of taxable income: evidence and implications," *Journal of Public Economics*, 84, 1–32.

Gustman, A. L. and T. L. Steinmeier (2008): "Projecting Behavioral Responses to the Next Generation of Retirement Policies," *Research in Labor Economics*, 28, 141–195.

Heathcote, J., F. Perri, and G. L. Violante (2010): "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006," *Review of Economic Dynamics*, 13, 15–51.

Heckman, J. J. (1993): "What Has Been Learned about Labor Supply in the Past Twenty Years?" *American Economic Review*, 83, 116–121.

Hendricks, L. and T. Schoellman (2012): "Student Abilities During the Expansion of US Education," *Working Paper*.

Huggett, M., G. Ventura, and A. Yaron (2011): "Sources of Lifetime Inequality," *American Economic Review*, 101(7), 2923–2954.

Immervoll, H., H. Kleven, C. Kreiner, and E. Saez (2007): "Welfare Reform in European Countries: A Microsimulation Analysis," *The Economic Journal*, 117, 1–44.

Immervoll, H., H. Kleven, C. Kreiner, and N. Verdelin (2011): "Optimal Tax and Transfer Programs for Couples with Extensive Labor Supply Responses," *Journal of Public Economics*, 95, 1485–1500.

Jacobs, B. and A. Bovenberg (2011a): "Optimal taxation of human capital and the earnings function," *Journal of Public Economic Theory*, 13, 957–971.

Jacobs, B. and A. L. Bovenberg (2011b): "Optimal Taxation of Human Capital and the Earnings Function," *Journal of Public Economic Theory*, 13, 957–971.

Jacobs, B. and D. Schindler (2012): "On the desirability of taxing capital income in optimal social insurance," *Journal of Public Economics*, 96, 853–868.

Jacobs, B., D. Schindler, and H. Yang (2012): "Optimal Taxation of Risky Human Capital," *Scandinavian Journal of Economics*, 114, 908–931.

Jacquet, L., E. Lehmann, and B. Van der Linden (2012): "Signing distortions in optimal tax and other adverse selection problems with random participation," *CESifo Working Paper 3766*.

Jappelli, T. and L. Pistaferri (2010): "The Consumption Response to Income Changes," *Annual Review of Economics*, 2, 479–506.

Kanbur, R., M. Keen, and M. Tuomala (1994): "Optimal non-linear income taxation for the alleviation of income-poverty," *European Economic Review*, 38, 1613 – 1632.

Kapicka, M. (2006): "Optimal Income Taxation with Human Capital Accumulation and Limited Record Keeping," *Review of Economic Dynamics*, 9, 612–639.

——— (2013): "Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach," *The Review of Economic Studies (forthcoming)*.

Kapicka, M. and J. Neira (2013): "Optimal taxation in a life-cycle economy with endogenous human capital formation," *Working paper, UC Santa Barbara.*

Kaplow, L. (2006): "On the undesirability of commodity taxation even when income taxation is not optimal," *Journal of Public Economics*, 90, 1235–1250.

——— (2007): "Optimal Income Transfers," *International Tax and Public Finance*, 14, 295–325.

Karahan, F. and S. Ozkan (2013): "On the persistence of income shocks over the life cycle: Evidence, theory, and implications," *Review of Economic Dynamics*, 26, 452–476.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2010): "The worldwide governance indicators: methodology and analytical issues," *World Bank Policy Research Working Paper*.

Keane, M. and K. Wolpin (1997): "The career decisions of young men," *Journal of Political Economy*, 105, 473–522.

Kleven, H., C. Kreiner, and E. Saez (2009): "The Optimal Income Taxation of Couples," *Econometrica*, 77, 537–560.

Kleven, H. J. and C. T. Kreiner (2006): "The marginal cost of public funds: Hours of work versus labor force participation," *Journal of Public Economics*, 90, 1955–1973.

KLEVEN, H. J. AND E. A. SCHULTZ (2012): "Estimating Taxable Income Responses using Danish Tax Reforms," .

KOCHERLAKOTA, N. R. (2005): "Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation," *Econometrica*, 73, 1587–1621.

KONRAD, K. A. (2001): "Privacy and time-consistent optimal labor income taxation," *Journal of Public Economics*, 79, 503–519.

KRUEGER, D. AND A. LUDWIG (2013): "Optimal Progressive Labor Income Taxation and Education Subsidies when Education Decisions and Intergenerational Transfers are Endogenous," *American Economic Review Papers & Proceedings*, 103, 496–501.

KYDLAND, F. E. AND E. C. PRESCOTT (1977): "Rules rather than discretion: The inconsistency of optimal plans," *The Journal of Political Economy*, 473–491.

LAROQUE, G. (2005): "Income Maintenance and Labor Force Participation," *Econometrica*, 73, 341–376.

LEMIEUX, T. (2006a): "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?" *The American Economic Review*, 461–498.

——— (2006b): "Postsecondary Education and Increasing Wage Inequality," *American Economic Review*, 96, 195–199.

LUTTMER, E. AND R. ZECKHAUSER (2008): "Schedule selection by agents: from price plans to tax tables," *NBER Working Paper No. 13808.*

MANKIW, N. AND M. WEINZIERL (2010): "The Optimal Taxation of Height: A Case Study of Utilitarian Income Redistribution," *American Economic Journal: Economic Policy*, 2, 155–176.

MANKIW, N., M. WEINZIERL, AND D. YAGAN (2009): "Optimal Taxation in Theory and Practice," *Journal of Economic Perspectives*, 23, 147–174.

MEGHIR, C. AND D. PHILLIPS (2010): "Labour Supply and Taxes," in *Dimensions of Tax Design: The Mirrlees Review*, ed. by Institute for Fiscal Studies, Oxford University Press.

MEGHIR, C. AND L. PISTAFERRI (2011): "Earnings, consumption and life cycle choices," *Handbook of Labor Economics*, 4, 773–854.

Mendoza, E. G. and V. Z. Yue (2012): "A general equilibrium model of sovereign default and business cycles," *The Quarterly Journal of Economics*, 127, 889–946.

Meyer, B. D. and D. T. Rosenbaum (2001): "Welfare, The Earned Income Tax Credit, and The Labor Supply of Single Mothers," *Quarterly Journal of Economics*, 116, 1063–1114.

Mirrlees, J. A. (1971): "An Exploration in the Theory of Optimum Income Taxation," *The Review of Economic Studies*, 38, 175–208.

Moffitt, R. (1982): "The Tobit Model, Hours of Work and Institutional Constraints," *The Review of Economics and Statistics*, 64, 510–515.

OECD (2011): "Pensions at a Glance 2011: Retirement-Income Systems in OECD and G20 Countries," URL: www.oecd.org/els/social/pensions/PAG.

Pavan, A., I. Segal, and J. Toikka (2012): "Dynamic Mechanism Design," *Working Paper.*

Pereira, J. (2009): "Essays on Time-Consistent Fiscal Policy," Ph.D. thesis, European University Institute.

Piketty, T. (1997): "La redistribution fiscale face au chomage," *Revue française d'économie*, 12, 157–201.

Piketty, T. and E. Saez (2013a): "Optimal labor income taxation," *Handbook of Public Economics, Vol 5.*

——— (2013b): "A Theory of Optimal Inheritance Taxation," *Econometrica*, forthcoming.

Pirttilä, J. and M. Tuomala (2004): "Poverty alleviation and tax policy," *European Economic Review*, 48, 1075–1090.

Poutvaara, P. (2003): "Investment in education and redistributive taxation without precommitment," *FinanzArchiv/Public Finance Analysis*, 177–188.

——— (2011): "The expansion of higher education and time-consistent taxation," *European Journal of Political Economy*, 27, 257–267.

Ramsay, C. M. (2006): "The distribution of sums of certain iid Pareto variates," *Communications in Statistics- Theory and Methods*, 35, 395–405.

RICHTER, W. F. (2009): "Taxing education in Ramsey's tradition," *Journal of Public Economics*, 93, 1254–1260.

