

ESSAYS IN DYNAMIC OPTIMAL TAXATION: WORKING  
EXPERIENCE, UNEMPLOYMENT SCARRING AND POVERTY

by

LOUIS MORAND PERRAULT

A thesis submitted to the  
Department of Economics  
in conformity with the requirements for  
the degree of Doctor of Philosophy

Queen's University  
Kingston, Ontario, Canada

June 2016

Copyright © Louis Morand Perrault, 2016

*À Daphné, l'amour de ma vie.*

# Abstract

This thesis investigates the design of optimal tax systems in dynamic environments. The first essay characterizes the optimal tax system where wages depend on stochastic shocks and work experience. In addition to redistributive and efficiency motives, the taxation of inexperienced workers depends on a second-best requirement that encourages work experience, a social insurance motive and incentive effects. Calibrations using U.S. data yield higher expected optimal marginal income tax rates for experienced workers for most of the inexperienced workers. They confirm that the average marginal income tax rate increases (decreases) with age when shocks and work experience are substitutes (complements). Finally, more variability in experienced workers' earnings prospects leads to increasing tax rates since income taxation acts as a social insurance mechanism.

In the second essay, the properties of an optimal tax system are investigated in a dynamic private information economy where labor market frictions create unemployment that destroys workers' human capital. A two-skill type model is considered where wages and employment are endogenous. I find that the optimal tax system distorts the first-period wages of all workers below their efficient levels which leads to more employment. The standard no-distortion-at-the-top result no longer holds due to the combination of private information and the destruction of human capital. I show this result analytically under the Maximin social welfare function and confirm it numerically for a general social welfare function. I also investigate the use of a training program and job creation subsidies.

The final essay analyzes the optimal linear tax system when there is a population of individuals whose perceptions of savings are linked to their disposable income and their family background through family cultural transmission. Aside from the standard equity/efficiency trade-off, taxes account for the endogeneity of perceptions through two channels. First, taxing labor decreases income, which decreases the perception of savings through time. Second, taxation on savings corrects for the misperceptions of workers and thus savings and labor decisions. Numerical simulations confirm that behavioral issues push labor income taxes upward to finance saving subsidies. Government transfers to individuals are also decreased to finance those same subsidies.

## **Co-authorship**

Chapter 4 of this thesis was co-authored with Pier-André Bouchard St-Amant, Researcher at Employment and Social Development Canada (ESDC).

## Acknowledgments

I am incredibly grateful to my supervisor Robin W. Boadway for his continued guidance, support and generosity throughout the long process of writing this thesis. I wish to thank the following faculty members at Queen's University for their comments: James Bergin, Charlie Beach, Marco Cozzi, Sumon Majumbar and Huw Lloyd-Ellis. I am also indebted to Étienne Lehmann for having generously given me his time, and for his valuable comments. I would also like to thank Laurence Jacquet for all her help and advice. I wish to thank my former colleagues at Queen's for the stimulating discussions that helped shaped this thesis, and for their comments on the essays found herein: Jean-Denis Garon, Nicolas Martineau, Alex Armstrong, Pier-André Bouchard St-Amand, Mike Cloutier, Jean-François Rouillard, Steve Kivinen and Afrasiab Mirza. A special thanks to my good friends Robert, Nicolas, Jean-François and Mathieu for being there whenever I needed to talk.

J'aimerais remercier Renée Lamontagne pour son soutien et pour m'avoir encouragé à emprunter la voie académique. Je suis infiniment reconnaissant à Daphné Lamontagne, ma partenaire, pour tout l'amour et le soutien qu'elle m'a donnés et tous les sacrifices qu'elle a dû faire durant ce long périple.

Finalement, je souhaite fortement remercier mes parents, Suzanne et Marc, pour avoir toujours encouragé ma curiosité et m'avoir encouragé à poursuivre mes passions jusqu'au bout.

# Contents

<b>Dedication</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Co-authorship</b>	<b>iv</b>
<b>Acknowledgments</b>	<b>v</b>
<b>Contents</b>	<b>vi</b>
<b>List of Tables</b>	<b>ix</b>
<b>List of Figures</b>	<b>xi</b>
<b>Chapter 1: Introduction</b>	<b>1</b>
<b>Chapter 2: Optimal Taxation with Work Experience as a Risky Investment</b>	<b>11</b>
2.1 The Model . . . . .	15
2.2 The Optimal Allocation: Optimal Wedges . . . . .	21
2.2.1 Definitions . . . . .	21
2.2.2 Optimal Wedges . . . . .	24
2.2.3 Uncertainty for older workers: the CES wage and log-normal distribution of shocks case . . . . .	31
2.2.3.1 No uncertainty in the second period . . . . .	32
2.2.3.2 Impact of complementarity and variance on the second period labor wedge . . . . .	33
2.3 Numerical Simulations . . . . .	36
2.3.1 Functional Forms, Calibration and Computational Strategy . . . . .	36
2.3.2 Results . . . . .	40
2.3.2.1 Changes in $\rho$ . . . . .	43
2.3.2.2 Changes in $\delta$ . . . . .	47
2.3.2.3 Changes in $\alpha$ . . . . .	50
2.3.2.4 Changes in $\sigma$ . . . . .	52
2.4 Conclusion . . . . .	54

<b>Chapter 3: Optimal Income Taxation with Unemployment Induced Loss of Human Capital</b>	<b>56</b>
3.1 The Model . . . . .	62
3.1.1 Individuals . . . . .	65
3.1.1.1 First Period . . . . .	65
3.1.1.2 Second Period . . . . .	67
3.1.2 Firms . . . . .	69
3.1.3 Wage Setting in First Period . . . . .	70
3.1.4 Planner and Perfect Information in First Period . . . . .	73
3.1.4.1 Planner . . . . .	73
3.1.4.2 Perfect Information in First Period . . . . .	75
3.2 Optimal Tax Policy . . . . .	78
3.2.1 Maximin Social Welfare Function . . . . .	80
3.2.2 Bergson-Samuelson Social Welfare Function . . . . .	87
3.3 Alternative Policy Tools . . . . .	91
3.3.1 Training Programs . . . . .	91
3.3.1.1 Perfect Information in the First Period . . . . .	93
3.3.1.2 Imperfect Information in the Second Period . . . . .	95
3.3.2 Job Creation Subsidies . . . . .	103
3.4 Concluding Comments . . . . .	110
<b>Chapter 4: Decision-Making In Poverty, Savings and Redistribution</b>	<b>112</b>
4.0.1 Literature Review . . . . .	114
4.1 Model . . . . .	117
4.1.1 Agents and Timing . . . . .	117
4.1.1.1 Behavior and Comparative Statics . . . . .	121
4.1.1.1.1 Effect of a change in time preference . . . . .	121
4.1.1.1.2 Effect of a change in disposable income . . . . .	123
4.1.1.1.3 Effect of taxes . . . . .	125
4.1.2 The Government . . . . .	126
4.2 Optimal Taxes . . . . .	127
4.3 Simulations . . . . .	130
4.4 Conclusion . . . . .	137
<b>Chapter 5: Conclusion</b>	<b>139</b>
<b>Bibliography</b>	<b>144</b>
<b>Appendix A: Mathematical Appendix</b>	<b>155</b>
A.1 Optimal taxation with work experience as a risky investment . . . . .	155
A.1.1 The Hamiltonians . . . . .	155
A.1.1.1 Labor wedge . . . . .	158
A.1.1.2 Intertemporal Wedge . . . . .	160
A.1.1.3 Proof of Corollary 2.a and 2.b . . . . .	161

A.1.2	Implementation: No savings in equilibrium . . . . .	166
A.1.3	Computational Information . . . . .	171
A.1.3.1	Further Numerical Simulations . . . . .	171
A.2	Optimal Income Taxation with Unemployment Induced Loss of Human Capital	172
A.2.1	Microfoundations of Wage-Setting Objective . . . . .	172
A.2.1.1	Nash-Bargaining . . . . .	172
A.2.1.2	Competitive Search Equilibrium . . . . .	173
A.2.2	Proof of Corollary 3 . . . . .	174
A.2.3	Optimal tax policy: Bergson-Samuelson case and Proof of Corollary 4.	176
A.2.4	Numerical Simulations . . . . .	178
A.3	Decision-Making In Poverty, Savings and Redistribution . . . . .	179
A.3.1	Proofs of various statements . . . . .	179
A.3.1.1	Comparative Statics of Consumption, Savings and Labor Al- locations . . . . .	180
A.3.1.1.1	First Sub-Period . . . . .	180
A.3.1.1.2	Effect of Time Preference on Labor Allocation . . . . .	181
A.3.1.1.3	Second Sub-Period . . . . .	181
A.3.1.2	Derivations of Tax Rates . . . . .	182
A.3.1.2.1	Derivation of the formula for $\tau_l$ : . . . . .	182
A.3.1.2.2	Derivation of the formula for $\tau_s$ : . . . . .	183

# List of Tables

2.1	Calibration Parameters . . . . .	39
2.2	Average Wedges in Each Period: All Calibrations . . . . .	44
2.3	Average Wedges in Each Period: Using Calibration of $\rho = 0.2$ with Changes in $\rho$ . . . . .	45
2.4	Average Wedges in Each Period: Using Calibration of $\rho = 1.2$ with Changes in $\rho$ . . . . .	47
2.5	Average Wedges in Each Period: $\rho = 0.5$ with Changes in $\delta$ . . . . .	49
2.6	Average Wedges in Each Period: $\rho = 0.5$ with Changes in $\alpha$ . . . . .	52
2.7	Average Wedges in Each Period: $\rho = 0.5$ with Changes in t=2's $\sigma$ . . . . .	54
3.1	Characteristics of the Optimal Allocation and Tax System: Changes in $\varepsilon$ . . . . .	89
3.2	Characteristics of the Optimal Allocation and Tax System: Various Inequality Aversion . . . . .	91
3.3	Characteristics of the Optimal Allocation and Tax System: Training Program	102
3.4	Characteristics of the Optimal Allocation and Tax System: Job Creation Subsidies . . . . .	109
4.1	Behavioral Elasticities . . . . .	127
4.2	Numerical Values Used For Default Simulations . . . . .	131
4.3	Parameters Reproducing Observed US Data . . . . .	132
4.4	Optimal Taxes For Various Proportions of Learning Agents (others at $\beta^*$ ) . . . . .	133
4.5	Optimal Taxes When Dynasties Keep the Same Productivities . . . . .	135

4.6	Optimal Taxes For Quasi-linear Preferences . . . . .	135
4.7	Optimal Taxes For Various Degree of Aversion to Inequality ( $\rho$ ) . . . . .	136
A.1	Average Wedges in Each Period: Full Calibration . . . . .	171
A.2	Characteristics of the Optimal Allocation and Tax System: Effect of Parameters	179

# List of Figures

2.1	Optimal Wedges from Calibration $\rho = 0.5$ . . . . .	41
2.2	Optimal Labor Wedge in t=1: Three Motives . . . . .	42
2.3	Optimal Wedges: $\rho = 0.2$ with Changes in $\rho$ . . . . .	46
2.4	Optimal Wedges: $\rho = 1.2$ with Changes in $\rho$ . . . . .	48
2.5	Optimal Wedges: $\rho = 0.5$ with Changes in $\delta$ . . . . .	50
2.6	Optimal Wedges: $\rho = 0.5$ with Changes in $\alpha$ . . . . .	51
2.7	Optimal Wedges: $\rho = 0.5$ with Changes in in t=2's $\sigma$ . . . . .	53
4.1	Phase diagram of time preference for different productivity levels . . . . .	122
A.1	Full Calibrations . . . . .	172

# Chapter 1

## Introduction

Individuals do not come into this life equal nor will they face the same chances, misfortunes and choices throughout their lives. They are not only born with different skills and abilities but they are also born into families that are themselves much different from family to family. These families can vary in wealth but also in attitudes towards hard work, education, and savings. During their lives, some individuals may have a chance to enter networks that will offer them good employment opportunities, others may not. Also some individuals will face unemployment, others chronic illness that forces them to cut back or even drop out of the labor market.

All these different circumstances and events will result in individuals living under very different levels of material ease, stress and health. They will also shape decisions that will have lasting effects on an individual's life. This implies that the tax system and public policies should have a redistributive and an insurance role to play in society.

Furthermore, circumstances affecting an individual today can have long term consequences on wages, employability, amount of wealth accumulated, education and so on. A child born in a poor family and living in a poor neighborhood will have its access to quality education and thus higher earning prospects in the future diminished (Heckman, 2008). A young worker who starts his career in a recession will likely have a lower starting wage and this wage gap will likely persist for a substantial part of his career (Oreopoulos et al., 2012).

A worker who is hit by an involuntary spell of employment can face an important wage penalty in the future (Arulampalam, 2001).

Therefore when designing policies aimed at redistributing and insuring, the government must also incorporate these dynamic considerations. It is in the interest of the government to make policies that allow or help maintain the productivity of workers throughout their lives. More productive workers can translate into more resources for the government as it affects its tax base. At the same time the dynamic implications of taxation and social insurance policies can put greater constraints on the extent to which the government is able to redistribute and insure. Greater levels of taxation can discourage investments in capital and education, and can reduce labor market participation. Offering greater insurance is known to raise issues of moral hazard which leads individuals to take on more risk and act in a very different manner than they would if they were exposed to the full amount of risk. In the case of unemployment insurance this may lead to greater aggregate unemployment levels which may have disastrous implications on the long term for many workers.

Redistribution has long been a focus of optimal tax theory. A seminal example is Mirrlees (1971) who considers the government's problem of redistributive taxation where workers are heterogeneous in their ability to generate labor income. It abstracts from many complications such as differences in preferences, the ability to transfer resources across periods of life and endogenous formation of market wages. Since then many extensions to this problem have been considered ranging from redistributive taxation with general equilibrium effects on wages, e.g. Stiglitz (1982), redistributive taxation and indirect taxation, e.g. Atkinson and Stiglitz (1976), redistributive taxation and education, e.g. Bovenberg and Jacobs (2005), redistributive taxation in an overlapping-generations model, e.g. Ordober and Phelps (1979), redistributive taxation and heterogeneous preferences, e.g. Boadway et al. (2002), and more.

Questions of insurance in dynamic models are often treated separately from the redistributive goal of the government in public economics literature. This is often by design

to clarify intuition but also for technical reasons. A good example of this is in the optimal unemployment insurance literature, e.g. Chetty (2006); Shimer and Werning (2006); Spinnewijn (2013). These papers usually feature representative agent models who will face different employment shocks during their life and the government will try to insure them against this risk. Some notable exceptions in this literature to this are Boadway et al. (2003); Hungerbühler et al. (2006) and Lehmann et al. (2011). Although designed to capture dynamic considerations these models can be considered to be more static in nature.

A strand of the optimal tax literature often referred to as the “New Dynamic Public Finance” (NDPF) tries to tackle the issue of redistribution and insurance together in a dynamic setting. Recent examples of this literature are Kapicka (2013); Farhi and Werning (2013) and Golosov et al. (2015). These papers abstract from the endogeneity of wages and labor market productivity by assuming exogenous ability that changes stochastically through life. In the same strand of literature some have relaxed the assumption on exogenous wages and have modeled them as incorporating a human capital element which is acquired through education and training, e.g. Bohacek and Kapicka (2008); Kapicka (May 2014) and Stantcheva (2014). In these papers both redistribution and insurance will have an impact on the productivity of workers and thus the distribution of wages.

This thesis explores the design of redistributive policies when labor market outcomes of workers have long lasting effects later in life, including on future wages, savings and on their families. It is composed of three essays. The first essay found in chapter 2 investigates the characteristics of an optimal tax system that offers redistribution and insurance throughout a worker’s life when wages are in part determined by work experience. The second essay in chapter 3 considers the optimal income tax system when involuntary unemployment causes wage scars through productivity destruction. Chapter 4 holds the last essay which analyzes the optimal redistributive linear tax system when resources available to households influence economic decision-making.

Chapter 2's essay contributes to the Mirrleesian optimal tax theory that considers the impact of taxation on wages when workers become more productive by acquiring work experience. The impact of work experience and tenure on future wages is a well established empirical fact (Blundell and MaCurdy, 1999; Farber, 1999). There are a few mechanism by which this could happen but a prominent one is that work experience is a form of human capital investment. Although, learning-by-doing fits in the general human capital framework it has mostly been ignored from the optimal taxation literature until very recently (Krause, 2009a; Best and Kleven, February 2013; Kapicka, 2014). This is in stark contrast to the very rich optimal tax literature that considers human capital accumulation through education and on-the-job training. As argued by Best and Kleven (February 2013), taxation models featuring formal education have quite different implications from models of taxation with learning-by-doing. This comes from the fact that the time required for formal education is a substitute to the time needed to work whereas the other form of human capital is a byproduct of work. This has an important consequence when designing tax policy as labor income taxation will always distort human capital accumulation through learning-by-doing which is not the case for accumulation coming from formal education (Eaton and Rosen, 1980; Stantcheva, 2014). And since formal education is technically observable, the government is able to use different tools to encourage or discourage acquiring human capital.

One important result from the learning-by-doing tax literature is that the age profile of labor income taxation is declining with age. This result is in part due to the lower labor supply elasticity with respect to labor income taxation in the younger period of life coming from the willingness of workers to work to obtain future wage gains. This is of particular interest as age-dependent taxation is seen as a potential near-term reform for developed economies (Banks and Diamond, 2010). This policy prescription is in sharp contrast to recent results from the NDPF literature which features persistent productivity shocks throughout a worker's lifetime. In these models the age profile of taxation is increasing. Thus, the age

profile of labor income taxes is still up for debate.

Since persistence in wages and earnings can be obtained with models of work experience, the essay in chapter 2 features a model where wages are a function of both a stochastic shock and a accumulated work experience. This contributes to the learning-by-doing literature as the latter has mostly been studied in cases where there is no risk on the future wage.<sup>1</sup> It also contributes to the NDPF literature as I model persistence directly through work experience. The methodology used to solve the model follows the first-order approach extended to a dynamic setting by Kapicka (2013) and Farhi and Werning (2013).

I find that the optimal marginal income tax rate when young includes three new additional motives. The first is a second-best motive that pushes down the marginal tax rate so as to encourage work experience accumulation since the government distorts the labor choice in future periods to raise revenues. The second motive is a social insurance motive that will encourage work experience if it insures the worker from wage risk or discourages it if it increases exposure to risk. This will depend on whether the shock and the stock of work experience are substitutes or complements. The third motive is an incentive rationale that takes into consideration the disincentive effect of taxation in the future period on work experience accumulation. Using numerical simulations based on data from the U.S. economy I find that a majority of workers should have a lower marginal labor income tax rate when young compared to the expected marginal labor income tax rate they will face when older. However, the age profile of the cross-sectional average of marginal income tax rates is increasing (decreasing) when the shock and work experience are substitutes (complements). An explanation for this result is that as the complementarity of the wage function increases, this leads young workers at the bottom of the distribution to have an increased desire to work and thus lowers the labor supply elasticity with respect to taxation when young.

---

<sup>1</sup>Stantcheva (December 2014) considers the case where training effort can be complimentary to working effort, learning-and-doing, with a limit case being learning-by-doing. The model studied in that paper does feature wage risk.

Unlike chapter 2, chapter 3's essay considers the case where human capital is changed by circumstances that are outside of the individual's control. To be more precise, workers lose a fraction of their productivity from involuntary unemployment. This phenomenon is often called 'unemployment scarring' and it relates to the long term consequences of experiencing unemployment. This can be from losing firm specific human capital or even a worker's general level of skill. Although related to learning-by-doing, it can be viewed as 'unlearning-by-not-doing'. The essay is an optimal taxation exercise where the government has a redistributive motive but also must take into consideration the impact taxes have on wages and the level of employment. The model used in the essay is a two-skill version of Hungerbühler et al. (2006) that incorporates dynamic considerations in the design of optimal taxes. In the model, the labor market is plagued by frictional unemployment coming from search and matching frictions (Diamond, 1982; Mortensen and Pissarides, 1999).

Related to this essay is the normative literature that also features depreciation of human capital due to unemployment (Coles and Masters, 2000; Shimer and Werning, 2006; Pavoni, 2009; Spinnewijn, 2013). I contribute to the literature by tackling the question of redistributive taxation in such a situation while most of the literature has focused on optimal unemployment insurance or how to maximize the total output of the economy. Income taxation in the model can influence employment through the 'wage-cum-labor-demand' margin (Lehmann et al., 2011). Briefly, an increase in the marginal income tax rate puts downward pressure on the equilibrium wage. With a high enough marginal tax rate compared to the average tax rate it is possible that the before-tax wage will be lower than in a no-tax economy and thus increasing the demand for workers.

I find that the optimal tax system would distort the skilled workers' labor market to decrease the equilibrium below the *laissez-faire* wage. This implies that employment of the skilled will be higher when the government has a redistributive motive and the workers can lose human capital due to unemployment. This comes first from the fact that the

government is also taxing workers in the future. The tax burden of the future has an impact on the equilibrium wage when young as it discourages them from accepting a lower wage to keep their human capital. The government must seek to undo this problem by increasing the marginal income tax rate and putting a downward pressure on the skilled wage when young. Second, because the government is raising tax revenues in the future it wishes to keep more skilled workers at their current skill level. This is similar to the second-best motive result found in chapter 2, but in this context it has a different policy prescription as it points to higher marginal income tax rates instead of lower ones. Plainly, by ensuring that more skilled workers obtain a job, and by so doing keep their human capital, more revenue is raised, the tax burden is shared by more skilled workers and the number of workers the government must redistribute to is reduced.

I also obtain the result found in Hungerbühler et al. (2006) and Lehmann et al. (2011) that redistribution to the unskilled workers is partly achieved by increasing the likelihood of them finding employment. This is achieved by distorting their wage through taxation below the *laissez-faire* level which then creates more employment for the unskilled. Finally, alternative labor policies such as a training program or a job creating subsidy can theoretically reduce the need to distort the labor markets and increase redistribution. However, using numerical simulations, I find that with plausible parameter values for the model only the training program can achieve its promise whereas job creating subsidies are redundant and would only be useful in the situation where the low skill wage would be driven to zero.

Chapter 4's essay, which is joint work with Pier-André Bouchard St-Amand, considers another aspect of the importance of labor market outcomes on the future. Recent optimal tax literature has investigated models where workers are heterogeneous along several dimensions. One reason for this is that positive models featuring only heterogeneity along one dimension cannot easily match empirical facts with respect to asset holdings and savings (Krusell and Smith, 1998). In light of this, optimal tax models featuring heterogeneity of both

preferences towards the future and ability have been of particular interest, e.g. Diamond and Spinnewijn (2011) and Golosov et al. (2013b). However, models featuring several dimensions of heterogeneity raise both technical and moral issues for optimal tax theory. For Mirrleesian type models issues of multidimensional screening will be present. This means that in an asymmetrical information model where workers are heterogeneous along several dimensions the government may not be able to easily separate all types of workers. More interesting are the philosophical issues that are raised when thinking of redistributive policies when workers are heterogeneous along multiple dimensions. Should the government redistribute or insure the workers that did not save for their older days irrespective of their preferences? Or should the government try to only compensate those that have different levels of productivity but refrain from interfering with the choices of those with different preferences?

In any case, models featuring exogenous distributions of time preferences and interpreting these preferences as falling under the responsibility of the individual can imply that some individuals have bad preferences. In the sense that the government should not necessarily be worried about individuals who have low levels of savings for their retirement as it is their fault if they chose poorly. Recent experimental and empirical evidence from the behavioral economics literature, e.g. Spears (2011); Shah et al. (2012), show that this maybe misleading as there appears to be a link between poverty and observed myopic behavior of the less fortunate. The essay presented in chapter 4 takes these findings and investigates the design of the optimal linear tax system when scarcity of resources has an effect on economic decision-making such as labor and savings choices. In such a context the moral intuition is simpler when it comes to redistributive preferences of the government. That is, poor individuals are not necessarily found responsible for their perceived low preference for the future, but it is more an extension of their situation of deprivation.

The essay is closely linked to the behavioral optimal taxation literature and by extension the non-welfarist optimal taxation literature. In this literature, individuals may make

decisions that they will regret later in life as they are based on temporary preferences coming from specific circumstances or altered states, e.g. O'Donoghue and Rabin (2006). In this context the government may have paternalistic preferences and will use its own preferences when judging the appropriateness of the individual's allocation. When this is the case, optimal tax formulas often have an extra term that captures the discrepancy between the government's preferences and those of the individual as it tries to correct the faulty behavior (Kanbur et al., 2006). The essay contributes to this literature as the optimal tax exercise considers the case where the extent of behavioral problems are a function of available resources and not just exogenously assigned as it is often done in the literature, e.g. Diamond and Spinnewijn (2011).

My coauthor and I find that the tax system must balance other considerations besides the more standard equity and efficiency considerations of the government. The behavioral problem of individuals raises two issues for the worker. He might not save enough and may make mistakes in his labor supply decision. Since both labor income taxes and saving taxes affect these two choices, the optimal tax formulas for each tax instrument will take their effects into considerations with regards to behavioral issues. Taxing labor income discourages labor supply and thus labor income generation. In this model this will have several effects. The more standard one is that by taxing labor this has an income effect on the savings decision. The novel effect is that by reducing income the tax modifies how individuals perceive the future. The tax on savings has a more direct effect on savings, and can thus better be used to change the marginal value of a dollar of savings. In a redistributive tax system where poverty can cause individuals to have myopic decision-making the government has a new tradeoff. It may want to subsidize savings or increase monetary transfers to individuals to reduce the behavioral problems, but to achieve this it may need to increase labor income taxes which can also increase behavioral problems.

Our numerical simulations suggest that savings should be subsidized and this subsidy

should increase more the more prevalent is the behavioral problem. This is funded by a reduction in transfers and an increase in labor income taxes. And as redistributive preferences of the government increase, the subsidies to savings, the labor income taxes and the transfer are increased. In this model the redistributive motive gives more weight to the behavioral issues of the poor. This is highlighted in the last result since even if more weight is given to the equity potential of saving taxation, the saving subsidies are made greater and thus the behavioral motive is not sacrificed for more redistribution through transfers.

## Chapter 2

# Optimal Taxation with Work Experience as a Risky Investment

Several empirical regularities, such as lower labor supply elasticity for young workers (Blundell and MaCurdy, 1999) and persistence in labor earnings (Storesletten et al., 2004) can be explained with models of work experience accumulation. Recent optimal tax literature has shown that these two facts call for conflicting policy prescriptions. For instance, models featuring persistent productivity shocks prescribe tax rates that increase with age (Farhi and Werning, 2013; Stantcheva, 2014). Contrastingly, models of learning-by-doing featuring no productivity shocks prescribe the exact opposite (Best and Kleven, February 2013).

Thus, features of the age structure of tax schedules are still an open question. This essay studies an optimal history-dependent tax system when future wage rates, while depending on accumulated work experience, are risky prospects. It adds to the recent strand of literature on age-dependent taxation (Kremer, 2002; Blomquist and Micheletto, 2008; Weinzierl, 2011) and it also enriches the new dynamic optimal tax literature of Golosov et al. (2003); Kocherlakota (2005); Golosov et al. (2007). In particular, I identify new key factors that should help us think through the optimal age-dependent tax structure.

The results I obtain, using data from the 2007 Panel Study of Income Dynamics (PSID), are first that in the optimum a majority of workers (roughly 66% to 93%) will face higher

expected marginal labor income tax rates when old.<sup>1</sup> Second, whether the cross-sectional average of marginal labor income tax rates is lower when young than when old depends crucially on the complementarity between the stochastic shock and the accumulated work experience. The information obtained on cross-sectional averages of the labor distortions can yield important insights in light of the results of Farhi and Werning (2013). They show that setting linear age-dependent taxes such that the tax rates at each age are the cross-sectional averages of the fully history-dependent tax system can capture almost all of the welfare gains of the second-best compared to the *laissez-faire* outcome. Using a Constant Elasticity of Substitution (CES) wage function, I illustrate that when the elasticity of complementarity between shock and accumulated work experience is below one, the cross-sectional average marginal income tax rate is lower when young, and when it is above one, the cross-sectional average is lower when old.<sup>2</sup> A possible explanation for this result is that a greater complementarity in the wage function leads to a higher wage elasticity with respect to work experience at the lower end of the shock distribution. In accordance with the results of Best and Kleven (February 2013) this would push marginal labor income taxes upwards for lower income young workers as their labor effort becomes less elastic to taxation.

On theoretical grounds, I provide an analytical characterization of the optimal history-dependent tax system using the characteristics of the optimal second-best allocation. In addition to the redistributive and efficiency motives, the optimal labor distortion (*wedge*) formula reflects a balance between three motives that capture the added effects of having risk and work experience determining the wage when old. The first of these is a second-best rationale that pushes the first period labor wedge downward to encourage work experience accumulation. The second captures the insurance goals of the planner that will either push

---

<sup>1</sup>Solving the problem of the planner yields the optimal labor and saving distortions. Under a specific implementation of the optimal allocation, the optimal labor distortions can be interpreted as marginal tax rates of a history-dependent tax system.

<sup>2</sup>Note that the elasticity of complementarity is the inverse of the elasticity of substitution. For the CES function used in the simulations, the elasticity of complementarity is also equal to the Hicksian complementarity coefficient.

upward or downward the labor wedge depending on whether the Hicksian complementary coefficient between shock and work experience is above or below one. The third is an incentive motive which takes into account the disincentive on work experience accumulation coming from taxation in the second period. This motive will either push downwards or upwards depending on whether the second period consumption and marginal benefit of work experience is positively correlated or negatively correlated.

The optimal labor wedge formula when old, although similar to the formula found in the Mirrleesian optimal tax literature, is supplemented by the wage elasticity with respect to the second period shock. The more elastic the wage is with respect to the shock the higher the second period wedge will be. As shown in Golosov et al. (2015), the shape of the optimal tax schedule is heavily influenced by the hazard ratio of the stochastic shock. I show that the behavior of the wage elasticity with respect to the shock, as the second period shock tends to infinity, either reinforces or drastically diminishes the impact of the hazard ratio on the labor wedge formula when old.<sup>3</sup> This new result also depends crucially on the complementarity between the shock and accumulated work experience. I further show that the riskier the second period is, the higher the second period labor wedge should be at the right tail of the distribution of shocks. The impact of risk on the whole tax system is investigated using numerical simulations. Increasing the volatility of stochastic shocks in the second period increases labor wedges in both period but much more drastically when old.

This project is closely related to the new public finance literature that looks at characteristics of the optimal tax system when skill shocks are persistent, such as Farhi and Werning (2013) and Golosov et al. (2015). It builds on these contributions as it follows the first-order approach to solve the planner's problem but models the persistence of productivity directly

---

<sup>3</sup>In fact, numerically, it can be observed that this result happens for values of the stochastic shocks slightly above the mean value of the shock.

by incorporating work experience.<sup>4</sup> By endogenizing the wage formation process, the optimal policies take into consideration the effect of taxes and how they effect persistence of earnings. It is also tied to recent work on optimal taxation with human capital and risky environments where human capital is acquired through either schooling or on-the-job training programs (Bohacek and Kapicka, 2008; Kapicka, May 2014; Kapicka and Neira, October 2014; Stantcheva, 2014).

Finally, the essay contributes to the literature on optimal taxation with learning-by-doing. Krause (2009a) considers optimal taxation in a two-type model similar to Stiglitz (1982), where the planner can commit or not to a two-period income tax schedule. The author finds that the no-distortion-at-the-top result no longer applies and that there are some cases where it is justified to tax the high skilled workers even if it depresses both labor supply and future wages. Best and Kleven (February 2013) also considers a Mirleesian economy with a continuum of types. They study both age-independent and age-dependent taxation. Their numerical simulations make a strong case for higher age-dependent income tax rates for the young. This is due to endogenous wage rates and a negative correlation between age and innate (first period) ability (conditional on earnings). In contrast to Krause (2009a) and Best and Kleven (February 2013), I incorporate uncertainty in the second period wage rate and the ability for workers to save.

Stantcheva (December 2014) also considers learning-by-doing in a risky environment. However, she is more concerned with the impact of the substitutability or complementarity of training effort with respect to labor effort on the optimal tax schedule. In that particular model, learning and labor efforts are dissociated, which allows for working without learning. My assumption is rather that labor supply decisions necessarily involve learning through the accumulation of work experience. In that respect, it is more in line with Krause (2009a) and

---

<sup>4</sup>As we will explain in more detail below, the planner is able to observe the accumulation of work experience. This assumption allows us to keep *common knowledge of preference*, and thus our analysis is also similar to Albanesi and Sleet (2006).

Best and Kleven (February 2013).

The rest of the essay is organized as follows. Section 2 presents the two-period model and writes the planner's problem recursively. Section 3 considers the characteristics of the optimal income tax systems derived from analytical results. Section 4 presents different numerical simulations to highlight further properties of the optimal allocation. Finally Section 5 is dedicated to concluding comments.

## 2.1 The Model

The economy is populated by workers who live for two periods  $t = 1, 2$ , each period they consume  $c_t$  and provide labor effort  $l_t$  from which they acquire work experience. Workers obtain wage  $w_t$  and earn gross income  $y_t = w_t l_t$ . The wage  $w_t(\theta_t, e_t)$  is a function of the time-varying shock  $\theta_t$  and of the stock of effective work experience  $e_t$ . It is assumed that:

$$\frac{\partial w_t}{\partial \theta_t} > 0; \quad \frac{\partial w_t}{\partial e_t} > 0 \quad \frac{\partial^2 w_t}{\partial \theta_t \partial e_t} \geq 0; \quad \frac{\partial^2 w_t}{\partial^2 x} \leq 0 \text{ for } x = \{\theta_t, e_t\}, \quad (2.1)$$

where the stock of worker experience evolves according to

$$e_{t+1} = \phi(y_t) + e_t, \quad (2.2)$$

where  $\phi_y > 0, \phi_{yy} < 0$  and the starting level of work experience,  $e_1$ , is identical for all workers. Work experience accumulation is assumed to be a function of income and not labor effort alone as it permits the accumulation to be observable to the planner. It also captures the idea that work experience would not only be a function of labor alone but also of the type of job or opportunity an individual would have.<sup>5</sup> In each period, worker's shock  $\theta_t$  is

---

<sup>5</sup>Making work experience a function of both income and present period shock  $\theta$ , i.e.  $\phi(\theta_t, y_t)$ , would remove the *common knowledge of preference* assumption as work experience would no longer be observable to the planner.

distributed according to the density  $f^t(\theta_t)$  with support  $[\underline{\theta}, \bar{\theta}]$ . The focus is restricted to shocks that are independent across periods and endogenize the persistence of wages through accumulated work experience. Let the per period preferences be represented by the following utility function

$$u(c_t) - h(l_t), \quad (2.3)$$

with  $u$  being strictly concave,  $h$  strictly convex, and also define  $\theta^t$  as the history of shocks up to period  $t$ . For a given allocation  $\{c(\theta^2), y(\theta^2)\}$  that specifies consumption and income for each history of shocks the worker's expected lifetime utility is

$$U(\{c, y\}) \equiv \sum_{t=1}^2 \beta^{t-1} \int \left[ u(c(\theta^t)) - h\left(\frac{y(\theta^t)}{w_t(\theta_t, e_t(\theta^{t-1}))}\right) \right] f^2(\theta_2) f^1(\theta_1) d\theta_2 d\theta_1, \quad (2.4)$$

where  $\beta$  is the discount factor.

**Incentive Compatibility and Planner's Problem:** The particularity of Mirrleesian optimal taxation is that worker's shocks are private observation and that the planner is constrained by the information he has, i.e. he can only observe labor income and savings.<sup>6</sup> By the revelation principle, it is possible to focus on direct mechanisms where workers will report their type every period. A reporting strategy  $\sigma = \{\sigma_1(\theta^1), \sigma_2(\theta^2)\}$  implies a history of per period reports  $\sigma^t(\theta^t)$ , for which the planner allocates consumption  $c(\sigma^t)$ , income  $y(\sigma^t)$  and by extension accumulation of work experience effective in the second period, i.e.

$$e_2(\sigma_1(\theta^1)) = \phi(y(\sigma_1(\theta^1))) + e_1. \quad (2.5)$$

For an allocation  $\{c, y\}$ , let  $\omega(\theta^t)$  denote the equilibrium continuation utility after history

<sup>6</sup>However, the distribution of shocks is known to the planner.

<sup>7</sup>Work experience  $e(\sigma_1(\theta^1))$  is also a function of  $e_1$  but since it is assumed that all workers start with the same level of work experience, for simplicity of exposition, we will omit to include this level of starting work experience for the rest of the essay.

$\theta^t$ , defined as the unique solution to

$$\omega(\theta^t) = u(c(\theta^t)) - h\left(\frac{y(\theta^t)}{w_t(\theta_t, e_t)}\right) + \beta \int \omega(\theta^t, \theta_{t+1}) f^{t+1}(\theta_{t+1}) d\theta_{t+1} \quad (2.6)$$

for  $t = 1, 2$  with  $\omega(\theta^3) \equiv 0$ .<sup>8</sup> For any reporting strategy  $\sigma$ , let the continuation value  $\omega^\sigma(\theta^t)$  be the unique solution to

$$\omega^\sigma(\theta^t) = u(c(\sigma^t(\theta^t))) - h\left(\frac{y(\sigma^t(\theta^t))}{w_t(\theta_t, e_t)}\right) + \beta \int \omega^\sigma(\theta^t, \theta_{t+1}) f^{t+1}(\theta_{t+1}) d\theta_{t+1}. \quad (2.7)$$

An allocation  $\{c, y\}$  is said to be incentive compatible if and only if

$$\omega(\theta^t) \geq \omega^\sigma(\theta^t) \quad \forall \theta^t, \forall \sigma. \quad (2.8)$$

Therefore an allocation is incentive compatible if truth telling, i.e.  $\sigma^* = \{\sigma^*(\theta^t)\}$  with  $\sigma_t^*(\theta_t) = \theta_t$  yields a weakly higher continuation utility.

Let the planner be an Utilitarian who promises workers, behind the veil of ignorance, expected lifetime utility  $v_0$ . Also suppose that the economy has a linear technology that transform effective labor into consumption and that the planner can transfer resources across periods at a gross interest rate of  $R$ . The planner's problem is then to minimize the cost of providing allocation  $\{c, y\}$  subject to the allocation being incentive compatible and offering expected lifetime utility  $v_0$ , i.e.

$$K(v_0, e_1) = \min_{\{c, y\}} \left[ \sum_{t=1}^2 \left(\frac{1}{R}\right)^{t-1} \int \{c(\theta^t) - y(\theta^t)\} f^2(\theta_2) f^1(\theta_1) d\theta_2 d\theta_1 \right] \quad (2.9)$$

$$s.t. \quad U(\{c, y\}) = v_0,$$

$$\omega(\theta^t) \geq \omega^\sigma(\theta^t) \quad \forall \theta^t, \forall \sigma,$$

$$e_2(\theta^1) = \phi(y(\theta^1)) + e_1.$$

---

<sup>8</sup>Note that  $\omega(\theta^t, \theta_{t+1})$  stands in for  $\omega(\theta^{t+1})$  for clarity of exposition.

## The relaxed problem

Following Farhi and Werning (2013), I use the first-order approach to write a relaxed problem of the planner’s problem. The approach relies on changing the incentive constraint (2.8) to a “temporal” incentive constraint which only considers one-shot deviations every period. They show that the set of allocations that satisfy these new incentive constraints from the relaxed problem is a subset of allocations that satisfy (2.8).<sup>9</sup>

I derive the temporal incentive constraints as in Farhi and Werning (2013). Working backwards and starting from period 2, let the continuation value under truthful revelation,  $\omega(\theta^2)$ , be:

$$\omega(\theta^2) = u(c(\theta^2)) - h\left(\frac{y(\theta^2)}{w_2(\theta_2, e_2(\theta^1))}\right). \quad (2.10)$$

Consider the deviation strategy  $r$  where the workers reports truthfully until  $t$  but not during period  $t$ , i.e.  $\sigma^{t-1}(\theta^{t-1}) = \theta^{t-1}$  and  $\sigma^t(\theta^t) = r$ , where  $r \neq \theta_t$ . The continuation utility under deviation strategy  $r$  in the second period is:

$$\omega^{\sigma^r}(\theta^2) = u(c(\theta^1, r)) - h\left(\frac{y(\theta^1, r)}{w_2(\theta_2, e_2(\theta^1))}\right). \quad (2.11)$$

An allocation is temporary incentive compatible in period 2, if for all histories  $\theta^2$ ,

$$\omega(\theta^2) = \max_r \omega^{\sigma^r}(\theta^2). \quad (2.12)$$

For period 1, let the continuation value under truthful revelation,  $\omega(\theta^1)$ , be the unique

---

<sup>9</sup>However, much like in the static nonlinear optimal tax case, the validity of the first-order approach in a dynamic setting is not guaranteed. In the case of a static optimal tax exercise, it can be shown that when the worker’s utility function satisfies the Spence-Mirrlees single-crossing property and the allocation satisfies a monotonicity condition, which corresponds to the second-order condition, the optimal allocation is incentive compatible. In this dynamic setting the envelope condition is a necessary condition, but there are no “simple” conditions like in the static case that guarantee incentive compatibility. Therefore, in the numerical simulations below, as in Farhi and Werning (2013) or Golosov et al. (2015), the incentive compatibility of the allocations from the solutions of the relaxed problems are verified ex-post numerically.

solution to:

$$\omega(\theta^1) = u(c(\theta^1)) - h\left(\frac{y(\theta^1)}{w_1(\theta_1, e_1)}\right) + \beta \int \omega(\theta^1, \theta_2) f^2(\theta_2) d\theta_2. \quad (2.13)$$

Again, considering a one shot-deviation strategy  $r$ , let the continuation utility be the unique solution to:

$$\omega^{\sigma^r}(\theta^1) = u(c(r)) - h\left(\frac{y(r)}{w_1(\theta_1, e_1)}\right) + \beta \int \omega^{\sigma^r}(e_2(\theta_1, r), \theta_2) f^2(\theta_2) d\theta_2, \quad (2.14)$$

where  $e(r) = \phi(y(r)) + e_1$ .<sup>10</sup> An allocation is temporary incentive compatible in period 1, if for all shock  $\theta^1$ ,

$$\omega(\theta^1) = \max_r \omega^{\sigma^r}(\theta^1).^{11} \quad (2.15)$$

An allocation is incentive compatible, i.e. constraints (2.8) hold, if for all histories of shocks both

$$\omega(\theta^1) = \max_r \omega^{\sigma^r}(\theta^1) \quad \text{and} \quad \omega(\theta^2) = \max_r \omega^{\sigma^r}(\theta^2) \quad (2.16)$$

are true. Following the first-order approach, these temporary incentive constraints are replaced by the following envelope conditions applied to (2.16):

$$\dot{\omega}(\theta^1) := \frac{\partial \omega(\theta^1)}{\partial \theta_1} = \frac{h'\left(\frac{y(\theta^1)}{w_1(\theta_1, e_1)}\right) y(\theta^1)}{[w_1(\theta_1, e_1)]^2} \frac{\partial w_1(\theta_1, e_1)}{\partial \theta_1}, \quad (2.17)$$

$$\dot{\omega}(\theta^2) := \frac{\partial \omega(\theta^2)}{\partial \theta_2} = \frac{h'\left(\frac{y(\theta^2)}{w_2(\theta_2, e_2(y(\theta^1)))}\right) y(\theta^2)}{[w_2(\theta_2, e_2(y(\theta^1)))]^2} \frac{\partial w_2(\theta_2, e_2(y(\theta^1)))}{\partial \theta_2}. \quad (2.18)$$

The problem is written recursively starting with the second period problem. Let

<sup>10</sup> $\omega^{\sigma^r}(r, e(r), \theta_2)$  can be any future strategy and has no link to the period 2 one-shot deviation utility  $\omega^{\sigma^r}(\theta^2)$ .

<sup>11</sup>Equivalently it can be written

$$\omega(\theta^1) = \max_r \left\{ u(c(r)) - h\left(\frac{y(r)}{w_1(\theta_1, e_1)}\right) + \beta \int \omega^{\sigma^r}(r, e_2(r), \theta_2) f^2(\theta_2) d\theta_2 \right\}.$$

$v(\theta^1) = \int \omega(\theta^1, \theta_2) f^2(\theta_2) d\theta_2$  be the expected continuation utility. The second period problem of the planner is to minimize the second period expected costs taking as given  $v(\theta^1)$  and  $e(\theta^1)$  (expressed as  $v$ , and  $e_2$  respectively) subject to the envelope condition,

$$\begin{aligned}
K(v, e_2, 2) &= \min_{\{c(\theta), y(\theta)\}} \int [c(\theta) - y(\theta)] f^2(\theta) d\theta & (2.19) \\
s.t. \quad \omega(\theta) &= u(c(\theta)) - h\left(\frac{y(\theta)}{w_2(\theta, e_2)}\right) \\
v &= \int \omega(\theta) f^2(\theta) d\theta \\
&\text{and (2.18).}
\end{aligned}$$

The first period problem taking as given  $v_0$  and starting work experience  $e_1$  is

$$\begin{aligned}
K(v_0, e_1, 1) &= \min_{\{c(\theta), y(\theta), \omega(\theta), v(\theta)\}} \int \left[ c(\theta) - y(\theta) + \frac{1}{R} K(v(\theta), e(\theta), 2) \right] f^1(\theta) d\theta \\
s.t. \quad \omega(\theta) &= u(c(\theta)) - h\left(\frac{y(\theta)}{w_1(\theta, e_1)}\right) + \beta v(\theta) & (2.20) \\
v_0 &= \int \omega(\theta) f^1(\theta) d\theta \\
e(\theta) &= \phi(y(\theta)) + e_1 \\
&\text{and (2.17).}
\end{aligned}$$

Because the planner is utilitarian  $v_0$  is the expected lifetime utility promised from beyond the veil of ignorance.<sup>12</sup> The value of  $v_0$  is chosen to be the highest value such that the expected costs of providing this level of expected lifetime utility is zero, i.e.  $K(v_0, e_1, 1) = 0$ .

---

<sup>12</sup>The veil of ignorance in the original position is a useful thought experiment that can help in deriving fundamental principles of justice. For example, see Rawls (1971).

## 2.2 The Optimal Allocation: Optimal Wedges

### 2.2.1 Definitions

To write and interpret the optimal wedge formulas I require the use of several terms which I define below.

#### Wedges

In the static non-linear optimal taxation literature differences in marginal utility of consumption and marginal disutility of labor are interpreted as marginal taxes. However, in a dynamic setting, these differences can no longer readily be interpreted in such a way. It is then convenient to define these differences as *wedges* which help in getting intuition from the solution of the optimal allocation problem. One of the issues with interpreting wedges as taxes is that there can be many different combinations of tax instruments to implement the optimal allocation. Furthermore, as argued by Golosov et al. (2007) each wedge corresponds to a particular choice of the worker taking all other choices fixed at a specific level. Since choices are made jointly in a decentralized economy setting, a particular tax rate equal to the wedge without further restrictions on the tax instruments may lead workers to deviate and the optimal allocation may not be implemented. Nonetheless, some implementations, like the one elaborated in Appendix A.1.2, can equate the labor wedges to marginal tax rates. For this reason, this essay sometimes uses those terms interchangeably.

For any allocation  $\{c, y\}$  after any history  $\theta^t$ , let the intertemporal wedges  $\tau_K(\theta^t)$  and

the labor wedges  $\tau_L(\theta^1), \tau_L(\theta^2)$  respectively be:

$$\tau_K(\theta^1) = 1 - \frac{1}{R\beta} \frac{u'(c(\theta^1))}{\mathbb{E}[u'(c(\theta^2))]} \quad \forall \theta^1, \quad (2.21)$$

$$\tau_L(\theta^1) = 1 - \frac{h' \left( \frac{y(\theta^1)}{w_1(\theta_1, e_1)} \right)}{u'(c(\theta^1))w_1(\theta_1, e_1)} + \beta \frac{\phi_y(y(\theta^1))}{u'(c(\theta^1))} \mathbb{E}[\mathcal{MB}(\theta^2)] \quad \forall \theta^1, \quad (2.22)$$

$$\tau_L(\theta^2) = 1 - \frac{h' \left( \frac{y(\theta^2)}{w_2(\theta_2, e(\theta^1))} \right)}{u'(c(\theta^2))w_2(\theta_2, e(\theta^1))} \quad \forall \theta^2, \quad (2.23)$$

where

$$\mathcal{MB}(\theta^2) \equiv \frac{h' \left( \frac{y(\theta^2)}{w_2(\theta_2, e_2(\theta^1))} \right) y(\theta^2)}{[w_2(\theta_2, e(\theta^1))]^2} \frac{\partial w_2(\theta_2, e(\theta^1))}{\partial e(\theta^1)}$$

is the realized marginal benefit of work experience and  $\mathbb{E}[\mathcal{MB}(\theta^2)]$  is the expected marginal benefit of work experience from the point of view of the first period.<sup>13</sup> The intertemporal wedge  $\tau_K(\theta^1)$  as written above is the difference between the expected marginal rate of intertemporal substitution and the return on savings. The labor wedges in both periods, i.e.  $\tau_L(\theta^1), \tau_L(\theta^2)$ , is the differences in marginal utility of consumption and marginal disutility of labor in a given period.

