# Income Inequality and Income Risk

Inaugural-Dissertation zur Erlangung des akademischen Grades
eines Doktors der Wirtschaftswissenschaft

Fachbereich Wirtschaftswissenschaft
Freie Universität Berlin

vorgelegt von

**Johannes König, M.Sc.**

Berlin, February 6, 2019

**Erstgutachter**
Prof. Dr. Dr. Giacomo Corneo, Freie Universität Berlin

**Zweitgutachter**
Prof. Dr. Carsten Schröder, DIW Berlin & Freie Universität Berlin

Datum der Disputation: 25.01.2019

## Collaboration with Coauthors and Publications

**Kapitel 1: Distributional effects of subsidizing retirement savings accounts: evidence from Germany**

- Kapitel 1 basiert auf einem im FinanzArchiv erscheinenden Artikel der zu gleichen Teilen mit Giacomo Corneo und Carsten Schröder verfasst wurde.
- Corneo, König und Schröder (2018)

**Kapitel 2: Inequality-minimization with a given public budget**

- Kapitel 2 basiert auf einem im Journal of Economic Inequality veröffentlichenten Artikel der zu gleichen Teilen mit Carsten Schröder verfasst wurde.
- König und Schröder (2018)

**Kapitel 3: Hours risk, wage risk, and life-cycle labor supply**

- Kapitel 3 basiert auf einem unveröffentlichten Artikel der zu gleichen Teilen mit Robin Jessen verfasst wurde.
- Jessen und König (2018)

**Kapitel 4: Earnings risk and tax policy**

- Kapitel 4 basiert auf einem unveröffentlichten Artikel der in Alleinarbeit angefertigt wurde.
- König (2018)

# Acknowledgements

Foremost, I would like to offer my gratitude to my two advisors Prof. Giacomo Corneo and Prof. Carsten Schröder. Prof. Corneo has contributed immensely to the development of my research interests. Prof. Corneo hired me for a position at his chair to work on a research project on the distributional impact of the Riester pension scheme. Our collaboration on the project planted the seeds for my research into structural life-cycle models of labor supply; the dominant topic in my recent research. Prof. Corneo is a vast fount of knowledge and has been able to direct my vague interests toward the newest relevant research on my topics of interest. He has been able to offer wonderful feedback on my research ideas and papers. Further, his seminar series on papers in public economics has been an extremely inspiring and challenging experience that I loved taking part in and that has spurred growth in both breadth and depth of my understanding of the field. He has also contributed to my development as a researcher by supporting my recent academic visit to University College London.

I have also immensely benefitted from the mentorship of Prof. Schröder. He too was a collaborator on the Riester project and has contributed to a immense extent to both my knowledge of theoretical and empirical economics. Most prominently, he taught me the subject of inequality measurement, to which he and I have now contributed in a joint paper. Prof. Schröder is an excellent teacher because of his approachable nature and the simple and concise way he can communicate complex subject matter. His acumen to reduce an issue to its essence makes him an excellent and highly appreciated critic of my work. Prof. Schröder has also helped my development as a researcher by hiring me at DIW Berlin to work on a project analyzing the German minimum wage reform and has fostered my international outreach by supporting my visit to UCL.

Further, I would like to thank Prof. Frank Fossen, who hired me as a research assistant while I was still a Master's student. He sparked my interest in empirical economics, especially microeconometrics, and coauthored a paper with me on the subject of health care costs' influence on the entry into self-employment, which is now published in Small Business Economics. Without his constant encouragement and patience I would likely never have pursued a dissertation. For this and the many enlightening discussion and feedback he has given on my research I am deeply indebted to him.

I also want to thank Prof. Viktor Steiner for supporting Robin Jessen's and my research and allowing me to participate in his seminar throughout the years and supporting my outreach to UCL.

Prof. Richard Blundell also deserves my thanks for inviting me to the economics department of UCL and enabling me to share my research with so many of the community at UCL and the Institute for Fiscal Studies. Prof. Blundell has also been very encouraging of my paper in sole authorship and helped it develop in its early stages for which I am especially thankful.

I would like to thank my coauthor and friend Robin Jessen for the inspiring and growth-inducing collaboration on our paper concerning wage and hours risk over the life-cycle. Robin has been an excellent partner for mutual intellectual development and is the first person I can always call on to flesh out new ideas. Furthermore, he has taught me how to appreciate the finer things in life.

Thanks also to Sebastian Schmitz for being a wonderful friend and colleague at Freie Universität Berlin and for the many discussions we had about each other's research and for the many positive changes he has caused in my life.

I want to thank Maria Metzing for being my office mate and trusted colleague at the SOEP. Maria is a joy to be around, makes working fun and is always open for discussion.

I'm thankful for the stimulating intellectual environments at both Freie Universität Berlin and DIW Berlin. In particular, I thank my fellow PhDs and colleagues at Freie Universität Berlin Pablo Anaya, Charlotte Bartels, Timm Bönke, Flora Budianto, Benjamin Fischer, Anja Gaentzsch, Dominik Hügle, Holger Lüthen, Theresa Neef, Frank Neher, Guido Neidhöfer, Davud Rostam-Afschar, Max Stockhausen, Marten von Werder, and Miriam Wetter. For his assistance and cooperation I want to thank Max Wenzel especially. Thanks also to Nelli Anan for being the backbone and soul of the public economics chair.

Same goes for my fellow PhDs and colleagues at the DIW: Sandra Bohmann, Patrick Burauel, Philipp Eisnecker, Alexandra Fedorets, Daniel Graeber, Christoph Halbmeier, Hannes Kröger, Felicitas Schikora, and Cortnie Shupe. I want to thank my coauthor Levent Neyse for being such an open minded, friendly and insightful collaborator and for being a good friend. I also want to thank Jian Li and Christos Koulovatianos at the University of Luxembourg for great discussions.

Many thanks to the people at UCL,IFS and other institutions in London, who have engaged with my work during my stay abroad. In particular, I want to thank Stuart Adam, Jonathan Cribb, Monica Costa-Dias, Frank Cowell, Giulio Fella, Michael Graber, Robert Joyce, Agnes Norris Keiller, Francois Maniquet, Rory McGee, Francesca Parodi, Ian Preston, Itay Saporta-Eksten, Josef Sigurdsson, Alexandros Theloudis, and Guglielmo Weber. I especially want to thank Jack Light for being a wonderful host at the department and a great partner for discussions on my research.

I want to thank my dear friends André, Marlen, Sky, Mathias, Leo and Anne for sticking it out with me, lending an ear when I needed it the most and always motivating me to participate in new and exciting activities.

Finally, I want to express my unending gratitude to my parents Conrad and Maryte. They have supported my academic career from the beginning and in every way possible. Without them I would never have achieved any of this.

# Contents

# 1 Distributional effects of subsidizing retirement savings accounts: evidence from Germany[1]

## 1.1 Introduction

Since the mid-1990s, governments have developed programs that provide financial aid to encourage private saving for retirement purposes, especially among low- and middle income households. Germany introduced its own program, called *Riester scheme*, in 2002. It promotes certified financial products for retirement saving by means of generous subsidies and tax deductions. Basically, all compulsorily insured employees in Germany, including public servants, are eligible for support under the Riester scheme.

In contrast with its goal of inducing people to save more, most empirical evaluations of the Riester scheme suggest that it hardly generates any effect on savings.[1] That is, the Riester scheme mainly displaces private savings from unsubsidized to subsidized assets, so that most of the subsidies amount to windfall gains for their beneficiaries.[2] In such a context, where real behavioral responses are negligible, it is imperative to investigate how those windfall gains are distributed. If low-income households were the main beneficiaries, the Riester scheme would be likely to contribute to reduce old-age poverty in the future. Assessing the distributional impact

[1]This is a post-peer-review, pre-copyedit version of an article published in Finanzarchiv/Public Finance Analysis. The final authenticated version is available online at: http://dx.doi.org/10.1628/fa-2018-0017

[1]See Coppola and Reil-Held (2009), Corneo et al. (2009), Corneo et al. (2010) and Pfarr and Schneider (2011) for analyses based on the German SOEP and the SAVE dataset. Börsch-Supan et al. (2008) find ambiguous results, while Börsch-Supan et al. (2012) discuss the validity of survey data on savings. Strong displacement effects are often found for similar programs in other countries. Engen et al. (1996) provide a plethora of arguments why the supposed stimulation effects of 401(k)s and IRAs in the U.S. are overstated or non-existent. A more recent example is Chetty et al. (2014) who find in the case of Denmark that each 1$ of government expenditure on saving subsidies increases savings by 1 cent. Engelhardt and Kumar (2011) document an incomplete crowding out in terms of types of wealth. Carbonnier et al. (2014), using French micro tax-data, show that savings demand is boosted for richer savers, but the effect is weak to nonexistent for poorer individuals.

[2]To the extent that subsidies are shifted to the supply side, insurers capture a fraction of them. In this paper, we abstract from shifting.

of the Riester scheme is thus an essential ingredient of a comprehensive evaluation of that policy.

At first glance, the Riester scheme is equality-enhancing. While the German PAYG system is of the Bismarckian variety and relies on the equivalence principle, the provisions of the Riester scheme entail distinctive elements of progressivity: a basic allowance that is equal for everybody and generous child allowances that favor multi-member households. In spite of those provisions, the overall distributional impact of the Riester scheme is *a priori* unclear. Participation is voluntary and persons are entitled to the government's financial aid only if they invest a certain minimum amount in so-called Riester contracts, that amount being determined as a fraction of a person's income liable to social security contributions. This boils down to requiring a minimum saving propensity in order to be able to invest the required amount into a Riester contract. If the saving propensity increases with income and poor households do not save enough to meet the participation requirement, self-selection in the program will generate a regressive effect. Furthermore, high income households are more likely to benefit from a special tax deduction provided by the Riester scheme. Whether the progressive effect from the subsidy provisions outweighs the regressive effect from self-selection and tax deductions is an empirical issue that we address in the current paper.

Determining the distributional impact of the Riester scheme requires a dataset that is representative of the German population and contains the necessary information to compute the total subsidy received by each household. Such a dataset, the Panel of Household Finance (PHF) was released by the German central bank in 2012. It is constructed along similar lines as the Survey of Consumer Finances (SCF) in the U.S. Its key advantage over alternative datasets like the German Socio-economic Panel (SOEP) is that the PHF records the amounts respondents contribute to their Riester contract. We are the first to exploit this dataset to investigate the Riester reform. We use the contribution information to estimate the monetary benefit received by each household and determine the effects of the Riester scheme on the current distribution of yearly incomes. A microsimulation model predicts for each household whether it receives a direct subsidy or a tax deduction. We carefully distinguish between tax unit - which is used by the public administration to set the type and level of the subsidy - and household - which is the unit from which we derive the equivalent incomes of the population.[3]

Our analysis pertains to first-round effects only, i.e. it merely captures the mechanical effect on current incomes from reaping the monetary gain delivered

[3]By observing the composition of the tax unit, i.e. determining how many Riester participants and relevant dependents are present in the tax unit, and by accounting for the individual saving efforts of the participants, we can determine how much direct allowance the tax unit receives. By determining the income tax liability of the tax unit we can perform the higher-yield test to ascertain whether the household receives a tax deduction or not. We then aggregate the Riester subsidies of the tax units at the household level for the distributional analysis.

by the subsidy. Our main results are as follows. Our main results are as follows. First, we find that about 38% of the subsidies accrue to the top quintile of the income distribution, while only about 7% accrues to the lowest quintile. Second, the progressive schedule of the subsidy is almost exactly offset by the regressive effect from self-selection into the program. As a result, measures of overall income inequality like the Gini coefficient are hardly affected by the Riester scheme. Its effect on measures of poverty is slightly worse, in particular on the share of the population below the poverty line: the Riester scheme increases that share by nearly one percentage point.

The issue analyzed in this paper is relevant for a number of countries beyond Germany – countries with similar programs where participation is voluntary and behavioral responses are small. So far, the distributional effects of these programs have received scant attention from the literature. Important exceptions are 401(k)s – a type of defined contribution plan – and individual retirement accounts (IRA) in the United States. Burman et al. (2004) examine the distribution of tax benefits from defined contribution plans and IRAs with data from the SCF and the SIPP (Survey of Income and Program Participation). When considering both defined contribution plans and IRAs, they find that 70% of the total tax benefit accrues to the top quintile and almost none to the lowest quintile. These results are robust to excluding IRAs. The pattern of self-selection into the programs is close to the one we uncover for the Riester scheme: while only 3% of households in the bottom quintile participate, 41% of households in the top quintile do so. Joulfaian and Richardson (2001) investigate the demographics of the population participating in defined contribution plans, IRAs and other subsidized savings vehicles using income tax data. They also briefly report on the distribution of the tax benefits along the income distribution of wage-earning households. They note that the lower half of the distribution receives less than 10% of the overall expenditure, while almost 55% of the expenses accrue to the top 10%. When restricting for eligibility for any of the subsidized savings programs, the bottom 50% receive 20% of the overall benefit and the top 10% receive 33%. Chernozhukov and Hansen (2004) use SIPP data to examine the impact of 401(k)-plans on wealth. Their findings indicate that the effect of 401(k)-participation is quite heterogeneous along the distribution; the largest positive effect is experienced by those in the upper tail of the wealth distribution. Even and Macpherson (2007) evaluate the impact of defined contribution plans on the distribution of pension wealth with SCF data. They suggest that the switch from defined benefit to defined contribution plans will widen the pension wealth gap between low and high earners.

We discuss our main empirical findings in section 4.7, after having presented the institutional details of the Riester scheme in section 1.2 and the data we employ in section 1.3. In section 1.5 we further scrutinize the regressive effect from participation by searching for its main determinants. In line with what Chernozhukov and

Hansen (2004) find for the US, we find that net household wealth has a distinctive positive effect on the probability to participate in the Riester scheme.

## 1.2 Institutional Background

The German Retirement Wealth Act (*Altersvermögensgesetz*) was adopted in June 2001 with the aim of reforming the statutory pension system and promoting funded pension plans (Riester contracts) by means of allowances and tax deductions. It went into effect in January 2002. The relevant unit for determining a Riester subsidy is the income tax unit, which can be a single individual or two individuals who file a tax return jointly (married couples and registered civil partnerships). Allowances and tax deductions depend on the number of children, the presence of a partner, income, and dedicated savings in the tax unit.

**Eligible Population** In Germany, every person in mandatory pension insurance is *directly* eligible to participate in the Riester scheme: dependent employees, civil servants, persons in vocational education, farmers, and the unemployed who receive unemployment benefits.[4] Individuals who are married to a directly eligible person and are not permanently separated, are also eligible (*indirect* eligibility). According to estimates by Fasshauer and Toutaoui (2009) from 2007, 71% (38.6 million) of individuals between 15-64 years are eligible. According to Stolz and Rieckhoff (2013), 10.2 million individuals received direct funding from the Riester scheme in 2010.

**Forms of Subsidization** There are two types of subsidies: direct allowances and tax deductions. Direct allowances comprise two parts, a personal allowance and child allowances. In 2010, the personal allowance was 154€ p.a. for singles and 308€ p.a. for couples with both parties eligible and participating. The child allowance was 185€ p.a. per child (300€ p.a. if the child was born after the 31st of December 2007). In addition, income tax deductions can be granted. The tax deduction is issued on the basis of a higher-yield test by the tax authority. The tax authority deducts the amount of own contributions including the sum of direct funding (up to a maximum of 2,100€ for singles/4,200€ for couples) from the personal income tax base and calculates an adjusted tax burden. It then adds the amount of direct allowance to the adjusted tax burden and compares it with the regular tax burden. The difference between the two tax burdens is the subsidy due to the tax deduction. The tax deduction is not applied if the difference is negative.[5] The sum of the

[4]Early pensioners are eligible if they have a fully reduced earnings capacity (*voll erwerbsgemindert*).
[5]Formally:

$$TAS = max(0, TB_{NoRiester} - TB_{Riester}),$$

direct funding received and the net tax savings due to the tax deduction, if any, give the overall Riester subsidy. As an example, suppose a childless single earns yearly income liable to social contributions of 60,000€ and the tax rate is 50%. The tax burden before Riester is 30,000€. The maximum subsidized saving amount is $2,100€ = min(60,000€ \times 0.04, 2100€)$. It consists of 1,946€ individual contribution and 154€ allowance. The adjusted tax burden is $(60000€ - 2100€) \times 0.5$. Adding the direct funding of 154€ yields a tax burden equal to $29,104€ < 30,000€$. Thus, this individual is granted a tax deduction of $30,000€ - 29,104€ = 896€$ and the overall subsidy the individual receives is $154€ + 896€ = 1,050€$.

**Minimum Saving Effort** A minimum saving effort is required to receive the full subsidy: the allowances and the personal saving effort must add up to 4% of an individual's income liable to social insurance contributions received in the last year (up to a maximum of 2,100€). Both must be invested into certified financial products called Riester contracts. The individual contribution must be at least 60€ per year. The funding is proportionally reduced if the sum of the allowance and the personal saving effort is less than the required 4%.[6]

## 1.3 Data

### 1.3.1 PHF and Tax Calculation

Our empirical analysis is mainly based on the Panel on Household Finances (PHF), a representative multiply-imputed survey dataset.[7] It covers the balance sheets, pension claims, savings, incomes and demographic characteristics of households living in Germany. The first wave of the PHF was collected in 2010 and 2011. Several variables were also asked retrospectively for 2009. Besides the surveyed variables, PHF provides 1,000 bootstrap weights to avoid problems of unresolved or unknown distributions for test statistics. By bootstrapping the variance estimates, users can rely on familiar routines for testing with (asymptotically) normally distributed random variables.

The PHF contains information on the amount an individual contributes to a Riester contract, but not on the financial support received by the same individual.

with $TAS$ the subsidy due to the tax allowance, $TB_{Riester}$ the tax burden with Riester tax allowance, and $TB_{NoRiester}$ the tax burden without.

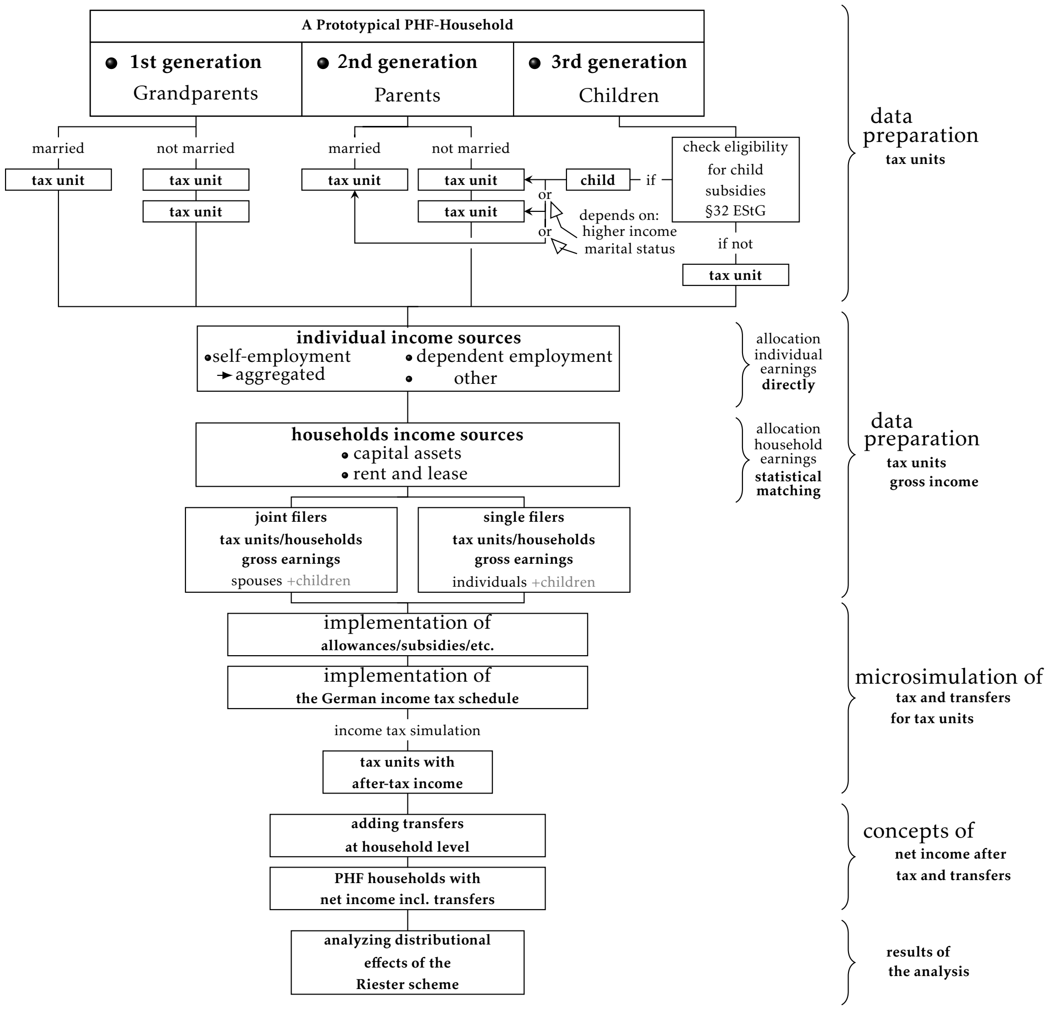
[6] Formally:

$$OCN = max(60, min(0.04 \times Y_{LSC}, 2100) - MDF),$$

where $OCN$ is the own contribution needed, $Y_{LSC}$ is the income liable to social contributions, and $MDF$ is the maximum direct funding. $MDF$ is the personal maximum allowance an individual can receive. E.g., $MDF$ is equal to 154€ for a childless single person.

[7] See von Kalckreuth et al. (2013) and HFCN (2013) for details. See Appendix 1.8.1 for detailed information on our treatment of the multiple imputations.

We compute the subsidy by taking into account information about the household context and by comparing the hypothetical benefits from direct funding with those from tax deduction. Tax units are the reference unit for computing the benefit from tax deduction. To apply the income tax law and calculate the complete Riester subsidies, we have constructed tax units from the PHF households. Afterwards, net incomes and subsidies are aggregated over all tax units forming a household.




*Note.* EStG is the abbreviation for the German income tax code.

**Figure 1.1:** *Data Preparation and Microsimulation*

Our technique of tax unit assignment is depicted in the first brace of figure 1.1. Our assignment method is equipped to deal with all household configurations of the PHF. There are two elementary types of tax units: single adults and couples (with respective children). Single adults are treated as complete tax units. Married couples are treated as a single tax unit, who files jointly. Non-married couples are treated as two separate tax units. In multi-generational households, we draw on the PHF relationship matrix to recover the above elementary types. Children are distinguished from adults by eligibility for child subsidies.

In order to compute the tax liability we aggregate all the taxable incomes at the tax unit level. One can verify from figure 1.1 (second brace) that information on income from self-employment, dependent employment and other income is provided at the individual level and can thus be assigned directly to the tax units. However, capital income and income from renting and leasing are provided at the household level. In households containing multiple tax units, assignment to each unit is not straightforward because information on individual ownership of the underlying assets is not available. To overcome that problem we exploit the information contained in another dataset, the German Socio-Economic Panel (SOEP).[8] In 2012 the SOEP asked individuals about the distribution of assets within their household. SOEP thus provides the relevant information on the within-household allocation of capital income and income from renting and leasing. We transfer this information to the PHF through statistical matching. The criterion for a match is the Mahalanobis distance measure.[9] We calculate the Mahalanobis distance measure based on certain covariates and assign a SOEP observation to each PHF observation according to this distance.[10]

After the match is complete and the ownership percentages are assigned, data on ownership of relevant assets is at the level of the individual in our dataset. Since the ownership percentages from the match do not necessarily sum to 100% in each household, we perform reweighting to achieve consistency. We employ the following reweighting factor,

$$p_{jk} = \frac{\sum_{i=1}^{I} p_{ijk}}{\sum_{j=1}^{J} \sum_{i=1}^{I} p_{ijk}}, \tag{1.1}$$

where $p_{ijk}$ is the percentage of ownership recovered from the match, $k$ indexes the household, $j$ the tax unit and $i$ the individual in that tax unit. Both the household level capital income and income from renting and leasing are then assigned to the tax units according to the reweighted percentages.

[8] For detailed information on the Socio-Economic Panel, see Wagner et al. (2007).

[9] $D_M(x,y) = \sqrt{(x-y)' CV^{-1}(x-y)}$ ,where $(x,y)$ may be points or vectors and $CV$ is the covariance matrix of $(x,y)$. We use the following variables in the calculation of the distance: household income variable to be assigned, individual-level income variables, number of household members, age. We restrict matches to certain slices of the data, wherein certain variables are in total agreement across matchable observations. These slicing variables are: filing jointly, gender and geographical region (north, east, south, west). The match is performed with replacement.

[10] The theory of statistical matching requires the two datasets to follow a joint distribution process. Then variables missing from either one or the other data set can be taken as independent. Missing information then does not play a role for matching the two datasets and matches are consistent with the joint distribution process. This paradigm for matching datasets has been termed the Conditional Independence Assumption (D'Orazio et al., 2006).

The calculation of the income tax liability relies on an adapted version of the microsimulation model STSM that was designed by Steiner et al. (2008) to calculate income tax liabilities for the SOEP. Since the design of the PHF is very similar to the SOEP, adapting the STSM for the PHF is straightforward (Appendix 1.8.2).

### 1.3.2 Descriptive Statistics

We investigate the redistributive effects of the Riester subsidies with respect to the overall population in Germany and the subset of the population eligible for the Riester scheme. Appendix 1.8.3 reports our results for the participating population. We are interested in the impact on the distribution of equivalent net household income. Net household income is defined as household income after taxes and transfers as described in Appendix 1.8.2. We use the household's OECD modified equivalence scale, $ES^{OECD} = 1 + 0.5 \times (n_{adults} - 1) + 0.3 \times n_{children}$, to adjust for needs.[11]

Our data show that 61.3% of households include at least one person who is eligible to participate in the Riester program. The fraction of households with at least one participating individual is 17%. In absolute terms this means 24,081,123 eligible households and 6,750,514 participating households. The average level of the Riester subsidy per household is only 70.38€ per year, but 36.7% of the beneficiaries receive a subsidy in excess of 500€. Riester subsidies turn out to increase household income by up to 17%.

We use two criteria to define the eligible population. First, households must contain at least one Riester eligible person; second, at least one household member must be below the age of 64. We impose the second criterion because older individuals had little incentive to enter into a Riester contract at the time of the reform or after.[12] Compared to the overall population, the eligible population is younger, has more married household heads and higher income. The fraction of households participating in the Riester scheme is 28% when looking at the eligible population and their average subsidy received is worth 115.94€. Tables 1.1 and 1.2 summarize the key statistics pertaining, respectively, to the overall population and the one eligible for the Riester subsidies.

In our distributional analysis we treat the Riester subsidies, computed as described above, as corresponding income increments for the participating households. Since this analysis rests on the premise that households' investments into Riester contracts mainly displace investments they would have made in other assets, we surmise that such a displacement does not affect other tax-favored assets – because

[11] To cope with outliers at the very bottom and top of the distribution and to limit biasing effects from multiple imputation or measurement error, we employ 98% Winsorization. This entails setting incomes below the first percentile (above the 99th percentile) to the value of the first (99th) percentile. See Hastings Jr. et al. (1947).

[12] According to table 4 in Stolz and Rieckhoff (2013), only 0.06% of the Riester recipients in 2010 were born before 1946.

**Table 1.1:** *Descriptive Statistics for the Overall Population*

| | mean | std. error | min | max | obs. |
|---|---|---|---|---|---|
| equivalent gross household income with transfers without Riester subsidy | 28957 | 450.756 | 850 | 324800 | 3565 |
| equivalent net household income with transfers without Riester subsidy | 25274 | 334.426 | 518 | 221772 | 3565 |
| number of household members | 2.044 | 0.005 | 1 | 8 | 3565 |
| married[c] | 0.495 | 0.008 | 0 | 1 | 3565 |
| age[c] | 52.28 | 0.127 | 18 | 90 | 3565 |
| female[c] | 0.350 | 0.006 | 0 | 1 | 3565 |
| completed vocational training[c] | 0.518 | 0.011 | 0 | | 3565 |
| completed extended vocational training[c] | 0.178 | 0.009 | 0 | 1 | 3565 |
| completed university degree[c] | 0.135 | 0.007 | 0 | 1 | 3565 |
| access to tertiary education[c] | 0.295 | 0.003 | 0 | 1 | 3565 |
| estimated subsidies and subsidy rates | | | | | |
| fraction of households participating in the Riester scheme[a] | 0.170 | 0.009 | 0 | 1 | 3565 |
| level of Riester subsidy[b] | 70.375 | 4.547 | 0 | 1764 | 3565 |
| ratio of subsidy to net household income in % | 0.184 | 0.017 | 0 | 17.111 | 3565 |

*Note.* PHF 2010. Own calculations. 1,000 bootstrap replicate weights used to compute standard errors.
[a] The participation variable is a dummy variable that indicates whether at least one household member currently pays into a Riester contract.
[b] The sum of the Riester subsidies of all tax units within a household.
[c] Variable refers to the household head.

in that case the income increment should be reduced by the foregone tax benefits on those assets. While our data do not allow for a thorough investigation of how shifting savings into Riester contracts actually occurs, they suggest that shifting from other tax-favored assets is rare. Households with tax-favored assets can be expected to use first their regularly taxed assets to invest in Riester contracts; hence, they would give up some of their tax-favored assets only if their regular assets are already nil. Yet, 93% of the participating households have a strictly positive level of wealth invested in bank and savings accounts, stocks, and bonds, i.e. non-tax-favored assets. If checking accounts are excluded, this percentage reduces to 83%, and if we additionally require that the sum of savings accounts, stocks and bonds be larger than the household's saving effort, the percentage of households fulfilling this

**Table 1.2:** *Descriptive Statistics for the Eligible Population*

| | mean | std. error | min | max | obs. |
|---|---|---|---|---|---|
| | | | | | |
| equivalent gross household income with transfers without Riester subsidy | 32168 | 644.275 | 850 | 324800 | 2106 |
| equivalent net household income with transfers without Riester subsidy | 27533 | 454.152 | 518 | 221772 | 2106 |
| number of household members | 2.364 | 0.018 | 1 | 8 | 2106 |
| married | 0.538 | 0.013 | 0 | 1 | 2106 |
| age | 43.29 | 0.210 | 18 | 90 | 2106 |
| female | 0.311 | 0.010 | 0 | 1 | 2106 |
| completed vocational training | 0.545 | 0.013 | 0 | 1 | 2106 |
| completed extended vocational training | 0.177 | 0.012 | 0 | 1 | 2106 |
| completed university degree | 0.146 | 0.010 | 0 | 1 | 2106 |
| access to tertiary education | 0.330 | 0.007 | 0 | 1 | 2106 |
| estimated subsidies and subsidy rates | | | | | |
| fraction of households participating in the Riester scheme | 0.280 | 0.014 | 0 | 1 | 2106 |
| level of Riester subsidy | 115.940 | 7.419 | 0 | 1764 | 2106 |
| ratio of subsidy to net household income in % | 0.303 | 0.028 | 0 | 17.111 | 2106 |

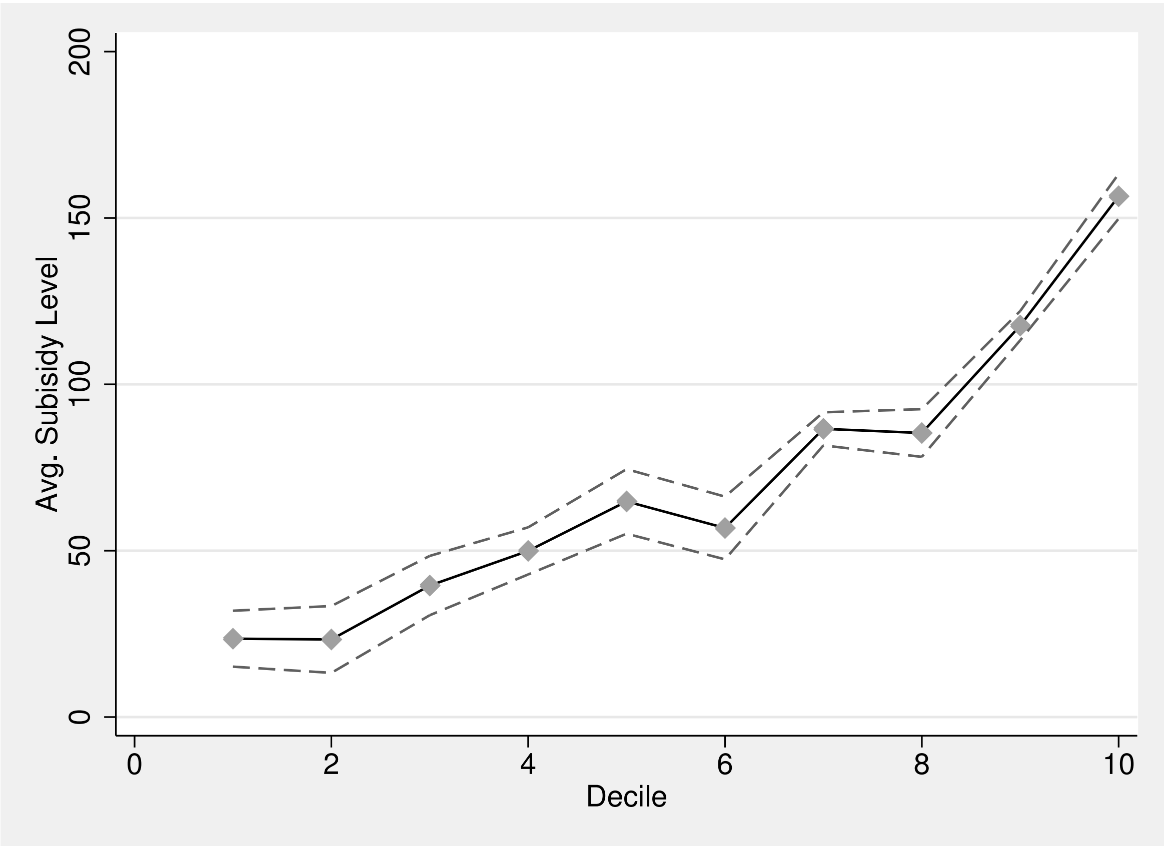
*Note.* PHF 2010. Own calculations. 1,000 bootstrap replicate weights used to compute standard errors. All previous notes on variables in Table 1.1 apply as in Table 1.1.

condition is 77%. As expected, the average income of the participating households that fulfill this condition is significantly larger than the average income of those that do not and may thus have sold tax-favored assets to pay into their Riester contract. All this suggests that the counterfactual we adopt in our distributional analysis is a reasonable first approximation. If investments in Riester contracts fully crowd-out other assets in households' portfolios, our approach is likely to overestimate the progressivity of the Riester scheme, since it appears that a larger fraction of households in the lower tail of the income distribution would lose out on the tax benefits of other assets when shifting their savings into Riester contracts.

# 1.4 Main Results

## 1.4.1 Subsidization along the Income Distribution

We compare two distributions - the one before and the one after receiving the Riester subsidies. The distribution *before* Riester is derived from the PHF data using the above-mentioned tax calculator. The distribution *after* Riester adds to it the simulated Riester subsidies.[13] This comparison captures first round effects and abstracts from any behavioral responses. As usual in the literature, we focus on the distribution of equivalent income to the individuals.




*Note.* Dashed lines indicate the 95% bootstrap confidence interval. Average subsidy level in € p.a. that accrues to households. Deciles are derived from the equivalent net household income distribution. Each decile comprises 10% of the weighted total of households.

**Figure 1.2:** *Subsidy Levels by Decile for the Overall Population*

[13]The effect of the Riester subsidy on a household's equivalent income is determined by summing up the Riester subsidies of all tax units within the household, adding the sum to net household income before Riester, and then dividing the total amount by the household's equivalence scale.

Figure 1.2 shows the decile-specific average subsidy levels for the overall population in Germany. Households are assigned to deciles according to their equivalent net household income before Riester subsidies. We find that the average subsidy increases over the deciles and the increase is sizable. In the bottom decile, the average subsidy is 23.56€. Up to the 6th decile, we find a moderate increase of the subsidy level to about 56.83€. Over the top four deciles, the subsidy level increases to 156€ in the top decile.[14]

By far the largest share of the total subsidy volume accrues to the upper part of the distribution. This can be seen from the concentration curve depicted in figure 1.3 along with the diagonal. The concentration curve of the Riester subsidy is the cumulative share of the Riester subsidy for the centiles of the cumulative distribution function (CDF) of equivalent net income.[15] The concentration curve, unlike the Lorenz curve, can cross the forty-five degree line, yet a concentration curve resting on the forty-five degree line would still imply equal distribution of the subsidy among the population.[16] We find that about 38% of the aggregate subsidy accrues to the top two deciles of the population, while only 7.3% accrues to the bottom two deciles.[17] Hence, the Riester scheme mainly subsidizes high-income households rather than the working poor.

### 1.4.2 Effects on Income Inequality and Poverty

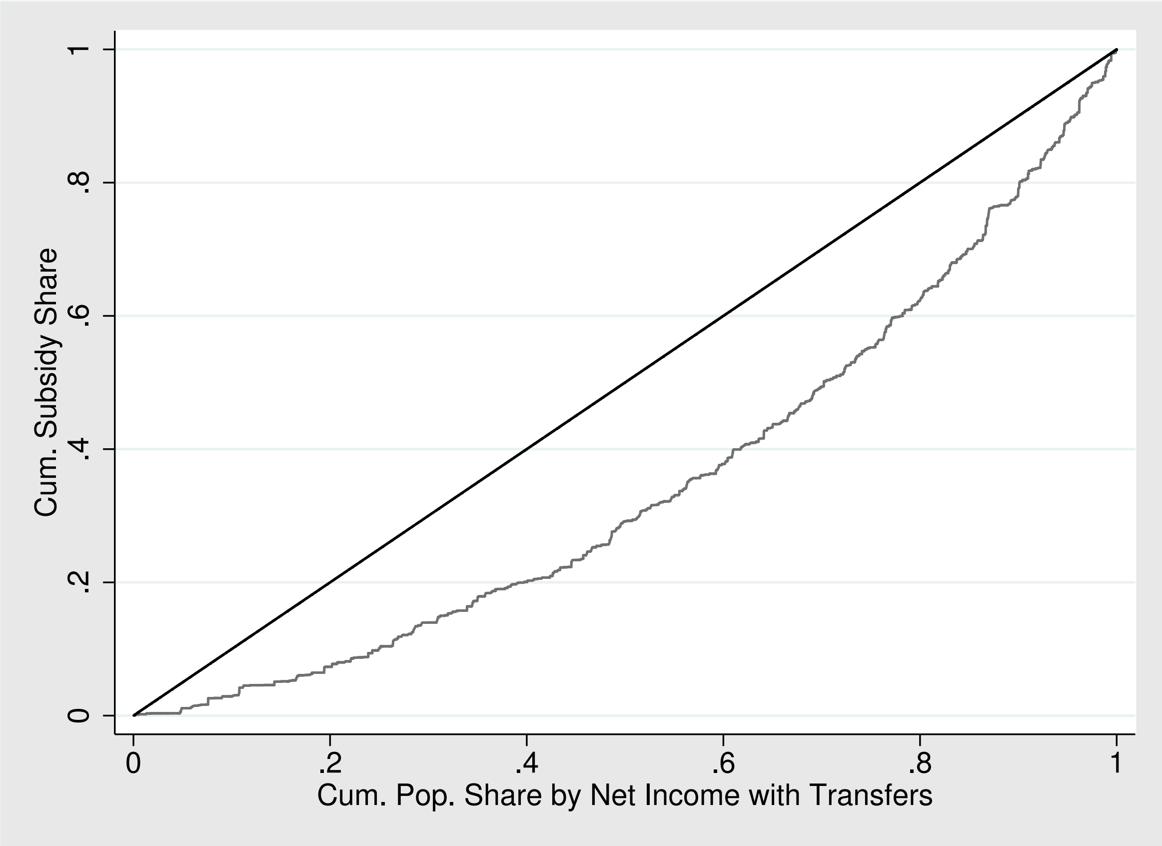
In order to evaluate the redistributive effect of the Riester scheme, we now compute inequality and poverty indices before and after Riester subsidies. Our distributional analysis relies on four inequality indices: the Gini index and three members of the generalized entropy class, namely the Theil index, the mean logarithmic deviation (MLD) and half the squared coefficient of variation ($GE(2)$). These entropy measures imply different levels of inequality aversion, with the $GE(2)$ putting the smallest weight on high incomes and comparatively the Theil the highest, with the MLD lying in-between. Furthermore, we make use of three poverty indices: the headcount ratio (HCR), the income gap ratio (IGR) and the Sen Index (Sen). The HCR is the percentage of individuals under the poverty line, while the IGR gives the average relative income gap of poor individuals from the poverty line. Accordingly, HCR ignores the severity of poverty, while IGR ignores the number of poor individuals, and both are uninformative about the extent of inequality among the poor. Since the

[14] In terms of equivalized subsidies, the average subsidy increases from 15.61€ for the bottom decile to 99.06€ for the top decile.

[15] Lambert (2001, p. 268 pp.) gives an account of how to construct concentration curves for subsidies.

[16] To determine a household's position in the income distributions over the five imputations we calculate the average location of the household in the CDF of income and the average amount of Riester subsidy received.

[17] In Appendix 1.8.4 we also calculate the decile graphs for other statistics like the average subsidy rate and the participation fraction, which are computed analogously to the average subsidy level.




*Note.* Population refers to the population of households sorted by equivalent household income before Riester-induced transfers. This curve gives the cumulated subsidy volume channeled to the poorest x% of households.

**Figure 1.3:** *Concentration Curve for the Overall Population*

Riester scheme may influence all those dimensions, we additionally make use of the Sen Index, $S = H\left[I + (1-I)\,G_{y_i<Z}\right]$, where $H$ is the HCR, $I$ is the product of HCR and IGR, and $G_{y_i<Z}$ is the Gini coefficient for individuals with income $y_i$ smaller than the poverty line $Z$. The poverty line is set at 50% of the median of equivalent net income in Germany. Before Riester subsidies, the poverty line as computed from the PHF is 10,965€, while it raises to 11,007€ after Riester subsidies.

All our inequality and poverty estimates are provided in table 1.3. Column *woR* shows the statistics for the baseline distribution, the income distribution without Riester subsidies. The adjacent column $wR - woR$ gives the change in the index when the Riester subsidies are taken into account. A positive (negative) difference indicates a regressive (progressive) effect of the Riester program.

As shown by the upper panel of table 1.3, the Riester scheme decreases income inequality. Despite the subsidies mainly accruing to the upper part of the distribution, the Gini coefficient and the GE-indices decline after taking the Riester scheme into account. The effects are small:

for instance, the Gini coefficient is reduced by 0.00014. This is not simply a small number, but it is also small as compared to what could be achieved if the budget used for the Riester subsidies was used instead to minimize inequality. In such a case, we compute that the Gini could be reduced by 0.00297, i.e. about twenty times as much. The algorithm used to minimize the Gini coefficient is described by König and Schröder (2018).

The effects of the Riester scheme on poverty are ambiguous. The HCR indicates a rise in the incidence of poverty. The Riester scheme increases the median income and thus the poverty line. Because participation is very low at the bottom of the income distribution, a higher proportion of the population falls below the new poverty line. At the same time, the average distance to the poverty line diminishes. This is reflected by a lower value of the IGR. However, our "overall" poverty index, the Sen Index, increases, pointing out that, while the effect of the Riester scheme on some of the poor is beneficial (as indicated by a lower IGR), it is offset by the changes in its two other components (HCR and $G_{y_i<Z}$).

The above assessment of the distributional consequences of the Riester scheme neglects the fact that the income distribution without the Riester scheme is associated with an improvement of the public budget equal to the total amount of the subsidies. The fiscal costs of the Riester scheme are sizable: at the current rate of participation (17%) we estimate its total volume at 2,790 million € (SD: 180 mil. €).[18] Neglecting the change in the budget allows one to avoid making assumptions about the way the government would use the resources made available by scrapping the Riester scheme. Now, we consider a counterfactual where budget neutrality holds and public expenditure is shifted from the Riester scheme to a hypothetical demogrant. More precisely, we assume that every household in the overall population receives the same amount of equivalized subsidy (48.23 €), with no regard shown to eligibility for the Riester scheme. The ensuing distribution ($wD$) is then compared with the distribution with the Riester scheme.[19] We view such a demogrant as a rough approximation of additional public expenditures on a vast array of publicly provided services that are rather uniformly consumed by the population. The same justification is usually offered when using a demogrant in theoretical models of income redistribution.

The result of our analysis is displayed in the fourth column of table 1.3. The Riester scheme turns out to be less progressive than a demogrant. In absolute terms, these differences are about two to four times larger than the baseline differences for the Riester scheme, $wR - woR$. This shows that even an untargeted instrument

[18] Stolz and Rieckhoff (2013) report the total of direct subsidies for 2010 to be 2,559 mil. €. Since the net gains from tax deductions can be imagined as resting on top of the direct subsidies that households have already received, we see rough agreement with our estimates.

[19] The poverty line is recalculated for the demogrant income distribution, amounting to 10,990€.

like a demogrant – or a general increase in the provision of public services – would redistribute income in a more egalitarian way than does the Riester scheme. A similar conclusion applies if the demogrant is received by the participating population.[20]

The lower panel of table 1.3 shows our results for the eligible population. As mentioned above, the eligible population is younger, has more married household heads and higher income than the overall population. It is thus no surprise that both inequality and poverty indices for the baseline distribution $woR$ are always lower. Because only eligible households remain in the sample, the progressive effect of the Riester scheme without budget balance is stronger. The differences $wR - woR$ for the inequality indices are about twice as large as those for the overall population. The differences for the poverty indices keep their signs, yet get smaller and insignificant. Part of this result is due to the definition of the sample which leads to a relatively strong exclusion of low-income households. Part is also due to the construction of the poverty line, which is determined by the income distribution in the overall population. When using the demogrant as the alternative tool for redistribution, our previous conclusions are confirmed, but the effects are smaller in absolute terms.

To sum up, the Riester scheme has mixed effects on income inequality and poverty that depend on the benchmark used for comparison. At first glance, this finding may seem to be at odds with the results from the incidence analysis, i.e. that most of the overall subsidy volume is channeled to the top of the distribution. But inequality measures are relative: equi-proportional changes of income leave the measured inequality unchanged. Thus, a progressive effect may obtain even if households at the bottom of the distribution receive markedly below-average subsidies. Key for the distributional impact is not how the subsidy *level* varies over the various deciles of the distribution but how the *ratio* of the subsidy to the income level (i.e. the subsidy rate) changes along the income distribution.

## 1.5 Proximate Causes

### 1.5.1 Decomposition

To better understand the drivers of the distributional impact of the Riester scheme, we break the subsidy rate of a given decile of the income distribution, $\sigma$, into its basic components. The subsidy rate is,

$$\sigma = \frac{\sum_{i=1}^{N} s_i}{\sum_{i=1}^{N} y_i}, \tag{1.2}$$

[20]Results can be obtained from the authors upon request.

where $s_i$ is the amount of equivalized subsidy received by each individual $i$ with equivalized income $y_i$ and $N$ is the number of individuals in a decile.[21] Let the members of the decile be ordered so that the first $M \leq N$ participate in the Riester scheme and the remainder does not. Accordingly, we can rewrite the subsidy rate as,

$$\sigma = \underbrace{\frac{\sum_{i=1}^{M} s_i}{\sum_{i=1}^{M} y_i}} \times \underbrace{\frac{M}{N}} \times \underbrace{\frac{N}{\sum_{i=1}^{N} y_i} \times \frac{\sum_{i=1}^{M} y_i}{M}} \tag{1.3}$$

$$= \sigma_M \times \mu \times \frac{\bar{Y}_M}{\bar{Y}}. \tag{1.4}$$

The intensity of subsidization among the group of the $M$ subsidized individuals is captured by $\sigma_M$. Participation is reflected in the participation rate, $\mu = \frac{M}{N}$, and $\bar{Y}_M/\bar{Y}$ is the ratio of the average income of the beneficiaries to the average income of the entire decile. Thus, equation (1.3) shows that the magnitude of the average relative income increase entailed by the Riester program for a given decile can be decomposed as the product of three terms: the average subsidy rate of those who participate, the share of participants within the decile, and the relative income of the participants.

Table 1.4 provides all four statistics and their standard errors for each decile of the income distribution. We first comment on the overall population. The subsidy rate of the decile, $\sigma$, displays a non-monotonic pattern along the income distribution and exhibits relatively small variations across deciles. As we saw in the previous section, this profile entails a small negative effect on inequality if budget neutrality is neglected. In turn, this effect is mainly driven by two opposing patterns concerning $\sigma_M$ and $\mu$. This is shown by the second and the third column of table 1.4. The subsidy rate of the beneficiaries, $\sigma_M$, is highest in the lowest decile and decreases over the income deciles. The participation rate, $\mu$, displays the opposite pattern, i.e. it tends to increase over the deciles. As it turns out, in terms of overall inequality, the progressive effect from $\sigma_M$ slightly dominates the regressive effect from $\mu$.

These results allow us to qualify our previous statement that the Riester scheme is an imprecise tool for redistribution: participation increases over the deciles, explaining why most of the total subsidy is channeled to the upper part of the distribution (figures 1.2 and 1.3), despite higher subsidy rates at the bottom for those who participate. For the eligible population the same basic pattern holds true. Accordingly, the trend in $\sigma$ and the underlying causes of that trend are the same as in the overall population.

[21] We calculate $\sigma$ based on equivalized household incomes and equivalized household subsidies and weight with the number of individuals in each household.

### 1.5.2 Drivers of Participation

If the Riester scheme puts cash on the table for the eligible households, why do so many of them – about 70% – refrain from taking it? While a comprehensive analysis of participation in the Riester scheme is beyond the scope of this paper, we close it by offering an econometric exploration of potential drivers in a simple multivariate framework.

We model the participation decision of household $i$, $C_i$ with $C_i \in \{0, 1\}$, by means of a logit model. The model builds on the form,

$$P(C_i = 1|\mathbf{X_i}) = \Lambda\left(\alpha + \boldsymbol{\gamma} \times \mathbf{X_i}\right), \tag{1.5}$$

where $\Lambda$ is the logistic cumulative density function, $\mathbf{X_i}$ is a set of control variables.

Our first variable of interest for explaining participation is equivalent *net income*. Higher income is expected to bring about a higher saving propensity and, hence, make it easier to surmount the hurdle of the 4% personal saving effort for full direct funding through the Riester scheme. Furthermore, higher income implies a higher marginal tax rate and, hence, a larger benefit from tax deductions. Therefore, we expect income to be a key driver of participation in the Riester scheme. This expectation is borne out by the estimation, as shown by the first row of table 1.5.[22] The coefficient on log income carries a positive sign and is strongly significant in all specifications.

When controlling for the *age* group, we find that people in the highest age bracket (56-64) are significantly less likely to participate in the Riester scheme. This can be explained by the fact that those individuals were relatively old when the Riester scheme was introduced – in 2002 – and had little to gain from entering the program because their accumulation period was short. The presence of *children* in the household increases the probability to participate in the Riester scheme – something that is expected in light of its generous child allowance. Neither the *gender* of the household head nor the *location* of the household in the western or the eastern part of Germany significantly affect the probability of benefiting from the Riester scheme.

In addition to the previously mentioned covariates, Specification (2) of table 1.5 includes dummies for the educational attainment of the household head. A priori, it is unclear how *education* should affect participation. While the better educated are more likely to diversify their portfolios and to be aware of the specific benefits offered by Riester scheme, it is also possible that the less educated are more easily taking up Riester contracts because they were heavily advertised. As it turns out,

[22] We also run specifications with income-decile dummies. After testing, we determine that there is only a significant difference in trend between the third and fourth decile, while all higher deciles (4-10) appear to have the same effect.

the coefficients of the education dummies are insignificant, meaning that we cannot reach any clear-cut conclusion.

Finally, Specification (3) adds a dummy for households that belong to the top quintile of the wealth distribution. The coefficient of the dummy is strongly significant and positive.[23] Its marginal effect at sample means is 0.122 (S.E.: 0.043). With other covariates at their means, belonging to top quintile of the net wealth distribution raises the probability of participating in the Riester scheme from 26% to 38 %.[24]

## 1.6 Qualifications and Extensions

Our analysis has focused on the distribution of annual income and how that distribution is directly affected by the Riester scheme. However, this program also affects the distribution of lifetime incomes. First, the long-term distributional effects will depend on the extent to which pay-outs will be taxed in case of the rich and credited against old-age assistance in case of the poor. Second, lifetime distributive effects might differ from cross-sectional effects if cross-sectional incomes are not tightly linked with permanent incomes. For example, some people have persistently low income so that annual income is a good measure of lifetime income, while other people might have low income for lifecycle reasons (early in the career) or transitory reasons (unemployment or time out of the labor market). Third, the relative importance of full allowance and tax deduction can change over the lifecycle. For example, in the beginning of the earnings career people may receive the full allowance and with rising earnings they may switch to tax deductions, thereby altering the distributional effects of the scheme. We plan to investigate these issues in future research.

Another drawback of the current analysis is that we are silent on behavioral effects. Even though, they are, as we have argued, likely to be small in terms of savings activity, there are other margins of adjustment like labor supply, that we have not considered. The issue of behavioral effects gains even more relevance when not just one period but the life-cycle is considered. A dynamic model of saving,

[23]We have tested down from the full set of net wealth decile dummies. We could not reject that all other dummies are jointly zero and a further test could not reject the equivalence of the coefficients for deciles 9 and 10. Results are available upon request.

[24]Due to the complex survey design and multiple imputation, assessing goodness of fit is nonstandard. McFadden's $R^2$ is unavailable, making us resort to Efron's $R^2$, which is not based on the log-likelihood. Efron's $R^2$ is calculated as

$$R_E^2 = 1 - \frac{\sum(y_i - \hat{p}_i)^2}{\sum(y_i - \bar{y})^2},$$

with $y_i$ the observed values, $\bar{y}$ their average and $\hat{p}_i$ the predicted probabilities from the model.

labor supply and consumption would be the candidate to give the pertinent answers regarding the distributional impact of the Riester scheme in the dynamic setting.

## 1.7 Conclusion

The Riester scheme is the main device used by the German government to subsidize retirement saving. As suggested by previous empirical studies, the Riester scheme largely fails to generate more savings. Rather, it generates windfall gains for a subset of the population. In this paper, we empirically investigate the distributional impact of the Riester scheme. We estimate that 38% of the subsidy volume accrues to the top quintile of the income distribution, but only 7.3% to the bottom quintile. The share of the population below the poverty line increases by nearly one percentage point. Nevertheless, the Riester scheme is almost distributionally neutral with respect to overall inequality measures like the Gini coefficient. Distributional neutrality results from two mutually offsetting effects: a progressive one stemming from the subsidy schedule, and a regressive one from voluntary participation. Participation is quite sparse in the lower deciles of the distribution, but, due to the low incomes at the bottom of the distribution, relative subsidization is high. In the upper part of the distribution participation is more widespread; yet, due to the rapid rise of incomes, subsidy rates cannot keep pace and fall off. We also show that uniformly redistributing the amount spent by the government on the Riester scheme by means of a demogrant would generate a significantly more equal distribution of income.

A simple multivariate regression analysis of take-up behavior of Riester subsidies confirms its correlation with the income of households, even when controlling for the presence of children in the household – another significant driver of participation. On top of that, take-up behavior is significantly explained by high household wealth: belonging to the top quintile of the distribution of net household wealth increases the probability to participate in the Riester scheme by about 12 percentage points.

# 1.8 Appendix

## 1.8.1 Multiple Imputation

Typical phenomena in survey data are unit and item non-response. The PHF exhibits a high unit non-response rate and a low item non-response rate.[25] Wealth and income data are assessed as highly reliable by the data providers, as respondents were willing to answer sensitive questions concerning these items.

To deal with item non-response, the dataset was multiply imputed, both for discrete and continuous variables.[26] If a variable is missing, five values are imputed. Otherwise the observed value is recorded in all imputations.

Multiple imputation entails that imputed values may differ across imputations. This is because an imputed value is the prediction of a regression model specific to that variable. In each imputation after the first, random noise is added to the prediction. This holds also for categorical variables. Accordingly, as an example, a person's status may be employed in one imputation and unemployed in another. Further, the framework of the imputation is hierarchical, meaning that the imputation of some variable depends on the imputed values of others. For example, work status is imputed before employment earnings. Non-uniformity of imputed variables across imputations complicates our analysis. For example, eligibility for Riester subsidies is determined by the employment status, but the employment status need not be the same across imputations. As a result, a household may appear as "eligible" in one imputation and "ineligible" in another. In such cases, we follow the guidelines of Rubin (2004) and define the status as ineligible.

Our analysis of the multiply imputed dataset follows the statistical procedures outlined in Rubin (2004). The point estimate of a variable is computed as the average of the point estimates over all imputations,

$$\bar{Q} = \frac{1}{m}\sum_{r=1}^{m}\hat{Q}_r, \tag{1.6}$$

for any desired point estimate $\bar{Q}$ and any within-imputation point estimate, $\hat{Q}_r$, with the number of imputations $r \in \{1,\ldots,m\}$.

The variance of the point estimate is the weighted sum of two components: the *between-imputation* and the *within-imputation* component. The between component is defined as,

$$B = \frac{1}{m-1}\sum_{r=1}^{m}\left(\hat{Q}_r - \bar{Q}\right)^2. \tag{1.7}$$

[25] See von Kalckreuth et al. (2012) .

[26] See Zhu and Eisele (2013) for details.

The within component, $\bar{U}$, is the average of the *within-imputation* variance estimates, $\hat{U}_r = SE_r^2 = \frac{1/N_r \sum_{i=1}^{N_r} \left(q_{ri} - \hat{Q}_r\right)^2}{N_r}$. The term $q_{ri}$ is the variable of interest for observation $i$, and $N_r$ is the number of observations in imputation $r$. Formally,

$$\bar{U} = \frac{1}{m} \sum_{r=1}^{m} \hat{U}_r. \tag{1.8}$$

The estimate of the total variance is,

$$T = \bar{U} + \left(1 + \frac{1}{m}\right) B, \tag{1.9}$$

which will conform to the Student's t-distribution with $\nu$ degrees of freedom,

$$\frac{\bar{Q} - Q}{\sqrt{T}} \sim t_\nu, \text{with } \nu = (m-1)\left[1 + \frac{\bar{U}}{1 + \frac{1}{m} B}\right]. \tag{1.10}$$

Since our samples are adequately large and item-nonresponse is generally low, making $\nu$ adequately large, we can use the simplifying assumption of the t-distribution approximating the Standard Normal. Thus we calculate 95% confidence intervals as $CI = \left(\bar{Q} \pm 1.96 \times \sqrt{T}\right)$.

## 1.8.2 Details on Tax Calculation

Following the German income tax law, our simulation proceeds in the following steps:

1. Calculating the sum of incomes.
2. Deducting allowances and calculating the taxable income.
3. Implementing progression reservation.
4. Calculating the tax.
5. Testing the higher yield of child allowance and adjusting tax liability.
6. Testing the higher yield of the Riester allowance and adjusting tax liability.
7. Calculating the withholding capital tax.
8. Calculating the solidarity tax.
9. Calculating net income.

10 Aggregating to household level and adding transfers.

Due to the great overlap between the SOEP and the PHF, we can restrict ourselves to two types of notable implementation differences concerning the tax calculation.

**Omissions**

Firstly, in calculating the sum of incomes (point 1), we collect all incomes relevant for the calculation of the tax base. We cannot allow for loss-compensation[27], because negative incomes are not recorded. This is also generally the case for the STSM, but in the PHF operating costs from rent and lease have also not been recorded. Therefore, we cannot deduct these costs from the rent and lease incomes. Secondly, due to lack of adequate information, the calculation of the sum of incomes omits rents of widows and orphans. Thirdly, for the same reason, we cannot deduct the so called *Entfernungspauschale*.[28]

Concerning the deduction of allowances (point 2), we are unable to implement the assessment of childcare costs[29] as a special expense, since there is no data on this expense in the PHF.

Furthermore, we disregard progression reservation (point 3), as it is unlikely to affect individuals relevant to our analysis.

**Improvements**

Concerning point 1, we impute *Werbungskosten*[30] from aggregate statistics[31] by grouping individuals with income from dependent employment. Considering point 10, a feature of the PHF data is its household-level variable on transfers. We simply add transfers to net or gross household income and do not need to model them.

Otherwise the simulation follows exactly the scheme of the STSM tax calculation.

### 1.8.3 Sample 3: Participating Population

The third sample in use is the eligible population holding an active Riester contract, meaning at least one actively contributing person in the household, see table 1.6. Compared to the previous samples, these households receive considerably higher average income. The average number of household members and married household heads is also higher. The level of the Riester subsidy is sizably larger than in the other samples, about 413.60€ on average. About 53% of the households in the participating population benefit from the tax deduction associated with the Riester scheme.

[27] A procedure that deducts losses - either across income sources or across periods - from earnings to lower the tax base.

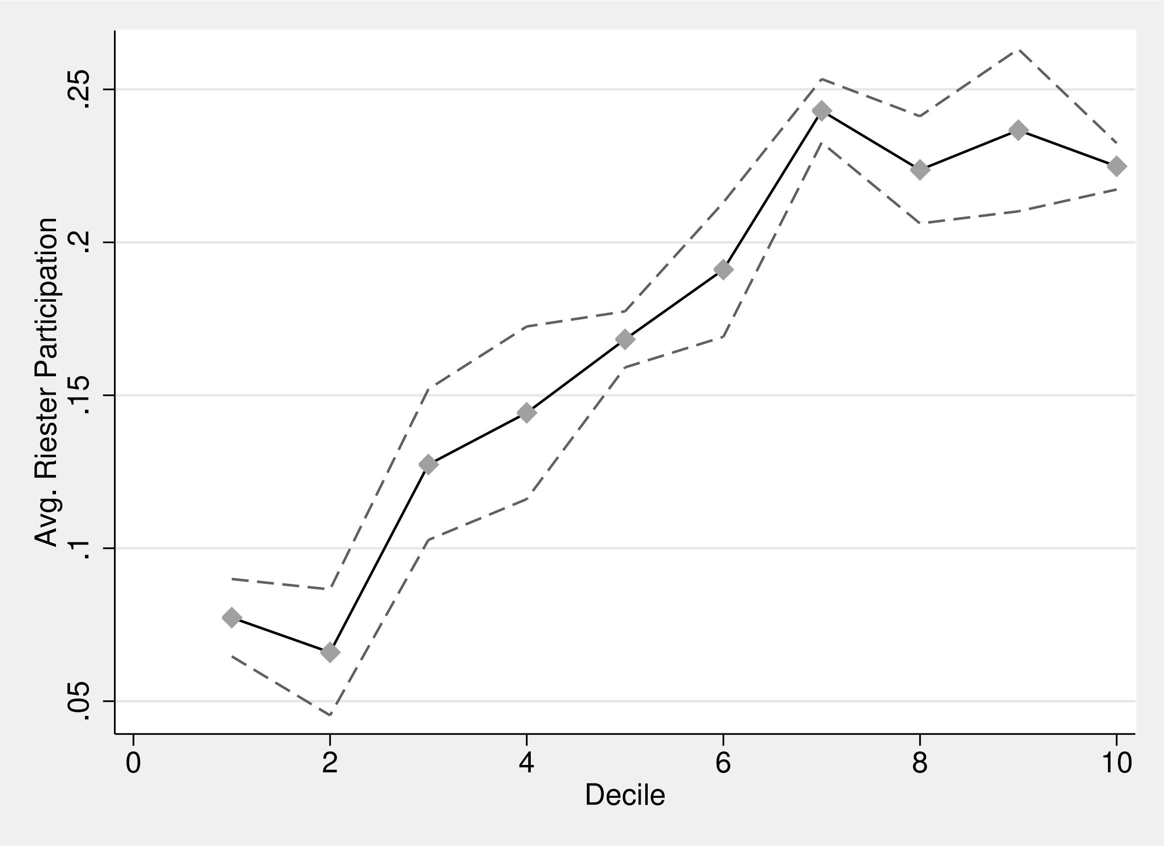
[28] A deduction of costs arising from commutes to one's workplace.

[29] See § 10 S. 1 Nr. 5 EStG.

[30] A deduction of costs that derive from expenses to maintain earnings. See § 9 EStG.

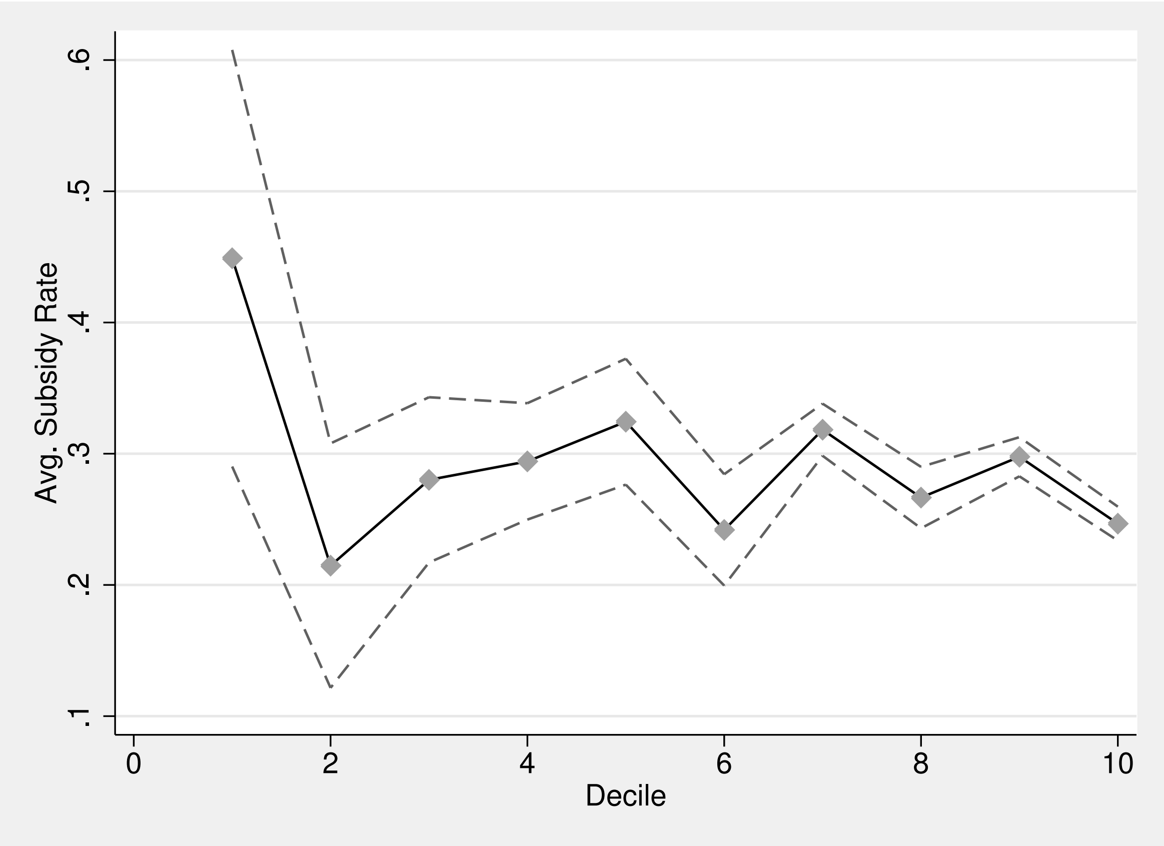
[31] Statistisches Bundesamt (2008), also with data from 2009 and 2010, which came from a special report on request.

## 1.8.4 Graphs




*Note.* Dashed lines indicate the 95% bootstrap confidence interval. Average participation in each decile. Deciles are derived from the equivalent net household income distribution. Each decile comprises 10% of the weighted total of households.

**Figure 1.4:** *Participation Fraction by Decile for the Overall Population*




*Note.* Dashed lines indicate the 95% bootstrap confidence interval. Average subsidy rate in each decile. Deciles are derived from the equivalent net household income distribution. Each decile comprises 10% of the weighted total of households. We define the subsidy rate as the sum of the subsidies divided by the sum of the incomes over all households in each decile, multiplied by 100. In that sense, we do not compute an average of the subsidy rates, but rather the subsidy rate of the decile.

**Figure 1.5:** *Subsidy Rate by Decile for the Overall Population*

**Table 1.3:** *Effects of the Riester scheme on inequality and poverty*

| Overall Population | | | | |
|---|---|---|---|---|
| Measure | $woR$ | $wR - woR$ | $wD$ | $wR - wD$ |
| Gini | 32.960 | -0.014* | 32.899 | 0.048* |
| | *(0.173)* | *(0.002)* | *(0.173)* | *(0.002)* |
| MLD | 20.516 | -0.025* | 20.377 | 0.114* |
| | *(0.347)* | *(0.006)* | *(0.342)* | *(0.009)* |
| Theil | 18.534 | -0.018* | 18.461 | 0.054* |
| | *(0.234)* | *(0.002)* | *(0.233)* | *(0.003)* |
| GE2 | 21.738 | -0.029* | 21.657 | 0.053* |
| | *(0.509)* | *(0.003)* | *(0.508)* | *(0.004)* |
| HCR | 12.237 | 0.798* | 12.052 | 0.983* |
| | *(0.166)* | *(0.158)* | *(0.196)* | *(0.124)* |
| IGR | 35.589 | -2.144* | 35.692 | -2.248* |
| | *(1.172)* | *(0.382)* | *(1.232)* | *(0.291)* |
| Sen | 6.236 | 0.153* | 6.145 | 0.244* |
| | *(0.205)* | *(0.036)* | *(0.202)* | *(0.032)* |
| **Eligible Population** | | | | |
| Measure | $woR$ | $wR - woR$ | $wD$ | $wR - wD$ |
| Gini | 31.750 | -0.031* | 31.693 | 0.026* |
| | *(0.112)* | *(0.003)* | *(0.112)* | *(0.003)* |
| MLD | 18.647 | -0.050* | 18.533 | 0.064* |
| | *(0.299)* | *(0.008)* | *(0.295)* | *(0.010)* |
| Theil | 17.131 | -0.035* | 17.067 | 0.029* |
| | *(0.173)* | *(0.003)* | *(0.172)* | *(0.004)* |
| GE2 | 19.947 | -0.046* | 19.876 | 0.025* |
| | *(0.604)* | *(0.005)* | *(0.603)* | *(0.005)* |
| HCR | 10.444 | 0.253 | 10.301 | 0.396* |
| | *(0.286)* | *(0.167)* | *(0.328)* | *(0.117)* |
| IGR | 33.010 | -0.875 | 33.030 | -0.895* |
| | *(2.155)* | *(0.491)* | *(2.258)* | *(0.344)* |
| Sen | 4.943 | 0.035 | 4.871 | 0.107* |
| | *(0.216)* | *(0.037)* | *(0.214)* | *(0.031)* |

*Note.* All entries were multiplied by 100. Statistical significance of the differences at the 5%-level is indicated by *. 1,000 bootstrap replicate weights used to compute standard errors. Standard errors are displayed in parentheses.
$woR$ refers to the income distribution without Riester subsidies. $wR$ refers to the income distribution with Riester subsidies. $wD$ refers to the income distribution with demogrant.

**Table 1.4:** *Decomposition of Subsidy Rates*

| Decile | Overall Population | | | | Eligible Population | | | |
|---|---|---|---|---|---|---|---|---|
| | $\sigma$ | $\sigma_M$ | $\mu$ | $\bar{Y}_M/\bar{Y}$ | $\sigma$ | $\sigma_M$ | $\mu$ | $\bar{Y}_M/\bar{Y}$ |
| 1 | 0.449 | 4.982 | 0.077 | 1.160 | 0.712 | 4.652 | 0.147 | 1.038 |
| | (*0.081*) | (*0.599*) | (*0.006*) | (*0.043*) | (*0.095*) | (*0.313*) | (*0.012*) | (*0.034*) |
| 2 | 0.215 | 3.166 | 0.066 | 1.021 | 0.505 | 2.749 | 0.182 | 1.013 |
| | (*0.048*) | (*0.292*) | (*0.011*) | (*0.011*) | (*0.054*) | (*0.125*) | (*0.018*) | (*0.009*) |
| 3 | 0.280 | 2.153 | 0.127 | 1.020 | 0.610 | 2.132 | 0.286 | 1.003 |
| | (*0.032*) | (*0.108*) | (*0.013*) | (*0.007*) | (*0.055*) | (*0.136*) | (*0.024*) | (*0.004*) |
| 4 | 0.294 | 2.049 | 0.144 | 0.998 | 0.493 | 1.742 | 0.282 | 1.001 |
| | (*0.023*) | (*0.131*) | (*0.014*) | (*0.006*) | (*0.054*) | (*0.130*) | (*0.013*) | (*0.004*) |
| 5 | 0.324 | 1.914 | 0.168 | 1.007 | 0.507 | 1.489 | 0.341 | 0.998 |
| | (*0.024*) | (*0.120*) | (*0.005*) | (*0.003*) | (*0.025*) | (*0.055*) | (*0.012*) | (*0.004*) |
| 6 | 0.242 | 1.286 | 0.191 | 0.984 | 0.417 | 1.352 | 0.306 | 1.008 |
| | (*0.022*) | (*0.069*) | (*0.011*) | (*0.003*) | (*0.025*) | (*0.062*) | (*0.013*) | (*0.002*) |
| 7 | 0.318 | 1.312 | 0.243 | 0.999 | 0.328 | 1.085 | 0.302 | 0.999 |
| | (*0.010*) | (*0.051*) | (*0.005*) | (*0.002*) | (*0.020*) | (*0.056*) | (*0.003*) | (*0.002*) |
| 8 | 0.267 | 1.187 | 0.224 | 1.004 | 0.423 | 1.261 | 0.336 | 0.998 |
| | (*0.012*) | (*0.038*) | (*0.009*) | (*0.003*) | (*0.025*) | (*0.035*) | (*0.019*) | (*0.004*) |
| 9 | 0.298 | 1.272 | 0.237 | 0.991 | 0.402 | 1.323 | 0.305 | 0.997 |
| | (*0.008*) | (*0.065*) | (*0.014*) | (*0.005*) | (*0.020*) | (*0.058*) | (*0.020*) | (*0.003*) |
| 10 | 0.247 | 1.098 | 0.225 | 1.000 | 0.337 | 1.068 | 0.317 | 0.996 |
| | (*0.007*) | (*0.044*) | (*0.004*) | (*0.018*) | (*0.010*) | (*0.038*) | (*0.011*) | (*0.018*) |
| Average | 0.293 | 2.042 | 0.170 | 1.018 | 0.473 | 1.885 | 0.280 | 1.005 |

*Note.* PHF 2010. Own calculations. There are slight deviations from the formula due to rounding errors. Both $\sigma$, $\sigma_M$ and their standard errors have been multiplied by 100. 1,000 bootstrap replicate weights used to compute standard errors. Standard errors displayed in parentheses. The row “Average” gives column-averages of the respective point estimates. The decomposition does not apply to that row.

**Table 1.5:** *Logit Models of Participation*

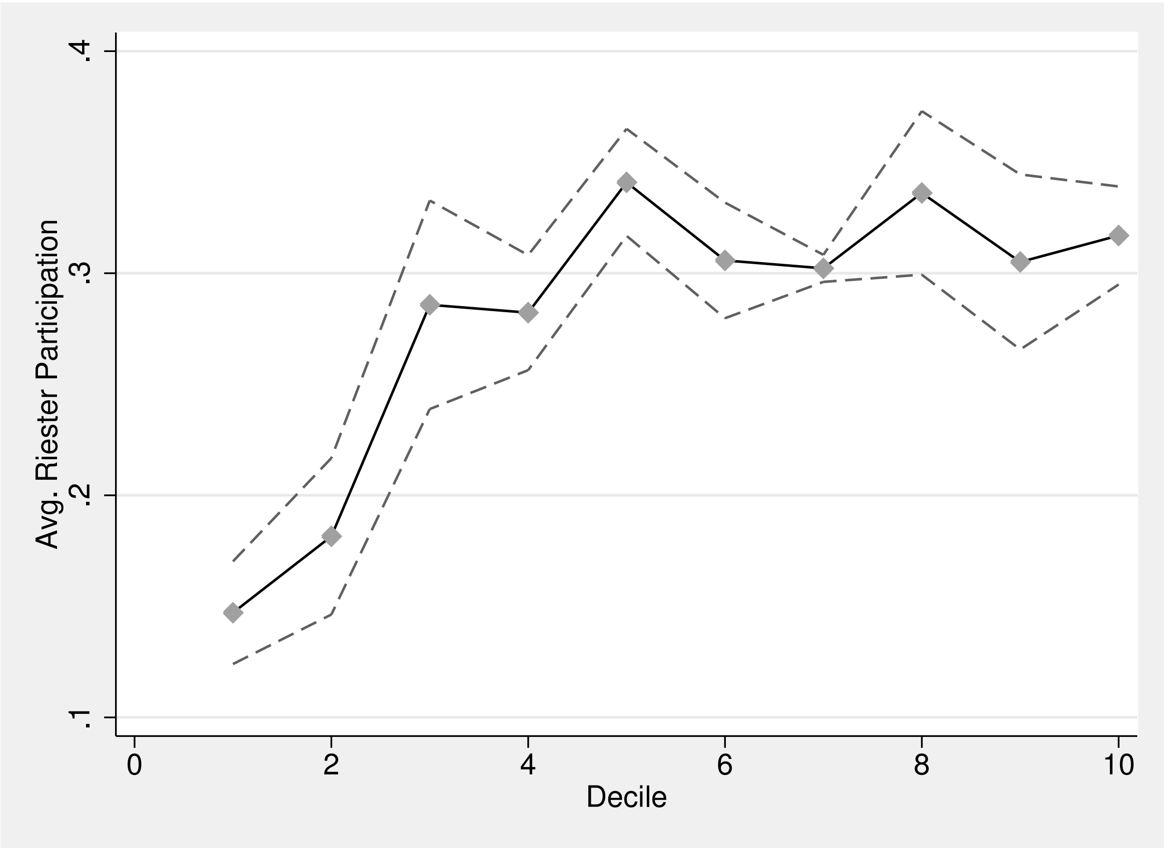
| | Specification (1) | Specification (2) | Specification (3) |
|---|---|---|---|
| log of equivalent net income | 0.5778*** | 0.5419*** | 0.4679*** |
| | (0.1347) | (0.1429) | (0.1430) |
| age: 36-45 | -0.2373 | -0.2140 | -0.2340 |
| | (0.1955) | (0.1951) | (0.1954) |
| age: 46-55 | -0.3157 | -0.2978 | -0.3355 |
| | (0.2084) | (0.2091) | (0.2103) |
| age: 56-64 | -1.2090*** | -1.1800*** | -1.2930*** |
| | (0.2229) | (0.2244) | (0.2336) |
| single w/ children | 0.5783 | 0.6016* | 0.5886* |
| | (0.3525) | (0.3492) | (0.3470) |
| couples | 0.0672 | 0.0938 | 0.0807 |
| | (0.2229) | (0.2229) | (0.2226) |
| couples w/ children | 0.6289*** | 0.6585*** | 0.6561*** |
| | (0.2091) | (0.2130) | (0.2115) |
| more than two adults | 0.2943 | 0.3774 | 0.3194 |
| | (0.2654) | (0.2635) | (0.2650) |
| female | 0.1004 | 0.0802 | 0.0774 |
| | (0.1683) | (0.1705) | (0.1730) |
| east | 0.1700 | 0.2031 | 0.2337 |
| | (0.1989) | (0.2044) | (0.2074) |
| sec. educ. completed | | 0.3011 | 0.2627 |
| | | (0.1985) | (0.1978) |
| tertiary educ. completed | | -0.2079 | -0.2165 |
| | | (0.2347) | (0.2320) |
| top quintile of net wealth | | | 0.6262*** |
| | | | (0.2230) |
| constant | -7.0285*** | -6.7657*** | -6.0048*** |
| | (1.3835) | (1.4415) | (1.4400) |
| observations | 2043 | 2043 | 2043 |
| Efron's $R^2$ | 0.065 | 0.066 | 0.069 |

PHF 2010. Own Calculations. 1000 bootstrap replicate weights used to compute standard errors. Standard errors in parentheses. (***) statistically significant at the 1%-level, (**) at the 5%-level, (*) at the 10%-level. We report the average of Efron's $R^2$ over all imputations, which may not be statistically appropriate when the statistic is not normally distributed.

**Table 1.6:** *Descriptive Statistics for the Participating Population*

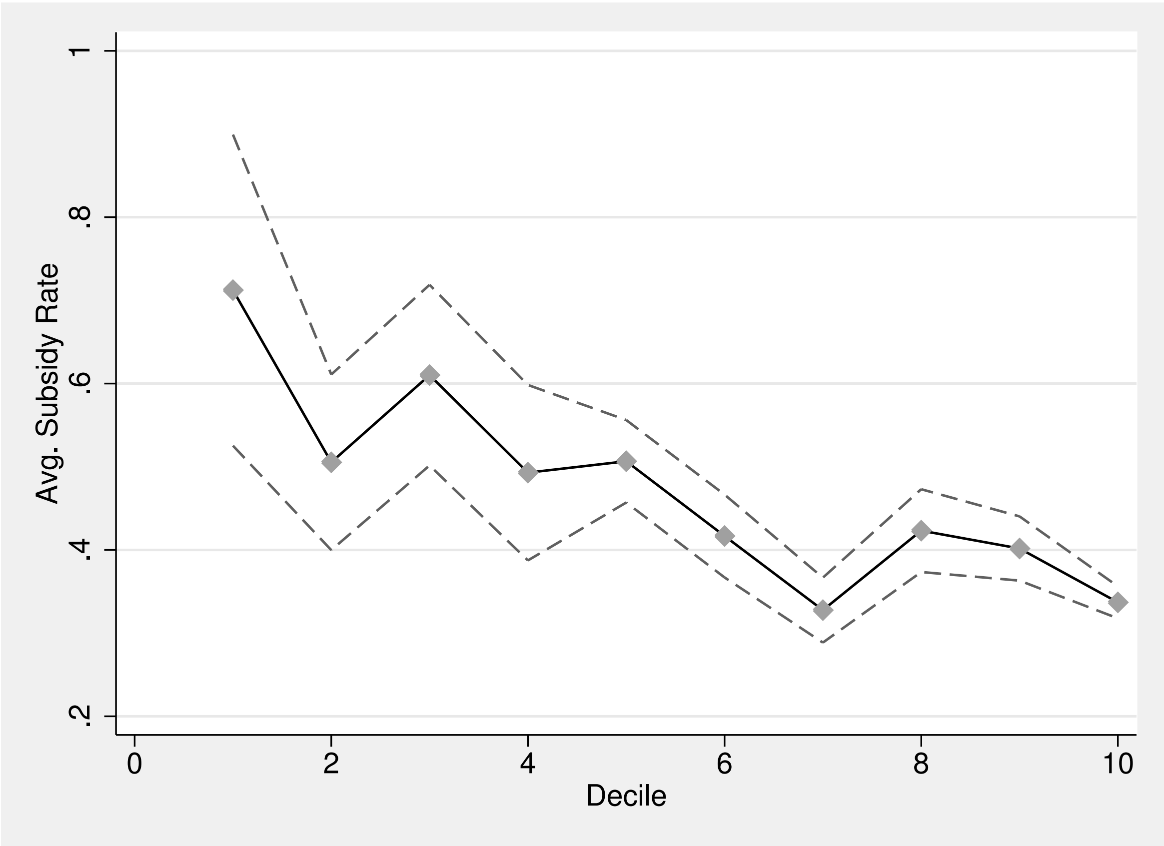
| | mean | std. error | min | max | obs. |
|---|---|---|---|---|---|
| equivalent gross household income with transfers without Riester subsidy | 34844 | 1209 | 1133 | 324800 | 628 |
| equivalent net household income with transfers without Riester subsidy | 29721 | 855.418 | 600 | 221772 | 628 |
| number of household members | 2.738 | 0.062 | 1 | 7 | 628 |
| married | 0.600 | 0.026 | 0 | 1 | 628 |
| age | 41.26 | 0.504 | 18 | 90 | 628 |
| female | 0.282 | 0.023 | 0 | 1 | 628 |
| completed vocational training | 0.554 | 0.025 | 0 | 1 | 628 |
| completed extended vocational training | 0.208 | 0.025 | 0 | 1 | 628 |
| completed university degree | 0.168 | 0.020 | 0 | 1 | 628 |
| access to tertiary education | 0.388 | 0.025 | 0 | 1 | 628 |
| Estimated subsidies and subsidy rates | | | | | |
| level of Riester subsidy | 413.593 | 15.427 | 0 | 1764 | 628 |
| ratio of subsidy to net household income in % | 1.082 | 0.080 | 0 | 17.111 | 628 |
| receiving Riester tax allowance | 0.534 | 0.029 | 0 | 1 | 628 |

*Note.* PHF 2010. Own calculations. 1,000 bootstrap replicate weights used to compute standard errors.



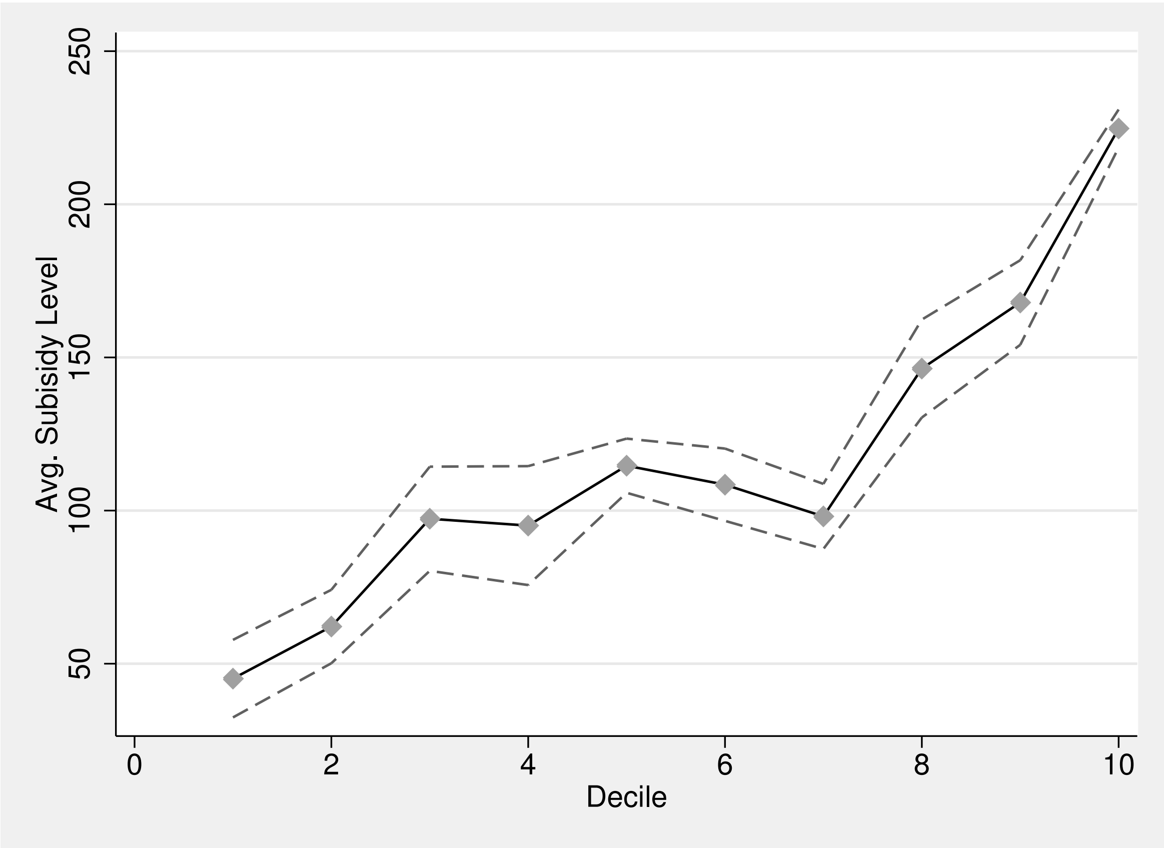

*Note.* Dashed lines indicate the 95% bootstrap confidence interval. Average participation in each decile. Deciles are derived from the equivalent net household income distribution. Each decile comprises 10% of the weighted total of households.

**Figure 1.6:** *Participation Fraction by Decile for the Eligible Population*



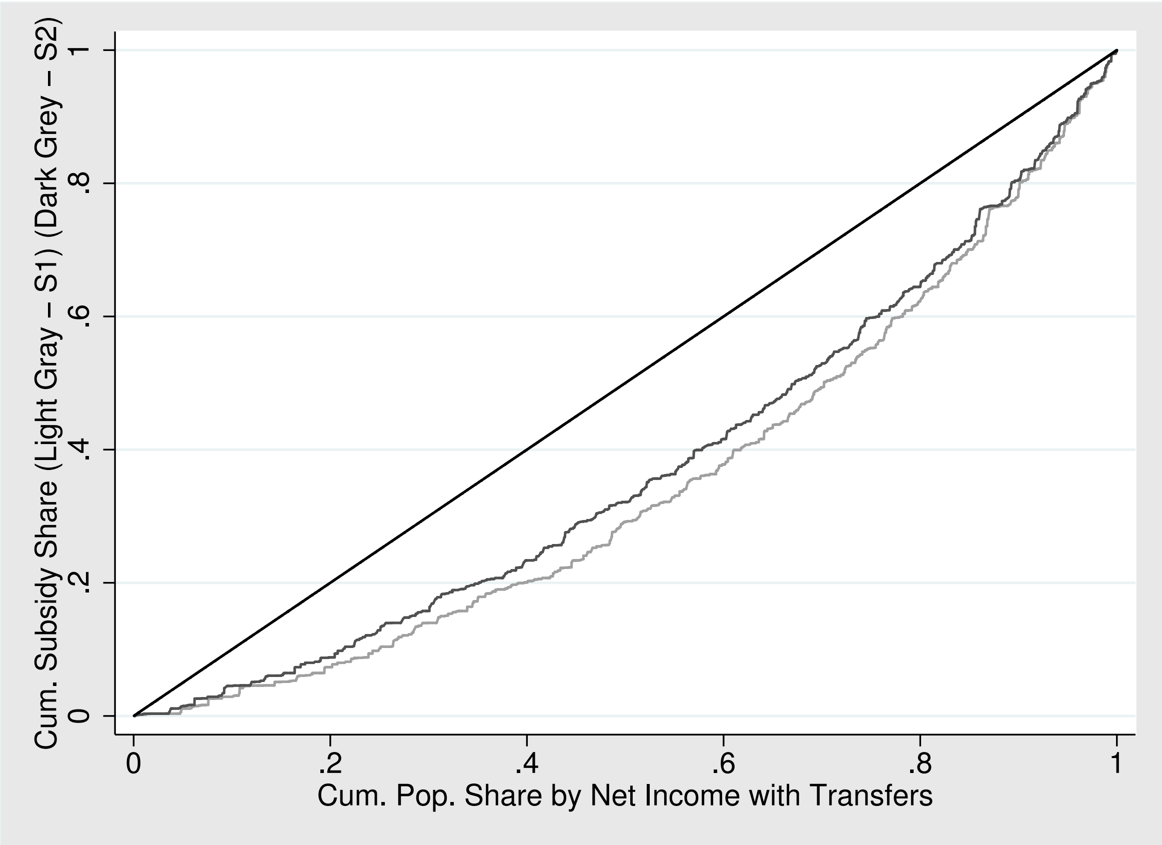

*Note.* Dashed lines indicate the 95% bootstrap confidence interval. Average subsidy rate in each decile. Deciles are derived from the equivalent net household income distribution. Each decile comprises 10% of the weighted total of households. We define the subsidy rate as the sum of the subsidies divided by the sum of the incomes over all households in each decile, multiplied by 100. In that sense, we do not compute an average of the subsidy rates, but rather the subsidy rate of the decile.

**Figure 1.7:** *Subsidy Rate by Decile for the Eligible Population*



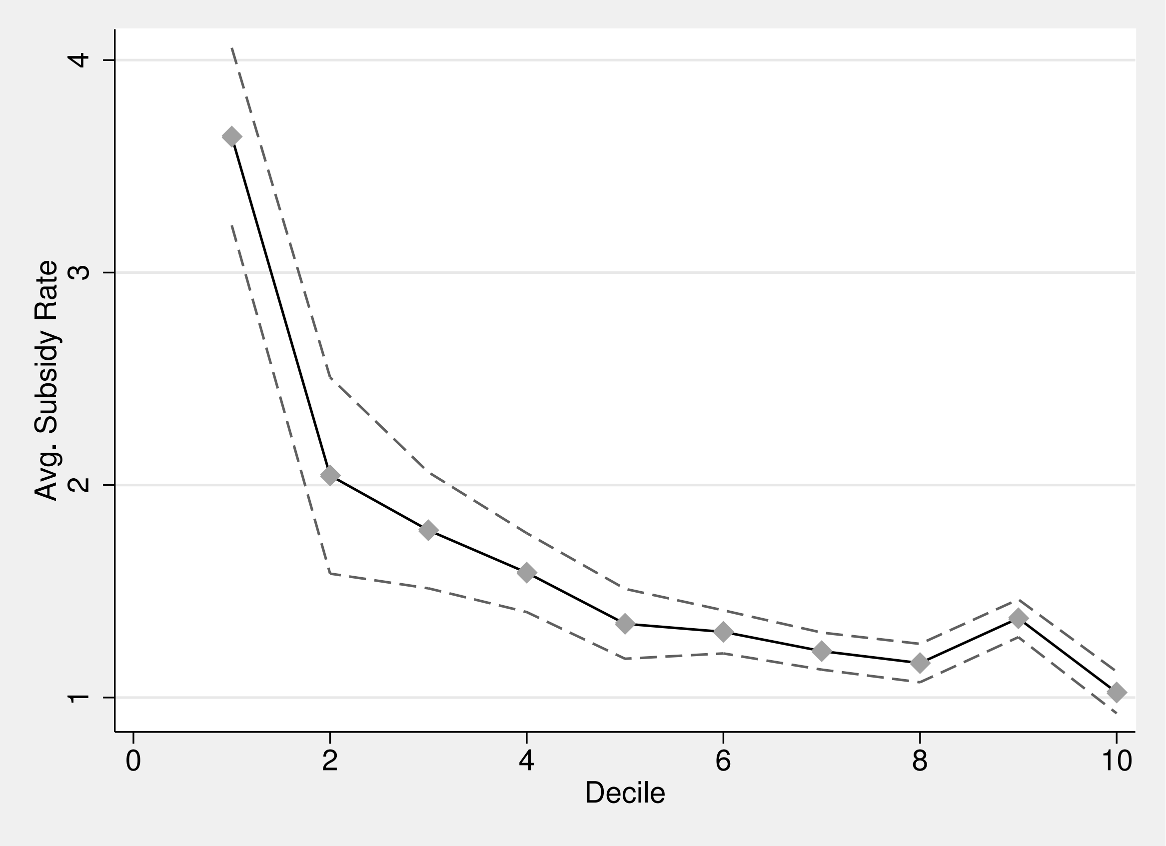

*Note.* Dashed lines indicate the 95% bootstrap confidence interval. Average subsidy level in € p.a. Deciles are derived from the equivalent net household income distribution. Each decile comprises 10% of the weighted total of households.

**Figure 1.8:** *Subsidy Level by Decile for the Eligible Population*



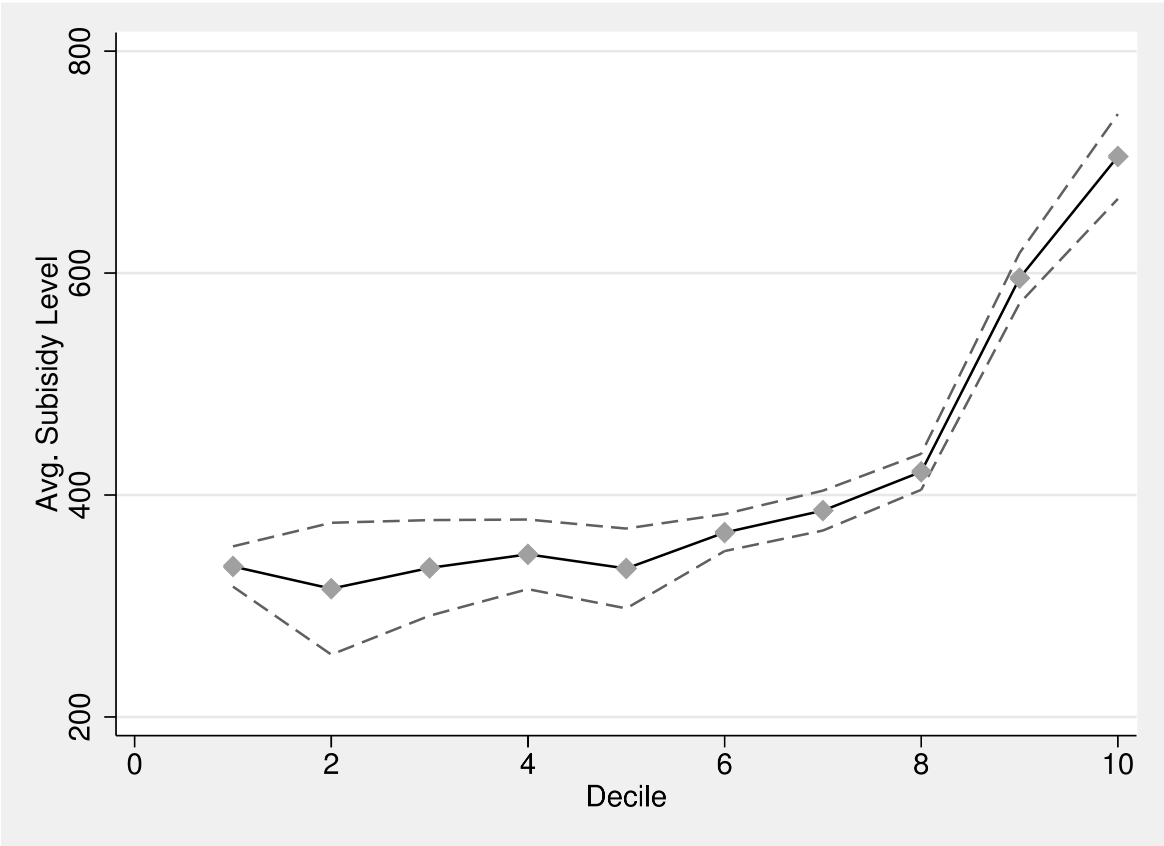

*Note.* Population refers to the population of households sorted by equivalent household income before Riester-induced transfers. This curve gives the cumulated subsidy volume channeled to the poorest x% of households.

**Figure 1.9:** *Concentration Curves for Overall and Eligible Population*



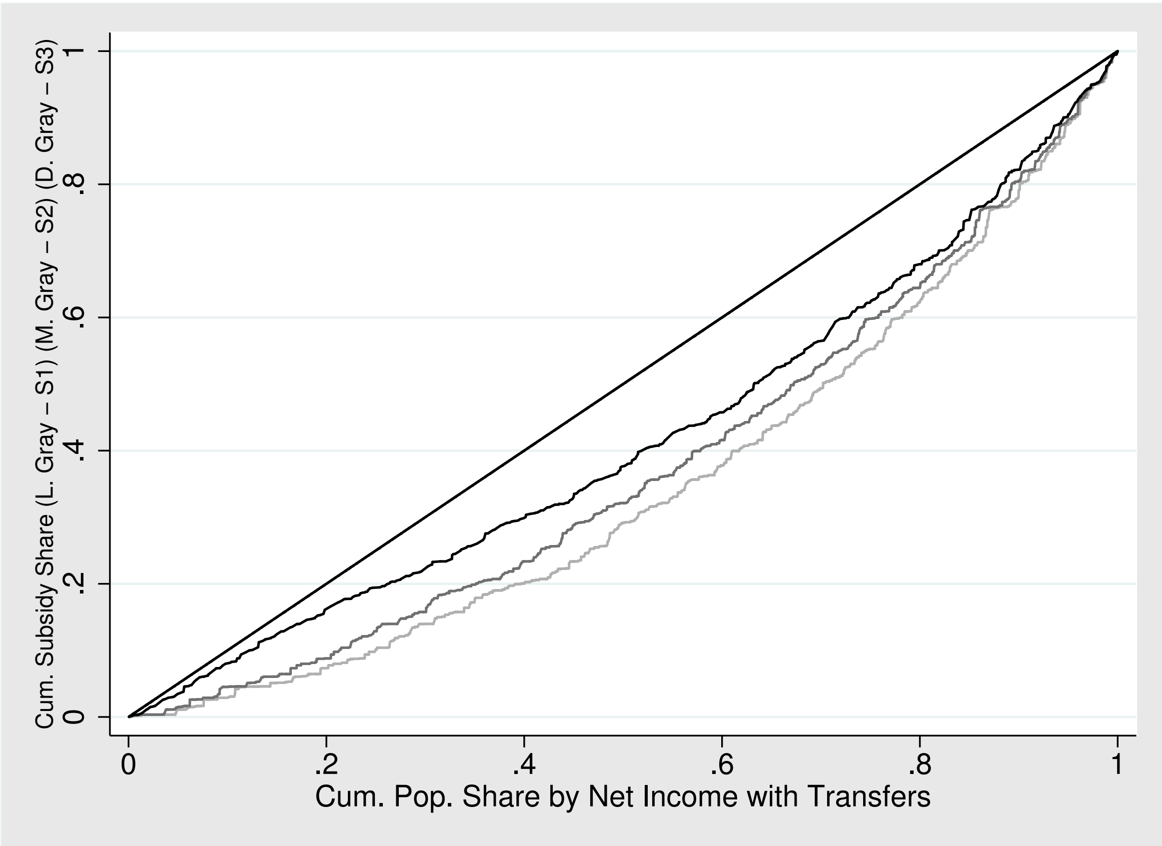

*Note.* Dashed lines indicate the 95% bootstrap confidence interval. Average subsidy rate in each decile. Deciles are derived from the equivalent net household income distribution. Each decile comprises 10% of the weighted total of households. We define the subsidy rate as the sum of the subsidies divided by the sum of the incomes over all households in each decile, multiplied by 100. In that sense, we do not compute an average of the subsidy rates, but rather the subsidy rate of the decile.

**Figure 1.10:** *Subsidy Rate by Decile for the Participating Population*




*Note.* Dashed lines indicate the 95% bootstrap confidence interval. Average subsidy level in € p.a. Deciles are derived from the equivalent net household income distribution. Each decile comprises 10% of the weighted total of households.

**Figure 1.11:** *Subsidy Level by Decile for the Participating Population*




*Note.* Population refers to the population of households sorted by equivalent household income before Riester-induced transfers. This curve gives the cumulated subsidy volume channeled to the poorest x% of households.

**Figure 1.12:** *Concentration Curves for Overall, Eligible and Participating Population*

# 2 Inequality-minimization with a given public budget[1]

## 2.1 Introduction

Globally, policy-makers are seeking to reduce inequalities in market incomes via transfers (Joumard et al. (2012)). Their distributional effects – their progressiveness – are usually evaluated by comparing inequality indices from the market and post-transfer distributions. Unanswered is the issue of the effectiveness of the transfer scheme: Is there an alternative scheme - for a given public budget - that yields a stronger inequality reduction, and what is the maximum feasible inequality reduction? The present paper shows that to minimize a relative inequality index, it is not always optimal to distribute income to units at the bottom of the distribution, resulting in a truncated distribution.[2] Instead, transferring to richer units with larger population weights could be more effective. This "puzzle", first demonstrated by Glewwe (1991), is solved in the present paper.

We identify the optimal transfer scheme by solving a constrained minimization problem. In our context, the problem is to minimize the inequality in a distribution of exogenously given incomes by means of non-negative transfers with a fixed public budget.[3] If the objective function, the index of interest, and the set of constraints is convex, then the problem can be solved by an interior-point algorithm. If the objective function is quasiconvex and the set of constraints is convex, then the problem can be solved by the bisection method.[4] For example, the variance is convex but the Gini index is quasiconvex.

First, consider the effect of a marginal monetary transfer. Its effect is determined by the first derivative of the objective function with respect to the income of the recipient: the larger the first derivative in absolute terms, the larger the inequality reduction. Hence, we need to know what determines the first derivative. Suppose the

[1] This is a post-peer-review, pre-copyedit version of an article published in The Journal of Economic Inequality. The final authenticated version is available online at: http://dx.doi.org/10.1007/s10888-018-9380-3

[2] Since Blackorby and Donaldson (1978), there is an established theoretical literature on the properties of social welfare functions implied by different inequality indices (see also Yitzhaki (1983)). The Atkinson (1970) index is directly constructed from a welfare function.

[3] A related problem is investigated in Prete et al. (2016). The authors identify socially desirable three brackets piecewise linear tax systems that allow collecting given revenue when the aim is to reduce inequality or income polarization.

[4] Convexity requires that the second derivative of the objective function is non-negative. Quasiconvexity requires that the function's sublevel sets are convex. For the definition of a sublevel set see Section 2.2.2.

population is homogeneous, meaning all units are of equal composition, thus having the same material needs. The first derivative of any standard inequality measure increases in the income or rank of the transfer recipient. This means that transfers should always be donated to the household units with the smallest pre-transfer income.

As an example, suppose the planner's objective is the minimization of the Gini index, the public budget is 1 monetary unit, and the income distribution is (14, 20, 30, 40, 80) with a Gini index of 0.33043. The optimal post-transfer distribution is (15, 20, 30, 40, 80) with a Gini index of 0.32432. If the budget were 20 units, the optimal post-transfer distribution would be (27, 27, 30, 40, 80) with a Gini index of 0.23333. The latter distribution is optimal as the entire budget is channeled to the households at the bottom of the distribution and marginal social utilities for all transfer recipients (first derivatives of the Gini w.r.t. transfers) are equal. We call the underlying transfer rule "bottom fill-up," as it minimizes income differences at the bottom of the distribution.

Now suppose the population is heterogeneous, meaning household units differ in composition (i.e., number of household members) and needs, with the latter being measured by an equivalence scale[5]. The common practice of measuring inequality in a heterogeneous population involves two steps. The first step is the needs adjustment of incomes by dividing a household unit's income by its equivalence scale. The second step involves the weighting of household units to construct the equivalent-income distribution. The traditional approach in inequality measurement is to weight households by the number of household members (size weighting). It is consistent with the welfarist's principle of normative individualism: each person is as important as any other. Then the first derivative of the objective function is not determined by income rank alone. Consider again the aforementioned income distribution (14, 20, 30, 40, 80) but now suppose household sizes are (1, 4, 1, 4, 1). Under size-weighting, the sorted equivalent income distribution under the square-root scale is ((10, 4) (14, 1), (20, 4), (30, 1), (80, 1)), with the first number giving equivalent income and the second the unit's weight. The resulting Gini index is 0.36215. For the public budget of 20 units, the bottom fill-up rule gives ((18, 4) (18, 1), (20, 4), (30, 1), (80, 1)) with a Gini of 0.24026. There is, however, a transfer scheme with an alternative feasible post-transfer distribution ((17, 1), (18.5, 4), (20, 4), (30, 1), (80, 1)) that results in a lower Gini index of 0.23940.

Bottom fill-up fails to produce the optimal distribution because, for the size-weighted heterogeneous distribution, the inequality-reducing effect of transfers not only depends on (a) the recipient household's rank in the equivalent income distribution (as in case of a homogeneous population), but also on (b) the household's

[5] An equivalence scale measures household-size economies and differences in needs across household members (e.g. of adults and children). For example, the square-root equivalence scale adjusts household income using the square root of the number of household members.

weight, and (c) the transfer-induced change in average equivalent income (which again depends on the recipient household's weight and equivalence scale). The bottom fill-up rule ignores channels (b) and (c).[6] The basic intuition why bottom fill-up fails for a size-weighted distribution is the following: Transferring income not to the poorest but to a less poor and larger household unit can imply a stronger increase of average equivalent income. This effect reduces any scale-invariant inequality index, and can overcompensate the potential inequality reduction that could have been achieved by shrinking income gaps at the very bottom of the distribution. In sum, the optimal assignment of transfers will hinge on how households of different types sort along the distribution of equivalent incomes, the type-specific equivalence scale, and the type-specific weighting factor. Further, because of the Pigou-Dalton principle, a transfer should always be provided to the poorest household of a certain type.

The optimal transfer scheme provides a lower bound for the feasible post-redistribution inequality. Its most general characterization is in terms of personalized transfers that may depend on the overall distribution of incomes and on household characteristics. Such a scheme is unlikely to be applied. However, our procedures are still useful for practical policy purposes as a benchmark: the optimal solution can be compared to those that could be used in practice in order to identify the second best policies that less differ from the first best one.[7]

Section 2 shows solutions to the aforementioned optimization problem for convex and quasi-convex inequality indices using constrained optimization techniques. Section 3 provides empirical applications. Section 4 concludes.

## 2.2 The Constrained Optimization Problem

Following Ebert and Moyes (2003), suppose households are defined by two attributes: household income $y_i \geq 0$ and household type $h_i \in \mathbb{H}$={1,...,H}. Household material well-being is defined by equivalent income, the ratio of household income and the household's equivalence scale $\text{ES}_i = \text{ES}(h_i) > 0$. In total, the population consists of $N$ households and $Q = \sum_{i=1}^{N} q_i$ individuals, with $q_i$ denoting the number of individuals in household $i$. Let $w_i$ denote the weight of a household. In case of a homogeneous

[6]Under size weighting, economies of scale create a wedge between household size and needs (equivalence scale). A possible way to avoid the wedge is to abandon the principle of normative individualism, and weight households by needs rather than size. Specifically, under needs weighting, the transfer-induced change of average equivalent income does not depend on the recipient household's composition. Characterizations of needs weighted distributions are found in the theoretical works of Ebert (1999), Ebert and Moyes (2003), and Shorrocks (2004). The downside is an "ethical dilemma because individuals who have less extensive needs would be given a lower weight" (Wodon and Yitzhaki (2005, p. 3)).

[7]De facto, however, the transfer schemes boil down to a bottom fill-up with type-specific truncation.

or size-weighted heterogeneous population $w_i = q_i$. In case of a needs-weighted heterogeneous population $w_i = ES_i$. Average equivalent income is $\bar{y} = (\sum_{i=1}^{N} w_i(y_i + t_i)/ES_i)/W$ with $W = \sum_{i=1}^{N} w_i$.

The aim of the social planner is to minimize inequality for a given income distribution, $Y$, via transfers, $t_i \geq 0$, with a given public budget, $B = \sum_{i=1}^{N} t_i$.[8] The planner assigns the same weight to each individual of every household (size-weighting). Inequality is inferred from the distribution of equivalent income - household income, $y_i$, divided by equivalence scale, $ES_i$. The planner's objective function is the inequality index $I : \mathcal{R}_+^{N\times N\times N\times N} \rightarrow \mathcal{R}$, which is at least twice continuously differentiable and convex or quasiconvex.[9] Thus the optimization problem is,

$$\underset{t_i}{\text{minimize}}\ I\left(\left\{w_i, \frac{y_i + t_i}{ES_i}\right\}_{i=1}^{N}\right), \tag{2.1}$$

subject to inequality constraints,

$$0 \leq t_i \ \forall i, \tag{2.2}$$

and

$$\sum_{i=1}^{N} t_i \leq B. \tag{2.3}$$

The corresponding Lagrangian is,

$$L = I\left(\left\{w_i, \frac{y_i + t_i}{ES_i}\right\}_{i=1}^{N}\right) + \sum_{i=1}^{N} \nu_i t_i + \lambda\left(B - \sum_{i=1}^{N} t_i\right), \tag{2.4}$$

implying the following Karush-Kuhn-Tucker optimality conditions (first-order and slack conditions),

[8] Extension to the case of the $t_i \gtreqqless 0$ is straightforward, which enables a full derivation of a tax and transfer system. Further, the investigator may introduce distortions. In the presence of such distortions, $\delta_i$, the transfer net of the distortion is $\tilde{t_i} = t_i - \delta_i$ and the post-transfer income is $y_i - \delta_i + t_i$. The public budget is $B = \sum_{i=1}^{N} t_i$.

[9] Shalit and Yitzhaki (2005) minimize Gini's mean difference subject to linear constraints in a finance context, while Yitzhaki (1982) minimizes the squares of the differences of pre- and post-reform after-tax income subject to non-linear constraints. Both papers consider a convex objective function, while we also provide solutions for quasiconvex functions.

$$\frac{\partial L}{\partial t_i} = \frac{dI\left(\left\{w_i, \frac{y_i+t_i}{ES_i}\right\}_{i=1}^N\right)}{dt_i} - \lambda + \nu_i \overset{!}{=} 0 \Leftrightarrow \left.\frac{dI\left(\left\{w_i, \frac{y_i+t_i}{ES_i}\right\}_{i=1}^N\right)}{dt_i}\right|_{t_i=t_i^*} = \lambda - \nu_i \;\forall i, \tag{2.5}$$

and $\sum_{i=1}^{N} t_i^* \leq B,\ 0 \leq t_i^* \;\forall i,\ \nu_i \geq 0,\ \nu_i(-t_i^*) = 0.$

If all the Lagrange parameters $\nu_i$ are zero and, consequently, not binding in the optimum, then the marginal social utilities of all transfer recipients should be equal to the shadow-price $\lambda$.[10] Take, for example, the Gini index in rank-based formulation, where households are ordered in a non-decreasing fashion with respect to their equivalent income, such that $\frac{y_i+t_i}{ES_i} > \frac{y_j+t_j}{ES_j} \Rightarrow i > j$ and $\frac{y_i+t_i}{ES_i} = \frac{y_j+t_j}{ES_j} \Rightarrow i = j$. Accordingly,

$$G = \frac{1}{W\sum_{i=1}^{N} w_i \frac{y_i+t_i}{ES_i}} \sum_{i=1}^{N}\sum_{i \geq j} w_i w_j \left(\frac{y_i+t_i}{ES_i} - \frac{y_j+t_j}{ES_j}\right). \tag{2.6}$$

For ease of notation, let $u = \frac{1}{W\sum_{i=1}^{N} w_i \frac{y_i+t_i}{ES_i}}$ and $\mathrm{v} = \sum_{i=1}^{N}\sum_{i \geq j} w_i w_j \left(\frac{y_i+t_i}{ES_i} - \frac{y_j+t_j}{ES_j}\right)$.

Suppose we are now implementing a rank-preserving marginal transfer. Replacing the general objective function $I$ in (2.4) with the Gini index, the first-order condition is,

$$\frac{\partial G}{\partial t_i} - \lambda = \frac{\partial u}{\partial t_i}\mathrm{v} + \frac{\partial \mathrm{v}}{\partial t_i}u - \lambda \overset{!}{=} 0, \tag{2.7}$$

with

$$\frac{\partial u}{\partial t_i} = \frac{-w_i/ES_i}{W\left(\sum_{i=1}^{N} w_i \frac{y_i+t_i}{ES_i}\right)^2},$$

and

$$\frac{\partial \mathrm{v}}{\partial t_i} = \sum_{i \geq j} w_i w_j \frac{1}{ES_i} - \sum_{j \geq i} w_j w_i \frac{1}{ES_i}.$$

[10]See Luenberger (1968) and Arrow and Enthoven (1961) for references on quasiconvex programming.

The derivative $\frac{\partial u}{\partial t_i}$ reflects the marginal effect of a transfer on average equivalent income. The derivative $\frac{\partial \mathrm{v}}{\partial t_i}$ shows that the transfer's effect on inequality depends on the recipient household's weight and equivalence scale $\frac{w_i}{ES_i}$ together with its rank in the equivalent income distribution.

For a homogeneous population

$$\frac{\partial u}{\partial t_i} = \frac{-1}{N\left(\sum_{i=1}^{N} y_i + t_i\right)^2},$$

$$\frac{\partial \mathrm{v}}{\partial t_i} = \sum_{i \geq j} 1 - \sum_{j \geq i} 1.$$

and $\frac{\partial \bar{y}}{\partial t_i} = \frac{\partial \bar{y}}{\partial t_j}\ \forall i, j$. The derivative $\frac{\partial \mathrm{v}}{\partial t_i}$ for the homogeneous population reveals that the transfer effect depends on the recipient household's rank (channel (a)). The effect is negative if the second of the two sums is larger in absolute value than the first, i.e. if the recipient household $i$ ranks below the median household. Because the household types are uniform, the effect is independent of the household's weight (channel (b) is irrelevant). For the same reason, the transfer-induced change in average income is independent of the transfer recipient (channel (c) is irrelevant).[11]

For a size-weighted heterogeneous population with $w_i = q_i$, all three channels are reflected in the first-order conditions. The "bottom fill-up"-rule therefore secures optimality only for the case of a homogeneous distribution. For heterogeneous distributions, the optimal transfer scheme can be derived using constrained optimization techniques. If $I(\cdot)$ is convex, the planner's problem can be solved with an interior-point algorithm. If $I(\cdot)$ is quasiconvex, the bisection method can be used. Table 2.1 shows the categorization of several well-known inequality measures with respect to whether they are convex or quasiconvex.[12]

[11] For a needs-weighted heterogeneous population $w_i = ES_i\ \forall i$, $\frac{\partial u}{\partial t_i} = \frac{-1}{W\left(\sum_{i=1}^{N} y_i + t_i\right)^2}$,

$$\frac{\partial \mathrm{v}}{\partial t_i} = \sum_{i \geq j} w_j - \sum_{j \geq i} w_j$$

and $\frac{\partial \bar{y}}{\partial t_i} = \frac{\partial \bar{y}}{\partial t_j}\ \forall i, j$. So, the redistributive effect of $t_i$ depends on channels (a) and (b) but not (c).

[12] See Appendix 2.6.1 for proofs. In the case of the Gini index, the bisection method is rather computer-time intensive. To avoid this computational burden, we exploit that the Gini is a linear-fractional function, meaning that an equivalent convex problem can be solved with the interior-point algorithm. This saves an immense amount of computer time. See Appendix 2.6.2 for details.

**Table 2.1:** Properties of Selected Inequality Indices

| Index | Convex | Quasiconvex |
|---|---|---|
| Variance | Yes | Yes |
| Absolute Mean Deviation | Yes | Yes |
| Relative Mean Deviation | No | Yes |
| Gini Index | No | Yes |
| Theil Index | No | Yes |
| Atkinson Index | No | Yes |

### 2.2.1 Solution for Convex Indices

Several interior-point algorithms to solve convex optimization problems are proposed in the literature. A general introduction to this literature is Boyd and Vandenberghe (2004). This section does not provide a comprehensive introduction; rather it provides a ready access to the literature and implementation for non-specialists.

Following Boyd and Vandenberghe (2004), an interior-point algorithm solves convex optimization problems of the type,

$$\begin{aligned} &\text{minimize } f\left(\{x_n\}_{n=1}^{M}\right) \\ &\text{subject to } g_j\left(\{x_n\}_{n=1}^{M}\right) \leq 0,\ j = 1,\ldots,J\ , \end{aligned} \tag{2.8}$$

where $f(\cdot)$ and the functions $g_j(\cdot)$ map from $\mathcal{R}^M$ to $\mathcal{R}$ and are twice continuously differentiable and convex. In our context, a social planner seeks to minimize inequality with non-negative transfers given a public budget $B$, i.e.

$$\begin{aligned} &\underset{t_i}{\text{minimize}}\ I\left(\left\{w_i, \frac{y_i + t_i}{ES_i}\right\}_{i=1}^{N}\right) \\ &\text{subject to } 0 \leq t_i,\ i = 1,\ldots,N \\ &\sum_{i=1}^{N} t_i \leq B. \end{aligned} \tag{2.9}$$

Ideally, we would proceed to optimize a modified Lagrangian, $L$, with first-order conditions given by,

$$\frac{\partial L}{\partial t_i} = \frac{dI\left(\left\{w_i, \frac{y_i + t_i^*}{ES_i}\right\}_{i=1}^{N}\right)}{dt_i} - \lambda^* + \nu_i^* = 0\ \forall i. \tag{2.10}$$

Because we have one first-order condition for each $i$, but three unknowns, $\nu_i^*$, $\lambda^*$ and $t_i^*$, the system is unidentified. Interior-point algorithms reformulate the problem to avoid underidentification by means of a so-called barrier function, $\iota$, which is integrated into the objective function. In our case, the inequality index is extended by a barrier function that takes large positive values if either the constraints for the transfers and/or the budget constraint are violated.

The reformulated unconstrained problem incorporating the barrier function is,

$$\text{minimize } \tilde{I}(\mathbf{t}) = I\left(\left\{w_i, \frac{y_i + t_i}{ES_i}\right\}_{i=1}^{N}\right) + \iota\left(p, \{t_i\}_{i=1}^{N}\right), \tag{2.11}$$

where $\iota\left(p, \{t_i\}_{i=1}^{N}\right) = -\sum_{i=1}^{N} \frac{1}{p}\text{Log}(t_i) - \frac{1}{p}\text{Log}\left(B - \sum_{i=1}^{N} t_i\right)$, and $p$ denoting the barrier parameter, a large number defined by the researcher.[13] In our setting, the barrier function puts a penalty on violations of the imposed constraints on transfers and the public budget. For example, consider the term $-\frac{1}{p}\text{Log}\left(B - \sum_{i=1}^{N} t_i\right)$: As the sum of transfers approaches the public budget, the term increases exponentially because of the logarithm, and $p$ scales the barrier function. Particularly, as $p$ approaches infinity, $\iota(\cdot)$ approaches the indicator function of our set of constraints,

$$\mathbf{I}_+\left(\{t_i\}_{i=1}^{N}\right) = \begin{cases} 0, t_i \geq 0 \text{ and } \sum_{i=1}^{N} t_i \leq B \\ \infty, t_i < 0 \text{ or } \sum_{i=1}^{N} t_i > B \end{cases}. \tag{2.12}$$

There is an immediate connection between the unconstrained problem (2.11) and the constrained problem (2.9). By multiplying (2.11) with $p$, we obtain an equivalent problem $pI(\cdot) - \sum_{i=1}^{N} \text{Log}(t_i) - \text{Log}\left(B - \sum_{i=1}^{N} t_i\right)$. Now assume that vector $\left\{t_i^*\right\}_{i=1}^{N}$ solves the optimization problem and is within the constraint set. Then the following holds,

$$p\frac{dI\left(\left\{w_i, \frac{y_i + t_i^*}{ES_i}\right\}_{i=1}^{N}\right)}{dt_i^*} - \frac{1}{t_i^*} - \left(\frac{-1}{B - \sum_{i=1}^{N} t_i^*}\right) = 0 \ \forall i. \tag{2.13}$$

To re-convert the first-order condition (2.13) into (2.5), set $\nu_i^* = -\frac{1}{pt_i^*}$ and $\lambda^* = \frac{-1}{p\left(B - \sum_{i=1}^{N} t_i^*\right)}$, giving

$$\frac{dI\left(\left\{w_i, \frac{y_i + t_i^*}{ES_i}\right\}_{i=1}^{N}\right)}{dt_i} - \lambda^* + \nu_i^* = 0 \ \forall i. \tag{2.14}$$

[13] In the applications we set $p = 10^7$.

Hence, there is an equivalence between the unconstrained and the constrained problem outlined in equations (2.11) and (2.9). This equivalence is important, as (2.11) can be solved by Newton's method.

The intuition of Newton's method is illustrated best in a one–dimensional setting following Judd (1998, pp. 96). Suppose our aim is to find the root of a complicated function $f(x)$. The general idea of numerical methods, including Newton's method, is to replace the complicated function with a simple approximation, say $f(x) = mx+b$ with root $x = -\frac{b}{m}$. We also know that the tangent line of a function is the "best" linear approximation of a function around its tangency point.

Since it is easy to find the root of the linear function, Newton's method assumes that the complicated function is a line and then finds the root of the line supposing that the line's crossing is a good approximation to the root of the complicated function. Formally, suppose we have the tangent line of $f(x)$ at $x = a$, where $a$ is the starting point, $f(x) \approx f(a) + f'(a)(x-a) = 0$ and $x = a - \frac{f(a)}{f'(a)}$. If the approximation is poor, the idea of Newton's method is to find the root of a new tangent line at $x = x_1$: $x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$. These so-called Newton steps are undertaken for $x_2, x_3, ..., x_n$ until $\frac{f(x_n)}{f'(x_n)}$ is small, meaning convergence to an approximation of the root.

In our setting, we do not search for a good approximation of the root of a function but for the minimum of the unconstrained problem. In such an optimization Newton's method is applied to $f'$, so that $x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$. This allows us to find $f'(\lim_{n\to\infty} x_n) = 0$. In particular, $x_n \to x$ is a critical point.

More general and for higher dimensions, Newton's method comprises four steps:

**Step 1** Start from a feasible transfer scheme. For example, a feasible scheme is $\mathbf{t}_0$ with $t_i = B/N \; \forall i$. This gives an initial value of the objective function of the unconstrained problem, $\tilde{I}(\cdot)$, $\tilde{I}_0$.

**Step 2** To improve $\tilde{I}_0$, calculate a Newton step. A Newton step takes the known information of the function at a given point (value, gradient, and Hessian), makes a quadratic approximation of that function, and minimizes that approximation. More specifically, the Newton step gives the direction toward the optimum, defined by the weighted negative gradient,

$$\Delta\mathbf{t} = -\left(\nabla^2 \tilde{I}(\mathbf{t})\right)^{-1} \nabla \tilde{I}(\mathbf{t}). \tag{2.15}$$

The Newton step is the product of two terms, the inverse of the Hessian, $\left(\nabla^2 \tilde{I}(\mathbf{t})\right)^{-1}$, and the negative gradient, $-\nabla \tilde{I}(\mathbf{t})$. $\Delta\mathbf{t}$ gives the change of the transfer schedule $\mathbf{t}_0$ (to be multiplied with step-size scalar, $s$). The Hessian is the higher-dimensional generalization of the second derivative and multiplying by its inverse is the non-commutative generalization of dividing by the second derivative $f''$ in the one-dimensional case. Weighting by the Hessian fits the steps to the shape of the contour sets of the function, as shown in Figure 2.1, thereby inducing faster convergence than other methods like *steepest descent*. The ellipse around $\mathbf{t}_0$ is given by the points of unit distance measured in terms of the norm of the Hessian. Step 2 selects the point on the boundary of the ellipse that gives the smallest value of the objective function.

**Step 3** The line through the initial point $\mathbf{t}_0$ and the point on the boundary of the ellipse is defined by $\mathbf{t}_0 + s\Delta\mathbf{t}$ (black line in Figure 2.1). Step 2 searches for the optimal $s$, $s^*$, that determines the point on the line with the smallest level of $\tilde{I}$. This is the so-called *line-search*. One line-search variant is called backtracking. It relies on the idea of approximating $\tilde{I}$ along the direction $\Delta\mathbf{t}$ with a first-order Taylor expansion.

Set $s = 1$ and compute $\tilde{I}$. If the transfer-induced reduction of the objective function is sufficiently large, the line-search algorithm terminates. Sufficiently large means that the new value of the objective function is smaller than the first-order Taylor expansion around $\mathbf{t}_0$, scaled by $\alpha$, i.e, $\tilde{I}(\mathbf{t} + s\Delta\mathbf{t}) \leq \tilde{I}(\mathbf{t}) + \alpha s \nabla \tilde{I}(\mathbf{t})' \Delta\mathbf{t}$. Parameter $\alpha \in (0,1)$ is set by the researcher to define the acceptable decrease of $\tilde{I}$ along the current direction $\Delta\mathbf{t}$.[14] Otherwise, rescale $s$ with parameter $\beta \in (0,1)$, and repeat the above procedure. Again, $\beta$ is determined by the researcher.[15] Completing Step 3 gives the new point $\mathbf{t}_1 = \mathbf{t}_0 + s^*\Delta\mathbf{t}$.

**Step 4** Here the decrease of $\tilde{I}$ between transfer scheme $\mathbf{t}_0$ and $\mathbf{t}_1$ in terms of the squared Newton decrement, defined as $\mathcal{ND}^2(\mathbf{t}) = \nabla\tilde{I}(\mathbf{t})' \left(\nabla^2 \tilde{I}(\mathbf{t})\right)^{-1} \nabla\tilde{I}(\mathbf{t})$, is evaluated. If the squared Newton decrement is smaller than the specific threshold, $2\varepsilon$, with $\varepsilon$ defined by the researcher, then the algorithm terminates. Otherwise, start with $\mathbf{t}_1$ and repeat Steps 2-4 again.

Newton's method with backtracking line search is summarized in the box **Algorithm 1**.

### 2.2.2 Solution for Quasiconvex Indices

Quasiconvex functions have convex sublevel sets. A sublevel set of a function, $f$, is,

$$L_k^-(f) = \{(x_1, \cdots, x_n) \mid f(x_1, \cdots, x_n) \leq k\},$$

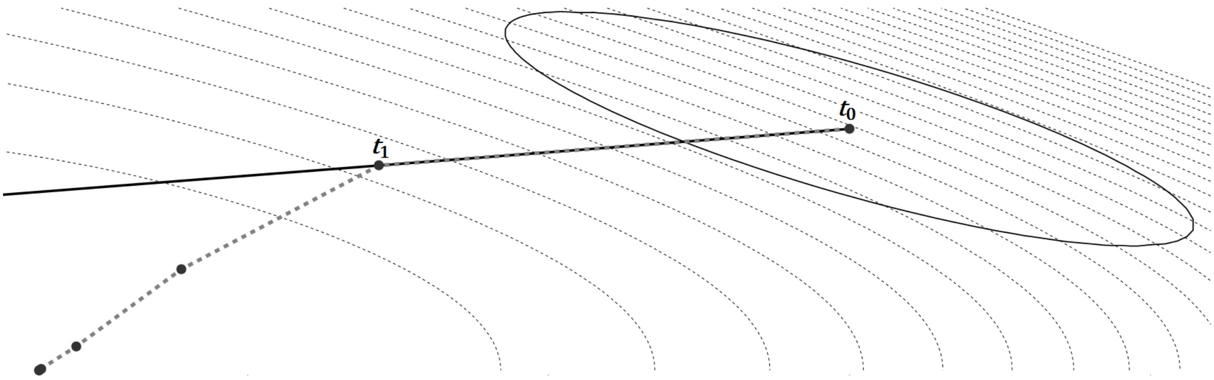
[14] The parameter $\alpha$ cannot be set to be larger than 1, since, at best, the condition $\tilde{I}(\mathbf{t} + s\Delta\mathbf{t}) = \tilde{I}(\mathbf{t}) + s\nabla\tilde{I}(\mathbf{t})'\Delta\mathbf{t}$ can be fulfilled with an infinitesimal $s$. The Taylor expansion is a lower bound on the function $\tilde{I}$.

[15] Backtracking is quite insensitive to the choice of $\alpha$ and $\beta$. See Boyd and Vandenberghe (2004).

**Algorithm 1** Newton's Method with Backtracking Line-Search

Start from some point $\mathbf{t}$ in the domain of $\tilde{I}(\cdot)$ and choose $\varepsilon > 0$ as the tolerance.

Compute $\Delta\mathbf{t}$ and $\mathcal{ND}^2(\mathbf{t})$ at the current point.
**while** $\mathcal{ND}^2(\mathbf{t})/2 > \varepsilon$ **do**
  1. Line-Search: Choose some $\alpha \in (0,1)$ and a $\beta \in (0,1)$ Set $s = 1$.
  **while** $\tilde{I}(\mathbf{t} + s\Delta\mathbf{t}) > \tilde{I}(\mathbf{t}) + \alpha s \nabla \tilde{I}(\mathbf{t})'\Delta\mathbf{t}$ **do**
    Set $s = \beta s$.
  **end while**
  2. Update: Change the current point $\mathbf{t}$ to be $\mathbf{t} = \mathbf{t} + s\Delta\mathbf{t}$
**end while**




*Note.* The ellipse around $t_0$ is given by $\left\{\mathbf{x} \middle| \|\mathbf{x} - \mathbf{t}_0\|_{\nabla^2 f(\mathbf{t}_0)} = 1\right\}$. The black line is given by $\mathbf{t}_0 + s\Delta\mathbf{t}$. Dotted lines are contour sets of $f(\mathbf{t})$.

**Figure 2.1:** Newton's Method with Backtracking Line-Search for the Function $f(\mathbf{t}) = e^{t_1+3t_2-0.1} + e^{t_1-3t_2-0.1} + e^{-t_1-0.1}$.

with $k$ denoting the upper bound of the sublevel set. According to the extreme value theorem, the function implied by the sublevel set has a minimum. The objective is to find the lowest $k$.

The search for the lowest value of $k$ is performed by solving a sequence of convex optimization problems. At every candidate value $k$, there exists a convex function $\Phi_k$, with the property that $\Phi_k(x) \le 0$ if $f(x) < k$. This property implies that a vector $x$ that guarantees $\Phi_k(x) \le 0$ also guarantees $f(x) \le k$. As a corollary, if no vector $x$ exists such that $\Phi_k(x) \le 0$, then $f(x) > k$. This procedure allows distinguishing between feasible and infeasible solutions.

The bisection method provides the lowest feasible value of $k$ by successively eliminating infeasible values. Below, we provide an illustration of the bisection method for identifying a transfer scheme that minimizes the Gini index, $G$, which is quasiconvex. Accordingly, there exists a function, $\Phi_k(G)$, that is convex. As shown in Appendix 2.6.1, $\Phi_k(G) = \sum_{j=1}^N w_j \sum_{i=1}^N w_i \left|\frac{y_i+t_i}{ES_i} - \frac{y_j+t_j}{ES_j}\right| - k\left(2W \sum_{i=1}^N w_i \frac{y_i+t_i}{ES_i}\right)$.

**Step 1** Determine an upper and lower bound of the Gini index, denoted $u$ and $l$. A reasonable choice of $u$ and $l$ is the theoretical minimum of the Gini index, $l = 0$, and the Gini of the pre-transfer distribution for $u$. Further, define a tolerance level as stopping criterion, $\varepsilon > 0$.

**Step 2** Bisect the interval between $l$ and $u$ to find a candidate value $k = \frac{l+u}{2}$.

**Step 3** Check if there exists a feasible transfer scheme $\mathbf{t}$ such that $\Phi_k(G(\mathbf{t})) \leq 0$, $\sum_{i=1}^{N} t_i \leq B$ , $t_i \geq 0\ \forall i$. The feasibility problem is solved by means of an interior-point algorithm. Specifically, the algorithm solves the following problem: minimize 0 s.t. $\Phi_k(G(\mathbf{t})) \leq 0$, $\sum_{i=1}^{N} t_i \leq B$ , $t_i \geq 0\ \forall i$. If a feasible transfer scheme is found, set $u = k$. Otherwise, set $l = k$.

**Step 4** Repeat steps 2-3 until $u - l < \varepsilon$.

The bisection method converges after $\mathrm{Log}_2\left(\frac{u-l}{\varepsilon}\right)$ iterations. The algorithm is detailed in box **Algorithm 2**.

**Algorithm 2** Bisection Method

Set the value $l$ as the lower bound of the function and $u$ as the upper bound. Define $\varepsilon > 0$ to be the tolerance.

```
while u − l ≥ ε do
    1. Set k = (l+u)/2.
    2. Find t s.t. Φ_k(G(t)) ≤ 0, Σ_{i=1}^{N} t_i ≤ B, t_i ≥ 0 ∀i
    if t is a solution of 2. then
        set u = k
    else
        set l = k
    end if
end while
```

Figure 2.2 illustrates the bisection method assuming three iterations. In the first two iterations, there exists a feasible solution to the optimization problem, providing two new upper bounds, $u_1$ and $u_2$. The third iteration fails to provide a feasible solution, leaving the interval $[l_3, u_3]$, which satisfies the tolerance criterion $\varepsilon$.

## 2.3 Application

This section presents two applications of constrained optimization techniques in the context of inequality. The first application relies on two synthetic datasets and

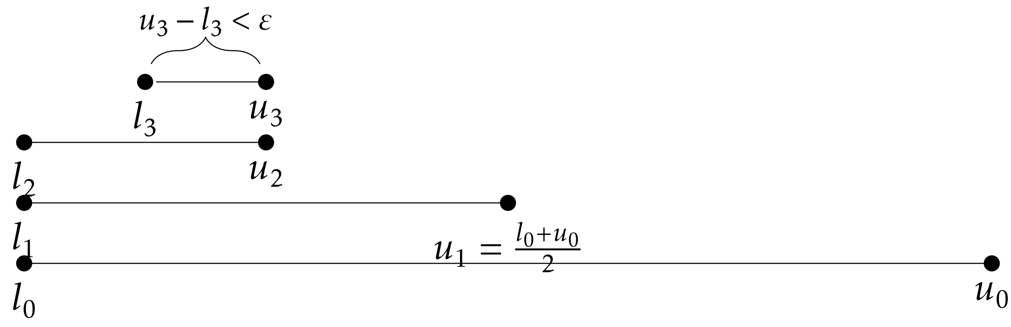



**Figure 2.2:** Bisection of Parameter Range $[l_0, u_0]$

serves two purposes: 1) to illustrate the difference between a transfer-scheme based on bottom fill-up and the optimal transfer scheme; and 2) to give an impression of the optimization problem's computational burden as a function of sample size. The second application is a real-world implementation.

### 2.3.1 Synthetic Data

The synthetic datasets are presented in Table 2.2. The two panels of the table distinguish two alternative pre-transfer distributions (distributions (1) and (2) ). Column $y_i$ provides the household pre-tax income, $q_i$ the number of household members, $ES_i$ the equivalence scale, and $y_i/ES_i$ the pre-tax equivalent income. We assume the planner's exogenous budget is 105 monetary units. Two indices serve as criteria: the quasiconvex Gini index and the convex absolute mean deviation (AMD). The adjacent two columns provide the nominal transfers and post-transfer equivalent incomes under bottom fill-up, while the last two columns provide the same information for the optimal transfer scheme.[16]

For the pre-transfer distribution (1) in the upper panel of Table 2.2 the AMD is 6.420 and 0.073 is the Gini index. The transfer budget suffices to equalize the post-transfer equivalent incomes among the three households at the bottom of the distribution, each getting an equivalent post-transfer income of 107.14 income units. For the post-transfer distribution under bottom fill-up, the AMD is 2.995 and 0.032 is the Gini index. Distribution (1) is constructed in a way that large household types with high weights and high within-household economies of scale cluster at the bottom of the distribution of pre-transfer equivalent income. In this constellation, the optimal transfer scheme coincides with bottom fill-up. This is because all three

[16]We provide the Matlab-code for the methods employed in the Online Appendix Implementation of the optimization of a convex measure is straightforward. One needs to code the inequality measure and then optimize using fmincon.

channels suggest that the households at the bottom of the distribution are the appropriate transfer recipients.

Distribution (2), in the bottom panel, is constructed in a way such that the bottom of the pre-transfer equivalent income distribution comprises heterogeneous household types: large and small households. For example, the lowest-ranked household is a one-member household, but the second-lowest ranked household is a nine-member household. Just focusing on the recipient household's rank (channel (a)), the one-member household should receive a transfer. However, it has a low weight and no or low household-size economies of scale arise (channels (b) and (c)). As a result, the optimal transfer scheme deviates from bottom fill-up. While under bottom fill-up four households at the bottom receive transfers that equalize their equivalent incomes at a level of 113.12 monetary units, the optimal transfer scheme does not grant transfers to the lowest and third lowest-ranked one-member households, but to the second- and fourth lowest ranked nine member households. For the latter two households, the assigned transfers equalize the post-transfer equivalent income. Thus, the optimal transfer scheme is a type-specific bottom fill-up, but the lowest-ranked households need not be the ones that benefit from transfers.

This application provides guidance concerning the practical implementation of an inequality-minimizing transfer scheme. It is apparent that transfers should be granted to the poorest household of each type (type-specific bottom fill up). Transfers should not necessarily be granted to households at the bottom of the distribution. This is the case when households at the bottom of the distribution have small weight and no or low household-size economies of scale. In this case channels (b) and (c) work in favor of higher-ranked multi-member household units. This is a general conclusion, which can be seen from the exercise detailed below.

To complete the exercise, we take the above pre-transfer distributions and identify the optimal post-transfer equivalent income distributions for various levels of the public budget. The resulting optimal post-transfer equivalent income distributions are provided in the two graphs in Figure 2.3. The upper graph refers to distribution (1) while the bottom graph refers to distribution (2). The abscissa provides the transferable public budget, $B$. The ordinate provides the optimal post-transfer equivalent incomes. The dark-grey lines with crosses connect the post-transfer incomes for the one-member households, the black lines with squares for the two, and the light-gray lines with circles for the nine member households.

For distribution (1), bottom-fill up coincides with the optimal transfer rule for all values of the public budget, supporting our above arguments: Starting with a public budget of zero, the first transfer units are allocated to the lowest-ranked households until it catches up with the second-lowest ranked household in terms of post-transfer equivalent income. For higher transfer volumes, transfers are split among the two households in a way that their post-transfer equivalent incomes are the same (truncation of the distribution), explaining why the slope of the recipients'

**Table 2.2:** Synthetic Data

| Distribution | $y_i$ | $q_i$ | $ES_i$ | $y_i/ES_i$ | $t_i^{\text{fill-up}}$ | $\frac{y_i+t_i^{\text{fill-up}}}{ES_i}$ | $t_i^{\text{opt}}$ | $\frac{y_i+t_i^{\text{opt}}}{ES_i}$ |
|---|---|---|---|---|---|---|---|---|
| | 270 | 9 | 3 | 90 | 51.42 | 107.14 | 51.42 | 107.14 |
| | 285 | 9 | 3 | 95 | 36.42 | 107.14 | 36.42 | 107.14 |
| | 134 | 2 | 1.414 | 95 | 17.17 | 107.14 | 17.17 | 107.14 |
| | 156 | 2 | 1.414 | 110 | 0 | 110 | 0 | 110 |
| (1) | 345 | 9 | 3 | 115 | 0 | 115 | 0 | 115 |
| | 120 | 1 | 1 | 120 | 0 | 120 | 0 | 120 |
| | 125 | 1 | 1 | 125 | 0 | 125 | 0 | 125 |
| | 130 | 1 | 1 | 130 | 0 | 130 | 0 | 130 |
| | 191 | 2 | 1.414 | 135 | 0 | 135 | 0 | 135 |
| | AMD | | | 6.420 | | 2.995 | | 2.995 |
| | Gini | | | 0.073 | | 0.032 | | 0.032 |
| | 90 | 1 | 1 | 90 | 23.12 | 113.12 | 0 | 90 |
| | 285 | 9 | 3 | 95 | 54.37 | 113.12 | 75 | 120 |
| | 95 | 1 | 1 | 95 | 18.12 | 113.12 | 0 | 95 |
| | 330 | 9 | 3 | 110 | 9.37 | 113.12 | 30 | 120 |
| (2) | 115 | 1 | 1 | 115 | 0 | 115 | 0 | 115 |
| | 169.7 | 2 | 1.414 | 120 | 0 | 120 | 0 | 120 |
| | 177 | 2 | 1.414 | 125 | 0 | 125 | 0 | 125 |
| | 184 | 2 | 1.414 | 130 | 0 | 130 | 0 | 130 |
| | 405 | 9 | 3 | 135 | 0 | 135 | 0 | 135 |
| | AMD | | | 7.037 | | 4.358 | | 3.53 |
| | Gini | | | 0.078 | | 0.04 | | 0.038 |

income-function gets smaller. If the transfer budget is large enough, such that the post-transfer equivalent incomes of the bottom three households are equalized, the budget is transferred to the three households again securing a truncated distribution, etc.

For distribution (2), bottom fill-up is not the optimal transfer rule for most values of the public budget. This is because it fails to consider the interplay of channels (a), (b), and (c). Due to this interplay, for a transfer budget up to 45, the transfer recipient is not the lowest-ranked household with an equivalent income of 90, a one-member household, but a nine-member household with a pre-transfer equivalent income of 95. Notice that there is another two-member household with the same pre-transfer equivalent income of 95 that does not receive transfers if the public budget is low. The nine-member household is the preferred transfer recipient as the channels (b) and (c) outweigh the gains in terms of inequality reduction that could be achieved by providing the transfers to lower-ranked households. Notice that the equivalent income of the recipient household increases with $B$ at rate $\frac{1}{ES}$.

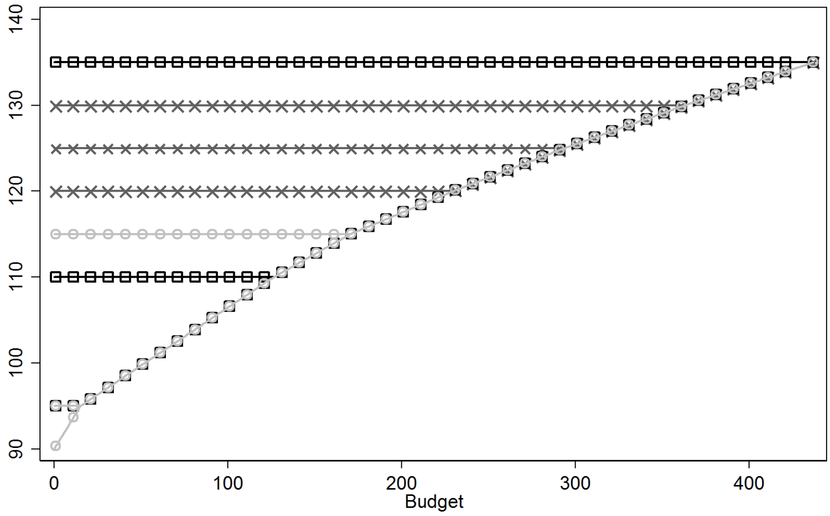
For a budget of $B = 45$, the transfer-receiving nine-member household has the same post-transfer equivalent income as the second-ranked household of its type. For a budget of up to 105, each additional transfer amount is equally split between these two households (household-type specific bottom fill-up). Notice that with a transfer volume of more than 75 there is a re-ranking between the two recipient households and the highest-ranked one-member household. The assignment of the marginal transfer changes when the two nine-member recipient households have an equivalent income of 120, the income of the lowest-ranked two-member household.

For the transfer range between 106 and 165, the one-member households are recipients of each additional transfer unit, and again there is a type-specific bottom fill-up, with the transfers first being allocated to the lowest-ranked one-member household. Once parity of equivalent incomes is reached for the lowest- and second-lowest ranked one-member household, additional transfers are again equally split among the two (household-type specific bottom fill-up), and among all three one-member households for a transfer volume between 150 and 165.

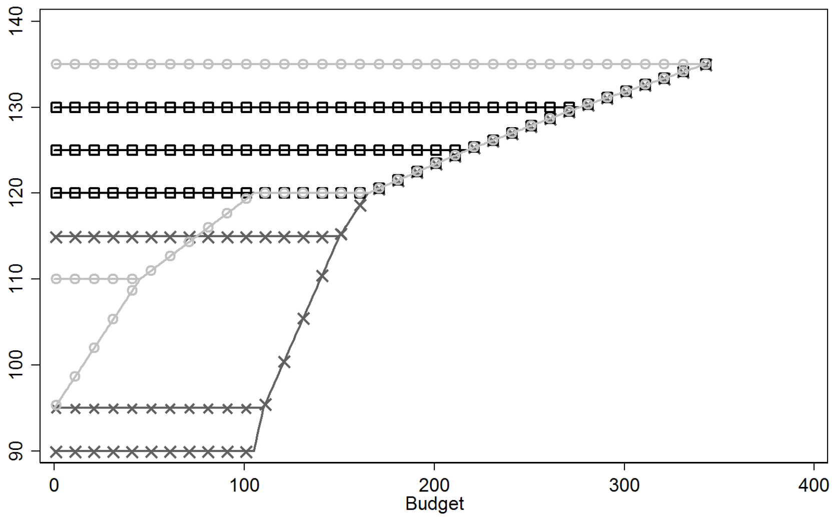
When all recipient households have a post-transfer equivalent income of 120, making them equally well-off with the lowest-ranked four-member household, each additional transfer unit is assigned in a way that the post-transfer-equivalent incomes of all the recipient households are equalized (general bottom fill-up).

**Computational Burden** To gain an impression of the computational burden of the optimization, we sequentially take n-folds of the original number of observations up to 1152 observations, while proportionally adjusting the transfer budget starting at 150 monetary units. Figure 2.4 shows the computational burden in seconds using a computer with an Intel i7-4770 (3.5GHz) processor, 8 GB RAM, and Matlab R2017b. We show results for the variance (convex), the Theil index and the Gini index (both quasiconvex). The comparison between the variance and the Theil reveals the typical differences between a convex and a quasiconvex measure. For the variance the computational burden is low, even for larger datasets. For example, the solution takes about 80 seconds for a sample size of 1152 observations. For the same number of observations, the computer time for the quasiconvex Theil index amounts to about 18500 seconds. The difference results from the fact that we need to apply the interior-point algorithm only once for the variance, while for the Theil index, bisection requires 20 iterations.

As detailed in Appendix 2.6.2, to minimize the Gini, we need to transform the problem to make it differentiable by introducing new variables and constraints that mimic the behavior of the nondifferentiable absolute values: For $n$ transfer recipients, one needs to introduce $n^2 - \frac{(n+1)n}{2}$ new variables and for each of them two new constraints. The computational burden for solving the transformed problem is enormous, e.g., more than 600 hours for a sample of 1152 observations. To reduce the computer time, we exploit that the Gini is a linear-fractional function (see Appendix




(a) Distribution (1)



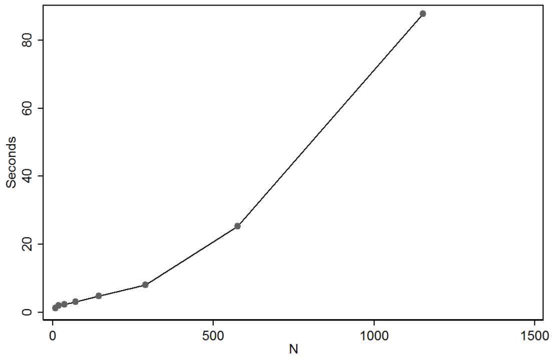

(b) Distribution (2)

**Figure 2.3:** Optimal Post-Transfer Distributions as Functions of Public Budget

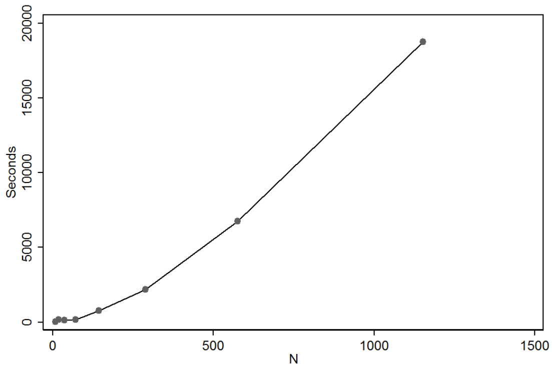
2.6.2), meaning an equivalent convex problem can be solved with the interior-point algorithm. The solution for 1152 observations requires 36 hours.

## 2.3.2 Real Data

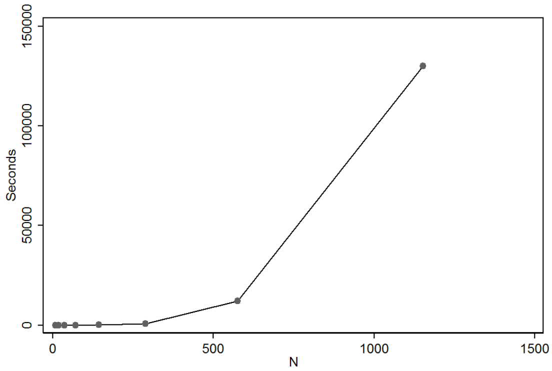
The real-world application deals with an assessment of the re-distributive effects of a pension reform in Germany, the introduction of the so-called Riester scheme. For dependently employed individuals, the scheme grants allowances and tax cuts




**(a)** Variance




**(b)** Theil




**(c)** Gini

**Figure 2.4:** Computational Burden in Seconds Depending on Number of Observations N

based on saving efforts and their household's composition. The policy was intended to have a progressive effect on the German income distribution. Corneo et al. (2018) find a small progressive effect of the Riester scheme on the distribution of equivalent incomes in year 2010: considering the Riester scheme reduces the Gini index from 0.32960 to 0.32946. The total Riester-related transfer volume is about 2.79 billion €. The above methods enable an assessment of the magnitude of the achieved inequality-reduction of 0.04 percent. Applying the bisection method gives an alternative scheme that lowers the Gini index by about 1 percent to a value of 0.32660.

## 2.4 Qualifications and Extensions

The current paper offers a solution to the problem of an inequality minimizing social planner in a setting with fixed incomes. Traditionally, in public economics the problem under consideration is that of a welfare maximizing social planner, who also has to consider the revenue side for the desired transfers. As laid out above, it is possible to introduce distortions into the minimization framework, but this is a way of stepping over the problem considered most relevant in the taxation literature ever since Mirrlees (1971): incentive compatibility. In problems of optimal income taxation of the Mirrleesian tradition there is an information asymmetry between the social planner and the individuals. Earned income is known but the productivity of individuals is not. Thus the planners objective is to offer a tax schedule that incentivizes individuals to earn income according that is commensurate with their productivity and not to imitate the behavior of an individual with differing productivity. In the current setting this complication is ignored since incomes are exogenous. Introducing information asymmetry into the framework is therefore an interesting extension.

Further, an assumption for this setting is the knowledge and simple form of heterogeneity between households. The households only differ in size, composition and income. The equivalence scale for a given household is a deterministic function of household size and composition only, as is standard in most empirical inequality studies. Both assumptions are not uncontroversial.

Assuming full knowledge of the heterogeneity is not an uncontroversial assumption, as the discussion above about the Mirrlees-model illustrates. Ranking in the distribution crucially depends on both the income as well as the size and composition of the household through the equivalence scale and the weight. If households can deceive the planner about these characteristics, then the current solution scheme breaks down. A possible extension is to compare the performance of the current solution technique with more simple methods, like basing transfers only on income rank or a simple demogrant, when there is uncertainty about the either the measured income or the size and composition of the household.

Finally, there is the problem of using, and thereby assuming the identification of equivalence scales. These are supposed to facilitate income comparisons between households of differing compositions. An equivalence scale is supposed to adjust household disposable income such that the equivalized income is that amount of income, which, if given to a single person, would allow them to reach the same level of welfare as a typical member of the household. As laid out in the review article by Chiappori (2016), there are numerous conceptual and econometric issues rooted in the use equivalence scales. Most pressing among the conceptual issues is the assumption that interpersonal welfare comparisons are feasible, stemming from the fact that the definition of an equivalence scale necessitates the definition of household welfare. To adhere to the concept of normative individualism, household welfare needs to be a function of the household members' utilities. Choosing the function that aggregates individual utilities is therefore another crucial issue that possibly restricts the finding of any analysis using the thus defined equivalence scale. Empirically, under strong assumption, equivalence scales can be identified from consumption behavior, which raises a whole host of other questions regarding data quality and identifying assumptions. For all of these reasons the consideration of the role of equivalence scales in this framework is another avenue for future research.

## 2.5 Conclusion

Transfer schemes implemented around the world seek to mitigate inequality in the distribution of market incomes, motivating a central question: How to design the transfer scheme in order to minimize a given inequality index given a particular public budget to distribute. The answer is particularly complex for heterogeneous distributions where households/individuals differ in characteristics or level of needs and may receive different weights, while their income should be equalized using equivalence scales. Then transferring income to the bottom of the distribution fails to guarantee a maximum inequality reduction.

The present paper provides general answers for convex and quasiconvex inequality indices building on mathematical methods developed for solving constrained minimization problems. The adequate solution method depends on the properties of the index of interest. For convex indices, the appropriate method is the interior-point algorithm, for quasiconvex indices it is the bisection method. In application, we show that computer time should not undermine the applicability of the detailed procedures.

In a broader perspective, the methods are suited to solving all kinds of convex or quasiconvex optimization problems. Possible applications include a generalization of our analysis when incomes are endogenous to social transfers and taxes, the minimization of the excess burden of a tax, the analysis of risky prospects, and the construction of optimal portfolios by means of the Mean-Gini approach (see Shalit and Yitzhaki (2005)).

# 2.6 Appendix

## 2.6.1 Proofs of Convexity and Quasiconvexity

The following proofs show the property of convexity or quasi-convexity of a variety of inequality indices. With $y_i$ we denote the income of an individual or a household and abstain from considering weighting factors in order to ease the proofs. However, the proofs are without loss of generality, as one could replace $y_i$ with $\frac{y_i+t_i}{ES_i}$ and scale the sums with weights $w_i$.[17]

### The Variance is Convex

The variance is defined as,

$$V\left(\{y_i\}_{i=1}^N\right) = \frac{1}{N}\sum_{i=1}^N \left(y_i - \frac{1}{N}\sum_{i=1}^N y_i\right)^2. \tag{2.16}$$

The functions $y_i - \frac{1}{N}\sum_{i=1}^N y_i$ are affine and, therefore, convex for all $i$. Squaring these functions and then summing preserves convexity. Therefore, the variance is convex.

### The Absolute Mean Deviation is Convex

The absolute mean deviation is defined as,

$$\text{AMD}\left(\{y_i\}_{i=1}^N\right) = \frac{1}{N}\sum_{i=1}^N \left|y_i - \frac{1}{N}\sum_{i=1}^N y_i\right|. \tag{2.17}$$

The functions $y_i - \frac{1}{N}\sum_{i=1}^N y_i$ are composed with the absolute value, which is a norm. Norms are convex and, therefore, the convexity of the expression is preserved.[18] As before with the variance, this implies that the absolute mean deviation is convex.

### The Gini Index is Quasiconvex

The Gini index can be written as,[19]

[17] If we optimize with respect to $t_i$, then $\frac{y_i+t_i}{ES_i}$ is just an affine transformation of the $t_i$ and therefore preserves concavity or convexity. Changing the sums to be weighted also preserves concavity or convexity. See Boyd and Vandenberghe (2004).

[18] See Boyd and Vandenberghe (2004).

[19] As shown in Yitzhaki and Schechtman (2012) there are more than a dozen alternative ways to define the Gini index.

$$G\left(\{y_i\}_{i=1}^{N}\right) = \frac{1}{2N\sum_{i=1}^{N} y_i} \sum_{j=1}^{N}\sum_{i=1}^{N} \left|y_i - y_j\right|. \tag{2.18}$$

To establish quasiconvexity of $G(\cdot)$, we need to establish that $G(\cdot)$ is quasiconvex in $y_i$.[20] Next we introduce the Gini's sublevel sets, $L_k^-(G)$,

$$L_k^-(G) = \{(y_1, \cdots, y_n) \mid G(y_1, \cdots, y_n) \leq k\}, \tag{2.19}$$

with $k$ denoting the upper bound of the sublevel set. If the elements of $L_k^-(G)$ are convex for every $k$, then $G(.)$ is quasiconvex. The sublevel set for an arbitrary $k$ may be denoted by,

$$\frac{1}{2N\sum_{i=1}^{N} y_i} \sum_{j=1}^{N}\sum_{i=1}^{N} \left|y_i - y_j\right| \leq k \tag{2.20}$$

$$\sum_{j=1}^{N}\sum_{i=1}^{N} \left|y_i - y_j\right| - k2\text{N}\sum_{i=1}^{N} y_i \leq 0. \tag{2.21}$$

This inequality condition holds because of the non-negativity of the mean and can be shown to describe a convex set in $y_i$ by establishing that the left-hand side is a convex function in $y_i$ for all $k$. This is sufficient, since any sublevel set of a convex function is a convex set and here we are studying the sublevel set of the function with level-value zero.

Next, we rewrite the left-hand side as a function that has known convexity properties. We note that the left-hand side can be expressed as the point-wise maximum of $2^{N-1}$ linear expressions in $y_i$ with another linear function in $y_i$ subtracted. For example if $N = 2$ :

$$\max\{2(y_1 - y_2), 2(y_1 - y_2)\} - k4\sum_{i=1}^{2} y_i \tag{2.22}$$

[20]Yitzhaki and Lambert (2013) investigate the relationships between Gini's mean difference (GMD), the mean absolute deviation, the least absolute deviation, and the absolute deviation from a quantile.

The point-wise maximum of linear expressions is convex, so the maximum term is convex.[21] The second term is linear in the $y_i$ and, thus, also convex. So the whole left-hand side is convex for any $k$. Ergo, the Gini index is quasiconvex in the $y_i$.[22]

### The Relative Mean Deviation is Quasiconvex

The following proof builds on the convexity of the absolute mean deviation (see proof above). Since $RMD = \frac{AMD(\{y_i\}_{i=1}^N)}{\frac{1}{N}\sum_{i=1}^N y_i}$, we can form the sublevel sets,

$$AMD\left(\{y_i\}_{i=1}^N\right) - k\frac{1}{N}\sum_{i=1}^N y_i \leq 0. \tag{2.23}$$

The lefthand side contains only convex terms. Hence, the sublevel sets of the RMD are convex. Therefore, the RMD is quasiconvex.[23]

### The Atkinson Index is Quasiconvex

The Atkinson-Index is defined as,

$$A_\epsilon\left(\{y_i\}_{i=1}^N\right) = 1 - \frac{1}{\frac{1}{N}\sum_{i=1}^N y_i}\left(\frac{1}{N}\sum_{i=1}^N (y_i)^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}}. \tag{2.24}$$

First, consider only the second term of $A_\epsilon$ and substitute $p = 1 - \epsilon$. Then,

$$s\left(\{y_i\}_{i=1}^N\right) = \left(\frac{1}{N}\sum_{i=1}^N (y_i)^p\right)^{\frac{1}{p}}. \tag{2.25}$$

This function is concave for $(p-1) < 0$ or equivalently $\epsilon \geq 0$.[24] To show quasiconvexity, we need to establish that the sublevel sets of $A_\epsilon$ are convex. This is sufficiently shown by verifying that the negative term of $A_\epsilon$ has convex sublevel sets, as the rest is just an affine transformation.

The sublevel sets are given by,

[21] See Boyd and Vandenberghe (2004).

[22] Lambert and Yitzhaki (2013) show that the absolute mean deviation is a special case of the between-group Gini mean difference (BGMD). In contrast to the Gini index, the BGMD is not normalized by the mean and is convex.

[23] The relative mean deviation is convex as, in contrast to the absolute mean deviation, the relative mean deviation is divided by the mean.

[24] See Boyd and Vandenberghe (2004, p. 87).

$$-\frac{1}{\frac{1}{N}\sum_{i=1}^{N} y_i}\left(\frac{1}{N}\sum_{i=1}^{N}(y_i)^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}} \leq k \tag{2.26}$$

$$-\left(\frac{1}{N}\sum_{i=1}^{N}(y_i)^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}} - k\frac{1}{N}\sum_{i=1}^{N} y_i \leq 0. \tag{2.27}$$

Next, we assess if the functions on the left-hand side are convex. If they generate sets that are convex given any $k$, quasiconvexity is implied. Since the first function is convex – the negative of $s\left(\{y_i\}_{i=1}^N\right)$ is convex – and the second function is affine, this is the case.

### The Theil Index is Quasiconvex

The definition of the Theil Index is,

$$T = \frac{1}{N}\sum_{i}^{N} \frac{N y_i}{\sum_i^N y_i} \text{Log}\left[\frac{N y_i}{\sum_i^N y_i}\right]. \tag{2.28}$$

For quasiconvexity the sublevel sets of the Theil Index need to be convex. Accordingly,

$$\frac{1}{N}\sum_{i}^{N} \frac{N y_i}{\sum_i^N y_i} \text{Log}\left[\frac{N y_i}{\sum_i^N y_i}\right] \leq k \tag{2.29}$$

$$\sum_{i}^{N} y_i \text{Log}\left[\frac{N y_i}{\sum_i^N y_i}\right] - k\sum_{i}^{N} y_i \leq 0. \tag{2.30}$$

The functions on the left-hand side induce convex sets if they are convex. The second term is affine and, thus, convex. The first term is convex if its Hessian is positive semi-definite. The second partial derivatives of $f(\{y_i\}_{i=1}^N) = \sum_i^N y_i \text{Log}\left[\frac{N y_i}{\sum_i^N y_i}\right]$ are,

$$f_{y_i,y_i} = \frac{1}{y_i} - \frac{1}{\sum_i^N y_i}, \quad f_{y_i,y_j} = -\frac{1}{\sum_i^N y_i}. \tag{2.31}$$

Then the Hessian of $f(\{y_i\}_{i=1}^N)$ is,

$$H_f = \begin{pmatrix} \frac{1}{y_1} - \frac{1}{\sum_i^N y_i} & -\frac{1}{\sum_i^N y_i} & \cdots \\ -\frac{1}{\sum_i^N y_i} & \frac{1}{y_2} - \frac{1}{\sum_i^N y_i} & \\ \vdots & & \ddots \end{pmatrix}. \tag{2.32}$$

We can reshape the matrix before we test for positive semi-definiteness as

$$H_f = \begin{pmatrix} \frac{1}{y_1} & 0 & \cdots \\ 0 & \frac{1}{y_2} & \\ \vdots & & \ddots \end{pmatrix} - \begin{pmatrix} \frac{1}{\sum_i^N y_i} & \frac{1}{\sum_i^N y_i} & \cdots \\ \frac{1}{\sum_i^N y_i} & \frac{1}{\sum_i^N y_i} & \\ \vdots & & \ddots \end{pmatrix}. \tag{2.33}$$

The Hessian is positive semi-definite iff for any vector $\boldsymbol{v}$,

$$\boldsymbol{v}' \begin{pmatrix} \frac{1}{y_1} & 0 & \cdots \\ 0 & \frac{1}{y_2} & \\ \vdots & & \ddots \end{pmatrix} \boldsymbol{v} - \boldsymbol{v}' \begin{pmatrix} \frac{1}{\sum_i^N y_i} & \frac{1}{\sum_i^N y_i} & \cdots \\ \frac{1}{\sum_i^N y_i} & \frac{1}{\sum_i^N y_i} & \\ \vdots & & \ddots \end{pmatrix} \boldsymbol{v} \geq 0. \tag{2.34}$$

To show that this is the case, we rely on the Cauchy-Schwarz-Inequality. It states that for any two vectors $\mathbf{a}$ and $\mathbf{b}$,

$$(\mathbf{a}'\mathbf{a})(\mathbf{b}'\mathbf{b}) \geq (\mathbf{a}'\mathbf{b})^2. \tag{2.35}$$

State the dot-product of the Hessian with $\boldsymbol{v}$ as summations,

$$\frac{1}{\sum_i^N y_i} \left( \left( \sum_i^N y_i \right) \left( \sum_i^N \frac{v_i^2}{y_i} \right) - \left( \sum_i^N v_i \right)^2 \right) \geq 0. \tag{2.36}$$

To complete the proof, pick $a' = (\sqrt{y_1}, \sqrt{y_1}, \ldots)$ and $b' = (\frac{v_1}{\sqrt{y_1}}, \frac{v_2}{\sqrt{y_2}}, \ldots)$, which establishes that the above sums are greater or equal to zero.

Since both functions determining the sublevel sets are convex for any $k$, the Theil is quasiconvex.

## 2.6.2 Implementation for the Gini Index

The optimization problem with the classic formulation of the Gini index is,

$$\begin{aligned} \underset{\mathbf{t}}{\text{minimize}} \quad & \frac{1}{2W\sum_{i=1}^{N} w_i \frac{y_i+t_i}{ES_i}} \sum_{i=1}^{N} w_i \sum_{j=1}^{N} w_j \left| \frac{y_i+t_i}{ES_i} - \frac{y_j+t_j}{ES_j} \right| \\ \text{subject to } & 0 \le t_i, \; i = 1, \ldots, N \\ & \sum_{i=1}^{N} t_i \le B \end{aligned} \tag{2.37}$$

Because of the absolute value function in the classic formulation of the Gini index, it is not differentiable at zero. To derive a differentiable reformulation, we introduce the variables $\Delta_{ij}$, which replace the absolute differences in the objective function, and impose linear constraints that require the $\Delta_{ij}$ to be non-negative: $-\Delta_{ij} + \left(\frac{y_i+t_i}{ES_i} - \frac{y_j+t_j}{ES_j}\right) \le 0$ and $-\Delta_{ij} - \left(\frac{y_i+t_i}{ES_i} - \frac{y_j+t_j}{ES_j}\right) \le 0$.[25] To see that this is the case, pick an income difference between any $i$ and $j$ and consider the following scenarios for $\Delta_{ij}$:

1. Let $\left(\frac{y_i+t_i}{ES_i} - \frac{y_j+t_j}{ES_j}\right)$ be non-negative. Then $\Delta_{ij}$ has to be greater than or equal to $\left(\frac{y_i+t_i}{ES_i} - \frac{y_j+t_j}{ES_j}\right)$ and, thus, will always be greater than or equal to $-\left(\frac{y_i+t_i}{ES_i} - \frac{y_j+t_j}{ES_j}\right)$. Accordingly, the $\Delta_{ij}$ will be non-negative.

2. Let $\left(\frac{y_i+t_i}{ES_i} - \frac{y_j+t_j}{ES_j}\right)$ be negative. Then $\Delta_{ij}$ has to be greater than or equal to $-\left(\frac{y_i+t_i}{ES_i} - \frac{y_j+t_j}{ES_j}\right)$ and, thus, will always be greater than $\left(\frac{y_i+t_i}{ES_i} - \frac{y_j+t_j}{ES_j}\right)$. So, the $\Delta_{ij}$ will again be non-negative.

Hence, we can convert (2.37) into an equivalent differentiable optimization problem,

[25] See Boyd and Vandenberghe (2004, p. 294).

$$\begin{aligned}
&\underset{\mathbf{t},\mathbf{\Delta_{ij}}}{\text{minimize}} \quad \frac{1}{2W\sum_{i=1}^{N} w_i \frac{y_i+t_i}{ES_i}} \sum_{i=1}^{N} w_i \sum_{j=1}^{N} w_j \Delta_{ij} \\
&\text{subject to } 0 \le t_i,\ i = 1,\dots,N \\
&\sum_{i=1}^{N} t_i \le B \\
&-\Delta_{ij} + \left(\frac{y_i + t_i}{ES_i} - \frac{y_j + t_j}{ES_j}\right) \le 0,\ \forall i,j = 1,\dots,N \\
&-\Delta_{ij} - \left(\frac{y_i + t_i}{ES_i} - \frac{y_j + t_j}{ES_j}\right) \le 0,\ \forall i,j = 1,\dots,N.
\end{aligned} \tag{2.38}$$

**A Linear-Fractional Problem** The objective function in the modified problem (2.38) has a specific form: an affine function in the numerator and an affine function in the denominator. Problems of this type are called *linear-fractional* (see Boyd and Vandenberghe (2004, p. 151) or originally Charnes and Cooper (1962)). The equivalent problem is,

$$\begin{aligned}
&\underset{\tilde{\mathbf{t}},\tilde{\Delta}_{ij},z}{\text{minimize}} \quad \sum_{i=1}^{N} w_i \sum_{j=1}^{N} w_j \tilde{\Delta}_{ij} \\
&\text{subject to } 0 \le \tilde{t}_i,\ i = 1,\dots,N \\
&\sum_{i=1}^{N} \tilde{t}_i \le zB \\
&-\tilde{\Delta}_{ij} + \left(\frac{zy_i + \tilde{t}_i}{ES_i} - \frac{zy_j + \tilde{t}_j}{ES_j}\right) \le 0,\ \forall i,j = 1,\dots,N \\
&-\tilde{\Delta}_{ij} - \left(\frac{zy_i + \tilde{t}_i}{ES_i} - \frac{zy_j + \tilde{t}_j}{ES_j}\right) \le 0,\ \forall i,j = 1,\dots,N. \\
&2W \sum_{i=1}^{N} w_i \frac{zy_i + \tilde{t}_i}{ES_i} = 1 \\
&z \ge 0,
\end{aligned} \tag{2.39}$$

where we obtain the desired transfer schedule $t_i = \frac{1}{z}\tilde{t}_i\ \forall i$. The major advantage of solving this problem, instead of performing bisection on (2.38), is the immense saving in computational effort: we need only one run of the interior-point algorithm to solve (2.39) instead of several, as in the case of bisection.

**Size of the Problem and Improving Performance** Further, there are two possibilities to reduce the number of variables and constraints for a given dataset: First, we can reduce the size of the dataset if two or more households are of the same type – in terms of their equivalence scale – and have the same income. Then we may simply add up their population weights and optimize the collapsed dataset. Second, we may restrict the number of households that may be recipients of a transfer in the optimization by performing the following procedure: 1. Perform a bottom fill-up procedure for every equivalence-scale-type, where the entire budget at disposal is distributed only among households of this type. 2. Mark those households that are recipients of a positive transfer. 3. Perform the optimization of (2.39) with free transfer variables for the marked households only.

The justification is that, even in the most extreme case, where just one type of household experiences a bottom fill-up, only the marked households can be transfer recipients. Other households of the same type have a weaker effect on the Gini index than the marked households.

Further, to save on memory and reduce computational effort, we provide the following guidelines to enhance the performance of the solver `fmincon` in Matlab:

1. The gradient of the objective function, the gradient of the constraints and the Hessian should be generated as sparse matrices to save memory.

2. The gradient of the objective function and the Hessian are zero everywhere and should be supplied directly by the user.

3. The gradient of the constraints is constant and can be generated before the execution of interior-point algorithm.

4. Parallel computations should be implemented in order to calculate the gradient of the constraints wherever possible.

# 3 Hours risk, wage risk, and life-cycle labor supply

## 3.1 Introduction

What drives the riskiness of earnings? A glance at the recent literature on life-cycle consumption, saving and labor supply suggests an implicit consensus: shocks to wages are the central source of risk. In this paper we re-open this discussion by starting from the natural decomposition of earnings into hours worked and wages. Thus, the main contribution of the paper is a decomposition of earnings risk along these lines: We tailor a structural model of life-cycle labor supply to feature earnings risk from both wage and hours shocks and assess the strength of their contributions to total earnings risk.

Knowing the extent to which hours and wage shocks contribute to total income risk is of general interest as it should inform future modeling decisions. Further, it informs us about the effectiveness of specific policy measures aimed at reducing income risk. For instance, if income risk was driven almost entirely by wage risk, devising policies to reduce the impact of shocks to hours would not be a fruitful endeavor.

In our model individuals face shocks to their productivity of market work, which result in wage shocks, as is standard. Our concise extension of the standard life-cycle model of consumption and labor supply is to model hours shocks as innovations in the disutility of work. Shocks to worked hours are conceptualized in an analogous fashion to wage shocks, namely as shocks to home production. For instance, when pressing needs of family members arise, they increase the opportunity costs of market work sharply. In terms of observed choices, one should then notice a shock to hours of work. Both types of shocks are decomposed into permanent (random walk) and transitory (MA(1)) components. Permanent wage shocks include the obsolescence of human capital, or the acquisition of new skills. Permanent hours shocks might, e.g., be caused by injuries or shocks to home production stemming from family members' needs.

While the extension of the model to include hours shocks is conceptually straightforward, the empirical analysis is complicated substantially. In our setting hours residuals contain hours shocks *in addition* to reactions to wage shocks. Thus, we need to separate the two. The solution is to utilize the covariance of hours and wages to estimate a transmission parameter that quantifies how permanent wage shocks affect the marginal utility of wealth. The parameter is allowed to vary between individuals. The larger this parameter is, the larger is the impact of shocks, and the lower is the degree of insurance against risk.[1] Estimation of the transmission parameter allows us to calculate the Marshallian labor supply elasticity eschewing consumption or asset data, the reliability of which has been hotly debated (Attanasio and Pistaferri, 2016). Thus, the second main contribution of our paper is the estimation of the transmission parameter without using consumption data, offering a new method to estimate the Marshallian elasticity.

We apply our framework to observations on married men in the US from the Panel Study of Income Dynamics (PSID) over the period 1970 to 1997, since at the end of this period the survey frequency turned bi-annual. Our estimate of the Frisch elasticity of labor supply is 0.36 and our estimate of the average Marshallian elasticity is -0.08, which is close to recent estimates in Blundell et al. (2016a); Heathcote et al. (2014a). We find that the standard deviation of permanent wage shocks is larger than the standard deviation of transitory shocks. The same holds for hours shocks, where the standard deviation of permanent shocks is about twice as large as that of transitory shocks. For most samples, the standard deviation of permanent hours shocks is slightly larger than that of permanent wage shocks.

However, the respective impact on earnings risk cannot directly be inferred from this evidence, as the reaction to shocks depends on the degree of insurance. The main exercise with the key components of earnings risk in hand is the variance decomposition. Here we shut down each of the stochastic components except for one in order to quantify their respective contributions to overall earnings risk. At the mean of the transmission parameter, permanent wage shocks explain about 18 percent of cross-sectional earnings growth risk, while permanent hours shocks explain 13 percent. Transitory wage shocks dominate their counterpart in the hours process. While transitory shocks are responsible for the lion's share of cross-section earnings growth risk, only permanent shocks have a substantial impact on life-time earnings. At the mean, a positive permanent hours shock of one standard deviation at age 30 increases life-time earnings by 124 000 Dollar compared to 150 000 Dollar for a permanent wage shock of one standard deviation. Thus, both types of shocks play an important role for life-time earnings.

We also consider a set of alternative models that resemble those applied in the extant literature. Crucially, a model abandoning hours shocks fits the data worse and leads to a substantial overestimation - in absolute terms - of the Marshallian elasticity. Finally, we show how our estimate of the transmission of wage shocks to the marginal utility of wealth can be used to calculate the pass-through of permanent wage shocks to consumption. Calibrating the parameter of relative risk aversion to two, we find that these pass-through parameters for different samples are roughly in line with those estimated in Blundell et al. (2008) using consumption and earnings

[1] When we speak of insurance, we mean precautionary measures like savings and labor supply and not direct insurance bought on the market. The financial market in our model only features a single bond with a fixed interest rate.

data. For the full sample this calculation implies that – on average – an increase in wages by one percent leads to an increase in consumption by .78 percent.

Our paper is related to studies that decompose total income risk into persistent and transitory components, which derive from ideas by Friedman (1957) and Hall (1978) (see MaCurdy, 1982; Abowd and Card, 1989; