European University Institute

DEPARTMENT OF ECONOMICS

Essays in Public Economics and Empirical Banking

Andreas Winkler

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 11 December 2018

European University Institute
Department of Economics

Essays in Public Economics and Empirical Banking

Andreas Winkler

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Examining Board

Prof. Árpád Ábrahám, EUI, Supervisor
Prof. Dominik Sachs, University of Munich
Prof. Andrés Erosa, Universidad Carlos III de Madrid
Prof. Marek Kapička, CERGE-EI

European University Institute

DEPARTMENT OF ECONOMICS

Researcher declaration to accompany the submission of written work

I, Andreas Winkler, certify that I am the author of the work *Essays in Public Economics and Empirical Banking* I have presented for examination for the PhD thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297).

The copyright of this work rests with its author. Quotation from it is permitted, provided that full acknowledgement is made. This work may not be reproduced without my prior written consent. This authorisation does not, to the best of my knowledge, infringe the rights of any third party.

Statement of inclusion of previous work (if applicable):

I confirm that chapter 3 was jointly co-authored with Matic Petricek, Carmen Garcia Galindo, and Alessandro Ferrari and I contributed 25% of the work.

Signature and Date: 28/11/2018

Acknowledgements

I would like to express my gratitude towards the EUI and the DAAD for allowing me to complete my PhD at such an exceptional institution. Especially, I am grateful towards my doctoral supervisors Árpád Ábrahám and Dominik Sachs for their support in making this thesis a reality. Their guidance and feedback have been invaluable over the past years. I would also like to thank the entire faculty of the EUI for their exceptional teaching, as well as many constructive comments on previous drafts of my work during various seminars.

I am grateful to all my friends and colleagues I got to meet during my time at the EUI. Without you, these last four years would have been so much less exciting. Special mentions belong to Björn Bremer, Lee Rüland, Alessandro Ferrari, Carmen Garcia Galindo, and Matic Petriçek, who introduced me to life outside of the Econ bubble, shamed me into running a marathon, and are the best co-authors I could have hoped for, respectively. Meeting any of you would have made this journey worthwhile by itself.

This thesis is dedicated to my parents, who taught me the curiosity and confidence to start this project as well as the necessary determination to see it through. You have always been there for me unconditionally, giving me the safety I needed to jump. Thank you for everything!

I would also like to dedicate this thesis to Hanna, who rode this roller coaster with me through all its ups and downs. Without your patience and support, all of this would have been infinitely harder.

Abstract

This thesis studies questions from the fields of public economics and empirical banking.

The first chapter studies the question of how the design of pension system affects household labour supply incentives. The effective marginal tax rate of the pension system is derived directly from the households' optimality conditions and decomposed into five components driven by inter- and intra-generational redistribution, demographics, borrowing constraints, and insurance. I provide quantitative results for the US, demonstrating that the effective tax rate lies significantly below the statutory rate. Eliminating progressivity from the US pension schedule significantly reduces effective marginal rates as it equalizes average and marginal replacement rates.

The second chapter considers the introduction of a Vickrey-style lifetime income tax in a heterogeneous agent model with idiosyncratic risk. In a model with perfect foresight, lifetime income taxation leads to unambiguous welfare gains as it redistributes resources from high-consumption to low-consumption households. A similar argument does not hold for the case with idiosyncratic risk. In a lifecycle model calibrated to the US economy, a transition to a tax on lifetime income leads to small welfare losses as workers increase their savings early in life in order to insure against uncertain future tax liabilities.

In the third chapter, we study the question how risk taking by banks responds to an exogenous change in leverage. We employ heterogeneity in the geographic distribution of banks' offices in order to introduce an exogenous variation in deposit supply based on local economic shocks. This variation is used to instrument banks' leverage. We measure bank risk taking by directly observing lending decisions on all residential mortgages in the US. In response to an exogenous decrease in their leverage, banks become more responsive to risk characteristics of residential mortgage loans, and the median predicted probability of default for issued loans decreases.

Contents

Chapter 1

Effective Marginal Tax Rates of Pension Systems

1.1 Introduction

For economists, marginal tax rates are central to understanding labor supply. A large part of the optimal tax literature is concerned with managing the trade-off between insurance and distortions to labor supply through high marginal tax rates. Following Prescott (2004), there is also a sizable literature that uses marginal tax rates in order to understand differences in labor supply patterns across countries.

However, measuring marginal tax rates properly is a more complicated problem than it appears at first glance. The reason for this is that taxes do not simply "vanish into the ether", but rather are used to finance expenditures, some of which are linked to current or past incomes. This implies, for example, a significantly higher effective marginal tax rate at the bottom of the income distribution due to benefit withdrawal, than what would be suggested by considering only the statutory tax rates.

Another area where the statutory rate is likely to be a poor estimate of the true distortion are social security contributions, especially pension taxes, as future benefits are often linked to past contributions, potentially significantly reducing the distortions below the statutory rate. To see this, consider two extreme examples. Both country A and B have a public pension system that levies a 10% payroll tax on earnings to fund this system. In country A, pensions are paid as a lump-sum transfer to all pensioners, while in country B, pensioners receive a pension that is equal to their contributions plus interest at the market rate. Thus, workers in country A see the statutory rate as a full *tax*, as their income is taxed at a 10% marginal rate, while pension benefits depend on aggregate labour supply, but do not increase with labour supply. On the other hand, workers in country B do not perceive their pension contributions as a tax at all, as the marginal tax they face is perfectly offset by the marginal increase in pension benefits. This, admittedly oversimplified, example demonstrates the importance of viewing the pension system *as a whole* when considering its impact on labor supply decisions.

In 2017, the OECD estimated that social security contributions, most of which are earmarked for the pension system, constitute 33% of the labour wedge for a US worker with average wages, and 42% for workers making less than 90% of the average wage[1]. Due to the looming solvency crisis of many pension systems, this fraction is likely to go up in the absence of fundamental reforms. Thus it is of primary importance to understand the specific distortions induced by the pension system and what steps might be taken to reduce them.

This paper demonstrates that the effective marginal tax rates of the US pension system can be directly derived from households' optimality conditions in a stochastic lifecycle model with ex-ante heterogeneity and idiosyncratic risk. Furthermore, I show that the effective tax rate can be decomposed into five intuitive components, that relate to *i*) inter-generational redistribution,

[1]OECD (2018), accessed at http://dx.doi.org/10.1787/data-00799-en

ii) intra-generational redistribution, *iii*) demographic conditions, *iv*) borrowing constraints, and *v*) the insurance provided through the pension system. This model is then calibrated to the US economy in order to quantify these net distortions, as well as their decomposition. Over the course of workers' lives, the effective marginal tax rate drops from an average of 8% at age 21 to -1% at age 63, the average effective rate over the lifecycle being significantly lower than the payroll tax of 11%, at 4%. Despite the very progressive US pension system, which sees replacement rates as high as 90% and as low as 15%, there is very little variation of effective tax rates by income, with the difference between average rates faced by high- and low-income workers staying below 2% points for most of their working lives. Furthermore, this variation is exactly opposite in sign to what one might first expect, with low income workers subject to the largest effective tax rates. Additionally, this relative homogeneity hides significant heterogeneity between income groups that can be uncovered by means of decomposing the effective tax rate into its components. While for low-income workers, borrowing constraints play a sizable role, high-income workers are subject to much higher intra-generational redistribution.

The decomposition of the effective tax rates also allows me to appraise the effects of the inherent progressivity of the pension system on labour supply in more detail. I show that this mainly impacts two of the components of the labour distortion: the intra-generational redistribution, and the insurance component. I demonstrate that the latter is quantitatively small for all agents, while the former is sizable on average. This is due to the fact that workers pay for the average replacement rate through the payroll tax, but take into account their expected *marginal* replacement rate when making labour supply decisions. Due to the strong progressivity of the current US system, the *average marginal* replacement rate lies significantly below the average replacement rate, thereby driving up the average effective tax rates.

Motivated by this finding, I consider a reform that replaces the current pension system with a linear pension system that provides the same average replacement rate. I find that such a reform would reduce the average effective marginal tax rate of the US pension system by 3% points, leading to a 1% increase in effective labour supply in the economy.

The remainder of the paper is structured as follows. The next section provides a brief overview of the related literature. In section 1.3, I demonstrate how the effective marginal tax rate can be derived from the household optimization problem in a standard lifecycle model with a realistic pension system, while section 1.4 demonstrates how the effective tax rate can be decomposed. Section 1.5 calibrates this model and discusses the current effective marginal tax rate of the US pension system quantitatively, while section 1.6 is concerned with the policy experiment. Finally, section 1.7 concludes.

1.2 Literature Review

This paper is not the first to consider the question of the net marginal tax rate associated with the social security payroll tax. Especially in the literature on the privatization of social security, authors have long cited the distortions introduced by the payroll tax as one of the ways in which a privatized system may be superior. Significant contributions to this discussion include Homburg (1990), Breyer and Straub (1993), as well as Feldstein and Samwick (1998). The first two consider social security in a simple two-period OLG economy, while the latter takes into account a more realistic lifecycle. What is common to this literature is the effective tax rate has been approximated very roughly by the differences in returns between private savings and the public pension system. While I show that this difference does indeed impact the effective tax rate of the pension system, it is not the quantitatively most important component.

Following work by Gordon (1982), Browning (1985), and Burkhauser and Turner (1985), Feldstein and Samwick (1992) were among the first to consider the net effective tax rate of social security by taking into account the specific design of the pension system. Similar to this study, they find that effective tax rates are decreasing over the lifecycle. However, they employ several simplifying assumptions, chiefly the absence of idiosyncratic risk as well as abstracting from borrowing constraints. This implies that their model-free approximation of the labor distortion significantly overstates the progressivity of the net marginal tax rate significantly. In an update to the original study, Cushing (2005) quantifies the spread between effective tax rates on high- and low-income earners to be 19%-points, which reduces to less than 3%-points when including uncertainty and borrowing constraints. Additionally, while certain reform proposals are discussed, none of the above papers carry out any sort of policy experiments based on their findings.

There are several other model-free approximations to the effective tax rate, all of which make assumptions in order to avoid dealing with uncertainty faced by workers. Goda et al. (2007) and Goda et al. (2011) assume that agents behave as if their current earnings are the last earnings in their lives when forming expectations on future benefits. This leads to effective tax rates that are increasing over this lifecycle due to the progressive design of the US pension system. When, instead, explicitly considering workers' expectations about their future labour supply, the net distortions are decreasing over the lifecycle. In another recent study on the impact of taxes and welfare benefits on labor supply in the UK, Brewer and Shaw (2018) abstract from any inter-temporal considerations (including pension entitlements) altogether.

Finally, there is a sizable quantitative literature on pension reform in OLG lifecycle models. Some of these, including Bagchi (2015), Huggett and Ventura (1999), Huggett and Parra (2010), and Brendler (2016) have included the effect of future pension benefits on labor supply. However, all of these treat the labor supply decision as a "black box", without discussing the effects of modeling decisions on effective labor distortions. As I demonstrate, assumptions on the risk faced by workers, as well as on their ability to borrow can have significant impact on the effective marginal tax rates induced by the pension system, which is in turn likely to influence any verdict on its desirability as an instrument of social insurance.

To summarize, this paper's contributions are three-fold. First, it is the first study to demonstrate that the net effective tax rates induced by the pension system can be formally derived from workers' optimality conditions in a lifecycle model with idiosyncratic risk. Second, it quantifies these net distortions in a realistic, calibrated model that allows judgment on the relative importance of influencing factors and modeling decisions. Finally, to my knowledge it is the first to implement a policy experiment aimed at reducing the effective tax rate of the pension system.

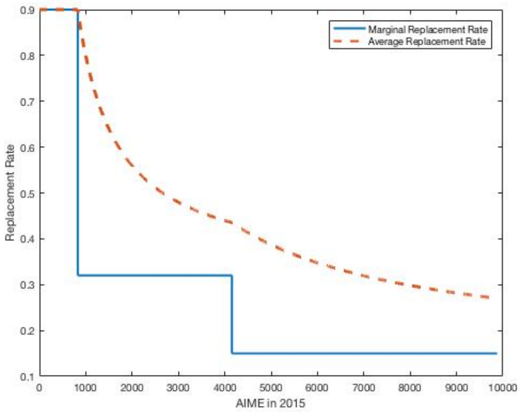
1.3 Deriving Net Effective Tax Rates

1.3.1 Institutional Framework

In the United States, pensions are administered by the Social Security Administration (SSA). Workers and firms contribute to the budget of the *Old Age Survivors and Disability Insurance* (OASDI) through a linear payroll tax on all earnings below a certain cap. In 2015, the total statutory rate was set at 6.2% for employers and workers each, coming to a total of 12.4%[2], while the cap on taxable earnings was set at $118,500$\$ (SSA (2017)). When workers retire, the SSA calculates their *Average Indexed Monthly Income* (AIME), by taking the indexed average 420 highest earning months (corresponding to 35 years). Indexing is done with the national average wage index in order to ensure that worker's contributions are not eaten up by inflation and that they profit from productivity growth during their working life. Based on the calculated AIME, the *Primary Insurance Amount* (PIA) is calculated by applying a highly progressive pension formula. In 2105, the PIA was calculated as 90% of the first 826\$ of AIME, plus 32% of the AIME between this *first bend-point*, and $4,154$\$, plus 15% of AIME above this *second bend-point* up to the cap of monthly taxable earnings at $9,875$\$. Above this cap, the marginal replacement

[2] Of these, 10.4% are earmarked for the old age pension

Figure 1.1: Average and Marginal Replacement Rate



Note: All information from the 2016 statistical supplement of the SSA (SSA (2017))

rate drops to zero. Figure 1.1 plots the average and marginal replacement rates as a function of AIME according to the 2015 SSA rules.

Since the 1977 *Social Security Amendments*, the marginal replacement rates of 90%, 32%, and 15% have stayed constant, while the two bend-points and the cap on taxable earnings has been indexed to the national average wage.

1.3.2 The Net Effective Tax Rate in an OLG Model

A key reason why economists care about the net effective rate in the first place is that it is this rate, rather than the statutory rate that measures the distortions of workers' labor supply decisions. However, instead of trying to approximate the distortions by calculating the net tax rates under simplifying assumptions, it is possible - and arguably simpler - to derive the distortions directly from a realistic model. By taking this approach, we automatically take into account the effects of uncertainty and borrowing constraints.

It is a standard result that the marginal distortion to any decision can be measured by a *wedge*. This wedge is defined as measuring the difference between an agents *marginal rate of* *substitution* (MRS) and their *marginal rate of transformation* (MRT). Formally, the wedge τ is defined as

$$\tau = 1 - \frac{MRS}{MRT}.$$

If a decision is undistorted, then an agent's marginal rates of substitution and transformation will be equalized and the wedge τ will be equal to zero. In case the decision is distorted, however, the two marginal rates will be different, leading to a non-zero wedge[3].

In the case of the labour wedges that this paper is concerned with, the relevant margin is that of exchanging leisure time for income through labor supply. Thus, the marginal rate of transformation is simply given by the *gross wage* that workers can earn in the market, given their productivity. On the other hand, their marginal rate of substitution is given by

$$MRS = \frac{u_l}{u_c},$$

where $u_i,\ i = l, c$ denotes the marginal utility of leisure and consumption, respectively. These two marginal utilities will be determined as part of workers' optimizing behavior, which is what I will use in order to derive a formulation for the net effective tax rate.

Thus, instead of trying to approximate the effective tax rate in the spirit of most of the literature, this paper will take the approach of solving for an allocation *induced* by the pension system and then using this allocation to derive the labour wedges, which equal the effective marginal tax rate. This approach is similar in spirit to the approach taken by the New Dynamic Public Finance literature (see for example Farhi and Werning (2013) and Golosov et al. (2016)) that solve for the constrained optimal allocation and use this to characterize optimal labour distortions.

[3] Note that this concept is general enough to also capture the effect of a *subsidy* which will lead to a negative wedge.

I consider a T period OLG economy with constant population growth, ex-ante heterogeneity, idiosyncratic productivity risk, and longevity risk. Workers enter the economy at age $t = 1$, work until $t = R - 1$, and then retire in period R until they die and exit the economy. They face mortality risk, and have an age specific survival probability ψ_t of surviving up until age t, conditional on having survived until age $t - 1$, and die for sure at age $T + 1$. During their working lives, they supply labor to the market, for which they are compensated at a rate θ, where θ denotes a worker's idiosyncratic productivity. At the start of their lives, each worker observes their ex-ante type θ_1, and then draws an idiosyncratic productivity shock $\theta_t,\ t \geq 2$ from a distribution $F(\theta_t|\theta_{t-1})$ in each period of their working life.

In each period of their working life, agents observe their current productivity, and make a labor-leisure, as well as a consumption-savings decision. Markets are incomplete, and workers only have access to a risk-less bond which pays an interest rate r. Upon entering retirement, agents cease to supply labor, instead only making a consumption-savings decision until they die and exit the economy. I abstract from any bequest motives, instead the government confiscates all estates and uses them to fund government consumption.

During their working life, agents are subject to a payroll tax τ on earnings up to the taxable maximum $\hat{y}$, an income tax $\mathbb{T}(y)$ on labor income, as well as a flat tax τ_a on capital income. During retirement, agents receive a pension payment $\mathbb{B}(\bar{y}_R)$, which depends on their indexed average earnings up to retirement. Pension contributions are indexed at a rate i. It should be noted, that I make one important simplifying assumption with regards to *indexing*. Specifically, I assume that the returns to pension contributions come from outside of the model instead of from wage growth. While it is possible to derive an implicit return to contributions within the model by including wage growth, this will lead to issues when solving the model, as the standard methods for re-introducing stationarity into a model with wage growth rely on the absence of income effects to the labour supply decision (see for example Fehr and Kindermann (2018)), which cannot be achieved in a setting with a general pension system. The pension schedule $\mathbb{B}$ follows the current US schedule presented in the preceding section[4]

Formally, workers solve the following problem upon observing their productivity draw

$$V_t\left(a, \bar{y}, \theta\right) = \max_{c, a', y} u\left(c, \frac{y}{\theta}\right) + \psi_{t+1} \beta \mathbb{E}_t\left[V_{t+1}\left(a', \bar{y}', \theta'\right)\right], \tag{1.1}$$

subject to the budget constraint

$$c + a' \leq y - \mathbb{T}(y) - min(y, \hat{y})\tau + \left(1 + \left(1 - \tau_a\right) r\right) a, \quad t < R \tag{1.2}$$

$$c + a' \leq \mathbb{B}(\bar{y}) + (1 + (1 - \tau_a) r) a, \quad t \geq R \tag{1.3}$$

as well as $a_0 = 0$, $a' \geq 0$ and the law of motion for the past average earnings $\bar{y}$

$$\bar{y}_{t+1} = \begin{cases} (1+i)\frac{(t-1)\bar{y}_t + min(y, \hat{y})}{t}, & t < R \\ \bar{y}_t, & t \geq R. \end{cases}$$

Here, $\mathbb{B}(\cdot)$ denotes the pension schedule as a function of past average earnings.

We can see directly from this formulation of the agents' problem that the impact of the pension system on labour supply will be two-fold. On the one hand, there is a marginal tax τ that is paid on all income below the cap, which will cause a decrease in labor supply. On the other hand, agents realize that working more now will increase their average earnings and hence lead to higher benefits during retirement.

The first order conditions of an optimizing household lead to the following inter- and intra-

[4] This formulation represents a good approximation of the current US pension system. The main difference is that in the US system, not all years of earnings count towards the calculation of the AIME, but rather the 35 years with highest earnings.

temporal optimality conditions

$$u_c(\cdot) = \beta\psi_{t+1}\left(1 + (1-\tau_a)\, r\right)\mathbb{E}_t u_c(c', l') + \mu \tag{1.4}$$

$$u_l(\cdot)\frac{1}{\theta} + u_c(\cdot)\left[1 - \mathbb{T}'(y) - \mathbb{1}_{y<\hat{y}}\tau\right] + \beta\psi_{t+1}\mathbb{E}_t\frac{\partial V_{t+1}(\cdot)}{\partial \bar{y}'}\frac{\partial \bar{y}'}{\partial y} = 0. \tag{1.5}$$

While the inter-temporal Euler Equation is not directly affected by the pension system, the intra-temporal condition now contains an extra term, compared to the case without considering the impact of labour supply on future benefits. Specifically, we will have an additional envelope condition that captures the effect of entering the next period with an additional unit of past average income.

We can use the fact that average income does not enter the agents' budget constraint until the retirement period in order to 'roll forward' the envelope condition until period R to write[5]

$$\frac{\partial V_{t+1}}{\partial \bar{y}'} = \frac{t}{R-1}(1+i)^{R-t-1}\beta^{R-t-1}\prod_{q=1}^{R-t-1}\psi_{t+q}\mathbb{E}_t\left[\mathbb{B}'(\bar{y}_R)\sum_{\tilde{s}=R}^{T}\beta^{\tilde{s}-R}\prod_{\tilde{q}=1}^{\tilde{s}-R}\psi_{R+\tilde{q}}u_c(c_{\tilde{s}})\right]. \tag{1.6}$$

Hence, the envelope condition for average income is given by the discounted marginal valuation of benefit payments throughout retirement. Substituting this into the intra-temporal optimality condition, and then in turn substituting this condition into our definition of the effective tax rate

$$\tau^t_{wedge} = 1 - \frac{MRT}{MRS},$$

yields the following result:

Proposition 1 *The effective marginal tax rate of the pension system at age t is given by*

$$\tau^t_{wedge} = \mathbb{1}_{y_t<\hat{y}}\left[\tau - \frac{1}{R-1}\beta^{R-t}(1+i)^{R-t}\sum_{\tilde{s}=R}^{T}\beta^{\tilde{s}-R}\prod_{q=1}^{\tilde{s}-t}\psi_{t+q}\mathbb{E}_t\left[\mathbb{B}'(\bar{y}_R)\frac{u_c(c_{\tilde{s}})}{u_c\left(c_t, l_t\right)}\right]\right].$$

[5] A detailed proof is given in the appendix

Thus, the effective marginal tax rate of the pension system has a formulation made up of two parts. The first term describes the instantaneous marginal tax due on any earnings below the cap. The second term, while somewhat unwieldy, denotes the discounted marginal benefits that are due to an extra unit of income in period t. Here, it is important to note that the correct discount factor is the *stochastic discount factor*, as workers weigh the marginal benefit streams during retirement with their associated marginal utilities of consumption.

In order to gain a better understanding of the above formula, it is helpful to come back to the leading example of this paper, namely the comparison of two countries, one of which pays pensions as a lump-sum benefit while the other pays pensions as contributions plus interest at the market rate. The case of the first country is simple, as we simply need to set $\mathbb{B}'(\cdot) = 0$ in order to see that any household with income below the cap on taxable earnings will see the payroll tax τ as a *full tax.* The case for the second country is easier to understand if we make some additional assumptions. Specifically, I assume that the lifecycle lasts only for two periods, there is no population growth or mortality risk, and that workers do not face any borrowing constraints. Given these assumptions, the above formula simplifies to

$$\tau^t_{wedge} = \mathbb{1}_{y_t < \hat{y}} \left[\tau - \beta(1+i)\mathbb{B}'(\bar{y}_2) \right],$$

since we can use the Euler equation to substitute in $\mathbb{E}\left[\frac{u_c(c_2)}{u_c(c_1,l_1)}\right] = 1$. Finally, since benefits are paid in a one-to-one relationship to contributions, we have that $\mathbb{B}' = \tau$, and since the pension system is indexed at the market rate, we will have

$$\tau^t_{wedge} = \mathbb{1}_{y_t < \hat{y}} \left[\tau - \beta(1+r)\tau \right] = 0,$$

leading to the anticipated 'zero marginal tax' result.

Clearly, all the above-mentioned simplifying assumptions are unlikely to hold in any realistic model of the lifecycle. Hence, I will devote the following section to decomposing the effective marginal tax rate of the pension system into its components in order to determine the factors influencing it. I will show that the effective marginal tax rate can be separated into five different components: *i*) inter-generational redistribution, *ii*) intra-generational redistribution, *iii*) a demographic component, *iv*) borrowing constraints, and *v*) the insurance value of non-linear pensions.

1.4 Decomposing Effective Tax Rates

It is instructive to start this discussion by considering a simplified version of the above model. To this end, consider an economy in which the lifecycle consists of two periods. During the first period agents work, and during the second they are retired and only consume. Agents discount the future at rate β and can save using a risk-free bond with interest rate $r = \frac{1}{\beta} - 1$. For notational simplicity, I abstract from longevity risk at this stage. Also, I assume that the payroll tax is applied to all earnings[6]. The population in this economy grows at a constant rate, with each generation being $1 + g$ times the size of the previous one.

In this simplified two-period economy, the effective marginal tax rate is given by

$$\tau_{wedge} = \tau - \beta(1+i)\frac{u_c(c_2)}{u_c(c_1, l_1)}\mathbb{B}'(y_1),$$

as in this case, first period earnings capture the entire history of earnings.

It should be noted that the remainder of this paper will employ a different view on the effective marginal tax rate of the pension system compared to the preceding section. Specifically,

[6]Both of these simplifying assumptions will be dropped when discussing the final result of the section

until now, we have compared the contemporaneous payroll tax τ with marginal benefits paid in the future. In order to better understand the effective tax rates, it is helpful to dispense with this 'dual view', instead expressing the effective marginal tax in terms of replacement rates, only. Specifically, I will use the budget balance of the pension system in order to replace the payroll tax with an expression in terms of the average replacement rate of the pension system.

Using this, we can relate the payroll tax and the pension benefits in the following manner[7]

$$(1+g)\int_{\theta_1} \tau y_1 dF(\theta_1) = \int_{\theta_1} \mathbb{B}(y_1) dF(\theta_1),$$

which immediately implies

$$\tau = \frac{1}{1+g}\frac{\bar{Y}}{\bar{\mathbb{B}}} = \frac{b}{1+g},$$

where b denotes the *average replacement rate*.

Thus, workers do not need to pay a payroll tax equal to the average replacement rate since the population is growing over time. This is exactly the effect captured by the third of our effects, the *demographic component*. The higher the population growth, the lower the payroll tax rate will have to be in order to finance a given level of benefits.

The first component of the effective tax rate, *inter-generational redistribution*, is driven by the fact that workers use two separate vehicles to save for retirement in this model. One of the vehicles, the pension system, pays a return of i, while the risk free bond that agents could use instead has a return of $r > i$. The larger this gap in returns, the larger the effective tax rate will be, as agents would have been better off by investing in the private asset instead. This can be interpreted as a form of redistribution between generations insofar as the lack of a market return is caused by the system not being pre-funded, implying that all generations are contributing

[7]Note that it is at this point that the assumption that resources for indexing of pension contributions come from outside of the model takes hold, as the right hand side of the equation contains the un-indexed earnings history rather than the indexed history.

to the windfall gain realized by the initial generation of pensioners, who were able to enjoy a pension without ever contributing to the system.

The second component of the effective tax, namely *intra-generational redistribution* is determined by the relationship between the worker's individual marginal replacement rate $\mathbb{B}'$, and the average replacement rate b in the economy. If a worker faces an individual replacement rate of exactly b, for example because the pension system is flat-rate, this component will be equal to zero, as she is not subject to any redistribution *within* her generation. If she has a replacement rate above b, this component will be negative, hence lowering the effective tax rate (and increasing it, respectively, for an individual replacement rate below b).

The fourth component of the tax, the effect of *borrowing constraints* represents the fact that pensions are only substitutes for savings in situations where workers would have chosen to save the same amount *in the absence of the pension system.* A worker who is heavily borrowing constrained will not view pension contributions as savings, but rather as a tax, since she would prefer to consume the income today. Obviously, this component will be zero for all workers who are not borrowing constrained.

The fifth and final component, the *insurance value* of the pension system will be zero for all workers in the simple two-period model, as there is no uncertainty left at the time when agents make their choices. This component will be discussed when we extend the model to three or more periods.

In order to separate out each of the four terms, we can substitute the budget balance condition into the above formula for the effective marginal tax rate, and collect terms.[8] Denoting the Lagrange multiplier on the borrowing constraint by μ, it is simple to transform the above formula for the effective tax rate in order to get the following result:

[8] Full proofs of all results in this section can be found in the appendix

Proposition 2 *In a two-period OLG economy with ex-ante heterogeneity and constant population growth g, the effective marginal tax rate of a PAYGO pension system can be additively separated into four components as*

$$\tau_{wedge} = \underbrace{\frac{r-i}{1+r}\left[\mathbb{B}'\frac{u'(c_2)}{u'(c_1)}\right]}_{\substack{\textit{inter-generational}\\ \textit{redistribution}}} + \underbrace{[b-\mathbb{B}']}_{\substack{\textit{intra-generational}\\ \textit{redistribution}}} - \underbrace{b\frac{g}{1+g}}_{\substack{\textit{demographic}\\ \textit{component}}} + \underbrace{\mathbb{B}'\frac{\mu_1}{u_c(c_1,l_1)}}_{\textit{borrowing constraint}} .$$

Considering the *inter-generational redistribution* component of the effective tax, we can see that another way to describe this would be as an opportunity cost of the pension system, as agents could have realized the same marginal benefit by investing in the private security at a higher rate. The larger the difference in rates $r-i$, the larger this component will be.