### Hicksian Complementarity

Let the Hicksian complementarity coefficient between the shock  $\theta_2$  and work experience  $e_2$

---

<sup>13</sup>The marginal benefit can also be written using labor instead of income:

$$\mathcal{MB}(\theta^2) \equiv \frac{h'(l(\theta^2)) l(\theta^2)}{w_2(\theta_2, e_2)} \frac{\partial w_2(\theta_2, e_2)}{\partial e}.$$

be defined as

$$\rho(\theta^2, e_2) \equiv \frac{\frac{\partial^2 w_2(\theta_2, e_2)}{\partial \theta \partial e_2} w_2(\theta_2, e_2)}{\frac{\partial w_2(\theta_2, e_2)}{\partial \theta} \frac{\partial w_2(\theta_2, e_2)}{\partial e_2}}. \quad (2.24)$$

This coefficient measures how complementary the shock and work experience are in the production of the second period wage. For any wage function that is additive in shock and work experience, this term is zero as they are perfect substitutes. In any situation where  $\frac{\partial^2 w_2}{\partial \theta \partial e} > 0$ , the Hicksian complementarity coefficient is positive and work experience will increase exposure to risk. As Stantcheva (2014) has demonstrated it is the relation of the Hicksian coefficient with respect to one that is more relevant to the planner for questions of risk and insurance.<sup>14</sup> More precisely, if the Hicksian coefficient is below 1, it implies that the wage elasticity with respect to the shock is decreasing in work experience, or alternatively the wage elasticity with respect to work experience is decreasing with respect to the shock. If the Hicksian coefficient is greater than 1, the wage elasticity with respect to the shock is increasing in work experience and whenever the coefficient equals 1 the wage elasticity is constant. Note that a CES wage function where  $w_2 = (\theta_2^{1-\rho} + e_2^{1-\rho})^{\frac{1}{1-\rho}}$  will have  $\rho(\theta^2, e_2) = \rho$ .

## Elasticities

Optimal tax formulas are generally written with elasticity parameters which capture the efficiency cost of taxation. Let

$$\alpha_t(\theta^t) \equiv \frac{h''(l(\theta^t))l(\theta^t)}{h'(l(\theta^t))} \text{ and } \eta_t(\theta^t) \equiv -\frac{u''(c(\theta^t))c(\theta^t)}{u'(c(\theta^t))},$$

be two elasticity measures.<sup>15</sup> The elasticity measure  $\alpha_t(\theta^t)$  is akin to the inverse of the Frisch

---

<sup>14</sup>This fact is demonstrated in the optimal tax literature in both Bovenberg and Jacobs (2011) and Stantcheva (2014).

<sup>15</sup>These elasticity measures are used in Golosov et al. (2015).

elasticity of labor supply.<sup>16</sup> The elasticity measure  $\eta_t(\theta^t)$  is the inverse of the elasticity of intertemporal substitution or alternatively the measure of relative risk aversion of the preferences.

### 2.2.2 Optimal Wedges

The subsection describes and analyzes the solution to the planner's problem. The labor choice distortion in both periods is considered first and the distortion to the saving decisions is considered later. To obtain further insights on the labor wedges, I make the following assumption:

**Assumption 1.** *The optimal allocation satisfies*

$$\dot{c}(\theta^t) \geq 0 \quad \forall \theta^t \text{ for } t = 1, 2.^{17}$$

**Proposition 1.** *The optimal labor wedges in periods  $t=1$  and  $t=2$  are:*

$$\begin{aligned} \tau_L^*(\theta^1) = & \frac{\mu(\theta^1)}{f^1(\theta_1)} \frac{h'(l_1)}{w_1} \frac{\epsilon_{\theta 1}}{\theta_1} (1 + \alpha_1(\theta^1)) - \frac{\phi_y}{R} E[\tau_L(\theta^2) w_{e2} l_2] \\ & - \frac{\phi_y}{R} \left\{ E\left[ \mathcal{MB}(\theta^2) \frac{\mu(\theta^2)}{f^2(\theta_2)} \frac{\epsilon_{\theta 2}}{\theta_2} (1 - \rho_{\theta e 2}) \right] + \text{Cov}\left( \frac{1}{u'(c_2)}, \mathcal{MB}(\theta^2) \right) \right\} \end{aligned} \quad (2.25)$$

$$\frac{\tau_L^*(\theta^2)}{1 - \tau_L^*(\theta^2)} = u'(c(\theta^2)) \frac{\mu(\theta^2)}{f^2(\theta_2)} \frac{\epsilon_{\theta 2}}{\theta_2} (1 + \alpha_2(\theta^2)) \geq 0, \quad (2.26)$$

---

<sup>16</sup>In this setting, in the first period it captures the elasticity of labor supply keeping future wage, consumption in both period and future labor supply constant. In the second period it also captures the elasticity of labor supply keeping consumption in both period and labor in the first period constant.

<sup>17</sup>It can be shown that incentive compatibility requires that

$$\frac{\partial[u(c(\theta)) + \beta\omega(\theta)]}{\partial\theta} \geq 0.$$

I find that Assumption 1 is satisfied in all of our numerical simulations. Assumption 1 is a standard assumption in the NDPF literature, for an example see Golosov et al. (2015).

where<sup>18</sup>

$$\mu(\theta^1) = \int_{\underline{\theta}}^{\bar{\theta}} (1 - g(\theta^1)) \frac{1}{u'(c(\theta^1))} f^1(\theta_1) d\theta_1, \text{ with } g(\theta^t) = u'(c(\theta^t)) \lambda_1,$$

$$\text{and } \lambda_1 = \int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{u'(c(\theta^1))} f^1(\theta_1) d\theta_1,$$

$$\mu(\theta^2) = \int_{\underline{\theta}}^{\bar{\theta}} (1 - g(\theta^1, \theta_2)) \frac{1}{u'(c(\theta^1, \theta_2))} f^2(\theta_2) d\theta_2,$$

$$\text{with } g(\theta^1, \theta_2) = u'(c(\theta^1, \theta_2)) \lambda_2, \text{ and } \lambda_2 = \int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{u'(c(\theta^1, \theta_2))} f^2(\theta_2) d\theta_2.$$

*Proof:* See Appendix A.1.1.1. Also note, that for ease of notation let  $l(\theta^t) = l_t$ ,  $c(\theta^t) = c_t$ ,  $e(\theta^1) = e_2$ ,  $w_t(\theta_t, e_t) = w_t$ ,  $\partial w_t / \partial e = w_{et}$ ,  $\partial w_t / \partial \theta = w_{\theta t}$ ,  $\epsilon_{\theta t} = w_{\theta t} * (\theta / w_t)$ ,  $\rho(\theta^2, e_2) = \rho_{\theta e 2}$ .

The multipliers associated with the envelope conditions in period 1 and period 2 are  $\mu(\theta^1)$  and  $\mu(\theta^2)$  respectively. The first period's multiplier captures the insurance motive or the redistributive goals of a utilitarian planner who promised lifetime expected utility  $v_0$  from behind the veil of ignorance.<sup>19</sup> And,  $\mu(\theta^2)$  is the insurance motive of the planner who promised continuation utility  $v(\theta^1)$  to a worker. Parameter  $g(\theta^t)$  is the value to the planner of giving one more dollar to an individual with history  $\theta^t$ ,  $\lambda_1$  is the marginal resource cost of providing a marginal increase in promised utility  $v_0$ , and  $\lambda_2$  is the same for the promised utility  $v$  for period 2 from period 1.

Starting with the second period labor wedge (2.26), notice that it is possible to rewrite

---

<sup>18</sup>I use the same definitions found in Stantcheva(2014b).

<sup>19</sup>If the planner had social preferences other than utilitarian and had redistributive preferences over lifetime utilities based on first period heterogeneity, then  $\lambda_1$  would need to be a function of parameter  $\theta_1$ . The alternative first period problem is not covered in this essay. See Golosov et al. (2015) for a treatment where a Planner has weighted utilitarian social preferences.

it in the ABC form found in the static optimal tax literature, e.g. Diamond (2003):

$$\frac{\tau_L(\theta^2)}{1 - \tau_L(\theta^2)} = A(\theta^2)B(\theta^2)C(\theta^2), \quad (2.27)$$

where

$$\begin{aligned} A(\theta^2) &= (1 + \alpha_2(\theta^2))\epsilon_{\theta^2}(\theta^2), \\ B(\theta^2) &= \frac{1 - F^2(\theta_2)}{\theta_2 f^2(\theta_2)}, \\ C(\theta^2) &= \int_{\theta_2}^{\bar{\theta}} \exp\left(\int_{\theta_2}^x \eta(\tilde{x}) \frac{\dot{c}(\tilde{x})}{c(\tilde{x})} d\tilde{x}\right) [1 - g(x)] \frac{f^2(x)}{1 - F^2(\theta_2)} dx. \end{aligned}$$

The intuition behind this optimal wedge formula goes as follows. Distorting labor supply at shock level  $\theta_2$  increases the marginal deadweight burden at this shock level which depends on the labor supply elasticity  $\alpha_2(\theta^2)$  and is amplified by the wage elasticity with respect to the shock  $\epsilon_{\theta^2}(\theta^2)$ . This effect is captured by  $A(\theta^2)$ . Note that  $\epsilon_{\theta^2}(\theta^2)$  is not present in most optimal income tax exercise since a shock (or type) represents the wage rate for a particular worker type.<sup>20</sup> Since wage in this essay is composed of both shock and work experience, the impact of a change in type on the wage will be present in the envelope condition and thus in the optimal labor wedge formula.

Increasing the distortion at shock level  $\theta_2$  transfers resources from workers with higher shock levels to the planner who's value for these resources is reflected by  $C(\theta^2)$ .<sup>21</sup> It is the insurance motive of the planner. And finally, the hazard ratio  $B(\theta^2)$  reflects the tradeoff between  $A(\theta^2)$  and  $C(\theta^2)$  as it captures the number of workers related to  $C(\theta^2)$  and the number of workers at  $A(\theta^2)$ . The fact that the wage elasticity, labor elasticity and the density are all positive combined with Assumption 1 leads to the result that the optimal

---

<sup>20</sup>The partial derivative  $\partial w/\partial\theta$  in the standard envelope condition is technically present but it is equal to one.

<sup>21</sup>As pointed out in Golosov et al. (2015), the term  $\exp\left(\int_{\theta_2}^x \eta(\tilde{x}) \frac{\dot{c}(\tilde{x})}{c(\tilde{x})} d\tilde{x}\right)$  represents the income effect on labor supply of transferring resources from workers who receive shocks above  $\theta_2$ .

second period wedge is non-negative.<sup>22</sup> From Assumption 1, the monotonicity assumption on consumption, it is possible to show that  $\mu(\theta^2) \geq 0$  and thus the result on the optimal wedge follows.<sup>23</sup>

The formula for the first period optimal wedge (2.25) is composed of four parts.<sup>24</sup> The first part is more familiar and the three others are linked to the accumulation of work experience through labor. This first part is composed of the redistributive motive, the shock density, the wage elasticity, and the labor choice elasticity parameter. By the same argument as above, this first part will also be non-negative as it can be shown that  $\mu(\theta^1) \geq 0$ .

The remaining three novel parts of the optimal first period labor wedge formula are all linked to the different motives the planner has for work experience accumulation. These motives are:

$$\underbrace{\mathbb{E}\left[\tau_L(\theta^2)w_{e2}l_2\right]}_{\text{Second-Best Motive}} + \underbrace{\mathbb{E}\left[\mathcal{MB}(\theta^2)\frac{\mu(\theta^2)}{f^2(\theta_2)}\frac{\epsilon_{\theta 2}}{\theta_2}(1 - \rho_{\theta e 2})\right]}_{t = 2\text{'s Social Insurance Motive}} + \underbrace{\text{Cov}\left(\frac{1}{u'(c_2)}, \mathcal{MB}(\theta^2)\right)}_{\text{Incentive Motive}}. \quad (2.28)$$

<sup>22</sup>From the boundary conditions, it is possible to obtain the classic no distortion at the top and bottom.

<sup>23</sup>See Golosov et al. (2011) and Stantcheva (2014). The demonstration is simpler than the one in these papers as it is not necessary to take into consideration the impact of persistence of shocks in the present setting.

<sup>24</sup>The first period labor wedge can also be partly written in the ABC form, but its interpretation is much less straightforward. Equation (2.25) can be rewritten in the following way:

$$\tau_L(\theta^1) = A(\theta^1)B(\theta^1)C(\theta^1)\left[\frac{h'(l_1)}{u'(c_1)w_1}\right] - \frac{\phi_y}{R} \left\{ \mathbb{E}\left[\tau_L(\theta^2)w_{e2}l_2\right] + \mathbb{E}\left[\mathcal{MB}(\theta^2)\frac{\mu(\theta^2)}{f^2(\theta_2)}\frac{\epsilon_{\theta 2}}{\theta_2}(1 - \rho_{\theta e 2})\right] + \text{Cov}\left(\frac{1}{u'(c_2)}, \mathcal{MB}(\theta^2)\right) \right\}$$

where

$$\begin{aligned} A(\theta^1) &= (1 + \alpha_1(\theta^1))\epsilon_{\theta 1}(\theta^1), \\ B(\theta^1) &= \frac{1 - F^1(\theta_1)}{\theta_1 f^1(\theta_1)}, \\ C(\theta^1) &= \int_{\theta_1}^{\bar{\theta}} \exp\left(\int_{\theta_1}^x \eta(\tilde{x})\frac{\dot{c}(\tilde{x})}{c(\tilde{x})}d\tilde{x}\right)[1 - g(x)]\frac{f^1(x)}{1 - F^1(\theta_1)}dx. \end{aligned}$$

<sup>25</sup>These three parts are multiplied by  $-\phi_y/R$ , which is removed for expositional simplicity.

The first part of (2.28) is referred to as the second-best motive which captures the expected marginal revenues the planner gains from increasing the worker's work experience. In fact, terms like this one can be found in most optimal tax literature where the planner distorts more than one market. The planner takes into account the impact of a change in a tax instrument in one market on the other markets that also face distortions by other tax instruments. Since  $\tau_L(\theta^2) \geq 0$ , the fact that  $w_{e2}l_2 \geq 0$ , and that for any given shock  $\theta_1$  there is at least one subsequent shock  $\theta_2$  where the planner requires  $l(\theta^2) > 0$  the first part must be positive. This implies that the second-best motive pushes for a lower optimal first period wedge which encourages work experience accumulation.

The second part of (2.28) captures the social insurance motive of the planner and how imperfect information in the second period affects pre and after-tax insurance. The second part of (2.28) is composed of the marginal benefit of work experience, the insurance motive in the second period  $\mu(\theta^2)$ , the density  $f^2(\theta^2)$ , the wage elasticity with respect the shock  $\epsilon_{\theta 2}$ , and the Hicksian complimentary coefficient  $\rho_{\theta e 2}$ . The sign of this part depends entirely on whether  $\rho_{\theta e 2}$  is smaller, equal or greater than one. The planner will want to encourage work experience when the elasticity of wage with respect to shocks will be decreasing in work experience, i.e.  $\rho_{\theta e 2} < 1$ , and discourage it when the elasticity of wage with respects to shocks is increasing in work experience, i.e.  $\rho_{\theta e 2} > 1$ . The rationale being that the planner is able to diminish exposure to risk by increasing work experience whenever  $\rho_{\theta e 2} < 1$ . The opposite is true when  $\rho_{\theta e 2} > 1$  making him want to discourage work experience. This result is similar to the optimal net wedge on human capital expenses in Stantcheva (2014). However in this model. it is the expected value of similar terms in the second period that matters as the effect of work experience is felt in the second period and not in the period it was

accumulated as in Stantcheva (2014).<sup>26</sup>

The third part can be seen as providing incentive in a second best framework. In contrast with the second part, it is linked to the marginal benefit of work experience in terms of labor disutility and not just benefits in terms of wage. Suppose that the planner was able to perfectly observe types in the second period, he could perfectly insure the worker, i.e.  $c(\theta^2) = \hat{c}$  for all  $\theta^2$ , which would imply that the covariance term would be zero. As the planner is incapable to do this in an asymmetric information set up, he must offer incentives for truthful reporting. Assumption 1 states that the optimal allocation features non-decreasing consumption in period two with respect to shock  $\theta_2$ , thus the covariance term will be positive if the marginal benefit of work experience is increasing in type. This will depend both two factors, one behavioral i.e. whether labor effort is generally increasing with type and the other technological, whether the Hicksian complementarity coefficient is greater or lesser than one as is shown in the following equations:

$$\frac{\partial \mathcal{MB}(\theta_1, \theta_2)}{\partial \theta_2} = (1 + \alpha_2(l_2)) \frac{\epsilon_{e2}}{e_2} i(\theta^2) - \frac{h'(l_2) l_2 w_{e2}}{w_2} \frac{\epsilon_{\theta 2}}{\theta_2} (1 - \rho_{\theta e 2}). \quad (2.29)$$

The more a worker works in the second period the more he will gain from an increase in his wage coming from an increase in the shock. Thus, if labor is increasing in type, the marginal benefit of work experience should, in part, also be increasing in type. But as it was argued above, whenever the Hicksian complementarity coefficient is below 1 the wage

---

<sup>26</sup>Note that the first and second terms of (2.28) can be combined. Using the definition of the optimal second period labor wedge it is possible to insert it in the insurance motive term. Rearranging and combining the first two terms of (2.28) it is possible to write the optimal first period labor wedge in the following way:

$$\begin{aligned} \tau_L^*(\theta^1) = & \frac{\mu(\theta^1)}{f^1(\theta_1)} \frac{h'(l_1)}{w_1} \frac{\epsilon_{\theta 1}}{\theta_1} (1 + \alpha_1(\theta^1)) \\ & - \frac{\phi_y}{R} \left\{ \mathbb{E} \left[ \tau_L^*(\theta^2) w_{e2} l_2 \left( \frac{2 + \alpha_2(\theta^2) - \rho_{\theta e 2}}{1 + \alpha_2(\theta^2)} \right) \right] + \text{Cov} \left( \frac{1}{u'(c_2)}, \mathcal{MB}(\theta^2) \right) \right\}. \end{aligned}$$

From this formulation two things come to light. First, the Hicksian complementarity parameter  $\rho(\theta^2, e_2)$  must be at least above 2 to push the value of the combined terms to be negative and thus push the first period labor wedge upwards. Second, it highlights the importance of the second period labor elasticity apart from the level of  $\tau_L^*(\theta^2)$  in determining the first period labor wedge.

elasticity with respect to work experience is decreasing in type. Thus in such a situation the marginal benefit of work experience will, in part, be decreasing in shock. So, in that case the effect can push in different directions. If for example labor effort is increasing in shock and the Hicksian complementarity coefficient is above 1 then the marginal benefit of work experience must be increasing in shock. If both second period consumption and the marginal benefit of work experience are increasing in shock, redistribution would reduce the incentive to acquire work experience. Hence the planner will want to decrease the optimal first period labor wedge to incentivize workers to acquire work experience to counteract the disincentive effect of redistribution in the second period. This logic is reversed if the covariance term becomes negative, in this case the planner will wish to increase the optimal first period labor wedge to discourage work experience accumulation.

### **Intertemporal Wedge:**

As utility is separable between consumption and labor effort in this model, the optimal condition (2.21) has the “inverse Euler equation” feature:

$$\frac{1}{u'(c(\theta^1))} = \frac{1}{R\beta} \int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{u'(c(\theta^2))} f^2(\theta_2) d\theta_2.$$

**Proposition 2.** *Suppose that the relaxed problem solves the original problem. Then the optimal intertemporal wedge is positive,  $\tau_K^*(\theta^1) > 0$ , and satisfies:*

$$\tau_K^*(\theta^1) = 1 - \frac{\left[ \int_{\underline{\theta}}^{\bar{\theta}} [u'(c(\theta^2))]^{-1} f^2(\theta_2) d\theta_2 \right]^{-1}}{\int_{\underline{\theta}}^{\bar{\theta}} u'(c(\theta^2)) f^2(\theta_2) d\theta_2} > 0.$$

The proof can be found in Appendix A.1.1.2. The result that the intertemporal wedge is positive is obtained by applying Jensen’s inequality to the definition of the optimal wedge.

This result is found in several NDPF papers such as Kocherlakota (2005); Farhi and Werning (2013); Stantcheva (December 2014).<sup>27</sup> The intuition for this distortion is that the planner seeks to discourage savings as savings make separating workers who received different shocks more difficult. For the rest of the essay I will sometimes refer to the intertemporal wedge as the savings wedge.

### 2.2.3 Uncertainty for older workers: the CES wage and log-normal distribution of shocks case

To further investigate the impact of introducing uncertainty in the later period of life on the age structure of the tax system, this subsection derives additional characteristics on labor wedges by specifying the functional form of the wage function and the probability distribution of the shocks.<sup>28</sup> Let the functional form of the wage function be

$$w_t(\theta_t, e_t) = (\theta_t^{1-\rho} + e_t^{1-\rho})^{\frac{1}{1-\rho}}. \quad (2.30)$$

For this wage function the Hicksian complementarity coefficient, the wage elasticity with respect to shock, and to work experience are

$$\rho(\theta_2, e_2) = \rho, \quad \epsilon_{\theta t} = \left(\frac{\theta_t}{w_t}\right)^{1-\rho} \quad \text{and} \quad \epsilon_{e t} = \left(\frac{e_t}{w_t}\right)^{1-\rho}. \quad (2.31)$$

The probability distribution of the shocks considered in this subsection is the log-normal distribution  $\ln \mathcal{N}(\mu, \sigma^2)$  where  $\mu$  and  $\sigma$  are, respectively, the mean and standard deviation of the random variables natural logarithm and the probability density function of the log-normal distribution is

$$f^t(\theta) = \frac{1}{\theta \sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln \theta - \mu)^2}{2\sigma^2}\right). \quad (2.32)$$

---

<sup>27</sup>See Golosov et al. (2015) for a more general result featuring non-separable preferences between consumption and labor effort.

<sup>28</sup>Many of the results in this subsection hold for more general wage functions as shown in the Appendix.

### 2.2.3.1 No uncertainty in the second period

As a benchmark, consider the characteristics of the optimal allocation when there is no uncertainty in the second period, i.e.  $\Pr\{\theta_2 = 0\} = 1$ . This implies that the wage of each worker is his accumulated work experience, i.e.  $w_2(\theta^1) = e_2(\theta^1)$ .

**Corollary 1.** *Supposing that the relaxed problem solves the original problem, when  $\Pr\{\theta_2 = 0\} = 1$  and the second period wage of workers is  $w_2(\theta^1) = e_2(\theta^1)$  the optimal intertemporal wedge is zero,  $\tau_K^*(\theta^1) = 0$ , the second period optimal labor wedge is zero,  $\tau_L^{2*}(\theta^1) = 0$  and the first period labor wedge is :*

$$\tau_L^{1*}(\theta^1) = \frac{\mu(\theta^1)}{f^1(\theta_1)} \frac{h'(l_1)}{w_1} \frac{\epsilon_{\theta_1}}{\theta_1} (1 + \alpha_1(\theta^1)).$$

In this special case there is no longer an information problem in the second period as the planner is able to observe work experience accumulation. In this situation the planner does not need to distort the second period labor decision. Because there is no need to distort the second period, there is also no reason for the planner to distort the savings decision as it will not relax the incentive problem in the second period (or in the first period). This implies that for each worker consumption in both periods is smoothed and follows  $u'(c_1(\theta^1)) = \beta R u'(c_2(\theta^1))$ . However, since there is still an information problem in the first period, consumption cannot be equalized in the first period for all workers and by extension second period consumption is also not equalized. This result is similar to the one obtained in Atkinson and Stiglitz (1976) where there is no incentive to tax savings when preferences are separable.

For the first period labor distortion, comparing (2.25) with the optimal first period wedge in Corollary 1, the difference between those formulas are the three terms of (2.28). As there is no distortion of the second period labor decisions nor any uncertainty, encouraging or discouraging work experience accumulation by distorting the first period labor decision is no

longer necessary. Only the redistribution motive and the efficiency motive of the first period remain.

Note that in this no uncertainty scenario, the results on the second period labor wedge are different than those of Krause (2009a) and Best and Kleven (February 2013). The reason for this is twofold. The first is that information available to the planner is different. In this model, if the worker deviates in the first period, the planner will know his second period wage as he can observe the worker's income. This is not the case in both Krause (2009a) and Best and Kleven (February 2013) as work experience is only a function of labor effort. The planner seeks to reduce the incentive problem by distorting the second period labor choice. Furthermore in Best and Kleven (February 2013), the planner must use age-dependent taxes, this restricts the tools he has to use to tackle the information problem. The informational structure of this essay offers the most contrasting results in the case of no uncertainty as the distortions when young are necessarily greater than the ones when old.

### 2.2.3.2 Impact of complementarity and variance on the second period labor wedge

In contrast to the riskless environment just highlighted above, uncertainty in the second period forces the planner to distort the second period labor market as shown by (2.26). The assumptions made on the wage function and the probability distribution allows to characterize in more detail the labor distortion at the right tail of the distribution. This is done by applying the methodology of Golosov et al. (2015) which consists of investigating the asymptotic behavior of (2.26) as  $\theta_2$  goes to infinity. For this, the assumption of a bounded distribution needs to be relaxed and assume that the shocks are non-negative, i.e.  $\theta_2 \in \Theta = \mathbb{R}_+$  in  $t = 2$ .

**Assumption 2.**  $\alpha_2(\theta)$ ,  $\eta_2(\theta)$  have finite, non-zero limits  $\bar{\alpha}$ ,  $\bar{\eta}$ ;  $\frac{c_2(\theta)}{y_2(\theta)}$  has a finite non-zero

limit;  $\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)}$  has a finite limit;  $\frac{\dot{c}_2(\theta)/c_2(\theta)}{\dot{y}_2(\theta)/y_2(\theta)}$  has a limit as  $\theta \rightarrow \infty$ .<sup>29</sup>

Assumption 2 is made to guarantee well-behaved cases. In addition, note that as  $\theta_2 \rightarrow \infty$  the CES wage function (2.30) has a limit  $\bar{\epsilon}_{\theta_2}$ . For parameter values of  $\rho \in [0, 1]$  the wage elasticity with respect to shock goes to 1 as the shock goes to infinity, i.e.  $\epsilon_{\theta_2}(\theta, e_2) \rightarrow 1$  ( $\theta \rightarrow \infty$ ). For a parameter value above 1 the wage elasticity with respect to shock goes to 0 as the shock goes to infinity, i.e.  $\epsilon_{\theta_2}(\theta, e_2) \rightarrow 0$  ( $\theta \rightarrow \infty$ ). These two facts lead to the following Corollaries.

**Corollary 2.a.** *Under Assumption 2, a CES wage function with  $0 \leq \rho \leq 1$  and  $f^2(\theta)$  distributed  $\ln \mathcal{N}(\mu, \sigma^2)$ , as  $\theta \rightarrow \infty$  the second period labor distortion is asymptotically equivalent to*

$$\frac{\tau_L^{2*}(\theta)}{1 - \tau_L^{2*}(\theta)} \sim A(\theta)B(\theta)C(\theta) \sim (1 + \bar{\alpha}) \bar{\epsilon}_{\theta_2} \left( \frac{\sigma^2}{\ln \theta - \mu} \right),$$

where

$$\bar{\epsilon}_{\theta_2} = 1.<sup>30</sup>$$

As can be seen from Corollary 2.a, the optimal second period wedge as  $\theta$  goes to infinity is shaped by the labor elasticity parameter  $\bar{\alpha}$ , the wage elasticity parameter with respect to second period shock  $\bar{\epsilon}_{\theta_2}$  and the variance parameter of the log-normal distribution  $\sigma^2$ . Thus the greater the variance parameter, capturing uncertainty to the worker, the higher the labor distortion when old. The intuition of this result can be obtained by taking each part of (2.27) individually. First, consider  $A(\theta)$  which measures the cost of distorting the labor decision. This part converges to a finite and positive limit as  $\theta$  goes to infinity. The redistributive part

<sup>29</sup>For ease of exposition, we replace  $c(\theta^2), y(\theta^2)$  by  $c_2(\theta), y_2(\theta)$  and  $\tau_L(\theta^2)$  by  $\tau_L^2(\theta)$ .

<sup>30</sup>Note that if the CES has a scaling factor  $\kappa$  as in the numerical simulations below the limit of the wage elasticity with respect to shock would be  $\kappa^{1-\rho}$ .

$C(\theta)$  can also be shown to have a finite limit of 1.<sup>31</sup> It can be shown numerically that for much of the workers on the domain of the shock distribution of young workers,  $C(\theta^2)$  climbs very quickly to its asymptotic value. Finally, the hazard ratio of the log-normal distribution can be shown to go to zero as  $\theta$  goes to infinity, i.e.  $B(\theta) \rightarrow 0$  ( $\theta \rightarrow \infty$ ). Even though  $B(\theta) \rightarrow 0$  as  $\theta$  goes to infinity this rate of convergence is rather slow in fact the tail behavior can be characterized by  $B(\theta) \sim \frac{\sigma^2}{\ln \theta - \mu}$ . All of these results combined give us Corollary 2.a.

In the case of  $\rho > 1$  the result of Corollary 2.a no longer hold has the limit of the wage elasticity with respect to shock is zero and thus  $\lim_{\theta \rightarrow \infty} A(\theta) = 0$ . The following result considers the behavior of the limit of this elasticity.

**Corollary 2.b.** *Under Assumption 2, a CES wage function with  $\rho > 1$  and  $f^2(\theta)$  distributed  $\ln \mathcal{N}(\mu, \sigma^2)$ , as  $\theta \rightarrow \infty$  the second period labor distortion is asymptotically equivalent to*

$$\frac{\tau_L^{2*}(\theta)}{1 - \tau_L^{2*}(\theta)} \sim A(\theta)B(\theta)C(\theta) \sim (1 + \bar{\alpha}) \left(\frac{e_2}{\theta}\right)^{\rho-1} \left(\frac{\sigma^2}{\ln \theta - \mu}\right).^{32}$$

Note that the change in the wage elasticity does not affect our characterization of  $\lim_{\theta \rightarrow \infty} C(\theta)$  nor its interpretation. In this particular context the asymptotic behavior of  $A(\theta)$  as  $\theta$  goes to infinity is  $A(\theta) \sim (1 + \bar{\alpha}) \left(\frac{e_2}{\theta}\right)^{\rho-1}$ . Importantly,  $A(\theta)$ , contrary to  $B(\theta)$ , can go quickly to zero. In fact, the higher the value of  $\rho$  above 1 the faster  $A(\theta)$  converges to 0. Comparing the results of Corollary 2.a and 2.b, for the right tail of the distribution of shocks,

---

<sup>31</sup> $C(\theta)$  can be written in the following way:

$$\lim_{\theta \rightarrow \infty} C(\theta) = \lim_{\theta \rightarrow \infty} [1 - \lambda_2 u'(c_2(\theta))] + \lim_{\theta \rightarrow \infty} \frac{\eta_2(\theta)}{\eta_2(\theta) + \alpha_2(\theta)} \lim_{\theta \rightarrow \infty} \frac{\tau_L^{2*}(\theta)}{1 - \tau_L^{2*}(\theta)}.$$

The limit of  $C(\theta)$  can be divided into two distinct parts. The first part captures the value of extracting a dollar from a worker who received shock  $\theta$  and as  $\theta$  goes to infinity, so does  $c_2(\theta)$  and thus the value of giving (or leaving) a dollar to shock type  $\theta$ , measured by  $g(\theta) \equiv \lambda_2 u'(c_2(\theta))$ , goes to zero. The second part measures the income effect on labor effort from distorting labor on type  $\theta$ . This income effect is of course influenced by the size of the limit of the second period labor distortion.

<sup>32</sup>We also obtain

$$\lim_{\theta \rightarrow \infty} C(\theta) = 1 + \frac{\bar{\eta}}{\bar{\eta} + \bar{\alpha}} \lim_{\theta \rightarrow \infty} \frac{\tau_L^{2*}(\theta)}{1 - \tau_L^{2*}(\theta)}$$

the optimal labor distortion when old, for a given realization of  $\theta_1$  in the first period, should be greater whenever  $0 \leq \rho \leq 1$  compared to when  $\rho > 1$ . The intuition is that whenever the complementarity coefficient  $\rho$  is greater than one the variance of the shock distribution has less influence on the shape of the optimal tax schedule when old.

## 2.3 Numerical Simulations

This section investigates the properties of the optimal wedges and how they are influenced by different key parameters. To compare the results derived with those of the literature, I proceed with a calibration exercise to match certain empirical moments of the wage distribution at each age and the wage elasticity with respect to work experience found in the meta-analysis of Best and Kleven (February 2013).

### 2.3.1 Functional Forms, Calibration and Computational Strategy

#### Functional Forms

The functional form for the per period utility function used is

$$\ln c - \frac{l^{1+\alpha}}{1+\alpha}, \quad (2.33)$$

where  $\alpha > 0$ . The function that transform income into effective work experience is

$$\phi(y) = y^\delta, \quad (2.34)$$

where  $\delta \in (0, 1)$  to obtain strict concavity of  $\phi(y)$ . The wage has a the following CES functional form

$$w_t(\theta_t, e_t) = \kappa_t * (\theta_t^{1-\rho} + \xi e_t^{1-\rho})^{\frac{1}{1-\rho}}, \quad (2.35)$$

where  $\kappa_t$  and  $\xi$  are scaling parameters with  $\rho$  being the Hicksian complementarity parameter. For simplicity, I assume that all workers start with no accumulated work experience, i.e.  $e_1 = 0$ , and that  $\kappa_1 = 1$  in the first period. This implies that in the first period

$$w_1(\theta_1) = \theta_1 \quad \text{and} \quad e_2(\theta_1) = [y(\theta_1)]^\delta. \quad (2.36)$$

Using both the wage function and the transformation function, the wage elasticity with respect to work experience and the wage elasticity with respect to shock, respectively, are

$$\epsilon_{e_2} = \kappa_2^{1-\rho} * \xi * \left(\frac{e_2}{w_2}\right)^{1-\rho}, \quad \epsilon_{\theta_2} = \kappa_2^{1-\rho} * \left(\frac{\theta_2}{w_2}\right)^{1-\rho}.^{33} \quad (2.37)$$

For the calibrations the wage elasticity with respect to first period labor effort is used. This elasticity also coincides with the wage elasticity with respect to first period income

$$\gamma(\theta_2, e_2) \equiv \frac{\partial w_2}{\partial l_1} \frac{l_1}{w_2} = \frac{\partial w_2}{\partial y_1} \frac{y_1}{w_2} = \delta * \kappa_2^{1-\rho} * \xi * \left(\frac{e_2}{w_2}\right)^{1-\rho}. \quad (2.38)$$

## Calibration

The different calibrations sets  $\delta$ ,  $\kappa_2$  and  $\xi$  to match three target moments. The first target moment is the mean of the estimated wage elasticity with respect to experience in Best and Kleven (February 2013), i.e.  $\hat{\gamma} = 0.29$ .<sup>34</sup> The second and third target moments are the mean and standard deviation of the wages of head of household workers age 41 and over in the 2007 round of the Panel Study of Income Dynamics (PSID). These moments were obtained by taking the hourly wages and age from the 2007 PSID and splitting the

---

<sup>33</sup>Thus for  $0 \leq \rho \leq 1$  this implies that  $\epsilon_{\theta_2} \rightarrow \kappa_2^{1-\rho}$  as  $\theta_2 \rightarrow \infty$  and for  $\rho > 1$  it implies that  $\epsilon_{\theta_2} \rightarrow 0$  as  $\theta_2 \rightarrow \infty$ .

<sup>34</sup>In their numerical simulations this elasticity is then taken to be the wage elasticity with respect to first period labor effort. They use three scenarios where  $\gamma = 0$ ,  $\gamma = 0.2$  and  $\gamma = 0.4$ . A similar logic is followed in taking the wage elasticity with respect to experience to be the wage elasticity with respect to first period labor.

sample in two where the cut off is the median age as in Best and Kleven (February 2013).<sup>35</sup> I then approximated the distribution  $F^{young}(w)$  of the first period of life by a log-normal distribution with  $(\mu_{young}, \sigma_{young}) = (2.805, 0.672)$  and the second period of life  $F^{old}(w)$  with  $(\mu_{old}, \sigma_{old}) = (3.165, 0.814)$ .<sup>36</sup> Again for comparability with other papers in the optimal tax literature such as Mankiw et al. (2009) and Bastani (May 2014), a fraction of 5% of disabled agents in the population was imposed in each period of life. For the calibration and a majority of the numerical simulations I imposed that  $F^1(\theta) = F^2(\theta) = F^{young}(w)$ . The estimated distribution  $F^{old}(w)$  is then used to obtain the mean and standard deviation of wages in the second period of life.<sup>37</sup>

I then created a *model economy* where heterogeneous workers, who will face risk in the second period of their life, decide how much to work and consume in both periods of their life, and how much to save at gross interest rate  $R$ .<sup>38</sup> The following key parameters were set to be  $\alpha = 2$ ,  $\beta = 0.6$  and  $R = 1/\beta$ .<sup>39</sup> The assumption on  $\beta$  and  $R$  implies that savings are for insurance purposes. The 2007 US tax system is approximated by a linear labor tax following the methodology of Jacquet et al. (2013). Due to limited empirical evidence on the complementarity between the shock and work experience, several calibrations were undertaken where  $\rho$  takes on different values.<sup>40,41</sup>

## Computational Strategy

<sup>35</sup>To make sure the sample is representative I only keep observations that are related to the original SRC(Survey Research Center) sample.

<sup>36</sup>To approximate the distribution the command *fitdistr* from the MASS package in R was used. See <http://cran.r-project.org/web/packages/MASS/index.html>.

<sup>37</sup>Recall that  $F^{old}(w)$  was adjusted to take into consideration the mass of disabled workers. The mean and standard deviation of the second period wage from distribution  $F^{old}(w)$  are 31.3 and 31.49 respectively.

<sup>38</sup>The debt limit was set to 0, i.e. savings must be non-negative.

<sup>39</sup>The length of time represents roughly 25 years. If  $\varrho$  is the one-year-ahead discount factor then  $\beta = \varrho^{25}$ . In the calibrations this assumes that  $\varrho$  is close to 0.98.

<sup>40</sup>As pointed out in Stantcheva (2014), there is some evidence of complementarity with respect to skill and on-the-job training. See and OECD (2004) and Huggett et al. (2011).

<sup>41</sup>In every calibration, the parameters are chosen as to minimize the following loss function

$$LOSS = \left| \frac{\tilde{\gamma}}{0.29} - 1 \right| + \left| \frac{\tilde{w}_2}{31.3} - 1 \right| + \left| \frac{\tilde{sd}_2}{31.49} - 1 \right|,$$

Table 2.1: Calibration Parameters

Parameter	Calibration 1	Calibration 2	Calibration 3	Calibration 4	Type
$\rho$	0.2	0.5	1.2	1.5	Exogenous
$\alpha$	2	2	2	2	Exogenous
$\beta$	0.6	0.6	0.6	0.6	Exogenous
$R$	$1/\beta$	$1/\beta$	$1/\beta$	$1/\beta$	Exogenous
$\delta$	0.8	0.9	0.7	0.3	Endogenous
$\xi$	1	0.45	0.3	0.26	Endogenous
$\kappa_2$	0.8	0.9	9	5.7	Endogenous

The computational strategy closely follows the strategy used in Golosov et al. (2015). The recursive formulation of the planner’s problem is used to solve a two period discrete-time dynamic program with a two-dimensional continuous state space. It is solved by backward induction. First, the second period value function is approximated by tensor products of Chebyshev orthogonal polynomials evaluated at root nodes. Each node problem is solved using an interior-point algorithm which consists of replacing the nonlinear programming problem by a series of barrier subproblems controlled by a barrier parameter.<sup>42</sup>

For each second period node problem, a discrete-type version of (2.19) is solved where only the downward incentive constraints linking two adjacent workers and a monotonicity condition on consumption is assumed.<sup>43</sup> As was shown by Hellwig (2007) the continuous type model and discrete type model are in a sense mathematically equivalent. Bastani (May 2014) has demonstrated that simulations using either continuous models or discrete type

---

where  $\tilde{\delta}$ ,  $\tilde{w}_2$ ,  $\tilde{s}d_2$  are, respectively, the mean of the wage elasticity with respect to first period income, the mean second period wage and the standard deviation of the second period wage. Note that  $\tilde{\gamma}$  is the discrete version of

$$\tilde{\gamma} = \int \int \delta \kappa_2^{1-\rho} \xi \left( \frac{e(\theta_1)}{w_2(\theta_2, e(\theta_1))} \right)^{1-\rho} f^1(\theta_1) f^2(\theta_2) d\theta_1 d\theta_2.$$

<sup>42</sup>We use algorithm 1 in KNITRO. See [http://www.artelys.com/tools/knitro\\_doc/](http://www.artelys.com/tools/knitro_doc/).

<sup>43</sup>I have also solved each problem without imposing the monotonicity condition and found the same solutions. This condition is imposed as it speeds up computation time required to solve each problem and it imposes the second order condition explicitly as a set of constraints.

models produced similar results as long as the number of types used to represent the skill distribution is large enough. These results lend confidence that the numerical simulations of this essay are similar to the continuous type problem of (2.19) while taking advantage of solving a well-behaved convex programming problem of the discrete type model.

Once the second period value function is approximated, it is incorporated in the first period problem (2.20). I imposed  $e_1 = 0$  and looked for the  $v_0$  such that  $K(v_0, 0, 1) = 0$ . Assuming that the first period value function is continuous, a value of  $v_0$  big enough is chosen so that  $K < 0$  and a discrete version of the first period problem is solved using the chosen  $v_0$ .<sup>44</sup> Following this step, another value of  $v_0$  small enough is chosen such that  $K > 0$ . I then proceed by bisection and solved a new first period problem with an updated  $v_0$ , and this is done until the value where  $K(v_0) = 0$  is found. With this solution, the optimal allocation is obtained by forward induction. The solution is then verified to be incentive compatible.

### 2.3.2 Results

I start by analyzing the results of the benchmark case of  $\rho = 0.5$ , i.e. calibration 2 as it features the clearest results.<sup>45</sup> The upper left graphic of Figure 2.1 illustrates many of the important results found in this essay. It features the first period labor wedge and the expected value of the second period labor wedge a worker of wage  $\theta_1$  will face.<sup>46</sup> Note that except for the very low wages and very high wages in the first period, workers usually face a lower first period wedge compared with the expected value of the second period wedge.<sup>47</sup> In fact, in this simulation, 92.9% of the workers in the population are in that situation.

It is possible to see this more clearly by looking at the lower left graphic where the whole

---

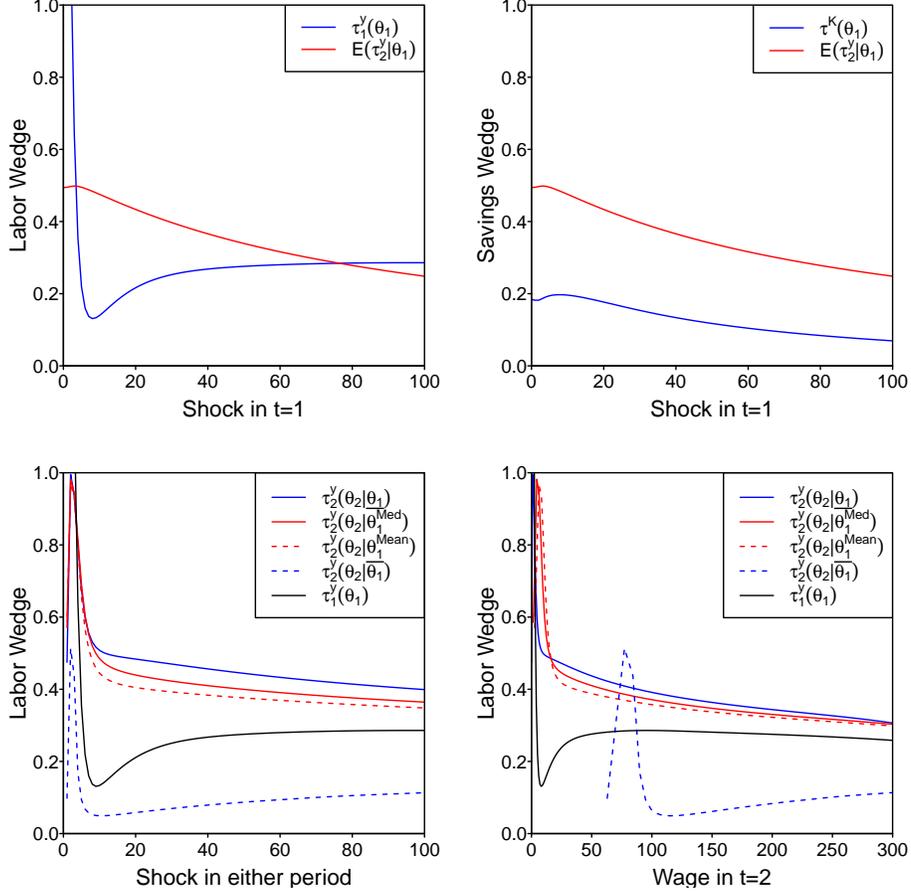
<sup>44</sup>In the first period, no monotonicity condition is imposed, but the downward incentive constraints are still used. The monotonicity and incentive compatibility is verified ex-post.

<sup>45</sup>The results for the other benchmark cases can be found in the Computational Information Appendix.

<sup>46</sup>In the simulations shocks go from 0.01 to 499.01. I restricted the x-axis since it allows a better look at what happens to the average skilled individuals and lower. Note that the share of the population that has a first period shock lower or equal to 100 is 99.6%.

<sup>47</sup>Very low skill workers face optimal first period wedges above 100% as it requires a high distortion to discount labor effort as work experience is valuable in the second period.

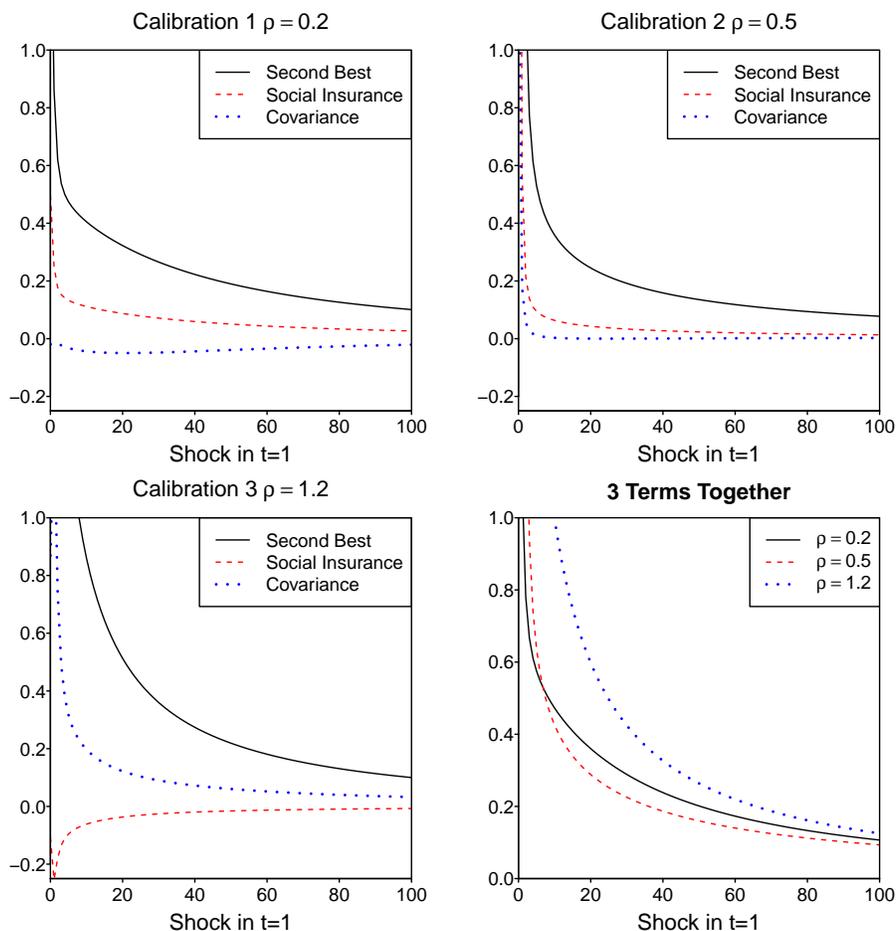
Figure 2.1: Optimal Wedges from Calibration  $\rho = 0.5$



second period optimal labor wedge schedule is graphed for workers that received the lowest, the median, the mean, and the highest shock, respectively  $\underline{\theta}_1$ ,  $\theta_1^{\text{Med}}$ ,  $\theta_1^{\text{Mean}}$  and  $\bar{\theta}_1$ .<sup>48</sup> This gives the upper and lower bound of the optimal wedge schedule in the second period for workers of different history  $\theta^2$ . The workers who received the median and mean shock in the first period should have second period labor wedges far above the labor wedge they face in the first period. In fact they should face a higher wedge than most workers in the first period except for those at the very bottom of the distribution in period 1. In the case of the highest shock in the first period, they will face a greater labor wedge in the second period

<sup>48</sup>In the simulations these values are  $\underline{\theta}_1 = 0.01\$$ ,  $\theta_1^{\text{Med}} = 16.01\$$ ,  $\theta_1^{\text{Mean}} = 20.01\$$  and  $\bar{\theta}_1 = 499.01\$$ .

Figure 2.2: Optimal Labor Wedge in  $t=1$ : Three Motives



unless they also receive the highest shock in the second period. However, by looking at the bottom right figure, it is apparent that the results from lower left figure does not imply that for an equivalent second period wage, the worker who received the high shock in the first period will always face a lower labor wedge in the second period compared to workers who received a lower shock in first period. The upper right figure plots the optimal savings wedge and the expected value of the second period labor wedge. These two curves feature similar patterns as the savings wedge is used to help in separating workers of different types in the second period.

To further illustrate the forces shaping the first period labor wedge, Figure 2.2 shows

the values of the three motives for each first period shock  $\theta_1$  for three calibrations. The first thing to notice is that the second-best motive appears to be dominant force out of all the three motives. As was shown in the section above the other two motives are influenced by whether the complementarity parameter is below or above one. The social insurance motive is positive for both calibrations where  $\rho$  is below one and negative for the calibration where  $\rho$  is above one. The incentive motive captured by the covariance of consumption and the marginal benefit of work experience appears to also be prominently determined by the complementarity coefficient. In Calibration 1, the covariance is slightly negative since  $\rho$  is smaller than one. On the contrary, in Calibration 3, the covariance is much more positive as the complementarity coefficient above one ensures that the marginal benefit of work experience is increasing. One of the interesting results comes from looking at the bottom right graph. It shows that even if the social insurance motive is negative, it is the calibration where  $\rho$  is above one that has the highest combined value. This would imply that it is the calibration where you would expect the optimal first period labor wedge to be the lowest. This turns out not to be the case, as shown in Table 2.2, in fact it is much higher than the first period labor wedges in the calibration where  $\rho$  is below one. One plausible explanation for this result could be that the higher  $\rho$  increases the wage elasticity with respect to labor effort and therefore renders the first period labor supply much less elastic, at least at very low levels of income, since in this situation the payoffs to work are much greater. This explanation would be in the spirit of the results found in Best and Kleven (February 2013). In fact, as illustrated by Table 2.2, parameter  $\rho$  appears to be a determining factor in whether the average labor distortion is increasing with age are not in the calibrations.