ROGERSON, W. P. (1985): "Repeated Moral Hazard," *Econometrica*, 53, 69–76.

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 68, 205–229.

——— (2002a): "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes," *Journal of Public Economics*, 83, 217–230.

——— (2002b): "Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses," *Quarterly Journal of Economics*, 117, 1039–1073.

SALANIÉ, B. (2003): *The Economics of Taxation*, The MIT Press.

SCHEUER, F. (2012): "Entrepreneurial Taxation with Endogenous Entry," *Working Paper, Stanford University.*

SHESHINSKI, E. (1972): "The optimal linear income-tax," *The Review of Economic Studies*, 39, 297–302.

SINN, H.-W., C. HOLZNER, W. MEISTER, W. OCHEL, AND M. WERDING (2006): "Aktivierende Sozialhilfe 2006: Das Kombilohn-Modell des ifo Instituts," *ifo Schnelldienst*, 59, 6–27.

STANTCHEVA, S. (2012): "Optimal Taxation and Human Capital Policies Over the Lifecycle," *Graduate Student Research Paper 12-02 MIT*.

——— (2013): "Optimal Income Taxation with Adverse Selection in the Labor Market," *MIT Department of Economics Graduate Student Research Paper*.

STEWART, M. B. AND J. K. SWAFFIELD (1997): "Constraints on the Desired Hours of Work of British Men," *The Economic Journal*, 107, 520–535.

STIGLITZ, J. E. (1982): "Self-selection and Pareto efficient taxation," *Journal of Public Economics*, 17, 213–240.

STORESLETTEN, K., C. TELMER, AND A. YARON (2004): "Consumption and risk sharing over the life cycle," *Journal of Monetary Economics*, 51, 609–633.

TABER, C. (2001): "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?" *Review of Economic Studies*, 68, 665–691.

TROSTEL, P. A. (1993): "The effect of taxation on human capital," *Journal of Political Economy*, 327–350.

WAGNER, G., J. FRICK, AND J. SCHUPP (2007): "The German Socio-Economic Panel Study – Scope, Evolution and Enhancements," *Schmoller's Jahrbuch*, 127, 626–654.

WEINZIERL, M. (2011): "The Surprising Power of Age-Dependent Taxes," *The Review of Economic Studies*, 78, 1490–1518.

——— (2012): "Why Do We Redistribute So Much But Tag So Little? Normative Diversity and Optimal Taxation," *Working Paper, Harvard Business School.*

WERNING, I. (2007): "Optimal Fiscal Policy with Redistribution," *The Quarterly Journal of Economics*, 122, 925–967.

——— (2011): "Nonlinear Capital Taxation," *Working Paper, MIT.*

Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit mit dem Thema

Optimal Social Insurance and Redistribution: Incentives for Educational Investment, Work and Savings

ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Weitere Personen, insbesondere Promotionsberater, waren an der inhaltlich materiellen Erstellung dieser Arbeit nicht beteiligt.[15] Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Konstanz, den 15. Juli 2013

Dominik Sachs

[15]Siehe hierzu die Abgrenzung auf der folgenden Seite.

Abgrenzung

Kapitel 1 entstammt einer gemeinsamen Arbeit mit Sebastian Findeisen (Universität Zürich und Centre for Equitable Growth, Berkeley). Meine individuelle Leistung bei der Erstellung des Kapitels beträgt 50 %.

Kapitel 2 entstammt einer gemeinsamen Arbeit mit Normann Lorenz (Universität Trier). Meine individuelle Leistung bei der Erstellung des Kapitels beträgt 50 %.

Kapitel 3 entstammt einer gemeinsamen Arbeit mit Sebastian Findeisen (Universität Zürich und Centre for Equitable Growth, Berkeley). Meine individuelle Leistung bei der Erstellung des Kapitels beträgt 50 %.

Kapitel 4 entstammt einer gemeinsamen Arbeit mit Sebastian Findeisen (Universität Zürich und Centre for Equitable Growth, Berkeley). Meine individuelle Leistung bei der Erstellung des Kapitels beträgt 50 %.