The *intra-generational redistribution* component is straightforward, as it is only driven by the difference (in multiple periods, the expected difference) between the private marginal and the average replacement rate in the economy. Here, it is important to note that the average of this component across the population will not necessarily be zero. This is due to the fact that we are comparing a marginal to an average rate. While workers will have to pay for the *average* rate through their contributions, they will only receive benefits on their last dollar earned at the *marginal* rate. If the system is highly progressive, as in the US, the average replacement rate will be above the marginal replacement rate for a large portion of the population, leading to a positive intra-generational redistribution component on average.

The *demographic component* is, in effect, a discount enjoyed by all contributors to the pension system that makes it more affordable to finance an average replacement rate of b, the higher the population growth is.

Finally, *borrowing constraints* increase the effective marginal tax due to the fact that a mandatory pension system may force some workers onto a suboptimal asset accumulation path, which will 'eat up' a fraction $\frac{\mu_1}{u_c(c_1,l_1)}$ of their private marginal benefit $\mathbb{B}'$, since a borrowing constrained worker would have chosen lower savings and more current consumption if given the opportunity.

After the decomposition of the effective tax rate in a two period model, I will now turn to a three-period model with two working periods. As previously stated, the three-period model will introduce the *insurance effect* to the effective tax rate, arising from the fact that workers may value the insurance across lifetime income histories afforded by a progressive pension system. This effect does not appear in the model with only one working period, as in this case the current income fully characterises the history of lifetime income.

In a three-period OLG model, the effective marginal tax rates will be changing with age and are given by

$$\tau^1_{wedge} = \tau - \beta^2(1+i)^2 \int_{\theta_2} \frac{u_c(c_3)}{u_c(c_1,l_1)} \frac{1}{2} \mathbb{B}'(\bar{y}) dF\theta_2$$
$$\tau^2_{wedge} = \tau - \beta(1+i) \frac{u_c(c_3)}{u_c(c_2,l_2)} \frac{1}{2} \mathbb{B}'(\bar{y})$$

Once again, we can use budget balance to express the payroll tax in terms of the average replacement rate b, as

$$\tau = b \frac{\bar{Y}}{(1+g)^2 \int_{\theta_1} y_1 dF(\theta_1) + (1+g) \int_{\theta_2} y_2 dF(\theta_2)} = b \frac{\bar{Y}}{\sum Y},$$

with $\bar{Y} = \frac{1}{2} \int_{\theta_1} y_1 dF(\theta_1) + \frac{1}{2} \int_{\theta_2} y_2 dF(\theta_2)$ being average lifetime earnings of a pensioner, and $\sum Y$ being aggregate earnings in the economy at the time of pension entry. Note that we can no longer replace this fraction with a simple function of g, as average income may differ between age groups. The principle, however, stays the same with $g > 0$ implying that $\frac{\bar{Y}}{\sum Y} < \frac{1}{2}$[9].

[9] Recall that there are now two generations contributing to the pension payments of the current old instead of just one

Once more, we substitute the budget balance condition into the above formulas for the effective tax rate, and employ the same technique as in the previous proposition.

Proposition 3 *In a three period OLG economy with constant population growth, ex-ante heterogeneity, and idiosyncratic risk, the effective marginal tax of a PAYGO pension system can be additively separated into five components as*

$$\tau^1_{wedge} = \frac{1}{2}\Bigg[\underbrace{\frac{(1+r)^2-(1+i)^2}{(1+r)^2}\mathbb{E}\left[\frac{u'(c_3)}{u'(c_1)}\mathbb{B}'\right]}_{\substack{\text{inter-generational}\\ \text{redistribution}}} + \underbrace{\mathbb{E}\left[b-\mathbb{B}'\right]}_{\substack{\text{intra-generational}\\ \text{redistribution}}} \underbrace{-b\frac{\sum\frac{1}{2}Y-\bar{Y}}{\sum\frac{1}{2}Y}}_{\substack{\text{demographic}\\ \text{component}}} + \underbrace{\frac{\mathbb{E}\left[\mu_1+\mu_2\right]}{u_c(c_1,l_1)}\mathbb{E}\left[\mathbb{B}'\right]}_{\text{borrowing constraint}} \underbrace{-Cov\left(\frac{u_c(c_3)}{u_c(c_1,l_1)},\mathbb{B}'(\bar{y})\right)}_{\text{insurance value}}\Bigg]$$

$$\tau^2_{wedge} = \frac{1}{2}\Bigg[\underbrace{\frac{r-i}{1+r}\left[\mathbb{B}'\left(1-\frac{\mu_2}{u_c(c_2,l_2)}\right)\right]}_{\substack{\text{inter-generational}\\ \text{redistribution}}} + \underbrace{\left[b-\mathbb{B}'\right]}_{\substack{\text{intra-generational}\\ \text{redistribution}}} \underbrace{-b\frac{\sum\frac{1}{2}Y-\bar{Y}}{\sum\frac{1}{2}Y}}_{\substack{\text{demographic}\\ \text{component}}} + \underbrace{\mathbb{B}'\frac{\mu_2}{u_c(c_2,l_2)}}_{\text{borrowing constraint}}\Bigg].$$

The most notable difference to the two period model is the introduction of the *insurance value* of the pension system. If the replacement rate is negatively correlated with consumption in the pension period (i.e. pensioners are receiving more marginal income in states of the world where they value it more), this will lead to an additional insurance value of the pension on top of the expected marginal replacement rate $\mathbb{B}'$. This effect only materializes if there is still uncertainty left to be resolved at the time that the labour supply decision is taken. This is why the effect is present in the first period (when the productivity draw of the second period is still uncertain), but not in the second period, when agents know their entire history upon making their labour supply decisions.

A second point of note relates to borrowing constraints. Here, it is important to recognise that not only current borrowing constraints matter, but also workers' expectations of being borrowing constrained in the future. This will be particularly important in the general model with multiple retirement periods and mortality risk, as it relates to an additional restriction of pensions not discussed previously. Since pensions are paid out as an annuity, agents cannot borrow against them during retirement. As, however, conditional survival probabilities drop sharply towards the end of agents' lives, it seems likely that they would choose to front-load consumption if given the chance. Since pensions do not provide this flexibility, there will be an effect of borrowing constraints even after all uncertainty with regards to lifetime income has been resolved.

Finally, we also need to take into account that there are now *two* generations financing benefit payments, implying that each worker will only have to pay half as much to finance the same level of benefits. Thus, in a more general model, the term pre-multiplying the respective effective tax rates will be driven by the ratio of working to retired agents. This has an immediate implication for the policy debate over retirement ages, as increasing the retirement age will decrease the ratio and hence decrease the average effective marginal tax of the pension system (keeping benefits constant).

The three period model captures all salient features of the full model, without burying any insights into the driving forces of effective tax rates below overly complicated notation. However, in order to later conduct a meaningful quantitative analysis of existing pension systems, a similar result for the full model is needed.

The underlying economy changes with regard to the three period model in two ways, both of which will lead to some notational changes which I will point out below. First we need to dispense with the assumption that pensioners are retired for only one period, as this clearly is not the case in reality. Second, any realistic lifecycle model will have to take into account longevity risk. Specifically, I return to the full model introduced in section 1.3, in which workers face ex-ante heterogeneity, idiosyncratic productivity risk, and longevity risk.

Proposition 4 *In the general model, the effective marginal tax rate of the pension system in* *period t can be additively decomposed into five components as*

$$\tau^t_{wedge} = \frac{\mathbb{1}_{y\leq\hat{y}}}{R-1}\Bigg[\textit{inter-generational redistribution} + \textit{intra-generational redistribution} - \textit{demographic component} + \textit{borrowing constraints} - \textit{insurance value}\Bigg],$$

with

$$\textit{inter-generational redistribution} = \sum_{\bar{s}=R}^{T}\left(\left[(1+(1-\tau_a)r)^{\bar{s}-t} - (1+i)^{R-t}\right]\beta^{\bar{s}-t}\Pi_{q=1}^{\bar{s}-t}\psi_{t+q}\right)\mathbb{E}\left[\frac{u'(c_{\bar{s}})}{u'(c_t)}\mathbb{B}'(y^R)\right]$$

$$\textit{intra-generational redistribution} = \sum_{\bar{s}=R}^{T}\mathbb{E}\left[b - \mathbb{B}'\left(y^R\right)\right]$$

$$\textit{demographic component} = \sum_{\bar{s}=R}^{T}\left(1-(1+g)^{-(\bar{s}-R)}\Pi_{\tilde{q}=1}^{\bar{s}-R}\psi_{R+\tilde{q}}\frac{\sum\tilde{Y}}{\sum Y}\right)b$$

$$\textit{borrowing constraints} = \sum_{\bar{s}=R}^{T}\mathbb{E}\left[\mathbb{B}'\left(y^R\right)\right]\mathbb{E}\left[\frac{u'(c_t) - (1+(1-\tau_a)r)^{\bar{s}-t}\beta^{\bar{s}-t}\Pi_{q=1}^{\bar{s}-t}\psi_{t+q}u'(c_{\bar{s}})}{u'(c_t)}\right]$$

$$\textit{insurance value} = \sum_{\bar{s}=R}^{T}(1+(1-\tau_a)r)^{\bar{s}-t}\beta^{\bar{s}-t}\Pi_{q=1}^{\bar{s}-t}\psi_{t+q}Cov\left(\frac{u'(c_{\bar{s}})}{u'(c_t)}, \mathbb{B}'\left(y^R\right)\right).$$

All of these terms are straightforward generalizations of their counterparts in the three period model, with notational differences arising from the introduction of multiple pension periods as well as mortality risk.

An important reason given for the presence of public pension systems is the presence of longevity risk, as it protects workers from out-living their savings. If workers value this insurance, we would expect it to be reflected in the marginal valuation of pension benefits and hence in the effective marginal tax rate. This expectation, it turns out, is met, as the effect of this insurance against longevity is, in fact, captured by the *demographic component.* While, at first, somewhat counter-intuitive, this makes sense when we consider that the advantage of annuities over private savings in insuring against longevity comes from the fact that, in the absence of annuities, workers would have to accumulate sufficient assets to cover the *maximum possible* lifespan. When annuities, or public pensions, are available however, workers only need to accumulate sufficient assets to cover their *expected* lifespan, as there is no aggregate longevity risk. Thus, the shorter the expected lifespan after entering retirement is (holding the maximum possible lifespan constant), the higher the demographic component will be, in turn leading to a lower effective tax rate.

Following this detailed theoretical analysis of the effective tax rate and its components, the next section will deal with their quantitative analysis. To this end, I will build a lifecycle model which will be calibrated to the US economy. Of specific interest in this section is the question how much the progressive nature of the pension system is reflected in the progressivity of the effective tax rate. Additionally, I will devote some time to discussing the relative sizes of the different components of the effective tax rate, as this will give some insight into how the distortions induced by a pension system may be reduced.

1.5 Quantitative Analysis

1.5.1 Setup

In order to analyze the quantitative importance of the pension wedges, I build a partial equilibrium OLG model with uncertain lifetime, ex-ante heterogeneity, idiosyncratic risk, endogenous labour supply, and a consumption savings decision.

Each period, a new cohort is born, whose size is $1+n$ times that of the previous cohort, with the population growth rate n being constant over time. Agents in my model live for a maximum of T periods. During their lives, they work for the first $R-1$ periods, and then retire at age R, from which time on they draw a pension and only consume. They face mortality risk throughout their lives - with the age specific survival probability ψ_{t+1} of surviving from age t to $t+1$. At age $T+1$ agents die for sure and exit the economy. During their life, agents can accumulate assets at an exogenous interest rate r, both to insure against idiosyncratic shocks to labour productivity and to save for retirement. I assume that they do not have any bequest motive.

Given that the demographic characteristics of the economy are stable over time, there will always be a fraction μ_t of age t individuals. Given the population growth rate n and the survival probabilities ψ, we can relate the relative sizes of cohorts as

$$\mu_{t+1} = \frac{\psi_{t+1}}{1+n}\mu_t.$$

For simplicity, I normalize these relative weights such that $\sum_{t=1}^{T} \mu_t = 1$.

1.5.2 The Government

The government employs three different instruments. First, there is a capital gains tax τ_a. Second, there is a non-linear income tax $\mathbb{T}(y)$ on labour income. These two taxes, together with the confiscation of accidental bequests are employed to fund some exogenous, and wasteful government spending G. Finally, the government runs a balanced social security system.

I model the social security system to be as close as possible to the current US system. Specifically, the government collects a payroll tax τ_{ssc} on all earnings below the social security cap $\hat{y}$. Contrary to the current US system, I model this tax as being exclusively levied on workers, instead of being shared between workers and firms as I am operating in partial equilibrium. During retirement, agents receive a pension as a function $\mathbb{B}$ of their average income $\bar{y}$. It is important to note that, in keeping with the current US system, only income up to the cap $\bar{y}$

counts toward calculating $\bar{y}$, i.e. that

$$\bar{y}_t = \frac{1}{min(t, R-1)} \sum_{s=1}^{min(t,R-1)} (1+i)^{min(t,R-1)-s} min(y_s, \hat{y}),$$

where i is the indexing applied to workers pension contributions. The function $\mathbb{B}$ is identical to the one currently used in the US. Writing the average income in the economy as $\hat{y}$, we have

$$\mathbb{B}_t(\bar{y}) = \begin{cases} 0.9 \cdot \bar{y}, & \bar{y} \leq 0.2 \cdot \hat{y} \\ 0.9 \cdot 0.2 \cdot \hat{y} + 0.32 \cdot (\bar{y} - 0.2 \cdot \hat{y}), & 0.2 \cdot \hat{y} < \bar{y} \leq 1.25 \cdot \hat{y} \\ 0.9 \cdot 0.2 \cdot \hat{y} + 0.32 \cdot (1.25 - 0.2) \cdot \hat{y} + 0.15 \cdot (\bar{y} - 1.25 \cdot \hat{y}), & 1.25 \cdot \hat{y} < \bar{y} \leq 2.5 \cdot \hat{y} \\ 0.9 \cdot 0.2 \cdot \hat{y} + 0.32 \cdot (1.25 - 0.2) \cdot \hat{y} + 0.15 \cdot (2.5 - 1.25) \cdot \hat{y}, & 2.5 \cdot \hat{y} < \bar{y} \end{cases}$$

for $t \geq R$, and $\mathbb{B}_t(\bar{y}) = 0$ otherwise.

1.5.3 Households

Upon being born, agents draw an initial productivity type θ_1 from the distribution $F(\theta_1)$. For $t > 1$, productivity follows a Markov process with conditional distribution $F_t(\theta_t|\theta_{t-1})$. Thus, there is both ex-ante heterogeneity between households as well as idiosyncratic risk.

During their working lives, agents are faced with wages $\omega_t = \gamma_t \theta_t$, where γ describes the common life-cycle component to productivity, while θ is the agent's idiosyncratic productivity. In retirement, the life-cycle productivity component drops to zero for all agents immediately implying an exit from the labour force. Upon observing her productivity, the agent makes a labour-leisure and a consumption-saving decision, taking the government's policy into account. They do so in order to maximize their expected discounted sum of utilities, with the flow utility in each period being given by

$$U_t = \frac{c_t^{1-\sigma}}{1-\sigma} + \xi \frac{(1-l_t)^{1-\nu}}{1-\nu},$$

where c_t denotes current consumption, l_t denotes the fraction of time devoted to the labour market, and ξ is the relative weight of consumption versus leisure in the utility function that is common to all agents.

As agents take the government's policies into account when making labour supply decisions, their information set needs to include a statistic used to form expectations on their marginal replacement rate in retirement. In order to be able to write the household problem recursively, I use the simple form of $\bar{y}$ to write a law of motion for it as

$$\bar{y}_{t+1} = \begin{cases} (1+i)\frac{(t-1)*\bar{y}_t + min(y_t,\bar{y})}{t}, & 1 < t \leq R-1 \\ \bar{y}_t, & t > R-1 \end{cases}$$

and $\bar{y}_1 = 0$. Thus, the state $\bar{y}$ describes an agent's *current* average past income that was subject to the social security payroll tax.

Denoting the value of an agent at age t by $V_t(a, \bar{y}, \theta)$, where a is their current asset holdings, $\bar{y}$ their past average income and θ their current idiosyncratic shock to productivity, we can write the household problem recursively as

$$V_t(a, \bar{y}, \theta) = \max_{c,a',l} \frac{c^{1-\sigma}}{1-\sigma} + \xi \frac{(1-l)^{1-\nu}}{1-\nu} + \beta\psi_{t+1} \int_{\theta'} V_t(a', \bar{y}', \theta') dF(\theta'|\theta) \tag{1.7}$$

subject to

$$c + a' \leq \gamma_t\theta l - \mathbb{T}(\gamma_t\theta l) - min(\gamma_t\theta l, \bar{y})\tau_{ssc} + \mathbb{B}_t(\bar{y}) + (1 + (1-\tau_a)r)a$$

$$\bar{y}_{t+1} = \begin{cases} (1+i)\frac{(t-1)*\bar{y}_t + min(y_t,\bar{y})}{t}, & t \leq R-1 \\ \bar{y}_t, & t > R-1 \end{cases}$$

as well as $a_0 = \bar{y}_0 = 0$, and the borrowing constraint $a' \geq 0$.

The result of the household's maximization problem will be two age-specific policy functions

$a'(a, \bar{y}, \theta, t)$ and $l(a, \bar{y}, \theta, t)$.

1.5.4 Equilibrium Definition

A *partial equilibrium* in the OLG economy is defined as two policy functions $a'(a, \bar{y}, \theta, t)$ and $l(a, \bar{y}, \theta, t)$, together with a tax schedule $\mathbb{T}$, a payroll tax τ_{ssc}, and a distribution $\Lambda(a, \bar{y}, \theta, t)$ such that, given the benefit schedule $\mathbb{B}$ and the interest rate r

1. The policy functions $a'(a, \bar{y}, \theta, t)$ and $l(a, \bar{y}, \theta, t)$ solve the households maximization problem (1.7)

2. The government runs a balanced budget, i.e.

$$\sum_{t=1}^{R-1} \mu_t \int \mathbb{T}(\gamma_t \theta l(a, \bar{y}, \theta, t)) d\Lambda(a, \bar{y}, \theta, t) + \sum_{t=1}^{T} \mu_{t-1} \psi_t \int r \tau_a a'(a, \bar{y}, \theta, t-1) d\Lambda(a, \bar{y}, \theta, t-1)$$
$$+ \sum_{t=1}^{T} \mu_{t-1} (1 - \psi_t) \int a'(a, \bar{y}, \theta, t-1) d\Lambda(a, \bar{y}, \theta, t) \geq G$$

3. The social security system runs a balanced budget, i.e.

$$\sum_{t=1}^{R-1} \mu_t \int min(\gamma_t \theta l(a, \bar{y}, \theta, t), \bar{y}) \tau_{ssc} d\Lambda(a, \bar{y}, \theta, t) \geq \sum_{t=R}^{T} \mu_t \int B(\bar{y}) d\Lambda(a, \bar{y}, \theta, t)$$

4. The distribution $\Lambda(a, \bar{y}, \theta, t)$ is consistent with individual choice, i.e.

$$\Lambda(\tilde{a}, \tilde{\bar{y}}, \theta', t+1) = \int_{\mathbb{X}_t} f(\theta' | \theta) d\Lambda(a, \bar{y}, \theta, t),$$

where $\mathbb{X}_t = \{(a, \bar{y}, \theta) \ s.t. \ a'(a, \bar{y}, \theta, t) = \tilde{a} \wedge \bar{y}'(l(a, \bar{y}, \theta, t)) = \tilde{\bar{y}}\}$, as well as with the initial conditions $a_0 = \bar{y}_1 = 0$.

1.5.5 Parameterization of the Model

I parameterize the model to reflect key aspects of the US economy.

Demographics: Agents enter the economy at age 21, and die for sure at age 101. I set the annual population growth rate n to 1.04%, which coincides with the US population growth rate since the $1960's$. The age specific survival probabilities are taken from the 2009 US life tables for males from Arias (2009). One period in my model corresponds to two years.

Taxation: I set the tax rate on capital income to $\tau_a = 0.2$ following Nardi et al. (2016). The level of government spending is taken from Huggett and Ventura (1999) and set at 19.5% of average income. In order to include the insurance effect of taxation into my model, I choose a non-linear income tax function $\mathbb{T}(y) = y - \lambda y^{1-\tau}$, which is frequently used in the literature and has been shown to approximate the current US tax system relatively well. I follow Heathcote et al. (2014) in setting $\tau = 0.151$ and calibrate λ to balance the governments budget. Indexing of workers' wages is done at an annual rate of 0.6%, which is consistent with the average *real* indexing of social security contributions since 1975[10]. Finally, the social security payroll tax is determined inside the model to balance the social security budget.

Productivity: For the age productivity profile, I take the age profile of earnings for US males, with the numerical values taken from Fehr and Kindermann (2018). For the idiosyncratic productivity process, I assume that $log(\theta)$ follows an $AR(1)$ process as

$$log(\theta_t) = \rho log(\theta_{t-1}) + \varepsilon_t, \ t \geq 2$$

with $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ and $log(\theta_1) \sim N(0, \sigma_0^2)$. Thus, ρ and σ_ε^2 drive the idiosyncratic risk, while σ_0^2 drives initial heterogeneity. In parameterizing this process, I follow Huggett and Ventura (1999) and set $\rho = 0.985$, $\sigma_\varepsilon^2 = 0.015$, and $\sigma_0^2 = 0.27$. For the numerical calculations, I discretise this process with $S = 19$ grid points. 18 of these are spaced evenly around zero between $-4\sigma_0^2$ and $4\sigma_0^2$, while I add an additional ultra-high income state at $5\sigma_0^2$. Finally, I rescale the state space

[10]Own calculations based on SSA data

Θ such that the average income during working life is equal to one.

Preferences: I set the risk preference parameter to $\sigma = 2$, while I set $\nu = 4$. The latter is chosen in order to yield a Frisch-elasticity of labour supply of 0.5 at the average labour supply. The relative weight on leisure in the utility function ξ is calibrated so that agents devote an average of $\bar{l} = 0.33$ of their time endowment to the labour market. The annual interest rate is set to 3.85%, which is the average US real interest rate since the 1960's (World Bank (2018)). Finally the discount factor is calibrated so that, in combination with the mortality risk, there is no systematic bias towards borrowing or saving over the lifecycle. This leads to an annual discount factor of 0.98.

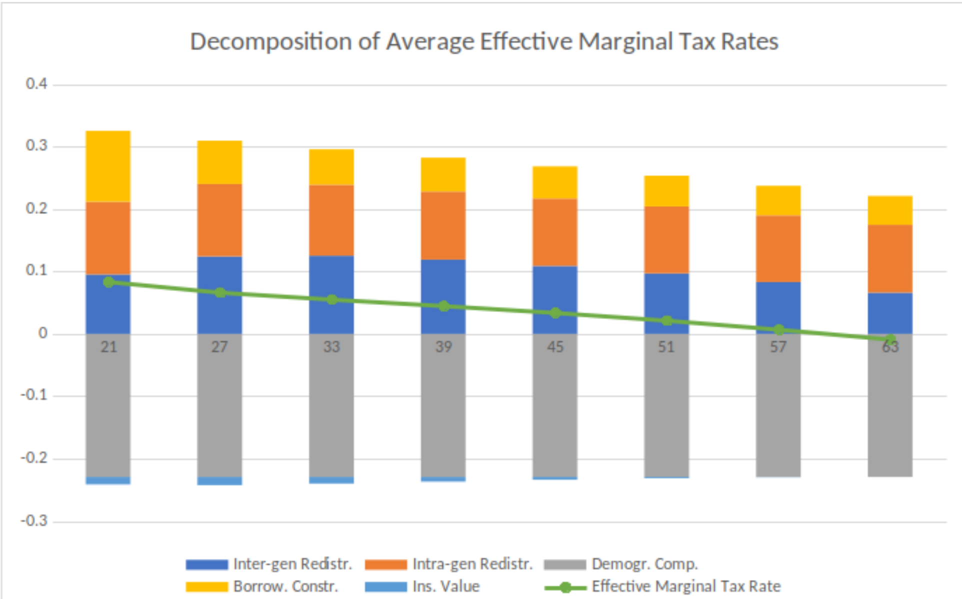
1.5.6 The Pension Wedge in the OLG Economy

Using the above baseline calibration, the equilibrium payroll tax that balances the social security budget is $\tau_{ssc} = 11.3\%$. If, however, one considers the *actual* distortion induced by the pension system, as measured by the effective marginal, the average distortion in the economy is reduced to 4.3%.

Figure 1.2 depicts the average effective tax rate over the lifecycle as well as it's decomposition. Over the course of a worker's life, the average effective tax in her cohort falls from 8.4% at age 21 to -0.8% at age 63, the last working period before retirement.

This strong decline is predominantly caused by variation in the impact of borrowing constraints over the lifecycle. As one would expect, young agents without sizable asset holdings are far more likely to be borrowing constrained, with the importance of borrowing constraints declining over the lifecycle. Even towards the end of the working life, this component does not disappear, despite the resolution of all idiosyncratic risk. This is due to the fact that agents expect to want to borrow against future pension payments in order to front-load consumption during retirement, which I do not permit in this model. As borrowing constraints decrease over

Figure 1.2: The average effective tax rate over the life-cycle



Note: Pension wedges at each age t are averaged using the distribution $\Lambda(a, \bar{y}, \theta, t)$

the working life, the private valuation of pension payments increases, leading to a larger inter-generational redistribution component, as a larger expected benefit implies larger opportunity cost of not financing this benefit through the private asset. Once borrowing constraints stabilize in the mid 40's, the importance of discounting takes over, causing the inter- generational redistribution component to fall again, as workers have successively less time up until retirement over which to discount. It does not, however, go towards zero at the end of the working life, as workers also discount over retirement years, when the pension system no longer pays any return to contributions, whereas the private asset would still pay the market return r.

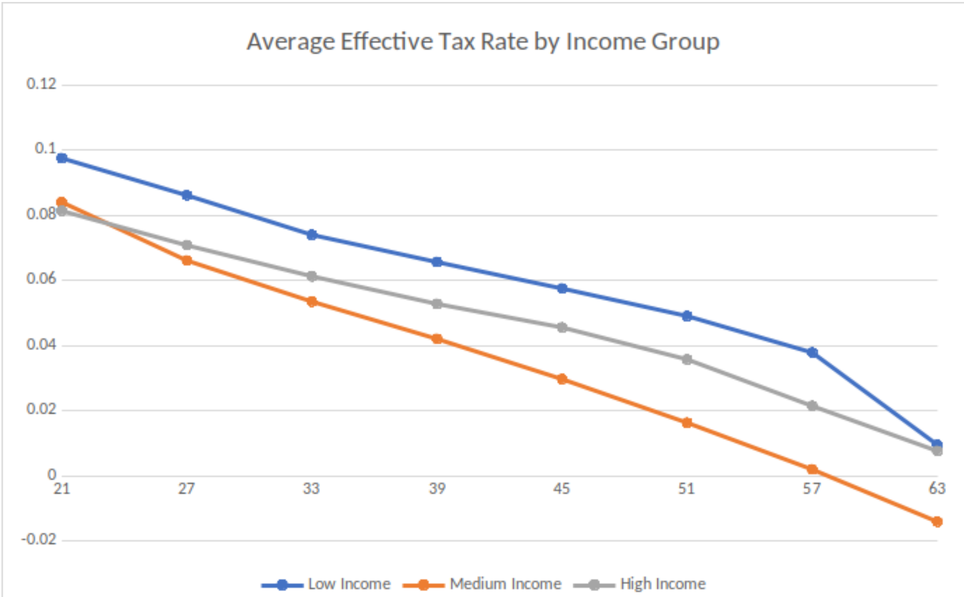
As previously discussed, the average intra-generational redistribution component is not zero, as workers compare the average replacement rate in the economy with their private expected replacement rate. Due to the progressive nature of the pension system, this difference will be positive on average, and stable over the life cycle. The second component related to redistribution, the insurance value of the progressive pension system reduces the effective tax rate slightly during the first half of the working life, and then tapers off quickly towards zero, as agents' receive more and more information about their eventual replacement rate, reducing the variance, and hence the insurance value of redistribution. It should also be noted that this insurance value is very small compared to the average intra-generational redistribution component, giving a first suggestion that, at least from the point of view of marginal tax rates, a linear pension system may have some advantages over a progressive system. Here it should be noted, however, that this only considers the desirability of *marginally more insurance*, and not the value of all insurance provided by the system. The way to consider this insurance value is as the marginal valuation agents attach to the insurance provided by the last dollar of their earnings, *given* all the insurance provided by the pension system through their average earnings.

Finally, the demographic component is sizable and reduces the effective rate by a constant amount for each cohort. As discussed previously, the component includes both the effect of population growth as well as the benefits of annuitisation as workers have to pay lower payroll taxes in order to insure themselves against longevity risk when annuities are available (or in this case provided though public pensions).

Due to the strongly progressive nature of the pension system, a natural question to ask is if the resulting effective tax rate will display a similar degree of progressivity. In order to answer this question, figure 1.3 displays the average effective tax rate by income group for each cohort. Income groups here are defined the same as in Feldstein and Samwick (1992), in that they correspond to the current 'bracket' of the pension system the worker would fall into if she were to earn this amount over her entire life.

Contrary to the previous literature, I find that there is no significant spread between the effective marginal tax rates of low-, medium-, and high-income workers. This, of course, is partially driven by the imperfect correlation of current to lifetime income. As the literature following Feldstein and Samwick (1992) does not consider idiosyncratic risk, current earnings are

Figure 1.3: The average effective tax rate over the life-cycle by current income

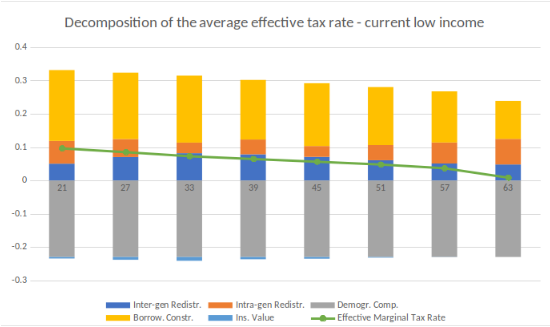


Note: Pension wedges at each age t are averaged using the conditional distribution $\Lambda(a, \bar{y}, \theta, t|y)$

the same as expected lifetime earnings. Thus, a low income worker at age 21 *knows* that they will enjoy a marginal replacement rate of 90%, while a worker in a more realistic lifecycle model will need to form expectations about their lifetime income given the presence of idiosyncratic shocks. This, however, cannot explain the second, more surprising insight, namely that the low-income workers do not, in fact, have the lowest effective marginal tax rate for most of their working lives. Given that there is at least some positive correlation between current and lifetime incomes, one would intuitively expect that low income workers should expect to receive the highest replacement rates and hence have the lowest effective marginal tax rates. However, as the graph shows, they actually have the *highest* effective rates over the entire lifecycle. Since this clearly cannot be explained by imperfect correlations between current and lifetime incomes, I will instead turn to the decompositions of these tax rates in order to gain a better understanding.

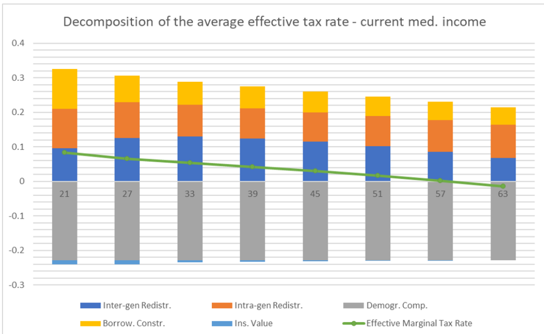
Considering figures 1.4 to 1.6, we can see that while the effective marginal tax rates may not vary strongly with income within each cohort, the decomposition thereof does vary significantly.

Figure 1.4: The average effective tax rate over the life-cycle - low income



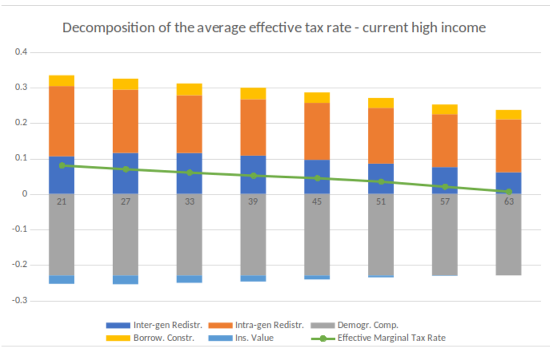
Note: Pension wedges at each age t are averaged using the conditional distribution $\Lambda(a, \bar{y}, \theta, t|y_l)$

Figure 1.5: The average effective tax rate over the life-cycle - medium income



Note: Pension wedges at each age t are averaged using the conditional distribution $\Lambda(a, \bar{y}, \theta, t|y_m)$

Figure 1.6: The average effective tax rate over the life-cycle - high income



Note: Pension wedges at each age t are averaged using the conditional distribution $\Lambda(a, \bar{y}, \theta, t|y_h)$

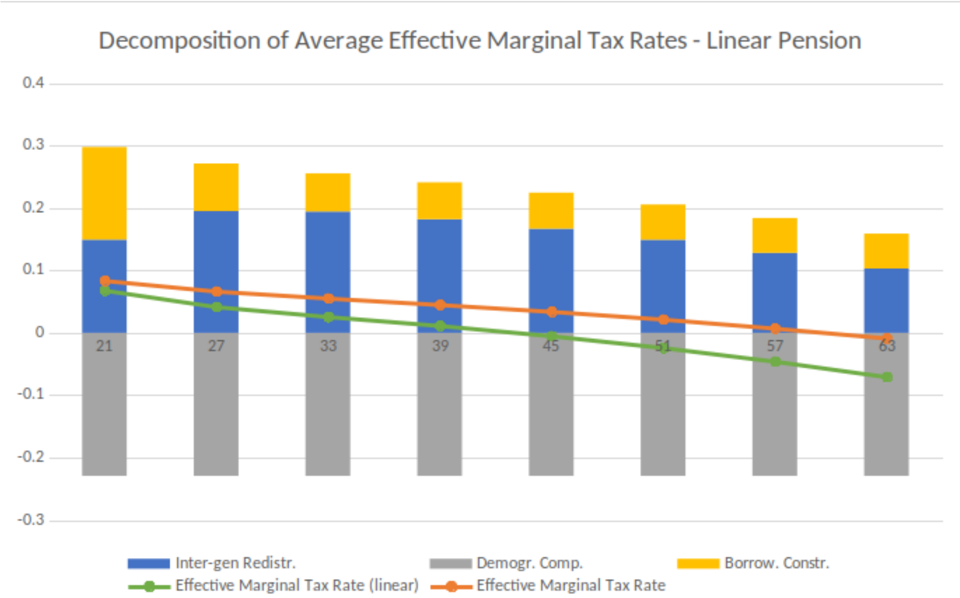
As we can see, the intra-generational redistribution component does indeed rise significantly with income. This suggests that the progressive nature of the pension system does indeed 'bleed through' into the effective marginal tax rates, as workers with high current incomes expect significantly lower marginal replacement rate than those with lower incomes. However, this effect is counteracted, and even overpowered, by the impact of borrowing constraints, which affect low-income workers much more severely than those with current medium or high incomes. High-income workers ,in particular, are almost unaffected by borrowing constraints during their working life, instead only caring about their inability to front-load consumption during retirement due to the annuity-like nature of pension payments. Finally, it should be noted that even for low-income workers, the intra-generational redistribution component is strictly positive over their entire working life, as even their expected replacement rate lies below the average replacement rate in the economy due to imperfect correlation between current and lifetime income.[11]

1.6 Policy Experiment - Equalizing Replacement Rates

The previous section demonstrates that out of the two effects of nonlinear pensions, a distributional component that, on average, increases the effective tax and an insurance component that decreases it, the distributional component is larger by far. This suggests that an intuitive reform targeted at reducing the labour supply distortion arising from the pension system would be to equalize replacement rates by replacing the current pension schedule with a flat schedule that pays the current average replacement rate to each worker regardless of income history. Intuitively, this will remove the intra-generational redistribution component, as each agent now perceives the average replacement rate as their expected marginal replacement rate.

[11] If one were to redo the above graphs in terms of current lifetime, instead of current annual income, the intra-generational redistribution component would, of course, turn negative for low lifetime income workers towards the end of their lives, as their expected marginal replacement rate would move ever closer to 90%, while the average replacement rate stays constant at 42%.

Figure 1.7: The average effective tax rate over the life-cycle - Linear Schedule



Note: Pension wedges at each age t are averaged using the distribution $\Lambda(a, \bar{y}, \theta, t)$

With the above parameterization, the average replacement rate in the economy is 41.9%. Thus, I define the new pension schedule

$$\tilde{\mathbb{B}}_t(\bar{y}) = \begin{cases} 0.419 \cdot min(\bar{y}, \hat{y}) & , \ t \geq R \\ 0 & , \ o.w. \end{cases}$$

Using the same parameterization as previously, I solve the lifecycle model again for this pension schedule. Once again, the average tax rate $1 - \lambda$ and the social security payroll tax τ_{ssc} will adjust in order to balance the social security, and government budget, respectively.

Figure 1.7 depicts the average effective tax rate of the pension system under a proportional benefit rule, as well as its decomposition. As we can see, the average effective tax rate lies below that of the current US pension system for every cohort. Averaging over all cohorts, this reform would reduce the average effective rate from 4.3% to 0.7%. This decline in the effective marginal rate causes the aggregate income in the economy to increase by 1.1%, as workers supply more efficiency units of labour to the market.

Considering the decomposition of the effective tax rate under a linear benefit rule, we find that, as expected, the intra-generational component as well as the insurance value are zero. Since the former outweighs the latter under the current system, this leads to a net decline in the effective rate. We do, however, also find that the impact of borrowing constraints, as well as of inter- generational redistribution is increased. This is simply due to the fact that both terms are multiplicative in workers' expectations of their eventual replacement rate. Since this rate increases on average when replacing the strongly progressive with a linear pension system in which everyone faces the average as their marginal replacement rate, so will the two aforementioned components. Finally, the demographic component is unchanged, as this reform will not impact any of the demographics of the model.

It should be noted, that these results do not immediately imply that such a reform would be desirable. The above analysis is only concerned with the effect of such a reform on effective marginal tax rates, and hence, on labour supply. This, however, is not the only effect of this reform, as the current progressive pension system also provides insurance to workers against bad lifetime income histories. As I have discussed previously, the insurance value under the current system represents the marginal valuation workers attach to the insurance provided by their last dollar of earned income, given the insurance already provided by their infra-marginal earnings through the pension system. Thus, moving from a currently progressive, to a linear pension system will have significant impact on consumption insurance, which is not captured by the marginal valuation of insurance.

In order to arrive at a final conclusion whether such a reform would be desirable, one would need to weigh the loss of consumption insurance over income histories against increased efficiency in the labour market, which, for the moment, I leave to future research.[12]

1.7 Conclusion

In this paper I theoretically and quantitatively investigate the distortions to labour supply induced by a pay-as-you-go pension system such as the US Social Security System.

In the first part of the paper, I demonstrate that, contrary to the approximation approaches taken by the previous literature, it is possible to calculate the effective marginal tax rates of the pension system directly from the first order conditions of an optimizing household in a heterogeneous agent lifecycle model. This approach has the advantage of allowing a decomposition of the effective tax rate into five principal components. As social security contributions pay a return less than the market rate, there is an inter-generational redistribution component, as all non-initial generations pay towards the windfall gain of the first generation of pensioners that got to enjoy the benefits of a PAYGO pension without making any contributions. Second, due to the nonlinear nature of the US pension system, there is an intra-generational redistribution component. In the presence of mortality risk and population growth, there is a demographic component to the effective marginal tax rate, which is driven by the relative measure of workers to pensioners. Fourth, since pensions and savings are only substitutes for a worker who is not borrowing constrained, there is a component driven by borrowing constraints. Finally, there is an insurance value to a nonlinear pension, as workers may value pension payments differently, depending on the income history they experienced.

In the second part of the paper, I calibrate the lifecycle to the US economy and quantify the effective marginal tax rates of the US pension system. The average effective rate is significantly

[12]A key consideration to keep in mind when conducting such a study is that removing intra-generational redistribution from the pension system will increase overall inequality in the economy. Thus, when deciding on a welfare criterion, it is important to ensure that any welfare gains or losses through this reform are not simply a result of the welfare weights, in the sense that they correct for (or exacerbate) an inefficient level of redistribution in the status quo.

below the statutory rate of 11%, at 4%. I find that, despite the highly progressive nature of the US pension system, the effective tax rates do not exhibit the large spread that has been suggested in the previous literature. Indeed, effective marginal rates are non-monotone in income with low income households facing the highest effective tax rate and medium income household the lowest. This relatively narrow spread, however, does not imply that there are no differences by income groups, as the decomposition of the effective marginal tax rate reveals significant heterogeneity. While for low income workers borrowing constraints play a very important role in determining the effective tax rate, the effective rate of high income workers is driven, to a large extend, by intra-generational redistribution. Finally, I note that the average insurance value is quantitatively insignificant, especially when compared to the second component that is directly driven by the progressive structure of the pension system, the intra-generational redistribution component. Since the latter increases the effective tax rate, on average, significantly more than the former reduces it, this suggests that the net effect of the progressive structure of the pension system on labour supply may be negative.

Guided by this last insight, I investigate the effects of a reform that replaces the current, progressive, pension schedule with a linear schedule paying the same average replacement rate to all workers. I find that this reform reduces the average effective tax rate of the pension system by more than 3% points, leading to an increase in effective labour supply of 1% over the status quo.

It is important to once again reiterate that this finding alone is not sufficient to make a convincing case for a reform of the pension system in this spirit. This is due to the fact that removing progressivity from the system will also impact the consumption insurance enjoyed by workers over the course of their lives. In order to adequately weigh these two effects, future research is warranted.

Appendix

Proof of Proposition 1

During retirement, agents solve

$$V_t\left(a, \bar{y}_R\right) = \max_{c,a'} u(c) + \beta\psi_{t+1} V_{t+1}\left(a', \bar{y}_R\right) \tag{1.8}$$

subject to

$$c + a' \leq \mathbb{B}\left(\bar{y}_R\right) + \left(1 + \left(1 - \tau_a\right) r\right) a \tag{1.9}$$

as well as $a' \geq 0$ and $V_{T+1}(\cdot, \cdot) = 0$.

The resulting first order conditions are

$$u_c(\cdot) - \lambda = 0 \tag{1.10}$$

$$-\lambda + \beta\lambda' + \mu = 0, \tag{1.11}$$

where λ denotes the multiplier of the budget constraint, μ denotes the multiplier on the borrowing constraint and the second equation uses the envelope condition for a'.

This leads to the standard inter-temporal optimality condition

$$u_c(c) \geq \psi_{t+1}\beta(1 + (1 - \tau_a))ru_c(c'), \tag{1.12}$$

which holds with equality when the borrowing constraint does not bind. There is no more risk (except for mortality) in this part of the agents' life and hence solving this model is equivalent to solving a standard cake eating problem with the trivial extension that agents receive an additional portion of cake (their pension) in each period. Note that upon entering retirement, the statistic of lifetime income stops updating and the relevant state is $\bar{y}_R$ for all retirement periods.

During working life, agents solve

$$V_t(a, \bar{y}, \theta) = \max_{c,a',y} u(c, l) + \psi_{t+1}\beta\mathbb{E}_t\left[V_{t+1}(a', \bar{y}', \theta')\right], \tag{1.13}$$

subject to the budget constraint

$$c + a' \leq y - \mathbb{T}(y) - min(y, \bar{y})\tau + (1 + (1 - \tau_a) r) a \tag{1.14}$$

as well as $a_0 = \bar{y}_0 = 0$, $a' \geq 0$, $l = \frac{y}{\theta}$, and the law of motion for $\bar{y}$

$$\bar{y}_{t+1} = \begin{cases} (1+i)\frac{(t-1)\bar{y}_t + min(y,\hat{y})}{t}, & t \leq R-1 \\ \bar{y}_t, & t > R. \end{cases}$$

The first order conditions of this problem are

$$u_c(\cdot) - \lambda = 0 \tag{1.15}$$

$$-\lambda + \beta\psi_{t+1}\mathbb{E}_t\frac{\partial V_{t+1}(\cdot)}{\partial a'} + \mu = 0 \tag{1.16}$$

$$u_l(\cdot)\frac{1}{\theta} + \lambda\left[1 - \mathbb{T}'(y) - \mathbb{1}_{y<\bar{y}}\tau\right] + \beta\psi_{t+1}\mathbb{E}_t\frac{\partial V_{t+1}(\cdot)}{\partial \bar{y}'}\frac{\partial \bar{y}'}{\partial y} = 0. \tag{1.17}$$

Substituting in for the multiplier on the budget constraint and using the envelope condition $\frac{\partial V_{t+1}(\cdot)}{\partial a'} = (1 + (1 - \tau_a) r) \lambda'$, the system of equations becomes

$$u_c(\cdot) = \beta\psi_{t+1}(1 + (1 - \tau_a) r) \mathbb{E}_t u_c(c', l') + \mu \tag{1.18}$$

$$u_l(\cdot)\frac{1}{\theta} + u_c(\cdot)\left[1 - \mathbb{T}'(y) - \mathbb{1}_{y<\bar{y}}\tau\right] + \beta\psi_{t+1}\mathbb{E}_t\frac{\partial V_{t+1}(\cdot)}{\partial \bar{y}'}\frac{\partial \bar{y}'}{\partial y} = 0. \tag{1.19}$$

While the inter-temporal optimality condition is not affected by the presence of the pension system, the optimal labour- leisure choice now has an added inter-temporal component. This is because current income affects the statistic $\bar{y}$, which in turn affects the pension to be paid during retirement. In order to make this second formula more intuitive, consider the envelope condition on $\bar{y}$. For working periods $t = 1, ..., R-1$, we have $\frac{\partial V_t}{\partial \bar{y}} = (1+i)\beta\psi_{t+1}\mathbb{E}_t\frac{\partial V_{t+1}}{\partial \bar{y}'}\frac{t-1}{t}$, because $\bar{y}$ does not enter agents' budget constraints in pre-retirement periods. Continuing to use this, we can forward substitute this equation into itself until we reach the first period of retirement, i.e we have

$$\frac{\partial V_t}{\partial \bar{y}} = (1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\mathbb{E}_t\frac{\partial V_R}{\partial \bar{y_R}}\frac{t-1}{R-1}. \tag{1.20}$$

On the other hand, during a retirement period $t = R, ..., T$, the envelope condition will be given by

$$\begin{aligned}\frac{\partial V_t}{\partial \bar{y}} &= \lambda\mathbb{B}(\bar{y})' + \beta\psi_{t+1}\frac{\partial V_{t+1}}{\partial \bar{y}'}\frac{\partial \bar{y}'}{\partial \bar{y}} \\ &= \lambda\mathbb{B}'(\bar{y}) + \beta\psi_{t+1}\lambda'\mathbb{B}'(\bar{y}')\frac{\partial \bar{y}'}{\bar{y}} + \beta^2\psi_{t+1}\psi_{t+2}\frac{\partial V_{t+2}}{\partial \bar{y}''}\frac{\partial \bar{y}''}{\partial \bar{y}'}\frac{\partial \bar{y}'}{\partial \bar{y}} \\ &= ...\end{aligned} \tag{1.21}$$

Now, using the fact that during retirement, there is no updating to the state $\bar{y}$, i.e. $\frac{\partial \bar{y}'}{\partial \bar{y}} = 1$, and using (1.10), we can write this as

$$\frac{\partial V_t}{\partial \bar{y}} = \mathbb{B}'(\bar{y})\sum_{s=t}^{T}\beta^{s-t}\prod_{q=1}^{s-t}\psi_{t+q}u_c(c_s), \tag{1.22}$$

and hence

$$\frac{\partial V_R}{\partial \bar{y}} = \mathbb{B}'(\bar{y})\sum_{s=R}^{T}\beta^{s-R}\prod_{q=1}^{s-R}\psi_{R+q}u_c(c_s). \tag{1.23}$$

Substituting this into (1.20), we get the full envelope condition for $\bar{y}$, namely

$$\frac{\partial V_t}{\partial \bar{y}} = \frac{t-1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\mathbb{E}_t\left[\mathbb{B}'(\bar{y}_R)\sum_{\tilde{s}=R}^{T}\beta^{\tilde{s}-R}\prod_{\tilde{q}=1}^{\tilde{s}-R}\psi_{R+\tilde{q}}u_c(c_{\tilde{s}})\right]. \tag{1.24}$$

While this may not look particularly intuitive at first sight, it has a lot of parallels to the envelope condition for assets. Primarily, both depend crucially on the marginal utility of consumption in future periods. This should not come as a surprise to readers as pensions can be interpreted as forced savings.

I measure labour distortions as the wedge between the marginal rate of substitution between labour and leisure and the marginal rate of transformation. I.e. I define

$$\tilde{\tau}^t_{wedge} = 1 - \frac{MRS}{MRT},$$

which in my context is given by

$$\tilde{\tau}^t_{wedge} = 1 - \frac{u_l(\cdot)}{u_c(\cdot)\theta}.$$

Substituting the first order condition (1.5), the envelope condition for $\bar{y}$ (1.24) into this formula, and using the fact that $\frac{\partial \bar{y}'}{\partial y} = \frac{1+i}{t}$, will yield the effective marginal tax rate as

$$\tilde{\tau}^t_{wedge} = \mathbb{T}'(y_t) + \mathbb{1}_{y_t<\bar{y}}\left[\tau - \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\mathbb{E}_t\left[\mathbb{B}'(\bar{y}_R)\sum_{\tilde{s}=R}^{T}\beta^{\tilde{s}-R}\prod_{\tilde{q}=1}^{\tilde{s}-R}\psi_{R+\tilde{q}}\frac{u_c(c_{\tilde{s}})}{u_c(c_t,l_t)}\right]\right]. \tag{1.25}$$

Thus, the effective marginal tax rate of workers can be separated additively into a tax and a pension component, the second one of which we denote by

$$\tau^t_{wedge} = \mathbb{1}_{y_t<\bar{y}}\left[\tau - \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\mathbb{E}_t\left[\mathbb{B}'(\bar{y}_R)\sum_{\tilde{s}=R}^{T}\beta^{\tilde{s}-R}\prod_{\tilde{q}=1}^{\tilde{s}-R}\psi_{R+\tilde{q}}\frac{u_c(c_{\tilde{s}})}{u_c(c_t,l_t)}\right]\right], \tag{1.26}$$

which gives the required result.

Proof of Proposition 2

The effective marginal tax rate of the pension system in the two period model is given by

$$\tau_{wedge} = \tau - \frac{1+i}{1+r}\frac{u_c(c_2)}{u_c(c_1, l_1)}\mathbb{B}'(y_1).$$

Replacing the payroll tax with its expression in terms of the average replacement rate, and substituting in the Euler equation $u_c(c_1, l_1) = u_c(c_2) + \mu$ yields

$$\begin{aligned}
\tau_{wedge} &= b\frac{1}{1+g} - \frac{1+i}{1+r}\mathbb{B}'(y_1) + \frac{1+i}{1+r}\mathbb{B}'(y_1) - \frac{1+i}{1+r}\frac{u_c(c_2)}{u_c(c_2)+\mu}\mathbb{B}'(y_1) \\
&= b\frac{1}{1+g} - \frac{1+i}{1+r}\mathbb{B}'(y_1) + \frac{1+i}{1+r}\frac{\mu}{u_c(c_2)+\mu}\mathbb{B}'(y_1) \\
&= b\frac{1}{1+g} - \frac{1+i}{1+r}[\mathbb{B}' - b + b] + \frac{1+i}{1+r}\frac{\mu}{u_c(c_2)+\mu}\mathbb{B}'(y_1) \\
&= b\frac{1}{1+g} - \frac{1+i}{1+r}b + \frac{1+i}{1+r}[b - \mathbb{B}'] + \frac{1+i}{1+r}\frac{\mu}{u_c(c_2)+\mu}\mathbb{B}'(y_1) \\
&= b\left[\frac{1}{1+g} - 1 + 1 - \frac{1+i}{1+r}\right] + \frac{1+i}{1+r}[b - \mathbb{B}'] + \frac{1+i}{1+r}\frac{\mu}{u_c(c_2)+\mu}\mathbb{B}'(y_1) \\
&= b\frac{r-i}{1+r} + \frac{1+i}{1+r}\left(b - \mathbb{B}'\right) - b\frac{g}{1+g} + \frac{1+i}{1+r}\mathbb{B}'\frac{\mu}{u_c(c_2)+\mu},
\end{aligned}$$

Finally, we want to remove the time discounting effects from the term for intra-generational redistribution and the borrowing constraints. We can do this by employing exactly the same technique as above to get

$$\begin{aligned}\tau_{wedge} &= b\frac{r-i}{1+r} + \frac{1+i}{1+r}(b-\mathbb{B}') - (b-\mathbb{B}') + (b-\mathbb{B}') - \\ &\quad b\frac{g}{1+g} + \frac{1+i}{1+r}\mathbb{B}'\frac{\mu}{u_c(c_2)+\mu} - \mathbb{B}'\frac{\mu}{u_c(c_2)+\mu} + \mathbb{B}'\frac{\mu}{u_c(c_2)+\mu} \\ &= \frac{r-i}{1+r}\left[\mathbb{B}'\left(1-\frac{\mu_1}{u_c(c_1,l_1)}\right)\right] + [b-\mathbb{B}'] - b\frac{g}{1+g} + \mathbb{B}'\frac{\mu_1}{u_c(c_2,l_1)} \\ &= \frac{r-i}{1+r}\left[\frac{u'(c_2)}{u'(c_1)}\mathbb{B}'\right] + [b-\mathbb{B}'] - b\frac{g}{1+g} + \mathbb{B}'\frac{\mu_1}{u_c(c_2,l_1)},\end{aligned}$$

which completes the proof.

Proof of Proposition 3

In a three-period model, the effective marginal tax rates at each age are given by

$$\begin{aligned}\tau^1_{wedge} &= \tau - (1+i)^2\beta^2\int_{\theta_2}\frac{u_c(c_3)}{u_c(c_1,l_1)}\frac{1}{2}\mathbb{B}'(\bar{y})dF\theta_2 \\ \tau^2_{wedge} &= \tau - (1+i)\beta\frac{u_c(c_3)}{u_c(c_2,l_2)}\frac{1}{2}\mathbb{B}'(\bar{y})\end{aligned}$$

Proving the result for the second period tax rate is a simple application of proposition 2, with the growth rate exchanged for the ratio of average lifetime income to total taxable income $\frac{\bar{Y}}{\sum Y}$. Hence, I will only discuss the result for the first period rate.

Replacing the payroll tax rate with its corresponding expression as a function of the average replacement rate, we get

$$\tau^1_{wedge} = b\frac{\bar{Y}}{\sum Y} - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\int_{\theta_2}\frac{u_c(c_3)}{u_c(c_1,l_1)}\mathbb{B}'(\bar{y})dF\theta_2.$$

First, we use a standard result from statistics

$$E[A \cdot B] = E[A]E[B] + Cov(A, B)$$

in order to separate the idiosyncratic replacement rate and the corresponding utility weights to get

$$\begin{aligned}
\tau^1_{wedge} = b\frac{\bar{Y}}{\sum Y} - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\int_{\theta_2}\frac{u_c(c_3)}{u_c(c_1,l_1)}dF\theta_2\int_{\theta_2}\mathbb{B}'(\bar{y})dF\theta_2 - \\
\frac{(1+i)^2}{(1+r)^2}\frac{1}{2}Cov\left(\frac{u_c(c_3)}{u_c(c_1,l_1)}, \mathbb{B}'(\bar{y})\right) \\
= b\frac{\bar{Y}}{\sum Y} - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\int_{\theta_2}\frac{u_c(c_3)}{u_c(c_1,l_1)}dF\theta_2\int_{\theta_2}\mathbb{B}'(\bar{y})dF\theta_2 + \\
\frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\int_{\theta_2}\mathbb{B}'(\bar{y})dF\theta_2 - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\int_{\theta_2}\mathbb{B}'(\bar{y})dF\theta_2 - \\
\frac{(1+i)^2}{(1+r)^2}\frac{1}{2}Cov\left(\frac{u_c(c_3)}{u_c(c_1,l_1)}, \mathbb{B}'(\bar{y})\right).
\end{aligned}$$

This step allows for a separation of the insurance premium component of the tax rate. Next, we can replace the integral over the utility weights with its corresponding expression in terms of the borrowing constraints by using the Euler equation

$$\begin{aligned}
\tau^1_{wedge} = b\frac{\bar{Y}}{\sum Y} - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\frac{\int_{\theta_2}u_c(c_3)dF\theta_2}{\int_{\theta_2}u_c(c_3)dF\theta_2 + \mu_1 + \int_{\theta_2}\mu_2 dF\theta_2}\int_{\theta_2}\mathbb{B}'(\bar{y})dF\theta_2 + \\
\frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\int_{\theta_2}\mathbb{B}'(\bar{y})dF\theta_2 - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\int_{\theta_2}\mathbb{B}'(\bar{y})dF\theta_2 - \\
\frac{(1+i)^2}{(1+r)^2}\frac{1}{2}Cov\left(\frac{u_c(c_3)}{u_c(c_1,l_1)}, \mathbb{B}'(\bar{y})\right) \\
= b\frac{\bar{Y}}{\sum Y} - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\mathbb{E}\left[\mathbb{B}'\right] + \\
\frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\frac{\mathbb{E}\left[\mu_1+\mu_2\right]}{\mathbb{E}\left[u_c(c_3)\right] + \mathbb{E}\left[\mu_1+\mu_2\right]}\mathbb{E}\left[\mathbb{B}'\right] - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}Cov\left(\frac{u_c(c_3)}{u_c(c_1,l_1)}, \mathbb{B}'(\bar{y})\right).
\end{aligned}$$

This step separates out the effect of borrowing constraints on the effective tax rate. All that remains at this stage is to separate the first line of the above expression into the inter- and intra-generational redistribution component as well as the demographic component.

To separate out the intra-generational redistribution component, we write

$$\begin{aligned}
\tau^1_{wedge} &= b\frac{\bar{Y}}{\sum Y} - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\mathbb{E}\left[\mathbb{B}'\right] + \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}b - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}b + \\
&\frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\frac{\mathbb{E}\left[\mu_1+\mu_2\right]}{\mathbb{E}\left[u_c(c_3)\right]+\mathbb{E}\left[\mu_1+\mu_2\right]}\mathbb{E}\left[\mathbb{B}'\right] - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}Cov\left(\frac{u_c(c_3)}{u_c(c_1,l_1)},\mathbb{B}'(\bar{y})\right) \\
&= b\frac{\bar{Y}}{\sum Y} + \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\left[b-\mathbb{E}\left[\mathbb{B}'\right]\right] - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}b + \\
&\frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\frac{\mathbb{E}\left[\mu_1+\mu_2\right]}{\mathbb{E}\left[u_c(c_3)\right]+\mathbb{E}\left[\mu_1+\mu_2\right]}\mathbb{E}\left[\mathbb{B}'\right] - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}Cov\left(\frac{u_c(c_3)}{u_c(c_1,l_1)},\mathbb{B}'(\bar{y})\right).
\end{aligned}$$

Finally, we can separate the inter-generational and the demographic component as follows

$$\begin{aligned}
\tau^1_{wedge} &= b\left[\frac{\bar{Y}}{2\sum\frac{1}{2}Y} - \frac{1}{2} + \frac{1}{2} - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\right] + \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\left[b-\mathbb{E}\left[\mathbb{B}'\right]\right] + \\
&\frac{(1+i)^2}{(1+r)^2}\frac{1}{2}\frac{\mathbb{E}\left[\mu_1+\mu_2\right]}{\mathbb{E}\left[u_c(c_3)\right]+\mathbb{E}\left[\mu_1+\mu_2\right]}\mathbb{E}\left[\mathbb{B}'\right] - \frac{(1+i)^2}{(1+r)^2}\frac{1}{2}Cov\left(\frac{u_c(c_3)}{u_c(c_1,l_1)},\mathbb{B}'(\bar{y})\right) \\
&= \frac{1}{2}\left[b\frac{(1+r)^2-(1+i)^2}{(1+r)^2} + \frac{(1+i)^2}{(1+r)^2}\left[b-\mathbb{E}\left[\mathbb{B}'\right]\right] - b\frac{\sum Y - \bar{Y}}{\sum Y} + \right. \\
&\left.\frac{(1+i)^2}{(1+r)^2}\frac{\mathbb{E}\left[\mu_1+\mu_2\right]}{\mathbb{E}\left[u_c(c_3)\right]+\mathbb{E}\left[\mu_1+\mu_2\right]}\mathbb{E}\left[\mathbb{B}'\right] - \frac{(1+i)^2}{(1+r)^2}Cov\left(\frac{u_c(c_3)}{u_c(c_1,l_1)},\mathbb{B}'(\bar{y})\right)\right].
\end{aligned}$$

To complete the proof, we now simply separate the time discounting factors, using the same technique as in the previous proof.

Proof of Proposition 4

In the general model, the effective marginal tax rate of the pension system in period t is given by

$$\tau^t_{wedge} = \mathbb{1}_{y_t<\hat{y}}\left[\tau - \frac{1}{R-1}\beta^{R-t}(1+i)^{R-t}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{q=1}^{\bar{s}-t}\psi_{t+q}\mathbb{E}_t\left[\mathbb{B}'(\bar{y}_R)\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right]\right].$$

In order to express the entire marginal tax rate in terms of benefits, rather than the payroll tax, I will first replace the payroll tax by its corresponding expression in terms of the average replacement rate.

Since we have constant population growth, I will normalize each cohort with the size of the cohort of *current fresh retirees*. Thus, the cohort of workers currently in their last working period is of size $(1+g)(2-\psi_R)$, while that of retirees of age $R+1$ is of size $\frac{1}{1+g}\psi_{R+1}$.

Thus, the revenue generated by the pension system is

$$\tilde{R} = \tau \sum_{s=1}^{R-1} (1+g)^s \prod_{q=0}^{s-1} (2-\psi_{R-q}) \int_{\Omega_{R-s}} min(y_{R-s}, \hat{y}) d\theta^{R-s},$$

where Ω denotes the set of histories of θ, which together with the starting conditions on the states for assets and average income fully characterise the state space.

On the other hand, the spending side of the pension system is given by

$$\int_{\Omega_R} \mathbb{B}(\bar{y}_R) d\theta^R \left[\sum_{\tilde{s}=0}^{T-R} (1+g)^{-\tilde{s}} \prod_{\tilde{q}=1}^{\tilde{s}} \psi_{R+\tilde{q}} \right],$$

which is simply the total pension benefits per cohort multiplied with the total measure of pensioners relative to 'fresh' retirees, taking into account population growth and mortality risk.

Using the budget balance equation $\tilde{R} = \tilde{S}$, we get

$$\tau = \frac{\int_{\Omega_R} \mathbb{B}(\bar{y}_R) d\theta^R \left[\sum_{\tilde{s}=0}^{T-R} (1+g)^{-\tilde{s}} \prod_{\tilde{q}=1}^{\tilde{s}} \psi_{R+\tilde{q}} \right]}{\sum_{s=1}^{R-1} (1+g)^s \prod_{q=0}^{s-1} (2-\psi_{R-q}) \int_{\Omega_{R-s}} min(y_{R-s}, \hat{y}) d\theta^{R-s}}.$$

In order to get a meaningful replacement rate, I will normalize the above with the average taxable lifetime income of the current generation of newly retired agents

$$\tau = \frac{1}{R-1} \frac{\int_{\Omega_R} \mathbb{B}(\bar{y}_R) d\theta^R}{\frac{1}{R-1} \sum_{t=1}^{R-1} \int_{\Omega_t} min(y_t, \hat{y}) d\theta^t} \cdot \\ \frac{\sum_{t=1}^{R-1} \int_{\Omega_t} min(y_t, \hat{y}) d\theta^t}{\sum_{s=1}^{R-1} (1+g)^s \prod_{q=0}^{s-1} (2 - \psi_{R-q}) \int_{\Omega_{R-s}} min(y_{R-s}, \hat{y}) d\theta^{R-s}} \cdot \\ \left[\sum_{\tilde{s}=0}^{T-R} (1+g)^{-\tilde{s}} \prod_{\tilde{q}=1}^{\tilde{s}} \psi_{R+\tilde{q}} \right].$$

Here, the term on the first line is the average replacement rate of the pension system, the term on the second line is the ratio of the total taxable income of current fresh retirees over their lifecycle to the taxable income to the economy right now. This ratio will generally be smaller than one due to population growth as well as mortality risk. Finally, the term on the third line denotes to total measure of pensioners in the economy. Denoting the average replacement rate by b, the total taxable income of current fresh pensioners by $\sum \tilde{Y}$, and the total taxable income by $\sum Y$, we can write

$$\tau = b \frac{1}{R-1} \left[\sum_{\tilde{s}=0}^{T-R} (1+g)^{-\tilde{s}} \prod_{\tilde{q}=1}^{\tilde{s}} \psi_{R+\tilde{q}} \right] \frac{\sum \tilde{Y}}{\sum Y}.$$

Just as in the proof for the previous result, we will use the identity

$$\mathbb{E}\left[A \cdot B\right] = \mathbb{E}\left[A\right] \cdot \mathbb{E}\left[B\right] + Cov\left(A, B\right)$$

to separate out the insurance component from the formulation for the effective tax rate. To keep the notation as concise as possible, I will substitute out for τ at a later date.

$$\begin{aligned}\tau^t_{wedge} = \mathbb{1}_{y\leq\hat{y}}\Bigg[\tau &- \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}} \\ &\cdot \mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\mathbb{E}\left[\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right] \\ &- \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\Bigg].\end{aligned}$$

In a next step, I will separate out the expected replacement rate from the borrowing constraints

$$\begin{aligned}\tau^t_{wedge} = \mathbb{1}_{y\leq\hat{y}}\Bigg[\tau &- \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}} \\ &\cdot \mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\mathbb{E}\left[\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right] \\ &+ \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\cdot\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right] \\ &- \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\cdot\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right] \\ &- \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\Bigg].\end{aligned}$$

$$\begin{aligned}\tau^t_{wedge} = \mathbb{1}_{y\leq\hat{y}}\Bigg[\tau &- \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right] \\ &+ \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\cdot\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\mathbb{E}\left[\frac{u_c(c_t,l_t)-u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right] \\ &- \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\Bigg].\end{aligned}$$

Following this, I will separate out the effects of intra-temporal redistribution

$$\begin{aligned}
\tau_{wedge}^{t} = \mathbb{1}_{y \leq \hat{y}} \Bigg[& \tau - \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right] \\
& + \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}b \\
& - \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}b \\
& + \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\cdot\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\mathbb{E}\left[\frac{u_c(c_t,l_t)-u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right] \\
& - \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\Bigg].
\end{aligned}$$

$$\begin{aligned}
\tau_{wedge}^{t} = \mathbb{1}_{y \leq \hat{y}} \Bigg[& \tau - \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}b \\
& + \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\mathbb{E}\left[b-\mathbb{B}'(\bar{y}_R)\right] \\
& + \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\cdot\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\mathbb{E}\left[\frac{u_c(c_t,l_t)-u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right] \\
& - \frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\Bigg].
\end{aligned}$$

Now we are ready to separate out the demographic component by substituting τ with its corresponding expression in terms of the average replacement rate.

$$
\begin{aligned}
\tau^t_{wedge} = \mathbb{1}_{y\leq\hat{y}} \Bigg[& b\frac{1}{R-1}\left[\sum_{\tilde{s}=0}^{T-R}(1+g)^{-\tilde{s}}\prod_{\tilde{q}=1}^{\tilde{s}}\psi_{R+\tilde{q}}\right]\frac{\sum\tilde{Y}}{\sum Y}+\sum_{\bar{s}=R}^{T}b-\sum_{\bar{s}=R}^{T}b \\
& -\frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}b \\
& +\frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\mathbb{E}\left[b-\mathbb{B}'(\bar{y}_R)\right] \\
& +\frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\cdot\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\mathbb{E}\left[\frac{u_c(c_t,l_t)-u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right] \\
& -\frac{1}{R-1}(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\Bigg].
\end{aligned}
$$

$$
\begin{aligned}
\tau^t_{wedge} = \frac{\mathbb{1}_{y\leq\hat{y}}}{R-1}\Bigg[& \left[\sum_{\bar{s}=R}^{T}1-(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\right]b \\
& +(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\mathbb{E}\left[b-\mathbb{B}'(\bar{y}_R)\right] \\
& -\left[\sum_{\bar{s}=R}^{T}1-\sum_{\tilde{s}=0}^{T-R}(1+g)^{-\tilde{s}}\prod_{\tilde{q}=1}^{\tilde{s}}\psi_{R+\tilde{q}}\frac{\sum\tilde{Y}}{\sum Y}\right]b \\
& +(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\cdot\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\mathbb{E}\left[\frac{u_c(c_t,l_t)-u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right] \\
& -(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\Bigg].
\end{aligned}
$$

In a final step, I separate out the inter-generational redistribution component, by collecting the discounting factors, indexing terms, and survival probabilities from the intra-generational, borrowing constraint, and insurance components.

$$
\begin{aligned}
\tau^t_{wedge} = \frac{\mathbb{1}_{y\leq\hat{y}}}{R-1}\Bigg[&\left[\sum_{\bar{s}=R}^{T} 1-(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\right] b \\
&+(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\mathbb{E}\left[b-\mathbb{B}'(\bar{y}_R)\right] \\
&-\sum_{\bar{s}=R}^{T}\mathbb{E}\left[b-\mathbb{B}'(\bar{y}_R)\right]+\sum_{\bar{s}=R}^{T}\mathbb{E}\left[b-\mathbb{B}'(\bar{y}_R)\right] \\
&-\left[\sum_{\bar{s}=R}^{T}1-\sum_{\tilde{s}=0}^{T-R}(1+g)^{-\tilde{s}}\prod_{\tilde{q}=1}^{\tilde{s}}\psi_{R+\tilde{q}}\frac{\sum\tilde{Y}}{\sum Y}\right] b \\
&+(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}\cdot\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\mathbb{E}\left[\frac{u_c(c_t,l_t)-u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right] \\
&-\sum_{\bar{s}=R}^{T}\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\mathbb{E}\left[\frac{u_c(c_t,l_t)-u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right]+\sum_{\bar{s}=R}^{T}\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\mathbb{E}\left[\frac{u_c(c_t,l_t)-u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right] \\
&-(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}}Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right) \\
&-\sum_{\bar{s}=R}^{T}Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)+\sum_{\bar{s}=R}^{T}Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\Bigg].
\end{aligned}
$$

In order to simplify this expression, I rewrite the following two sums

$$
\begin{aligned}
&\sum_{\bar{s}=R}^{T}1-(1+i)^{R-t}\beta^{R-t}\prod_{q=1}^{R-t}\psi_{t+q}\sum_{\bar{s}=R}^{T}\beta^{\bar{s}-R}\prod_{\bar{q}=1}^{\bar{s}-R}\psi_{R+\bar{q}} \\
&=\sum_{\bar{s}=R}^{T}\left[1-(1+i)^{R-t}\beta^{\bar{s}-t}\prod_{q=1}^{\bar{s}-t}\psi_{t+q}\right] \\
&\sum_{\bar{s}=R}^{T}1-\sum_{\tilde{s}=0}^{T-R}(1+g)^{-\tilde{s}}\prod_{\tilde{q}=1}^{\tilde{s}}\psi_{R+\tilde{q}}\frac{\sum\tilde{Y}}{\sum Y} \\
&=\sum_{\bar{s}=R}^{T}\left[1-(1+g)^{-(\bar{s}-R)}\prod_{q=1}^{\bar{s}-R}\psi_{R+q}\frac{\sum\tilde{Y}}{\sum Y}\right]
\end{aligned}
$$

Then the above becomes

$$
\begin{aligned}
\tau^t_{wedge} = \frac{\mathbb{1}_{y\leq\hat{y}}}{R-1}\Bigg[\sum_{\bar{s}=R}^{T}&\left[1-(1+i)^{R-t}\beta^{\bar{s}-t}\prod_{q=1}^{\bar{s}-t}\psi_{t+q}\right]\cdot\left[\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\left(1-\mathbb{E}\left[\frac{u_c(c_t,l_t)-u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\right]\right.\\
&\left.+\,Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\right]\\
&+\sum_{\bar{s}=R}^{T}\mathbb{E}\left[b-\mathbb{B}'(\bar{y}_R)\right]\\
&-\sum_{\bar{s}=R}^{T}\left[1-(1+g)^{-(\bar{s}-R)}\prod_{q=1}^{\bar{s}-R}\psi_{R+q}\frac{\sum\tilde{Y}}{\sum Y}\right]b\\
&+\sum_{\bar{s}=R}^{T}\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\mathbb{E}\left[\frac{u_c(c_t,l_t)-u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right]\\
&-\sum_{\bar{s}=R}^{T}Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\Bigg].
\end{aligned}
$$

While the above formulation already separates the effective marginal tax rate into its five components, their interpretation does not yet carry over one-to-one from the case with a single pension period without mortality. This is due to the fact that in this case, the simple identity $\beta(1+r) = 1$ no longer holds. Thus, the fourth term in the above equation can no longer be interpreted directly as a borrowing constraint, since the Euler equation is no longer given by $u_c(c_t) \geq \mathbb{E}\left[u_c(c_{t+1})\right]$, but rather by

$$
u_c(c_t) \geq (1+(1-\tau_a)r)\beta\psi_{t+1}\mathbb{E}\left[u_c(c_{t+1})\right].
$$

Thus, in order to get the correct formulation of the borrowing constraints, I transform the above as follows

$$\begin{aligned}
\tau^t_{wedge} = \frac{\mathbb{1}_{y \leq \hat{y}}}{R-1} \Bigg[\sum_{\bar{s}=R}^{T} \left[1 - (1+i)^{R-t} \beta^{\bar{s}-t} \prod_{q=1}^{\bar{s}-t} \psi_{t+q} \right] \cdot \Bigg[\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right] \left(1 - \mathbb{E}\left[\frac{u_c(c_t, l_t) - u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right) \right] \\
&+ Cov\left(\mathbb{B}'(\bar{y}_R), \frac{u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right) \Bigg] \\
&+ \sum_{\bar{s}=R}^{T} \mathbb{E}\left[b - \mathbb{B}'(\bar{y}_R)\right] \\
&- \sum_{\bar{s}=R}^{T} \left[1 - (1+g)^{-(\bar{s}-R)} \prod_{q=1}^{\bar{s}-R} \psi_{R+q} \frac{\sum \tilde{Y}}{\sum Y} \right] b \\
&+ \sum_{\bar{s}=R}^{T} \mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right] \mathbb{E}\left[\frac{u_c(c_t, l_t) - u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right] \\
&+ \sum_{\bar{s}=R}^{T} \mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right] \mathbb{E}\left[\frac{u_c(c_t, l_t) - (1+(1-\tau_a)r)^{\bar{s}-t} \beta^{\bar{s}-t} \prod_{q=1}^{\bar{s}-t} \psi_{t+q} u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right] \\
&- \sum_{\bar{s}=R}^{T} \mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right] \mathbb{E}\left[\frac{u_c(c_t, l_t) - (1+(1-\tau_a)r)^{\bar{s}-t} \beta^{\bar{s}-t} \prod_{q=1}^{\bar{s}-t} \psi_{t+q} u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right] \\
&- \sum_{\bar{s}=R}^{T} Cov\left(\mathbb{B}'(\bar{y}_R), \frac{u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right) \Bigg],
\end{aligned}$$

which can be simplified to

$$
\begin{aligned}
\tau^t_{wedge} = \frac{\mathbb{1}_{y \leq \hat{y}}}{R-1} \Bigg[\sum_{\bar{s}=R}^{T} & \left[1 - (1+i)^{R-t} \beta^{\bar{s}-t} \prod_{q=1}^{\bar{s}-t} \psi_{t+q} \right] \cdot \left[\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right] \left(1 - \mathbb{E}\left[\frac{u_c(c_t, l_t) - u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right) \right] \right. \\
& \left. + Cov\left(\mathbb{B}'(\bar{y}_R), \frac{u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right) \right] \\
& + \sum_{\bar{s}=R}^{T} \mathbb{E}\left[b - \mathbb{B}'(\bar{y}_R) \right] \\
& - \sum_{\bar{s}=R}^{T} \left[1 - (1+g)^{-(\bar{s}-R)} \prod_{q=1}^{\bar{s}-R} \psi_{R+q} \frac{\sum \tilde{Y}}{\sum Y} \right] b \\
& + \sum_{\bar{s}=R}^{T} \mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right] \mathbb{E}\left[\frac{u_c(c_t, l_t) - (1+(1-\tau_a)r)^{\bar{s}-t} \beta^{\bar{s}-t} \prod_{q=1}^{\bar{s}-t} \psi_{t+q} u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right] \\
& + \sum_{\bar{s}=R}^{T} \left[(1+(1-\tau_a)r)^{\bar{s}-t} \beta^{\bar{s}-t} \prod_{q=1}^{\bar{s}-t} \psi_{t+q} - 1 \right] \mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right] \mathbb{E}\left[\frac{u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right] \\
& - \sum_{\bar{s}=R}^{T} Cov\left(\mathbb{B}'(\bar{y}_R), \frac{u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right) \Bigg].
\end{aligned}
$$

In addition to adjusting the borrowing constraint, we will also have to adjust the insurance value. In this term, we are normalizing the marginal utility of consumption in period $\bar{s}$ with the current marginal utility of consumption in period t. However, since we now have mortality risk as well as capital gains taxation, the correct Euler equation to use when comparing marginal utilities is actually

$$u_c(c_t, l_t) \geq (1+(1-\tau_a)r)^{\bar{s}-t} \beta^{\bar{s}-t} \prod_{q=1}^{\bar{s}-t} \psi_{t+q} u_c(c_{\bar{s}}).$$

Hence, we rewrite the above as

$$
\begin{aligned}
\tau^t_{wedge} = \frac{\mathbb{1}_{y \leq \hat{y}}}{R-1} \Bigg[& \sum_{\bar{s}=R}^{T} \left[1 - (1+i)^{R-t} \beta^{\bar{s}-t} \prod_{q=1}^{\bar{s}-t} \psi_{t+q} \right] \cdot \left[\mathbb{E} \left[\mathbb{B}'(\bar{y}_R) \right] \left(1 - \mathbb{E} \left[\frac{u_c(c_t, l_t) - u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right) \right] \right. \\
& \left. + Cov \left(\mathbb{B}'(\bar{y}_R), \frac{u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right) \right] \\
& + \sum_{\bar{s}=R}^{T} \mathbb{E} \left[b - \mathbb{B}'(\bar{y}_R) \right] \\
& - \sum_{\bar{s}=R}^{T} \left[1 - (1+g)^{-(\bar{s}-R)} \prod_{q=1}^{\bar{s}-R} \psi_{R+q} \frac{\sum \tilde{Y}}{\sum Y} \right] b \\
& + \sum_{\bar{s}=R}^{T} \mathbb{E} \left[\mathbb{B}'(\bar{y}_R) \right] \mathbb{E} \left[\frac{u_c(c_t, l_t) - (1 + (1-\tau_a) r)^{\bar{s}-t} \beta^{\bar{s}-t} \prod_{q=1}^{\bar{s}-t} \psi_{t+q} u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right] \\
& + \sum_{\bar{s}=R}^{T} \left[(1 + (1-\tau_a) r)^{\bar{s}-t} \beta^{\bar{s}-t} \prod_{q=1}^{\bar{s}-t} \psi_{t+q} - 1 \right] \mathbb{E} \left[\mathbb{B}'(\bar{y}_R) \right] \mathbb{E} \left[\frac{u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right] \\
& - \sum_{\bar{s}=R}^{T} Cov \left(\mathbb{B}'(\bar{y}_R), \frac{u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right) \\
& + \sum_{\bar{s}=R}^{T} (1 + (1-\tau_a) r)^{\bar{s}-t} \beta^{\bar{s}-t} \prod_{q=1}^{\bar{s}-t} \psi_{t+q} Cov \left(\mathbb{B}'(\bar{y}_R), \frac{u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right) \\
& - \sum_{\bar{s}=R}^{T} (1 + (1-\tau_a) r)^{\bar{s}-t} \beta^{\bar{s}-t} \prod_{q=1}^{\bar{s}-t} \psi_{t+q} Cov \left(\mathbb{B}'(\bar{y}_R), \frac{u_c(c_{\bar{s}})}{u_c(c_t, l_t)} \right) \Bigg],
\end{aligned}
$$

which can be simplified to

$$\tau^t_{wedge} = \frac{\mathbb{1}_{y\leq\hat{y}}}{R-1}\left[\sum_{\bar{s}=R}^{T}\left[(1+(1-\tau_a)r)^{\bar{s}-t}-(1+i)^{R-t}\right]\beta^{\bar{s}-t}\prod_{q=1}^{\bar{s}-t}\psi_{t+q}\cdot\left[\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\left(1-\mathbb{E}\left[\frac{u_c(c_t,l_t)-u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\right]\right.\right.$$
$$+Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\Bigg]$$
$$+\sum_{\bar{s}=R}^{T}\mathbb{E}\left[b-\mathbb{B}'(\bar{y}_R)\right]$$
$$-\sum_{\bar{s}=R}^{T}\left[1-(1+g)^{-(\bar{s}-R)}\prod_{q=1}^{\bar{s}-R}\psi_{R+q}\frac{\sum\tilde{Y}}{\sum Y}\right]b$$
$$+\sum_{\bar{s}=R}^{T}\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\mathbb{E}\left[\frac{u_c(c_t,l_t)-(1+(1-\tau_a)r)^{\bar{s}-t}\beta^{\bar{s}-t}\prod_{q=1}^{\bar{s}-t}\psi_{t+q}u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right]$$
$$\left.-\sum_{\bar{s}=R}^{T}(1+(1-\tau_a)r)^{\bar{s}-t}\beta^{\bar{s}-t}\prod_{q=1}^{\bar{s}-t}\psi_{t+q}Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\right],$$

Finally, in order to arrive at the desired result, I once again use the identity

$$\mathbb{E}\left[A\cdot B\right]=\mathbb{E}\left[A\right]\cdot\mathbb{E}\left[B\right]+Cov\left(A,B\right)$$

to write

$$\tau^t_{wedge} = \frac{\mathbb{1}_{y\leq\hat{y}}}{R-1}\left[\sum_{\bar{s}=R}^{T}\left[(1+(1-\tau_a)r)^{\bar{s}-t}-(1+i)^{R-t}\right]\beta^{\bar{s}-t}\prod_{q=1}^{\bar{s}-t}\psi_{t+q}\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right]\right.$$
$$+\sum_{\bar{s}=R}^{T}\mathbb{E}\left[b-\mathbb{B}'(\bar{y}_R)\right]$$
$$-\sum_{\bar{s}=R}^{T}\left[1-(1+g)^{-(\bar{s}-R)}\prod_{q=1}^{\bar{s}-R}\psi_{R+q}\frac{\sum\tilde{Y}}{\sum Y}\right]b$$
$$+\sum_{\bar{s}=R}^{T}\mathbb{E}\left[\mathbb{B}'(\bar{y}_R)\right]\mathbb{E}\left[\frac{u_c(c_t,l_t)-(1+(1-\tau_a)r)^{\bar{s}-t}\beta^{\bar{s}-t}\prod_{q=1}^{\bar{s}-t}\psi_{t+q}u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right]$$
$$\left.-\sum_{\bar{s}=R}^{T}(1+(1-\tau_a)r)^{\bar{s}-t}\beta^{\bar{s}-t}\prod_{q=1}^{\bar{s}-t}\psi_{t+q}Cov\left(\mathbb{B}'(\bar{y}_R),\frac{u_c(c_{\bar{s}})}{u_c(c_t,l_t)}\right)\right],$$

which completes the proof. ■

Chapter 2

Lifetime Taxation in a Heterogenous Agent Framework

2.1 Introduction

A key characteristic of the progressive income tax system employed in most developed economies is that households with volatile incomes pay a higher amount of taxes over their lives than those who earn the same amount of lifetime income at a constant rate. As I demonstrate in the context of a standard lifecycle model, households with the same level of lifetime income pay average lifetime taxes that can differ by up to 5%-points. This significant variation is due to the fact that steady income earners maximise the share of their lifetime income that is taxed at lower rates, while households with volatile incomes will see a larger share of their income fall into high tax brackets.

This unequal treatment of households with the same lifetime resources prompted Vickrey (1939) and Vickrey (1947) to propose a tax averaging procedure to equalise the tax burden between households of equal lifetime income. However, while various tax averaging procedures have been in place throughout the 20'th century[1], to my knowledge no such provision is currently in operation in a major economy today.

What has been missing from the discussion about the desirability of such an averaging procedure to date is a thorough investigation in a formal economic model that makes credible assumptions about the sources of the underlying income volatility. While Liebman (2002) makes a start into this direction, he does so in a model without uncertainty, with households enjoying perfect foresight over their future, volatile, incomes. As I show below, this modeling assumption is not without loss of generality, both when it comes to discussing the equity, as well as the efficiency implications of a tax on lifetime income.

The first half of this paper is concerned with the equity implications of a tax on lifetime income. To this end, I reconsider the classical argument for lifetime taxation in a model with perfect foresight. I show that the unambiguous welfare gains found in such a model depend crucially on two key characteristic implied by the perfect foresight assumption. First, if households enjoy perfect foresight, then a household's lifetime income together with the spread of this income is a sufficient statistic to characterize the solution to the household's optimization problem. Second, in the case of annual taxation, it is possible to use the spread of lifetime income to rank households by their consumption, with low-spread households enjoying the highest level of consumption. Since a transition to a lifetime income tax will equalize consumption within each level of lifetime income in a world with perfect foresight, such a reform will lead to unambiguous welfare gains. As similar arguments cannot be made, however, for the case where households face idiosyncratic risk, I then conduct a quantitative analysis of the welfare effect in a lifecylce model with idiosyncratic risk. I show that, contrary to the 'perfect foresight' benchmark, a tran-

[1]See Vickrey (1939) for a description of a short-lived averaging procedure in Wisconsin, Copland (1924) for tax averaging in Australia, and Schmalbeck (1984) and Liebman (2002) for a discussion of US tax averaging from 1964 to 1986.

sition to a tax on lifetime income will lead to a small welfare loss equivalent to 0.16% of average consumption. This comes from the fact that workers increase savings, and hence back-load their consumption in order to insure against uncertain future tax payments.

In the second half of the paper, I turn to the question of whether a shift to lifetime income taxation can have positive effects on efficiency.

In a first step, I once again revisit the world of perfect foresight and demonstrate that the assumption of perfectly constant marginal tax rates over the lifecycle in Liebman (2002) only holds in the absence of borrowing constraints. If workers are borrowing constrained, they value the increase in contemporaneous consumption through increased earnings more strongly than the future consumption they will have to forgo in order to pay taxes on these increased earnings in the future. This effect will depress effective marginal tax rates below the level of the ex-post marginal tax rate faced by the household.

In a second step I then turn to effective marginal tax rates of a lifetime income tax in a model with idiosyncratic risk. I derive a formula for these and show that they can be decomposed into *i*) an expected ex-post marginal tax rate, *ii*) the effect of borrowing constraints, and *iii*) the effect of insurance against adverse lifetime incomes which will lead households to value lost consumption through taxes less strongly if this occurs in a state where they have enjoyed an income history leading to high lifetime incomes. Additionally, I demonstrate that in the presence of idiosyncratic risk, the effective tax rate is no longer independent of the question of *when* taxes are collected. If the government chooses to collect taxes not only at the end of the working life, but also at intermediate points, intermediate borrowing constraints will start to impact the households labour supply decision, leading to an increase in the effective marginal tax rate.

I then present some approximations of the effective marginal tax rates and their decompositions, demonstrating the the effects of borrowing constraints and insurance will depress the effective marginal tax rate below the expected ex-post rate during the first part of the working life. As more and more uncertainty is resolved, the effective marginal tax rate converges towards workers' expectation about their ex-post tax rates.

The remainder of the paper is structured as follows. Section 2.2 briefly reviews the related literature. Section 2.3 considers the equity implications of a move to lifetime income taxation, while 2.4 considers the effects of such a reform on the distortions to households labour supply decisions. Finally, section 2.5 concludes and provides an outlook on future work.

2.2 Literature Review

My paper is chiefly related to the literature on *tax averaging* started by Vickrey (1939) and Vickrey (1947). In these texts, Vickrey argues for the case of tax averaging over multiple years mainly on equity grounds. He develops specific and very involved averaging procedures to achieve equity between workers regardless of the volatility of their income. It should be noted in this context that Vickrey is writing with the US tax system in mind and thus is not only concerned with labour income, but also with capital income. To my knowledge, Liebman (2002) is the first, and so far only, paper that considers the welfare implications of such a "Vickrey-tax" on labour income[2] in a formal economic model. In this paper, the author considers a sample of the US population over the course of their lives and asks the question how much better they would have been off under a tax regime in the spirit of Vickrey, finding significant positive welfare effects. However, this paper makes several simplifying assumptions, chiefly among which is that all uncertainty is resolved at the beginning of the lifecycle and that workers have perfect foresight with regards to their (potentially volatile) income histories. As I will demonstrate, this assumption will lead to certain results that do not generalize to a more realistic model with idiosyncratic risk.

[2] Although Vickrey's ideas have received some attention in the literature on capital gains taxation. See for example Auerbach (1991), Bradford (1994), Auerbach and Bradford (2004)

In recent years, there has been a growing interest in lifetime taxation in the New Dynamic Public Finance Literature. Three papers are of special interest here as they concern themselves with lifetime taxation. First, Michau (2014) derives the optimal lifetime tax schedule in a setting with one tax payment at the end of the working life and a deterministic productivity schedule. Second, Farhi and Werning (2013), as well as Golosov et al. (2016) derive optimal labour wedges in a lifecycle setting setting with uncertainty. While they demonstrate that the implementation of these optimal distortions leads significant positive welfare gains, this literature is silent with regards to how such wedges could be implemented.

2.3 Equity Considerations

2.3.1 The Classical Argument for Lifetime Income Taxation

In this section, I will briefly discuss and formalise the classic argument for equity gains through lifetime income taxation in a model without idiosyncratic risk that has been central of the previous literature.

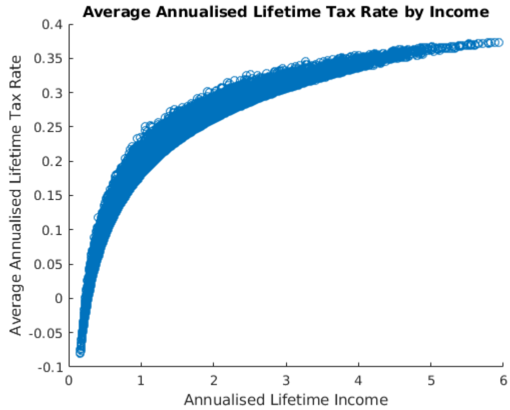
As Vickrey (1939) writes in his original paper,

> It has long been considered one of the principal defects of the graduated individual income tax that fluctuating incomes are, on the whole, subjected to much heavier tax burdens than incomes of comparable average magnitude which are relatively steady from year to year. (p.379)

Figure 2.1 illustrates this point by depicting the average annualised lifetime tax rate[3] paid by 5 million agents with simulated earnings histories from a standard lifecycle model with an $AR(1)$

[3] For every level of lifetime income (which must, of course be calculated in NPV terms), I define annualised lifetime income as that income that would have yielded this level of lifetime income if earned at a perfectly steady rate over the lifecycle. Annualised lifetime taxes are defined equivalently. The average annualised lifetime tax rate is then simply the quotient of the two.

Figure 2.1: Spread of Average Lifetime Tax Rates



The average annualised lifetime tax rate is calculated by taking the annualised NPV of taxes paid over the lifecycle and dividing it by the annualised NPV of lifetime income

income process[4]. As can be seen, there is significant heterogeneity in lifetime tax payments between agents of the same level of lifetime income. At most levels of lifetime income, this spread between the lower contour of the figure, representing those with perfectly steady incomes, and those with the highest volatility approaches 5%-points.

Vickrey (1939), Vickrey (1947), Vickrey (1972), and Liebman (2002) all propose reforms that are targeted at eliminating this spread. In this paper, I will follow Liebman (2002) in considering a reform that taxes each earnings history at the average of the lifetime taxes paid on earnings histories with the same lifetime income under the current, annual, tax system. This type of reform has two distinct advantages: first, it is revenue neutral up to adjustments in labour supply, as it collects the same amount of taxes at each level of lifetime income as the pre-reform tax system. Second, it is distributionally neutral in the sense that it does not redistribute resources between different levels of lifetime income. Previous reform proposals such as Vickrey (1939), who suggested taxing every earnings history as if income had been earned at a perfectly steady rate do not meet this requirement. Since such a reform would move every agent to the lower

[4] Taxes on annual income y are calculated as $T(y) = y - \lambda y^{1-\tau}$ with $\tau = 0.151$ following Heathcote et al. (2014)

contour of figure 2.1, the revenue to the government would obviously be lower than pre-reform. Additionally, the benefits of the reform would accrue disproportionately to those lifetime income levels that exhibit particularly large volatility in earnings.

In order to understand why such a reform will lead to welfare improving equity gains in the 'perfect foresight' world considered in the previous literature, I will present a two-period example that formalizes and generalizes the argument in Liebman (2002). This example will also demonstrate why this reasoning cannot be applied directly to a heterogeneous agent model, necessitating a more thorough investigation by means of a quantitative model.

Specifically, consider a population distribution over earnings histories $G(y_1, y_2)$, which describes the measure of agents that have experienced the earnings history $y^2 = \{y_1, y_2\}$. This distribution over the population can be the outcome of (at least) two types of income processes. First, and this is the world view underlying the previous literature on lifetime taxation, agents can draw their type $y^2 = \{y_1, y_2\}$ directly upon being born. Then, there is no more uncertainty to be resolved over the lifecycle, and all households have perfect foresight over their earnings histories. Second, and this is will be the focus of the next section, agents can draw a type y_1 from a distribution $F(y_1)$ upon being born, and then draw a second-period type y_2 from a conditional distribution $F(y_2|y_1)$ in the second period. If, for every history $y^2 = \{y_1, y_2\}$ we have $G(y_1, y_2) = F(y_2|y_1)F(y_1)$, the two income processes are clearly observationally equivalent *ex-post*[5]. In the remainder of this section, I will consider the first of the two scenarios, as it allows for a very strong result, and is the basis upon which the classic argument for lifetime income taxation has been built.

Given an income tax schedule $T(y)$ on annual labour income, I define the lifetime tax schedule

[5]In the sense that it is not possible to determine which of the two processes generated the population by simply considering the the observed earnings histories within it

$\hat{T}(\bar{y})$ on annualised lifetime income $\bar{y} = \frac{1}{1+(1+r)}\left((1+r)y_1 + y_2\right)$ by

$$\hat{T}(\bar{y}) = \frac{1}{1+(1+r)} \int_{\mathbb{X}(\bar{y})} \left[(1+r)T(y_1) + T(y_2)\right] dG(y_1, y_2),$$

where $\mathbb{X}(\bar{y})$ denotes the set of earnings histories that will lead to an annualised lifetime income of $\bar{y}$, i.e.

$$\mathbb{X}(\bar{y}) = \left\{ \{y_1, y_2\} \ s.t. \ \frac{1}{1+(1+r)}\left((1+r)y_1 + y_2\right) = \bar{y} \right\}.$$

Thus, under the lifetime tax schedule, a worker with lifetime income $\bar{y}$ will face a tax burden equal to the average lifetime tax burden faced by all workers of lifetime income $\bar{y}$ under the current, annual, tax system.

In order to see why a move from annual tax schedule $T(\cdot)$ to lifetime tax schedule $\hat{T}(\cdot)$ can lead to welfare gains in a setting where all uncertainty is resolved in the initial period, consider the optimization problem facing a household that has drawn type $y^2 = \{y_1, y_2\}$ and faces the annual tax schedule $T(\cdot)$

$$V(y_1, y_2) = \max_{c_1, c_2, a_2} u(c_1) + \beta u(c_2)$$

subject to the budget constraints

$$c_1 = y_1 - T(y_1) - a_2$$

$$c_2 = y_2 - T(y_2) + (1+r)a_2.$$

Abstracting from borrowing constraints, and making the standard assumption $\beta(1+r) = 1$, we can immediately use the Euler equation to write

$$c_1 = c_2 = \frac{1}{1+(1+r)}\left((1+r)y_1 + y_2\right) - \frac{1}{1+(1+r)}\left((1+r)T(y_1) + T(y_2)\right).$$

Due to the assumption on the process generating this income history, households have perfect foresight and can perfectly smooth consumption. Thus, in each period, they will consume their annualised *net lifetime resources*, which is the difference between lifetime income and lifetime taxes paid.

Now, if the tax schedule $T(\cdot)$ is twice continuously differentiable and progressive in the sense that $T'' > 0$, then it is possible to show that for any given lifetime income $\bar{y}$, consumption will be decreasing in the difference between first-, and second-period income. Specifically, a straightforward application of Taylor's theorem (see for example Binmore (1982)) yields[6]

$$c_1 = c_2 \approx \bar{y} - T(\bar{y}) - T''(\bar{y})x^2 \frac{(1+r) + (1+r)^2}{2(1 + (1+r))^3},$$

where $x = y_1 - y_2$. Thus, within each level of lifetime income, it is possible to rank households' consumption (and hence welfare) based on the amount of volatility of their earnings. Households with the lowest volatility in their earnings enjoy the highest level of consumption (for $x = 0$, the third term of the above formula disappears altogether), while those with the highest volatility face the lowest level of consumption.

On the other hand, under the lifetime tax schedule described above, the budget constraints are given by

$$c_1 = y_1 - a_2$$

$$c_2 = y_2 - (1 + (1+r))\hat{T}((1+r)y_1 + y_2) + (1+r)a_2,$$

[6] A similar argument will hold when the marginal tax rate is a weakly increasing step function. However, for simplicity I will focus on the abovementioned case.

and repeating the above analysis will yield a constant level consumption of

$$c_1 = c_2 = \bar{y} - \hat{T}(\bar{y})$$

within each level of lifetime income. Following the definition of the tax on annualised lifetime income, we can express this level of consumption in terms of the original, annual, tax schedule in order to determine who wins and who loses through the reform

$$c_1 = c_2 = \bar{y} - \frac{1}{1+(1+r)} \int_{\mathbb{X}(\bar{y})} \left[(1+r)T(\tilde{y}_1) + T(\tilde{y}_2)\right] dG(\tilde{y}_1, \tilde{y}_2).$$

Now, if for every pair $\{\tilde{y}_1, \tilde{y}_2\}$, we express the taxes paid on this income stream under the annual tax system in terms of average annualised income $\bar{y}$ and the difference $\tilde{x} = \tilde{y}_1 - \tilde{y}_2$, a second application of Taylor's theorem will allow us to write

$$\begin{aligned} \frac{1}{1+(1+r)} \int_{\mathbb{X}(\bar{y})} \left[(1+r)T(\tilde{y}_1) + T(\tilde{y}_2)\right] dG(\tilde{y}_1, \tilde{y}_2) \approx \\ T(\bar{y}) + T''(\bar{y}) \frac{(1+r)+(1+r)^2}{2(1+(1+r))^3} \int_{\mathbb{X}(\bar{y})} \tilde{x}^2 dH(\tilde{x}|\bar{y}). \end{aligned}$$

Therefore, it is possible to write the consumption gain through the reform as

$$\delta(y_1, y_2) \approx T''(\bar{y}) \frac{(1+r)+(1+r)^2}{2(1+(1+r))^3} \left[(y_1 - y_2)^2 - \int_{\mathbb{X}(\bar{y})} (\tilde{y}_1 - \tilde{y}_2)^2 dG(\tilde{y}_1, \tilde{y}_2) \right].$$

Hence, households with earnings that exhibit lower than average volatility for their level of lifetime income will lose consumption due to the reform while those with above average volatility will gain consumption. But as this exactly shifts consumption from high-consumption to low-consumption households without changing their relative ranking, (as post-reform consumption will be equalized within each lifetime income level), such a reform will lead to welfare gains under standard assumptions on the curvature of $u(\cdot)$. The following proposition will summarize our findings so far

Proposition 5 *In a two-period economy with volatile incomes, perfect foresight, and without borrowing constraints, a move from an annual tax schedule to the lifetime income tax schedule $\hat{T}$ will lead to unambiguous welfare gains as long as there are diminishing returns to consumption.*

It is important to note that these unambiguous welfare gains are the result of two features of the model without any idiosyncratic risk. First, agents have perfect foresight, and hence can perfectly smooth consumption over the lifecycle. Second, and relatedly, in such a model, an agent's lifetime income, as well as it's volatility are good statistics for characterizing the household problem[7]. To the author's knowledge, no such claims can be made in the case where the distribution $G(y_1, y_2)$ is the result of ex-ante heterogeneity as well as idiosyncratic risk. For this reason, the next section will concern itself with the quantitative analysis of a shift to lifetime income taxation in a heterogeneous agent model.

2.3.2 Lifetime Taxation in a Heterogeneous Agent Framework

Setup

In order to conduct the analysis, I build a partial equilibrium lifecycle model with ex-ante heterogeneity, idiosyncratic productivity risk, exogenous income, and incomplete markets.

Agents in this economy live for a maximum of T periods. They work for the first $R-1$ periods, in each period inelastically supplying one period of labor to the market and then retire for the remainder of their lives. Upon entering retirement, they face a risk of dying with probability

[7] In the case of a step-wise linear tax function, lifetime income and its spread $x = y_1 - y_2$ are actually *sufficient statistics* for earnings history, as the third moment of the tax function is zero everywhere, and a discrete version of the Taylor approximation used above will hold with equality. This will allow us to fully characterize the solution to each household's problem in terms of the two statistics $\bar{y}$ and x. Incidentally, if we restrict the tax function to be linear, then lifetime income itself is a sufficient statistic for the history of earnings.

$1 - \psi_{t+1}$ between periods t and $t+1$, with $\psi_{T+1} = 1$. Given these population dynamics, there will be a measure μ_t of age t agents, and I normalise this measure such that $\sum_{t=1}^{T} \mu_t = 1$.

Markets are incomplete, allowing households to save in a one period bond at an interest rate r. I assume that agents do not have any bequest motives.

The government: In order to focus the discussion, I assume that the government only taxes labour income in order to finance some level of wasteful government expenditure. Specifically, I assume that the government wants to raise a constant net present value of G from each generation over the course of their working life. It does so by collecting taxes $\mathbb{T}_t(y^t)$ in period t from a household with income history y^t. In the case of an annual tax (pre-reform), this will be given by

$$\mathbb{T}_t(y^t) = \begin{cases} T(y_t) & , \ t \leq R-1 \\ 0 & , \ t > R-1 \end{cases}$$

as taxes depend only on the last realization of income. Post-reform, taxes will be raised through a tax on annualised lifetime income that is a generalisation of the two-period case introduced above. Specifically, for each period t I define *current annualised lifetime income* $\bar{y}_t$ as

$$\bar{y}_t = \frac{1}{\sum_{s=1}^{t}(1+r)^{t-s}} \sum_{s=1}^{t}(1+r)^{t-s} y_s.$$

Taxes on annualised lifetime income are then defined as

$$\hat{T}_t(\bar{y}_t) = \frac{1}{\sum_{s=1}^{t}(1+r)^{t-s}} \int_{\mathbb{X}(\bar{y}_t)} \left[\sum_{s=1}^{t}(1+r)^{t-s} T(\tilde{y}_s)\right] dG(\tilde{y}^t),$$

where $\mathbb{X}(\bar{y})$ is once again the set of histories $\tilde{y}^t$ such that

$$\frac{1}{\sum_{s=1}^{t}(1+r)^{t-s}} \sum_{s=1}^{t}(1+r)^{t-s} \tilde{y}_s = \bar{y}_t.$$

Having defined *how* taxes on lifetime income are to be calculated, next we need to define *when* tax payments are due. In the simplest case, there will only be one tax payment at the end of the households' working lives. In this case, taxes will be collected as

$$\mathbb{T}_t(y^t) = \begin{cases} 0 & , \ t < R-1 \\ \sum_{s=1}^{t}(1+r)^{t-s}\hat{T}_t(\bar{y}_t) & , \ t = R-1 \\ 0 & , \ t \geq R, \end{cases}$$

and I will refer to this case as the lifetime tax schedule *without* intermediate collections. It is easy to confirm that this tax schedule will be revenue neutral as well as distributionally neutral over levels of lifetime income compared to the underlying annual tax schedule.

However, the assumption of a single tax payment at the end of the working life is not without loss of generality in the presence of borrowing constraints. If these constraints affect young agents more strongly than those at the end of their working lives, then such a shift in the timing of tax collections might lead to a welfare gain that is unrelated to the question whether or not it is possible to provide better insurance through lifetime taxation.

In order to forestall such concerns, I will define a second tax schedule, referred to as the lifetime tax schedule *with* intermediate collections, in which taxes are raised in each period and then rebated back to the household next period as tax credits towards their future tax payments. This will lead to a tax schedule

$$\mathbb{T}_t(y^t) = \begin{cases} 0 & , \ t = 0 \\ \sum_{s=1}^{t}(1+r)^{t-s}\hat{T}_t(\bar{y}_t) - (1+r)\sum_{s=1}^{t-1}(1+r)^{t-1-s}\hat{T}_{t-1}(\bar{y}_{t-1}) & , \ 1 \leq t \leq R-1 \\ 0 & , t > R-1. \end{cases}$$

It is important to remark that the total NPV of taxes collected on annualised lifetime income $\bar{y}_{R-1}$ at retirement does not differ between the case with and without intermediate collections. It is simply the timing of tax collections that is altered.

In addition to raising revenue in this way, the government will also collect all accidental bequests and distribute them as a lump-sum payment d to each household.

Households: When they enter the economy, agents draw their initial income type θ_1 from the distribution $F(\theta_1)$. For any period $t > 1$, income follows a Markov process with the conditional distribution $F(\theta_t|\theta_{t-1})$.

In each period, workers face wages $w_t = \gamma_t\theta_t$, where γ is the lifecycle component of productivity that is common to all workers, whereas θ is the idiosyncratic shock. We have $\gamma_t = 0$ for all $t > R-1$, implying zero wages during retirement. Upon observing her wage, an agent makes a consumption-saving decision in order to maximise her expected discounted sum of utilities, where to flow utility in each period is given by

$$U_t = \frac{c_t^{1-\sigma}}{1-\sigma}.$$

Here, c_t denotes consumption in period t.

Thus, the problem of a household with current assets a_t, past annualised lifetime earnings $\bar{y}_{t-1}$, and current productivity θ_t is given by

$$V_t(a_t, \bar{y}_{t-1}, \theta_t) = \max_{c_t, a_{t+1}} \frac{c_t^{1-\sigma}}{1-\sigma} + \psi_{t+1}\beta\mathbb{E}V_{t+1}(a_{t+1}, \bar{y}_t, \theta_{t+1}) \tag{2.1}$$

subject to the budget constraint

$$c_t + a_{t+1} \leq y_t - \mathbb{T}(y^t) + (1+r)a_t + d,$$

the law of motion for annualised lifetime income

$$\bar{y}_t = \frac{(1+r)\sum_{s=1}^{t-1}(1+r)^{t-1-s}\bar{y}_{t-1} + y_t}{\sum_{s=1}^{t}(1+r)^{t-s}},$$

as well as the borrowing constraint $a_{t+1} \geq 0$, and the initial conditions $a_1 = \bar{y}_0 = 0$.

Equilibrium definition: a partial equilibrium in the OLG economy is defined as a policy function $a_{t+1}(a_t, \bar{y}_{t-1}, \theta_t)$, together with a tax schedule $\mathbb{T}$, a lump-sum transfer d, and a distribution $\lambda_t(a_t, \bar{y}_{t-1}, \theta_t)$ such that, given the interest rate r

1. The policy function solves the household problem (2.1)

2. The government collects the required amount of taxes from every generation over their lifecycle, i.e.
$$\sum_{t=1}^{R-1}(1+r)^{R-1-t}\int \mathbb{T}_t(y^t)d\lambda_t(a_t, \bar{y}_{t-1}, \theta_t) = G$$

3. The lump-sum transfer is equal to the accidental bequests collected by the government
$$d = \sum_{t=R}^{T}\mu_{t-1}(1-\psi_{t-1})\int a_t(a_{t-1}, \bar{y}_{t-2}, \theta_{t-1})d\lambda(a_{t-1}, \bar{y}_{t-2}, \theta_{t-1})$$

4. The distribution λ is consistent with individual optimality, i.e.
$$\lambda_{t+1}(\tilde{a}_{t+1}, \tilde{\bar{y}}_t, \theta_{t+1}) = \int_{\mathbb{Z}_t} f(\theta_{t+1}|\theta_t)d\lambda(a_t, \bar{y}_{t-1}, \theta_t),$$
where $\mathbb{Z}_t = \{(a_t, \bar{y}_{t-1}, \theta) \ s.t. \ a_{t+1}(a_t, \bar{y}_{t-1}, \theta) = \tilde{a}_{t_1}\}$, as well as with the initial conditions $a_1 = \bar{y}_0 = 0$.

Parameterization

The parameters of the model are chosen in order to broadly reflect key characteristics of the US economy. Future versions of this paper will include a more careful calibration, especially with regards to the income process and the borrowing constraints.

Demographics: Agents enter the economy at age 21 and die for sure at age 101. They do not face mortality risk during their working life, i.e. $\psi_{t+1} = 0,\ t < R - 1$. Age specific survival probabilities following the entry into retirement are taken from the US life table for males (Arias (2009)).

Taxation: The exogenous level of government spending is taken from Huggett and Ventura (1999) and set at 19.5%. The progressivity parameter τ in the tax function on annual income, $T(y) = y - \lambda y^{1-\tau}$ is set at 0.151 for the pre-reform state following Heathcote et al. (2014). The parameter λ will adjust to balance the government budget. Note that no budget adjustment is required for the case of lifetime income taxation as the reform is revenue neutral.

Productivity: The common lifecycle profile of productivity follows that of US males with numerical values taken from Fehr and Kindermann (2018). The idiosyncratic productivity process is assumed to follow a log-normal process as

$$log(\theta_t) = \rho log(\theta_{t-1}) + \varepsilon_t,\ t > 1,$$

with $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ and the initial state distributed as $log(\theta_1) \sim N(0, \sigma_0^2)$. The parameterization of this process follows Huggett (1996) with $\rho = 0.96$, $\sigma_\varepsilon^2 = 0.045$, and $\sigma_0^2 = 0.38$. This process is discretized following the method developed by Tauchen (1986) with 19 grid points. Finally, the state space is rescaled to normalize the average gross annualised lifetime income in the pre-reform economy to one.

Preferences: The risk preference parameter is set to $\sigma = 2$. I parameterize the discount factor β such that the identity $\beta(1 + r) = 1$ holds, where the interest rate r is set to the average real US interest rate since the 1960's at $r = 3.85\%$ following World Bank (2018). This results in a value for the discount factor of $\beta = 0.963$.

Results

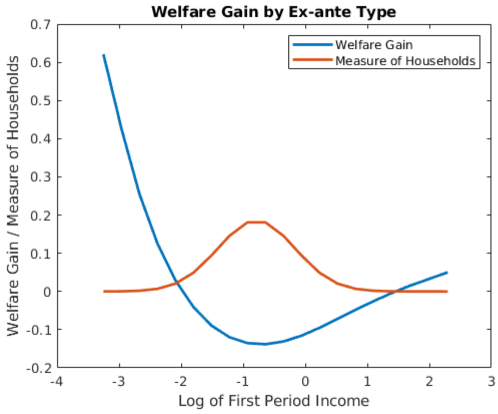
I will only discuss the quantitative results for the case where lifetime taxes are collected *with* intermediate tax payments in each period. This is for two reasons: first, the collection procedure *without* intermediate collection would change the timing of tax collections compared to the baseline model with an annual tax system. No taxes would be collected during the course of the working life with one large payment coming due before retirement. If there is a non-zero measure of borrowing constrained households in the baseline model, this will lessen the impact of those constraints and lead to welfare gains that, in my view, cannot be attributed to the change of the tax base, but rather to that in the timing of collections. Second, a single large tax payment at the end of the working life gives rise to several implementability concerns such as agents dying prematurely, or leaving the country toward the end of the working life in order to escape their tax burden. The case with intermediate collections is free from such concerns as workers are 'even' with the tax authorities at the end of each period.

Contrary to the case of perfect foresight discussed in the previous section, such a reform will not lead to welfare gains in a model with idiosyncratic risk. Given the above specification, a move from an annual tax system to a tax on lifetime income will lead to a small welfare loss that is equivalent to a decrease in average consumption of 0.16%[8].

First, it is important to note that this welfare loss is not uniform, i.e. not everybody loses due to this reform. Figure 2.2 shows the welfare gains of households by their ex-ante income types. As can be seen in the figure, workers who start their lives with very low, or very high, incomes, and hence expect to have fairly high volatility in their lifetime earnings, will gain through the

[8]All calculations assume a utilitarian welfare function with the weights on ex-ante types given by their population weights.

Figure 2.2: Welfare Gains Through the Reform



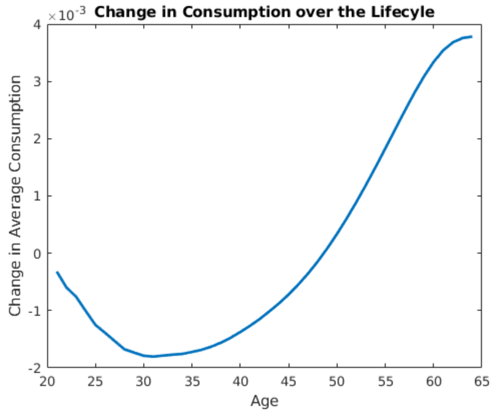
reform, while those that start their lives with closer to the median of the distribution lose welfare. While the gain is considerable for low-income households, the overall welfare gain is still negative due to the small weight of such households in the population[9].

Several reasons for this drop in average welfare can be excluded on the basis of careful modeling choices. First, as the reform is revenue neutral, average consumption does not decrease due to the change in the tax system, as income is exogenous and the government collects the same amount of taxes. Second, due to the way the intermediate collections have been defined, there is no change in the *timing* of tax collections, either. This implies that any change in welfare must come from a change in the abilities of households to smooth consumption over their lifecycle.

In order to investigate this further, figure 2.3 depicts the change in average consumption over the lifecycle. It shows that, on average, households will consume less during the first part of their lives and then increase consumption over the pre-reform level towards the end of their working lives. This suggests that the change in the tax system will lead to larger expected volatility due to uncertainty about the size of future tax payments. Households react to this increase in volatility by accumulating assets which they then run down at a faster rate towards the end of

[9]If, however, one were to consider a Rawlsian welfare function that seeks to maximize the welfare of the 'least well off' member of the society, this reform would lead to overall welfare gains

Figure 2.3: Change in Average Consumption over the Lifecycle



Average consumption at each age is calculated for both the pre- and post-reform steady state with the respective distribution.

their working lives when their expectations over their ex-post tax burden have stabilized. Since the life-cycle profile of earnings will lead to an upward sloping consumption profile even in the pre-reform steady state, such a 'back-loading' of consumption is detrimental to welfare.

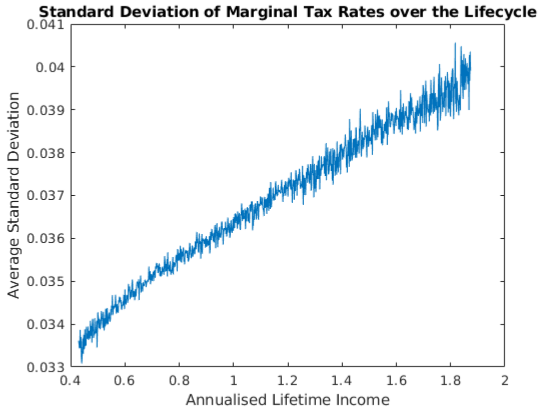
2.4 Efficiency Considerations

2.4.1 Effective Marginal Tax Rates with Perfect Foresight

In addition to the classic argument based on equity considerations discussed in section 2.3.1, Liebman (2002) introduced the possibility of efficiency gains through a transition to a lifetime income tax. In this section, I will briefly discuss his argument, as it is instructive when considering how to think about labour distortions under tax on lifetime income.

Figure 2.4 illustrates the basis for his argument, as it depicts the average standard deviation of marginal tax rates faced by households of a given lifetime income over their lifecycle. Given that income fluctuates significantly over the lifecycle, the average standard deviation of marginal tax rates is consistently between 3 and 4%-points. Liebman (2002) remarks that, since the

Figure 2.4: Spread of Marginal Tax Rates



Average standard deviation for each lifetime income level are calculated based on the simulation of 5 million earnings histories from an $AR(1)$ income process, and for each draw calculating the standard deviation of marginal tax rates faced over the lifecycle.

deadweight loss of taxation rises with the square of the effective marginal tax rate rather than linearly, it should be more efficient if one could eliminate this dispersion and instead have each household face a single, constant marginal tax rate over the lifecycle.

Since the meaning of a 'marginal tax rate' is not immediately clear in a setting in which taxes are levied on lifetime- rather than annual income, I will at this point introduce the concept of an effective marginal tax rate. Formally, the effective marginal tax rate is equivalent to the wedge introduced into the labor-leisure decision through the tax system, and can be defined as

$$\tau_{eff} = 1 - \frac{MRS}{MRT}, \tag{2.2}$$

where MRS is the marginal rate of substitution between labour and leisure, while MRT denotes the marginal rate of transformation, in this case the gross wage.

Consider the following simple two period model, where agents once again draw their entire history at the start of their lives. As we are now considering labour supply, I will assume that this history consists of *productivities*, instead of incomes. Given a draw $\theta^2 = \{\theta_1, \theta_2\}$, the household, that now cares about leisure as well as consumption solves

$$\max_{c_1,c_2,y_1,y_2} u\left(c_1, 1-\frac{y_1}{\theta_1}\right)+\beta u\left(c_2, 1-\frac{y_2}{\theta_2}\right)$$

subject to

$$c_1 = y_1 - T(y_1) - a_2$$

$$c_2 = y_2 - T(y_2) + (1+r)a_2$$

in the case of annual taxation, and

$$c_1 = y_1 - a_2$$
$$c_2 = y_2 - (1+(1+r))\hat{T}\left(\frac{1}{1+(1+r)}\left((1+r)y_1+y_2\right)\right)+(1+r)a_2$$

in the case of a lifetime tax.

Taking first-order conditions in the case of annual taxation yields the intra-temporal optimality conditions

$$\frac{u_l}{u_c} = (1-T'(y_t))\theta_t.$$

Since the left hand side of this equation is exactly the marginal rate of substitution between labour and leisure, we can substitute the above into the definition of the effective marginal tax rate to yield

$$\tau_{eff} = T'(y_t),\ t=1,2.$$

Thus, in the case of annual taxation, the effective marginal tax rate coincides with the 'usual' definition of the marginal tax rate and differs across periods for those households whose income changes between period.

On the other hand, the intra-temporal optimality conditions for the case with lifetime income taxation are given by

$$\begin{aligned}\frac{u_l(c_1, 1-l_1)}{u_c(c_1, 1-l_1)} &= \left[1 - \frac{u_c(c_2, 1-l_2)}{u_c(c_1, 1-l_1)}\hat{T}'(\bar{y})\right]\theta_1 \\ \frac{u_l(c_2, 1-l_2)}{u_c(c_2, 1-l_2)} &= \left[1 - \hat{T}'(\bar{y})\right]\theta_2,\end{aligned}$$

which will lead to an effective marginal tax rate of

$$\begin{aligned}\tau^1_{eff} &= \frac{u_c(c_2, 1-l_2)}{u_c(c_1, 1-l_1)}\hat{T}'(\bar{y}) \\ \tau^1_{eff} &= \hat{T}'(\bar{y}).\end{aligned}$$

Thus, even in a world with perfect foresight, effective marginal tax rates are not necessarily equalized under a lifetime income tax schedule. If a household is borrowing constrained, then the effective marginal tax rate is depressed in the first period, as the household cares more about the contemporaneous consumption afforded by an extra unit of income than about the consumption they will have to forgo in the next period to pay the taxes on this unit of income. Thus, the constant marginal tax rate employed by Liebman (2002) is an upper bound for the distortion induced by a lifetime tax schedule, implying that the reduction in deadweight loss from a transition to such a lifetime tax calculated in that paper actually constitutes a lower bound on the true reduction in deadweight loss.

However, the above result once again depends on the assumption that households have perfect foresight over their future productivity types (and hence their future earnings). If future earnings are uncertain due to idiosyncratic risk, workers have to form expectations about their eventual lifetime earnings, which will drive their effective marginal tax rate. Additionally, as I will demonstrate in the next section, insurance against bad earnings histories afforded by the tax system will start to impact the effective marginal tax rate faced by households over their lifecycle, in addition to the effect of borrowing constraints discussed above.

2.4.2 Effective Marginal Tax Rates with Idiosyncratic Risk

In order to analyse the properties of effective marginal tax rates under lifetime income taxation in the presence of idiosyncratic risk, I will add a labour-leisure choice to the model introduced in section 2.3.2. Similar to the analysis in that section, I will once again consider both a case with intermediate tax collection as well as a case where all taxes are paid at the end of the working life.

I now assume that households no longer draw their income directly, but rather draw a productivity type θ. Again, they draw their initial type θ_1 at the start of their lives, and then make a new draw each period from a distribution $F(\theta_t|\theta_{t-1})$. Upon observing their type in period t, households solve the following problem

$$V_t(a_t, \bar{y}_{t-1}, \theta_t) = \max_{c_t, a_{t+1}, y_t} u\left(c_t, \frac{y_t}{\theta_t}\right) + \psi_{t+1}\beta\mathbb{E}_t\left[V_{t+1}\left(a_{t+1}, \bar{y}_t, \theta_{t+1}\right)\right] \tag{2.3}$$

subject to the budget constraints

$$c_t + a_{t+1} \leq y - \mathbb{T}(y^t) + (1+r)a_t, \; t < R \tag{2.4}$$

$$c_t + a_{t+1} \leq (1+r)a_t, \; t \geq R \tag{2.5}$$

the borrowing constraint $a' \geq 0$, the initial conditions $a_1 = \bar{y}_0 = 0$, as well as the law of motion for the annualised lifetime income

$$\bar{y}_t = \frac{(1+r)\sum_{s=1}^{t-1}(1+r)^{t-1-s}\bar{y}_{t-1} + y_t}{\sum_{s=1}^{t}(1+r)^{t-s}}. \tag{2.6}$$

Finally, the tax payment $\mathbb{T}(y^t)$ due after history y^t depends on the way that taxes are collected. Without intermediate collection, we will have

$$\mathbb{T}(y^t) = \begin{cases} 0 & , \ t < R-1 \\ \sum_{s=1}^{t}(1+r)^{t-s}\hat{T}_t(\bar{y}_t) & , \ t = R-1, \end{cases} \tag{2.7}$$

while in the case *with* intermediate collection we have

$$\mathbb{T}(y^t) = \begin{cases} \hat{T}_t(\bar{y}_t) & , \ t = 1 \\ \sum_{s=1}^{t}(1+r)^{t-s}\hat{T}_t(\bar{y}_t) - (1+r)\sum_{s=1}^{t-1}(1+r)^{t-1-s}\hat{T}_{t-1}(\bar{y}_{t-1}) & , \ t > 1. \end{cases} \tag{2.8}$$

Since in both cases, tax payments depend only on the annualised lifetime income instead of the entire history of earnings, we can simplify the notation and write $\mathbb{T}(y^t) = \mathbb{T}_t(\bar{y}_t)$.

The households' first order conditions during working periods are given by

$$u_c(c_t, l_t) \geq \beta(1+r)\mathbb{E}\left[u_c(c_{t+1}, l_{t+1})\right] \tag{2.9}$$

$$u_l(c_t, l_t)\frac{1}{\theta} + u_c(c_t, l_t)(1 - \mathbb{T}_t'(\bar{y}_t)) + \beta\mathbb{E}\left[\frac{\partial V_{t+1}\left(a_{t+1}, \bar{y}_t, \theta_{t+1}\right)}{\partial \bar{y}_t}\frac{\partial \bar{y}_t}{\partial y_t}\right] = 0. \tag{2.10}$$

Note that, since workers face no mortality risk during their working lives, i.e. we have $\psi_{t+1} = 0, \ t < R$, I have dropped this parameter for notational convenience.

Considering the first oder condition for labour supply, we can see that taxes now have an instantaneous, as well as a dynamic impact on the labour supply decision. First, if taxes are collected in period t, workers take into account the *current* marginal tax rate $\mathbb{T}'_t(\bar{y}_T)$. Additionally, since income earned today will impact the annualised lifetime income state, there will be a dynamic effect through the envelope condition $\frac{\partial V_{t+1}(a_{t+1},\bar{y}_t,\theta_{t+1})}{\partial \bar{y}_t}$, that describes how a household's continuation value will change when starting the next period with slightly more past annualised lifetime income.

In order to simplify the discussion, I will consider the cases with and without intermediate tax payments separately, starting with the case *without* intermediate collections. In this case, since taxes are only collected in the final period, we can immediately see that the above-mentioned effects of taxes on labour supply are exclusive to each other. In any non-final working period, there will be no instantaneous effect of taxation, i.e. $\mathbb{T}'_t(\bar{y}_t) = 0$, while in the final working period we will have $\frac{\partial V_{t+1}(a_{t+1},\bar{y}_t,\theta_{t+1})}{\partial \bar{y}_t} = 0$, since taxes do not matter during retirement.

Thus, in the final working period, the first order condition is given by

$$
\begin{aligned}
& u_l(c_t,l_t)\frac{1}{\theta} + u_c(c_t,l_t)(1-\mathbb{T}'_t(\bar{y}_t)) = 0 \\
\leftrightarrow\ & u_l(c_t,l_t)\frac{1}{\theta} + u_c(c_t,l_t)\left(1-\sum_{s=1}^{t}(1+r)^{t-s}\hat{T}'(\bar{y}_t)\frac{\partial \bar{y}_t}{\partial y_t}\right) = 0 \\
\leftrightarrow\ & u_l(c_t,l_t)\frac{1}{\theta} + u_c(c_t,l_t)\left(1-\hat{T}'(\bar{y}_t)\right) = 0,
\end{aligned}
$$

where the first step employs equation (2.7), while the second follows from (2.6).

Turning to non-final working periods $t < R-1$, we will need to consider the envelope condition. For any $t < R-2$, we will have

$$
\frac{\partial V_{t+1}\left(a_{t+1},\bar{y}_t,\theta_{t+1}\right)}{\partial \bar{y}_t} = \beta\mathbb{E}\left[\frac{\partial V_{t+2}\left(a_{t+2},\bar{y}_{t+1},\theta_{t+2}\right)}{\partial \bar{y}_{t+1}}\frac{\partial \bar{y}_{t+1}}{\partial \bar{y}_t}\right],
$$

which can then be 'rolled forward' until the last working period to give

$$\frac{\partial V_{t+1}\left(a_{t+1}, \bar{y}_t, \theta_{t+1}\right)}{\partial \bar{y}_t} = \beta^{R-1-(t+1)} \mathbb{E}\left[\frac{\partial V_{R-1}\left(a_{R-1}, \bar{y}_{R-2}, \theta_{R-1}\right)}{\partial \bar{y}_{R-2}} \frac{\partial \bar{y}_{R-2}}{\partial \bar{y}_t}\right].$$

Since taxes come due in the last working period and depend not only on current, but also on past income, the envelope will be non-trivial in this case. Specifically, we have

$$\frac{\partial V_{R-1}\left(a_{R-1}, \bar{y}_{R-2}, \theta_{R-1}\right)}{\partial \bar{y}_{R-2}} = -u_c(c_{R-1}, l_{R-1}) \mathbb{T}'_{R-1}(\bar{y}_{R-1}) \frac{\partial \bar{y}_{R-1}}{\partial \bar{y}_{R-2}},$$

implying

$$\frac{\partial V_{t+1}\left(a_{t+1}, \bar{y}_t, \theta_{t+1}\right)}{\partial \bar{y}_t} = -\beta^{R-1-(t+1)} \mathbb{E}\left[u_c(c_{R-1}, l_{R-1}) \mathbb{T}'_{R-1}(\bar{y}_{R-1}) \frac{\partial \bar{y}_{R-1}}{\partial \bar{y}_t}\right].$$

Finally, by repeatedly employing equation (2.6) as well as (2.7), we get the final expression for the envelope condition as

$$\begin{aligned} \frac{\partial V_{t+1}\left(a_{t+1}, \bar{y}_t, \theta_{t+1}\right)}{\partial \bar{y}_t} &= -(1+r)^{R-1-t} \beta^{R-1-(t+1)} \mathbb{E}\left[u_c(c_{R-1}, l_{R-1}) \hat{T}'(\bar{y}_{R-1})\right] \\ &= (1+r) \mathbb{E}\left[u_c(c_{R-1}, l_{R-1}) \hat{T}'(\bar{y}_{R-1})\right], \end{aligned} \tag{2.11}$$

where the second step makes use of the standard assumption $\beta(1+r) = 1$.

Substituting the envelope into the first order condition, we get

$$u_l(c_t, l_t) \frac{1}{\theta} + u_c(c_t, l_t) - \mathbb{E}\left[u_c(c_{R-1}, l_{R-1}) \hat{T}'(\bar{y}_{R-1})\right] = 0. \tag{2.12}$$

Thus, in non-final working periods, labour supply is impacted not by the current marginal tax rate, but rather by the expected marginal tax rate at the end of the working life, weighted with the marginal utility of consumption in the last working period.

Having derived the first order conditions for the household, we can now substitute all of our preliminary results into the definition of the effective tax rate to get

$$\begin{aligned} \tau_{eff}^{t} &= 1 - \frac{MRS}{MRT} = 1 - \frac{u_l}{u_c \theta} \\ &= \mathbb{E}\left[\frac{u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)} \hat{T}'(\bar{y}_{R-1})\right]. \end{aligned} \tag{2.13}$$

Thus, for the case without intermediate collections, the effective marginal tax rate faced by households is given by the marginal utility weighted expected marginal tax rate at the end of the working life.

In order to better understand this object, it is worthwhile to consider how lifetime taxation differs from annual taxation, and how this will impact the effective marginal tax rate. First, since workers do not enjoy perfect foresight in a world with idiosyncratic risk, they will have to form expectations about their ex-post marginal tax rate, which they will update in each period of their working lives as new information arrives. Hence, contrary to Liebman (2002), households in a heterogeneous agent model will not face constant marginal tax rates over their lives. Given that, however, they receive successively more information in each period, one should suspect that effective tax rates might be smoothed significantly over the lifecycle compared with annual taxation. Second, workers will now face gaps between the time that income is earned, and when tax on that income comes due. This disparity will become important if households value income differently at different times of their lifecycle. A household that is borrowing constrained, or expects to be borrowing constrained at some point between the current period and the period in which taxes are due, will value income more today than at the point when taxes are due. This will, in effect, cause them to discount future tax payments more strongly, lowering the effective marginal tax rate. Finally, lifetime taxes are able to provide some level of insurance that annual taxes cannot. Specifically, they insure households against the risk of receiving a *history of bad shocks*. Thus, we should expect this to be reflected in the effective marginal tax rates, as households may care less about high marginal tax rates in states that follow a good earnings history compared to those that follow a poor earnings history.

Returning to the formulation of the effective marginal tax rate, we can in fact find all these three elements. To make this more clear, I will decompose the effective tax rate into three components, each corresponding to one of the above mentioned effects.

Using an identity from probability theory, namely $\mathbb{E}[A \cdot B] = \mathbb{E}[A]\,\mathbb{E}[B] + Cov(A, B)$, we can write the effective marginal tax rate as

$$
\begin{aligned}
\tau^t_{eff} &= \mathbb{E}\left[\frac{u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)} \hat{T}'(\bar{y}_{R-1})\right] \\
&= \mathbb{E}\left[\frac{u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)}\right] \cdot \mathbb{E}\left[\hat{T}'(\bar{y}_{R-1})\right] + Cov\left(\frac{u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)}, \hat{T}'(\bar{y}_{R-1})\right) \\
&= \mathbb{E}\left[\frac{u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)}\right] \cdot \mathbb{E}\left[\hat{T}'(\bar{y}_{R-1})\right] - \mathbb{E}\left[\hat{T}'(\bar{y}_{R-1})\right] + \mathbb{E}\left[\hat{T}'(\bar{y}_{R-1})\right] + Cov\left(\frac{u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)}, \hat{T}'(\bar{y}_{R-1})\right) \\
&= \mathbb{E}\left[\hat{T}'(\bar{y}_{R-1})\right] - \mathbb{E}\left[\frac{u_c(c_t, l_t) - u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)}\right] \mathbb{E}\left[\hat{T}'(\bar{y}_{R-1})\right] + Cov\left(\frac{u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)}, \hat{T}'(\bar{y}_{R-1})\right).
\end{aligned}
$$

The first term in this decomposition denotes the effect of the expected marginal tax rate at the end of the working life. As workers receive more information about their lifetime income, this term will update over the lifecycle and grow ever closer to the actual ex-post marginal tax rate. The second term captures the effect of borrowing constraints. A worker who currently is, or expects to be borrowing constrained in the future will have a higher marginal utility of consumption in the current period compared to the final working period. Because of this, she cares more about income today and discounts the expected ex-post marginal tax rate accordingly. Finally, the last term captures the effect of the insurance over income histories. If consumption and marginal tax rates are positively correlated, as should be expected when the tax system is progressive, this term will be negative, and hence lower the effective tax rate.

Proposition 6 summarises the results so far

Proposition 6 *In a lifecycle model with idiosyncratic risk and lifetime income taxation without tax collection in intermediate periods, the effective tax rate faced by households is given by*

$$\tau^t_{eff} = \mathbb{E}\left[\frac{u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)}\hat{T}'(\bar{y}_{R-1})\right],$$

and can be decomposed into three components as

$$\tau^t_{eff} = \underbrace{\mathbb{E}\left[\hat{T}'(\bar{y}_{R-1})\right]}_{\textit{Expected ex-post rate}} \underbrace{- \mathbb{E}\left[\frac{u_c(c_t, l_t) - u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)}\right]\mathbb{E}\left[\hat{T}'(\bar{y}_{R-1})\right]}_{\textit{Borrowing constraints}} + \underbrace{Cov\left(\frac{u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)}, \hat{T}'(\bar{y}_{R-1})\right)}_{\textit{Insurance component}}.$$

Next, I will turn to the somewhat more involved case of intermediate tax collections. In this case, income earned in the current period will not only affect the tax payment due in the final period, but also in all intermediate periods. Hence, we can no longer claim that the instantaneous effect of taxation, and the envelope condition will be mutually exclusive as in the previous case. Instead, consider the first order condition in period t

$$u_l(c_t, l_t)\frac{1}{\theta} + u_c(c_t, l_t)\left(1 - \mathbb{T}'_t(\bar{y}_t)\frac{\partial \bar{y}_t}{\partial y_t}\right) + \beta\mathbb{E}\left[\frac{\partial V_{t+1}\left(a_{t+1}, \bar{y}_t, \theta_{t+1}\right)}{\partial \bar{y}_t}\frac{\partial \bar{y}_t}{\partial y_t}\right] = 0.$$

By the definitions (2.8) and (2.6), we will have

$$\mathbb{T}'_t(\bar{y}_t)\frac{\partial \bar{y}_t}{\partial y_t} = \hat{T}'(\bar{y}_t),$$

i.e. the instantaneous marginal tax rate. Additionally, we can no longer 'roll forward' the envelope condition as in the previous case, since past average income now directly impacts the workers budget constraint in each period through the tax rebate as well as the taxes due in that period. Specifically, we have

$$\begin{aligned}
\frac{\partial V_{t+1}\left(a_{t+1}, \bar{y}_t, \theta_{t+1}\right)}{\partial \bar{y}_t} &= u_c(c_{t+1}, l_{t+1}) \left(\sum_{s=1}^{t+1} (1+r)^{t+1-s} \hat{T}'(\bar{y}_{t+1}) \frac{\partial \bar{y}_{t+1}}{\partial \bar{y}_t} - (1+r) \sum_{s=1}^{t} (1+r)^{t-s} \hat{T}'(\bar{y}_t) \right) \\
&\quad + \beta \mathbb{E}\left[\frac{\partial V_{t+2}\left(a_{t+2}, \bar{y}_{t+1}, \theta_{t+2}\right)}{\partial \bar{y}_{t+1}} \frac{\partial \bar{y}_{t+1}}{\partial \bar{y}_t} \right] \\
&= u_c(c_{t+1}, l_{t+1})(1+r) \sum_{s=1}^{t} (1+r)^{t-s} \left(\hat{T}'(\bar{y}_{t+1}) - \hat{T}'(\bar{y}_t) \right) \\
&\quad + \frac{\sum_{s=1}^{t} (1+r)^{t-s}}{\sum_{s=1}^{t+1} (1+r)^{t+1-s}} \mathbb{E}\left[\frac{\partial V_{t+2}\left(a_{t+2}, \bar{y}_{t+1}, \theta_{t+2}\right)}{\partial \bar{y}_{t+1}} \right].
\end{aligned}$$

Iterating this equation forward one period and collecting terms, we get

$$\begin{aligned}
\frac{\partial V_{t+1}\left(a_{t+1}, \bar{y}_t, \theta_{t+1}\right)}{\partial \bar{y}_t} &= -(1+r) \sum_{s=1}^{t} (1+r)^{t-s} \hat{T}'(\bar{y}_t) u_c(c_{t+1}, l_{t+1}) \\
&\quad + (1+r) \sum_{s=1}^{t} (1+r)^{t-s} \hat{T}'(\bar{y}_{t+1}) \mathbb{E}\left[u_c(c_{t+1}, l_{t+1}) - u_c(c_{t+2}, l_{t+2}) \right] \\
&\quad + (1+r) \sum_{s=1}^{t} (1+r)^{t-s} \mathbb{E}\left[\hat{T}'(\bar{y}_{t+2}) u_c(c_{t+2}, l_{t+2}) \right] \\
&\quad + \frac{\sum_{s=1}^{t} (1+r)^{t-s}}{\sum_{s=1}^{t+2} (1+r)^{t+2-s}} \mathbb{E}\left[\frac{\partial V_{t+3}\left(a_{t+3}, \bar{y}_{t+2}, \theta_{t+3}\right)}{\partial \bar{y}_{t+2}} \right].
\end{aligned}$$

Once again, we know that the envelope in period R will be zero, as annualised lifetime income does not enter the household problem during retirement. Hence, we can repeatedly substitute the above into itself to get the following expression for the envelope condition

$$\begin{aligned}
\frac{\partial V_{t+1}\left(a_{t+1}, \bar{y}_t, \theta_{t+1}\right)}{\partial \bar{y}_t} &= -(1+r) \sum_{s=1}^{t} (1+r)^{t-s} \hat{T}'(\bar{y}_t) u_c(c_{t+1}, l_{t+1}) \\
&\quad + (1+r) \sum_{s=1}^{t} (1+r)^{t-s} \sum_{q=t+1}^{R-2} \mathbb{E}\left[\hat{T}'(\bar{y}_q) \left(u_c(c_q, l_q) - u_c(c_{q+1}, l_{q+1}) \right) \right] \\
&\quad + (1+r) \sum_{s=1}^{t} (1+r)^{t-s} \mathbb{E}\left[\hat{T}'(\bar{y}_{R-1}) u_c(c_{R-1}, l_{R-1}) \right]
\end{aligned}$$

By substituting the above into the first order condition and once more employing (2.6), we finally get

$$\begin{aligned} u_l(c_t,l_t)\frac{1}{\theta} + u_c(c_t,l_t) - \sum_{q=t}^{R-2}\mathbb{E}_t\left[\hat{T}'(\bar{y}_q)\left(u_c(c_q,l_q) - u_c(c_{q+1},l_{q+1})\right)\right] \\ - \mathbb{E}\left[\hat{T}'(\bar{y}_{R-1})u_c(c_{R-1},l_{R-1})\right] = 0. \end{aligned} \tag{2.14}$$

Compared to the corresponding first order condition in the case without intermediate tax payments, we can immediately see that now it is not only the marginal utility weighted expected *ex-post* marginal tax rate that matters for labour supply decisions, but rather all intermediate rates as well. Specifically, each intermediate rate will matter in the case that a household expects to value income differently at the time that the tax is paid and that the rebate is granted in the following period.

Once again employing the definition of the effective marginal tax rate, we can immediately write it as

$$\tau_{eff}^t = \sum_{q=t}^{R-2}\mathbb{E}\left[\hat{T}'(\bar{y}_q)\frac{u_c(c_q,l_q) - u_c(c_{q+1},l_{q+1})}{u_c(c_t,l_t)}\right] + \mathbb{E}\left[\hat{T}'(\bar{y}_{R-1})\frac{u_c(c_{R-1},l_{R-1})}{u_c(c_t,l_t)}\right]$$

While the second term is identical to the case without intermediate collections, the first term is novel. This term captures the aforementioned intermediate tax rates that become relevant precisely whenever a household values income more in the current than in the next period, i.e. whenever it is borrowing constrained. Hence, we will have an additional component in this case that captures *intermediate borrowing constraints* and will increase the effective marginal tax rate. For a household that expects to never be borrowing constrained, this term will be zero. Additionally, since the tax function is non-linear, households will care more about borrowing constraints in states where they pay a high marginal tax compared to those where they pay low marginal taxes. This will lead to an additional *intermediate insurance component*, where households are not insured against bad income histories, but rather against becoming borrowing constrained in the future.

Applying the same technique as in the previous proposition, we arrive at the final result of this section

Proposition 7 *In a lifecycle model with idiosyncratic risk and lifetime income taxation with tax collection in intermediate periods, the effective tax rate faced by households is given by*

$$\tau_{eff}^{t} = \sum_{q=t}^{R-2} \mathbb{E}\left[\hat{T}'(\bar{y}_q)\frac{u_c(c_q, l_q) - u_c(c_{q+1}, l_{q+1})}{u_c(c_t, l_t)}\right] + \mathbb{E}\left[\hat{T}'(\bar{y}_{R-1})\frac{u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)}\right],$$

and can be decomposed into five components as

$$\begin{aligned}\tau_{eff}^{t} = \; & \underbrace{\mathbb{E}\left[\hat{T}'(\bar{y}_{R-1})\right]}_{\textit{Expected ex-post rate}} \; \underbrace{- \mathbb{E}\left[\frac{u_c(c_t, l_t) - u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)}\right]\mathbb{E}\left[\hat{T}'(\bar{y}_{R-1})\right]}_{\textit{Borrowing constraints}} + \underbrace{Cov\left(\frac{u_c(c_{R-1}, l_{R-1})}{u_c(c_t, l_t)}, \hat{T}'(\bar{y}_{R-1})\right)}_{\textit{Insurance component}} \\ & + \underbrace{\sum_{q=t}^{R-2} \mathbb{E}\left[\hat{T}'(\bar{y}_q)\right]\left[\frac{u_c(c_q, l_q) - u_c(c_{q+1}, l_{q+1})}{u_c(c_t, l_t)}\right]}_{\textit{Intermediate borrowing constraints}} + \underbrace{\sum_{q=t}^{R-2} Cov\left(\frac{u_c(c_q, l_q) - u_c(c_{q+1}, l_{q+1})}{u_c(c_t, l_t)}, \hat{T}'(\bar{y}_q)\right)}_{\textit{Intermediate Insurance}}.\end{aligned}$$

Having derived, as well as decomposed, the effective marginal tax rates that would be induced by a tax on lifetime income, the next section will concern itself with the quantitative analysis of this effective marginal tax rate and its components.

2.4.3 Approximating Effective Tax Rates

This section will present a quantitative approximation of the effective marginal tax rates under lifetime taxation with intermediate collections, thereby quantifying the results developed in proposition 7. In order to avoid having to solve the full model with endogenous labour supply developed in section 2.4.2, I will use the model with exogenous income from section 2.3.2 as an 'auxiliary model'. Specifically, I will take the solution to this model and employ it to calculate the expectations and covariances in proposition 7[10].

The derivative of the lifetime tax function $\hat{T}$ with respect to lifetime income is calculated numerically after deriving this function from the baseline model with annual taxes by averaging the NPV of taxes paid by all agents within a certain income group.

Figure 2.5 shows the average effective marginal tax rate for each age group in the case of annual and lifetime income taxation, as well as the decomposition of the effective tax rate in the case of annual income taxation.

As can be seen from the figure, the overall size of the effective tax rate in the case of lifetime taxation in driven by workers' expectations about their ex-post marginal tax rate on lifetime income. Surprisingly, this expectation lies above the average effective tax rate of the pre-reform state in every period, suggesting that there exists a non-trivial relationship between the functions $T(y)$ and $\hat{T}(\bar{y})$ that should be explored more closely in future work. Driven by this, the average effective marginal tax rate in the post-reform state lies above that of the pre-reform state in every age group.

Turning to the decomposition, we find that the four components capturing the effects of borrowing constraints and insurance matter only during the first part of the working life. Young workers in this model are very likely to be borrowing constrained, leading to a significant reduction in the effective tax rate in the first years of labour market activity. This is because a borrowing constrained worker cares less about taxes they will have to pay at some point in the future and more about boosting his current consumption through increased labour earnings. At the same time, this is somewhat counteracted by the effect of intermediate borrowing constraints. While taxes paid in period t will be rebated to the worker in period $t + 1$, this only removes the effect of intermediate collection for those workers who are not currently borrowing

[10]This approximation is necessary as the effective marginal tax rate is not always well-defined outside of equilibrium, making the solution of the full model difficult. I aim to include the results of the model with endogenous labour supply in future versions of this paper.

constrained. Since, however, the marginal taxes on lifetime income are lower during the first part of the earnings life, borrowing constraints have a negative net effect on the effective marginal tax rate.

Turning to the effect of the two insurance components we find that they both reduce the effective marginal tax rate during the first half of the working life. The first, and quantitatively more significant, stems from the fact that workers expect to face low ex-post marginal tax rates on their lifetime income in states where their consumption is low (and correspondingly their marginal utility of consumption is large). Thus, when forming expectations over their ex-post marginal tax rate they weigh these not evenly, but rather give larger weight to low tax states and lower weight to those states in which both taxes and consumption are high, reducing their effective marginal rate below the unconditional expectation of the ex-post rate.

Finally, there is a small intermediate insurance component that captures the effect that agents are more likely to be borrowing constrained in states in which they are relatively poor, and are consequently facing a lower tax on their current lifetime income.

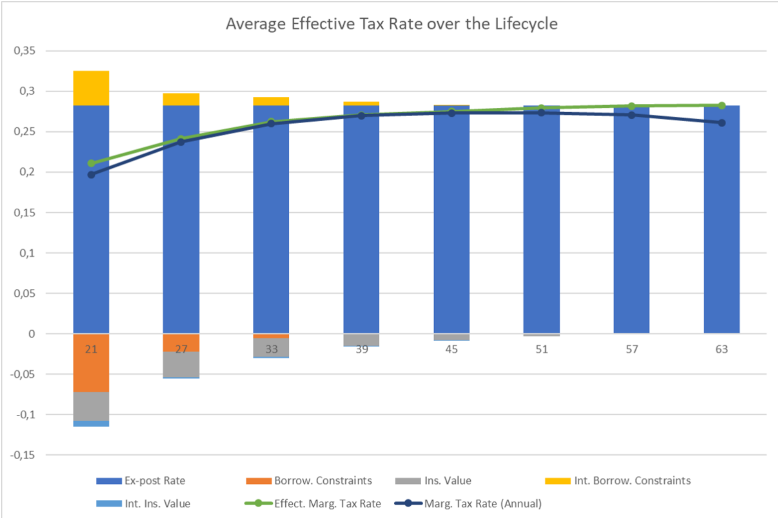
Both of these insurance effects disappear over the course of the working life. The former goes to zero as workers' expectations about their ex-post marginal rates as well as their ex-post consumption stabilize when more and more uncertainty is resolved. The latter goes to zero simply because agents are less and less likely to be borrowing constrained as they grow older.

2.5 Conclusion

In this paper, I explore the equity- as well as efficiency considerations of lifetime income taxation in a heterogeneous agent model with idiosyncratic risk.

I demonstrate that the classical argument for equity gains of lifetime income taxation relies on the particular structure of perfect foresight models employed to analyse such reforms previously.

Figure 2.5: Average Effective Tax Rates over the Lifecycle



The average effective marginal tax rate at each age is calculated for both the pre- and post-reform steady state with the respective distribution.

As such an argument cannot be made in a model with idiosyncratic risk, answering this question will necessitate quantitative analysis. To this end, I build a lifecycle model with idiosyncratic risk. I parameterize this model to reflect key aspects of the US economy and show that a move from annual to lifetime taxation would lead to a small welfare loss that is equivalent to 0.16% of average consumption. This loss is mainly due to increased savings by workers in order to insure against possibly large tax burdens in the future. This leads to 'back-loading' of consumption which is detrimental to welfare.

Turning to efficiency considerations, I demonstrate that the 'constant marginal tax rate' result is only valid in the absence of borrowing constraints. If borrowing constraints are present, the effective marginal tax rate is not constant over the lifecycle even in a model with perfect foresight. Rather, in periods where households are borrowing constrained, the marginal tax rate is depressed. I then derive a formula for the effective marginal tax rate induced by a lifetime income tax in a heterogeneous agent model with idiosyncratic risk and show that it can be separated into the expected ex-post marginal tax rate, the effects of borrowing constraints, and the effects of insurance against lifetime income risks.

I then approximate the effective marginal tax rate under lifetime taxation and demonstrate that the effects of insurance and borrowing constraints are quantitatively important, leading, in aggregate, to a depression of the effective marginal tax rate below the expected ex-post rate. Additionally, preliminary calculations demonstrate that the average effective marginal tax rate under a lifetime income tax will be above that of an annual income tax which is in contrast to the results of the previous literature.

As I demonstrate, the conclusions drawn about the desirability of lifetime income taxation in a perfect foresight world do not generalize to a mode with idiosyncratic risk. I show that there is evidence that such a reform may be undesirable for both equity and efficiency reasons. Future research should focus on gaining a better understanding of the relationship between marginal taxes under annual and lifetime taxation in order to better understand why such a reform may be undesirable on efficiency grounds.

Chapter 3

Bank Funding Structure and Risk Taking

JOINT WITH M. PETRICEK, C. GARCIA GALINDO, AND A. FERRRARI

3.1 Introduction

There are two established, opposing theoretical results about the effect of leverage on risk taking by banks. First, due to limited liability, expected returns on equity investment increase with an increased riskiness of the portfolio. As a bank's equity holders are protected from the left tail of the returns to assets distribution by limited liability, they have an incentive to increase the variance of the distribution by taking on more risk. On the other hand, callable demand deposits constitute a substantial part of bank debt. On average, almost 20 percent of these are above the amount insured by the Federal deposit insurance corporation. This provides depositors with a strong incentive to monitor as well as with tools to punish excessive risk taking.

There is inconclusive empirical evidence on which of the two effects prevails. Some authors such as Koudstaal and van Wijnbergen (2012) find that higher leverage leads to less risk taking, while other such as Acosta Smith et al. (2017) find that an increase in capital requirements leads to a decrease in risk taking. Finally, there are some authors such as Jacques and Nigro (1997) that find no effect at all. As we argue below, the results in all these papers are, however, influenced by endogeneity issues. Our aim in this paper is then to estimate the causal effect of leverage on risk taking behaviour of banks. In doing so we add to the literature by proposing a novel approach to addressing endogeneity in this setting.

We identify two sources of endogeneity and propose a method to address them. First, an increase in leverage may incentivise banks to pursue riskier investment, but at the same time the demand and supply of deposits, which drive a bank's leverage given equity, can also be affected by a bank's risk taking, giving rise to reverse causality. Such an effect may arise if a bank becomes known for making risky investment choices and is consequently avoided by depositors. Second, shocks, observable or non-observable, common to both assets and liabilities of a bank, if omitted, can cause a bias in the estimate of the effect of leverage on risk taking. This, in particular, is an issue that will arise whenever one measures risk as realised risk in the portfolio, as is common in the previous literature.

We conduct our empirical analysis by using instrumental variable estimation. We first address the issue of reverse causality by making use of bank office level data on deposits for US banks and geographically granular unemployment data. We argue that local unemployment rates are exogenous to the risk taking of a bank as a whole, and construct an instrument based on bank's exposure to deposit supply shocks caused by changes in local unemployment rates, to use in our final regression of leverage on risk. However, as is often the case in the empirical literature, if risk taking is approximated by using a risk measure of existing portfolios, the issue of omitted shocks, which may be common to both assets and liabilities, remains. To address this, we construct a measure of risk taking based solely on newly issued mortgage loans, by considering the universe of mortgage loan applications from 1999 to 2016. We argue that, while the existing portfolio of a bank can be affected by geographical area specific shocks which affect deposits and leverage at the same time, newly issued loans are chosen by banks after the shocks have realized. Hence, the riskiness of newly issued loans is a choice by the bank, unaffected by local area shocks.

Our results confirm that limited liability induces banks to take on more risk after an exogenous increase in leverage, with more leveraged banks being significantly more likely to issue risky mortgage loans. We find that decreasing a bank's leverage ratio by 1 percentage point (which corresponds to an increase in leverage), will lead to this bank originating a loan of 'average' risk with a 3.8% higher probability. In a second estimation we find that this translates into a 1% increase in the predicted median probability of default of loans issued by this bank.

The remainder of this paper is structured as follows: section 3.2 provides a review of the relevant literature, sections 3.3 explains the methodology and the data. Section 3.4 presents the results, while section 3.5 discusses their robustness with respect to the measure of risk taking. Finally, section 3.6 concludes and discusses the policy implications of the findings.

3.2 Literature Review

There are several theoretical papers on how banks' funding structure should impact their risk taking. Jensen and Meckling (1976) show that for a firm the decision to take on debt is equivalent to buying a call option from its creditors. When the debt is due, they can either choose to redeem the bond (buy back the firm, so to speak) or not to. Since the value of such an option is increasing in volatility (see for example Black and Scholes (1973)), firms have an interest to take on higher amount of risk than without debt financing. On the other hand, as Laeven and Levine (2009) point out, requiring banks to hold more capital may not necessarily reduce these incentives if this capital is raised by issuing equity to new shareholders. Simply adding more shareholders with the same incentives may not actually alleviate the issue. Instead, shareholders may decide to make up for the higher cost of capital by taking on even more risky projects in the spirit of Koehn and Santomero (1980). Finally, Diamond and Rajan (2001) show that demand deposits can have a disciplining effect on banks. Our paper contributes to the literature by providing a well-identified answer about the causal effect of leverage on risk taking.

In the empirical literature, there are two ways that previous work has addressed the question of how leverage impacts bank risk taking. The first strain of papers seeks to provide an answer through a direct regression of a measure of leverage on some measure of risk. Altunbas et al. (2007) aim to identify the relationship between leverage and risk by means of a seemingly unrelated regression design that relates changes in capital and risk. They use loan-loss provisions as a proxy for the risk taken on by the bank and find that in their whole sample, banks with a higher equity to asset ratio will take on more risk, while the relationship is negative for the most efficient banks in the sample. Jacques and Nigro (1997) employ an approach that uses a regulatory pressure variable, a measure of how far a bank's equity holdings are from the regulatory threshold, to identify the effect of leverage on risk, but can not refute the null hypothesis that leverage has no effect on risk taking. Koudstaal and van Wijnbergen (2012) aim to identify the effect of leverage on risk by regressing the standard deviation of returns on assets on lagged leverage while controlling for market volatility. They find that higher leverage leads to less risk taking, but that this result is entirely driven by low leverage banks. Highly leveraged banks, they find, do not react to changes in leverage. Similarly to the above paper, Shrieves and Dahl (1992) find that banks take on more risk when there is a positive shock to capital by using a simultaneous regression framework. As we will argue in our methodological section below, all these papers have shortcomings in two ways: first, they fail to provide a convincing identification in the sense that leverage cannot be seen as exogenous in any of the above models. Second, as all the papers employ some measure of portfolio risk, they consider a rather noisy measure of risk taking which is also impacted by market conditions, which may in turn be impacting banks leverage decisions.

The second strain of papers in the literature attempts to provide exogenous variation to leverage by evaluating the effect of policies that impact banks ability to leverage out. Laeven and Levine (2009) find that requiring banks with an owner that holds a significant voting share to hold more capital has the effect of reducing risk taking, while the opposite is true for widely held banks. Acosta Smith et al. (2017) find that the introduction of the leverage ratio requirements in the Basel III framework did cause banks to increase their capital holdings and reduce their risk taking. Finally, Ashraf et al. (2016) show that the introduction of risk weighted capital standards led to a reduction of bank portfolio risk in Pakistan. We add to the empirical literature by providing clean identification of the causal effect of leverage on risk taking, both by adding a new instrument for leverage and by using a measure of risk taking (a banks decision to issue certain loans) that is much less likely to be subject to outside influences and previous choices than the portfolio based measure currently used in the literature. While these papers suffer from the endogeneity associated with leverage to a smaller degree, we believe that our identification is superior as we do not need to rely on the assumption that banks did not react to news or rumors of potential policy changes prior to the implementation of the reform. Additionally, all of these papers rely once more on portfolio based measures of risk, while the present work employs a much more direct measure of risk taking.

In parallel work to this, Ohlrogge (2017) uses an approach that is similar to the one taken in this paper and comes to results that confirm our analysis.

Methodologically, our paper is related to a paper by Bartik (1993) that employs local industry shares to identify the impact of labour supply on wages. This approach has been recently formalized and discussed by Goldsmith-Pinkham et al. (2017).

3.3 Methodology

3.3.1 Overview

In this section we explain the endogeneity issues that plague the analysis of leverage and risk taking, as well as our methodology to tackle them. Before diving deeper into those issues, some definitions will prove useful in the following discussion. First, we will use the term **risk taking (behaviour)** as an act of making new investments (issuing new loans) with different degree of riskiness attached to them. This is the subject of our analysis. It is important to distinguish it from the term **riskiness of the portfolio**, which is defined by the riskiness attached to loans which have been issued in the past. Variation in riskiness of the portfolio can be caused by both risk taking behaviour and by current and past shocks absorbed by the portfolio. Although the riskiness of the portfolio is often used to proxy risk taking behaviour in the literature, the distinction will be important in understanding the issue of endogeneity and our identification.

Most commonly, the literature defines an increase in leverage as an increase in debt financing relative to equity financing by a bank. Basel III however defines a leverage ratio as core equity relative to total assets. For the purpose of easy application to the most recent regulation, we use the latter as a measure of leverage in our econometric analysis. Thus, an *increase* in leverage due to an increase in debt financing corresponds to a *decrease* in the leverage ratio.

We identify two sources of endogeneity which prevent a causal interpretation of a simple regression of leverage on some commonly used measure of riskiness of the portfolio.

- Simultaneity/reverse causality: For a given level of equity, more deposits may incentivise banks to undertake riskier investments, but the demand and supply of deposits are also affected by the riskiness of the portfolio and a bank's risk taking behavior.
- Omitted shocks common to the portfolio and the deposits.

To tackle the first source of endogeneity we build an instrument for leverage that is exogenous to a bank's risk choices. To this end, we use data on deposits at the office level provided by the FDIC. This detailed geographical information on deposits enables us to compute bank exposure to local unemployment variation, which we use as an instrument providing us with variation in leverage exogenous to bank behaviour. This measure, however, can still be endogenous if it relies on the existing portfolio. If some exogenous shocks to economic activity occur, they are likely to not only affect deposits (and leverage) through unemployment but also through the riskiness of the existing portfolio. To tackle this issue, we construct a measure of risk taking based on the new issuance of mortgage loans. We argue that while riskiness of the portfolio may be affected by some shocks which are common to both deposits and assets, the riskiness of new issuances are a choice for banks.

Following the procedure described above, we conduct our empirical analysis by, first, constructing an instrument to assure variation in leverage that is independent of bank risk taking and constructing a measure of risk taking based on newly issued mortgage loans, which is independent of local area shocks. We then regress our instrumented leverage ratio on the riskiness of new issuances in order to estimate the causal effect of leverage on bank risk taking. The methodology is explained in more detail after a discussion of the data employed.

3.3.2 Data

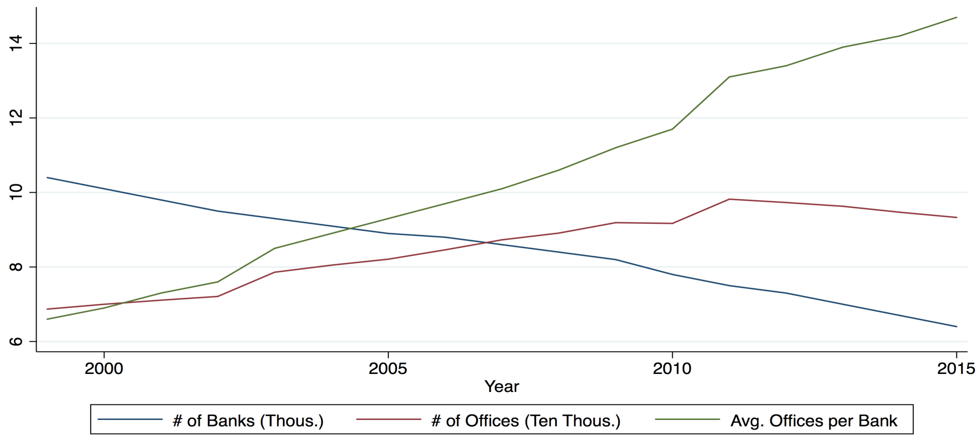
There are four main sources of data that we use in the analysis. The bank-office level deposit data from the FDIC, and the local unemployment data from the Bureau of Labour Statistics are used in constructing the instruments. We use the Home Mortgage Disclosure Act data on the universe of mortgage applications from the FFIEC to construct a measure of risk taking. Finally, we add balance sheet data from the FDIC to calculate the leverage ratio in the IV estimation of the effect of leverage on risk taking.

Summary of deposits (FDIC)

Summary of deposits data is annual data on the level of deposits at the bank office level. For every office, for every bank operating in the US and insured by the FDIC, the amount of deposits as of June 31st is reported. Along with the deposit level, the data contains detailed geographic and demographic information for every office as well as an identifier for the owning bank. This identifier is then used to merge this data with balance sheet data also provided by the FDIC.

For the purpose of our analysis, the data for every bank is collapsed at a relevant geographical area level. We will be using the Core Based Statistical Areas. A Core Based Statistical Area (CBSA) consists of one or more counties (or equivalents) anchored by an urban center of at least 10,000 people plus adjacent counties that are socioeconomically tied to the urban center by commuting. Not all counties are a part of a CBSA. Around 10% of all observations come from counties which are not part of any CBSAs adding up to around 5% of all deposits. We aggregate these counties at the state level into CBSA equivalents and brand them as *rural state areas*. Figure 3.1 presents some descriptive statistics from the Summary of Deposits data. Two features of the data are noteworthy. First, over our sample period of 1999 to 2015, there is a significant consolidation in the banking market, with the number of banks decreasing by about 40%, while the number of offices per bank more than double. We take care that this development does not impact our analysis by removing an office sold by bank i to bank j in year t from the sample in years t and $t-1$, so that the change in deposits between $t-1$ and t for neither bank i nor bank j is affected by the transfer of ownership of that office.

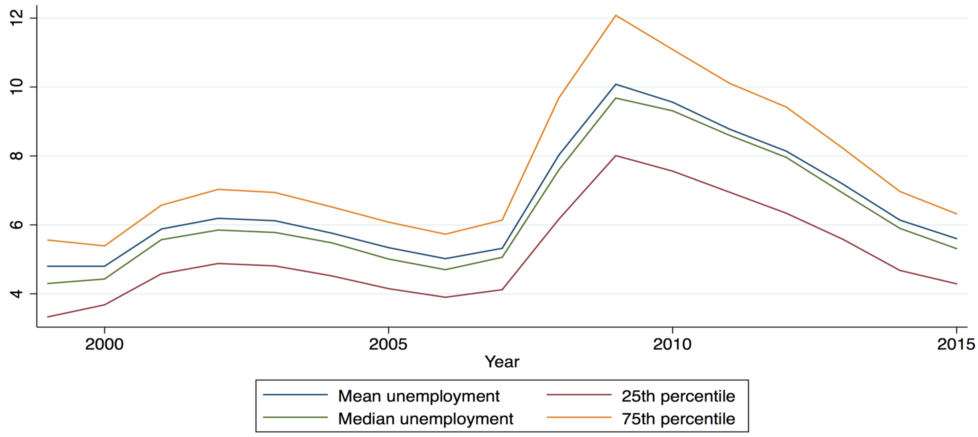
Figure 3.1: Bank offices data



Local area unemployment statistics (BLS)

Local area unemployment statistics provide monthly data on unemployment at the county level. Since the relevant geographical area in constructing the instruments is the CBSA, we aggregate the statistics to the CBSA level at yearly frequency. Figure 3.2 presents the descriptive statistics of unemployment rate of the CBSAs in the US within our sample period. This graph shows that there is indeed significant heterogeneity between CBSAs, when it comes to unemployment, providing variation to exploit when constructing our instrument. As we show later, this heterogeneity in unemployment rates will have heterogenous effects on banks, which are differentiated by the geographic composition of their deposit holdings. Finally, the first stage regression of our instrumental variable approach shows that this variation is strongly correlated with leverage at the bank level.

Figure 3.2: CBSA unemployment statistics



Home mortgage disclosure act data (FFIEC)

The Home Mortgage Disclosure Act (HMDA) mandates that banks above a set threshold of assets issue detailed reports on their mortgage applications, lending and purchases. The reporting is done through the Loan Application Registries (LAR) and includes all mortgage loan applications within a year. Moreover, the registries contain some characteristics of the applicant and potential co-applicant (ethnicity, race, gender, income), as well as characteristics of the loan (amount, type, purpose, rate spread for some, occupancy), the property (type, census tract, etc.), the census tract in which the property is located (income relative to the MSA, minority population, number of housing units, etc.), as well as the action taken by the bank (origination, denial and its reason, sale to an institution like Freddie Mac).

Our construction of the measure of risk taking loosely follows DellAriccia et al. (2012) and relies on the loan to income ratio. The loan to income ratio is computed as the total loan amount in the application over the total gross annual income an institution relied upon in making the credit decision[1]. To add to the methodology on a measure of risk taking we also use the data

[1] Gross annual income is not registered in HMDA due to four possible reasons: (i.) multi-family dwellings, (ii.) income was not registered in the loan purchase documentation, (iii.) loans to bank employees, (iv.) loans to non natural persons. These cases are excluded from the estimation as described in the methodology section

on origination. We define origination as an application which has been accepted and then either originated or refused by the applicant, a purchase of a loan, or a preapproved request. We define a non-origination as an application denied by the bank or a denied prerequest. We ignore all applications withdrawn by the applicants or applications closed for incompleteness.

The LAR data reports all applications, accepted or rejected. Table 3.1 provides the statistics on the origination ratio between 2004 and 2012. The share of originated loan applications (see column (1)) decreased from 74% in 2004 to 67% in 2007. In 2009, the origination ratio increased sharply and then gradually increased to 80% in 2012. The sharp increase reflects the crisis, which has decreased the demand for loans and forced the worse potential borrowers out of the market. The remaining pool was of higher quality which increased banks willingness to lend to remaining applicants.

Table 3.1: Reason for denial

Year	Origin. ratio	Reason for denial					
		DtI	Empl. hist.	Cred. hist.	Coll.	Dwnpay.	Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
2004	.744	.131	.010	.299	.113	.0154	.432
2005	.729	.122	.0111	.267	.122	.012	.466
2006	.715	.150	.0129	.284	.154	.017	.381
2007	.677	.173	.0123	.272	.193	.017	.332
2008	.681	.205	.0114	.265	.250	.018	.249
2009	.767	.227	.0129	.209	.310	.020	.220
2010	.776	.222	.0134	.202	.252	.021	.288
2011	.766	.214	.013	.223	.241	.021	.285
2012	.794	.213	.013	.233	.218	.024	.301

Table 3.1 also reports the shares of prevailing reasons for rejecting a loan. Insufficient collateral (column (5), high debt-to-income (column (2)) and poor credit history (column (4)) explain the bulk of the rejection decisions. The effect of the crisis is evident in the spike of the share of rejections due to insufficient collateral in 2009 when house prices collapsed. We take into account this crisis effect by including time fixed effects in our estimation of the risk measures.

On top of the loan application data, the LAR reports also the information about the loans purchased by banks. Tables 3.2 provides the statistics about the characteristics of all applications, the accepted applications, the rejected applications and the purchased loans. The table shows that accepted loans have, on average, a lower loan-to-income ratio than rejected loans.

Table 3.2: Loan-to-Income ratio

	All		Accepted		Rejected		Purchased	
	mean	p50	mean	p50	mean	p50	mean	p50
2004	2.274	2.083	2.227	2.103	2.408	2.000	2.502	2.390
2005	2.281	2.098	2.217	2.106	2.452	2.077	2.480	2.400
2006	2.188	1.974	2.109	1.951	2.385	2.038	2.322	2.233
2008	2.304	2.098	2.184	2.059	2.553	2.200	2.547	2.451
2008	2.433	2.209	2.311	2.174	2.687	2.300	2.735	2.607
2009	2.573	2.243	2.420	2.211	3.060	2.372	2.539	2.427
2010	2.430	2.139	2.309	2.115	2.846	2.244	2.673	2.540
2011	2.349	2.023	2.222	2.000	2.761	2.083	2.587	2.435
2012	2.384	2.054	2.291	2.051	2.735	2.065	2.573	2.409

Balance sheet data (FDIC)

To construct the leverage measure of banks, we use balance sheet data provided by the FDIC. The data is available at quarterly level and includes income statements as well as several performance ratios.

3.3.3 Exogenous Variation in Deposits: Two Instruments

The aim here is to construct an instrument to assure that the variation in leverage is independent of risk taking. We do so in two ways: i) estimating shocks to local deposits caused by unemployment changes which we then aggregate for each bank by computing the weighted average of these local shocks; ii) or directly computing the weighted average of unemployment changes and using this as an instrument for leverage.

To this end we use data on deposits at the bank-office level, administered by the Federal Deposit Insurance Corporation (FDIC) and the Local Area Unemployment Statistics administered by the Bureau of Labour Statistics. The first dataset contains yearly information on the level of deposits for all offices of all banks insured by the FDIC, together with the demographic information on the office and the bank which owns it. The second dataset provides monthly unemployment figures at county level. The relevant geographical definition in our analysis is the

Core Based Statistical Area.

The rationale behind these instruments for bank level deposit growth rates follows closely Bartik (1993), whose approach has been extensively analysed in Goldsmith-Pinkham et al. (2017). The standard idea behind this approach is that when one is interested in a parameter, say, the elasticity of labour supply, using changes in wages and employment growth rates at local level, one should be concerned with the endogeneity of local employment growth. To solve this issue, the Bartik approach suggests to define an instrument as the local employment growth predicted by interacting local industry employment shares with national industry employment growth rates.

In our setting we follow a similar logic but we apply it to a different level of granularity of the data. Our potentially endogenous object is the deposit growth rate of banks. We therefore build as an instrument the predicted change in deposits for a bank in a given period as the interaction between the bank's geographical area deposit share and the change in deposits in the geographical area.

We do so in two different ways to avoid any further endogeneity concern or feedback loop between bank and area level deposit changes: i) we predict the change in deposits in a geographical area in a given period based on the change of local unemployment in that period and use this fitted value as our instrument; ii) we use the change in unemployment in the geographical area directly (not using it to predict deposits) as the instrument.

Before discussing the two approaches in further detail, key differences between our strategy and the standard Bartik instruments should be highlighted. As mentioned above, the literature employs this approach to solve endogeneity problems. In contrast, we use the geographical area shares as a mean of aggregation, not as a solution to endogeneity per se. We adopt as instrumental variables for the leverage the "relevant" changes in local unemployment or deposit supply, where "relevant" is to be read as weighted by geographical area deposit composition. The adoption of this specific aggregation strategy serves two distinct purposes: first and foremost is an appealing way of aggregating geographical area specific changes to the bank level; second it eliminates any further endogeneity concerns.

The two approaches are formalized below.

Instrumenting deposits: In building the first instrument we regress the growth rate of total deposits in a geographical area on the change in the local unemployment rate. We brand the fitted values from this model at the geographical area level as shocks to deposit supply at the geographical area level. For each bank in each year we then compute the exposure to this variation in deposit supply as a weighted average of these shocks using the deposits each bank holds in a particular area as weights. This implies the following procedure,

$$\Delta dep_{i,t} = \alpha_0 + \gamma_i + \eta_t + \beta \Delta unemp_{i,t} + \epsilon_{i,t} \tag{3.1}$$

where $\Delta dep_{i,t}$ denotes the growth rate of deposits in a geographical area i in period t, γ_i and η_t denote geographical area and time fixed effects, and $\Delta unemp_{i,t}$ denotes a change in unemployment rate in geographical area i in period t. We call the fitted values from the model above local deposit supply shocks. To compute the exposure of a particular bank in a particular period to these shocks, which will serve as an instrument for leverage in our final estimation, we compute the weighted average of these shocks for every bank, where then we use the deposit this particular bank holds in different areas as weights. For bank b, operating in areas $i = 1..I$, this implies:

$$\Delta \hat{dep}_{b,t} = \frac{\sum_{i=1}^{I} dep_{b,i,t} \Delta \hat{dep}_{i,t}}{\sum_{i=1}^{I} dep_{b,i,t}} \tag{3.2}$$

where $dep_{b,i,t}$ denotes the deposits bank b holds in geographical area i in period t. We use the measure $\Delta \hat{dep}_{b,t}$ as one of the possible instruments in the final estimation of the effect of leverage on risk taking.

Table 3.3.3 presents the estimation results for equation 3.1. Results, as expected, prove a negative and highly significant effect of changes in unemployment on deposit growth rates at the CBSA level. An increase in unemployment change in a CBSA by one percentage point decreases the deposit growth rate in that area by 0.43 percentage points after controlling for the CBSA and year fixed effects.

Table 3.3: First preliminary stage

	(1)
	Δln(DEPOSITS)
Δ UNEMP	-0.432****
	(0.117)
CONSTANT	0.0310****
TIME FE	YES
CBSA FE	YES
N	21450
R^2	0.014

STANDARD ERRORS IN PARENTHESES
** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

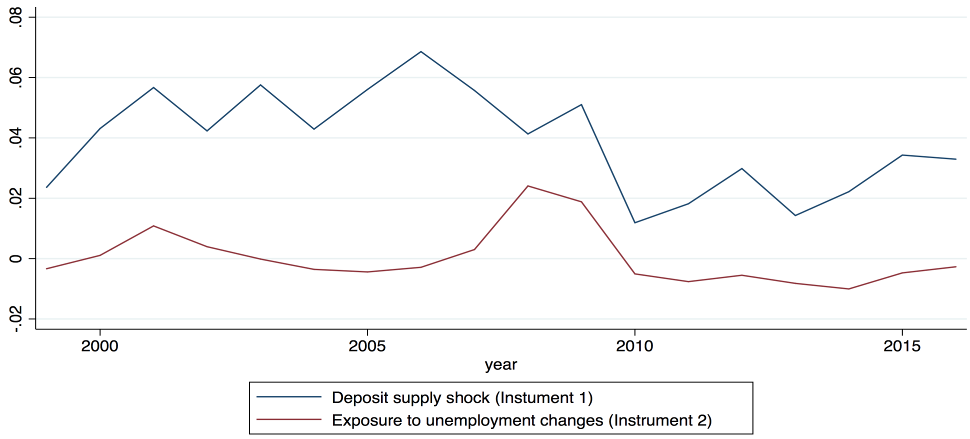
Direct local unemployment exposure: In building the alternative instrument we directly estimate the exposure of each bank to changes in local unemployment rates, using, as before, the deposits a bank holds as weights. For bank b, operating in areas $i = 1..I$, this exposure, $\Delta exp_{b,t}$, is given by:

$$\Delta exp_{b,t} = \frac{\sum_{i=1}^{I} dep_{b,i,t} \Delta unemp_{i,t}}{\sum_{i=1}^{I} dep_{b,i,t}} \tag{3.3}$$

where, as before, $dep_{b,i,t}$ denotes the deposits bank j holds in geographical area i in period t, and $\Delta unemp_{i,t}$ denotes a change in unemployment rate in geographical area i in period t. We use $\Delta exp_{b,t}$ as the second instrument in the final estimation of the effect of leverage on risk taking.

Figure 3.3 plots the mean of the bank level deposit supply shock, $\hat{dep}_{j,t}$ across time, and the mean bank level exposure to unemployment, $\Delta exp_{b,t}$. The figure reveals a sharp drop in deposit supply coinciding with a spike in unemployment across the US.

Figure 3.3: Mean bank exposure to local unemployment changes and deposit supply shocks



The two instruments exploit the same variation of changes in unemployment at the local level. They are not numerically equivalent due to different specifications of the fixed effects. It is also worth noting that the second instrument does not require any estimation since it is only built through aggregation of local areas changes at the bank level. This is relevant because one may be concerned that our first instrument may suffer from generated regressor problems. As we will show later we obtain fairly similar results with the two instruments.

3.3.4 A Measure of Risk Taking

Remaining endogeneity

The procedure explained above describes constructing a measure of an exogenous change in deposits and a measure of an exposure of banks to changes in local unemployment rates. Both these measures are exogenous to risk taking, but not exogenous to a measure of riskiness of the portfolio. To exemplify the issue, consider a shock to deposits in a certain area as estimated in the previous section. Such a shock is likely to impact the income of depositors. It cannot, however, be excluded to have impacted also the borrowers, private or corporate, in an area, which may or may not have borrowed from the banks operating in that areas. Any measure of risk, which is based on the performance of the existing portfolio might be subject to this sort of residual endogeneity.

New issuances of loans are not subject to this endogeneity concern since new issuances can only be affected by the existing pool of potential loans. A geographical area shock can affect the existing local pool of borrowers, while it does not affect the entire pool of potential borrowers. New issuance of a loan is a choice for a bank and the riskiness of new issuances proxies risk taking behaviour.

Creating a measure of risk taking

To this end we construct a measure of risk taking based on the issuances of new mortgage loans based on the Home Mortgage Disclosure Act (HMDA) dataset, administered by the Federal Financial Institutions Examination Council (FFIEC). It is a yearly dataset on the population of mortgage applications to banks and other mortgage lenders with detailed information on the borrower and loan characteristics. We take the riskiness of new mortgage lending as representative of risk taking on the entire portfolio.

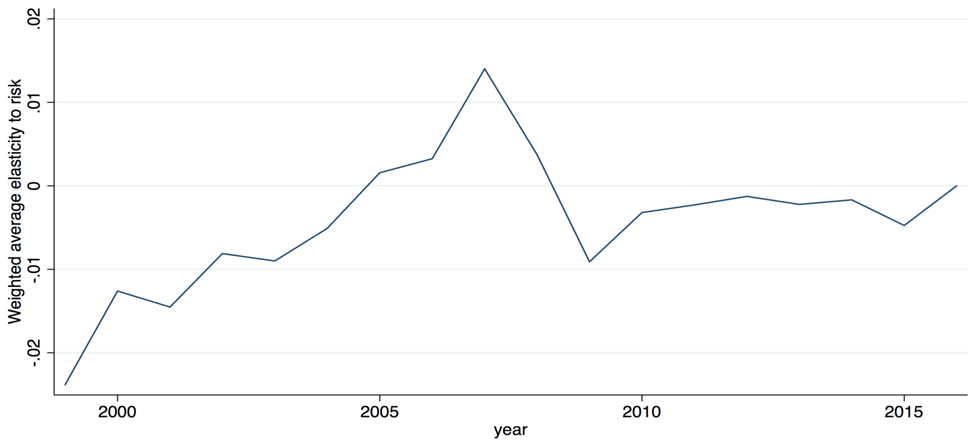
To construct a measure of risk taking behaviour by banks, we estimate the responsiveness of loan issuance of each bank in each year to riskiness of the borrower and the loan. As a measure of riskiness of the loan and the borrower we use the Loan-to-Income (LtI) ratio computed from the HMDA dataset for every loan application. This follows loosely DellAriccia et al. (2012), where LtI is used directly as a measure of risk in their analysis of lending standards. This methodology implies the following model[2,3].

$$Origin_{t,b,j} = \gamma_t^0 + \gamma_{t,b}^1 LtI_{t,b,j} + \epsilon_{t,b,j} \tag{3.4}$$

where $Origin_{t,b,j}$ denotes a binary loan origination variable which takes the value $Origin_{t,b,j} = 1$ if the application in period t to a bank b by a borrower j is accepted and loan is originated, and takes the value $Origin_{t,b,j} = 0$ if the application is rejected and the loan is not originated. γ_t^0 captures the effect of the macroeconomic situation in period t for all banks, such as market conditions and regulation, which may cause banks to have differing appetites for risk over time. Finally, for every bank b in every period t we also obtain an estimate of the risk responsiveness $\hat{\gamma}_{t,b}^1$ based on Loan-to-Income of all applicants j, which serves as a measure of risk taking behavior by banks.

Figure 3.4 plots the risk measure for the banks included in the analysis over the years. The distribution has a mean of $-.0039622$.

Figure 3.4: Responsiveness of origination to risk



[2] In order for γ_t^0 to capture the macroeconomic conditions affecting the origination choices, we estimate the model for all banks reporting to the HMDA dataset but only use the $\gamma_{t,i}^1$ for banks included in the final regressions. This implies including all the loan applications in the HMDA reporting in the estimations. The number varies between 17 million applications and 40 million application which constrains us to estimating the model as a linear probability model.

[3] This measure is joint work with Lpez-Quiles and Petricek (2018)

3.3.5 IV Estimation of the Effect of Leverage on Risk Taking

In estimating the effect of leverage on risk taking we use the two instruments, explained in detail above. As argued before the instrumented deposits and the direct measure of exposure to local unemployment shocks are exogenous to risk taking. The instruments allow us to estimate two effects: (i.) the effect of the two instruments on leverage, and (ii.)the effect of leverage on risk taking.

More specifically we run the following two specifications:

$$\hat{\gamma}_{b,t} = \beta_0 + \beta_1 lev_{b,t} + \eta_b + \delta_t + \epsilon_{bt} \tag{3.5}$$

Where lev_{bt} is the endogenous variable, measured as the leverage ratio, η_b are bank fixed effects and δ_t are time fixed effects which are included to control for different baseline risk preferences of banks as well as a regulatory environment that may change over time. This equation is then estimated by IV, where the endogenous variable lev_{bt} is instrumented with one of the two instruments: either $\Delta \hat{dep}_{bt}$ or Δexp_{bt}, depending on the model. It is also estimated by OLS, in order to compare the coefficients of interest.

The results of the estimations are presented and discussed in the next section.

3.4 Results

Table 3.4 presents the results of the estimations for both instruments. Column (1) shows the biased OLS estimate of regressing the risk taking measure on leverage ratio. Columns (2) and (3) present the results using the deposit supply shocks as an instrument, while columns (4) and (5) present the results employing the exposure to changes in unemployment. Columns (2) and (4) show the first stage of the two IV regressions, while Columns (3) and (5) show the second stage. Both sets of results are consistent in terms of sign, so we will focus on the latter in explaining them. All standard errors are clustered at the bank level.

First we find that the naive OLS approach severely underestimates the effect of leverage on risk taking behaviour, the result being close to zero and statistically insignificant. This bias may go some way towards explaining findings in the previous literature that leverage leads to less risk taking, or has no effect (see for example Altunbas et al. (2007) and Jacques and Nigro (1997)).

All our results provide evidence for a positive effect of leverage on banks' risk taking due to limited liabilities[4].

Table 3.4: Risk Measure

	(1) Risk	(2) Leverage	(3) Risk	(4) Leverage	(5) Risk
Leverage	-0.0000902		-0.0118**		-0.00564**
	(0.0000801)		(0.00557)		(0.00240)
IV Δ Unemployment		5.462****			
		(1.287)			
IV Δ Deposits				-25.12****	
				(2.817)	
Non-Current Loans		-0.163****	-0.00255***	-0.164****	-0.00155****
		(0.00404)	(0.000913)	(0.00403)	(0.000400)
Constant	0.0461****	9.776****	0.160***	10.35****	0.101****
	(0.000977)	(0.0290)	(0.0544)	(0.0713)	(0.0234)
Time FE	Yes	Yes	Yes	Yes	Yes
N	65811	65660	65660	65660	65660
R^2	0.118	0.056		0.057	
$F(instr.excl.)$		17.72		79.00	

Standard errors in parentheses
** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

We also investigate the possibility of nonlinear effects of leverage on risk taking by running our IV regression through 2SLS and interacting the predicted endogenous variable with dummy

[4]Since leverage is measured as leverage ratio (core capital over total average assets), a negative sign implies that an increase in leverage (a decrease in leverage ratio) increases risk taking by banks

variables denoting different deciles of the leverage distribution. We find no significant pattern in the estimation and that most coefficient on these dummies are not statistically significant. The takeaway of this analysis is that the effect of leverage in our data does not vary significantly along the distribution of leverage.

Quantitatively our results state that a one point increase in the leverage ratio[5] generates an .0118 decrease in our risk taking measure. The mean of the risk measure in our data is .038, which implies that a one point increase in the leverage ratio produces a 31% decrease risk taking when compared to the average.

For a specific example, assume that two banks are identical except for their leverage ratios, which differ by one percentage point. Assume that they receive the same application for a mortgage loan with average Loan-to-Income. Our estimates suggest that this application has an expected probability of being originated in the bank with the higher leverage (i.e. lower leverage ratio) of $\gamma_0 + \gamma_i LTI$ whereas the expected probability of origination for the less leveraged bank is $\gamma_0 + (\gamma_i - .0118)LTI$. Evaluating these probabilities at the average of our estimates and at the average loan-to-income we obtain that the more leveraged bank has a 3.1% higher probability of originating the loan.

Note also that this wedge between the probabilities of acceptance increases with the loan-to-income ratio. Meaning that the higher the loan-to-income of the applicant the larger the difference in expected acceptance probabilities between the more and less leveraged bank.

Our results have policy relevant implications in terms of the aggregate level of risk in the banking system. The estimations show that more levered banks are more likely to take on riskier projects due to limited liability incentives which implies that curbing leverage has the added

[5]Leverage ratio is defined as core capital divided by assets. Thus an *increase* in the leverage ratio is associated with a *decrease* in bank leverage.

benefit of reducing banks' risk taking, thereby producing a more resilient banking system.

3.5 Robustness

In our main analysis we use as the outcome a risk measure based directly on banks' lending decisions. The novelty of this measure lies in the fact that it captures the responsiveness of origination behaviour to a proxy of risk, namely the loan to income ratio. One could however be concerned that LtI being our only proxy for risk in our estimation the γ parameter may be capturing correlations of LtI with unobserved loan characteristics. In order to address this potential problem redo the above analysis with a second measure for bank risk taking.

We resort to a further data source (Consumer Financial Protection Bureau) to obtain mortgage delinquency rates at the county level. This data covers 470 counties for the period 2009-2015. The sample is representative and covers approximately 85% of all mortgage lending in the HMDA data. We use the 90 days delinquency rate for closed-end, 1-4 family residential mortgages. The full list of variables and sources is in Table 3.1 in the Appendix.

The idea of this procedure is that we can observe the delinquency rate at the county level and we obtain a number of predictors with the same level of disaggregation. Once we have established a prediction model for our outcome we can predict the delinquency rate at the loan level by using more disaggregated data from the HMDA.

From the county data we build a 3 years ahead average delinquency rate at the county level, which will be used as the outcome of our prediction model. We obtain a number of explanatory variables from the American Community Survey, the Internal Revenue Service and the Bureau of Labour Statistics.

Since the goal is to be able to predict the delinquency rate at the loan level we run the model including two sets of explanatory variables: i) regressors that we observe both at the county and at the loan level; 2) regressors we only observe at the county level.

In order to find the best prediction model we estimate approximately 1000 models including combinations of the predictors. For each model we split our sample keeping 75% of the counties to estimate the model and 25% as out of sample fit and we estimate it 10 times. We evaluate the models by the out of sample Mean Squared Prediction Error and pick the one with the lowest MSPE. All models are estimated with and without time fixed effects.

More formally we estimate

$$DR_{i,t} = \beta_0 + \beta_1 X^1_{i,t} + \beta_2 X^2_{i,t} + \beta_3 X^1_{i,j,t} X^2_{i,t} + \delta_t + \epsilon_{i,t} \tag{3.6}$$

Where i denotes the county and X^1 is a set of predictors that is aggregated at the county level from the loan level data, whereas X^2 is purely county level data. We include the interaction term between the loan aggregated and the county level regressors to account for the covariances between the two and eliminate potential bias when we move to the loan level prediction.

Once we have estimated this model we use the coefficient and produce the following loan level prediction

$$\hat{D}R_{i,j,t} = \hat{\beta}_0 + \hat{\beta}_1 X^1_{i,j,t} + \hat{\beta}_2 X^2_{i,t} + \hat{\beta}_3 X^1_{i,j,t} X^2_{i,t} + \hat{\delta}_t, \tag{3.7}$$

where j denotes a loan. The specification of the prediction model is displayed in Table 3.2 in the Appendix.

This procedure gives us an expected probability of default for all the loans in our sample.

In order to use this as an outcome in our final stage we take the median forecasted probability of default for each bank in every year. We estimate this model using the IV discussed in the methodology section. The results of this procedure are displayed in Table 3.5.

Table 3.5: Regression table

	(1)	(2)	(3)	(4)	(5)
	Median PD	Leverage	Median PD	Leverage	Median PD
Leverage	-0.0208**		-1.174***		-1.025***
	(0.00861)		(0.434)		(0.374)
IV Δ Unemployment		6.099****			
		(1.541)			
IV Δ Deposits				-15.22****	
				(3.510)	
Non-Current Loans		-0.125****	-0.122**	-0.126****	-0.104**
		(0.00527)	(0.0550)	(0.00527)	(0.0475)
Constant	2.592****	10.05****	14.26***	10.95****	12.74****
	(0.0884)	(0.0378)	(4.416)	(0.181)	(3.802)
Time FE	Yes	Yes	Yes	Yes	Yes
N	25050	25040	25040	25040	25040
R^2	0.253	0.092		0.092	
$F(instr.excl.)$		37.05		44.04	

Standard errors in parentheses

** $p < 0.05$, *** $p < 0.01$, **** $p < 0.001$

Using this measure for risk taking yields results that are in line with the ones discussed in the previous section. OLS underestimates the effect of leverage on risk taking for this measure as well. Using the IV procedure yields a negative and significant effect of leverage ratio on risk taking. This once again translates into a positive relationship between bank leverage and risk taking. Our results do not differ between the two IV procedures. Using this risk measure, we find that an exogenous one point increase in the leverage ratio of a bank will lead to roughly a one percentage point decrease in the median probability of default of loans issued by that bank.

3.6 Conclusions

This paper addresses the question of the causal effect of changes in leverage on banks' risk taking behaviour. We do so by constructing two instruments to overcome the endogeneity problems resulting from the potential simultaneity and reverse causality between risk decisions and the deposit market conditions. We instrument exogenous changes in leverage by building two instruments: i) one based on the geographical area unemployment changes; ii) one based on the geographical area deposit supply changes. In both cases we aggregate them using the local deposit share of banks.

We then build a new measure of risk taking behaviour based on the responsiveness of origination decisions to a measure of risk of loan applications (loan-to-income). We compute this measure at the bank/year level and use it as our outcome.

Our empirical analysis suggests that exogenous increases in leverage incentivises banks to take on more risk, i.e. to originate loans with riskier characteristics. We also employ a second measure of risk, the median predicted probability of default of originated loans, and find that higher levererage leads to a higher probability of default of issued loans. These results are consistent with a limited liability and moral hazard story put forth by some of the theoretical literature. They are novel in the empirical literature on leverage and risk taking, as previous work has found no relation or a negative relationship between leverage and risk.

These results have relevant policy implications in that they suggest that any measure that would reduce banks' leverage would also decrease incentives to invest in risky assets, thereby considerably reducing systemic risk.

Appendix

Variable list

Table 3.1: Data for Prediction Model

VARIABLE	SOURCE	AVAILABILITY
median lti of originated loans for purchases or refinancing	HMDA	loan
median income of borrowers	HMDA	loan level
median borrowed amount	HMDA	loan level
median housing cost	ACS	county level
median housing cost to median income	ACS	county level
median income	ACS	county
share households with a second mortgage	ACS	county level
median value of a housing unit	ACS	county level
average monthly mortgage payment	IRS	ZIP code level
unemployment rate	BLS	county level

Data sources are:

HMDA - home mortgage disclosure act data

ACS - American Community Survey data

IRS - Internal Revenues Service

BLS - Bureau of Labor Statistics

Delinquency rate prediction model

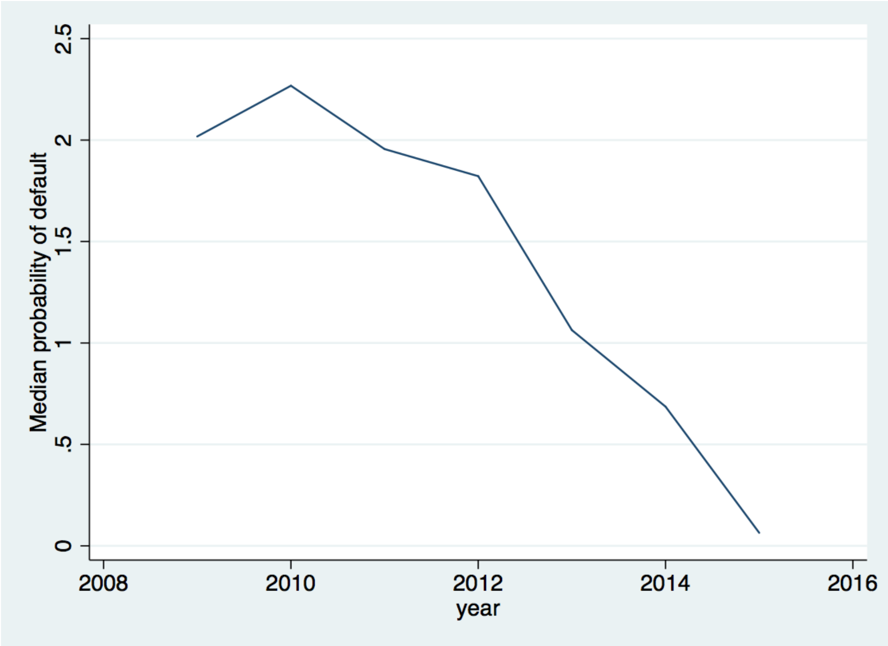
Table 3.2: Prediction Model

	(1)
	dr3c
Median lti of origin. loans	-0.482
	(0.686)
Median income of borrowers	-0.0621***
	(0.0145)
Median household income	0.0000254
	(0.0000144)
Med. housing cost to household inc.	315.1***
	(71.32)
% of households with 2nd mortg.	0.0417
	(0.0343)
Average mortgage payments	0.000217*
	(0.000110)
Unemployment	0.0000428***
	(0.00000523)
Interaction terms	YES
State fixed effects	YES
Constant	-1.223
	(1.485)
N	1606
R^2	0.793
Adj. R^2	0.783

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 3.5: Median predicted probability of default; (average across banks in the final estimation sample)



Bibliography

Acosta Smith, J., Grill, M., and Lang, J. H. (2017). The Leverage Ratio, Risk-Taking and Bank Stability. SSRN Scholarly Paper ID 2990215, Social Science Research Network, Rochester, NY.

Altunbas, Y., Carbo, S., Gardener, E. P., and Molyneux, P. (2007). Examining the Relationships between Capital, Risk and Efficiency in European Banking. *European Financial Management*, 13(1):49–70.

Arias, E. (2009). United States Life Tables, 2009. page 64.

Ashraf, B. N., Arshad, S., and Hu, Y. (2016). Capital Regulation and Bank Risk-Taking Behavior: Evidence from Pakistan. *International Journal of Financial Studies*, 4(3):16.

Auerbach, A. J. (1991). Retrospective Capital Gains Taxation. *The American Economic Review*, 81(1):167–178.

Auerbach, A. J. and Bradford, D. F. (2004). Generalized cash-flow taxation. *Journal of Public Economics*, 88(5):957–980.

Bagchi, S. (2015). Labor supply and the optimality of Social Security. *Journal of Economic Dynamics and Control*, 58:167–185.

Bartik, T. J. (1993). Who Benefits from Local Job Growth: Migrants or the Original Residents? *Regional Studies*, 27(4):297–311.

Binmore, K. G. (1982). *Mathematical Analysis: a straightforward approach.* Cambridge University Press.

Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637–654.

Bradford, D. F. (1994). Fixing realization accounting: Symmetry, consistency and correctness in the taxation of financial instruments. *Tax L. Rev.*, 50:731.

Brendler, P. (2016). Lifetime earnings inequality and income redistribution through social security. *Working Paper*.

Brewer, M. and Shaw, J. (2018). How Taxes and Welfare Benefits Affect Work Incentives: A Life-Cycle Perspective. *Fiscal Studies*, 39(1):5–38.

Breyer, F. and Straub, M. (1993). Welfare effects of unfunded pension systems when labor supply is endogenous. *Journal of Public Economics*, 50(1):77–91.

Browning, E. K. (1985). The Marginal Social Security Tax on Labor. *Public Finance Quarterly*, 13(3):227–251.

Burkhauser, R. V. and Turner, J. A. (1985). Is the Social Security Payroll Tax a Tax? *Public Finance Quarterly*, 13(3):253–267.

Copland, D. B. (1924). The Australian Income Tax. *The Quarterly Journal of Economics*, 39(1):70–95.

Cushing, M. J. (2005). Net Marginal Social Security Tax Rates over the Life Cycle. *National Tax Journal*, 58(2):227–245.

DellAriccia, G., Igan, D., and Laeven, L. U. (2012). Credit booms and lending standards: Evidence from the subprime mortgage market. *Journal of Money, Credit and Banking*, 44(2-3):367–384.

Diamond, D. W. and Rajan, R. G. (2001). Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking. *Journal of Political Economy*, 109(2):287–327.

Farhi, E. and Werning, I. (2013). Insurance and Taxation over the Life Cycle. *The Review of Economic Studies*, 80(2):596–635.

Fehr, H. and Kindermann, F. (2018). *Introduction to Computational Economics Using Fortran*. Oxford University Press, New York, NY.

Feldstein, M. and Samwick, A. (1992). Social Security Rules and Marginal Tax Rates. *National Tax Journal*, 45(1):1–22.

Feldstein, M. and Samwick, A. (1998). The Transition Path in Privatizing Social Security. In *Privatizing Social Security*, pages 215–264. The University of Chicago Press, Chicago.

Goda, G. S., Shoven, J. B., and Slavov, S. N. (2011). Implicit Taxes on Work from Social Security and Medicare. *Tax Policy and the Economy*, 25(1):69–88.

Goda, G. S., Slavov, S. N., and Shoven, J. B. (2007). Removing the Disincentives in Social Security for Long Careers. SSRN Scholarly Paper ID 986960, Social Science Research Network, Rochester, NY.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2017). Bartik instruments: What, when, why and how. *Unpublished Working Paper*.

Golosov, M., Troshkin, M., and Tsyvinski, A. (2016). Redistribution and Social Insurance. *American Economic Review*, 106(2):359–386.

Gordon, R. H. (1982). Social Security and Labor Supply Incentives. *Contemporary Economic Policy*, 1(3):16–22.

Heathcote, J., Storesletten, K., and Violante, G. L. (2014). Optimal Tax Progressivity: An Analytical Framework. Working Paper 19899, National Bureau of Economic Research.

Homburg, S. (1990). The Efficiency of Unfunded Pension Schemes. *Journal of Institutional and Theoretical Economics (JITE) / Zeitschrift fr die gesamte Staatswissenschaft*, 146(4):640–647.

Huggett, M. (1996). Wealth distribution in life-cycle economies. *Journal of Monetary Economics*, 38(3):469–494.

Huggett, M. and Parra, J. (2010). How Well Does the U.S. Social Insurance System Provide Social Insurance? *Journal of Political Economy*, 118(1):76–112.

Huggett, M. and Ventura, G. (1999). On the Distributional Effects of Social Security Reform. *Review of Economic Dynamics*, 2(3):498–531.

Jacques, K. and Nigro, P. (1997). Risk-based capital, portfolio risk, and bank capital: A simultaneous equations approach. *Journal of Economics and Business*, 49(6):533–547.

Jensen, M. C. and Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4):305–360.

Koehn, M. and Santomero, A. M. (1980). Regulation of Bank Capital and Portfolio Risk. *The Journal of Finance*, 35(5):1235–1244.

Koudstaal, M. and van Wijnbergen, S. (2012). On Risk, Leverage and Banks: Do highly Leveraged Banks take on Excessive Risk? Tinbergen Institute Discussion Paper 12-022/2/DSF31, Tinbergen Institute.

Laeven, L. and Levine, R. (2009). Bank governance, regulation and risk taking. *Journal of Financial Economics*, 93(2):259–275.

Liebman, J. B. (2002). Should taxes be based on lifetime income? Vickrey taxation revisited. In *Harvard University Working Paper*.

Lpez-Quiles, C. and Petricek, M. (2018). Deposit insurance and bank risk-taking. *ADEMU working paper series*, (2018/101).

Michau, J.-B. (2014). Optimal redistribution: A life-cycle perspective. *Journal of Public Economics*, 111:1–16.

Nardi, M. D., Fella, G., and Pardo, G. P. (2016). The Implications of Richer Earnings Dynamics for Consumption and Wealth. Working Paper 21917, National Bureau of Economic Research.

OECD (2018). Tax wedge decomposition.

Ohlrogge, M. (2017). Bank Capital and Risk Taking: A Loan Level Analysis. *SSRN Electronic Journal*.

Prescott, E. C. (2004). Why do Americans Work so Much More than Europeans? Working Paper 10316, National Bureau of Economic Research.

Schmalbeck, R. (1984). Income Averaging After Twenty Years: A Failed Experiment in Horizontal Equity. *Duke LJ*, page 509.

Shrieves, R. E. and Dahl, D. (1992). The relationship between risk and capital in commercial banks. *Journal of Banking & Finance*, 16(2):439–457.

SSA (2017). Annual Statistical Supplement, 2016.

Tauchen, G. (1986). Finite state markov-chain approximations to univariate and vector autoregressions. *Economics Letters*, 20(2):177–181.

Vickrey, W. (1939). Averaging of Income for Income-Tax Purposes. *Journal of Political Economy*, 47(3):379–397.

Vickrey, W. (1972). Cumulative averaging after thirty years. In *Modern fiscal issues*. Toronto Press.

Vickrey, W. S. (1947). *Agenda for progressive taxation*. Ronald Press Company.

World Bank (2018). Real interest rate (%) | Data. `https://data.worldbank.org/indicator/FR.INR.RINR?view=chart`. Accessed: 2018-09-04.