### 2.3.2.1 Changes in $\rho$

To explore in greater details the effects of the value of the Hicksian complementarity parameter two sets of simulations are presented. One set of simulation has the values of the

Table 2.2: Average Wedges in Each Period: All Calibrations

Parameter	$E_0(\tau_L(\theta^1))$	$E_0(\tau_L(\theta^2))$	% Lower	$E_0(\tau_K(\theta^1))$
$\rho = 0.2$	0.262	0.406	89.4%	0.161
$\rho = 0.5$	0.245	0.439	92.9%	0.175
$\rho = 1.2$	0.430	0.425	79.3%	0.188
$\rho = 1.5$	0.439	0.356	66.1%	0.166
No WE	0.365	0.533	92.9%	0.206

Source: Authors' calculations. Numbers rounded.

coefficient  $\rho$  below one and the other above one. This is done because the behaviour of the wage function changes drastically as  $\rho$  goes from values below one to above one.<sup>49</sup> The first set of simulations take the calibrated parameters  $\delta$  and  $\xi$  from Calibration 1, i.e.  $\rho = 0.2$ , and increase  $\rho$  to 0.5 and 0.8. For each new change in  $\rho$ ,  $\kappa_2$  is adjusted to match the mean of the estimated mean of the second period wage.<sup>50</sup> The second set of simulations take Calibration 3, i.e.  $\rho = 1.2$  as its basis and  $\rho$  is increased to 1.5 and 1.8. For the later simulations, the parameter  $\delta$  and  $\xi$  from Calibration 3 are kept and  $\kappa_2$  is adjusted to match the first moment of the second period wage. In addition, the results obtained when there is no work experience (No WE) in the model are reported. The second period wage wage is  $w_2 = \kappa_2 * \theta_2$ . For this model,  $\kappa_2$  was calibrated again to match the estimated mean second period wage.

Looking at Figure 2.3 several things jump out as  $\rho$  is increased in the first set of simulations, i.e. where  $\rho < 1$ . The first is how the first period wedge decreases significantly for workers who have a first period wage below 20\$ an hour, roughly 60% of the population. For these workers the lowest labor wedge is about 19%, in the  $\rho = 0.2$  case, and drops to close to 0% in the  $\rho = 0.8$  scenario. As shown in Table 2.3, the mean first period labor wedge goes from 26.2% to 18.7%.

<sup>49</sup>This fact extends to the optimal labor wedge from both our analytical results of the asymptotic behaviour of the second period labor wedge and the numerical results found in Figure 2.2 in the Appendix.

<sup>50</sup>As  $\rho$  increases from 0.2 to 0.8 the calibrated  $\kappa_2$  is decreased from 1 to 0.12. In the end this reduces the wage elasticity with respect to labor.

The change in  $\rho$  also has a profound impact on the second period labor wedge. In the bottom left figure of Figure 2.3, the average second period labor wedge schedule flattens with an increase in  $\rho$ . Furthermore, the disabled worker in the first period's average second period labor wedge goes from the highest in the population to the lowest for almost all workers in the economy. Thus an increase in  $\rho$  lowers the expected labor distortion in the second period for a majority of the population. A similar pattern emerges for the savings wedge schedule, but the impact is much less on the average savings wedge. From the bottom right figure, it is possible to partially see the process of how the average labor wedge in second period is changed. The second period labor wedge schedule is lower around the median second period shock and it gets lower as  $\rho$  is increased.<sup>51</sup> As shown in Table 2.3, even if both labor wedges decrease with an increase in  $\rho$ , the cross-sectional average of the labor distortion is increasing with age.

Table 2.3: Average Wedges in Each Period: Using Calibration of  $\rho = 0.2$  with Changes in  $\rho$

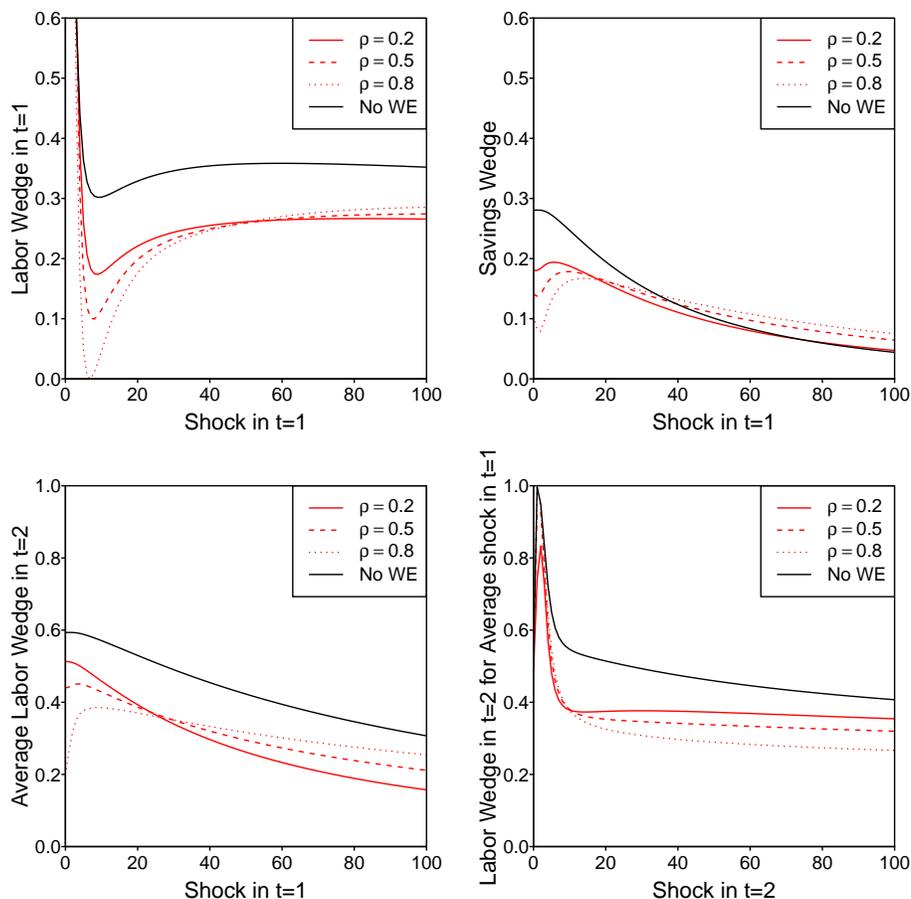
Parameter	$E_0(\tau_L(\theta^1))$	$E_0(\tau_L(\theta^2))$	% Lower	$E_0(\tau_K(\theta^1))$
$\rho = 0.2$	0.262	0.406	89.4%	0.161
$\rho = 0.5$	0.227	0.392	91.8%	0.158
$\rho = 0.8$	0.187	0.358	92.9%	0.149
No WE	0.365	0.533	92.9%	0.206

Source: Authors' calculations. Numbers rounded.

The second set of simulations, i.e.  $\rho > 1$ , gives similar results but are much less contrasting. The top left figure shows the results of increasing  $\rho$  from 1.2 to 1.5. It lowers the optimal first period labor wedge for workers with low skills but also the average and slightly higher shocks. However, this is no longer true once we go from 1.5 to 1.8. In that case, the labor wedge is increased in the first period for workers who roughly earn less than 20\$ an hour. This result is also reflected in Table 2.4 where the lowest average first period labor wedge is in the  $\rho = 1.5$  simulation.

<sup>51</sup>The general shape is almost identical for the worker who obtained the median shock in the first period.

Figure 2.3: Optimal Wedges:  $\rho = 0.2$  with Changes in  $\rho$



The second period labor wedge follows a similar pattern as the first set of simulations with  $\rho < 1$ . Increasing  $\rho$  lowers the average second period labor wedge each shock level will face in the second period. The difference can be quite significant as seen in Table 2.4, the average second period labor wedge is about 8% lower in simulation with  $\rho = 1.8$  compared to the one with  $\rho = 1.2$ . And since the second period labor distortion is decreased with an increase in  $\rho$  a similar pattern emerges for the savings wedge and can be seen in the top right figure of Figure 2.4 and the average savings wedge in Table 2.4. One thing to note from this second set of simulations is that the fraction of the population that will face a lower labor wedge in the first period of their life compared to their expected second period is

decreasing as the shock and work experience becomes more and more complimentary. The results on the second period labor wedge are inline with the analytical results presented above. Because  $\rho > 1$ , the wage elasticity with respect to shock found in (2.26) goes to zero as  $\theta_2$  increases. The greater  $\rho$  is the faster is the convergence and thus pushes the optimal second period labor wedge down much quicker. But as mentioned above, a higher value of  $\rho$  increases the wage elasticity with respect to labor effort at the bottom of the distribution which may explain the reason why the cross-sectional average appear to be higher in the first period of life compared to the one in the second period. This would also account for the decrease in the average second period labor wedge as it would increase the elasticity of second period labor with respect to second period taxation.

Table 2.4: Average Wedges in Each Period: Using Calibration of  $\rho = 1.2$  with Changes in  $\rho$

Parameter	$E_0(\tau_L(\theta^1))$	$E_0(\tau_L(\theta^2))$	% Lower	$E_0(\tau_K(\theta^1))$
$\rho = 1.2$	0.430	0.425	79.3%	0.188
$\rho = 1.5$	0.378	0.385	79.3%	0.169
$\rho = 1.8$	0.431	0.349	66.8%	0.157
No WE	0.365	0.533	92.9%	0.206

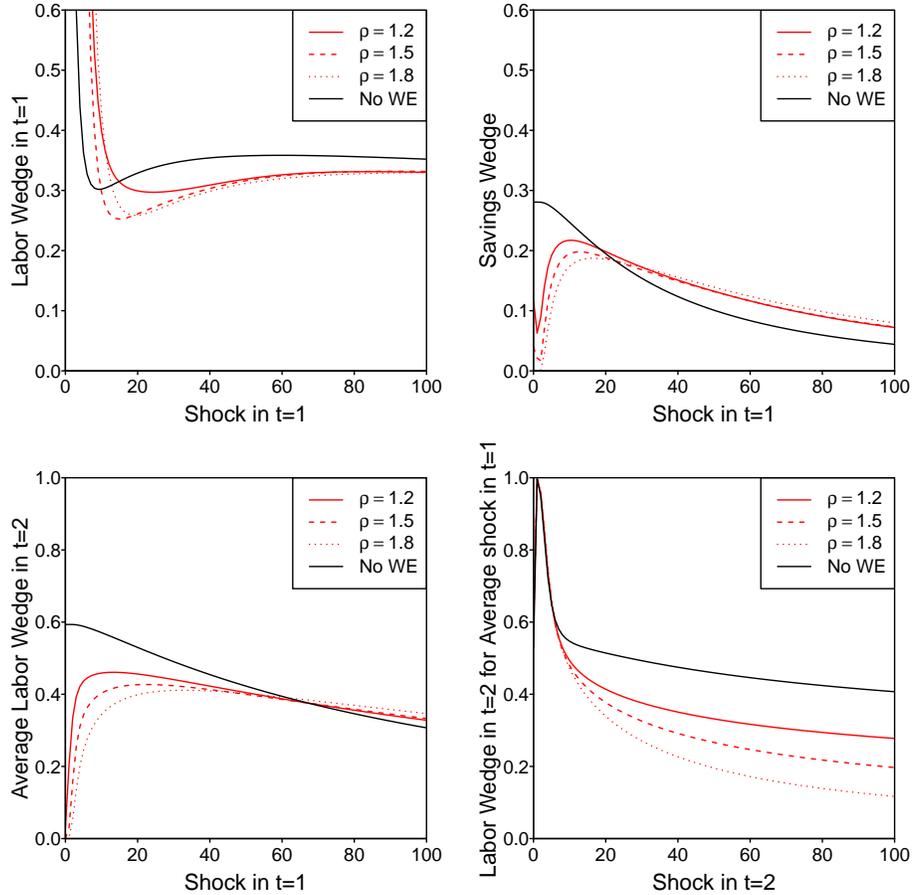
Source: Authors' calculations. Numbers rounded.

### 2.3.2.2 Changes in $\delta$

The next set of simulations investigates the role of the parameter that influences how income is transformed into work experience. It is the parameter that more directly influences the elasticity of future wages with respect to labor effort. Contrary to changes in  $\rho$  its effect is felt by all workers not more keenly felt by specific parts of the distribution. Taking Calibration 2 as the basis, the value of  $\delta$  is lowered to 0.6 and then 0.3.<sup>52</sup> For each reduction of  $\delta$  an extra calibration is done where  $\kappa_2$  is adjusted such that the mean second period wage matches the

<sup>52</sup>Recall that  $\delta = 0.9$  is used in Calibration 2.

Figure 2.4: Optimal Wedges:  $\rho = 1.2$  with Changes in  $\rho$



estimated mean of the second period wage.<sup>53</sup>

Results from decreasing  $\delta$  can be observed in Figure 2.5 and Table 2.5. I find that the higher  $\delta$  is, and hence the more elastic the second period wage is with respect to labor, the lower the optimal wedge schedule in both periods. From the top left figure, it can be seen that the effect is more pronounced in the first period for the workers close and around the median and mean first period wage. The two bottom figures of Figure 2.5 shows what happens in the second period of life. As the bottom right figures demonstrates, for the

<sup>53</sup>This results in a gradual increase of  $\kappa_2$  with the decrease in  $\delta$ . Note that the value of  $\delta * \kappa_2^{1-\rho}$  is always close to the value of  $\delta$  as the calibration value of  $\kappa_2$  is always slightly above 1. This also slightly effect the behaviour of the second period wedge as  $\lim_{\theta \rightarrow \bar{\theta}} \epsilon_{\theta 2}$  will be slightly pushed upwards with an increase in  $\kappa_2$ .

worker who received the mean wage in the first period, the labor wedge in the second period is very similar for most of the second period shocks below the value of the median shock.<sup>54</sup> This changes as the second period shock gets higher but then gets close once again. In these simulations, irrespective of the value of  $\delta$ , an overwhelming majority of workers will have a lower first period labor wedge when compared to the expected second period wedge they will face knowing their first period shock.

A major effect of a change in  $\delta$  is on the shape of the savings wedge schedule. The more elastic the second period wage is with respect to labor the lower the savings wedge is for roughly 80% of the population whereas it slightly increases for the rest of the population. So a higher wage elasticity also lowers the distortion made to the savings decision but as observed in Table 2.5, this does not necessarily translate to a lower average savings wedge.

The results here are in contrast with the results of Best and Kleven (February 2013) who find that an increase in  $\delta$  usually leads to higher labor distortions for the young than for the old. In the setting presented here, it appears that this is not case. One explanation is that the parameter in our model does not entirely capture the wage elasticity with respect to labor and an important parameter for this elasticity is  $\rho$ . And as shown in the analytical results  $\rho$  also influences other considerations of the planner that are not only related to labor elasticities. One example of this would be the social insurance motive of the planner coming from uncertainty in the second period.

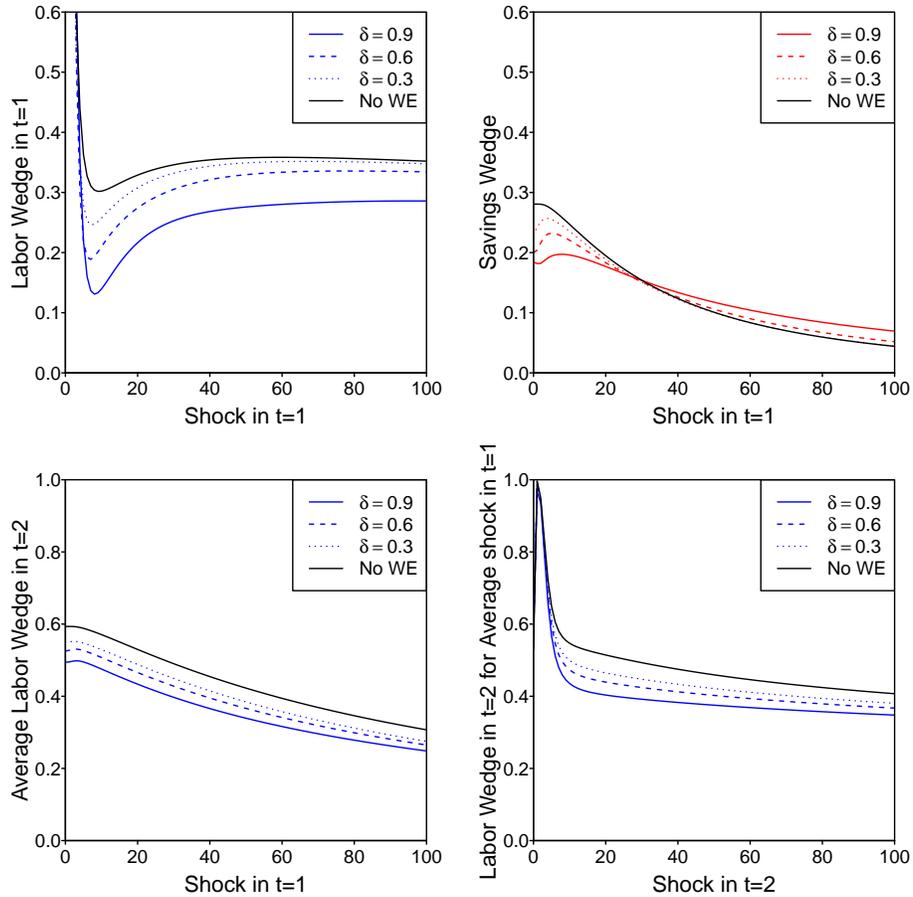
Table 2.5: Average Wedges in Each Period:  $\rho = 0.5$  with Changes in  $\delta$

Parameter	$E_0(\tau_L(\theta^1))$	$E_0(\tau_L(\theta^2))$	% Lower	$E_0(\tau_K(\theta^1))$
$\delta = 0.3$	0.336	0.492	92.5%	0.156
$\delta = 0.6$	0.300	0.470	92.5%	0.187
$\delta = 0.9$	0.249	0.439	92.9%	0.175
No WE	0.365	0.533	92.9%	0.206

Source: Authors' calculations. Numbers rounded.

<sup>54</sup>This pattern is also similar for the workers who receive the median shock in the second period.

Figure 2.5: Optimal Wedges:  $\rho = 0.5$  with Changes in  $\delta$

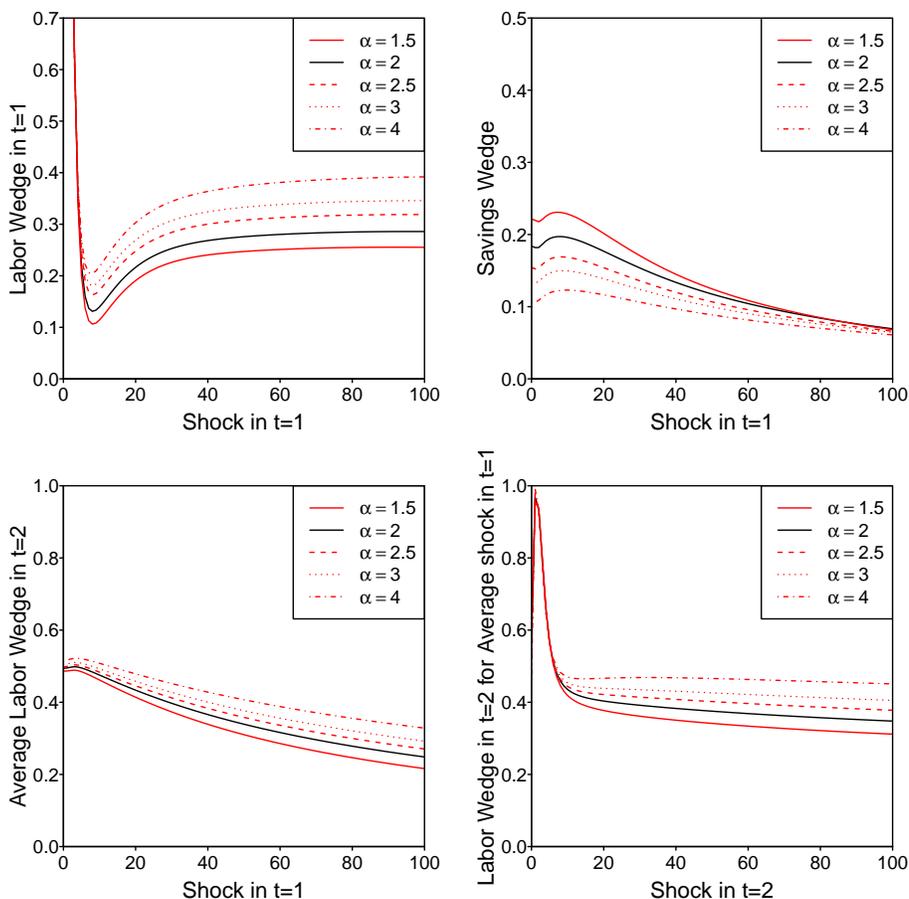


### 2.3.2.3 Changes in $\alpha$

These simulations look at the impact of changing the parameter that controls the cost of labor effort. Calibration 2 is taken as a basis and  $\kappa_2$  is adjusted to match the estimated mean of the second period wage for each new value of  $\alpha$ .

Figure 2.6 and Table 2.6 show the effect on the optimal wedges as  $\alpha$  is increased. As the value of  $\alpha$  is increased, which results in a decrease in the labor supply elasticity in each period, the optimal distortions are increased in both periods. The smaller labor supply elasticity allows the planner to redistribute/insure more as the efficiency costs of such policies

Figure 2.6: Optimal Wedges:  $\rho = 0.5$  with Changes in  $\alpha$



are reduced. The increase in labor distortion seem to impact slightly more the first period, as the increase in average labor wedge increases by about 10% when  $\alpha$  is taken from 1.5 to 4. Whereas the increase in second period average labor wedge is of roughly 6%. In these simulations the average labor wedge when old remain higher than the average labor wedge when young. As the labor supply elasticity decreases, the usefulness of distorting savings to help separate workers in the second period is also decreased. Notice that as  $\alpha$  is taken from 1.5 to 4 the average savings wedge goes from 20% to about 11%.

Table 2.6: Average Wedges in Each Period:  $\rho = 0.5$  with Changes in  $\alpha$

Parameter	$E_0(\tau_L(\theta^1))$	$E_0(\tau_L(\theta^2))$	% Lower	$E_0(\tau_K(\theta^1))$
$\alpha = 1.5$	0.225	0.420	92.9%	0.200
$\alpha = 2$	0.249	0.439	92.9%	0.175
$\alpha = 2.5$	0.278	0.450	92.5%	0.152
$\alpha = 3$	0.297	0.462	92.3%	0.136
$\alpha = 4$	0.327	0.481	91.8%	0.114

Source: Authors' calculations. Numbers rounded.

### 2.3.2.4 Changes in $\sigma$

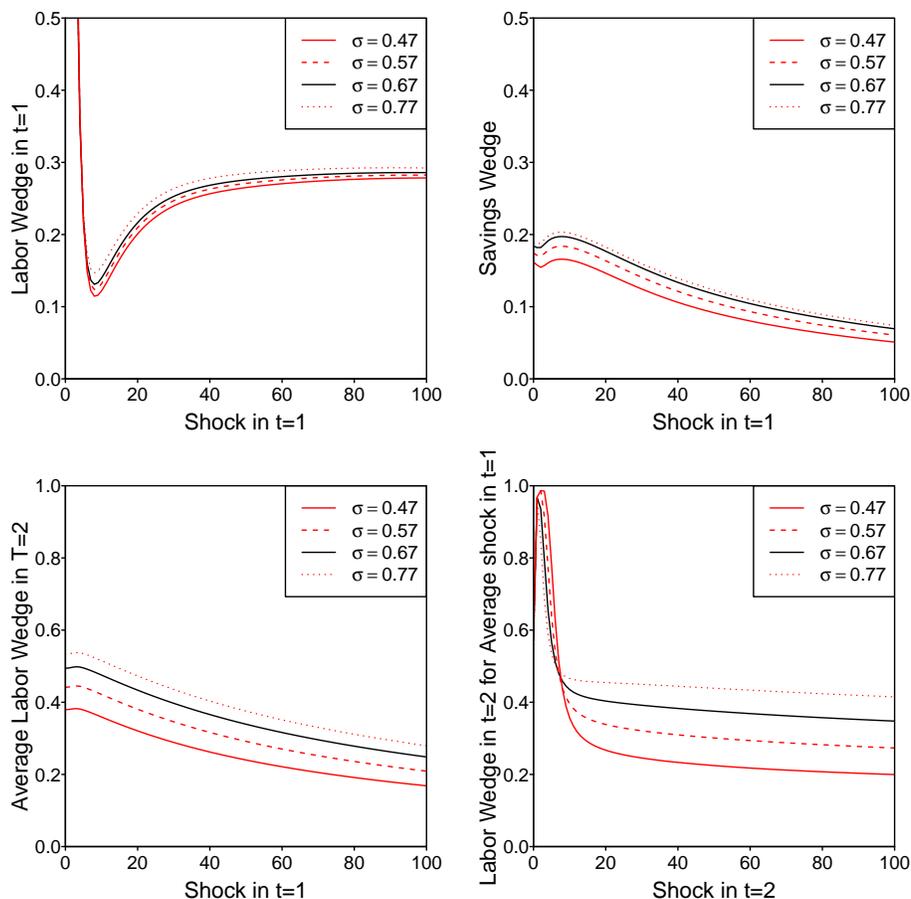
The last set of simulations focuses on the change in the *riskiness* of the second period. This is done by changing the value of the scale parameter  $\sigma$  of the log-normal distribution of the second period distribution of shocks. The distribution of shocks is no longer identical in both period. The first period distribution is the same as in all other simulations above, i.e.  $F^1(\theta) = F^{young}(\theta)$ . The second period distribution will be a modification of  $F^{young}(\theta)$  as the new scale parameter  $\sigma$  is changed to obtain  $F^2(\theta)$ .<sup>55</sup> An increase in  $\sigma$  results in an increase in the variance of the shock and hence the variance of the second period wage for a given level of work experience. These changes will have an impact on the mean of the shock but not on the median shock. Again, Calibration 2 is used for most of the parameters but  $\kappa_2$  is changed to keep the mean of the second period wage matching the same as the above simulations.<sup>56</sup>

Figure 2.7 and Table 2.7 illustrates that little changes for the first period labor wedge as  $\sigma$ , i.e. the riskiness, is increased. The labor wedge slightly increases and that is reflected by the 2.5% increase in the average first period labor wedge. The most important change happens in the second period of life. This is inline with the analytical results that showed that an increase in  $\sigma$ , when the distribution of shocks is log-normal, results in an increase in

<sup>55</sup>The distribution is also rescaled as to impose the 5% of disabled agents as in the first period distribution.

<sup>56</sup>Note that  $\sigma = 0.67$  simulation is the same found in Figure 2.1.

Figure 2.7: Optimal Wedges:  $\rho = 0.5$  with Changes in in  $t=2$ 's  $\sigma$



the second period labor wedge. As shown in Table 2.7, the mean second period labor wedge increases by roughly 15% as  $\sigma$  is taken from 0.47 to 0.77. As risk increase the different between the average labor distortion when young and when old increases as well.<sup>57</sup> This added uncertainty is an important explanation for the the age structure of taxation to be increasing. Furthermore, as an increase in  $\sigma$  implies a greater labor distortions in the second period, the savings wedge is also increased to facilitate the separation of types.

<sup>57</sup>It is quite possible that the log-normal distribution used in the numerical simulations of this essay overestimates the risk workers face in the second period of life. Considering a distribution constructed from a mixture of log-normals as in Golosov et al. (2015), would not alter the results as the highest  $\sigma$  of the different log-normal distribution would be the one driving the results, at least at the right tail of the distribution.

Table 2.7: Average Wedges in Each Period:  $\rho = 0.5$  with Changes in  $t=2$ 's  $\sigma$

Parameter	$E_0(\tau_L(\theta^1))$	$E_0(\tau_L(\theta^2))$	% Lower	$E_0(\tau_K(\theta^1))$
$\sigma = 0.47$	0.236	0.327	85.8%	0.145
$\sigma = 0.57$	0.243	0.387	91.0%	0.162
$\sigma = 0.67$	0.245	0.439	92.9%	0.175
$\sigma = 0.77$	0.261	0.478	93.5%	0.181

Source: Authors' calculations. Numbers rounded.

## 2.4 Conclusion

This essay studies the design of an optimal dynamic tax system when wages depend on both accumulated work experience and a stochastic shock. The analysis is made under the assumption that work experience is observable to the planner. Using a first-order approach to solve the planner's problem, I find that in addition to standard considerations found in the standard optimal taxation, the first period labor tax is a balance of three motives that captures the added effects of having wages be a function of risk and work experience.

The main contribution of the essay is to highlight the importance of the parameter determining the complementarity between the stochastic shock and the accumulated work experience on the age structure of the optimal tax system. I find that a majority of workers, in numerical simulations calibrated to US data, would expect to face a higher marginal labor income tax when old compared to the one they faced when young. Nevertheless, whether the cross-sectional average of the marginal labor income tax increases or decreases with age depends on the complementarity parameter. A parameter value smaller than one leads to increasing average labor taxes with age whereas a parameter value above one leads to declining average labor taxes with age. The above results illustrate the importance of obtaining estimates of the value of the complementarity parameter if policy makers are considering tax reforms towards an age-dependent tax system.

In future work, I plan on relaxing the assumption of observable work experience accumulation and consider the case where work experience is a function of first period shock and labor effort. This would allow for different types of job opportunities and rewards to be part of the model. However, this comes at the cost of losing the *common knowledge of preference* assumption as work experience becomes unobservable. In the model above, work experience in one profession is in a way perfectly transferable to another profession. Considering the impact of labor income taxation on job-specific work experience may relax the incentive constraints on workers that decide to change occupation in response to labor income taxes.

## Chapter 3

# Optimal Income Taxation with Unemployment Induced Loss of Human Capital

An important issue in the optimal redistributive tax literature has been the role of income taxation and the accumulation of human capital. When considering human capital (Bovenberg and Jacobs, 2005; Jacobs, 2005), the focus has mainly been on income taxation, education policy and the educational choice of the individual. Despite a large body of empirical evidence linking work experience with wage rates (Blundell and MaCurdy, 1999) validating the idea that past labor market participation improves the worker's productivity, human capital acquisition through learning-by-doing has been ignored in the optimal tax literature until recently (Krause, 2009a; Best and Kleven, February 2013; Kapicka, 2014). Both the education and learning-by-doing models assume that workers fully control the level of either education or work effort that will result in an increase in human capital. However, in the presence of involuntary unemployment, the decision to work or how much to work can be entirely irrelevant in the accumulation of human capital if one is unable to find work. Thus for many workers the actual level of human capital acquired is determined by factors that are out of their control.

The phenomenon called 'unemployment scarring' relates to the destruction of human

capital or productivity while unemployed. This reflects the long-term consequences of experiencing a period of unemployment. Although there can be many types of unemployment scarring that can potentially affect an individual's working life, the one connecting unemployment to wage rates can be drastic. Empirical evidence suggest that after experiencing a spell of unemployment, workers, and especially young workers, will face the prospect of lower wages in their next employment opportunity (Arulampalam, 2001; Arulampalam et al., 2001; Gregg and Tominey, 2005). These negative effects on wages can even last decades after the last spell of unemployment. One prominent explanation for this phenomenon is that unemployment spells preclude the worker from acquiring work experience but also brings about the deterioration of the worker's general skill level. This 'unlearning-by-not-doing' is a result of the labor market environment over which the worker has no control which is the situation at the heart of this essay.

To take account of this reality, I propose a model of optimal income taxation with non-Walrasian labor markets. In the model income taxation affects wages, unemployment and at the same time the future distribution of human capital levels. I use a two-period model where frictional unemployment destroys a fraction of a skilled worker's human capital. The model of the labor market is based on the search and matching literature developed by Diamond (1982), Mortensen (1982), Pissarides (2000). The essay extends the optimal tax framework with search frictions developed by Hungerbühler et al. (2006, 2008); Lehmann et al. (2011) and Jacquet et al. (2014) to incorporate dynamic considerations. To emphasize the importance of non-employment being a possibility despite a worker's best efforts, I assume that individuals always participate (or search in the labor market) and that voluntary unemployment is not an option. Therefore, the loss of human capital is never a conscious choice. This model is in contrast to the models of Krause (2009a), Best and Kleven (February 2013) and Kapicka (2014) as there is no intensive margin decision on labor. Although every individual participates in the labor market, the model of this essay is closer to the extensive

margin analyses of Diamond (1980) and Saez (2002b).

I find that the optimal allocation does not feature the standard result of no-distortion at the top, which in this case would be not to distort the before-tax wage and consumption choice in the skilled labor market.<sup>1</sup> In fact, the skilled workers before-tax wage is distorted downward away from the efficient level. The fact that the planner is only able to observe income places limits on his ability to redistribute as he wishes, and it forces the planner to adopt redistributive taxation in the second period as well as the first. This has two effects.

The first effect is the incentive effect which can also be broken down in two parts. More redistribution from the skilled to the unskilled workers in the second period makes being a skilled worker in that period less appealing. This has a direct impact on wage formation in the first period, because a skilled worker would be willing to accept a lower pre-tax wage in exchange for a higher probability of retaining his human capital level. By resorting to redistributive taxation in the second period, the planner lowers the reward of keeping the high human capital, thus it reduces the skilled worker's willingness to accept a lower pre-tax wage which has an upward effect on the equilibrium wage. To counteract this increase in the pre-tax wage above the efficient level, the planner must reduce the ability of the firm to transfer utility through the before-tax wage to the worker by increasing the marginal tax rate.<sup>2</sup> This results in lowering the equilibrium pre-tax wage and raising the employment level closer to the efficient level. This increase in the employment level is the other part of the incentive effect. The planner will seek to trade off the reduction in the value of keeping the human capital against a higher probability of keeping it by lowering the before-tax wage in the first period.

The second effect is the fiscal effect or a second-best effect that is absent from other

---

<sup>1</sup>Best and Kleven (February 2013) side-step the issue of no-distortion at the top by assuming an unbounded distribution of skills. As for the static models of Hungerbühler et al. (2006) and Lehmann et al. (2011), the optimal allocations are characterized by the no-distortion-at-top result.

<sup>2</sup>This effect can also be achieved by reducing the marginal net gain of keeping the high human capital level. This will be made clearer in section 4.1.

human capital and income tax results in the literature but found in the literature where the planner distorts more than one market. In fact, the results found in this essay are more in line with those of Stiglitz (1982) who relies on general equilibrium effects to redistribute. Here the difference stems from the fact that in our model distorting the allocation of the skilled types has fiscal benefits that come from modifying the distribution of types in the second period. Because the planner must redistribute from the skilled to the unskilled in the second period, he has now an incentive to create more employment in the skilled sector. The reason for this is threefold: having more skilled workers raises more revenue, the tax burden can be shared over more workers and allows the planner to make it smaller, and finally it reduces the number of workers the planner must redistribute to.

The important message to take away from these results is that both the incentive and fiscal effects are dependent on the presence of both the informational problem and the human capital destruction problem. Taken individually, each problem will not generate these effects. If there is only the human capital destruction problem, the planner is able to set up his tax system to achieve the first-best wage level. With only the informational problem, second period policies will not affect the wage determination process in the first period and the policies of the first period will not affect the second period.

I show that, in a similar fashion to Hungerbühler et al. (2006) and Lehmann et al. (2011), the wage of the unskilled workers will be distorted downwards from the efficient level to facilitate redistribution. To achieve this, the marginal tax rate faced by the unskilled must be raised higher than the average tax rate (in this case, the ratio between the transfer he receives and his market wage). Redistribution of income occurs through both direct transfers and a higher probability of employment of the unskilled workers resulting from the lower equilibrium wage.

I demonstrate that introducing a training program that could retrain unemployed previously skilled workers can put downward pressure on the first-period wage of the unskilled

worker. Since the presence of the training program can reduce the information problem it also reduces the need to distort the wage downward. Having a training program will increase the amount of redistribution undertaken in the second period because there are more skilled workers now for the same employment level.

Introducing job creation subsidies can also relax the informational problem of the planner because each labor market will react differently to a given level of subsidy. Job creation subsidies are found to be useless for reasonable degrees of market frictions. The optimal tax system is enough to achieve the redistributive goals of the planner in most cases. When market frictions achieve dramatic levels, i.e. employment levels below 50% in our numerical simulations, which leads the planner to distort the unskilled worker's wage to zero, then it is optimal to have positive levels of subsidies.

This essay contributes to the literature that features human capital depreciation caused by unemployment. So far, the literature has concentrated mainly on the design of the optimal unemployment insurance system (Shimer and Werning, 2006; Pavoni, 2009; Spinnewijn, 2013) with the exception of Coles and Masters (2000) who look at the optimal labor policies that maximize the surplus generated in the steady state.

Related to this article, Stiglitz (1982) and Naito (1999) investigate the property of the optimal nonlinear tax system in a two-type economy when wages are endogenous due to the imperfect substitutions of types in the production function. Engström (2002) and Boone and Bovenberg (2004) both analyze optimal redistributive nonlinear taxation in a search framework but feature exogenous wages. Boadway et al. (2003) consider optimal employment and redistributive policies with frictional unemployment and observable skills.

The essay is closely linked to papers where non-linear taxation in a search frictions framework influences wages and employment as in Hungerbühler et al. (2006, 2008); Lehmann et al. (2011) and Jacquet et al. (2014). Taxes in these models affects the economy through what Lehmann et al. (2011) call the 'wage-cum-labor-demand' margin. An example of a

framework featuring this margin is a matching model where before-tax wages are determined by Nash bargaining.<sup>3</sup> A rise in the marginal tax rate reduces the reward to negotiating aggressively on the part of the worker because an increase in before-tax wage results in a lower increase in the after-tax wage. Thus the ‘wage moderation’ effect of the increasing marginal tax rate will lower the negotiated before-tax wage which will increase labor demand (and employment). However, this effect on employment is not free and can come with an efficiency cost, since in a model of search and matching, too much employment can mean that the total output minus the costs needed to create all these employment opportunities can be lower than under a situation with lower levels of employment.

Another related strand of literature is the recent ‘New Public Finance’ (Goloso et al., 2003; Kocherlakota, 2005; Goloso et al., 2007) which focuses on dynamic optimal taxation in which the wage rate can change over time due to a random process, but never depends on the policy choices of the planner. Bohacek and Kapicka (2008) investigate the optimal income tax and optimal schooling subsidies in a dynamic private information economy with observable human capital accumulation. They find that schooling subsidies can be greatly welfare improving if the income tax is not set optimally. However if the optimal income tax is set optimally their effects on welfare are small and the marginal schooling subsidies are positive and smaller than the marginal tax rates. It is also tied to work on optimal taxation with human capital in risky environments where human capital is acquired through either schooling or on-the-job training programs (Kapicka, May 2014; Kapicka and Neira, October 2014; Stantcheva, 2014).

The essay also contributes to the literature on optimal taxation with learning-by-doing. Krause (2009a) considers optimal taxation in a two-type model where the planner can set a

---

<sup>3</sup>Other models of labor markets with matching frictions giving rise to such a margin include, for example, the Competitive Search Equilibrium (Moen, 1997) and the monopoly union model of Mortensen and Pissarides (1999).

two-period income tax schedule. One finding of the paper is that the no-distortion-at-the-top result does not apply. In addition, there are cases where it is justified to tax the high skilled workers even if it depresses both labor supply and future wages. Best and Kleven (February 2013) considers an economy with a continuum of types. They study both age-independent and age-dependent taxation and their numerical simulations make a strong case for higher age-dependent income tax rates for the young. Chapter 2 features learning-by-doing through the work experience accumulation. Like Krause (2009a) and Best and Kleven (February 2013) the above chapter considers the labor choice along the intensive margin whereas the essay of chapter 3 considers a scenario that is closer to an extensive margin framework.

### 3.1 The Model

The economy is populated by a continuum of risk-neutral individuals divided into two types who live for two periods ( $t = 1, 2$ ). Each type of individual is characterized by their productivity level. A skilled individual has productivity level  $a_h$  and an unskilled individual has level  $a_\ell \equiv \delta a_h$ , where  $0 < \delta < 1$ . In the first period there is a fraction  $\pi_\ell$  of unskilled individuals and a fraction  $\pi_h$  of skilled individuals with  $\pi_\ell + \pi_h = 1$ .

Each period features a distinct model of the labor market. In the first period the labor market is plagued by search and matching frictions *à la* Mortensen and Pissarides (1999). In this period, individuals of productivity  $a_i$  direct their search in markets of jobs of skill type  $i$ . For simplicity, assume that individuals of skill  $i$  who search in markets other than  $i$  and find a job will be unable to produce any output. This assumption, made for tractability, will be relaxed in the next period.<sup>4</sup> As in the search and matching literature workers and firms get together according to a constant-return-to-scale function  $M(U_i, V_i)$  which determines the number of matches in a type  $i$  labor market. This function is twice-continuously

---

<sup>4</sup>Although a strong assumption, segmentation is more plausible than the other extreme case of one single labor market for all skill levels.

differentiable, increasing and concave in the number of individuals searching ( $U_i$ ) and the number of vacancies opened ( $V_i$ ), and also satisfies

$$M(0, V_i) = M(U_i, 0) = 0, \text{ and } M(U_i, V_i) < \min(U_i, V_i).$$

Matches last for only one period as in Pissarides (1992). This is assumed in order to keep the number of wages in the second period to two which makes the analysis simpler. Due to matching frictions, there will be unemployment in the first period. This unemployment is especially costly for skilled individuals since it destroys a fraction of their human capital, which in this essay is modeled as their productivity level. I assume that skilled individuals who become unemployed lose enough human capital to be considered an unskilled individual in the future. Equivalently unemployment destroys  $(1 - \delta)$  of the skilled individual's productivity.

In the second period, there are no longer any matching frictions and individuals face a perfectly elastic labor demand. In this period everyone will be employed, but contrary to the previous period, individuals are able to choose in which skill sector they want to work. To be more precise, skilled individuals can decide to work in either the skilled or in the unskilled sector. This is modeled as the occupational choice framework of Piketty (1997) and Saez (2002b, 2004), and has a similar interpretation of Saez (2004) as the long-term choice of the individual. What determines this choice of occupation is a nonnegative individual cost  $\alpha_h$  of working in the skilled sector that the skilled individual discovers after working in that sector in the first period. Unskilled individuals lack the necessary skill to work in the skilled sector. Thus they are bound to work in the unskilled sector as in Diamond (2006) and Diamond and Spinnewijn (2011).

I assume that the planner is unable to observe the skill type of individuals and the second period individual cost of working in the skilled sector. However, he is able to observe the

wage  $w_{ti}$  offered to workers of type  $i$  in each period  $t$ .<sup>5</sup> With the assumption that the planner is able to commit to his tax policy in both periods, the planner sets up a history-dependent tax system that is a function of observable wages and observable unemployment. In the first period, the worker faces a tax function  $T_1(w_{1i})$  and in the second period he faces a tax function  $T_2(w_{1i}, w_{2i})$ , where unemployment is represented by  $w_{10} = 0$ . In this model, due to our assumption of risk-neutrality and no participation decision the model has no unemployment benefits.<sup>6</sup>

Furthermore, in addition to not being able to observe the skill type of individuals, I assume that the planner is unable to observe the matching process or the skill type of the vacancies offered by firms. Therefore the planner is unable to infer the skill level of the individuals searching in these markets. This informational constraint also puts a limit on the tools of the planner to promote employment in specific skill sectors. The timing of the two-period model is the following:

### **T=1**

1. The planner credibly commits to tax functions  $T_1(\cdot)$  that depends on wage  $w_{1i}$  and  $T_2(\cdot, \cdot)$  that depends on both the first and second period wage .
2. For each skill level  $i$ , firms open up vacancies at costs  $\kappa_i$ . Each individual costlessly searches in their type-specific markets.
3. For each labor market, the matching process determines the number of filled jobs and wage level.

---

<sup>5</sup>Note that in this model, the wage an individual receives for his labor is also equal to his income.

<sup>6</sup>Without a participation decision and unemployment benefits, the planner, in this model, would favor allocations that reduces after-tax income to almost zero and increase unemployment insurance benefits to a very high level. This would raise social welfare without making the incentive constraint bind. Adding several skill types and having a unique participation cost across skill types as the model of Hungerbühler et al. (2006) or having a distribution of participation costs per type as Lehmann et al. (2011) both solve the problem. This, however, complicates the model significantly and is left for future research.

4. Each employed individual of skill  $i$  supplies a fixed amount of labor and produces  $a_i$  units of goods. They receive wage  $w_{1i}$  and pay taxes or receive transfers  $T_1(w_{1i})$ . Skilled individuals who did not find a job, loses  $(1 - \delta)$  of productivity and become unskilled individuals. Individuals consume their after-tax income as they do not save.<sup>7</sup>
5. Each match is dissolved. Period ends.

## **T=2**

6. Skilled individuals that found a job in the first period learn their cost  $\alpha_h$  of working in the skilled sector.
7. There are no longer any matching frictions and everyone is employed. The skilled individuals with  $\alpha_h$  below a certain threshold decide to work in the skilled sector, the rest decide to work in the unskilled sector. The unskilled individuals (including the skilled individuals that were unemployed in T=1) all work in the unskilled sector.
8. Each individual employed in sector  $j$  supplies a fixed amount of labor and produce  $a_j$  units of goods. Workers in the skilled sector receive wage  $w_{2h}$  and pay taxes  $T_2(w_{1h}, w_{2h})$ , workers in the unskilled sector get  $w_{2\ell}$  and pay taxes  $T_2(\cdot, \cdot)$  depending on their labor market history. Individuals consume their after-tax income.

### **3.1.1 Individuals**

#### **3.1.1.1 First Period**

The first period of life is the early decision to look for a job corresponding to one's skill. Workers costlessly search for firms in their skill specific labor market. In this period, the worker's actions influences the wage he will receive from the firm.<sup>8</sup> The probability of finding

---

<sup>7</sup>This assumption on savings is also made in Krause (2009a) and Best and Kleven (February 2013).

<sup>8</sup>The model can accommodate different microfoundations on how wages are determined. For example, a worker looks for a good wage in the case of directed search by skill and wage, or to negotiate a wage with the firm in random matching setting. More information is provided in subsection 3.1.3.

a job (or the ‘labor demand’) for type  $i$  will be  $L_i$ . If they match, they will receive after-tax income  $w_{1i} - T_1(w_{1i})$ .<sup>9</sup> For notational convenience let expected after-tax income in the first period for type  $i$  to be  $u_{1i} = L_i[w_{1i} - T_1(w_{1i})]$ . In addition to the after-tax income, the employed skilled workers each keep their human capital which guarantees them expected utility  $v_{hh}$  in the second period.

If a skilled individual does not find a job, he loses a fraction of his human capital. This loss entails that he will have utility  $u_{0\ell}$  in the second period which represents the value of working in the unskilled sector knowing that the individual was unemployed in the first period. In the case of the unskilled worker, whether he is employed or not, he does not lose any human capital. However, because the planner is able to observe both wages and employment status in the first period, he can tax differently the unskilled type in the second period depending on his employment history. Thus, an unskilled worker that was employed in the first period will have  $u_{\ell\ell}$  utility in the second period and the ones that were unemployed will have utility  $u_{0\ell}$  in the second period. From this, the expected utilities in the first period for each type are:

$$\begin{aligned}
v_h &= L_h[w_{1h} - T_1(w_{1h}) + \beta v_{hh}] + (1 - L_h)\beta u_{0\ell}, \\
&= L_h[w_{1h} - T_1(w_{1h}) + \beta \Psi_h] + \beta u_{0\ell} \\
&= u_{1h} + L_h\beta \Psi_h + \beta u_{0\ell},
\end{aligned}$$

and

$$\begin{aligned}
v_l &= L_l[w_{1\ell} - T_1(w_{1\ell}) + \beta u_{\ell\ell}] + (1 - L_l)\beta u_{0\ell}, \\
&= L_l[w_{1\ell} - T_1(w_{1\ell}) + \beta \Psi_l] + \beta u_{0\ell} \\
&= u_{1\ell} + L_l\beta \Psi_l + \beta u_{0\ell},
\end{aligned} \tag{3.1}$$

where  $\Psi_h \equiv v_{hh} - u_{0\ell}$ , and  $\Psi_l \equiv u_{\ell\ell} - u_{0\ell}$ ,  $\beta$  is the worker’s discount rate.  $\Psi_i$  is the net

---

<sup>9</sup>Note that in this model after-tax income in a given period is also the consumption of the individual.

future gain of matching in the first period for skill  $i$ . In the case of the skilled worker, this captures the net gain of keeping his human capital.

### 3.1.1.2 Second Period

In the second period, the worker's choice is different. He can now decide to choose another 'occupation' than the one corresponding to his skill level. In the two skill level context of the model only the skilled workers that found a job in the first period can decide to work in the other sector.

Suppose that for a skilled individual, the cost of working in the skilled sector is  $\alpha_h$ . The cost is assigned randomly to them at the start of the second period according to the cumulative distribution function  $F(\alpha_h)$  with  $\alpha_h \in [0, \bar{\alpha}]$ . Interpretation of this cost is the knowledge gained through working in the first period about how difficult or enjoyable the skilled occupation is. Therefore, when making their decision in the first period, individuals are unaware of this future value, only its distribution. In addition, there is no cost to working in the unskilled sector, regardless of skill or employment history.<sup>10</sup> As mentioned above, unskilled workers, i.e. unskilled workers in the first period and the skilled workers that experienced unemployment in the first period, have no choice but to be in the unskilled sector since they do not have the human capital to work in the skilled sector.

Because there is no uncertainty with regards to employment in the second period the utility a skilled worker gets from participating in the skilled sector is equal to his consumption utility (or after-tax income)  $u_{hh} = w_{2h} - T_2(w_{1h}, w_{2h})$  minus the cost of working  $\alpha_h$ . If he chooses to participate in the unskilled market, he gets utility  $u_{hl} = w_{2l} - T_2(w_{1h}, w_{2l})$ . Therefore the skilled workers that will choose to work in the skilled sector are those who

---

<sup>10</sup>Adding a cost to working in the unskilled sector changes nothing to the analysis. Instead of the distribution of  $\alpha_h$  that would be important, it would be the distribution of the difference between the two costs.

have a low enough  $\alpha_h$  such that:

$$u_{hh} - \alpha_h \geq u_{hl},$$

$$\Delta u_{hh} \equiv u_{hh} - u_{hl} \geq \alpha_h.$$

This implies that the proportion of skilled workers in the second period that will decide to work in the skilled sector is  $F(\Delta u_{hh})$ .

Taking this decision into account, the expected value of being the skilled individual is:

$$v_{hh} = \int_0^\infty \max \{u_{hh} - \alpha_h; u_{hl}\} f_h(\alpha_h) d\alpha_h,$$

$$= u_{hl} + \int_0^{\Delta u_{hh}} [\Delta u_{hh} - \alpha_h] f_h(\alpha_h) d\alpha_h,$$

From this definition and recalling that  $\Psi_h = v_{hh} - u_{0\ell}$ , the expression for the net gain of keeping the skilled human capital is:

$$\Psi_h = \Delta u_{h\ell} + \int_0^{\Delta u_{hh}} [\Delta u_{hh} - \alpha_h] f_h(\alpha_h) d\alpha_h, \quad (3.2)$$

where  $\Delta u_{h\ell} = u_{hl} - u_{0\ell}$  is the difference between the utility that a skilled individual gets by working in the unskilled sector and the utility that a previously unemployed worker gets by working in the unskilled sector which is  $u_{0\ell} = w_{2\ell} - T_2(0, w_{2\ell})$ . For the unskilled worker that found a job in the previous period, the utility (after-tax income) the he will receive in the second period is  $u_{\ell\ell} = w_{2\ell} - T_2(w_{1\ell}, w_{2\ell})$ . I define  $\Psi_l \equiv \Delta u_{\ell\ell} \equiv u_{\ell\ell} - u_{0\ell}$ .

Because of the history-dependent tax function and the ability of the skilled to work in the unskilled sector, the individuals in the second period can be divided in four distinct groups. The first group is the skilled individuals working in the skilled sector, the second is the skilled working in the unskilled sector, the third is the unskilled that has worked in the unskilled sector in both periods and the last group is the one where the individuals have

experienced unemployment in the first period and are now working in the unskilled sector. The measure of individuals in each group are respectively

$$\begin{aligned}\phi_{hh} &= \pi_h \cdot L_{1h} \cdot F(\Delta u_{hh}), \\ \phi_{hl} &= \pi_h \cdot L_{1h} \cdot [1 - F(\Delta u_{hh})], \\ \phi_{ll} &= \pi_\ell \cdot L_{1l}, \\ \phi_{0l} &= \pi_\ell \cdot [1 - L_{1l}] + \pi_h \cdot [1 - L_{1h}],\end{aligned}$$

where  $\phi_{hh} + \phi_{hl} + \phi_{ll} + \phi_{0l} = 1$ .

### 3.1.2 Firms

Firms are risk-neutral and are able to enter freely into labor markets. In the first period, the matching technology is the same in both labor markets, and therefore the analysis of a particular firm in a particular labor market can be applied to all firms in the first period. Define tightness  $\theta_i$  as the ratio  $V_i/U_i$ . The probability of filling a type- $i$  vacancy (resp. the probability of finding a job) is  $m(\theta_i) = M(U_i, V_i)/V_i = M(1/\theta_i, 1)$  (resp.  $L_i = \theta_i m(\theta_i) = M(U_i, V_i)/U_i = M(1, \theta_i)$ ). Thus, the probability of filling a vacancy decreases with an increase in vacancies ( $V_i$ ) and increases in the number of job seekers ( $U_i$ ), while the probability of finding a job increases with the number vacancies and decreases with the number of job seekers.

In the first period a firm opens up a vacancy in market  $i$  at the search cost  $\kappa_i$ . Each filled job of type  $i$  will produce an amount  $a_i$ . Therefore the firm's expected profits are  $m(\theta_i)[a_i - w_i] - \kappa_i$ . Since firms can enter freely, vacancies will be created when profits are positive. Because an increase in vacancies decreases the probability of filling a job, firms will open up vacancies until  $m(\theta_i)[a_i - w_i] = \kappa_i$ . This is the 'free-entry condition' which pins down the value of market tightness, i.e  $\theta_i = m^{-1}(\kappa_i/(a_i - w_i))$ , and leads to the following

the labor demand:

$$L(a_i, w_i) = m^{-1} \left( \frac{\kappa_i}{a_i - w_i} \right) \times \frac{\kappa_i}{a_i - w_i}, \quad (3.3)$$

which is decreasing in the level of wage  $w_i$  and increasing in the productivity level  $a_i$ . The constant-return-to-scale property implies that only the productivity level and wage matter and not the number of individuals looking for a job. For ease of exposition, from now on I will write  $L_i(w_j)$  the labor demand for the market of skill type  $i$ , i.e. where the value of production is  $a_i$  and cost of vacancy  $\kappa_i$ , for a wage of skill type  $j$ . For example, if the pre-tax wage prevailing in the unskilled market would be  $w_{1h}$ , then I will write the labor demand in the following way  $L_\ell(w_{1h}) \equiv L(a_\ell, w_{1h})$ .

In the second period, because there are no longer any market frictions, the number of matches in each skill level is determined by the short side of the market, i.e.  $M(U_i, V_i) = \min(U_i, V_i)$ . Since, there is free entry and a fixed number of individuals in each skill level, this implies that all individuals find a job. As there are no matching frictions, firms no longer need to pay search cost  $\kappa_i$  to find workers. Combining this with free-entry of firms, wages in each market are competed up until  $w_{2i} = a_i$ .

### 3.1.3 Wage Setting in First Period

Due to search frictions, a match between a firm and a worker creates a surplus. The search and matching literature has highlighted many mechanisms by which this surplus is shared. In this essay I consider sharing rules, i.e. wage-setting mechanisms, that leave only a redistributive role of taxation, i.e. the no-tax economy is efficient. To get the property that the no-tax economy is efficient, the wage-setting mechanism must be such that it maximizes the expected utility of workers in the no-tax economy. Under the assumptions of risk-neutrality and free-entry condition, it can be shown that the wage chosen by these mechanisms will

also maximize the output value in both periods net of vacancy costs in the no-tax economy.<sup>11</sup> The standard approach to gain this property is to assume the wage-setting mechanism to be the Nash-bargaining solution where the worker's bargaining power is equal to the elasticity of the matching function with respect to unemployment, but there are other justifications for the functional specification of the 'wage-setting objective' that gives an efficient no-tax economy.<sup>12</sup> Directed search by wage and skill as in the Competitive Search Equilibrium of Moen (1997) or a utilitarian monopoly union set up as in Mortensen and Pissarides (1999) can also lead to the wage-setting objective used in this essay.

An important feature of the wage-setting objective in this model is that it is increasing in after-tax surplus  $x_i = w_{1i} - T_1(w_{1i}) + \beta\Psi_i(w_{1i})$  and decreasing in pre-tax wage  $w_{ti}$ . For both types in period 1, the 'wage-setting objective' is:

$$\mathcal{WS}(x, w, a) \equiv x \cdot L(a, w), \quad (3.4)$$

which is the expected surplus of matching with a firm in period 1 for an individual of skill  $i$ . Thus the wage chosen in each skill sector will be

$$w_{1i} = \arg \max_w [w - T_1(w) + \beta\Psi_i(w)] \cdot L_i(w). \quad (3.5)$$

---

<sup>11</sup>To see this, first notice that in a no-tax economy,  $\Psi_h = \int_0^{a_h - a_\ell} [(a_h - a_\ell) - \alpha_h] f_h(\alpha_h) d\alpha_h$  and  $\Psi_l = 0$ , since  $u_{hh} = w_{2h} = a_h$  and  $u_{hl} = u_{\ell\ell} = u_{0\ell} = w_{2\ell} = a_\ell$ . By the free-entry condition it follows that  $m(\theta_i)a_i - \kappa_i = m(\theta_i)w_i$ , and by multiplying both sides by  $\theta_i$  I obtain  $L_i(w_i)a_i - \theta_i\kappa_i = L_i(w_i)w_i$ . For the unskilled workers I can stop here, because the unskilled worker will produce  $a_\ell$  and get paid  $a_\ell$  regardless if he finds a job or not. Hence maximizing output net of vacancy cost in the unskilled market is equivalent of maximizing the first period expected income of the unskilled worker when there are no taxes. This is slightly different for the skilled individuals. In the first period, the discounted expected value of production in the second period is  $\beta(L_i(w_i) \int_0^\infty \max\{a_h - \alpha_h; a_\ell\} f_h(\alpha_h) d\alpha_h + [1 - L_i(w_i)]a_\ell)$ , which is equivalent to  $\beta[L_i(w_i)\Psi_h + a_\ell]$  in the no tax economy. Therefore, maximizing the output value net of vacancy cost and work effort in the skilled market, i.e.  $L_i(w_i)[a_i + \beta\Psi_h] + \beta a_\ell - \theta_i\kappa_i$ , is equivalent to maximizing the expected lifetime utility of the individual, i.e.  $L_i(w_i)[w_i + \beta\Psi_h] + \beta a_\ell$ , when the individuals do not have to pay taxes. This is equivalent to maximizing expected surplus from matching with a firm in the no-tax economy, i.e.  $L_i(w_i)[w_i + \beta\Psi_h]$ .

<sup>12</sup>See Hosios (1990). To see more clearly how the wage-setting objective relates to the Nash-bargaining solution see Appendix A.

Because of the history-dependent taxes featured in the model  $\Psi_i$  is written as a function of  $w_{1i}$ . Choosing  $w_{1i}$  influences the payoff in the future and thus the surplus from matching. Due to the two-type model set up this function is discontinuous. To obtain intuition for the problem, I concentrate on the continuous portion of  $\Psi_i$  in first period wage  $w$  in a similar fashion as Stiglitz (1982) with the income tax function. To make (3.5) well-behaved and give the property that  $w_{1i}$  is increasing in skill  $i$  in the no-tax economy. I suppose that the wage elasticity  $\frac{\partial L(a,w)}{\partial w} \frac{w}{L(a,w)}$  is decreasing in  $w_i$  and increasing in  $a_i$  as in Lehmann et al. (2011).<sup>13</sup> The first-order condition of (3.5) can be written in this following way:

$$-\frac{\partial L_i(w_{1i})}{\partial w_{1i}} \frac{w_{1i}}{L_i(w_{1i})} = \frac{1 - T_1'(w_{1i}) + \beta \Psi_i'(w_{1i})}{1 - \left[ \frac{T_1(w_{1i}) - \beta \Psi_i}{w_{1i}} \right]}. \quad (3.6)$$

The left-hand side is the negative of the the wage elasticity of the employment probability, it measures the reduction in the probability of employment as wage increases. The right-hand side is the wage elasticity of the surplus going to the worker once he is matched. The chosen optimal wage has these two elasticities at equal value. And since labor demand is a decreasing function of wage and that the surplus of matching is positive, the FOC implies that  $T_1'(w_{1i}) - \beta \Psi_i'(w_{1i}) \leq 1$ . Similar to the New Dynamic Finance literature that features history-dependent tax systems,  $T_1'(w_{1i}) - \beta \Psi_i'(w_{1i})$  is the wedge that will cause a distortion in the labor market and affects the decision of the worker.

Equation (3.6) highlights the impact of the tax function  $T_1(w)$  on the equilibrium wage of a given skill and at the same time the effect it has on employment for the labor market. Suppose that  $T_1'(w)$ ,  $\frac{T_1(w)}{w}$ ,  $\frac{\beta \Psi_i}{w}$  and  $\Psi_i'(w_{1i})$  are parameters. Keeping other parameters constant, an increase in  $T_1'(w)$  (or a decrease in  $\Psi_i'(w_{1i})$ ) will lower the value of the wage elasticity of the surplus, this implies that wage elasticity of employment must also follow.

---

<sup>13</sup>The last two assumptions are not limiting, as Lehmann et al. (2011) argue, the assumptions are easily satisfied with a CES matching functions under a broad range of the parameter value determining the the elasticity of substitution and also is  $\frac{\partial \kappa(a)}{\partial a} \frac{a}{\kappa} \leq 1$ .

From the assumptions made on the matching function, this translates to a reduction in wage and thus an increase in employment for that skill level. The reverse happens if  $\frac{T_1(w)}{w}$  is increased. This will increase the value of the wage elasticity of the surplus and thus will lead to an increase in the wage and therefore a reduction in employment. An increase in  $\frac{\beta\Psi_i}{w}$  has the same effect as an increase in the marginal tax rate; it lowers the chosen wage. All three of these effects come from the surplus sharing rule. Increasing the marginal tax rate makes it harder to transfer utility from the firm to the worker, because some of it is taxed away. Increasing  $\Psi_i$  makes the worker much more willing to find and accept a lower wage because he gets this value only if he is employed. Increasing the average tax rate  $\frac{T_1(w)}{w}$  decreases the surplus of the worker of finding a job. It lowers the value of finding and keeping a job and thus makes them unwilling to accept a lower wage.

### 3.1.4 Planner and Perfect Information in First Period

#### 3.1.4.1 Planner

The planner's preference is over lifetime expected utilities of individuals and can be expressed by the following Bergson-Samuelson social welfare function:

$$SWF = \pi_\ell W\left(u_{1\ell} + \beta L_\ell(w_{1\ell})\Psi_\ell + \beta u_{0\ell}\right) + \pi_h W\left(u_{1h} + \beta L_h(w_{1h})\Psi_h + \beta u_{0h}\right), \quad (3.7)$$

where  $W'(\cdot) > 0$  and  $W''(\cdot) \leq 0$ . This social welfare function is quite general and can represent different aversions to inequality from the utilitarian criterion ( $W''(\cdot) = 0$ ) to the Maximin (or Rawlsian) criterion that exhibits extreme aversion to inequality of expected utilities.

In the 'occupational choice' model of the labor market of the second period, for all informational cases considered throughout the essay, i.e. the planner being or not being able to observe the type of worker in the first period, the planner in the second period is able to

observe the productivity of all jobs but he is unable to observe the cost of working  $\alpha_h$ . These informational assumptions are made to clarify the analysis and to focus on the workings of the first period labor market and tax system. The framework of the occupational model of Saez (2002b, 2004) makes it possible to investigate the impact of changing the number of skilled individuals on the optimal allocation in the second period, which permits one to identify characteristics of the tax system in the first period.<sup>14</sup>

Finally, the planner must balance his budget every period. He is unable to save or borrow to help smooth consumption between periods.<sup>15</sup>

Since the planner is unable to transfer resources between periods, he faces two budget constraints, one in the first period and another in the second. The problem is written as an allocation problem, therefore the first period budget constraint of the planner can be written in the following way:

$$\pi_\ell \cdot [L_\ell(w_{1\ell}) \cdot w_{1\ell} - u_{1\ell}] + \pi_h \cdot [L_h(w_{1h}) \cdot w_{1h} - u_{1h}] = 0, \quad (3.8)$$

recalling that  $u_{1\ell} = L_\ell(w_{1\ell}) \cdot [w_{1\ell} - T_1(w_{1\ell})]$  and  $u_{1h} = L_h(w_{1h}) \cdot [w_{1h} - T_1(w_{1h})]$ .

Also, the second period budget constraint becomes:

$$\begin{aligned} \phi_{0\ell}(w_{1\ell}, w_{2\ell})[w_{2\ell} - u_{0\ell}] + \phi_u(w_{1\ell})[w_{2\ell} - u_{\ell\ell}] \\ + \phi_{h\ell}(w_{1h}, \Delta u_{hh})[w_{2\ell} - u_{h\ell}] + \phi_{hh}(w_{1h}, \Delta u_{hh})[w_{2h} - u_{hh}] = 0, \end{aligned}$$

---

<sup>14</sup>It also prevents from having the trivial answer that all redistribution should take place in the second period.

<sup>15</sup>This assumption is made due to the risk-neutrality of workers that guarantees the *laissez-faire* outcome to be efficient. In the case of risk-averse preferences the *laissez-faire* outcome would not be efficient. See the Appendix of Hungerbühler et al. (2008).

which can be rewritten in the following way:

$$\phi_{0l}w_{2l} + \phi_{ll}[w_{2l} - \Delta u_{ll}] + \phi_{hl}[w_{2l} - \Delta u_{hl}] + \phi_{hh}[w_{2h} - \Delta u_{hl} - \Delta u_{hh}] = u_{0l}, \quad (3.9)$$

where  $u_{ll} = \Delta u_{ll} + u_{0l}$ ,  $u_{hl} = \Delta u_{hl} + u_{0l}$  and  $u_{hh} = \Delta u_{hh} + u_{hl}$ .

### 3.1.4.2 Perfect Information in First Period

I investigate in this subsection the characteristics of the perfect information allocation, i.e. the planner is able to observe the productivity of a worker-firm pair in the first period. He maximizes his social preferences by picking the first period wage for both skill levels  $w_{1i}$ , the first period expected income  $u_{1i}$ , the utility level  $u_{0l}$  and the utility differences  $\Delta u_{ll}$ ,  $\Delta u_{hl}$  and  $\Delta u_{hh}$ . The problem of the planner is:

$$\begin{aligned} \max_{\substack{\{w_{1i}, u_{1i}\}_{\forall i} \\ u_{0l}, \Delta u_{ll} \\ \Delta u_{hl}, \Delta u_{hh}}} & \pi_l W\left(u_{1l} + \beta L_\ell(w_{1l}) \Psi_l(\Delta u_{ll}) + \beta u_{0l}\right) + \pi_h W\left(u_{1h} + \beta L_h(w_{1h}) \Psi_h(\Delta u_{hl}, \Delta u_{hh}) + \beta u_{0l}\right) \\ \text{s.t. } & (\lambda_1) \quad \pi_l \cdot [L_\ell(w_{1l}) \cdot w_{1l} - u_{1l}] + \pi_h \cdot [L_h(w_{1h}) \cdot w_{1h} - u_{1h}] = 0, \\ & (\lambda_2) \quad \phi_{0l}(w_{1l})w_{2l} + \phi_{ll}(w_{1l})[w_{2l} - \Delta u_{ll}] \\ & \quad + \phi_{hl}(w_{1h}, \Delta u_{hh})[w_{2l} - \Delta u_{hl}] + \phi_{hh}(w_{1h}, \Delta u_{hh})[w_{2h} - \Delta u_{hl} - \Delta u_{hh}] = u_{0l}. \end{aligned} \quad (3.10)$$

Each solution must satisfy the following necessary conditions:

$$W'(u_{1\ell} + \beta L_\ell(w_{1\ell})\Psi_\ell(\Delta u_{\ell\ell}) + \beta u_{0\ell}) = W'(u_{1h} + \beta L_h(w_{1h})\Psi_h(\Delta u_{h\ell}, \Delta u_{hh}) + \beta u_{0\ell}), \quad (3.11)$$

$$\Delta u_{hh} = w_{2h} - w_{2\ell}, \quad (3.12)$$

$$-\frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} \frac{w_{1\ell}}{L_\ell(w_{1\ell})} = 1, \quad (3.13)$$

$$-\frac{\partial L_h(w_{1h})}{\partial w_{1h}} \frac{w_{1h}}{L_h(w_{1h})} = \frac{1}{1 + \frac{\beta \Psi_h^{LF}}{w_{1h}}}, \quad (3.14)$$

where

$$\Psi_h^{LF} = \int_0^{w_{2h} - w_{2\ell}} [(w_{2h} - w_{2\ell}) - \alpha_h] f_h(\alpha_h) d\alpha_h = \int_0^{a_h - a_\ell} [(a_h - a_\ell) - \alpha_h] f_h(\alpha_h) d\alpha_h,^{16}$$

since in the second period  $w_{2h} = a_h$  and  $w_{2\ell} = a_\ell$ . Condition (3.11) says that expected utility of each type must be equalized.<sup>17</sup> Condition (3.12) indicates that the tax paid by the skilled types irrespective of their occupational choice must be the same, i.e  $T_2(w_{1h}, w_{2h}) = T_2(w_{1h}, w_{2\ell}) = T_2(w_{1h})$  which also leads to  $\Psi_h = \Delta u_{h\ell} + \Psi_h^{LF}$ .

The last two conditions are the ones related to the optimal first period wages. The first thing to notice is that both (3.13) and (3.14) are at their *laissez-faire* levels. To see this, we just have to look at (3.6), remove taxes and remember that  $\Psi_\ell^{LF} = 0$  and  $\Psi_h^{LF} = 0$ . The assumptions made on the wage-setting mechanism are such that the *laissez-faire* allocation would be efficient, and since the planner has perfect information in the first period he can redistribute and set the tax system in such a way that it is not distortionary in that period.

To decentralize the optimal wage, the first-order condition of the wage-setting problem

---

<sup>16</sup>The FOCs have been partly solved to remove multipliers.

<sup>17</sup>This result also could have been derived using a Maximin social welfare function for the preferences of the planner.

must be used much like the first-order condition to the individual's problem is used in a standard optimal income tax exercise. Combining (3.13) and (3.6) to get:

$$T_1'(w_{1\ell}) - \beta\Psi_l'(w_{1\ell}) = \frac{T_1(w_{1\ell})}{w_{1\ell}} - \beta\frac{\Psi_l}{w_{1\ell}}, \quad (3.15)$$

where  $\Psi_l = \Delta u_{\ell\ell} = T_2(0, w_{2\ell}) - T_2(w_{1\ell}, w_{2\ell})$ . Decentralizing the optimal skilled worker's wage, uses (3.14), (3.6) and the fact that in this informational case  $\Psi_h = \Delta u_{h\ell} + \Psi_h^{LF}$ :

$$T_1'(w_{1h}) - \beta\Psi_h'(w_{1h}) = \frac{T_1(w_{1h})}{w_{1h} + \beta\Psi_h^{LF}} - \frac{\beta\Delta u_{h\ell}}{w_{1h} + \beta\Psi_h^{LF}}, \quad (3.16)$$

where  $\Delta u_{h\ell} = T_2(0, w_{2\ell}) - T_2(w_{1h}, w_{2\ell}) = T_2(0, w_{2\ell}) - T_2(w_{1h}, w_{2h})$ .

Conditions (3.13) and (3.14) can also be used to determine the conditions for chosen wages to be below or above their efficient level. For the unskilled worker's wage, (3.13) gives following conditions:

$$\begin{aligned} -\frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} \frac{w_{1\ell}}{L_\ell(w_{1\ell})} < 1 &\implies w_{1\ell}^* < \bar{w}_{1\ell}, \\ -\frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} \frac{w_{1\ell}}{L_\ell(w_{1\ell})} > 1 &\implies w_{1\ell}^* > \bar{w}_{1\ell}, \end{aligned}$$

where  $\bar{w}_{1\ell}$  is the efficient level wage and  $w_{1\ell}^*$  is the chosen wage. If the wage elasticity of employment is lower (resp. higher) than one, this implies that the wage is below (resp. above) the efficient level. It is easier to rewrite condition (3.14) to do the same exercise with the skilled worker's wage:

$$\begin{aligned} -\frac{\partial L_h(w_{1h})}{\partial w_{1h}} \frac{[w_{1h} + \beta\Psi_h^{LF}]}{L_h(w_{1h})} < 1 &\implies w_{1h}^* < \bar{w}_{1h}, \\ -\frac{\partial L_h(w_{1h})}{\partial w_{1h}} \frac{[w_{1h} + \beta\Psi_h^{LF}]}{L_h(w_{1h})} > 1 &\implies w_{1h}^* > \bar{w}_{1h}, \end{aligned}$$

where again  $\bar{w}_{1h}$  is the efficient level wage and  $w_{1h}^*$  is the chosen wage.

### 3.2 Optimal Tax Policy

In this section I consider the case where the planner is unable to observe the productivity of a worker-firm pair. He is only able to observe the income (negotiated wage) of workers. This puts a constraint on the ability of the planner to redistribute income across types of individuals, since taxes can no longer be a function of ability and income: they can only be a function of incomes (wage) and labor market experience. In this essay the worker-firm pair's behavior is modeled as a single agent maximizing the wage-setting objective (3.5) which acts as the preferences over different bundles.<sup>18</sup> Because of the history-dependent tax system in place, choosing the optimal wage for a particular skill level implies the worker gets surplus  $x_{1i} = w_{1i} - T_1(w_{1i}) + \beta\Psi_i(w_{1i})$  from a match with a firm. From this he has expected surplus  $E_{1i} = x_{1i} \cdot L(a_i, w_{1i})$ . Again, notice that the wage-setting objective is increasing in the surplus  $x_{1i}$  and decreasing in the wage  $w_{1i}$ . Thus a tax system  $\{T_1(\cdot), T_2(\cdot, \cdot)\}$  will induce a set of allocations  $\{w_{1i}, x_{1i}, E_{1i}\}_{i=l,h}$ , and from the *Taxation principle* (Rochet, 1985; Guesnerie, 1995) we know that the set of allocations will correspond to the set of allocations that satisfy the following incentive constraints:

$$x_{1l} \cdot L_\ell(w_{1\ell}) \geq \widehat{x}_{1h} \cdot L_\ell(w_{1h}), \quad (3.17)$$

$$x_{1h} \cdot L_h(w_{1h}) \geq x_{1l} \cdot L_h(w_{1\ell}). \quad (3.18)$$

where  $\widehat{x}_{1h} = w_{1h} - T_1(w_{1h}) + \widehat{\Psi}_h$  with  $\widehat{\Psi}_h$  being the value to a mimicking unskilled type of the difference in utility levels. Since, he is unable to work in the skilled market, similar to a skilled worker with a very high  $\alpha_h$ , he will have to take bundle  $u_{hl}$  if he mimics. There is no need for such a modification for the mimicking skilled worker as he would be required to work in the unskilled sector in the second period when he mimics.

The notion of a worker-firm pair being a single agent with the wage-setting objective as

---

<sup>18</sup>I assume that side-payments are not allowed.

preferences has several microfoundations. For the wage-setting objective used in this essay, and one that has a very natural interpretation for mimicking, is the Competitive Search Equilibrium when search is directed by skill and wage. In this context there is one potential skill market for each skill level and each wage level. So when it is time to either look for jobs and open up vacancies workers and firms decide in which submarket, and only one, to participate. As in Moen (1997), this is a non-cooperative game between firms and workers and using the Nash equilibrium concept gives us the Competitive Search Equilibrium. The behavior of the workers and firms end up maximizing the wage-setting objective (3.4).<sup>19</sup>

In the asymmetrical information scenario investigated in this section the wage-setting objective defined above has an important property. Recall that the wage-setting objective is:

$$\mathcal{WS}(x, w, a) \equiv x \cdot L(a, w),$$

so the marginal rate of substitution between the matching surplus  $x$  and wage  $w$  is:

$$\left. \frac{dx}{dw} \right|_{\mathcal{WS}(a, \cdot)} = - \frac{x}{L(a, w)} \frac{\partial L(a, w)}{\partial w}.$$

From the assumptions made in the above section on  $L(a, w)$ , note that the marginal rate of substitution is decreasing in  $a$ , concluding from this that the marginal rate of substitution for the skilled worker-firm pair evaluated at a specific bundle will be smaller than the marginal rate of substitution of an unskilled worker-firm pair. Therefore in this model there is a Spence-Mirrlees single-crossing property.<sup>20</sup>

For the rest of the essay, I will concentrate on what Stiglitz (1982) calls the ‘normal case’ where incentive constraint (3.18) binds at the optimum but not (3.17). This reflects the idea that the planner wishes to redistribute from the skilled types who have a greater

---

<sup>19</sup>See Appendix A.2.1 for a more details.

<sup>20</sup>The single-crossing property can fail for the cases of less sophisticated tax system, i.e. age-dependent or history(age)-independent tax systems.

expected lifetime utility to the unskilled types which induces the skilled types to ‘lie’ about their actual type.

### 3.2.1 Maximin Social Welfare Function

I start by investigating the Maximin social welfare function case as it offers sharp analytical results. The characteristic of the Maximin social preference is that it features extreme inequality aversion and the planner only cares about the least-well off. The planner wishes to maximize the lifetime expected utility of the unskilled workers. Using the same definitions as in section 3.1.4, the incentive constraint (3.18) can be written in the following way:

$$u_{1h} + L_h(w_{1h})\beta\Psi_h(\Delta u_{h\ell}, \Delta u_{hh}) \geq \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})}u_{1\ell} + \beta L_h(w_{1\ell})\Delta u_{\ell\ell}.$$

Under strong redistributive preferences it is possible that under specific values of productivity and population size of each productivity type that the planner would wish to shutdown the labor market of the unskilled in the first period. He would do so to obtain more from the skilled workers and redistribute it to the unskilled. This particular outcome is not investigated in this essay. I make the following assumptions:

**Assumption 3.** *The values of the parameters  $\pi_\ell, \pi_h, a_\ell, a_h$  and distribution  $F(\cdot)$  are such that the optimal allocation has both skill types working in the first period and that*

$$L_\ell(\tilde{w}) < L_h(\tilde{w}) \quad \forall \tilde{w} \in (0, a_\ell).$$

The planner's problem is then:

$$\begin{aligned}
& \max_{\substack{\{w_{1i}, u_{1i}\}_{\forall i} \\ u_{0\ell}, \Delta u_{\ell\ell}, \Delta u_{h\ell}, \Delta u_{hh}}} & u_{1\ell} + \beta L_\ell(w_{1\ell}) \Delta u_{\ell\ell} + \beta u_{0\ell} \\
& s.t. \quad (\lambda_1) \quad \pi_\ell \cdot [L_\ell(w_{1\ell}) \cdot w_{1\ell} - u_{1\ell}] + \pi_h \cdot [L_h(w_{1h}) \cdot w_{1h} - u_{1h}] = 0, \\
& \quad (\lambda_2) \quad \phi_{0l}(w_{1\ell}, w_{1h}) w_{2\ell} + \phi_u(w_{1\ell}) [w_{2\ell} - \Delta u_{\ell\ell}] \\
& \quad \quad + \phi_{hl}(w_{1h}, \Delta u_{hh}) [w_{2\ell} - \Delta u_{h\ell}] + \phi_{hh}(w_{1h}, \Delta u_{hh}) [w_{2h} - \Delta u_{h\ell} - \Delta u_{hh}] = u_{0\ell} \\
& \quad (\mu) \quad u_{1h} + L_h(w_{1h}) \beta \Psi_h(\Delta u_{h\ell}, \Delta u_{hh}) \geq \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} u_{1\ell} + \beta L_h(w_{1\ell}) \Delta u_{\ell\ell}, \\
& \quad u_{1\ell} \geq 0, u_{1h} \geq 0, w_{1\ell} \geq 0, w_{1h} \geq 0, u_{0\ell} \geq 0, \Delta u_{\ell\ell} \leq 0, \Delta u_{h\ell} \leq 0. \quad (3.19)
\end{aligned}$$

This nonlinear programming problem can have different solutions depending on parameter values. First it can be show that for any solution where  $u_{1h} > 0$ , it is impossible for this solution to also have  $\Delta u_{\ell\ell} \leq 0$  or  $\Delta u_{h\ell} \leq 0$  or both of them together.<sup>21</sup> A solution with  $u_{1h} = 0$  is possible only if the incentive constraint is not binding. In this model  $u_{1h} = 0$  either means the income of the skilled individual is entirely taxed or  $w_{1h} = a_h$  and there are no skilled sector jobs. Assumption 1 puts aside the later case. The former case means that the skilled worker can be entirely taxed and the difference in expected utility he gets from matching is enough to make the worker-firm pair unwilling to take the unskilled worker-firm's bundle. For this to be true,  $\Delta u_{\ell\ell}$  needs to be very negative,  $\Delta u_{h\ell}$  to be very positive and  $\Delta u_{hh} = w_{2h} - w_{2\ell}$ . This means that there is now severe taxation on the group of unskilled individuals who have worked in both periods and it is given to both skilled groups working in the second period. If there is a big enough difference in  $w_{2h} - w_{2\ell}$  and or a high enough number of skilled individuals, this solution is not possible and the more standard solution with a binding incentive constraint is the only case possible. From this point on, I will

---

<sup>21</sup>Actually, this extends to solutions where  $\Delta u_{\ell\ell} \neq 0$  or  $\Delta u_{h\ell} \neq 0$ .

only consider the solution under which (3.18) is binding. This implies that the tax function, although history-dependent, has no insurance component.

To determine characteristics of the optimal first period wages in the imperfect information case we need to look at the following necessary conditions for  $\Delta u_{hh}$ ,  $w_{1\ell}$  and  $w_{1h}$  respectively:

$$w_{2h} - w_{2\ell} - \Delta u_{hh}^* = (1 - \lambda_1) \frac{F(\Delta u_{hh}^*)}{f(\Delta u_{hh}^*)}, \quad (3.20)$$

$$-\frac{\partial L_\ell(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{w_{1\ell}^*}{L_\ell(w_{1\ell}^*)} = 1 - \frac{\mu u_{1\ell}}{\lambda_1 \pi_\ell} \frac{L_h(w_{1\ell}^*)}{[L_\ell(w_{1\ell}^*)]^2} \left[ \frac{\partial L_h(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_h(w_{1\ell}^*)} - \frac{\partial L_\ell(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_\ell(w_{1\ell}^*)} \right], \quad (3.21)$$

$$-\frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \frac{w_{1h}^*}{L_h(w_{1h}^*)} = 1 + \frac{\beta}{L_h(w_{1h}^*)} \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \left\{ \Psi_h^* + \frac{F(\Delta u_{hh}^*) [w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1} \right\}. \quad (3.22)$$

From the FOC of  $u_{1\ell}$  and  $u_{1h}$ , we can get the following expression for the lagrange multiplier of the first period budget constraint:

$$0 < \lambda_1 = \frac{L_\ell(w_{1\ell}^*)}{\pi_\ell L_\ell(w_{1\ell}^*) + \pi_h L_h(w_{1\ell}^*)} < 1.$$

Condition (3.20) and (3.21) together imply:

**Proposition 3.a.** *Under imperfect information in the first period, the optimal tax system under Maximin has the following properties: there is redistribution of income in the second period from workers in the skilled sector to workers in the unskilled sector and the wage of the unskilled workers in the first period is distorted below the efficient level. Thus in the first period, marginal tax rates for the unskilled is greater than the average tax rate, i.e.*

$$T_1'(w_{1\ell}) > \frac{T_1(w_{1\ell})}{w_{1\ell}}$$

The first part of the above proposition uses condition (3.20) and the fact that  $\Delta u_{\ell\ell} = 0$

and  $\Delta u_{h\ell} = 0$ . The later implies that  $u_{il}^* = u_{hl}^* = u_{0l}^* \equiv u_{2l}$ . In terms of taxes this implies  $T_2(w_{1\ell}, w_{2\ell}) = T_2(w_{1h}, w_{2\ell}) = T_2(0, w_{2\ell}) = T_2(w_{2\ell})$ , and thus  $\Delta u_{hh} = u_{hh} - u_{2l}$ . Since the right-hand side of condition (3.20) is positive, this means that there is redistribution from those who work in the skilled sector in the second period to those who work in the unskilled sector. Furthermore, notice that  $\Delta u_{hh}^* < \Delta u_{hh}^{LF} = w_{2h} - w_{2\ell}$ .

The second part of the proposition uses condition (3.21) and the conditions derived in the perfect information case. For the unskilled worker's wage to be distorted, the right-hand side of (3.21) must be smaller than 1. This is true only if

$$\frac{\mu u_{1\ell}}{\lambda_1 \pi_\ell} \frac{L_h(w_{1\ell}^*)}{[L_\ell(w_{1\ell}^*)]^2} \left[ \frac{\partial L_h(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_h(w_{1\ell}^*)} - \frac{\partial L_\ell(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_\ell(w_{1\ell}^*)} \right] > 0.$$

Notice that the term multiplying what is inside the big square brackets is positive. From the assumptions made on  $L_i(\cdot)$ , i.e. that  $\frac{\partial L(a,w)}{\partial w} \frac{w}{L(a,w)}$  is increasing in  $a$ , it follows that the term inside the bracket is positive. The intuition behind this result is to make it less appealing for the skilled worker-firm pair to mimic the unskilled worker-firm pair. This happens because the skilled worker-firm pair using the 'wage-setting objective' as preferences would, at the margin, prefer a higher level of  $w$ .

In addition, note that redistribution to the unskilled partly goes through an increase in the probability of matching as, from (3.21), the unskilled wage is distorted  $-\frac{\partial L_\ell(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{w_{1\ell}^*}{L_\ell(w_{1\ell}^*)} < 1$ . Using this result and (3.6), the tax system that can decentralize this allocation will have the following feature:

$$\frac{1 - T_1'(w_{1\ell})}{1 - \frac{T_1(w_{1\ell})}{w_{1\ell}}} < 1 \implies \frac{T_1(w_{1\ell})}{w_{1\ell}} < T_1'(w_{1\ell}).$$

Because there is redistribution in the first period and the budget constraint must balance, we have that  $T_1(w_{1\ell}) < 0$ .

The necessary condition (3.22) determines some characteristics of the optimal wage of the skilled worker. The terms inside the brackets on the right-hand side are the undiscounted

marginal gains to society of increasing the probability of skilled individuals matching in the first period. The first term in this bracket is  $\Psi_h^*$  which is the marginal gain to a skilled worker of matching. Note that since  $\Delta u_{hh}^* < \Delta u_{hh}^{LF}$ , this implies  $\Psi_h^* < \Psi_h^{LF}$ . The second term is the social marginal welfare weighted fiscal gain in the second period from increasing the matching probability in the second period. In the *Laissez-faire* case this fiscal gain does not exist because there is no redistribution, and in the perfect information case this value is null since the optimal tax system requires  $\Delta u_{hh}^* = w_{2h} - w_{2\ell}$ . Therefore there are no fiscal gains from sending more skilled workers in the second period.

**Proposition 3.b.** *Under imperfect information in the first period, the optimal first period skilled wage will be distorted above or below the efficient wage according to the following equation:*

$$\left[ \Psi_h^* + \frac{F(\Delta u_{hh}^*)[w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1} \right] - \Psi_h^{LF} \geq 0 \implies w_{1h}^* \leq w_{1h}^{LF}.$$

The logic of Proposition 3.b is that the value of the marginal gains to society of increasing the probability of skilled individuals matching in the first period must be compared with the value of  $\Psi_h^{LF}$ . To see the effect on the wage, I subtract both sides of 3.22 by

$$\frac{\partial L_h(w_{1h})}{\partial w_{1h}} \frac{\beta \Psi_h^{LF}}{L_h(w_{1h})},$$

which gives

$$-\frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \frac{[w_{1h}^* + \beta \Psi_h^{LF}]}{L_h(w_{1h}^*)} = 1 + \frac{\beta}{L_h(w_{1h}^*)} \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \left\{ \left[ \Psi_h^* + \frac{F(\Delta u_{hh}^*)[w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1} \right] - \Psi_h^{LF} \right\}. \quad (3.23)$$

The wage of the skilled workers is distorted below or above the efficient value if the right-hand side of the above equation is below or above 1, or more precisely if the second term of the right-hand side is negative or positive. It is straightforward to see that what is multiplying the terms inside the brackets is negative and thus to know if the right-hand side of (3.23) is

below or above 1 which is stated in Proposition 3.b.

Under the present general assumptions on the problem it is not possible to have a sharper result than that of Proposition 3.b. Thus, I must make some assumptions on the distribution function  $F(\cdot)$ .

**Corollary 3.** *Under imperfect information in the first period, Maximin social welfare function and assuming that  $\alpha_h$  is distributed according to  $F(\alpha_h) = \left(\frac{\alpha_h}{\alpha}\right)^\varepsilon$  with  $\varepsilon > 0$ , the optimal tax system distorts the wages of the skilled workers below the efficient level. Thus employment of skilled individuals is higher in the imperfect information case than in the perfect information (or Laissez-faire) case.*

The proof of this Corollary is in Appendix A.2.2.

This result comes about for two reasons: the informational problem and the human capital destruction. The information problem in the first period forces the planner to rely more on the second period to redistribute income. On its own this fact would not result in a downward distortion of the skilled worker's wage because there are no gains in doing so. When there is no human capital destruction, any policy affecting the second period will not have an effect on how wages are determined. Furthermore, the policies in the first period cannot influence the distribution of workers in the second period. In addition, there are no informational gains from distorting the wage of the skilled worker as is standard in the principal-agent literature. Therefore there are no reasons to distort the skilled worker's wage when there is no human capital destruction.

When human capital destruction is introduced, redistributive policies in the second period have an impact on the first period wage and thus the employment levels in the first period. And policies in the first period have an impact on how much redistribution can happen in the second period. By distorting the skilled worker's wage downward, the planner does four things. He counteracts the upward effect on the skilled worker's wage caused by increasing redistribution in the second period. He increases the likelihood of being skilled in the second

period and thus making mimicking less appealing. He creates more employment of the skilled insuring that there will be more of them to split this new redistributive burden, which at the same time permits him to make it smaller. Finally, he also insures that there are less people to redistribute to.

Using (3.6) to decentralize the result implies:

$$\frac{1 - T_1'(w_{1h}) + \beta\Psi_h'}{1 - \left[ \frac{T_1(w_{1h}) - \beta\Psi_h}{w_{1h}} \right]} < 1 \implies T_1'(w_{1h}) - \beta\Psi_h' > \frac{T_1(w_{1h})}{w_{1h}} - \beta\frac{\Psi_h^*}{w_{1h}}.$$

From the last inequality, determining exactly if the marginal tax rate is higher or below the average tax rate is not possible. But this inequality says that the greater  $\Psi_h^*$ , the lower the marginal tax rate or the greater  $\Psi_h'^*$  can be relative to the average tax rate, on the contrary the more redistribution there is in the second period, the higher the ratio between the marginal tax rate and the average tax rate must be.

### 3.2.2 Bergson-Samuelson Social Welfare Function

In this subsection I will consider the problem with the more general social welfare function.

With the Bergson-Samuelson social-welfare function, the planner's problem is

$$\begin{aligned}
& \max_{\{w_{1i}, u_{1i}\}_{\forall i}} \quad \pi_\ell W\left(u_{1\ell} + \beta L_\ell(w_{1\ell}) \Delta u_{\ell\ell} + \beta u_{0\ell}\right) + \pi_h W\left(u_{1h} + \beta L_h(w_{1h}) \Psi_h(\Delta u_{h\ell}, \Delta u_{hh}) + \beta u_{0\ell}\right) \\
& \quad u_{0\ell}, \Delta u_{\ell\ell}, \Delta u_{h\ell}, \Delta u_{hh} \\
& \text{s.t. } (\lambda_1) \quad \pi_\ell \cdot [L_\ell(w_{1\ell}) \cdot w_{1\ell} - u_{1\ell}] + \pi_h \cdot [L_h(w_{1h}) \cdot w_{1h} - u_{1h}] = 0, \\
& \quad (\lambda_2) \quad \phi_{0\ell}(w_{1\ell}) w_{2\ell} + \phi_{u\ell}(w_{1\ell}) [w_{2\ell} - \Delta u_{\ell\ell}] \\
& \quad \quad + \phi_{h\ell}(w_{1h}, \Delta u_{hh}) [w_{2\ell} - \Delta u_{h\ell}] + \phi_{hh}(w_{1h}, \Delta u_{hh}) [w_{2h} - \Delta u_{h\ell} - \Delta u_{hh}] = u_{0\ell} \\
& \quad (\mu) \quad u_{1h} + L_h(w_{1h}) \beta \Psi_h(\Delta u_{h\ell}, \Delta u_{hh}) \geq \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} u_{1\ell} + \beta L_h(w_{1\ell}) \Delta u_{\ell\ell}, \\
& \quad u_{1\ell} \geq 0, u_{1h} \geq 0, w_{1\ell} \geq 0, w_{1h} \geq 0, u_{0\ell} \geq 0. \tag{3.24}
\end{aligned}$$

As above, I look at the case where (3.18) is binding in the optimum and also use Assumption

1. Thus cases where  $u_{1h}^* = 0, \Delta u_{\ell\ell} \leq 0, \Delta u_{h\ell} \leq 0$  is a possible solution are not investigated.

From the first-order condition of the above problem, the solution is characterized by the following necessary conditions<sup>22</sup>:

$$w_{2h} - w_{2\ell} - \Delta u_{hh}^* = \frac{\mu\Omega}{[\lambda_1 + \mu\Omega]} \frac{F(\Delta u_{hh}^*)}{f(\Delta u_{hh}^*)}, \tag{3.25}$$

$$-\frac{\partial L_\ell(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{w_{1\ell}^*}{L_\ell(w_{1\ell}^*)} = 1 - \frac{\mu u_{1\ell}}{\lambda_1 \pi_\ell} \frac{L_h(w_{1\ell}^*)}{[L_\ell(w_{1\ell}^*)]^2} \left[ \frac{\partial L_h(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_h(w_{1\ell}^*)} - \frac{\partial L_\ell(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_\ell(w_{1\ell}^*)} \right], \tag{3.26}$$

$$-\frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \frac{w_{1h}^*}{L_h(w_{1h}^*)} = 1 + \frac{\beta}{L_h(w_{1h}^*)} \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \left\{ \Psi_h^* + \frac{[\lambda_1 + \mu\Omega]}{\lambda_1} F(\Delta u_{hh}^*) [w_{2h} - w_{2\ell} - \Delta u_{hh}^*] \right\}, \tag{3.27}$$

---

<sup>22</sup>See Appendix A.2.2 for more details.

where  $\Omega = \left[ \frac{L_h(w_{1\ell}^*)}{L_\ell(w_{1\ell}^*)} - 1 \right] > 0$  since  $L_h(w_{1\ell}^*) > L_\ell(w_{1\ell}^*)$ .<sup>23</sup> Note that from the FOC of  $u_{1\ell}$  and  $u_{1h}$  one gets that  $\pi_\ell W'(v_l^*) + \pi_h W'(v_h^*) = \lambda_1 + \mu\Omega$ .<sup>24</sup>

Condition (3.25) determines that there is redistribution in the second period, i.e.  $\Delta u_{hh}^* < \Delta u_{hh}^{LF}$ . From (3.26), which is identical to the condition from the above subsection, the result that the unskilled worker's wage in the first period is distorted downward follows. Also as above, to know if the wage of the skilled workers in the first period is distorted below or above the efficient level, the value of the marginal gains to society of increasing the probability of skilled individuals of matching in the first period with the value of  $\Psi_h^{LF}$  must be compared:

$$\left\{ \Psi_h^* + \frac{[\lambda_1 + \mu\Omega]}{\lambda_1} F(\Delta u_{hh}^*) [w_{2h} - w_{2\ell} - \Delta u_{hh}^*] \right\} - \Psi_h^{LF} \geq 0.$$

Notice that the social marginal welfare weight is different in the Bergson-Samuelson case: it is  $\frac{[\lambda_1 + \mu\Omega]}{\lambda_1}$  which is the dollar equivalent value of giving an extra dollar to each individual in the economy in the first period.<sup>25</sup>

**Corollary 4.** *Under imperfect information in the first period and assuming that  $\alpha_h$  is uniformly distributed between 0 and  $\bar{\alpha} > 0$ , the optimal tax system distort the wages of the skilled workers below the efficient level. Thus employment of the skilled individuals is higher in the imperfect information case than in the perfect information (or Laissez-faire) case.*

The proof of Corollary 4 is located in Appendix A.2.3. The restriction to the distribution function  $F(\cdot)$  is necessary to derive analytical results when a more general social welfare function is considered. The intuition for this downward pressure on the skilled worker's wage is identical to the Maximin case.

To see if the downward distortion of the skilled worker's wage in the first period result

---

<sup>23</sup>See Appendix A.2.3 for more details.

<sup>24</sup>From the FOC of  $u_{1\ell}$  and  $u_{1h}$  and the fact that (3.18) is binding, we can show that  $W'(v_l) > W'(v_h)$ ,  $W'(v_l) > \lambda_1 > W'(v_h)$  and both  $\lambda_1 > 0$ ,  $\mu > 0$ .

<sup>25</sup>Notice that in the Maximin case,  $\lambda_1 + \mu\Omega = 1$ , since the planner only cares about the expected lifetime utility of the unskilled worker.

holds for the more general distribution function used in the Maximin subsection, i.e.  $F(\alpha_h) = \left(\frac{\alpha_h}{\bar{\alpha}}\right)^\varepsilon$ , I turn to numerical simulations. For these simulations, I use a CES matching function and use the following isoelastic social welfare function:

$$SWF = \pi_\ell \left(\frac{v_\ell}{\sigma}\right)^\sigma + \pi_h \left(\frac{v_h}{\sigma}\right)^\sigma,$$

where  $\sigma$  is the parameter that determines inequality aversion. The higher it is, the less inequality aversion the planner has. The two extremes are  $\sigma \rightarrow -\infty$  which are the Maximin preferences and  $\sigma = 1$  which are the standard utilitarian preferences featuring no inequality aversion. The details of the functional form and parameter values assumptions can be found in Appendix A.2.4 alongside some sensitivity analyzes. The downward distortion result is a feature of all the solutions verified where the incentive constraint is binding and that  $\sigma < 1$ .<sup>26</sup> Examples of simulations results are reported in Table 3.1.

Table 3.1: Characteristics of the Optimal Allocation and Tax System: Changes in  $\varepsilon$

	Isoelastic ( $\sigma \rightarrow 0$ )			
	$\varepsilon = 0.1$	$\varepsilon = 0.3$	$\varepsilon = 0.6$	$\varepsilon = 0.9$
$L_\ell^*$	0.8832	0.8823	0.881	0.8803
$L_h^*$	0.9384	0.9354	0.9319	0.9294
$\% \Delta w_\ell^*$	-2.17	-2.02	-1.85	-1.75
$\% \Delta w_h^*$	-4.11	-4.42	-4.79	-5.1
$T_1(w_{1\ell})$	-2.2648	-2.0522	-1.83	-1.6838
$T_1'(w_{1\ell})$	-1.8027	-1.6418	-1.4754	-1.3601
$T_1(w_{1h})$	2.1062	1.7484	1.3734	1.1261
$T_1'(w_{1h}) - \beta \Psi'_h$	0.564	0.5091	0.4434	0.3945
$u_{0\ell}$	1.0266	1.0085	1.0035	1.0022
$\Delta u_{hh}^*$	1.9606	1.9863	1.9936	1.9953
$v_\ell^*$	3.5066	3.3022	3.1001	2.9692
$v_h^*$	3.8479	3.6224	3.3983	3.2525
$v_h/v_\ell$	1.0973	1.097	1.0962	1.0954

For ease of exposition only the percentage change from the *laissez-faire* wage level is

<sup>26</sup>When  $\sigma = 1$  due to the risk-neutral preferences there is no longer any redistributive motive.

reported and not the actual wage levels.<sup>27</sup> Increases in  $\varepsilon$  imply that the reaction to changes in  $\Delta u_{hh}$  will be greater, or it can be interpreted as the likelihood of deciding to work in the skilled sector in the second period is reduced. One thing that is clear from Table 3.1 is that the wage of the skilled in the first period is always distorted downward like in the Maximin case. As it would be expected an increase in  $\varepsilon$  leads to a reduction in redistribution in the second period. This results in an increased distortion of the skilled worker's wage. Although, less redistribution makes the second period revenue collected lower, the higher value of the social marginal welfare weight caused by the lower utility of both type makes it more valuable. Also, an increase in  $\varepsilon$  radically reduces the value of  $\Psi_h$  which pushes the wage upwards, but since the planner still wants to separate types, he wants to increase the likelihood of the skilled to to keep his human capital. Although not reported, in the last case there is a difference of 10% employment between the *Laissez-faire* outcome and the second best policy. Following this logic, note that as  $\varepsilon$  increases the planner relies less on the distortion of the unskilled worker's wage in the first period to achieve his goals.<sup>28</sup>

Using the results of Table 3.2 to investigate the impact of changing the redistributive tastes of the planner, it is possible to see that as they are lowered, the downward distortion on the skilled workers is not much affected.<sup>29</sup> This is true despite the fact that redistribution in the second period also goes down, almost to null in the last case. Thus the planner lowers the wedge, i.e.  $T'_1(w_{1h}) - \beta\Psi'_h$ , to compensate and keep the wage at around the same level. The important distortion in these four examples is the unskilled worker's wage. The high level of distortion in the Maximin case is analogous to the one found in the standard optimal tax literature, which wants to heavily discourage the skilled to mimic the unskilled. Here it has the added benefit of increasing the employment of the unskilled. But as the redistributive tastes go down, so does the distortion on that labor market. Interestingly, the transfer in

---

<sup>27</sup> $\% \Delta w_i^* = 100 * (w_i^* - w_i^{LF}) / w_i^{LF}$

<sup>28</sup>This result holds for other redistributive tastes, for example Maximin reported in Table A.2.

<sup>29</sup>Here we consider the case when  $\varepsilon = 0.25$ . Also, changes in the parameter determining the search frictions both have a strong impact on the downward distortion of the skilled wage. See Table A.2.

Table 3.2: Characteristics of the Optimal Allocation and Tax System: Various Inequality Aversion

	Maximin	Isoelastic		
		$\sigma \rightarrow 0$	$\sigma = 0.25$	$\sigma = 0.5$
$L_\ell^*$	0.969	0.8825	0.8793	0.8756
$L_h^*$	0.9361	0.9361	0.9361	0.9361
$\% \Delta w_l^*$	-35.36	-2.06	-1.61	-1.12
$\% \Delta w_h^*$	-4.34	-4.34	-4.34	-4.34
$T_1(w_{1\ell})$	-2.1748	-2.0996	-2.101	-2.1021
$T_1'(w_{1\ell})$	0.6966	-1.6777	-1.8025	-1.947
$T_1(w_{1h})$	1.8178	1.8284	1.8294	1.8298
$T_1'(w_{1h}) - \beta \Psi_h'$	0.5452	0.5219	0.5214	0.5208
$u_{0\ell}$	1.0969	1.011	1.0083	1.0059
$\Delta u_{hh}^*$	1.8440	1.9827	1.9868	1.9906
$v_l^*$	3.4875	3.3465	3.3393	3.3311
$v_h^*$	3.5621	3.6715	3.6741	3.677
$v_h/v_l$	1.0214	1.0971	1.1002	1.1038

the first period to the unskilled stays steady despite the reduction in redistributive tastes. Because overall redistribution is going down, it is less interesting for the skilled worker-firm pair to mimic the unskilled market, and this allows the planner to lower the marginal tax rates of the unskilled to get closer to the efficient level of output in this labor market.

### 3.3 Alternative Policy Tools

#### 3.3.1 Training Programs

In this subsection I consider the case when the planner has access to a training technology. At the end of the first period, the planner is able to retrain a fraction  $\rho$  of the skilled workers that did not obtain a job. He is able to do this at cost  $c(z)$  where  $z = \rho\pi_h[1 - L_h(w_{1h})]$  is the number of workers being retrained and the cost function is convex in  $z$ , i.e.  $c'(z) > 0$ ,  $c''(z) > 0$ , and satisfies  $c(0) = 0$ . Unskilled workers in this particular exercise cannot be trained to be skilled workers in the second period. This assumption is made to focus on the destruction of human capital due to unemployment and not human capital acquisition.

However, the assumption could be justified by assuming some underlying characteristic of the unskilled workers that makes training them to attain the skill level of the skilled workers prohibitively costly.

A training technology will have an impact on the potential labor market history of a skilled worker, which can now be split into three events in the first period. The first is that he was employed and worked, the second is that he was unemployed but got retrained and the last is that he was unemployed but was not retrained. These different labor market experience will modify the expected lifetime utility of the skilled worker defined in section 4.1, and the budget constraints and the incentive constraint faced by the planner in the imperfect information scenario. Supposing that the fraction of the skilled worker that will be retrain is assigned randomly to the skilled unemployed, the expected lifetime utility of the skilled worker is:

$$\begin{aligned} v_h &= L_h(w_{1h})[w_{1h} - T_1(w_{1h}) + \beta v_{hh}] + [1 - L_h(w_{1h})]\beta[\rho v_{hh} + (1 - \rho)u_{0\ell}] \\ &= u_{1h} + L_h(w_{1h})\beta(1 - \rho)\Psi_h + \beta[\rho\Psi_h + u_{0\ell}]. \end{aligned}$$

For the budget constraints, not only is there an additional cost in the first period but there is also a modification on how policy affects the number of individual working in the unskilled and skilled market in the second period. Using the same definition of the four groups of workers in the second period as above based on labor market history, we take into account the training program and modify the equations determining the numbers of workers in each

groups:

$$\begin{aligned}
\phi_{hh} &= \pi_h \{L_h(w_{1h}) + \rho[1 - L_h(w_{1h})]\}F(\Delta u_{hh}), \\
\phi_{hl} &= \pi_h \{L_h(w_{1h}) + \rho[1 - L_h(w_{1h})]\}[1 - F(\Delta u_{hh})], \\
\phi_{ll} &= \pi_l L_l(w_{1l}), \\
\phi_{0l} &= \pi_l [1 - L_l(w_{1l})] + \pi_h(1 - \rho)[1 - L_h(w_{1h})].
\end{aligned}$$

From this the budget constraint of the second period can be written to be:

$$\phi_{0l}w_{2l} + \phi_{ll}[w_{2l} - \Delta u_{ll}] + \phi_{hl}[w_{2l} - \Delta u_{hl}] + \phi_{hh}[w_{2h} - \Delta u_{hl} - \Delta u_{hh}] = u_{0l}.$$

### 3.3.1.1 Perfect Information in the First Period

The planner's problem when he is able to observe the skill level of each match in the first period is:

$$\begin{aligned}
\max_{\substack{\{w_{1i}, u_{1i}\}_{\forall i}, u_{0l} \\ \Delta u_{ll}, \Delta u_{hl}, \Delta u_{hh}, \rho}} \quad & \pi_l W(v_l(u_{1l}, w_{1l}, \Delta u_{ll}, u_{0l})) + \pi_h W(v_h(u_{1h}, w_{1h}, \rho, \Delta u_{hl}, \Delta u_{hh}, u_{0l})) \\
s.t. \quad & (\lambda_1) \quad \pi_l \cdot [L_l(w_{1l}) \cdot w_{1l} - u_{1l}] + \pi_h \cdot [L_h(w_{1h}) \cdot w_{1h} - u_{1h}] - c(\rho\pi_h[1 - L_h(w_{1h})]) = 0, \\
& (\lambda_2) \quad \phi_{0l}(w_{1l}, w_{1h}, \rho)w_{2l} + \phi_{ll}(w_{1l})[w_{2l} - \Delta u_{ll}] \\
& \quad + \phi_{hl}(w_{1h}, \Delta u_{hh}, \rho)[w_{2l} - \Delta u_{hl}] + \phi_{hh}(w_{1h}, \Delta u_{hh}, \rho)[w_{2h} - \Delta u_{hl} - \Delta u_{hh}] = u_{0l}.
\end{aligned} \tag{3.28}$$

Similar to the perfect information case from above, each solution must satisfy the following necessary conditions:

$$u_{1\ell} + \beta L_\ell(w_{1\ell})\Psi_\ell(\Delta u_{\ell\ell}) = u_{1h} + L_h(w_{1h})\beta(1-\rho)\Psi_h(\Delta u_{h\ell}, \Delta u_{hh}) + \beta\rho\Psi_h(\Delta u_{h\ell}, \Delta u_{hh}), \quad (3.29)$$

$$\Delta u_{hh} = w_{2h} - w_{2\ell}, \quad (3.30)$$

$$\beta\Psi^{LF} = c'(\rho\pi_h[1 - L_h(w_{1h})]), \quad (3.31)$$

$$-\frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} \frac{w_{1\ell}}{L_\ell(w_{1\ell})} = 1, \quad (3.32)$$

$$-\frac{\partial L_h(w_{1h})}{\partial w_{1h}} \frac{w_{1h}}{L_h(w_{1h})} = \frac{1}{1 + \frac{\beta\Psi_h^{LF}}{w_{1h}}}. \quad (3.33)$$

Note that from the first condition the expected lifetime utilities of both types of workers are equalized. Condition (3.30) which helps derive the rest of the other conditions is similar to the one found above, i.e. there is no difference in tax burden of the skilled workers irrelevant of their occupational choice in the second period. This feature of the solution implies that there is no fiscal gains of having the number of skilled workers in the second period above the efficient value. This has a direct impact on (3.31), since increasing  $\rho$  will increase the expected lifetime utility of the skilled but also the number of skilled workers in the second period. Since there are no fiscal gains to having more skilled workers than is efficient, the efficient level of  $\rho$  will be set where the marginal cost of increasing  $\rho$  is equal the discounted value of the expected net utility gain of being skilled in the second period.

From conditions (3.32) and (3.33) it is possible to see that in the perfect information scenario, adding the training technology has no impact on the optimal wages of each skill level, they are set at their *laissez-faire* level. In this context, for the wage of the unskilled worker the result is not surprising since  $\rho$  does not directly enter the lifetime expected

---

<sup>30</sup>The FOCs have been partly solved to remove multipliers.

utility of an unskilled worker. The same result for the skilled worker is less straightforward. Decreasing the skilled worker's first-period wage increases the number of skilled workers that will find a job and keep their human capital, but the effect of this wage decrease on the future is mitigated by the presence of the training technology, since there will be a fraction of those that did not find a job that will be retrained regardless of the market wage. As for the budget constraints, the standard effect on total production is still there, but the interesting part is that a decrease in the wage will lower the cost of retraining for a given  $\rho$  since there are now less skilled unemployed to retrain. However this effect is not one for one since only a fraction  $\rho$  will be retrained. And since there are no fiscal gains in the second period there is no value for that period to lowering the wage below the efficient value. Taking all the effects into account the first order condition with respect to the wage of the skilled in the first period can be written in the following way:

$$-\frac{\partial L_h(w_{1h})}{\partial w_{1h}} \frac{[w_{1h} + \beta \Psi_h^{LF}]}{L_h(w_{1h})} = 1 + \frac{1}{L_h(w_{1h})} \frac{\partial L_h(w_{1h})}{\partial w_{1h}} [\rho c'(\cdot) - \rho \beta \Psi_h^{LF}] = 0.^{31}$$

The benefit of lowering the wage on the cost of training and lowered impact of manipulating the wage caused by  $\rho$  can be seen in the right-hand side of the above equation. Condition (3.31) says that both of these effects are equal and thus cancel each other out which results in (3.33).

### 3.3.1.2 Imperfect Information in the Second Period

To keep in line with the results derived in section 3.2, I consider only the optimal tax system that does not offer insurance, i.e.  $\Delta u_{\ell\ell} = \Delta u_{h\ell} = 0$ , in concert with the presence of the training technology. By the same logic, the planner will not use the information gathered while training the unemployed skilled workers and offer them compensation in the second period for not working in the first period.

---

<sup>31</sup>Here the lagrange multiplier has been solved out.

The presence of the training program has an impact on the incentive constraint the planner faces when he is unable to observe the skill level of a match. The major impact on the expected lifetime utility of a ‘mimicker’ comes from when he does not find a job, or in the case of bargaining willingly decides not to work. Since I have assumed that skilled workers can get retrained with some probability and that if the worker that mimicked did not find a job, there is no way for the planner to know that this worker tried to mimic in the first period. In that situation the mimicker is able to be retrained, and since it is costless for him to do so he will choose to do it. One way of thinking about this mimicking would be in a Competitive Search Equilibrium setting where the expected lifetime utility of a skilled worker that searches in a market with the unskilled worker’s wage prevailing is:

$$\begin{aligned}
\widehat{v}_h &= L_h(w_{1\ell})[w_{1\ell} - T_1(w_{1\ell}) + \beta u_{0\ell}] + [1 - L_h(w_{1\ell})]\beta[\rho v_{hh} + (1 - \rho)u_{0\ell}], \\
&= L_h(w_{1\ell})[w_{1\ell} - T_1(w_{1\ell}) + \beta(u_{0\ell} - \rho v_{hh} - (1 - \rho)u_{0\ell})] + \beta[\rho\Psi_h + u_{0\ell}], \\
&= L_h(w_{1\ell})[w_{1\ell} - T_1(w_{1\ell}) - \beta\rho\Psi_h] + \beta[\rho\Psi_h + u_{0\ell}].
\end{aligned}$$

The incentive constraint of the planner under imperfect information is then:

$$u_{1h} + \beta L_h(w_{1h})(1 - \rho)\Psi_h(\Delta u_{hh}) \geq \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})}u_{1\ell} - L_h(w_{1\ell})\beta\rho\Psi_h(\Delta u_{hh}).^{32}$$

Rewriting the incentive constraint is instructive on how introducing a training program alters the other policy choices’ effect on the incentive to mimic:

$$u_{1h} + \beta L_h(w_{1\ell})\Psi_h(\Delta u_{hh}) + \rho\beta[L_h(w_{1\ell}) - L_h(w_{1h})]\Psi_h(\Delta u_{hh}) \geq \frac{L_h(w_{1h})}{L_\ell(w_{1\ell})}u_{1\ell}.$$

Looking at the left-hand side, it is possible to see that it is composed of three parts with the two first part being the one found in the incentive constraint of the problem without

---

<sup>32</sup>For more details on the Competitive Search Equilibrium can be found in Appendix A.2.1.

a training program. The right-hand side is also the same as in the above section with the no-insurance tax system. The difference now comes from the third part on the left-hand side which combines the expected value of not finding a job for a ‘mimicker’ and the reduction in value of the matching surplus caused by the training program. Under reasonable parameter values  $w_{1h} > w_{1\ell}$  is expected, which would indicate that this third part has a positive value since  $L_h(w_{1\ell}) > L_h(w_{1h})$ .

The analysis of the training program is done using the Maximin social welfare function. The problem of the planner is:

$$\begin{aligned}
& \max_{\{w_{1i}, u_{1i}\}_{\forall i}} \quad u_{1\ell} + \beta u_{0\ell} \\
& \quad u_{0\ell}, \Delta u_{hh}, \rho \\
& \text{s.t. } (\lambda_1) \quad \pi_\ell \cdot [L_\ell(w_{1\ell}) \cdot w_{1\ell} - u_{1\ell}] + \pi_h \cdot [L_h(w_{1h}) \cdot w_{1h} - u_{1h}] - c(\rho\pi_h[1 - L_h(w_{1h})]) = 0, \\
& \quad (\lambda_2) \quad \phi_{0\ell}(w_{1\ell}, w_{1h}, \rho)w_{2\ell} + \phi_u(w_{1\ell})w_{2\ell} \\
& \quad \quad + \phi_{hl}(w_{1h}, \Delta u_{hh}, \rho)w_{2\ell} + \phi_{hh}(w_{1h}, \Delta u_{hh}, \rho)[w_{2h} - \Delta u_{hh}] = u_{0\ell} \\
& \quad (\mu) \quad u_{1h} + \beta L_h(w_{1h})(1 - \rho)\Psi_h(\Delta u_{hh}) \geq \frac{L_h(w_{1h})}{L_\ell(w_{1\ell})}u_{1\ell} - L_h(w_{1\ell})\beta\rho\Psi_h(\Delta u_{hh}).
\end{aligned}$$

It is possible to see that increasing  $\rho$  will slacken the incentive constraint. The effect on the incentive constraint will be greater the less redistribution there is in the second period and the bigger the difference between the labor demand for the skilled worker at the different wage levels. A reduction in redistribution, i.e.  $\Delta u_{hh}$ , also slackens the incentive constraint. It does so in the same manner as before by increasing the potential payoff of being skilled in the second period if one chooses the market with the skilled worker’s wage. But the presence of the training program increases the effect of a reduction in redistribution since it makes the unskilled worker’s bundle and actually finding a job at the unskilled’s wage, which implies

that there is no possibility of getting the skilled worker's income in the second period, less appealing.

Another effect is the one on the unskilled worker's wage, it slackens the incentive constraint in an additional way. The lower the unskilled's wage is the higher the probability of a mimicking skilled worker to find a job and thus not have access to retraining. The effect is reversed for the skilled worker's wage. Although lowering the skilled worker's wage increases the probability of finding a job and keeping the human capital of the worker, this effect on the left-hand side is mitigated, as mentioned in the perfect information case, by the training program. For the skilled worker, matching is no longer the only way to keep his human capital. He is less willing to take a lower wage since this technology reduces the surplus of matching. In the end the total effect of the skilled worker's wage on the incentives is difficult to evaluate from the above equation, but there is now this upward effect on the wage because training lowers the expected surplus of matching.

As it turns out the effect of  $\rho$  on the optimal allocation will strongly depend on the skilled labor demand at the first period unskilled wage  $w_{1\ell}$ , i.e.  $L_h(w_{1\ell})$ . First start with the first-order condition<sup>33</sup> for the decision of  $\Delta u_{hh}$ ,  $\rho$  and  $w_{1\ell}$ :

$$[w_{2h} - w_{2\ell} - \Delta u_{hh}^*] = (1 - \lambda_1 \Upsilon) \frac{F(\Delta u_{hh}^*)}{f(\Delta u_{hh}^*)}, \quad (3.34)$$

$$\beta \left[ \frac{L_h(w_{1\ell}) - L_h(w_{1h})}{1 - L_h(w_{1h})} \Psi_h^* + \frac{F(\Delta u_{hh})[w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1} \right] = c'(\rho \pi_h [1 - L_h(w_{1h})]), \quad (3.35)$$

$$-\frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} \frac{w_{1\ell}}{L_\ell(w_{1\ell})} = 1 - \frac{\mu}{\lambda_1 \pi_\ell} \left\{ \frac{L_h(w_{1\ell}^*) u_{1\ell}}{[L_\ell(w_{1\ell}^*)]^2} \left[ \frac{\partial L_h(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_h(w_{1\ell}^*)} - \frac{\partial L_\ell(w_{1\ell}^*)}{\partial w_{1\ell}} \frac{1}{L_\ell(w_{1\ell}^*)} \right] - \beta \frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} \rho \Psi_h \right\}, \quad (3.36)$$

---

<sup>33</sup>With both the lagrange multiplier of the second period budget constraint and the one of the incentive constraint,  $\mu = \lambda_1 \pi_h$ .

where

$$\Upsilon = \frac{L_h(w_{1h})(1 - \rho) + \rho L_h(w_{1\ell})}{L_h(w_{1h})(1 - \rho) + \rho}.$$

Notice that (3.34) is slightly different than (3.25) due to the presence of  $\Upsilon$ .  $\Upsilon$  captures the tradeoff between the incentive effect of decreasing redistribution (the numerator), and loss of tax revenue of this reduction in redistribution (the denominator). Assuming again that  $F(\alpha_h) = \left(\frac{\alpha_h}{\bar{\alpha}}\right)^\varepsilon$ , then  $d\Delta u_{hh}/d\Upsilon > 0$ , this implies that there is less redistribution the greater  $\Upsilon$  is. This is where the importance of  $L_h(w_{1\ell})$  comes in. If  $w_{1\ell}$  is high enough and  $L_h(w_{1\ell}) < 1$ , then  $\Upsilon$  must be smaller than 1 which hints at more redistribution than without a training technology. Although without being able to compare the values of  $\lambda_1$  for both cases, there is no definite answer. But what it says is that the effect on the budget of the training program is greater than the incentive effect. But if  $w_{1\ell}$  is low enough so that  $L_h(w_{1\ell})$  tends to or equals 1, then both effects are equal and we recover (3.20).

The tradeoff between the budget effects and the incentive effects are also present in (3.35) which is the first-order condition of  $\rho$ . The left-hand side is the gains in expected utility and fiscal revenues mixed with the incentive effect of increasing the number of skilled workers in the second period. The right-hand side is the marginal cost of increase  $\rho$ . The left-hand side looks similar to gains brought by reducing the number of skilled workers by lowering the skilled worker's wage, but there is a positive term multiplying  $\Psi_h$ .

This term, like  $\Upsilon$ , measures in part the tradeoff between the budget concerns (denominator) and the incentive concerns (numerator). Since by marginally increasing  $\rho$  the fiscal benefit is multiplied by number of potential skilled workers, i.e.  $1 - L_h(w_{1h})$ , but this is also true for the marginal cost. What is left is the incentive effect captured by  $L_h(w_{1\ell}) - L_h(w_{1h})$ . Again, the value of  $L_h(w_{1\ell})$  becomes important in determining if  $\rho$  will be higher than in the perfect information case. Starting with the case when  $L_h(w_{1\ell}) < 1$ , this means that the term multiplying  $\Psi_h$  is below 1, which reduces the weight of the value of keeping the level of human capital. Because of this, it is quite possible that the left-hand side of (3.35) is smaller

in value than  $\beta\Psi_h^{LF}$ , which means that  $\rho$  can be smaller than in the perfect information case. Alternatively the more  $L_h(w_{1h})$  approaches 1, the more likely it is that  $\rho$  is above the one in perfect information. With  $L_h(w_{1h}) = 1$  then with a similar proof as in the above section it follows that  $\Psi_h^* + \frac{F(\Delta u_{hh})[w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1} > \Psi_h^{LF}$  and thus  $\rho$  in the imperfect information setting is always greater than in the perfect information setting.

Condition (3.36) is similar to the one derived above. In fact it is the same with the added term  $-\beta\frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}}\rho\Psi_h$ , which is positive. Thus, there is the difference in elasticities helping to separate the two types, but also the incentive effect brought by the training program through a decrease in wage which increases the probability of matching of a mimicker, making the mimicking bundle less attractive since he must forgo the benefit brought on by keeping his human capital. But one more effect must be taken in consideration. Since  $\rho$  is able to reduce the information problem, the value of the lagrange multiplier related to the incentive constraint ( $\mu$ ) can be lower. This pulls the wage towards *Laissez-faire* level, thus upwards.

The skilled worker's wage in the first period has many effects, from the fiscal gain, to the incentive effect and also the reduction in cost of the training program. To be able to say more on the optimal level of  $w_{1h}$  under imperfect information, I incorporate (3.35) in the first-order condition of  $w_{1h}$  to get:

$$-\frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \frac{w_{1h}^*}{L_h(w_{1h}^*)} = 1 + \frac{\beta}{L_h(w_{1h}^*)} \frac{\partial L_h(w_{1h}^*)}{\partial w_{1h}} \left\{ \left( 1 + \rho \frac{L_h(w_{1\ell}) - 1}{1 - L_h(w_{1h})} \right) \Psi_h^* + \frac{F(\Delta u_{hh}^*)[w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1} \right\}. \quad (3.37)$$

The impact of having the training program on the first-order condition can be seen by looking at the term inside the curly bracket that is multiplying  $\Psi_h$ . It incorporates both the negative effect on incentive brought on by the training program but also the positive benefit on lowering the cost of the training program. If  $L_h(w_{1\ell})$  is smaller than one, then the entire term multiplying  $\Psi_h$  is smaller than one. The closer  $L_h(w_{1\ell})$  gets to 1, the more the

first-order condition looks identical to the one without the training program. To determine if the skilled worker's wage in the first period is distorted below the *Laissez-faire* level, it must be determined if what is inside the curly bracket is bigger than  $\Psi_h^{LF}$ .

Considering the case when  $L_h(w_{1\ell}) < 1$ , i.e.  $\Upsilon < 1$ , which implies that redistribution can be quite substantial in the second period and thus lowering  $\Psi_h$ , this also makes the term multiplying  $\Psi_h$  smaller than one. Assuming the functional form for the distribution of  $\alpha_h$  above or even the uniform distribution cannot help us rule out the case when

$$\Psi_h^{LF} > \left(1 + \rho \frac{L_h(w_{1\ell}) - 1}{1 - L_h(w_{1h})}\right) \Psi_h^* + \frac{F(\Delta u_{hh}^*)[w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1}.$$

In this case, the skilled worker's wage could be pushed above the *Laissez-faire* level. This comes about for several reasons. The first one is that the training technology makes the skilled workers in the first period much less willing to take a lower wage since the surplus from matching is both lower from more redistribution but is also lower since there is always a probability that if he becomes unemployed he will have a chance to regain the human capital lost. Thus distorting the wage below the *Laissez-faire* level becomes quite costly. Since the skilled worker's wage is higher, it also means that the fraction  $\rho$  affects a larger pool of workers, and thus the program is much more expensive. Hence we can better see the tradeoff between having a lower  $\rho$  and a bigger pool of potential candidates to retrain. But if  $w_{1\ell}$  is low enough such that  $L_h(w_{1\ell})$  approaches or is equal to 1, then  $\Upsilon = 1$  and the term multiplying  $\Psi_h$  is also 1, and therefore  $w_{1h}$  is distorted below the *Laissez-faire* level. The tradeoff between wage and  $\rho$  shows up, since  $\rho$  is now bigger than the perfect information case and the pool of skilled unemployed is now lower due to the skilled worker's wage being below the *Laissez-faire* level. The dependence of the result on  $L_h(w_{1\ell})$  stems from when  $L_h(w_{1\ell}) < 1$  the incentive effects are less important compared to the budget considerations of taxing more people in the second period. When  $L_h(w_{1\ell}) = 1$ , this is no longer the case, and the incentive effect dominates.

To untangle all of these effects, I use numerical simulations which the results are summarized in Table 3.3.<sup>34</sup> In the numerical simulations I consider different redistributive tastes.

Table 3.3: Characteristics of the Optimal Allocation and Tax System: Training Program

	Maximin		Isoelastic SWF $\sigma \rightarrow 0$	
	without	with	without	with
$\rho^*$	-	0.314	-	0.329
$L_\ell^*$	0.969	0.9691	0.8825	0.8817
$L_h^*$	0.9361	0.936	0.9361	0.9361
$\% \Delta w_\ell^*$	-35.36	-35.43	-2.06	-1.95
$\% \Delta w_h^*$	-4.34	-4.32	-4.34	-4.34
$T_1(w_{1\ell})$	-2.1748	-2.1787	-2.0996	-2.0923
$T_1'(w_{1\ell})$	0.6966	0.6972	-1.6777	-1.6994
$T_1(w_{1h})$	1.8178	1.8269	1.8284	1.8181
$T_1'(w_{1h}) - \beta \Psi_h'$	0.5452	0.6323	0.5219	0.621
$u_{0\ell}$	1.0969	1.1029	1.011	1.0294
$\Delta u_{hh}^*$	1.8440	1.8377	1.9827	1.9545
$v_\ell^*$	3.4875	3.4959	3.3465	3.3537
$v_h^*$	3.5621	3.5715	3.6715	3.6815
$v_h/v_\ell$	1.0214	1.0216	1.0971	1.0977

Incorporating the training has only slight effects on the optimal wages and thus the employment levels. In the maximin case, the incentive effect of incorporating  $\rho$  dominates the reduction in the value of  $\mu$ . But this result is overturned when a lower aversion to inequality preference is considered. In both cases the planner attempts much more redistribution in the second period. This should have a positive effect on the skilled worker's wage, but the optimal income tax system requires a higher wedge to put a downward pressure to counteract the upward pressure of this increase in redistribution in the second period. Overall, this technology benefits both skill levels as their lifetime expected utility is greater, and this is so even if the skilled worker his taxed more.

<sup>34</sup>The functional forms and parameter values are the same as those highlighted in Appendix A.2.4. The only addition is the cost function of of the training program which is chosen to be the following quadratic function  $c(z) = (\frac{75 \cdot z}{2})^2$ .

### 3.3.2 Job Creation Subsidies

In this section job creation subsidies offered to any firm that opens a vacancy are considered and I investigate if they can be welfare improving. We only consider the case of a uniform<sup>35</sup> subsidy  $s \geq 0$  which is given to a firm after they have proved they paid a cost of  $\kappa_i$ .<sup>36</sup>

Before going on with the planner's problem must first determine how labor market tightness and the labor demand reacts to an increase in the subsidy  $s$ . First, recall that the zero-profit condition helps pin down market tightness, this condition is slightly modified and becomes  $m(\theta_i)[a_i - w] = \kappa_i - s$ , which means that

$$\theta_i = m^{-1}\left(\frac{\kappa_i - s}{a_i - w}\right), \quad L_i(w, s) = m^{-1}\left(\frac{\kappa_i - s}{a_i - w}\right) \frac{\kappa_i - s}{a_i - w}.$$

From this we get

$$\frac{\partial \theta_i}{\partial s} = -\frac{1}{m'(\cdot)} > 0, \quad \text{and} \quad \frac{\partial L_i(w, s)}{\partial s} = -\frac{1}{m'(\cdot)} \frac{\kappa_i - s}{a_i - w} - m^{-1}\left(\frac{\kappa_i - s}{a_i - w}\right) \frac{1}{a_i - w}.$$

The first implies that by reducing the cost of opening up a vacancy through a subsidy would lead to an increase in the number of firms opening vacancies and increasing market tightness. The effect of increasing  $s$  on labor demand is not straightforward and as we can see it is composed of two terms. The first term is positive but the second one is negative. To analyze the incorporation of job creation subsidies, more structure on the labor demand function must be added. Therefore I assume a CES matching function  $M(U_i, V_i) = [U_i^{-\gamma} + V_i^{-\gamma}]^{-\frac{1}{\gamma}}$

---

<sup>35</sup>The case where  $s$  would vary with the actual cost would automatically reveal the skill level of each job if that specific firm would meet with a worker and the information problem of the planner would not exist. Although an interesting tool, I leave for further research a more complicated scenario where a subsidy would not reveal automatically the skill level of a match.

<sup>36</sup>Only positive subsidies are considered, negative values of  $s$  would be a tax and since firms only last for one period in this model and there are no financial markets, this tax revenue would not be modeled properly.

which gives the following labor demand for skill level  $i$ :

$$L_i(w, s) = \left[ 1 - \left( \frac{a_i - w}{\kappa_i - s} \right)^{-\gamma} \right]^{\frac{1}{\gamma}},$$

where  $\gamma \geq 0$ . With the functional form assumption and the standard assumption that  $\frac{\partial \kappa(a)}{\partial a} \frac{a}{\kappa(a)} \leq 1$ , the result follows:

$$\frac{\partial L_i(w, s)}{\partial s} > 0.$$

Another feature of this functional form is that the elasticity of labor demand with respect to  $s$ , i.e.  $\frac{\partial L_i(w, s)}{\partial s} \frac{s}{L_i(w, s)}$ , is decreasing with skill since  $\kappa_i$  increases with skill.

The planner's problem in both informational cases is similar with the exception that  $s$  will now enter the labor demand in each market and that revenue must be raised to pay for the subsidy. Since  $s$  is going to be given for each vacancy in both markets the total cost of this policy is  $(V_l + V_h)s$ . Recall that  $\theta_i = \frac{V_i}{U_i}$  and that  $U_i = \pi_i$ , thus we get  $V_i = \pi_i \theta_i$ . The new budget constraint in the first period faced by the planner is:

$$\pi_\ell [L_\ell(w_{1\ell}, s)w_{1\ell} - u_{1\ell} - \theta_\ell(w_{1\ell}, s)s] + \pi_h [L_h(w_{1h}, s)w_{1h} - u_{1h} - \theta_h(w_{1h}, s)s] = 0.$$

Lets first consider the case of perfect information, the planner's problem is:

$$\begin{aligned}
& \max_{\substack{\{w_{1i}, u_{1i}\}_{\forall i} \\ u_{0\ell}, \Delta u_{\ell\ell}, \Delta u_{h\ell} \\ \Delta u_{hh}, s}} \pi_\ell W\left(v_\ell(u_{1\ell}, w_{1\ell}, s, \Delta u_{\ell\ell}, u_{0\ell})\right) + \pi_h W\left(v_h(u_{1h}, w_{1h}, s, \Delta u_{h\ell}, \Delta u_{hh}, u_{0\ell})\right) \\
& \text{s.t. } (\lambda_1) \pi_\ell \cdot [L_\ell(w_{1\ell}, s) \cdot w_{1\ell} - u_{1\ell} - \theta_\ell(w_{1\ell}, s)s] \\
& \quad + \pi_h \cdot [L_h(w_{1h}, s) \cdot w_{1h} - u_{1h} - \theta_h(w_{1h}, s)s] = 0, \\
& (\lambda_2) \phi_{0\ell}(w_{1\ell}, w_{1h}, s)w_{2\ell} + \phi_{\ell\ell}(w_{1\ell}, s)[w_{2\ell} - \Delta u_{\ell\ell}] \\
& \quad + \phi_{h\ell}(w_{1h}, \Delta u_{hh}, s)[w_{2\ell} - \Delta u_{h\ell}] + \phi_{hh}(w_{1h}, \Delta u_{hh}, s)[w_{2h} - \Delta u_{h\ell} - \Delta u_{hh}] = u_{0\ell}.
\end{aligned} \tag{3.38}$$

To investigate if adding employment subsidies is welfare improving, I evaluate the gains to social welfare of increasing  $s$  at  $s = 0$ . First, when  $s = 0$  the necessary conditions for the planner's problem are identical to the ones derived in section 4.1. This means that there is no difference in the tax treatment of the skilled worker in the second period even when working in different labor markets, i.e.  $\Delta u_{hh} = w_{2h} - w_{2\ell}$  and that wages in the first period are at their *Laissez-faire* levels. The effect of increasing  $s$  on welfare is

$$\begin{aligned}
\left. \frac{\partial \mathcal{L}}{\partial s} \right|_{s=0} &= \pi_\ell \left[ \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} w_{1\ell} - \theta_\ell(w_{1\ell}, s) \right] + \pi_h \left[ \frac{\partial L_h(w_{1h}, s)}{\partial s} [w_{1h} + \beta \Psi_h^{LF}] - \theta_h(w_{1h}, s) \right] \\
& \quad + \beta \pi_h \frac{\partial L_h(w_{1h}, s)}{\partial s} F(\Delta u_{hh}) [w_{2h} - w_{2\ell} - \Delta u_{hh}], \\
&= \pi_\ell \left[ \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} w_{1\ell} - \theta_\ell(w_{1\ell}, s) \right] + \pi_h \left[ \frac{\partial L_h(w_{1h}, s)}{\partial s} [w_{1h} + \beta \Psi_h^{LF}] - \theta_h(w_{1h}, s) \right].
\end{aligned}$$

Due to  $\Delta u_{hh} = w_{2h} - w_{2\ell}$ , the effects that are left are the first period resources gains in each skill level. We have the resources gains from more individuals finding jobs in each labor market, including the more production in the second period when there are more

skilled workers, but also the revenue loss for each subsidized vacancy. Using the zero-profit condition in both labor market when  $s > 0$ , we have that  $L_\ell(w, s)a_\ell - \theta\kappa_\ell = L_\ell(w, s)w - \theta s$  and  $L_h(w, s)[a_h + \beta\Psi^{LF}] - \theta\kappa_h = L_h(w, s)w - \theta s$ . At  $s = 0$ , both wages are set at their *Laissez-faire* level, which is the level that maximizes the output net of vacancy costs. Using the envelope theorem this result must be true:

$$\begin{aligned}\frac{\partial L_\ell(w_{1\ell}^*, s)}{\partial s} w_{1\ell}^* - \theta_\ell(w_{1\ell}^*, s) &= 0. \\ \frac{\partial L_h(w_{1h}^*, s)}{\partial s} [w_{1h}^* + \beta\Psi_h^{LF}] - \theta_h(w_{1h}^*, s) &= 0.\end{aligned}$$

From this I can conclude that  $\frac{\partial \mathcal{L}}{\partial s}|_{s=0} = 0$ , implying that there are no welfare gains from adding a subsidy to employment under perfect information. This result can be obtained in this context as taxation can redistribute until lifetime expected utility is equal while at the same time ensuring that the economy produces to its net-output maximizing level. Therefore using the additional and less precise tool of the employment subsidy does not bring any welfare gains.

The perfect information in the first period result does not as a rule carry to the imperfect information case because the job creation subsidy can help in separating different skill types. In this subsection I consider the maximin case where it is possible to derive more general properties of the solution. I also only consider the tax system that features no insurance for

market outcomes along the working life. The planner's problem is then:

$$\begin{aligned}
& \max_{\{w_{1i}, u_{1i}\}_{\forall i}, u_{0\ell}, \Delta u_{hh}, s} && u_{1\ell} + \beta u_{0\ell} \\
& \text{s.t. } (\lambda_1) && \pi_\ell \cdot [L_\ell(w_{1\ell}, s) \cdot w_{1\ell} - u_{1\ell} - \theta_\ell(w_{1\ell}, s)s] \\
& && + \pi_h \cdot [L_h(w_{1h}, s) \cdot w_{1h} - u_{1h} - \theta_h(w_{1h}, s)s] = 0, \\
& (\lambda_2) && \phi_{0l}(w_{1\ell}, w_{1h}, s)w_{2\ell} + \phi_{ll}(w_{1\ell}, s)w_{2\ell} \\
& && + \phi_{hl}(w_{1h}, \Delta u_{hh}, s)w_{2\ell} + \phi_{hh}(w_{1h}, \Delta u_{hh}, s)[w_{2h} - \Delta u_{hh}] = u_{0\ell} \\
& (\mu) && u_{1h} + L_h(w_{1h}, s)\beta\Psi_h(\Delta u_{hh}) \geq \frac{L_h(w_{1\ell}, s)}{L_\ell(w_{1\ell}, s)}u_{1\ell}.
\end{aligned}$$

Evaluating the problem at  $s = 0$ , we have the same necessary conditions as in section 3.2.

The first-order condition with respect to  $s$  evaluated at  $s = 0$  is:

$$\begin{aligned}
\left. \frac{\partial \mathcal{L}}{\partial s} \right|_{s=0} = & \lambda_1 \left\{ \pi_\ell \left[ \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} w_{1\ell} - \theta_\ell(w_{1\ell}, s) \right] + \pi_h \left[ \frac{\partial L_h(w_{1h}, s)}{\partial s} w_{1h} - \theta_h(w_{1h}, s) \right] \right\} \\
& + \beta \pi_h \frac{\partial L_h(w_{1h}, s)}{\partial s} F(\Delta u_{hh}) [w_{2h} - w_{2\ell} - \Delta u_{hh}] \\
& + \mu \left\{ \beta \frac{\partial L_h(w_{1h}, s)}{\partial s} \Psi_h - u_{1\ell} \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} \left[ \frac{\partial L_h(w_{1\ell}, s)}{\partial s} \frac{1}{L_h(w_{1\ell})} - \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} \frac{1}{L_\ell(w_{1\ell})} \right] \right\}.
\end{aligned}$$

As above, the equation is the sum of the effect on the first period budget constraint, the fiscal gains of adding more skilled types in the second period, but also the effect on the incentive constraint which is the term on the third line inside the curly brackets. The first term is the value of increasing the probability of keeping the worker's human capital, the second term is the difference in the marginal rate of substitution between after-tax income and subsidy  $s$  coming from the wage-setting objective. Using the assumption made on the matching function this marginal rate of substitution can be shown to be increasing in the skill level. This implies that the planner can decrease  $u_{1\ell}$  and increase  $s$  and this will make the unskilled's bundle less appealing to the skilled mimickers. This is because the increase

in the probability of matching of the unskilled workers due to the subsidy is greater than the one of the mimicking skilled worker-firm, which permits lowering slightly the after-tax income of the unskilled worker while keeping the expected utility of the unskilled worker, and at the same relaxing the incentive constraint.

Rearranging and substituting in  $\mu$ , we have:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} \Big|_{s=0} = & \lambda_1 \left\{ \pi_h \left( \frac{\partial L_h(w_{1h}, s)}{\partial s} \left[ w_{1h} + \beta \left( \Psi_h + \frac{F(\Delta u_{hh})[w_{2h} - w_{2\ell} - \Delta u_{hh}]}{\lambda_1} \right) \right] - \theta_h(w_{1h}, s) \right. \right. \\ & \left. \left. - u_{1\ell} \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} \left[ \frac{\partial L_h(w_{1\ell}, s)}{\partial s} \frac{1}{L_h(w_{1\ell})} - \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} \frac{1}{L_\ell(w_{1\ell})} \right] \right) \right. \\ & \left. + \pi_\ell \left( \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} w_{1\ell} - \theta_\ell(w_{1\ell}, s) \right) \right\}. \end{aligned}$$

At  $s = 0$  and with the maximin social welfare function, the necessary condition for  $w_{1h}$  is identical to (3.22). Using this condition, and the functional form assumption it is straightforward to show that:

$$\frac{\partial L_h(w_{1h}, s)}{\partial s} \left[ w_{1h} + \beta \left( \Psi_h + \frac{F(\Delta u_{hh})[w_{2h} - w_{2\ell} - \Delta u_{hh}]}{\lambda_1} \right) \right] - \theta_h(w_{1h}, s) = 0.$$

Again, using (3.21) and the functional form assumption, I get that

$$\frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} w_{1\ell} - \theta_\ell(w_{1\ell}, s) < 0.$$

Combining both these results, the potential welfare gains are:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial s} \Big|_{s=0} = & \lambda_1 \pi_h \left\{ -u_{1\ell} \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} \left[ \frac{\partial L_h(w_{1\ell}, s)}{\partial s} \frac{1}{L_h(w_{1\ell})} - \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} \frac{1}{L_\ell(w_{1\ell})} \right] \right. \\ & \left. + \frac{\pi_\ell}{\pi_h} \left( \frac{\partial L_\ell(w_{1\ell}, s)}{\partial s} w_{1\ell} - \theta_\ell(w_{1\ell}, s) \right) \right\}. \end{aligned}$$

The above equation is composed of two terms, the first term is positive and represents the gains of slacking the incentive constraint by increasing  $s$ , the second term is the net effect

on first period resources in the unskilled labor market of increasing  $s$ . Since this net effect on resources is negative, it does not imply automatically that the global effect of increasing  $s$  is positive.

Using numerical simulations, I find that the optimal level of subsidies  $s$  is zero unless the labor market frictions are such that the optimal allocation requires the unskilled worker's wage be distorted all the way to zero. At this point, the wage can no longer be used as a tool to separate types. The planner then uses the job creation subsidy to separate types further. This is illustrated in Table 3.4, as  $\gamma$  decreases, i.e. making the  $U$  and  $V$  less complimentary and therefore increasing market frictions, the optimal plan calls for distorting the unskilled worker's wage dramatically. In the example, when  $\delta \leq 0.5$  the planner sets  $w_l = 0$  and introduces small subsidies. An interesting characteristic of the allocation is that as the frictions become worse, that does not mean that the planner will want to distort the skilled worker's wage more. Actually the contrary is possible as can be seen when we compare the case at  $\gamma = 0.5$  and the one at  $\gamma = 0.3$ .

Table 3.4: Characteristics of the Optimal Allocation and Tax System: Job Creation Subsidies

	Maximin			
	Benchmark	$\gamma = 1$	$\gamma = 0.5$	$\gamma = 0.3$
$s^*$	-	0	0.0119	0.0248
$L_\ell^*$	0.969	0.881	0.4944	0.1281
$L_h^*$	0.9361	0.8178	0.4627	0.1361
$\% \Delta w_l^*$	-35.36	-76.65	-100	-100
$\% \Delta w_h^*$	-4.34	-9.26	-15.22	-14.45
$T_1(w_{1\ell})$	-2.1748	-1.9498	-1.2842	-0.721
$T_1'(w_{1\ell})$	0.6966	0.6609	0.2289	1
$T_1(w_{1h})$	1.8178	1.6167	1.237	1.1048
$T_1'(w_{1h}) - \beta \Psi'_h$	0.5452	0.5516	0.369	-0.3836
$u_{0\ell}$	1.0969	1.1936	1.2238	1.0843
$\Delta u_{hh}^*$	1.8440	1.6325	1.187	0.8769
$v_l^*$	3.4875	2.8337	1.6348	0.9783
$v_h^*$	3.5621	2.9815	1.8098	1.0363
$v_h/v_l$	1.0214	1.0522	1.1070	1.0593

To obtain intuition for the results it is important to look at the necessary condition of  $w_{1h}$  when  $s > 0$ :

$$\frac{\partial L_h(w_{1h}^*, s)}{\partial w_{1h}} \frac{[w_{1h}^* + \beta \Psi_h^{LF}]}{L_h(w_{1h}^*, s)} = 1 + \frac{\beta}{L_h(w_{1h}^*, s)} \frac{\partial L_h(w_{1h}^*, s)}{\partial w_{1h}} \left\{ \left[ \Psi_h^* + \frac{F(\Delta u_{hh}^*)[w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1} \right] - \Psi_h^{LF} \right\} - \frac{\partial \theta_h(w_{1h}^*, s)}{\partial w_{1h}} s. \quad (3.39)$$

Condition (3.39) is similar to (3.22) with the exception that there is now a new term that takes into account the revenue effect of reducing  $w_{1h}$ .<sup>37</sup> Since  $\frac{\partial \theta_h(w_{1h}^*, s)}{\partial w_{1h}}$  has a negative value, we can see that it puts an upward pressure on the optimal wage. The reason for this is simple. By lowering wages which will lead to an increased number of vacancies being opened, it will make the job creation subsidy program more expensive for a given level of  $s$ .

Thus, job creation subsidies in this model are useless unless I consider unrealistic levels of market frictions.

### 3.4 Concluding Comments

I have derived characteristics of an optimal redistributive tax system when unemployment destroys a fraction of a worker's human capital in a model where wages and employment are endogenous. Using a two type model, I find that the optimal tax system under imperfect information distorts the first period wage of both types away from their net-output maximizing levels (*laissez-faire levels*). So in both cases employment is increased compared to the *laissez-faire* outcome which guarantees more redistribution for the unskilled workers and a higher level of human capital in the future. Furthermore, I investigate how other policy tools such as training programs and job creation subsidies impact the optimal tax system. I find that both tools can be used in tandem with the optimal tax system to relax the informational constraint faced by the government but both these tools affect the optimal tax

---

<sup>37</sup>Also since  $s$  does not enter the FOC of the choice of  $\Delta u_{hh}$  and with the functional form assumption on the distribution function used above, we recover the result that  $\Psi_h^* + \frac{F(\Delta u_{hh}^*)[w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1} > \Psi_h^{LF}$ .

system in opposite ways. In the case of the training program, it can add further downward pressure on the unskilled worker's wage in the first period, but the effect is mitigated by the relaxing of the constraint brought about by the training technology. Numerical results show that distortion of the skilled worker's wage is similar in the presence or absence of the training technology. In the case of the job creation subsidies, the policy is useless unless matching frictions are quite high. This can have an upward effect on the skilled worker's wage since having this wage being too low increases the amount of vacancies being opened and thus making the subsidy program much more expensive.

Extending the analysis to more types such as Hungerbühler et al. (2006); Lehmann et al. (2011) may permit one to find other properties of the tax schedule, especially how progressive or regressive it is under different circumstances. In addition, incorporating a participation decision would allow incorporating income-based unemployment benefits which has the potential to modify the second period redistributive tax scheme. Finally, a richer dynamic of human capital accumulation could be considered if the model is extended to more than two periods.

## Chapter 4

# Decision-Making In Poverty, Savings and Redistribution

In recent years, much attention has been given in the optimal taxation literature on models featuring heterogeneity along multiple dimensions. A good motivation for this assumption is that models with heterogeneity solely along the labor productivity dimension fail to generate empirically plausible distributions of wealth while models with heterogeneity of both preferences and earning abilities are able to do so (Krusell and Smith (1998)).

In the optimal tax literature with multidimensional heterogeneity, the distribution of preferences is taken as given. Our paper instead investigates the case where the distribution of those preferences are shaped by available economic resources and socio-economic background. In a first step, individuals inherit time preferences from their parents and these are used to determine their labor allocation.<sup>1</sup> This idea is based on the theoretical and empirical literature on cultural transmission and the dynamics of preferences. This literature tries to explain differences in taste for leisure, education, entrepreneurship and also labor market participation (Bisin and Verdier (2001); Doepke and Zilibotti (2008); Saez-Marti and Zilibotti (2008); Doepke and Zilibotti (December 2012); Dahl et al. (2014)). It also relates to adaptative preferences talked about by Sen (1985) and Fleurbaey (2008), where individuals

---

<sup>1</sup>It can also be viewed as if individuals inherit a discounting factor.

come to identify themselves with preferences that are not truly their own.

The second step relies on recent findings in psychology, behavioral and experimental economics of the link between poverty and the perceived myopic behavior of the poor e.g. Spears (2011) and Shah et al. (2012). In the model of the paper individuals' time preferences (discount factor) will shift based on their disposable income. The higher is the income of an individual – which depends on both labor decisions and productivity – the less myopic his preference will be.<sup>2</sup> Individuals in poorer socio-economic backgrounds thus regret their savings decisions when they retire as they can truly see their own needs. This perception of savings is then transmitted to the individuals's children, who must then make a labor decision in the following period, and so on.

We then conduct an optimal tax exercise and focus on the effects of these perceptions (and scarcity) on the decision to work and save for future consumption. The central question is how an optimal linear tax system should be designed to mitigate these adverse effects of poverty.

Aside from the standard trade-off between equity and efficiency, the behavioral reaction of individuals introduce two additional motives that the optimal tax system must take into consideration. The first motive accounts for the misperception leading to under saving. By changing the marginal value of savings, the government can influence the attractiveness of savings and thus correct for the misallocation. The second effect is related to the impact of time preferences on the choice of labor supply. The planner takes into account the discrepancy between the realized time preference and the inherited time preference of each worker and how it affects the labor supply.

In this model, the labor income tax introduces a new inefficiency: because individuals face a higher tax, their total income decreases. Hence, the tax worsens the ability of future

---

<sup>2</sup>In this paper the extent of myopia is measured as the difference between the endogenous discount factor of individuals and a true discount factor shared by all individuals in the economy. We interpret the endogenous discount factor as the ability or capacity individuals have in foreseeing their retirement needs. In this sense, all individuals prefer having perfect foresight *ex-post*.

generations to anticipate their retirement needs, through cultural transmission, and thus has a dynamic effect.

As a general result of our simulations, we find first that a tax instrument on savings should be used as a subsidy. By doing so, the government influences the trade-off between the (ill)-perceived value of savings and current consumption. The higher is the number of individuals with this problem in the economy, the higher is the subsidy. The funds to pay for this subsidy come from an increase in the taxation on labor and a decrease in the lump-sum transfer.

We present various simulations of a calibrated model based on the U.S. economy that show how these effects are balanced in an optimal tax setup. Tax estimates suggest an increase of the tax rate on labor from 42% to roughly 48% from the case of no behavioral issue to all workers facing issues. The subsidy on savings increases from 0% to about 40% when the number of individuals with behavioral problems is maximal. The lump-sum transfer although positive is decreased by 21% from the no behavioral case.

The rest of this paper is organized as follows. In section 4.0.1, we present an overview of the literature on optimal taxation and heterogeneous preferences. We then present in section 4.1 the setup of the model and, in section 4.2 a theoretical analysis of the tax formulas. In section 4.3, we present some simulation results. A brief conclusion follows.

#### **4.0.1 Literature Review**

The analysis of the paper is related to two strands of literature. The first one is optimal redistributive taxation where individuals have heterogeneous preferences and heterogeneous abilities, and the second one is the taxation and behavioral economics literature.

Heterogeneous preferences raise two issues for redistributive taxation: how much to redistribute and to whom to redistribute? Although both issues are not independent, a set of papers Saez (2002a); Diamond (2003); Judd and Su (2006); Blomquist and Christiansen

(2008); Choné and Laroque (2010) has focused on the efficient way of redistributing given more standard social welfare functions with individual's utility has arguments. Another set of papers Cuff (2000); Boadway et al. (2002); Fleurbaey and Maniquet (2006); Jacquet and Van de gaer (2011) has instead focused more on the impact of differing redistributive tastes of the planner, representing alternative moral judgements on the optimal tax system.

From the above strand of literature, this paper is more closely related to the work of Diamond and Spinnewijn (2011) and Golosov et al. (2013b). In contrast to the present paper, these two papers consider non-uniform taxation of capital. Diamond and Spinnewijn (2011) build a model of four worker types, two possible skill levels and two possible time preferences, who choose labor effort along the extensive margin and face earning dependent labor and saving taxes. Furthermore, they assume that there is a positive correlation between willingness to work and savings. In their model the optimal scheme leaves each skill level at their skill level earnings. To keep the high-skill/low-preferences for saving worker in the high skill job, the planner introduces a positive linear tax on savings for the high earners and a linear subsidy on savings for the low earners. The last part makes the low earning job less attractive to the more impatient high skilled worker.

The paper by Golosov et al. (2013b) considers a model with a continuum of types in a Mirrleesian framework where workers choose labor effort along the intensive margin. Workers are heterogeneous in both time preference and earning ability, but there is a perfect correlation between preferences and ability, which makes it a unidimensional problem. The optimal tax system features a nonlinear tax on savings such that it discourages high earners from mimicking low earners. This is achieved in part by deterring low earners from savings since future consumption is preferred by high earners. This result follows a similar logic as the standard result on commodity taxation where there is complementarity between leisure and a specific good Atkinson and Stiglitz (1976). In the paper by Golosov et al. (2013b) ,

the positive correlation between preferences for savings and skill means that high skill workers will want to save more. If the high skill worker mimics the income of a lower skilled individual, he will nonetheless desire to save more than the skill type he tries to mimic. By taxing savings, the lower income bundle becomes less desirable.

The second strand of literature is behavioral optimal taxation. Behavioral optimal taxation is related to non-welfarist optimal taxation as it can replace the individual's preferences over bundles of goods with the preferences of the planner over the same bundles in the social decision function. As is demonstrated in the review paper by Kanbur et al. (2006), optimal tax formulas in this literature usually share two common terms. One term is a more standard term found in welfarist taxation papers and a novel second term which captures the paternalistic motive for taxation.<sup>3</sup> This last term is the social value of the difference between the individuals and the planner's preferences, and similarly to a pigovian tax, it tries to correct the behavior as to more reflect the social preferences.

Myopic and other types of time-inconsistent preferences have been used in a variety of optimal tax exercises. O'Donoghue and Rabin (2006) demonstrate that when a fraction of the population has myopic preferences towards a "sin good" imposing a tax on such a good can increase welfare. Myopic individuals consume more of this sin good and can thus be taxed, the revenue being used to compensate non-myopic agents. Farhi and Gabaix (2015) consider general tax problems with a variety of behavioral problems. In a close application to ours, they consider the case where richer agents make less behavioral mistakes and look at redistributive linear income taxation. They find that for high redistributive preferences, large behavioral problems lead to more redistribution, whereas for lower redistributive taste large behavioral lead to less redistribution. An exercise closer to ours considers time-inconsistent preferences on pensions and the design of social security, e.g. Diamond (2003); Cremer et al. (2008); Fehr et al. (2008); Bouchard St-Amand and Garon (July 2014). In a review article

---

<sup>3</sup>They also refer to this term as the first-best motive for taxation.

on this strand of literature, Cremer and Pestieau (2011) highlight the interplay between paternalistic and redistributive considerations and come up with two analytical results. The first is that the higher the proportion of myopic individuals in the economy the more social security is desirable. Second, constrained to a linear pension scheme, the higher the proportion of myopic individuals the more the pension system becomes redistributive.

The present article relates to the above literature as it undertakes an optimal tax exercise where workers are heterogeneous in both time preferences and in abilities. Because the model features endogenous preferences that evolve throughout the lifetime of an individual, the objective of the planner is chosen to be paternalistic and thus is linked to the non-welfarist optimal taxation literature. However, in contrast to the papers by Diamond and Spinnewijn (2011) and Golosov et al. (2013b), we investigate a tax structure composed of a linear tax on labor income, a linear tax on savings and a demogrant. These tax instruments are chosen to make our analysis tractable in a complex dynamic environment and thus brings our approach in line with the one of Piketty and Saez (2013a).

## 4.1 Model

### 4.1.1 Agents and Timing

Each generation (indexed by  $t$ ) in the economy is composed of individuals who live for two periods. In the first period, they work, consume and save. In the second period, they retire and consume their savings. Two elements distinguishes individuals of a generation: their productivity  $w$  and their inherited discount factor  $\beta$ . We index by  $i$  individuals with such a given pair of characteristics and their marginal density in generation  $t$  is given by  $d\mu_t(i(w, \beta))$ . When they die, individuals are replaced by an equal number of descendants. Each of them inherit the discount factor of their parents, but productivities are inherited according to a probability distribution  $f(w_t|w_{t-1})$ , with the distribution on  $w_0$  fixed. We assume that  $f$  satisfies some “mixing conditions”, meaning that the successive applications

of  $f(w_t|w_{t-1})$  over generations leads to a unique steady state distribution<sup>4</sup>  $f(w)$ .

The first period of each individual is divided in two sub-periods. In the first sub-period, they must decide how much time they are willing to work (denoted  $L$ ) during their active life, how much consumption they plan to enjoy (denoted  $cp$ ) and how much they plan to save for consumption in the second period (denoted  $sp$ ). Although these values ( $cp, sp$ ) are planned, they will never be realized as individuals will face a shift in their perception of savings in the next sub-period.

Savings are taxed at the rate  $(1 - \tau_s)$ , but accrue at the rate of  $(1 + r)$ , where  $r$  is the (exogenous) interest rate. The consumption a worker can plan to enjoy at retirement is thus given by  $(1 - \tau_s)(1 + r)sp$ . The allocation decision is guided by the following utility function:

$$u(cp) + \beta_{t-1}v((1 - \tau_s)(1 + r)sp) - h(L), \quad (4.1)$$

where  $u(\cdot), v(\cdot)$  are strictly concave and differentiable functions that measure the utility in each period derived from consumption. The function  $h(\cdot)$  is differentiable and convex and gives the disutility of providing an amount of labor  $L$ . The function  $h$  is made separable from  $u$  and  $v$  because it is sufficient, with its convexity, to guarantee a decreasing labor supply with the tax rates. The discount rate of utility in the second period  $\beta_{t-1} \in [0, 1]$  measures the preference for future consumption, or savings, and is inherited from the parents. This parameter will change once the labor decision is made. Income derived from labor is taxed and individuals disposable income in the first period is:

$$z \equiv (1 - \tau_l)wL + a. \quad (4.2)$$

The symbol  $\tau_l$  denotes the linear tax rate and  $a$  denotes a lump sum transfer (or demogrant). Taken together, this means that their savings for consumption in the second period must

---

<sup>4</sup>Specifically, this means that there exists a unique eigenvalue equal to one on the spectral operator  $dF(w_{t+1}) = \int f(w_{t+1}|w_t)dF(w_t)$  and all other eigenvalues have a smaller modulus.

satisfy the following budget constraint:

$$sp \leq z - cp \tag{4.3}$$

Their first sub-period allocation thus maximizes their utility given the previous budget constraint:

$$\begin{aligned} (cp^*, sp^*, L^*) \equiv \arg \max_{cp, sp, L} & u(cp) + \beta_{t-1}v((1 - \tau_s)(1 + r)sp) - h(L) \\ & \text{subject to } sp \leq (z - cp) \end{aligned} \tag{4.4}$$

Throughout the paper, we will append variables with a star ( $cp^*$ , etc.) to distinguish the chosen quantities by individuals from their generic counterpart.

In the second sub-period, individuals start to work and can no longer change their labor allocation. When they receive their income, their valuation of retirement consumption in the second sub-period shifts: the discount factor changes according to a strictly increasing, concave and differentiable function:

$$\beta_t = \Gamma(z^*), \text{ with } \lim_{z \rightarrow \infty} \Gamma(z) = \beta^* \tag{4.5}$$

This functional form captures the idea that individuals with higher income have better resources to ascertain the value of retirement compared to those with lower income.<sup>5</sup> If all individuals had an infinite amount of income, they could perfectly anticipate their enjoyment of retirement and use  $\beta^*$  as a valuation for time preferences. Time preference  $\beta^*$  captures perfect foresight and this level is shared by all individuals. As they have however a finite amount of income  $z^*$ , they can only plan the future according to  $\Gamma(z^*)$ .<sup>6</sup>

---

<sup>5</sup>The resources affect the capacity of individuals to make important economic decisions. See Spears (2011).

<sup>6</sup>Note that the assumptions made on  $\Gamma(z)$  does not rule out the case where  $\beta^*$  could be reached with a lower level of  $z^*$ .

This shift in preferences from  $\beta_{t-1}$  to  $\beta_t$  implies that individuals will modify their planned allocation to find a new allocation of consumption and savings,  $c^*$  and  $s^*$ , in the second sub-period:

$$\begin{aligned} \max_{c,s} \quad & u(c) + \beta_t v((1 - \tau_s)(1 + r)s), \\ \text{subject to} \quad & s \leq z^* - c \end{aligned} \tag{4.6}$$

Hence, the allocation that each individual realizes is given by  $L^*(\tau_l, \tau_s, a)$ ,  $c^*(\tau_l, \tau_s, a)$ ,  $s^*(\tau_l, \tau_s, a)$  and the planned consumption bundle from the first sub-period ( $cp^*$ ,  $sp^*$ ) is never realized.

This setup captures two key ideas that we want to explore. First, individuals can regret their past decision of allocation on labor because their tastes have evolved over time. Second, preferences are the consequence of the wealth background of the parents. Children from families with lower income will inherit a lower  $\beta$  and will thus prefer immediate consumption to savings for retirement and a lower labor supply. They will however regret it in the second sub-period. Individuals from higher income families will also change their mind in the second sub-period but the size of the effect will be smaller through the concavity of  $\Gamma$ . This model thus gives a better ability to anticipate retirement needs to individuals with a higher income.

If one lets the economy evolve over generations, the density of individuals with a given pair  $(w, \beta)$  will move away from the initial distribution and reach a steady state:

$$\mu_t(i(w, \beta)) = \mu_{t+1}(i(w, \beta)) \equiv \mu_{ss}(i(w, \beta)) \quad \forall i. \tag{4.7}$$

The number of people exiting from a pair  $(w, \beta)$  to another pair due to endogenous shifts in preferences and random assignments of productivity is exactly compensated by the entry of individuals from all other pairs, leaving the global distribution unchanged. Although there are individuals constantly entering and exiting a particular category  $(w, \beta)$ , the global distribution of the population remains fixed. The proof of existence of this steady state can

be found in Huggett (1993), by applying the theorem 2 to the functional operator:

$$\mu_{t+1}(i(\tilde{w}, \tilde{\beta})) = \int_j \chi(\Gamma(z_j^*(w, \beta)) = \tilde{\beta}) f(\tilde{w}|w) d\mu_t(j(w, \beta)), \quad (4.8)$$

where  $\chi(\Gamma(z_j^*(w, \beta)) = \tilde{\beta})$  is the indicator function.

#### 4.1.1.1 Behavior and Comparative Statics

##### 4.1.1.1.1 Effect of a change in time preference

Workers inherit time preference  $\beta_{t-1}$ , which influences the labor decision in the first subperiod and by extension the realized consumption choices. So keeping everything else constant, an increase in inherited time preference  $\beta_{t-1}$  results in an increase in labor effort, i.e.  $dL/d\beta_{t-1} > 0$ . This implies that a worker that inherits more patient preferences will also have a higher disposable income. This means in turn that the next generation is less prone to myopia than a generation equally skilled, but with a smaller inherited time preference.

In the second subperiod, an increase in  $\beta_t$  keeping disposable income constant leads to an increase in savings, i.e.  $ds/d\beta_t > 0$  and a reduction in first period consumption  $dc/d\beta_t < 0$ . It is then important to understand how time preference evolves through time. From (4.5) and the definition of disposable income we have:

$$\beta_t = \Gamma((1 - \tau_L)wL(w, \tau_L, \tau_s, a, R, \beta_{t-1}) + a) = \Gamma(w, \tau_L, \tau_s, a, R, \beta_{t-1}), \quad (4.9)$$

which is a simple autonomous difference equation<sup>7</sup>. From this equation, it is possible to determine that the phase line is upward sloping since:

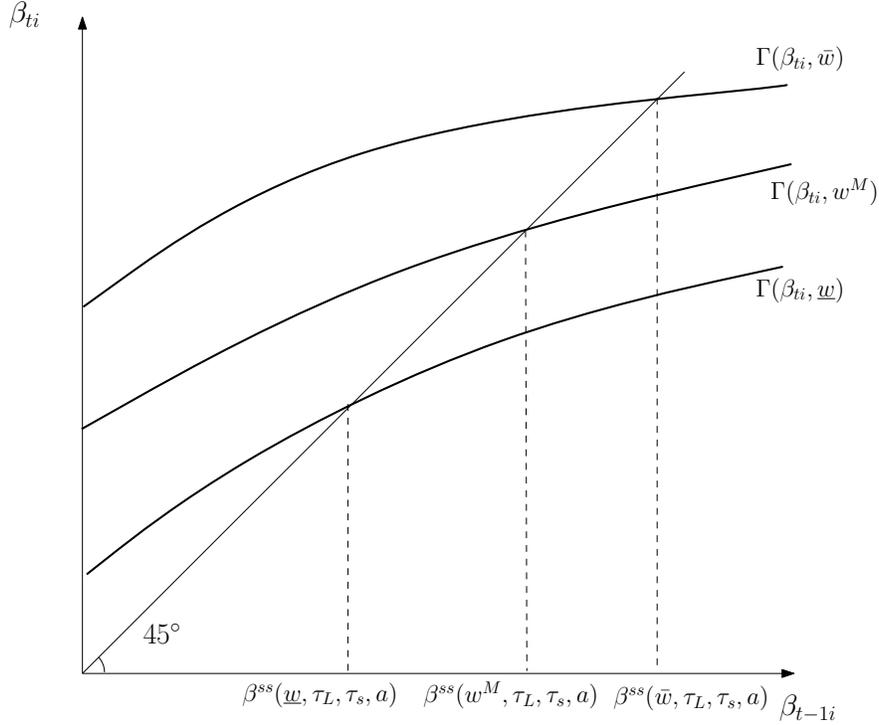
$$\frac{\partial \beta_t}{\partial \beta_{t-1}} = \Gamma'(z_t)(1 - \tau_L)w \frac{\partial L}{\partial \beta_{t-1}} > 0. \quad (4.10)$$

---

<sup>7</sup>See Appendix A.3.1 for all the formal derivations.

When  $(1 - \tau_L)$  is positive, the strict concavity of  $\Gamma$ , combined with (4.5), implies the existence of a fixed point  $\Gamma(z_t(\beta)) = \beta$ .<sup>8</sup> An example of a phase diagram for different levels of productivity can be seen in Figure 4.1.

Figure 4.1: Phase diagram of time preference for different productivity levels



It is easier to understand the dynamics by explaining first what happens when the productivity is fixed along the same dynasty. In that case, figure 4.1 shows how the properties of  $\Gamma$  leads to a fixed point. When a dynasty keeps the same level of productivity  $w$  for several generations, it reaches a long-run level of time preference  $\beta^{ss}(w)$ . Indeed, when a dynasty has a value  $\beta_{t-1} < \beta^{ss}(w)$  in generation  $t$ , it will see an increase over generations, as work increases  $\beta_t > \beta_{t-1}$ . Eventually, the dynasty reaches the fixed point. The converse is also true. If a dynasty starts with  $\beta_{t-1} > \beta^{ss}(w)$ , then the subsequent generations receive a lower

<sup>8</sup>The following condition

$$\left| \Gamma'(w_i, \tau_L, \tau_s, a, R, \beta_i^{ss})(1 - \tau_L)w_i \frac{\partial L(\beta_i^{ss})}{\partial \beta_{t-1}} \right| < 1$$

is also satisfied.

time preference, until  $\beta^{ss}(w)$  is reached.

Of course, productivity levels changes through time in our model. Although this introduces additional dynamics, what is important remains the position of the inherited time preference *vis-à-vis* the steady state preference of the worker's if it kept the same skill level.

For a given  $\beta_{t-1}$ , there are three possible cases:

- Case 1:  $\beta_{t-1} < \beta^{ss}(w_i)$ ,
- Case 2:  $\beta_{t-1} > \beta^{ss}(w_i)$ ,
- Case 3:  $\beta_{t-1} = \beta^{ss}(w_i)$ .

Case 3 represents the fixed point of figure 4.1. In case 1, a dynasty see an increase in  $\beta$  and thus reduce consumption ( $cp^* > c^*$ ) and increase savings ( $sp^* < s^*$ ). In the second case, the opposite happens. Time preferences decrease, leading to an upward revision of consumption ( $cp^* < c^*$ ) and a downward decrease in savings ( $sp^* > s^*$ ). The first two cases are important since the changes from  $\beta_{t-1}$  to  $\beta_t$  implies a difference between the planned marginal rate of substitution between consumption and labor and the realized marginal rate of substitution. Individuals facing case 1 end up wishing they had worked more. On the contrary, workers of case 2 end up wishing they had worked less.

#### 4.1.1.1.2 Effect of a change in disposable income

Disposable income affects the ability of individuals to anticipate the future. Hence, an increase in disposable income does more than just increase the amount of resources to allocate, but changes the behavior of the individual in the second sub-period. To highlight the differences between a model with fixed time preferences and this model, we present below the reaction of an individual with fixed preferences to disposable income.<sup>9</sup> If we denote his

---

<sup>9</sup>Note that this behavior is also the one corresponding to the “naïve” first subperiod worker.

first period consumption  $\hat{c}$  and savings  $\hat{s}$ , we find:

$$\frac{d\hat{c}}{dz} = \frac{\beta\tilde{R}^2 v''(\tilde{R}\hat{s})}{u''(\hat{c}) + \beta\tilde{R}^2 v''(\tilde{R}\hat{s})} > 0 \quad \text{and} \quad \frac{d\hat{s}}{dz} = \frac{u''(\hat{c})}{u''(\hat{c}) + \beta\tilde{R}^2 v''(\tilde{R}\hat{s})} > 0, \quad (4.11)$$

where  $\tilde{R} = (1 - \tau_s)(1 + r)$ . As it can be seen, both first period consumption and savings increase with disposable income. Now consider the effect on savings of a change in disposable income in our model:

$$\frac{ds}{dz} = \frac{u''(c)}{u''(c) + \beta_i \tilde{R}^2 v''(\tilde{R}s)} + \underbrace{\frac{-\Gamma'(z)\tilde{R}v'(\tilde{R}s)}{u''(c) + \beta_i \tilde{R}^2 v''(\tilde{R}s)}}_{\text{Endogenous Preference Effect}} > 0. \quad (4.12)$$

The increase in disposable income also increases the value of the second period consumption, this added weight then translates into added demand for savings.

When time preferences are endogenous, the first period consumption is no longer guaranteed to be a normal good. In other words, an increase in disposable income does not necessarily translate into an increase in first period consumption. From the budget constraint, we know that  $dc/dz + ds/dz = 1$ , and hence for first period consumption to be a normal good, it is required that  $ds/dz < 1$ . This is true solely when the following condition holds:

$$-\frac{v''(\tilde{R}s)}{v'(\tilde{R}s)} > \frac{\Gamma'(z)}{\Gamma(z)} \frac{1}{\tilde{R}}, \quad (4.13)$$

that is when the coefficient of absolute risk aversion, a measure of the curvature of the second period utility of consumption, is greater than the ratio of the marginal increase in time preference from an increase in disposable income on the value of time preference weighted by the net gross interest rate. This last term measures the marginal increase in the desire for retirement consumption, a measure of the shift in preferences.

### 4.1.1.1.3 Effect of taxes

How the allocation of consumption, savings and labor reacts to taxation will depend on the functional form employed. By making labor disutility,  $h$ , separable from consumption, the labor supply is always decreasing with the labor tax ( $\tau_l$ ). From the fact that

$$\frac{dz^*}{d\tau_l} = -w_i L^* + (1 - \tau_L) w_i \frac{dL^*}{d\tau_L} < 0, \quad (4.14)$$

planned consumption ( $cp^*$ ) and planned savings ( $sp^*$ ) also decrease with an increase in labor income taxation.

The effect of labor taxation on actual consumption ( $c^*$ ) and savings ( $s^*$ ) is determined by the impact it has on disposable income. This effect is, however, ambiguous as it must account for the trade-off between the change in time preference and the curvature of consumption preferences (as explained in (4.13)).

The sign of the reaction of labor supply to the savings tax depends on:

$$\text{sign} \left[ \frac{dL^*}{d\tau_s} \right] = \text{sign} \left[ - \left( 1 + \tilde{R} sp^* \frac{v''}{v'} \right) \right].^{10} \quad (4.15)$$

If it is negative, the marginal disutility of labor remains higher than the disutility of decreasing savings (and labor supply decreases). Notice that if a CRRA form is used, as in our simulations, this turns out to be always true.

The variation of consumption with respect to the savings tax is also ambiguous, as it depends on the trade-off explained in (4.13).

---

<sup>10</sup>For details of the derivation see Appendix A.3.1.1.1.

### 4.1.2 The Government

The government seeks to maximize the steady-state welfare of all individuals in the economy. It does so by considering the steady state distribution of generations. Its objective function is given by:

$$W(\tau_s, \tau_l, a) \equiv \left( \int_i (u(c_i^*) + \beta^* v((1+r)(1-\tau_s)s_i^*) - h(L_i^*))^{1-\rho} d\mu_{ss}(i) \right)^{1/(1-\rho)}. \quad (4.16)$$

The parameter  $\rho$  is a measure of inequality aversion. If the parameter  $\rho$  is equal to zero, the government maximizes utility in a Benthamite sense while if  $\rho$  goes to infinity, it maximizes utility in a Rawlsian sense. Values of  $\rho$  in between can be interpreted as particular degrees of aversion to inequalities.

We point out that this function depends on the sum of products of the functions  $\Gamma(z)v(\tilde{R}(z-c))$ . For this welfare function to remain concave in  $\tau_l$ , it must be that the joint product of these functions remains concave. A necessary and sufficient condition on  $\Gamma(z)v(\tilde{R}(z-c))$  is the Gonzi condition.<sup>11</sup> This condition imposes some restrictions on the functional form of  $\Gamma$ , given a function  $v$ . We assume this condition throughout.

Notice that the government is paternalistic and does not use the discount factor of individuals, but rather uses  $\beta^*$ , the value that individuals wish they had *ex-post*, meaning that they would have preferred perfect foresight ( $\beta^*$ ) in planning their retirement instead of their own inherited value.<sup>12</sup> Hence, when the government maximizes social welfare, it designs a tax system such that individuals who imperfectly planned their retirement savings will eventually adopt allocations that approaches their desired behavior. When choosing tax rates, it must also respect budget balance to finance the demogrant  $a$ . The sum of revenues

<sup>11</sup>A pair of functions  $f(x), g(x)$  satisfies the Gonzi condition if  $(f(x) - f(y))(g(x) - g(y)) \geq 0 \forall x, y$ .

<sup>12</sup>One way to tie this assumption with the previous behavioral economics literature, is to interpret  $\beta^*$  as the realized value of the weight on second period consumption. More clearly,  $\beta^*v(\tilde{R}s)$  will be the cardinal utility that a worker will derive from consumption in the second period  $\tilde{R}s$ , and  $\beta_t$  will be the imperfect time preference with which he will make his economic decisions. Hence,  $\beta^*$  can be seen as the planner taking the *ex-post* welfare of workers into consideration.

in equilibrium must therefore equal the cost of the demogrant:

$$\begin{aligned}
a &= \int_i [\tau_s s_i^* + \tau_l w_i L_i^*] d\mu_{ss}(i(w, \beta)), \\
&= \underbrace{\tau_s \int_i s_i^* d\mu_{ss}(i(w, \beta))}_{\equiv S^*} + \underbrace{\tau_l \int_i w_i L_i^* d\mu_{ss}(i(w, \beta))}_{\equiv Y^*}, \\
&= \tau_s S^* + \tau_l Y^*
\end{aligned} \tag{4.17}$$

Taken altogether, this defines a program that the government seeks to solve:

$$\begin{aligned}
W^* &\equiv \max_{\tau_l, \tau_s} W(\tau_s, \tau_l, a) \\
&\text{subject to: } a = \tau_s S^* + \tau_l Y^*
\end{aligned} \tag{4.18}$$

This program defines taxes  $(\tau_l^*, \tau_s^*)$ . Their characterization is discussed in the next section.

## 4.2 Optimal Taxes

Table 4.1: Behavioral Elasticities

Elasticity of individual savings with respect to $\tau_l$ and $\tau_s$	$\epsilon_{\tau_l}^{s_i} \equiv \frac{\partial s_i}{\partial(1-\tau_l)} \frac{(1-\tau_l)}{s_i}$ ,	$\epsilon_{\tau_s}^{s_i} \equiv \frac{\partial s_i}{\partial(1-\tau_s)} \frac{(1-\tau_s)}{s_i}$
Elasticity of individual income with respect to $\tau_l$ and $\tau_s$	$\epsilon_{\tau_l}^{y_i} \equiv \frac{\partial y_i}{\partial(1-\tau_l)} \frac{(1-\tau_l)}{y_i}$ ,	$\epsilon_{\tau_s}^{y_i} \equiv \frac{\partial y_i}{\partial(1-\tau_s)} \frac{(1-\tau_s)}{y_i}$
Elasticity of aggregate savings with respect to $\tau_l$ and $\tau_s$	$\epsilon_{\tau_l}^S \equiv \frac{\partial S}{\partial(1-\tau_l)} \frac{(1-\tau_l)}{S}$ ,	$\epsilon_{\tau_s}^S \equiv \frac{\partial S}{\partial(1-\tau_s)} \frac{(1-\tau_s)}{S}$
Elasticity of aggregate income with respect to $\tau_l$ and $\tau_s$	$\epsilon_{\tau_l}^Y \equiv \frac{\partial Y}{\partial(1-\tau_l)} \frac{(1-\tau_l)}{Y}$ ,	$\epsilon_{\tau_s}^Y \equiv \frac{\partial Y}{\partial(1-\tau_s)} \frac{(1-\tau_s)}{Y}$

We now turn to the analysis of the optimal tax rates.<sup>13</sup> First, we introduce both the behavioral elasticities, found in Table 4.1 and a shorthand for the “social marginal paternalistic

<sup>13</sup>We derive in appendix A.3.1.2 the expressions for tax rates.

welfare weights” that the government uses to weight each group of individuals  $i$ :

$$g_i \equiv \frac{\left[ u(c_i^*) + \beta^* v(\tilde{R}s_i^*) - h(L_i^*) \right]^{-\rho} u'(c_i^*)}{\int \left[ u(c_j^*) + \beta^* v(\tilde{R}s_j^*) - h(L_j^*) \right]^{-\rho} u'(c_j^*) d\mu(j)}. \quad (4.19)$$

As in Piketty and Saez (2013a), the weights are normalized to sum to 1 and capture the value to the planner of distributing \$1 to an individual of type  $i$  relative to distributing equally the same dollar \$1 to everyone in the economy. With these definitions, an expression for the tax-rate on labor can be found:

$$\frac{\tau_l}{1 - \tau_l} = \frac{\overbrace{\int_i g_i [Y - y_i^*] d\mu}^{\text{Equity}} - \overbrace{\int_i g_i \left[ \frac{\beta^* - \beta_i}{\beta_i} \frac{\partial s_i^*}{\partial(1 - \tau_l)} + \frac{u'(c_i^*) - u'(cp_i^*)}{u'(c_i^*)} y_i^* \epsilon_{\tau_l}^{y_i^*} \right] d\mu}^{\text{Behavioral Motives}} - \overbrace{\left[ \tau_s \frac{\partial S}{\partial(1 - \tau_l)} \right]}^{\text{Efficiency}}}{\underbrace{\epsilon_{\tau_l}^Y Y}_{\text{Efficiency}}}. \quad (4.20)$$

It is useful to begin the description of the formula by assuming there is no myopic behavior and worker’s have perfect foresight. In this case, only the first, the last terms of the numerator and the denominator remain. These remaining terms then show the classical trade-off between equity (the first term) and efficiency (the denominator) and a second-best effect of labor income tax on the other taxed base i.e. savings.<sup>14</sup> Increasing taxation pushes the equity term towards zero (perfect equality), which is what measures the “equity term”. However, increasing taxation reduces the incentives to work and, in our model, to save and thus, the size of the taxed bases. These effects are captured by the term at the denominator (which is expressed in terms of the aggregate elasticity) and the last term on the numerator. All things being equal the greater the aggregate labor elasticity the lower the optimal linear

<sup>14</sup>The tax function would be similar for the case where  $\Gamma(z)$  reaches  $\beta^*$  faster. The equity and efficiency terms would be identical, although the global elasticity would need to take into account the behavioral effect. The big difference would be the measure of workers where we need to take into account the behavioral motives. And this measure would be a function of disposable income.

labor income tax should be.<sup>15</sup> In the case of the last term of the numerator, it captures the effects on its tax revenues coming from saving taxation from increasing the labor income tax.

When poorer individuals have difficulties to anticipate the value of their retirement, it introduces behavioral effects through the two undiscussed terms. The impact of the tax rate on savings is amplified by the difference between perfect foresight ( $\beta^*$ ) and their current perception of retirement ( $\beta_i$ ). Although taxation reduces available income, it also reduce the perceived value of future savings, which is exactly what the tax system is trying to compensate. Hence, increasing taxation has also a dynamic effect that goes against an increase in savings. This effect becomes more acute when individuals weight their retirement period less (e.g.: when  $\beta^* - \beta_i$  is high) which are the poorer workers in the economy. Due to the redistributive preferences of the planner, the negative effect of labor taxation on savings of the poor bears more weight, as  $g_i$  is decreasing in income.

The other term accounts for the impact of a decrease in labor effort due to labor income taxation measured in terms of difference between planned and realized marginal rates of utility. Recall that there is a wedge between the planned marginal rate of substitution between consumption and labor effort and the realized marginal rate of substitution between consumption and labor effort. Hence, workers will regret their choice of labor by either finding that they have worked too little or too much. By increasing labor taxation, the planner discourages labor effort for all workers which will help those who worked too much but exacerbate the problem of the workers that worked too little.

---

<sup>15</sup>If the sum of the numerator is positive.

The optimal tax rate on savings can be described in similar terms:

$$\tau_s^* = \frac{\overbrace{\int_i g_i \left( S - \frac{\beta^*}{\beta_i} s_i^* \right) d\mu}^{\text{Equity+}} - \overbrace{\int_i g_i \left[ \frac{\beta^* - \beta_i}{\beta_i} s_i^* \epsilon_{\tau_s}^* + \frac{u'_i(c_i^*) - u'_i(cp_i^*)}{u'_i(c_i^*)} (1 - \pi) y_i^* \epsilon_{\tau_s}^* \right] d\mu}^{\text{Behavioral Motives}} - \overbrace{\pi Y \epsilon_{\tau_s}^*}^{\text{Efficiency}}}{\underbrace{(1 + \epsilon_{\tau_s}^S) S}_{\text{Efficiency}}}.$$

(4.21)

The global interpretation is similar in spirit. The first term at the numerator measures equity while the denominator measures efficiency through the elasticity of the tax base, provided that taxing savings reduces savings. The difference with the tax on savings is that the equity term does not exclusively capture equity concerns, which is why we call this term ‘Equity plus’ as the term also captures behavioral issues. In this case, equity and behavioral issues push the optimal tax rate in two different directions. The ‘pure’ equity motive would push for higher saving taxes if the high income individual save more than the average individual. However, the behavioral motive in the ‘Equity plus’ term pushes for lower savings’ tax as  $\beta^*/\beta_i$  is always greater than one. The starkest scenario is when there is no redistribution ( $g_i = 1$ ). In this case, the whole term is negative and pushes for a subsidy. Other terms have a similar interpretation as in the tax rate on labor but coming from the effect on realized choices of savings taxation.

### 4.3 Simulations

To further highlight the characteristics of the optimal tax system, we provide numerical simulations of the model and discuss the results found. All simulations, but one, are done with identical utility of consumption functions in both periods ( $u(c) = v(c) = \ln c$ ). We used

a disutility of labor of the constant relative risk type:

$$h(L) \equiv \phi_1 \frac{L^{\phi_2}}{\phi_2}. \quad (4.22)$$

We model the ability to foresee correctly the future through a square root function:

$$\Gamma(z) \equiv \begin{cases} \frac{\beta^*}{\phi_3} (\phi_3^2 - (\phi_3 - z)^2)^{1/2} & \text{if } z \leq \phi_3, \\ \beta^* & \text{if } z > \phi_3 \end{cases} \quad (4.23)$$

For values of  $z \in [0, \phi_3]$ , the ability to foresee the retirement needs increases with income. In this interval,  $\Gamma$  is increasing and concave. However, at  $z = \phi_3$ , individuals reach the perfect foresight threshold ( $\Gamma(\phi_3) = \beta^*$ ). Above this value, agents remain with perfect foresight and an increase in income has no impact. The parameter  $\phi_3$  can thus be interpreted as the required income to reach perfect foresight. This functional form was chosen because it suits the theoretical properties we described in the previous sections of the paper, but also because  $\Gamma$  and  $v$  satisfy jointly the Gonzi condition. The default numerical values used for other parameters of the simulation can be found in Table 4.2. The numerical value of wages and transition probabilities are taken from Bakis et al. (2012).

Table 4.2: Numerical Values Used For Default Simulations

Parameter	Value																
$r$	2.0715																
$\rho$	0.0																
$\beta_i$	{0.300, 0.314, 0.328, ..., 1.0}																
$\mu_0(\beta_i)$	$\frac{1}{50} \forall i$																
$w_1, \dots, w_4$	{9.7, 14.9, 20.2, 33.0}																
$F(w_j w_i)$	<table border="1" style="margin-left: auto; margin-right: auto;"> <tr> <td>0.34</td> <td>0.33</td> <td>0.17</td> <td>0.16</td> </tr> <tr> <td>0.18</td> <td>0.31</td> <td>0.34</td> <td>0.17</td> </tr> <tr> <td>0.12</td> <td>0.29</td> <td>0.31</td> <td>0.28</td> </tr> <tr> <td>0.07</td> <td>0.16</td> <td>0.24</td> <td>0.53</td> </tr> </table>	0.34	0.33	0.17	0.16	0.18	0.31	0.34	0.17	0.12	0.29	0.31	0.28	0.07	0.16	0.24	0.53
0.34	0.33	0.17	0.16														
0.18	0.31	0.34	0.17														
0.12	0.29	0.31	0.28														
0.07	0.16	0.24	0.53														

We fitted the free parameters  $\phi_1, \phi_2$  and  $\phi_3$  so that model minimizes the sum of squares between the mean and variance of hours worked in the 2007 PSID and the first two moments (mean and variance) of the distribution of discount factors obtained using the same data and procedure as Golosov et al. (2013b). In order to do so, we followed the methodology of Jacquet et al. (2013) and chose a tax system that constitutes a reduced form of the tax system in the United States. We then performed the fitting by minimizing the sum of squares between the model’s statistics and those found in the data. The parameters found are summarized in Table 4.3.

Table 4.3: Parameters Reproducing Observed US Data

Parameter	Fitted Value
$\phi_1$	456.11
$\phi_2$	6.70
$\phi_3$	26.4

Source: Authors’ calculations. Numbers rounded.

Using the model, we first performed an experiment where we vary the quantity of agents whose discount factor evolves over generations. For the purpose of discussion, we call them learners. Conceptually, the population is divided in to two groups: a fraction  $\pi$  of learners, who have a pair  $(w_i, \beta_i)$  following the model described in the previous section and the remaining fraction  $1 - \pi$  who have perfect knowledge ( $\beta_i = \beta^*$ ) and random productivity assignment.<sup>16</sup> Each fraction has an identical initial distribution  $\mu(\beta, w)$  which is a discretized uniform distribution. Given this initial setup, learners evolve over time according to the function  $\Gamma(z)$ , their labor decision and the transition probabilities on  $w$ . Individuals with perfect knowledge remain with perfect knowledge ( $\beta^*$ ). Thus, by varying  $\pi$ , we vary the fraction of individuals who learn in the economy. For this simulation, the planner’s aversion

<sup>16</sup>This does not alter the social welfare function as the non-learners already have the same time preference as the government.

to inequality parameter is  $\rho = 0$ . We report in Table 4.4 the optimal taxes for various values of  $\pi$ .

Table 4.4: Optimal Taxes For Various Proportions of Learning Agents (others at  $\beta^*$ )

$\pi$	$\tau_l^*$	$\tau_s^*$	$a^*$
0.0	0.420004	0.0	3.6524
0.2	0.426241	-0.05482	3.4738
0.4	0.429872	-0.10546	3.3134
0.6	0.44045	-0.17414	3.1699
0.8	0.450346	-0.25612	3.0119
1.0	0.476541	-0.40144	2.8591

Source: Authors' calculations. Numbers rounded.

The first thing to note in this table is the general pattern of taxation: the tax on labor is positive, between 42% and 48% of income and the tax on savings is negative, acting therefore as a subsidy. As the tax rate on savings is the only instrument that can influence the perception of the second period, it is used to counteract the worker's myopia. As the fraction of individuals with small discounting factors grows, the subsidy on savings increases the attractiveness of savings to counteract the increasing myopia in the economy. Likewise, a high tax rate on income allows for a positive demogrant, which acts also as an equalizer for utility in the first period. As the fraction of learners increase the demogrant is reduced as more and more of the resources raised by income taxation are diverted to the savings subsidy.

In this situation, when the fraction of learners is zero, there are no paternalistic motives to tax savings since all time preferences are identical across workers. In this case, the only motive for taxation is the inequalities generated by the differences in productivities ( $w_i$ ). Because labor and savings are independent, there are no motives to tax savings, as shown by Corlett and Hague (1953). As the number of learners becomes non-zero, this no longer

holds, and both paternalistic and redistributive motives become important. As the number of individuals who learn grows, the more the planner wishes to subsidize savings, even if preference for savings and income becomes correlated. Thus, the labor tax rate increases as more revenue is required, in excess of redistributive motive, to fund the increasing subsidies dedicated to the growing number of workers with smaller time preferences.

To tease out the different effects shaping the results obtained in the above simulation, we propose three other simulations trying to shut down the different motives of the planner so as to quantify the magnitude of each effect in our model. The first of these simulation alters the model by giving a single productivity to each dynasty (or family) that will stay unchanged in time. The distribution of these productivities matches the long run distribution of productivities in the above simulation. Under such a set up each dynasty reaches the unique long run steady-state value of time preference linked to their dynasty's productivity level. In this circumstance, there is no difference between the planned bundle of consumption and saving and the realized one. Hence there is no longer any difference in the marginal rate of substitution between labor and consumption and thus no longer any motive for the planner to correct the issue. We are then left with the redistributive motive and the paternalistic motive towards savings. Again, we vary the quantity of learners and non-learners and report our results in Table 4.5.

As can be seen, the results between this simulation and the full model 4.4 are very close. In the single productivity level for each dynasty we can observe slightly higher labor income tax rates. This result might suggest that in the full model some workers regret their labor decision and thus, the planner wishes to distort labor less. The level of the subsidy is fairly close, but as the labor income tax is slightly higher it helps finance a marginally greater level of subsidies. Thus in this model, taxing labor to redistribute more is less in conflict with the planner's motive to correct the behavioral issue.

We now remove the redistributive motive entirely and isolate the impact of endogenous

Table 4.5: Optimal Taxes When Dynasties Keep the Same Productivities

$\pi$	$\tau_l^*$	$\tau_s^*$	$a^*$
0.0	0.420013	0.0	3.6596
0.2	0.427862	-0.056246	3.4823
0.4	0.427602	-0.098069	3.3237
0.6	0.442633	-0.170981	3.2003
0.8	0.453614	-0.258358	3.0353
1.0	0.480382	-0.402601	2.8911

Source: Authors' calculations. Numbers rounded.

myopia on the tax system. To do this, we assume linearity of utility in first period consumption, i.e.  $u(c) = c$ , and  $\rho = 0$ .<sup>17</sup> In this simulation the planner only seek to raise revenue to correct the myopic decision making of workers, which even under no redistributive motive, will put more weight on low skill worker's problem. We report our results in Table 4.6.

Table 4.6: Optimal Taxes For Quasi-linear Preferences

$\pi$	$\tau_l^*$	$\tau_s^*$	$a^*$
0.0	0.0	0.0	0.0
0.2	0.0	-0.039738	-0.0380
0.4	-0.0	-0.083281	-0.0768
0.6	-0.000197	-0.131052	-0.1182
0.8	0.000446	-0.171621	-0.1393
0.99	0.001113	-0.240076	-0.1796

Source: Authors' calculations. Numbers rounded.

The case without any learners shows that there are no motives for taxation and thus both instruments are optimally set at 0. As we increase the number of learners in the economy, we can see that subsidies to savings increase as well. However, the planner does

<sup>17</sup>In this case, labor decisions are just a function of the worker's after-tax wage rate.

not use labor income taxation to raise revenues as it distorts labor and saving choices, the last through  $\Gamma(z)$ . In fact, labor decisions are only a function of the after-tax wage, so the worker will never regret his labor choice. In this situation the government raises revenue using the demogrant which becomes negative, and thus acts as a lump sum tax which does not influence the labor effort decision. However, it does have a direct effect on  $\Gamma(z)$ . The intuition behind this result is that the planner wishes to subsidize savings and correct the behavioral issue and this at the smallest behavioral cost.

Finally we investigate how aversion to inequality influences the optimal tax system. The model is the same used for the results of 4.4 but now we only look at the case where all workers are learners. In this we vary  $\rho$  from 0 to 1 and observe how tax rates evolve. These results are reporter in Table 4.7.

Table 4.7: Optimal Taxes For Various Degree of Aversion to Inequality ( $\rho$ )

$\rho$	$\tau_l^*$	$\tau_s^*$	$a^*$
0.0	0.471818	-0.353969	2.9384
0.2	0.49189	-0.375989	3.0398
0.4	0.509934	-0.388706	3.1474
0.6	0.524553	-0.397873	3.2364
0.8	0.538128	-0.405995	3.3189
0.95	0.538516	-0.406557	3.3204

Source: Authors' calculations. Numbers rounded.

When the inequality aversion parameter increases, the weight  $g_i$  on less productive individuals becomes more important. This results in the redistributive motive taking a more prominent role but also increases the impact of the distributional issue at the heart of the behavioral problem. As the planner's inequality aversion grows the labor income tax rate, the saving subsidies and the demogrant increases. The labor income tax rate rises as it is

used more and more to finance the demogrant and equalize after-tax revenues. But the increase in the labor income tax rate does more than just finance the demogrant. An increase in the redistributive motive would suggest that savings should be taxed more, or subsidized less, which would come from the fact that the higher income workers save more. However, the government will also put more weight on the behavioral issues of the poorest individuals which would suggest that savings should be even more subsidized. In the end, the weight given to the behavioral problems of the poor has more influence than the potential increase in transfers that would of been made possible by taxing savings which would of hit the higher income workers more. Hence the increase in the labor income tax rate also serves to fund the increase saving subsidies.

#### **4.4 Conclusion**

We studied an optimal taxation scheme when individuals who inherit their time preferences through the socio-economic background of their families. When they make their own labor supply decisions, their perception of savings shifts and they thus regret their initial decisions on labor supply. When this is so, the optimal design of a tax scheme changes to account for two additional behavioral effects. First, the ill-perceived value of savings is corrected through a subsidy on savings. When savings are subsidized (a “negative tax”), individuals are given incentives to save more and thus correct for the incorrect perception of savings inherited from their parents and/or low disposable income. In addition, labor taxation lowers income which both influences willingness to save but also lowers the importance of savings for the individual and its future descendants. Second, the shift in time preferences during an individual’s lifetime creates regret towards the labor decision of the individual. Both optimal tax instrument take into consideration this last effect and tries to diminish it.

We performed numerical simulations on a calibrated model of the American economy and our results suggests the first effect dominates the second one. A subsidy in savings, financed

through an increase in labor taxation can be used to increase general welfare. These results suggest that a policy that subsidize savings has some grounds if individuals regret their past savings decisions.

## Chapter 5

### Conclusion

This thesis has explored the design of optimal tax policies when those same policies have long lasting impacts on individuals and their family. In the first chapter, I investigated the characteristics of the optimal tax system where individual's wages are a function of accumulated work experience and a stochastic shock in a two-period dynamic Mirrleesian framework. This shock can be interpreted as a random event happening during the individual's life such as ability, health issues, labor market mismatch or plain luck.

Using numerical simulations based on U.S. data, I find that inexperienced workers should expect to face higher marginal labor income tax rates when they are older. However, whether the average marginal labor income tax rate when inexperienced is lower or greater than when older depends crucially on whether the accumulated work experience and stochastic shock are substitutes or complements. To be more precise whenever the elasticity of complementarity is below one the age profile of the average marginal labor income tax rate is increasing and is decreasing if the elasticity is above one. The intuition for this result is that as the accumulated work experience and the stochastic shock become more complementary the marginal benefit of acquiring work experience increases substantially for the low skilled young individuals. Because work experience becomes very valuable for these workers, they are much more willing to work and their labor supply becomes much less elastic. This leads the planner to increase taxation on those young workers.

I show that the optimal marginal income tax rate formula when young is composed of the usual terms found in the Mirrleesian literature such as the terms capturing the redistributive motive and the efficiency motive as well as three other terms resulting from incorporating work experience considerations. The first captures a second-best motive which results from the government taxing labor income in the future period of life and that taxing labor of the young workers affects their ability to generate labor income. The second is a social insurance motive which may encourage work experience in the case where work experience offers some insurance to worker whereas it is discouraged when it increases the exposure to risk. This depends again on the complementarity between the accumulated work experience and the stochastic shock. Finally, the last term captures an incentive motive aimed at correcting the disincentive effect brought upon the insuring role of the tax system in the second period of **life** on the accumulation of work experience.

The optimal marginal income tax formula for the experienced workers is also similar to the Mirrleesian optimal tax literature with a slight difference. This difference is that the optimal tax formula is multiplied by the wage elasticity with respect to the stochastic shock. This turns out to have a profound effect on the tax rates of high income earners. By giving more structure to the wage function and to the stochastic process, I show that when the parameter controlling the elasticity of complementarity is above one, the optimal marginal labor income tax rates is pushed to zero much more rapidly.

The model of the first essay assumes that work experience levels are observable through the accumulation mechanism. Dropping this assumption by considering a work experience accumulation mechanism that takes into account both the opportunities of the worker and the labor effort which would be unobservable to the government would raise technical issues. By making work experience unobservable the *common knowledge of preferences* is lost and the recursive formulation of the government's problem used above is no longer valid to find the optimal tax system.

In the second essay, I characterize features of the optimal income tax system in non-Walrasian labor markets where involuntary unemployment causes productivity scars in the future. The analyze is conducted in a two-period and two-skill type Mirrleesian framework where unemployment is a consequence of the search and matching frictions of the labor market. In such a model the income tax has an effect on wages of workers and thus on employment. I find that the optimal tax system will seek to distort the wages of skilled and unskilled young workers alike below their efficient levels. The government seeks to do this for two very different reasons and it is able to do this since marginal income tax rates affect the ability of the employer to transfer utility to the worker.

Two motive pushes the government to distort the wage of the skilled young workers below the efficient level. The first is that the government will redistribute from the skilled to the unskilled in the second period of life which results in applying an upward pressure on the skilled wages in the first period of life. Without any corrections the skilled wage would be above its efficient value and there would be more unemployment in the skilled sector than in the *laissez-faire* outcome. The second motive is a second-best motive akin to the one found in the first essay. Since the government is distorting the second period labor market, it then has a fiscal benefit of having more skilled workers in the second period. Put simply, having more skilled workers raises more revenue, the tax burden can be shared over more workers, and finally it reduces the number of workers the planner must redistribute to.

The purpose of distorting the wage of the unskilled young workers away from its efficient level is redistributive. Although redistribution to the young unskilled workers is also done through direct transfers, by distorting the labor market and depressing the equilibrium before-tax wage the government creates more employment than in the *laissez-faire*. By increasing the demand for labor the government increases the likelihood of the young unskilled individuals to find employment. Alternative labor policies such as a training program or a job creating subsidy can help reduce the need to distort the labor markets. Results from

numerical simulations imply that for plausible parameter values of the considered model only the training program appear to do so whereas job creating subsidies are redundant except in cases where the optimal allocation calls for the unskilled wage to be driven to zero.

The model featured in the chapter 3 is very stylized to obtain clean and clear intuitions for the results obtained in the essay. Considering the design of the optimal insurance in the presence of an optimal redistributive tax system in a dynamic model would yield interesting results especially in the presence of unemployment scarring. Tackling such an issue is however fraught with technical difficulties as the wage-setting mechanism used in the essay does not lend itself well to the recursive formulation used in the first essay of this thesis. The issue also applies to modeling the unemployment benefit as something distinct from the transfer received for the low skilled in the recursive formulation. The essay of chapter 3 has made progress in analyzing the design of redistributive policies in the presence of unemployment scarring but more work is still required to tackle this complex problem.

The last essay of the thesis considers the design of the optimal linear tax system where available resources, or more precisely scarcity of resources, affects the economic decision-making of the individuals. The fiscal instruments considered are a linear tax on labor income, a linear tax on savings and a demogrant. It builds on the experimental economics literature that finds that through various mechanisms poverty can impede the economic decision-making process. The essay takes a general reduced-form approach by assuming that the workers' time preferences with regards to future consumption is a function of disposable income. In such a model, income taxation affects the savings decision both through a standard income effect but also through the effect of disposable income on time preferences. The optimal tax exercise is non-welfarist in the sense that the government substitutes its preferences for the decision preferences of the workers.

My coauthor and I find that the optimal tax system must take into consideration, besides the usual efficiency and redistributive motive of the government, the effects of taxes on the

already misaligned decisions of workers on labor and savings caused by the behavioral problem. In this model the Corlett and Hague (1953) result of the redundancy of indirect taxes no longer applies and savings should be taxed or subsidized. The government faces a novel tradeoff as it may want to raise revenues to subsidize savings or to increase the demogrant so as to reduce the behavioral problem, but to do this it may need to increase labor income taxation which would exacerbate the behavioral problem. Numerical simulations shows that savings should be subsidized and that as the behavioral problem becomes widespread these subsidies are financed by a decrease in the demogrant and an increase in the labor income tax.

These results suggest that a redistributive tax system should try to encourage savings and that if for other reasons savings must be taxed, the government should keep in mind under-savings brought about by the mechanism used in this essay. One further issue to consider is that the model considers fairly limited tax instruments. A tax system with varying saving tax rates based on either income or saving levels may offer a different picture. These varying rates would also be a function of the shape and prevalence of the behavioral problem which are both empirical questions that remain to be investigated. As optimal taxation with behavioral agents has garnered attention in the last few years, the implication of poverty as the source of these issues has remained largely unexplored in this literature. I believe that the third essay presented in this thesis is a starting point.

## Bibliography

- ALBANESI, S. AND C. SLEET (2006): “Dynamic Optimal Taxation with Private Information,” *Review of Economic Studies*, 73(1), 1–30.
- APPS, P. AND R. REES (2006): “Repeated Optimal Nonlinear Income Taxation,” *Mimeo*.
- ARULAMPALAM, W. (2001): “Is Unemployment Really Scarring? Effects of Unemployment Experiences on Wages,” *The Economic Journal*, 111, F585–F606.
- ARULAMPALAM, W., A. L. BOOTH, AND M. P. TAYLOR (2000): “Unemployment Persistence,” *Oxford Economic Paper*, 52, 24–50.
- ARULAMPALAM, W., P. GREGG, AND M. GREGORY (2001): “Unemployment Scarring,” *The Economic Journal*, 111, F577–F584.
- ATKINSON, A. B. AND J. STIGLITZ (1976): “The Design of Tax Structure: Direct vs. Indirect Taxation,” *Journal of Public Economics*, 6, 55–75.
- BAKIS, O., B. KAYMAK, AND M. POSHKE (2012): “On the Optimality of Progressive Income Redistribution,” *Working Paper*.
- BANKS, J. AND P. DIAMOND (2010): *The Base of Direct Taxation*, Oxford University Press, chap. 6.
- BASTANI, S. (May 2014): “Using the Discrete Model to Derive Optimal Income Tax Rates,” *Working Paper*.
- BERLIANT, M. AND J. LEDYARD (2011): “Optimal Dynamic Nonlinear Income Taxes with No Commitment,” *mimeo*.
- BESLEY, T. (1988): “A Simple Model for Merit Good Arguments,” *Journal of Public Economics*, 35, 371–383.

- BEST, M. AND H. J. KLEVEN (2012): “Optimal Income Taxation with Career Effects of Work Effort,” *Working Paper*.
- BEST, M. C. AND H. J. KLEVEN (February 2013): “Optimal Income Taxation with Career Effects of Work Effort,” *Working Paper*.
- BISIN, A. AND T. VERDIER (2001): “The Economics of Cultural Transmission and the Dynamics of Preferences,” *Journal of Economic Theory*, 97, 298–319.
- BLOMQUIST, S. AND V. CHRISTIANSEN (2008): “Taxation and Heterogeneous Preferences,” *FinanzArchiv: Public Finance Analysis*, 64(2), 218–244.
- BLOMQUIST, S. AND L. MICHELETTO (2008): “Age-Related Optimal Income Taxation,” *The Scandinavian Journal of Economics*, 110(1), 45–71.
- BLUNDELL, R. AND T. MACURDY (1999): “Labor Supply: A Review of Alternative Approaches,” in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, North Holland: Elsevier, vol. 3, chap. 27, 1559–1695.
- BOADWAY, R., K. CUFF, AND N. MARCEAU (2003): “Redistribution and Employment Policies with Endogenous Unemployment,” *Journal of Public Economics*, 87, 2407–2430.
- BOADWAY, R., M. MARCHAND, P. PESTIEAU, AND M. DEL MAR RACIONERO (2002): “Optimal Redistribution with Heterogeneous Preferences for Leisure,” *Journal of Public Economic Theory*, 4 (4), 475–498.
- BOHACEK, R. AND M. KAPICKA (2008): “Optimal human capital policies,” *Journal of Monetary Economics*, 55, 1–16.
- BOONE, J. AND L. BOVENBERG (2002): “Optimal Labor Taxation and Search,” *Journal of Public Economics*, 85, 53–97.
- (2004): “The Optimal Taxation of Unskilled Labor with Job Search and Social Assistance,” *Journal of Public Economics*, 88(11), 2227–2258.
- BOUCHARD ST-AMAND, P.-A. AND J.-D. GARON (July 2014): “Optimal Redistributive Pensions and the Cost of Self-Control,” *International Tax and Public Finance*.
- BOVENBERG, A. L. AND B. JACOBS (2005): “Redistribution and Education Subsidies are Siamese Twins,” *Journal of Public Economics*, 89(11-12), 2005–2035.

- (October 2001): “Reditribution and Education Subsidies Are Siamese Twins,” *CentER Discussion Paper*.
- BRETT, C. AND J. A. WEYMARK (2003): “Financing Education Using Optimal Redistributive Taxation,” *Journal of Public Economics*, 87, 2549–2569.
- (January 2008): “Optimal Nonlinear Taxation of Income and Savings without Commitment,” *Working Paper No. 08-W05*.
- (May 2010): “How Optimal Nonlinear Income Taxes Change When the Distribution of the Population Changes,” *Working Paper*.
- (November 2007): “The Impact of Changing Skill Levels on Optimal Nonlinear Income Taxes,” *Working Paper No. 07-W08R*.
- CHETTY, R. (2006): “A General Formula for the Optimal Level of Social Insurance,” *Journal of Public Economics*, 90, 1879–1901.
- CHONÉ, P. AND G. LAROQUE (2010): “Negative Marginal Tax Rates and Heterogeneity,” *American Economic Review*, 100(5), 2532–47.
- COLES, M. AND A. MASTERS (2000): “Retraining and Long-term Unemployment in a Model of Unlearning by Not Doing,” *European Economic Review*, 44, 1801–1822.
- CONESA, J. C., S. KITAO, AND D. KRUEGER (2009): “Taxing Capital? Not a Bad Idea after All!” *American Economic Review*, 99 (1), 2548.
- CORLETT, W. AND D. HAGUE (1953): “Complementarity and the excess burden of taxation,” *Review of Economic Studies*, 21, 21–30.
- CREMER, H. AND P. PESTIEAU (2011): “Myopia, Redistribution and Pensions.” *European Economic Review*, 55, 165–175.
- CREMER, H., P. PESTIEAU, AND M. RACIONERO (2011): “Unequal Wages for Equal Utilities,” *International Tax and Public Finance*, 18, 383–398.
- CREMER, H. AND DE DONDER, P., D. MALDONADO, AND P. PESTIEAU (2008): “Designing an Optimal Linear Pension Scheme with Forced Savings and wage Heterogeneity,” *International Tax and Public Finance*, 15, 547–562.

- CUFF, K. (2000): “Optimality of Workfare with Heterogeneous Preferences,” *Canadian Journal of Economics*, 33(1), 149–174.
- DAHL, G. B., A. R. KOSTOL, AND M. MOGSTAD (2014): “Family Welfare Cultures,” *The Quarterly Journal of Economics*, 1711–1752.
- DIAMOND, P. (1982): “Wage Determination and Efficiency in Search Equilibrium,” *Review of Economic Studies*, 49(2), 217–227.
- (2006): “Optimal Tax Treatment of Private Contributions for Public Goods with and without Warm Glow Preferences,” *Journal of Public Economics*, 90(4-5), 897-919.
- DIAMOND, P. AND J. MIRRLEES (1971): “Optimal Taxation and Public Production I: Production Efficiency and II: Tax Rules,” *American Economic Review*, 61, 827 and 261–278.
- DIAMOND, P. AND J. SPINNEWIJN (2011): “Capital Income Taxes with Heterogeneous Discount Rates,” *American Economic Journal: Economic Policy*, 3, 527-6.
- DIAMOND, P. A. (1980): “Income Taxation with Fixed Hours of Work,” *Journal of Public Economics*, 13, 101–110.
- (2003): *Taxation, Incomplete Markets, And Social Security*, Cambridge, MA: MIT Press.
- DOEPKE, M. AND F. ZILIBOTTI (2008): “Occupational Choice and the Spirit of Capitalism,” *The Quarterly Journal of Economics*, 123(2), 747–793.
- (December 2012): “Parenting with Style: Altruism and Paternalism in Intergenerational Preference Transmission,” *Working Paper*.
- EATON, J. AND H. S. ROSEN (1980): “Taxation, Human Capital and Uncertainty,” *American Economic Review*, 70, 705–715.
- ENGSTRÖM, P. (2002): “Optimal Non-Linear Taxation In Search Equilibrium,” (*Working Paper 2002:18, Uppsala University*).
- FARBER, H. (1999): *Handbook of Labor Economics, vol. 3*, Amsterdam: North Holland, chap. Chap 37: Mobility and Stability: The Dynamics of Job Change in Labor Markets.

- FARHI, E. AND X. GABAIX (2015): “Optimal Taxation with Behavioral Agents,” *Working Paper*.
- FARHI, E. AND I. WERNING (2012): “Insurance and Taxation Over the Life Cycle,” *Working Paper*.
- (2013): “Insurance and Taxation Over the Life Cycle,” *Review of Economic Studies*, 80(2), 596–635.
- FEHR, H., C. HABERMANN, AND F. KINDERMANN (2008): “Social Security with Rational and hyperbolic Consumers,” *Review of Economic Dynamics*, 11(4), 884–903.
- FERNANDES, A. AND C. PHELAN (2000): “A Recursive Formulation for Repeated Agency with History Dependence,” *Journal of Economic Theory*, 91, 223–247.
- FLEURBAEY, M. (2008): *Fairness, Responsibility and Welfare*, Oxford University Press.
- FLEURBAEY, M. AND F. MANIQUET (2006): “Fair Income Tax,” *Review of Economic Studies*, 73, 5583.
- GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKY (2003): “Optimal Indirect and Capital Taxation,” *Review of Economic Studies*, 70, 569–587.
- GOLOSOV, M., M. TROSHKIN, AND A. TSYVINSKI (2011): “Optimal Dynamic Taxes,” *Working Paper*.
- (2013a): “Redistribution and Social Insurance,” *Working Paper*.
- (2015): “Redistribution and Social Insurance,” *Working Paper*.
- GOLOSOV, M., M. TROSHKIN, A. TSYVINSKI, AND M. WEINZIERL (2013b): “Preference Heterogeneity and Optimal Capital Income Taxation,” *Journal of Public Economics*, 97, 160–175.
- GOLOSOV, M., A. TSYVINSKY, AND I. WERNING (2007): “New Dynamic Public Finance: A Users Guide,” *NBER Macroannual 2006*, 317–387.
- GREGG, P. (2001): “The Impact of Youth Unemployment on Adult Unemployment in the NCDS,” *The Economic Journal*, 111, F626–F653.
- GREGG, P. AND E. TOMINEY (2005): “The Wage Scar from Youth Unemployment,” *Labour Economics*, 12, 487–509.

- GUESNERIE, R. (1995): *A Contribution to the Pure Theory of Taxation*, Cambridge University Press.
- GUO, J.-T. AND A. KRAUSE (January 2011): “Dynamic Income Taxation without Commitment: Comparing Alternative Tax Systems,” *Working Paper*.
- (October 2011): “Optimal Dynamic Nonlinear Income Taxation under Loose Commitment,” *Working Paper*.
- HECKMAN, J. J. (2008): “Role of Income and Family Influence on Child Outcomes,” *Annals of the New York Academy of Sciences*, 1136, 307–323.
- HELLWIG, M. F. (2007): “A contribution to the theory of optimal utilitarian income taxation,” *Journal of Public Economics*, 91, 1449–1477.
- HOPENHAYN, H. A. AND E. C. PRESCOTT (1992): “Stochastic Monotonicity and Stationary Distributions for Dynamic Economies,” *Econometrica*, 60(6), 1387–1406.
- HOSIOS, A. (1990): “On the Efficiency of Matching and Related Models of Search and Unemployment,” *Review of Economic Studies*, 57, 279–298.
- HUGGETT, M. (1993): “The Risk-Free Rate in Heterogeneous-Agent Incomplete-Insurance Economies,” *Journal of Economic Dynamics and Control*, 17, 953–969.
- HUGGETT, M., G. VENTURA, AND A. YARON (2011): “Sources of Lifetime Inequality,” *American Economic Review*, 101(7), 2923–54.
- HUNGERBÜHLER, M., E. LEHMANN, A. PARMENTIER, AND B. VAN DER LINDEN (2006): “Optimal Redistributive Taxation in A Search Equilibrium Model,” *Review of Economic Studies*, 73, 743–767.
- (2008): “A Simple Theory of Optimal Redistributive Taxation with Equilibrium Unemployment,” *Economie Publique/Public Economics*, 1/2 22-23, 203–218.
- (March 2010): “A Simple Theory of Optimal Redistributive Taxation with Equilibrium Unemployment,” *IZA Discussion Paper Series*, No. 4832.
- JACOBS, B. (2005): “Optimal Income Taxation with Endogenous Human Capital,” *Journal of Public Economic Theory*, 7(2), 295–315.

- JACOBS, B. AND A. L. BOVENBERG (2009): “Human capital and optimal positive taxation of capital income,” *International Tax and Public Finance*, 10.1007/s10797-009-9120-5, 1–28.
- JACQUET, L., E. LEHMANN, AND B. V. DER LINDEN (2014): “Optimal Income Taxation with Kalai wage bargaining and endogenous participation,” *Social Choice and Welfare*, 42(2), 381–402.
- (September 2011): “Optimal Redistributive Taxation with both Labor Supply and Labor Demand Responses,” *Working Paper*.
- JACQUET, L., E. LEHMANN, AND B. VAN DER LINDEN (2013): “The Optimal Marginal Tax Rates with both Extensive and Intensive Responses,” *Journal of Economic Theory*, 148(5), 1770–1805.
- JACQUET, L. AND D. VAN DE GAER (2011): “A Comparison of Optimal Tax Policies when Compensation or Responsibility Matter,” *Journal of Public Economics*, 95(11-12), 1248–1262.
- JUDD, K. AND C.-L. SU (2006): “Optimal Income Taxation with Multidimensional Taxpayer Types,” *Working Paper*.
- KANBUR, R., J. PIRTIL, AND M. TUOMALA (2006): “Non-Welfarist Optimal Taxation And Behavioural Public Economics,” *Journal of Economic Surveys*, 20(5), 849–868.
- (2008): “Moral Hazard, Income Taxation and Prospect Theory,” *The Scandinavian Journal of Economics*, 110(2), 321–337.
- KAPICKA, M. (2013): “Efficient Allocations in Dynamic Private Information Economies with Persistent Shocks: A First-Order Approach,” *Review of Economic Studies*, 80(3), 1027–1054.
- (2014): “Pareto Efficient Taxation with Learning by Doing,” *Working Paper*.
- (May 2014): “Optimal Mirrleesian Taxation with Unobservable Human Capital Formation,” *American Economic Journal: Macroeconomics*, Forthcoming.
- KAPICKA, M. AND J. NEIRA (October 2014): “Optimal Taxation in a Life-Cycle Economy with Endogenous Human Capital Formation,” *Working Paper*.

- KOCHERLAKOTA, N. (2005): “Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation,” *Econometrica*, 73, 1587-1621.
- KOCHERLAKOTA, N. R. (2010): *The New Dynamic Public Finance*, Princeton University Press.
- KRAUSE, A. (2009a): “Optimal nonlinear income taxation with learning-by-doing,” *Journal of Public Economics*, 93, 1098-1110.
- (2009b): “Optimal nonlinear income taxation with learning-by-doing,” *Journal of Public Economics*, 93, 1098–1110.
- KREMER, M. (2002): “Should Taxes be Independent of Age,” *Mimeo*.
- KRUSELL, P. AND A. SMITH (1998): “Income and Wealth Heterogeneity in the Macroeconomy,” *Journal of Political Economy*, 106(5), 867–896.
- KUHNEN, C. AND B. KNUTSON (September 2011): “The Influence of Affect on Beliefs, Preferences and Financial Decisions,” *MPRA Paper No. 10410*.
- LEHMANN, E., A. PARMENTIER, AND B. V. DER LINDEN (2011): “Optimal Income Taxation with Endogenous Participation and Search Unemployment,” *Journal of Public Economics*, 95, 1523–1537.
- LEHMANN, E. AND B. VAN DER LINDEN (2007): “On the Optimality of Search-Matching Equilibrium when Workers are Risk Averse,” *Journal of Public Economic Theory*, 9, 867–884.
- LJUNGVIST, L. AND T. J. SARGENT (1998): “The European Unemployment Dilemma,” *Journal of Political Economy*, 106(3), 514–550.
- MANKIW, N. G., M. WEINZIERL, AND D. YAGAN (2009): “Optimal taxation in theory and practice,” *Journal of Economic Perspectives*, 23(4), 147–74.
- MIRRELES, J. (1971): “An exploration in the theory of optimal income taxation,” *Review of Economic Studies*, 38, 175–208.
- MOEN, E. (1997): “Competitive Search Equilibrium,” *Journal of Political Economy*, 105(2), 385–411.

- MORTENSEN, D. AND C. A. PISSARIDES (1999): *Handbook of Labour Economics, Vol.3*, Amsterdam: North Holland, chap. New Developments in Models of Search in the Labor Market, 2567–2627.
- NAITO, H. (1999): “Re-examination of uniform commodity taxes under a non-linear income tax system and its implication for production efficiency,” *Journal of Public Economics*, 71, 165–188.
- O’DONOGHUE, T. AND M. RABIN (1999): “Doing It Now or Later,” *The American Economic Review*, 89(1), 103–124.
- (2006): “Optimal Sin Taxes,” *Journal of Public Economics*, 90(10), 1825–1849.
- OECD (2004): “Improving Skills for More and Better Jobs: Does Training Make a Difference?” *Discussion Paper, OECD Employment Outlook*.
- ORDOVER, J. A. AND E. S. PHELPS (1979): “The Concept of Optimal Taxation in the Overlapping-generations Model of Capital and Wealth,” *Journal of Public Economics*, 12, 1–26.
- OREOPOULOS, P., T. VON WACHTER, AND A. HEISZ (2012): “The Short- and Long-Term Career Effects of Graduating in a Recession,” *American Economic Journal: Applied Economics*, 4(1), 1–29.
- PAVONI, N. (2009): “Optimal unemployment insurance, with human capital depreciation, and duration dependence,” *International Economic Review*, 50(2), 323–362.
- (September 2008): “Optimal Unemployment Insurance, with Human Capital Depreciation, and Duration Dependence,” *Working Paper*.
- PERRAULT, L. (January 2015): “Optimal Income Taxation with Unemployment Induced Loss of Human Capital,” *Working Paper*.
- PIKETTY, T. (1997): “La Redistribution Fiscale Face au Chomage,” *Revue Francaise d’Economie*, 12, 157–201.
- PIKETTY, T. AND E. SAEZ (2012a): “A Theory of Optimal Capital Taxation,” *NBER Working Paper 17989*.
- (2012b): “A Theory of Optimal Inheritance Taxation,” *CEPR Discussion Paper No. 9241*.

- (2013a): “Optimal Labor Income Taxation,” in *Handbook of Public Economics, Volume 5*, Elsevier.
- (2013b): “A Theory of Optimal Inheritance Taxation,” *Econometrica*, 81(5), 1851–1886.
- PISSARIDES, C. A. (1992): “Loss of Skill During Unemployment and the Persistence of Employment Shocks,” *The Quarterly Journal of Economics*, 107(4), 1371–1391.
- (2000): *Equilibrium Unemployment Theory*, 2nd edn (Cambridge, MA: MIT Press).
- RAWLS, J. (1971): *A Theory of Justice*, Cambridge, Massachusetts: Belknap Press.
- ROCHET, J.-C. (1985): “The Taxation Principle and Multi-time Hamilton-Jacobi Equations,” *Journal of Mathematical Economics*, 14, 113–128.
- SAEZ, E. (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.
- (2004): “Direct or Indirect Tax Instruments for Redistribution: Short-run versus Long-run,” *Journal of Public Economics*, 88, 503–518.
- SAEZ, E. AND S. STANTCHEVA (January 2013): “Generalized Social Marginal Welfare Weights for Optimal Tax Theory,” January 2013.
- SAEZ-MARTI, M. AND F. ZILIBOTTI (2008): “Preferences as Human Capital: Rational Choice Theories of Endogenous Preferences and Socioeconomic Changes,” *Finnish Economic Papers*, 21(2), 81–94.
- SALANIÉ, B. (2003): *The Economics of Taxation*, The MIT Press.
- SEADE, J. (1980): *Optimal Non-Linear Policies for Non-Utilitarian Motives*, Bristol: Science Technica.
- SEN, A. (1985): *Commodities and Capabilities*, Amsterdam: North Holland.
- SHAH, A. K., S. MULLAINATHAN, AND E. SHAFIR (2012): “Some Consequences of Having Too Little,” *Science*, 338, 682–85.

- SHIMER, R. AND I. WERNING (2006): “On the Optimal Timing of Benefits with Heterogeneous Workers and Human Capital Depreciation,” *Working Paper*.
- SØRENSEN, P. B. (1999): “Optimal Tax Progressivity in Imperfect Labor Markets,” *Labour Economics*, 6, 435–452.
- SPEAR, S. AND S. SRIVASTAVA (1987): “On Repeated Moral Hazard with Discounting,” *Review of Economic Studies*, 54(4), 599–617.
- SPEARS, D. (2011): “Economic Decision-Making in Poverty Depletes Behavioral Control,” *The B.E. Journal of Economic Analysis and Policy*, 11(1).
- SPINNEWIJN, J. (2013): “Training and Search during Unemployment,” *Journal of Public Economics*, 99, 49–65.
- (November 2010): “Training and Search during Unemployment,” *Working Paper*.
- STANTCHEVA, S. (2014): “Optimal Taxation and Human Capital Policies over the Life Cycle,” *Working Paper*.
- (December 2014): “Learning and (or) Doing: Human Capital Investments and Optimal Taxation,” *Working Paper*.
- STIGLITZ, J. (1982): “Self-selection and pareto efficient taxation,” *Journal of Public Economics*, 17, 213–240.
- STOKEY, N. L., R. E. LUCAS, AND E. C. PRESCOTT (1989): *Recursive Methods in Economic Dynamics*, Cambridge, MA: Harvard University Press.
- STORESLETTEN, K., C. TELMER, AND A. YARON (2004): “Consumption and Risk Sharing Over The Life Cycle,” *Journal of Monetary Economics*, 51(3), 609–633.
- THOMAS, J. AND T. WORRALL (1990): “Income Fluctuation and Asymmetric Information: An Example of a Repeated Principal-Agent Problem,” *Journal of Economic Theory*, 51(2), 367–390.
- WEINZIERL, M. (2011): “The Surprising Power of Age-Dependent Taxes,” *Review of Economic Studies*, 78, 1490–1518.
- WERNING, I. (2001): “Moral Hazard with Unobserved Endowments: A Recursive Approach,” *Working Paper*.
- (September 2011): “Nonlinear Capital Taxation,” *Working Paper*.

## Appendix A

### Mathematical Appendix

#### A.1 Optimal taxation with work experience as a risky investment

##### A.1.1 The Hamiltonians

First, use the first constraint in the first period problem and manipulate it such that

$$c(\theta) = u^{-1} \left( \omega(\theta) + h \left( \frac{y(\theta)}{w_1(\theta, e_1)} \right) - \beta v(\theta) \right).$$

The Hamiltonian for period  $t = 1$  is:

$$\begin{aligned} & \left[ u^{-1} \left( \omega(\theta) + h \left( \frac{y(\theta)}{w_1(\theta, e_1)} \right) - \beta v(\theta) \right) - y(\theta) + \frac{1}{R} K(v(\theta), e(\theta), 2) \right] f^1(\theta) \\ & + \lambda_1 [v_0 - \omega(\theta) f^1(\theta)] + \mu(\theta) \left[ \frac{h' \left( \frac{y(\theta)}{w_1(\theta, e_1)} \right) y(\theta)}{[w_1(\theta, e_1)]^2} \frac{\partial w_1(\theta, e_1)}{\partial \theta_1} \right], \end{aligned}$$

where

$$e(\theta) = \phi(y(\theta)) + e_1,$$

and the Hamiltonian for period  $t = 2$  is

$$\begin{aligned} & \left[ u^{-1} \left( \omega(\theta) + h \left( \frac{y(\theta)}{w_2(\theta, e_2)} \right) \right) - y(\theta) \right] f^2(\theta) \\ & + \lambda_2 [v - \omega(\theta) f^2(\theta)] + \mu(\theta) \left[ \frac{h' \left( \frac{y(\theta)}{w_2(\theta, e_2)} \right) y(\theta)}{[w_2(\theta, e_2)]^2} \frac{\partial w_2(\theta, e_2)}{\partial \theta_2} \right]. \end{aligned}$$

Furthermore using the envelope theorem we get:

$$K_v(v, e_2, 2) = \lambda_2,$$

and so from the point of view of the first period, we have:

$$K_v(v(\theta), e(\theta), 2) = \lambda(\theta).$$

The FOCs for the second period problem are:

$y(\theta)$ :

$$\left[ -1 + \frac{h' \left( \frac{y(\theta)}{w_2(\theta, e_2)} \right)}{u'(c(\theta)) w_2(\theta, e_2)} \right] f^2(\theta) + \mu(\theta) \left[ \frac{h'' \left( \frac{y(\theta)}{w_2(\theta, e_2)} \right) y(\theta)}{[w_2(\theta, e_2)]^3} + \frac{h' \left( \frac{y(\theta)}{w_2(\theta, e_2)} \right)}{[w_2(\theta, e_2)]^2} \frac{\partial w_2(\theta, e_2)}{\partial \theta} \right] = 0. \quad (\text{A.1})$$

The law of motion for the co-state variable  $\omega(\theta)$ :

$$-\dot{\mu}(\theta) = \frac{f^2(\theta)}{u'(c(\theta))} - \lambda_2 f^2(\theta). \quad (\text{A.2})$$

The FOCS for the first period are:

$y(\theta)$ :

$$\left[ -1 + \frac{h'(l(\theta))}{u'(c(\theta))w_1(\theta, e_1)} + \frac{1}{R} \phi'(y(\theta)) \frac{\partial K}{\partial e(\theta)} \right] f^1(\theta) + \mu(\theta) \left[ \frac{h''(l(\theta))l(\theta)}{[w_1(\theta, e_1)]^2} + \frac{h'(l(\theta))}{[w_1(\theta, e_1)]^2} \right] \frac{\partial w_1(\theta, e_1)}{\partial \theta} = 0, \quad (\text{A.3})$$

where

$$\begin{aligned} \frac{\partial K}{\partial e(\theta)} = & - \int \frac{h' \left( \frac{y(\theta)}{w_2(\theta, e_2)} \right) y(\theta)}{u'(c(\theta)) [w_2(\theta, e_2)]^2} \frac{\partial w_2(\theta, e_2)}{\partial e} f^2(\theta) d\theta \\ & - \int \frac{\mu(\theta)}{f^2(\theta)} \left[ \frac{h'' \left( \frac{y(\theta)}{w_2(\theta, e_2)} \right) y(\theta)}{[w_2(\theta, e_2)]^3} + \frac{h' \left( \frac{y(\theta)}{w_2(\theta, e_2)} \right)}{[w_2(\theta, e_2)]^2} \right] \frac{\partial w_2(\theta, e_2)}{\partial \theta} \frac{\partial w_2(\theta, e_2)}{\partial e(\theta)} \frac{y(\theta)}{w_2(\theta, e_2)} f^2(\theta) d\theta \\ & - \int \frac{\mu(\theta)}{f^2(\theta)} \frac{h' \left( \frac{y(\theta)}{w_2(\theta, e_2)} \right) y(\theta)}{[w_2(\theta, e_2)]^3} \frac{\partial w_2(\theta, e_2)}{\partial \theta} \frac{\partial w_2(\theta, e_2)}{\partial e(\theta)} \left[ 1 - \frac{\frac{\partial^2 w_2(\theta, e_2)}{\partial \theta \partial e(\theta)}}{\frac{\partial w_2(\theta, e_2)}{\partial \theta} \frac{\partial w_2(\theta, e_2)}{\partial e(\theta)}} \right] f^2(\theta) d\theta. \end{aligned} \quad (\text{A.4})$$

$v(\theta)$ :

$$-\frac{\beta}{u'(c(\theta))} + \frac{1}{R} \frac{\partial K}{\partial v(\theta)} = 0, \quad (\text{A.5})$$

The law of motion for the co-state variable  $\omega(\theta)$ :

$$-\dot{\mu}(\theta) = \frac{f^1(\theta)}{u'(c(\theta))} - \lambda_1 f^1(\theta). \quad (\text{A.6})$$

### A.1.1.1 Labor wedge

**Labor wedge in Period 2** Starting for  $t=2$ , using the FOC for  $y(\theta^2)$  and rearranging, we get:

$$1 - \frac{h'(l(\theta^2))}{u'(c(\theta^2))w_2(\theta_2, e_2)} = \frac{\mu(\theta^2)}{f^2(\theta_2)} \frac{h'(l(\theta^2))}{[w_2(\theta_2, e_2)]^2} \left[ 1 + \frac{h''(l(\theta^2))l(\theta^1)}{h'(l(\theta^2))} \right] \frac{\partial w_2(\theta_2, e_2)}{\partial \theta}. \quad (\text{A.7})$$

Using the definition our elasticity measure and

$$\epsilon_{\theta t} \equiv \frac{\partial w_t(\theta_t, e(\theta^{t-1}))}{\partial \theta_t} \frac{\theta_t}{w_t(\theta_t, e(\theta^{t-1}))},$$

we obtain the formula (2.26) from Proposition 1.

To write the optimal second period wedge formula (2.26) into the ABC format use the definition from above, and see that

$$\mu(\theta^2)u'(c(\theta_2)) = \int_{\theta_2}^{\bar{\theta}} \left( \frac{u'(c(\theta_2))}{u'(c(x))} \right) [1 - \lambda_2 u'(c(x))] f^2(x) dx,$$

and

$$\begin{aligned} \frac{u'(c(\theta))}{u'(c(x))} &= \exp \left( \ln \frac{u'(c(\theta))}{u'(c(x))} \right) = \exp \left( - \int_{\theta}^x \frac{du'(c(\tilde{x}))}{u'(c(\tilde{x}))} \right) \\ &= \exp \left( - \int_{\theta}^x \frac{u''(c(\tilde{x}))dc(\tilde{x})}{u'(c(\tilde{x}))} \right) = \exp \left( - \int_{\theta}^x \frac{u''(c(\tilde{x}))c(\tilde{x}) \dot{c}(\tilde{x})}{u'(c(\tilde{x}))c(\tilde{x})} d\tilde{x} \right) \\ &= \exp \left( \int_{\theta}^x \eta(\tilde{x}) \frac{\dot{c}(\tilde{x})}{c(\tilde{x})} d\tilde{x} \right). \end{aligned}$$

Multiplying on both sides by  $(1 - F^2(\theta))/(1 - F^2(\theta))$  (2.26) and using the new definitions,

I obtain

$$\frac{\tau_L^*(\theta^2)}{1 - \tau_L^*(\theta^2)} = (1 + \alpha(\theta^2))\epsilon_{\theta^2} \left[ \frac{1 - F^2(\theta_2)}{\theta_2 f^2(\theta_2)} \right] \frac{\int_{\theta_2}^{\bar{\theta}} \exp\left(\int_{\theta}^x \eta(\tilde{x}) \frac{\dot{c}(\tilde{x})}{c(\tilde{x})} d\tilde{x}\right) [1 - \lambda_2 u'(c(\tilde{x}))] f^2(x) dx}{1 - F^2(\theta_2)}.$$

**Labor wedge in Period 1** Using the FOC for  $y(\theta^1)$  and rearranging, we get:

$$1 - \frac{h'(l(\theta^1))}{u'(c(\theta^1))w_1(\theta_1, e_1)} - \frac{1}{R} \phi'(y(\theta^1)) \frac{\partial K}{\partial e(\theta^1)} = \frac{\mu(\theta^1)}{f^1(\theta_1)} \frac{h'(l(\theta^1))}{[w_1(\theta_1, e_1)]^2} \left[ 1 + \frac{h''(l(\theta^1)) l(\theta^1)}{h'(l(\theta^1))} \right] \frac{\partial w_1(\theta_1, e_1)}{\partial \theta_1}. \quad (\text{A.8})$$

Notice that

$$\begin{aligned} \int \frac{h'(l(\theta^2)) l(\theta^2)}{u'(c(\theta^2))w_2(\theta_2, e_2)} \frac{\partial w_2(\theta_2, e_2)}{\partial e} f^2(\theta_2) d\theta_2 &= \int \frac{1}{u'(c(\theta^2))} f^2(\theta_2) d\theta_2 \\ &\times \int \frac{h'(l(\theta^2)) l(\theta^2)}{w_2(\theta_2, e_2)} \frac{\partial w_2(\theta_2, e_2)}{\partial e} f^2(\theta_2) d\theta_2 \\ &+ \text{Cov}\left(\frac{1}{u'(c(\theta^2))}, \frac{h'(l(\theta^2)) l(\theta^2)}{w_2(\theta_2, e_2)} \frac{\partial w_2(\theta_2, e_2)}{\partial e}\right), \end{aligned}$$

and using the result from the intertemporal wedge, we have:

$$\begin{aligned} \int \frac{h'(l(\theta^2)) l(\theta^2)}{u'(c(\theta^2))w_2(\theta_2, e_2)} \frac{\partial w_2(\theta_2, e_2)}{\partial e} f^2(\theta_2) d\theta_2 &= \frac{R\beta}{u'(c(\theta^1))} \int \frac{h'(l(\theta^2)) l(\theta^2)}{w_2(\theta_2, e_2)} \frac{\partial w_2(\theta_2, e_2)}{\partial e} f^2(\theta_2) d\theta_2 \\ &+ \text{Cov}\left(\frac{1}{u'(c(\theta^2))}, \frac{h'(l(\theta^2)) l(\theta^2)}{w_2(\theta_2, e_2)} \frac{\partial w_2(\theta_2, e_2)}{\partial e}\right). \end{aligned}$$

Reinserting this in  $\partial K/\partial e$ , and using the rewritten FOC, we have:

$$\begin{aligned}
& 1 - \frac{h'(l(\theta^1))}{u'(c(\theta^1))w_1(\theta_1, e_1)} + \beta \frac{\phi'(y(\theta^1))}{u'(c(\theta^1))} \int \frac{h'(l(\theta^2))l(\theta^2)}{w_2(\theta_2, e_2)} \frac{\partial w_2(\theta_2, e_2)}{\partial e} f^2(\theta_2) d\theta_2 = \\
& \frac{\mu(\theta^1)}{f^1(\theta_1)} \frac{h'(l(\theta^1))}{[w_1(\theta_1, e_1)]^2} \left[ 1 + \frac{h''(l(\theta^1))l(\theta^1)}{h'(l(\theta^1))} \right] \frac{\partial w_1(\theta_1, e_1)}{\partial \theta_1} \\
& - \frac{\phi'(y(\theta^1))}{R} \left\{ \text{Cov} \left( \frac{1}{u'(c(\theta^2))}, \frac{h'(l(\theta^2))l(\theta^2)}{w_2(\theta_2, e_2)} \frac{\partial w_2(\theta_2, e_2)}{\partial e} \right) + \int \tau_L(\theta^2) \frac{\partial w_2(\theta_2, e_2)}{\partial e} l(\theta^2) f^2(\theta_2) d\theta_2 \right. \\
& \left. + \int \frac{\mu(\theta^2)}{f^2(\theta_2)} \frac{h'(l(\theta^2))l(\theta^2)}{[w_2(\theta_2, e_2)]^2} \frac{\partial w_2(\theta_2, e_2)}{\partial \theta} \frac{\partial w_2(\theta_2, e_2)}{\partial e(\theta^1)} \left[ 1 - \frac{\frac{\partial^2 w_2(\theta_2, e_2)}{\partial \theta \partial e(\theta^1)}}{\frac{\partial w_2(\theta_2, e_2)}{\partial \theta_2} \frac{\partial w_2(\theta_2, e_2)}{\partial e(\theta^1)}} \right] f^2(\theta_2) d\theta_2 \right\}.
\end{aligned} \tag{A.9}$$

Using the definitions of the Hicksian complementarity, the labor elasticity and the wage elasticity, we obtain the results of Proposition 1.

#### A.1.1.2 Intertemporal Wedge

Using the FOC on  $\omega(\theta)$  for period t=2 and the following boundary conditions:

$$\lim_{\theta \rightarrow \underline{\theta}} \mu(\theta) = 0 \quad \text{and} \quad \lim_{\theta \rightarrow \bar{\theta}} \mu(\theta) = 0,$$

we get

$$\mu(\theta) = \int_{\underline{\theta}}^{\bar{\theta}} \left[ \frac{1}{u'(c(\theta))} - \lambda_2 \right] f^2(\theta) d\theta,$$

and

$$\int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{u'(c(\theta))} f^2(\theta) d\theta = \lambda_2.$$

Using the FOC on  $v(\theta)$  in  $t=1$  and the envelope result from above, we have

$$\lambda(\theta^1) = \frac{\beta R}{u'(c(\theta^1))}.$$

Combining both we obtain the inverse Euler equation:

$$\frac{1}{u'(c(\theta^1))} = \frac{1}{R\beta} \int_{\underline{\theta}}^{\bar{\theta}} \frac{1}{u'(c(\theta^2))} f^2(\theta_2) d\theta_2.$$

### A.1.1.3 Proof of Corollary 2.a and 2.b

The proof of Corollary 2.a and 2.b follows the proof of Corollary 1 in Golosov et al. (2015).<sup>1</sup>

Properties of the log-normal distribution: Use the fact that

$$\frac{\theta f^{2'}(\theta)}{f^2(\theta)} = 1 + \frac{\ln \theta - \mu}{\sigma^2},$$

and by L'Hôpital's rule, it is possible to obtain

$$\begin{aligned} \lim_{\theta \rightarrow \infty} \frac{1 - F^2(\theta)}{\theta f^2(\theta)} &= \lim_{\theta \rightarrow \infty} \frac{-1}{\theta f^{2'}(\theta)/f^2(\theta) + 1} = 0, \\ \lim_{\theta \rightarrow \infty} \frac{(1 - F^2(\theta))(\ln \theta - \mu)/\sigma^2}{\theta f^2(\theta)} &= \lim_{\theta \rightarrow \infty} \frac{(\ln \theta - \mu)/\sigma^2 + (1 - F^2(\theta))/(\theta f^2(\theta)\sigma^2)}{\theta f^{2'}(\theta)/f^2(\theta) + 1} = 1. \end{aligned}$$

From this the following result follows:  $\frac{1 - F^2(\theta)}{\theta f^2(\theta)} \sim \frac{\sigma^2}{\ln \theta - \mu}$ .

### Proof of Corollary 2.a

Using Assumption 2, and also specifically the fact that both  $\frac{c_2(\theta)}{y_2(\theta)}$  and  $\frac{\dot{c}_2(\theta)/c_2(\theta)}{\dot{y}_2(\theta)/y_2(\theta)}$  have finite

---

<sup>1</sup>However, I do not prove every lemma as I do not require all of them for the proof of Corollary 2.a and 2.b.

limits, it follows that there exist a limit for  $\lim_{\theta \rightarrow \infty} \frac{\dot{c}_2(\theta)}{\dot{y}_2(\theta)}$  since

$$\frac{\dot{c}_2(\theta)}{\dot{y}_2(\theta)} = \frac{\dot{c}_2(\theta)/c_2(\theta)}{\dot{y}_2(\theta)/y_2(\theta)} \frac{c_2(\theta)}{y_2(\theta)}. \quad (\text{A.10})$$

Also note that  $c_2(\theta), y_2(\theta) \rightarrow \infty$  as  $\theta \rightarrow \infty$  must hold. If this was not the case, it would imply that  $1 - \tau_L^2(\theta) = \frac{h'(y_2(\theta)/w_2(\theta, e))}{u'(c(\theta))w_2(\theta, e)} \rightarrow 0$  as  $\theta$  goes to infinity which would contradict the assumption that  $\frac{\tau_L^2(\theta)}{1 - \tau_L^2(\theta)}$  has a finite limit.

Using L'Hôpital's rule

$$\lim_{\theta \rightarrow \infty} \frac{c_2(\theta)}{y_2(\theta)} = \lim_{\theta \rightarrow \infty} \frac{\dot{c}_2(\theta)}{\dot{y}_2(\theta)},$$

and using (A.10) I obtain

$$1 = \frac{\lim_{\theta \rightarrow \infty} \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}}{\lim_{\theta \rightarrow \infty} \frac{\dot{y}_2(\theta)\theta}{y_2(\theta)}} = \frac{\lim_{\theta \rightarrow \infty} \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}}{\lim_{\theta \rightarrow \infty} \epsilon_{\theta 2}(\theta, e_2) + \frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}}. \quad (\text{A.11})$$

Using the fact that  $\lim_{\theta \rightarrow \infty} \tau_L^2(\theta) = \bar{\tau}_L^2 < 1$ , and using  $(1 - \bar{\tau}_L^2) = \frac{h'(l_2(\theta))}{u'(c(\theta))w_2(\theta, e)}$ , and by L'Hôpital's rule it is possible to get the following result

$$\begin{aligned} 1 &= \frac{\lim_{\theta \rightarrow \infty} \frac{h'(l_2(\theta))}{u'(c(\theta))w_2(\theta, e)}}{1 - \bar{\tau}_L^2} = \frac{1}{1 - \bar{\tau}_L^2} \lim_{\theta \rightarrow \infty} \frac{h''(l_2(\theta))\dot{l}_2(\theta)}{u''(c(\theta))\dot{c}(\theta)w_2(\theta, e) + u'(c(\theta))\frac{\partial w_2}{\partial \theta}}, \\ &= \frac{1}{1 - \bar{\tau}_L^2} \lim_{\theta \rightarrow \infty} \frac{h'(l_2(\theta))}{u'(c(\theta))w_2(\theta, e)} \frac{\frac{h''(l_2(\theta))\dot{l}_2(\theta)}{h'(l_2(\theta))}}{\frac{u''(c(\theta))\dot{c}(\theta)}{u'(c_2(\theta))} + \frac{\partial w_2}{\partial \theta} \frac{1}{w_2(\theta, 2)}} \\ &= \frac{1}{1 - \bar{\tau}_L^2} \lim_{\theta \rightarrow \infty} \frac{h'(l_2(\theta))}{u'(c(\theta))w_2(\theta, e)} \lim_{\theta \rightarrow \infty} \frac{\alpha(\theta) \frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}}{\epsilon_{\theta 2}(\theta, e_2) - \eta(\theta) \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}} \\ &= \lim_{\theta \rightarrow \infty} \frac{\alpha(\theta) \frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}}{\epsilon_{\theta 2}(\theta, e_2) - \eta(\theta) \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}}. \end{aligned} \quad (\text{A.12})$$

From the above assumptions let  $\alpha(\theta) \rightarrow \bar{\alpha}$ ,  $\eta(\theta) \rightarrow \bar{\eta}$ , and  $\epsilon_{\theta 2}(\theta, e_2) \rightarrow \bar{\epsilon}_{\theta 2}$  as  $\theta \rightarrow \infty$ , then

if  $\lim_{\theta \rightarrow \infty} \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}$  and  $\lim_{\theta \rightarrow \infty} \frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}$  are finite (A.11) and (A.12) can be used to obtain

$$\lim_{\theta \rightarrow \infty} \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)} = \lim_{\theta \rightarrow \infty} \left( \frac{1 + \alpha(\theta)}{\eta(\theta) + \alpha(\theta)} \right) \epsilon_{\theta 2}(\theta, e_2) = \left( \frac{1 + \bar{\alpha}}{\bar{\eta} + \bar{\alpha}} \right) \bar{\epsilon}_{\theta 2}, \quad (\text{A.13})$$

$$\lim_{\theta \rightarrow \infty} \frac{\dot{l}_2(\theta)\theta}{l_2(\theta)} = \lim_{\theta \rightarrow \infty} \left( \frac{1 - \eta(\theta)}{\eta(\theta) + \alpha(\theta)} \right) \epsilon_{\theta 2}(\theta, e_2) = \left( \frac{1 + \bar{\eta}}{\bar{\eta} + \bar{\alpha}} \right) \bar{\epsilon}_{\theta 2}. \quad (\text{A.14})$$

Using assumption 2 it can be shown that these limits are finite. If either  $\lim_{\theta \rightarrow \infty} \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}$  or  $\lim_{\theta \rightarrow \infty} \frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}$  are infinite, (A.11) would imply that

$$\lim_{\theta \rightarrow \infty} \frac{\frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}}{\frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}} = 1.$$

Suppose that  $\lim_{\theta \rightarrow \infty} \left| \frac{\dot{l}_2(\theta)\theta}{l_2(\theta)} \right| = \infty$ , then by (A.12) this result follows

$$1 = \lim_{\theta \rightarrow \infty} \frac{\alpha(\theta) \frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}}{\epsilon_{\theta 2}(\theta, e_2) - \eta(\theta) \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}} = \lim_{\theta \rightarrow \infty} \frac{\alpha(\theta)}{\frac{\epsilon_{\theta 2}(\theta, e_2)}{\frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}} - \eta(\theta)} \frac{\frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}}{\frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}} = -\frac{\bar{\alpha}}{\bar{\eta}} < 0$$

which is a contradiction. This result also relies on the fact that  $\epsilon_{\theta 2}(\theta, e_2)$  was assumed to have a finite limit.

Using the result on the limit of  $\lim_{\theta \rightarrow \infty} \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}$  the behavior of  $C(\theta)$  can be characterized as  $\theta \rightarrow \infty$ . First define  $q(\tilde{x}) \equiv \eta(\tilde{x}) \frac{\dot{c}_2(\tilde{x})\tilde{x}}{c_2(\tilde{x})}$ , and rewrite  $C(\theta)$  with the assumption that the distribution is unbounded:

$$C(\theta) = \frac{\exp\left(-\int_0^\theta \frac{q(\tilde{x})}{\tilde{x}} d\tilde{x}\right) \int_\theta^\infty \exp\left(\int_0^x \frac{q(\tilde{x})}{\tilde{x}} d\tilde{x}\right) [1 - \lambda_2 u'(c_2(x))] f^2(x) dx}{1 - F^2(\theta)}.$$

By L'Hôpital's rule

$$\begin{aligned}
\lim_{\theta \rightarrow \infty} C(\theta) &= \lim_{\theta \rightarrow \infty} \frac{-\exp\left(-\int_0^\theta \frac{q(\tilde{x})}{\tilde{x}} d\tilde{x}\right) \frac{q(\theta)}{\theta} \int_\theta^\infty \exp\left(\int_0^x \frac{q(\tilde{x})}{\tilde{x}} d\tilde{x}\right) [1 - \lambda_2 u'(c_2(x))] f^2(x) dx}{-f^2(\theta)} \\
&\quad + \frac{-\exp\left(-\int_0^\theta \frac{q(\tilde{x})}{\tilde{x}} d\tilde{x}\right) \exp\left(-\int_0^\theta \frac{q(\tilde{x})}{\tilde{x}} d\tilde{x}\right) [1 - \lambda_2 u'(c(\theta))] f^2(\theta)}{-f^2(\theta)} \\
&= \lim_{\theta \rightarrow \infty} [1 - \lambda_2 u'(c(\theta))] + \lim_{\theta \rightarrow \infty} \frac{q(\theta) C(\theta) [1 - F^2(\theta)]}{\theta f^2(\theta)}.
\end{aligned}$$

And since,  $c_2(\theta) \rightarrow \infty$ ,  $u'(c_2(\theta)) \rightarrow 0$  and thus

$$\lim_{\theta \rightarrow \infty} C(\theta) = 1 + \lim_{\theta \rightarrow \infty} q(\theta) \frac{\tau_L^2(\theta)}{1 - \tau_L^2(\theta)} \frac{1}{A(\theta)}.$$

The above assumptions, (A.13) and the above result are used to obtain

$$\lim_{\theta \rightarrow \infty} C(\theta) = 1 + \frac{\bar{\eta}}{(1 + \bar{\alpha}) \bar{\epsilon}_{\theta 2}} \left( \frac{1 + \bar{\alpha}}{\bar{\eta} + \bar{\alpha}} \right) \bar{\epsilon}_{\theta 2} \lim_{\theta \rightarrow \infty} \frac{\tau_L^2(\theta)}{1 - \tau_L^2(\theta)}.$$

Because  $\lim_{\theta \rightarrow \infty} \frac{\tau_L^2(\theta)}{1 - \tau_L^2(\theta)}$  was assumed to have a finite limit,  $C(\theta)$  must have a finite limit. From assumption 2,  $A(\theta)$  also has a finite limit, using the above result on  $C(\theta)$  and the property of the log-normal distribution such that  $\lim_{\theta \rightarrow \infty} \frac{1 - F^2(\theta)}{\theta f^2(\theta)} = 0$ , which implies that  $B(\theta) \rightarrow 0$  as  $\theta \rightarrow \infty$ , then this results follows

$$\lim_{\theta \rightarrow \infty} \frac{\tau_L^2(\theta)}{1 - \tau_L^2(\theta)} = \lim_{\theta \rightarrow \infty} A(\theta) B(\theta) C(\theta) = 0.$$

The last result implies that with the log-normal distribution  $\lim_{\theta \rightarrow \infty} C(\theta) = 1$ . Now consider  $\lim_{\theta \rightarrow \infty} \frac{\tau_L^2(\theta)}{1 - \tau_L^2(\theta)} \frac{\ln \theta - \mu}{\sigma^2}$ ,

$$\begin{aligned}
\lim_{\theta \rightarrow \infty} \frac{\tau_L^2(\theta)}{1 - \tau_L^2(\theta)} \frac{\ln \theta - \mu}{\sigma^2} &= \lim_{\theta \rightarrow \infty} (1 + \alpha(\theta)) \epsilon_{\theta 2}(\theta, e_2) \lim_{\theta \rightarrow \infty} \frac{1 - F^2(\theta)}{\theta f^2(\theta)} \frac{\ln \theta - \mu}{\sigma^2} \lim_{\theta \rightarrow \infty} C(\theta), \\
&= (1 + \bar{\alpha}) \bar{\epsilon}_{\theta 2}
\end{aligned} \tag{A.15}$$

From the results above I get

$$\frac{\tau_L^2(\theta)}{1 - \tau_L^2(\theta)} \sim (1 + \bar{\alpha})\bar{\epsilon}_{\theta 2} \left( \frac{\sigma^2}{\ln \theta - \mu} \right) \text{ as } \theta \rightarrow \infty.$$

### Proof of Corollary 2.b

The proof of Corollary 2.b is almost identical to the one of Corollary 2.a. Using Assumption 2 and the wage function

$$w_2(\theta, e_2) = (\theta^{1-\rho} + e_2^{1-\rho})^{\frac{1}{1-\rho}},$$

I get

$$\lim_{\theta \rightarrow \infty} \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)} = \lim_{\theta \rightarrow \infty} \left( \frac{1 + \alpha(\theta)}{\eta(\theta) + \alpha(\theta)} \right) \epsilon_{\theta 2}(\theta, e_2) = \lim_{\theta \rightarrow \infty} \left( \frac{1 + \alpha(\theta)}{\eta(\theta) + \alpha(\theta)} \right) \left( \frac{w_2(\theta, e_2)}{\theta} \right)^{\rho-1}, \quad (\text{A.16})$$

$$\lim_{\theta \rightarrow \infty} \frac{\dot{l}_2(\theta)\theta}{l_2(\theta)} = \lim_{\theta \rightarrow \infty} \left( \frac{1 - \eta(\theta)}{\eta(\theta) + \alpha(\theta)} \right) \epsilon_{\theta 2}(\theta, e_2) = \lim_{\theta \rightarrow \infty} \left( \frac{1 - \eta(\theta)}{\eta(\theta) + \alpha(\theta)} \right) \left( \frac{w_2(\theta, e_2)}{\theta} \right)^{\rho-1}, \quad (\text{A.17})$$

since with the above CES function the wage elasticity with respect to shock is  $\epsilon_{\theta 2} = \left( \frac{w_2(\theta, e_2)}{\theta} \right)^{\rho-1}$ . A particularity of this function is that whenever  $\rho > 1$  the limit of this elasticity is zero. This implies that both  $\lim_{\theta \rightarrow \infty} \frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}$  and  $\lim_{\theta \rightarrow \infty} \frac{\dot{l}_2(\theta)\theta}{l_2(\theta)}$  have a finite limit of zero.

In this situation, the limit of  $C(\theta)$  can still be written in the following way:

$$\lim_{\theta \rightarrow \infty} C(\theta) = 1 + \left( \frac{\bar{\eta}}{\bar{\eta} + \bar{\alpha}} \right) \lim_{\theta \rightarrow \infty} \frac{\tau_L^2(\theta)}{1 - \tau_L^2(\theta)},$$

since

$$\lim_{\theta \rightarrow \infty} \frac{\frac{\dot{c}_2(\theta)\theta}{c_2(\theta)}}{A(\theta)} = \lim_{\theta \rightarrow \infty} \frac{1}{\eta(\theta) + \alpha(\theta)}.$$

The limit of  $C(\theta)$  is still finite by the assumption on the finite limit of  $\frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)}$ . However, the asymptotic properties of  $A(\theta)$  have changed since the wage elasticity goes to zero. This also does not change the following result

$$\lim_{\theta \rightarrow \infty} \frac{\tau_L^2(\theta)}{1 - \tau_L^2(\theta)} = \lim_{\theta \rightarrow \infty} A(\theta)B(\theta)C(\theta) = 0,$$

but it changes the result on  $\lim_{\theta \rightarrow \infty} \frac{\tau_L^2(\theta)}{1-\tau_L^2(\theta)} \frac{\ln \theta - \mu}{\sigma^2}$ ,

$$\begin{aligned} \lim_{\theta \rightarrow \infty} \frac{\tau_L^2(\theta)}{1 - \tau_L^2(\theta)} \frac{\ln \theta - \mu}{\sigma^2} &= \lim_{\theta \rightarrow \infty} (1 + \alpha(\theta))\epsilon_{\theta 2}(\theta, e_2) \lim_{\theta \rightarrow \infty} \frac{1 - F^2(\theta)}{\theta f^2(\theta)} \frac{\ln \theta - \mu}{\sigma^2} \lim_{\theta \rightarrow \infty} C(\theta), \\ &= 0 \times 1 \times 1 = 0. \end{aligned}$$

Notice that with the CES function with  $\rho > 1$ , this holds

$$\lim_{\theta \rightarrow \infty} \epsilon_{\theta 2}(\theta, e_2)\theta^{\rho-1} = e_2^{\rho-1}.$$

Now consider

$$\begin{aligned} \lim_{\theta \rightarrow \infty} \frac{\tau_L^2(\theta)}{1 - \tau_L^2(\theta)} \frac{\theta^{\rho-1}(\ln \theta - \mu)}{\sigma^2} &= \lim_{\theta \rightarrow \infty} (1 + \alpha(\theta))\epsilon_{\theta 2}(\theta, e_2)\theta^{\rho-1} \lim_{\theta \rightarrow \infty} \frac{1 - F^2(\theta)}{\theta f^2(\theta)} \frac{\ln \theta - \mu}{\sigma^2} \lim_{\theta \rightarrow \infty} C(\theta), \\ &= (1 + \bar{\alpha})e_2^{\rho-1}. \end{aligned} \tag{A.18}$$

Using this result

$$\frac{\tau_L^2(\theta)}{1 - \tau_L^2(\theta)} \sim (1 + \bar{\alpha}) \left(\frac{e_2}{\theta}\right)^{\rho-1} \left(\frac{\sigma^2}{\ln \theta - \mu}\right) \text{ as } \theta \rightarrow \infty.$$

### A.1.2 Implementation: No savings in equilibrium

In this section, I consider the decentralization of the optimal allocation through a tax system.

We follow Werning (September 2011) methodology which augments any given mechanism

and allows a choice over savings subject to a nonlinear savings tax. I also follow Kapicka and Neira (October 2014) where the the design of the tax on savings is first considered and than proceed to incorporate this tax in the full tax system.

Recall that the equilibrium values  $\omega(\theta^1)$  and  $\omega(\theta^2)$  from an incentive compatible allocation  $\{c, y\}$  satisfy

$$\begin{aligned}\omega(\theta^1) &= u(c(\theta^1)) - h\left(\frac{y(\theta^1)}{w_1(\theta_1, e_1)}\right) + \beta \int \omega(\theta^1, e(\theta^1), \theta_2) f^2(\theta_2) d\theta_2, \\ \omega(\theta^2) &= u(c(\theta^2)) - h\left(\frac{y(\theta^2)}{w_2(\theta_2, e(\theta^1))}\right).\end{aligned}$$

And as above, let the worker's optimization problem be

$$\begin{aligned}\omega^r(\theta_1) &= \max_{r^1} \left\{ u(c(r^1)) - h\left(\frac{y(r^1)}{w_1(\theta_1, e_1)}\right) + \beta \int \omega^{\sigma^r}(r^1, e(r^1), \theta_2) f^2(\theta_2) d\theta_2 \right\}, \\ \omega^r(r^1, \theta_2) &= \max_{r^2} \left\{ u(c(r^2)) - h\left(\frac{y(r^2)}{w_2(\theta_2, e(r^1))}\right) \right\},\end{aligned}$$

where  $r^1$  and  $r^2$  are the reports of their types given in each period to the planner. Therefore incentive compatibility implies

$$\begin{aligned}\omega^r(\theta_1) &= \omega(\theta^1), \\ \omega^r(r^1, \theta_2) &= \omega(r^1, \theta_2) = \omega(\theta^2).\end{aligned}$$

So, for any incentive compatible allocation  $\{c, y\}$ , consider these two budget constraints

$$\begin{aligned}\tilde{c}_1 + M(x_2, r^1) &\leq c(r^1), \\ \tilde{c}_2 &\leq x_2 + c(r^2),\end{aligned}$$

where  $x_2$  is after-interest savings and  $M(x_2, \sigma^1)$  is referred to as a tax function based on

after-interest savings and the worker's report of his type  $r^1$  in period 1. Again, the goal is to augment the direct mechanism by allowing the worker to save. If I suppose that the net interest rate is  $i$  then  $M(x_2, r^1) - x_2/(1+i)$  is a nonlinear tax on savings in period 1. An important part is that  $M(0, r^1) = 0$  so that a worker that decides not to save pays no taxes.

Under the augmented mechanism, the worker's problem in the first and second period is:

$$V(x_1, e_1, \theta_1) = \max_{r^1, x_2} \left\{ u(c(r^1) - M(x_2, r^1)) - h\left(\frac{y(r^1)}{w_1(\theta_1, e_1)}\right) + \beta \int V(x_2, r^1, e(r^1), \theta_2) f^2(\theta_2) d\theta_2 \right\},$$

$$V(x_2, r^1, e(r^1), \theta_2) = \max_{r^2} \left\{ u(c(r^2) + x_2) - h\left(\frac{y(r^2)}{w_2(\theta_2, e(r^1))}\right) \right\},$$

where  $x_1 = 0$ . Now define  $W(x_2, r^1, \theta_1)$  has the right hand side of the first equation above, i.e.:

$$W(x_2, r^1, \theta_1) \equiv u(c(r^1) - M(x_2, r^1)) - h\left(\frac{y(r^1)}{w_1(\theta_1, e_1)}\right) + \beta \int V(x_2, r^1, e(r^1), \theta_2) f^2(\theta_2) d\theta_2.$$

Now I impose that

$$W(x_2, r^1, \theta_1) \leq \omega(\theta_1) \quad \forall x_2, r^1, \theta_1$$

$$W(0, \theta_1, \theta_1) = \omega(\theta_1) \quad \forall \theta_1.$$

The inequality is there to ensure that in period 1, it is optimal for the individual to report truthfully,  $r^1 = \theta_1$  and not save,  $x_2 = 0$ . The equality is there to make sure that this mechanism delivers the same utility as the original mechanism. Imposing these inequalities is equivalent to imposing

$$M(x_2, r^1) \geq M^*(x_2, r^1),$$

where

$$M^*(x_2, r^1) \equiv \max_{\theta_1} \tilde{M}^*(x_2, r^1, \theta_1),$$

and

$$\tilde{M}^*(x_2, r^1, \theta_1) \equiv c(r^1) - u^{-1} \left( \omega(\theta^1) + h \left( \frac{y(r^1)}{w_1(\theta_1, e_1)} \right) - \beta \int V(x_2, r^1, e(r^1), \theta_2) f^2(\theta_2) d\theta_2 \right). \quad (\text{A.19})$$

The last definition,  $\tilde{M}^*$ , is a hypothetical tax function that ensures that individuals with shock  $\theta_1$  are indifferent to any savings or report, i.e.  $W(x_2, r^1, \theta_1) = \omega(\theta_1) \forall x_2, r^1, \theta_1$ . The innovation of Werning (September 2011) is to show that since this tax function is implausible since it would need to depend on both  $\theta_1$  and report  $r^1$ , the upper envelope over true types  $\theta_1$  can be used which gives the function  $M^*$  which is only conditioned on report  $r^1$ . This then rules out misreporting in period 1 and 2.<sup>2</sup>

Assuming  $M^*$  is differentiable at  $x_2 = 0$ , the first-order condition of the worker augmented-mechanism problem at zero savings and truth-telling, when  $R$  is the gross rate of return is

$$u'(c(\theta^1)) M_x^*(0, \theta_1) = \beta R \int \frac{\partial V(0, \theta_1, \theta_2)}{\partial x_2} f^2(\theta_2) d\theta_2,$$

and this implies that

$$M_x^*(0, \theta_1) = \beta R \frac{E(u'(c(\theta^2)))}{u'(c(\theta^1))} = \frac{1}{1 - \tau_K(\theta^1)},$$

where  $\tau_K(\theta^1)$  is the intertemporal wedge as defined above.

Now, consider the history-dependent tax system with labor income tax,  $T = (T_1(y_1), T_2(y_1, y_2))$ ,

---

<sup>2</sup>Tax function  $M^*$  is the lowest possible tax that prevents a deviation.

and savings tax ,  $M(x_2, y_1)$ , where  $y_t$  is labor income in a decentralized economy. The worker has then the following budget constraints

$$c_1 + M(x_2, y_1) \leq y_1 - T_1(y_1), \quad (\text{A.20})$$

$$c_2 \leq x_2 + y_2 - T_2(y_1, y_2), \quad (\text{A.21})$$

The worker's problem is then to maximize his lifetime utility subject to (A.20) and (A.21). The market allocation of this decentralized economy is given by vectors  $\{\tilde{c}, \tilde{y}, \tilde{x}\}$ . Following Kapicka and Neira (October 2014) it is possible to prove a version of the *taxation principle*

**Modified Taxation Principle:** If an allocation,  $\{c, y\}$  is incentive compatible, i.e. (2.8) is satisfied, then there exist a tax system such that  $M(0, y_1) = 0$  for all  $y_1$  and an allocation  $\{c, y, 0\}$  solves the worker's decentralized problem. Reciprocally, consider a tax system, composed of  $T$  and  $M$ , that is such that  $\{c, y, x_2\}$  solves the worker's decentralized problem, then the allocation is also incentive compatible.

The proof of this goes as follows, let

$$\begin{aligned} T_1(y_1(\theta^1)) &= c_1(\theta^1) - y_1(\theta^1) \\ T_2(y_1(\theta^1), y_2(\theta^2)) &= c_2(\theta^2) - y_2(\theta^2) \\ M(x_2, y_1(\theta^1)) &= M^*(x_2, \theta_1), \end{aligned} \quad (\text{A.22})$$

with  $T_1$  and  $T_2$  set very high for values of income not considered in the optimal plan such that no worker would ever choose them. Defining  $M^*$  as before and thus by construction I have

$$\omega(\theta^1) = W(0, \theta_1, \theta_1) \geq \max_{x_2} W(x_2, r^1, \theta_1) \geq W(0, r^1, \theta_1) \quad \forall r^1 \in [\underline{\theta}, \bar{\theta}],$$

and

$$\omega(\theta^2) = V(0, \theta_2, \theta_2).$$

So choosing  $\{c, y, 0\}$  yields lifetime utility  $\omega(\theta^1)$  for an individual of type  $\theta^1$ , and therefore any other choice would lead to lower or equal lifetime utility. So this implies that  $\{c, y, 0\}$  is the solution to the worker's problem.

In the opposite way, if you take a tax system  $(T, M)$ , and let  $\{c, y, x_2\}$  be the solution to the worker's decentralized problem. Then it must be that a worker with history  $\theta^2$  will prefer allocation  $\{c(\theta^2), y(\theta^2), x_2(\theta^1)\}$  over any other allocation  $\{c(\tilde{\theta}^2), y(\tilde{\theta}^2), x_2(\tilde{\theta}^2)\}$ . This implies that allocation  $\{c, y\}$  is incentive compatible.

From there it is straightforward to show that the optimal marginal tax rate on savings will be  $\frac{1}{1-\tau_K(\theta^1)}$  and that the marginal labor tax rates  $T'_1$  and  $T'_2$  evaluated at the optimal allocation must coincide with the optimal labor wedges  $\tau_L(\theta^1)$  and  $\tau_L(\theta^2)$ .

### A.1.3 Computational Information

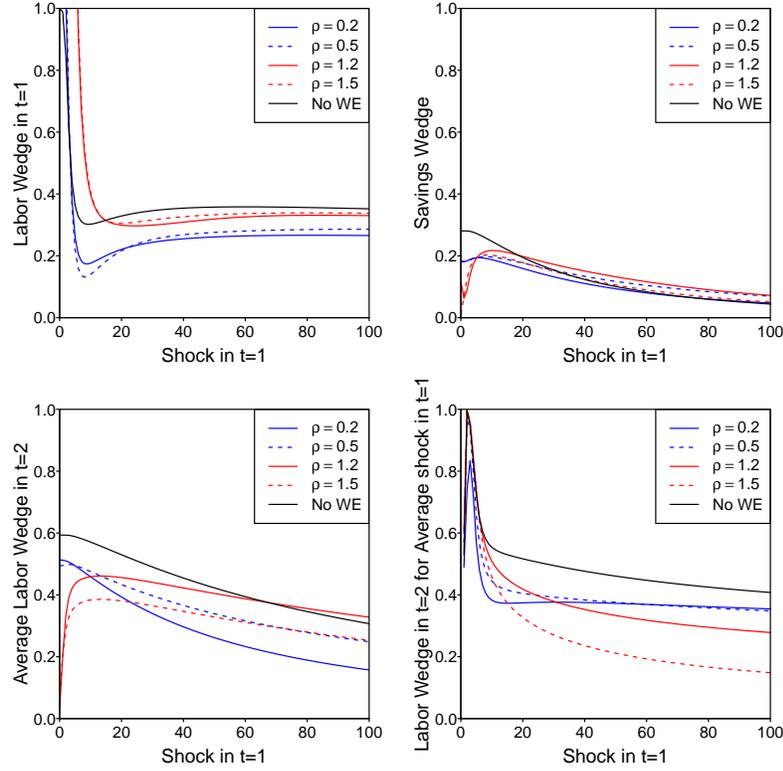
#### A.1.3.1 Further Numerical Simulations

Table A.1: Average Wedges in Each Period: Full Calibration

Parameter	$E_0(\tau_L(\theta^1))$	$E_0(\tau_L(\theta^2))$	% Lower	$E_0(\tau_K(\theta^1))$
$\rho = 0.2$	0.262	0.406	89.4%	0.161
$\rho = 0.5$	0.245	0.439	92.9%	0.175
$\rho = 1.2$	0.430	0.425	79.3%	0.188
$\rho = 1.5$	0.439	0.356	66.1%	0.166
No WE	0.365	0.533	92.9%	0.206

Source: Authors' calculations. Numbers rounded.

Figure A.1: Full Calibrations



## A.2 Optimal Income Taxation with Unemployment Induced Loss of Human Capital

### A.2.1 Microfoundations of Wage-Setting Objective

#### A.2.1.1 Nash-Bargaining

Suppose Nash Bargaining and  $M(U_i, V_i) = U_i^\gamma V_i^{1-\gamma}$ .

$$w_{ti} = \arg \max_w [w - T(w)]^\gamma [a_i - w]^{1-\gamma}$$

The FOC is:

$$\frac{\gamma}{1-\gamma} [1 - T'(w)] [a_i - w] = w - T(w).$$

With the assumption on matching function we have  $L(a_i, w_i) = \left(\frac{a_i - w_i}{\kappa_i}\right)^{\frac{1-\gamma}{\gamma}}$ .

$$w_{ti} = \arg \max_w [w - T(w)] \left(\frac{a_i - w}{\kappa_i}\right)^{\frac{1-\gamma}{\gamma}},$$

with FOC

$$\frac{\gamma}{1-\gamma} [1 - T'(w)] [a_i - w] = w - T(w).$$

Imposing that the bargaining power of the worker and the elasticity of matching be the same parameter  $\gamma$ , it follows that the wage-setting objective used in the paper can be a simple transformation of the Nash-bargaining product that results in the same solution.

### A.2.1.2 Competitive Search Equilibrium

For the following exposition, suppose that there are a continuum of skill levels and wages. Using the same notation as above, let  $\theta_{iw} = V_{iw}/U_{iw}$  be the market tightness in submarket  $(i, w)$ . Since both firms and workers are atomistic they take the market tightness and by implication the probability of matching  $m(\theta_{iw})$  for firms and  $L(a_i, w) \equiv \theta_{iw} m(\theta_{iw})$  for workers as given. In such a set up, in every submarket  $(i, w)$ , due to free-entry, the zero-profit condition holds. If not, it would be profitable to enter at that specific wage and skill market and profits would be driven to zero again. It is important to remember that when a firm enters a specific submarket at a specific wage  $w$ , it credibly commits to offer that wage to workers and can't offer another one after the match has happened.

Workers are limited to search in markets that are specific to their skill level, however they can enter one market at any posted wage  $w$ . I will ignore that participation decision of the worker and assume that the expected lifetime utility of worker that participate in a specific submarket is always greater or equal to zero. Take for example the case of a skilled worker, and consider the case of two submarket  $(a_h, w)$  and  $(a_h, w')$ . If a worker enters and

look in submarket  $(a_h, w)$ , he expects  $L(a_h, w)[w - T_1(w) + \Psi_h(w)] + \beta u_{0\ell}$ , and if enters in submarket  $(a_h, w')$ , he expects  $L(a_h, w')[w' - T_1(w') + \Psi_h(w')] + \beta u_{0\ell}$ . Let suppose that  $L(a_h, w)[w - T_1(w) + \Psi_h(w)] > L(a_h, w')[w' - T_1(w') + \Psi_h(w')]$  is true. Thus, it would never be in the interest of any worker of skill  $a_h$  to enter submarket  $(a_h, w')$ . Since firms want to attract workers and that workers will only enter the submarket where they get the highest expected surplus, a worker will only enter the submarket  $(a_h, w)$  if  $w$  maximizes  $L(a_h, w^*)[w^* - T_1(w^*) + \Psi_h(w^*)]$  in  $w^*$ . This logic also holds for unskilled workers (or any skill level considered). Therefore, when the planner is unable to observe skill levels and can only condition his policies on wages and market experience, the allocations that can be reached by the planner must be consistent with constraints

$$[w_{1\ell} - T_1(w_{1\ell}) + \beta\Psi_\ell(w_{1\ell})] \cdot L(a_\ell, w_{1\ell}) \geq [w_{1h} - T_1(w_{1h}) + \beta\Psi_\ell(w_{1h})] \cdot L(a_\ell, w_{1h}), \quad (\text{A.23})$$

$$[w_{1h} - T_1(w_{1h}) + \beta\Psi_h(w_{1h})] \cdot L(a_h, w_{1h}) \geq [w_{1\ell} - T_1(w_{1\ell}) + \beta\Psi_h(w_{1\ell})] \cdot L(a_h, w_{1\ell}), \quad (\text{A.24})$$

which are exactly (3.17) and (3.18).

### A.2.2 Proof of Corollary 3

Suppose that

$$F(\alpha_h) = \left(\frac{\alpha_h}{\bar{\alpha}}\right)^\varepsilon$$

which gives a constant elasticity of participation of  $\varepsilon$  in the skilled sector in the second period. Recall that  $\Delta u_{hh}^{LF} = w_{2h} - w_{2\ell}$ , and use (3.20) to get

$$\Delta u_{hh}^* = \frac{\varepsilon}{1 + \varepsilon - \lambda_1} \Delta u_{hh}^{LF} < \Delta u_{hh}^{LF}.$$

Also recall that when  $\Delta u_{h\ell} = 0$ ,  $\Psi_h = \int_0^{\Delta u_{hh}} [\Delta u_{hh} - \alpha_h] f_h(\alpha_h) d\alpha_h$ , thus

$$\Psi_h^{LF} = \frac{(\Delta u_{hh}^{LF})^{(\varepsilon+1)}}{\bar{\alpha}^\varepsilon[\varepsilon + 1]}, \Psi_h^* = \frac{\left[\frac{\varepsilon}{1+\varepsilon-\lambda_1} \Delta u_{hh}^{LF}\right]^{(\varepsilon+1)}}{\bar{\alpha}^\varepsilon[\varepsilon + 1]}.$$

For the skilled worker's wage to be distorted downward it is required that

$$\Psi_h^* + \frac{F(\Delta u_{hh}^*)[w_{2h} - w_{2\ell} - \Delta u_{hh}^*]}{\lambda_1} > \Psi_h^{LF}.$$

Using the definitions I have just derived using the functional form of the CDF, the above inequality is true if the following inequality is true

$$\left(\frac{\varepsilon}{1 + \varepsilon - \lambda_1}\right)^\varepsilon > \lambda_1.$$

Because  $0 < \lambda_1 < 1$ , the last inequality holds for any  $\varepsilon > 0$ . The problematic case of this inequality is when  $\lambda_1$  approaches 1 and that  $\varepsilon$  approaches 0. To show that the inequality is true, we rewrite the inequality has

$$g(\lambda_1, \varepsilon) = \exp\left\{\varepsilon \ln\left[\frac{\varepsilon}{1 + \varepsilon - \lambda_1}\right]\right\} - \lambda_1 > 0.$$

This function can be shown to be globally convex on the domain  $\varepsilon > 0, 0 < \lambda_1 \leq 1$  and that the minimums of that function are when  $\lambda_1 = 1$  (or  $\varepsilon = -1$ ). At those minimums the function is equal to 0. Thus for  $0 < \lambda_1 < 1$  and  $\varepsilon > 0$ , the function must be positive. Thus the inequality is always true.

Therefore as long as the elasticity of participation parameter is above 0 in the Maximin case, the skilled worker's wage in the first period is distorted downwards away from the efficient value.

### A.2.3 Optimal tax policy: Bergson-Samuelson case and Proof of Corollary 4.

In this section of the appendix, I only write the necessary condition for the case  $u_{1\ell} > 0$ ,  $u_{1h} > 0$ ,  $w_{1\ell} > 0$ ,  $w_{1h}$ ,  $u_{0\ell} > 0$ ,  $\Delta u_{hh} > 0$ ,  $\Delta u_{\ell\ell} = \Delta u_{h\ell} = 0$ . These are:

$$\begin{aligned} \pi_\ell W'(v_\ell) - \lambda_1 \pi_\ell - \mu \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} &= 0, \\ \lambda_1 \pi_\ell \left[ \frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} w_{1\ell} + L_\ell(w_{1\ell}) \right] - \mu u_{1\ell} \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} \left[ \frac{\partial L_h(w_{1\ell})}{\partial w_{1\ell}} \frac{1}{L_h(w_{1\ell})} - \frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} \frac{1}{L_\ell(w_{1\ell})} \right] &= 0, \\ \pi_h W'(v_h) - \lambda_1 \pi_h + \mu &= 0, \\ \pi_h W'(v_h) \beta \frac{\partial L_h(w_{1h})}{\partial w_{1h}} \Psi_h + \lambda_1 \pi_h \left[ \frac{\partial L_h(w_{1h})}{\partial w_{1h}} w_{1h} + L_h(w_{1h}) \right] \\ + \lambda_2 \pi_h \left\{ \frac{\partial L_h(w_{1h})}{\partial w_{1h}} F(\Delta u_{hh}) [w_{2h} - w_{2\ell} - \Delta u_{hh}] \right\} + \mu \beta \frac{\partial L_h(w_{1h})}{\partial w_{1h}} \Psi_h &= 0, \\ \pi_\ell W'(v_\ell) \beta + \pi_h W'(v_h) \beta - \lambda_2 &= 0, \\ [\pi_h W'(v_h) + \mu] \beta L_h(w_{1h}) \frac{\partial \Psi_h}{\partial \Delta u_{hh}} + \lambda_2 L_h(w_{1h}) \{ f(\Delta u_{hh}) [w_{2h} - w_{2\ell} - \Delta u_{hh}] - F(\Delta u_{hh}) \} &= 0. \end{aligned}$$

Using the FOC relating to  $u_{1\ell}$  and  $u_{1h}$  you get that:

$$\pi_\ell W'(v_\ell) + \pi_h W'(v_h) = \lambda_1 + \mu \left[ \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} - 1 \right] = \lambda_1 + \mu \Omega,$$

where  $\Omega = \left[ \frac{L_h(w_{1\ell})}{L_\ell(w_{1\ell})} - 1 \right] > 0$ . From this we have  $\lambda_2 = \beta[\lambda_1 + \mu\Omega]$ .

Using the FOC of  $\Delta u_{hh}$ , and using the fact that  $\pi_h W'(v_h) + \mu = \lambda_1 \pi_h$ ,  $\lambda_2 = \beta[\lambda_1 + \mu\Omega]$  and  $\frac{\partial \Psi_h}{\partial u_{hh}} = F(\Delta u_{hh})$  you get:

$$[\lambda_1 + \mu\Omega] f(\Delta u_{hh}) [w_{2h} - w_{2\ell} - \Delta u_{hh}] = [\lambda_1 + \mu\Omega] F(\Delta u_{hh}) - \lambda_1 F(\Delta u_{hh}),$$

rewriting it, you get (3.20).

For (3.21), one just needs divide both sides by  $\lambda_1 \pi_\ell L_\ell(w_{1\ell})$  and move  $\frac{\partial L_\ell(w_{1\ell})}{\partial w_{1\ell}} \frac{w_{1\ell}}{L_\ell(w_{1\ell})}$  to the right-hand side.

For (3.22), again need  $\pi_h W'(v_h) + \mu = \lambda_1 \pi_h$ ,  $\lambda_2 = \beta[\lambda_1 + \mu\Omega]$ , the FOC on  $w_{1h}$ , and rearranging we get:

$$-\frac{\partial L_h(w_{1h})}{\partial w_{1h}} \frac{w_{1h}}{L_h(w_{1h})} = 1 + \frac{\beta}{L_h(w_{1h})} \frac{\partial L_h(w_{1h})}{\partial w_{1h}} \left\{ \Psi_h + \frac{[\lambda_1 + \mu\Omega]}{\lambda_1} F(\Delta u_{hh}) [w_{2h} - w_{2\ell} - \Delta u_{hh}] \right\}.$$

Subtracting both sides by

$$\frac{\partial L_h(w_{1h})}{\partial w_{1h}} \frac{\beta \Psi_h^{LF}}{L_h(w_{1h})},$$

we get (3.22).

Suppose that

$$F(\alpha_h) = \left( \frac{\alpha_h}{\bar{\alpha}} \right)^\varepsilon.$$

Using (3.25) and the CDF, we find that

$$\Delta u_{hh}^* = \Delta u_{hh}^{LF} \frac{[\lambda_1 + \mu\Omega]\varepsilon}{[\lambda_1 + \mu\Omega]\varepsilon + \mu\Omega},$$

giving us

$$\Psi_h^{LF} = \frac{(\Delta u_{hh}^{LF})^{(\varepsilon+1)}}{\bar{\alpha}^\varepsilon [\varepsilon + 1]}, \quad \Psi_h^* = \frac{[\Delta u_{hh}^{LF} \frac{[\lambda_1 + \mu\Omega]\varepsilon}{[\lambda_1 + \mu\Omega]\varepsilon + \mu\Omega}]^{(\varepsilon+1)}}{\bar{\alpha}^\varepsilon [\varepsilon + 1]}.$$

If we suppose that the wage will be distorted downward, we need this inequality to be true

$$\Psi_h^* + \frac{[\lambda_1 + \mu\Omega]}{\lambda_1} F(\Delta u_{hh}^*) [w_{2h} - w_{2\ell} - \Delta u_{hh}^*] > \Psi_h^{LF}.$$

Using the definitions and assumptions we have made and after some manipulations, the

above inequality is true if

$$\frac{[\lambda_1 + \mu\Omega]^{(\varepsilon+1)} \varepsilon^\varepsilon}{([\lambda_1 + \mu\Omega]^\varepsilon + \mu\Omega)^\varepsilon} > \lambda_1.$$

Without knowing much more on the values of  $\mu\Omega$  other than it positive, it is impossible to know if this inequality holds for any values of  $\varepsilon$ . However, if we assume that  $\varepsilon = 1$ , i.e. the distribution function is uniform, the inequality becomes

$$\mu\Omega > 0,$$

therefore, if the incentive constraint (3.18) is binding and that  $w_{1\ell}$  is not too low such that  $L_\ell(w_{1\ell}) = L_h(w_{1\ell}) = 1$  then this is always true and thus the skilled worker's wage is distorted below the efficient level.

#### A.2.4 Numerical Simulations

All numerical simulations in this paper are made assuming the matching function has the following CES form  $M(U_i, V_i) = [U_i^{-\gamma} + V_i^{-\gamma}]^{-\frac{1}{\gamma}}$ . I also assume that the vacancy cost has the following iso-elastic shape  $k(a) = b * a^v$ . The value  $\gamma = 1.8$ ,  $b = 0.1$  and  $v = 0.3$  are chosen so that in the *Laissez-faire* the employment of both skill levels are close to 90%.

The productivity parameters and the share of the population are taken from the numerical simulations of Lee and Saez (2008) which uses a simple two type model to investigate the optimality of the minimum wage. So I take  $a_\ell = 1$  and  $a_h = 3$  to be the productivities for each type and assume that the fraction of the population of low types is  $\pi_\ell = 0.25$  and the fraction of skilled types is  $\pi_h = 0.75$ . This is natural since in our simple model, the skilled should be considered to be a majority of the population.

The benchmark model uses  $\varepsilon = 0.25$  which represents the elasticity of participation of the skilled worker in the second period. This low value is chosen because the model is much more responsive to changes in other parameter which makes getting intuition much easier.

The value of  $\bar{\alpha}$  is chosen to be 3, so that even in the *Laissez-faire* some individuals would still decide to work in the unskilled market.

I assume the following isoelastic general social welfare function as the planner's preferences :

$$SWF = \pi_\ell \left( \frac{v_\ell}{\sigma} \right)^\sigma + \pi_h \left( \frac{v_h}{\sigma} \right)^\sigma .$$

Table A.2: Characteristics of the Optimal Allocation and Tax System: Effect of Parameters

	Maximin				
	benchmark	participation		substitution	
		$\epsilon = 0.5$	$\epsilon = 0.75$	$\gamma = 1$	$\gamma = 2$
$L_\ell^*$	0.969	0.9679	0.967	0.881	0.9754
$L_h^*$	0.9361	0.9329	0.9305	0.8178	0.9469
$\% \Delta w_\ell^*$	-35.36	-26.36	-25.60	-52.41	-23.83
$\% \Delta w_h^*$	-4.34	-4.68	-4.95	-9.26	-3.78
$T_1(w_{1\ell})$	-2.1748	-2.0032	-1.8709	-1.9498	-2.2019
$T_1'(w_{1\ell})$	0.6966	0.6989	0.7009	0.6609	0.6984
$T_1(w_{1h})$	1.8178	1.5267	1.3021	1.6167	1.8423
$T_1'(w_{1h}) - \beta \Psi_h'$	0.5452	0.4902	0.4455	0.5516	0.5426
$u_{0\ell}$	1.0969	1.0471	1.0295	1.1936	1.0820
$\Delta u_{hh}^*$	1.8440	1.9158	1.9414	1.6325	1.8701
$v_\ell^*$	3.4875	3.287	3.1497	2.8337	3.5647
$v_h^*$	3.5621	3.3598	3.2211	2.9815	3.6279
$v_h/v_\ell$	1.0214	1.0221	1.0227	1.0526	1.0177

### A.3 Decision-Making In Poverty, Savings and Redistribution

#### A.3.1 Proofs of various statements

Throughout the appendix, we use  $u', u'', v', v''$  and so on to shorten the notation of first and second order derivatives. We further denote  $\tilde{R} \equiv (1 - \tau_s)(1 + r)$ ,  $\tilde{w}_i \equiv (1 - \tau_l)w_i$ ,  $c_2^* \equiv \tilde{R}s^*$  as respectively the net return on savings, the net return on wages and consumption at retirement.

### A.3.1.1 Comparative Statics of Consumption, Savings and Labor Allocations

#### A.3.1.1.1 First Sub-Period

The first-order conditions of an individual are given by:

$$0 = \tilde{R}\tilde{w}_i\beta_{t-1}v' - h' \quad (\text{A.25})$$

$$0 = u' - \tilde{R}\beta_{t-1}v' \quad (\text{A.26})$$

The hessian matrix is given by:

$$H = \begin{bmatrix} \tilde{R}^2\tilde{w}_i^2\beta_{t-1}v'' - h'' & -\tilde{R}^2\tilde{w}_i\beta_{t-1}v'' \\ -\tilde{R}^2\tilde{w}_i\beta_{t-1}v'' & u'' + \tilde{R}^2\beta_{t-1}v'' \end{bmatrix} \quad (\text{A.27})$$

The first minor is negative and the second minor is given by:

$$\begin{aligned} |H| &= \left( \tilde{R}^2\tilde{w}_i^2\beta_{t-1}v'' - h'' \right) \left( u'' + \tilde{R}^2\beta_{t-1}v'' \right) - \left( \tilde{R}^2\tilde{w}_i\beta_{t-1}v'' \right)^2, \\ &= \underbrace{\tilde{R}^2\tilde{w}_i^2\beta_{t-1}v''u'' - \tilde{R}^2\beta_{t-1}v''h'' - h''u''}_{>0}, \end{aligned}$$

which implies a global minimum.

The derivatives of  $cp^*$  and most importantly  $L^*$  are given by:

$$\frac{\partial cp^*}{\partial \tau_l} = \frac{\tilde{R}^2\beta_{t-1}v''w_iL^*}{|H|} < 0, \quad \frac{\partial cp^*}{\partial \tau_s} = \underbrace{\frac{\tilde{R}\beta_{t-1}}{(1-\tau_s)v'|H|}}_{+} \left[ 1 + \tilde{R}sp^*\frac{v''}{v'} \right], \quad (\text{A.28})$$

$$\frac{\partial L^*}{\partial \tau_l} = \frac{-w_i\tilde{R}v' + \tilde{R}^2\tilde{w}_i v'' w_i L^*}{|H|} < 0, \quad \frac{\partial L^*}{\partial \tau_s} = -\frac{\tilde{R}\beta_{t-1}\tilde{w}_i}{(1-\tau_s)v'|H|} \left[ 1 + \tilde{R}sp^*\frac{v''}{v'} \right]. \quad (\text{A.29})$$

The derivative of labor with respect to  $\tau_l$  is negative, but the derivative with respect to  $\tau_s$  cannot be signed without imposing a structural form on  $v$ . If  $v$  is iso-elastic (as in the simulations), the derivative with respect to  $\tau_s$  is then negative in the case of labor and

positive in the case of planned consumption.

### A.3.1.1.2 Effect of Time Preference on Labor Allocation

From the workers problem, we obtain:

$$\frac{dL}{d\beta_{t-1}} = \frac{w(1 - \tau_L)u''(cp)\tilde{R}v'(\tilde{R}sp)}{[u''(cp) + \beta_{t-1}\tilde{R}^2v''(\tilde{R}s)]h''(L) - \beta_{t-1}\tilde{R}^2u''(cp)v''(\tilde{R}sp)[w(1 - \tau_L)]^2} > 0. \quad (\text{A.30})$$

In the case of quasi-linearity in consumption,  $u(c) = c$ , which is considered in one of the simulations, labor is only a function of after-tax wage and not influence by inherited preferences.

### A.3.1.1.3 Second Sub-Period

The first order condition is given by:

$$0 = u' - \tilde{R}\Gamma v'. \quad (\text{A.31})$$

The hessian scalar associated with this problem is given by:

$$H = u'' + \tilde{R}^2\Gamma v'' < 0 \quad (\text{A.32})$$

and the solution is therefore a maximum. To find the derivatives of consumption with respect to tax arguments, one must take the effect of  $\tau_l$  on  $\Gamma$ :

$$\frac{\partial c_1}{\partial \tau_l} = \underbrace{\frac{\tilde{R}\beta_t}{Hv'} \frac{\partial z_i^*}{\partial \tau_l}}_+ \left[ \frac{\Gamma'}{\Gamma} + \tilde{R} \frac{v''}{v'} \right], \quad \frac{\partial c_1}{\partial \tau_s} = \underbrace{-\frac{\tilde{R}\beta_t}{Hv'}}_+ \left[ \frac{\Gamma'}{\Gamma} + \frac{c_2^*}{1 - \tau_s} \frac{v''}{v'} \right] + \frac{\tilde{R}\beta_t}{Hv'} \frac{\partial z_i^*}{\partial \tau_s} \left[ \frac{\Gamma'}{\Gamma} + \tilde{R} \frac{v''}{v'} \right] \quad (\text{A.33})$$

### A.3.1.2 Derivations of Tax Rates

#### A.3.1.2.1 Derivation of the formula for $\tau_l$ :

The budget constraint of the government is given by  $a = \tau_l Y + \tau_s S$ . If we substitute in the SWF, one finds:

$$SWF(\tau_l, \tau_s) = \left( \int_i \left[ u((1 - \tau_l)y_i^* + \tau_l Y + \tau_s S - s_i^*) + \beta^* v(\tilde{R}s_i^*) - h(y_i^*/w_i) \right]^{1-\rho} d\mu(i) \right)^{1/(1-\rho)} \quad (\text{A.34})$$

Recall that individuals do not account for the change in  $a$  when  $\tau_l$  changes, which implies that  $\frac{\partial Y}{\partial a} \frac{\partial a}{\partial \tau_l} = 0$  when changing  $\tau_l$ . Therefore, the first-order condition is given by:

$$0 = \left( \int_i \dots d\mu(i) \right)^{1/(1-\rho)-1} \int_i [\dots]^{-\rho} u'_i d\mu(i) \left[ Y + \tau_l \frac{\partial Y}{\partial \tau_l} + \tau_s \frac{\partial S}{\partial \tau_l} \right] + \dots \\ \dots + \left( \int_i \dots d\mu(i) \right)^{1/(1-\rho)-1} \int_i [\dots]^{-\rho} \left[ u'_i \left[ -y_i^* + (1 - \tau_l) \frac{\partial y_i^*}{\partial \tau_l} - \frac{\partial s_i^*}{\partial \tau_l} \right] + \beta^* \tilde{R} v' \frac{\partial s_i^*}{\partial \tau_l} - \frac{h'}{w_i} \frac{\partial y_i^*}{\partial \tau_l} \right] d\mu(i) \quad (\text{A.35})$$

Define  $g_i \equiv \frac{[u(c_i^*) + \beta^* v(\tilde{R}s_i^*) - h(L_i^*)]^{-\rho} u'(c_i^*)}{\int [u(c_j^*) + \beta^* v(\tilde{R}s_j^*) - h(L_j^*)]^{-\rho} u'(c_j^*) d\mu(j)}$  and substitute the first-order conditions of individuals to obtain:

$$0 = - \int_i g_i [Y - y_i^*] d\mu(i) - \mathbb{E}[g_i] \left[ \frac{\tau_l}{1 - \tau_l} Y \epsilon_{\tau_l}^Y + \tau_s \frac{\partial S}{\partial (1 - \tau_l)} \right] + \dots \\ \dots - \int_i g_i \left[ \frac{u'(c_i^*) - u'(cp_i^*)}{u'(c_i^*)} y_i^* \epsilon_{\tau_l}^{y_i^*} + \frac{\beta^* - \beta_i}{\beta_i} \frac{\partial s_i^*}{\partial (1 - \tau_l)} \right] d\mu(i), \quad (\text{A.36})$$

which implies:

$$\frac{\tau_l}{(1 - \tau_l)} = \frac{\int_i g_i [Y - y_i^*] d\mu(i) - \int_i g_i \left[ \frac{u'(c_i^*) - u'(cp_i^*)}{u'(c_i^*)} y_i^* \epsilon_{\tau_l}^{y_i^*} + \frac{\beta^* - \beta_i}{\beta_i} \frac{\partial s_i^*}{\partial (1 - \tau_l)} \right] d\mu(i) - \left[ \tau_s \frac{\partial S}{\partial (1 - \tau_l)} \right]}{\epsilon_Y^{\tau_l} Y}. \quad (\text{A.37})$$

### A.3.1.2.2 Derivation of the formula for $\tau_s$ :

The first order condition with respect to  $\tau_s$  is given by:

$$\begin{aligned}
0 &= \mathbb{E}[g_i] \left[ \tau_l \frac{\partial Y}{\partial \tau_s} + S + \tau_s \frac{\partial S}{\partial \tau_s} \right] + \dots \\
&\dots + \int_i [\dots]^{-\rho} \left[ u'_i \left[ (1 - \tau_l) \frac{\partial y_i^*}{\partial \tau_s} - \frac{\partial s_i^*}{\partial \tau_s} \right] - \beta^* R s_i^* v' + \beta^* (1 - \tau_s) R v' \frac{\partial s_i^*}{\partial \tau_s} - \frac{h'}{w_i} \frac{\partial y_i^*}{\partial \tau_s} \right] d\mu(i),
\end{aligned} \tag{A.38}$$

which leads to:

$$\begin{aligned}
0 &= -\mathbb{E}[g_i] \left[ \tau_l Y \epsilon_{\tau_s}^Y + \tau_s S \epsilon_{\tau_s}^S \right] + \int_i g_i \left( (1 - \tau_s) S - \frac{\beta^*}{\beta_i} s_i^* \right) d\mu(i) \dots \\
&\dots - \int_i g_i \left[ \frac{u'_i(c_i^*) - u'_i(cp_i^*)}{u'_i(c_i)} (1 - \tau_l) y_i^* \epsilon_{\tau_s}^{y_i^*} - \frac{\beta^* - \beta_i}{\beta_i} s_i^* \epsilon_{\tau_s}^{s_i^*} \right] d\mu(i),
\end{aligned} \tag{A.39}$$

and therefore:

$$\tau_s = \frac{\int_i g_i \left( S - \frac{\beta^*}{\beta_i} s_i^* \right) d\mu(i) - \int_i g_i \left[ \frac{u'_i(c_i^*) - u'_i(cp_i^*)}{u'_i(c_i)} (1 - \tau_l) y_i^* \epsilon_{\tau_s}^{y_i^*} - \frac{\beta^* - \beta_i}{\beta_i} s_i^* \epsilon_{\tau_s}^{s_i^*} \right] d\mu(i) - \tau_l Y \epsilon_{\tau_s}^Y}{(1 + \epsilon_{\tau_s}^S) S}. \tag{A.40}$$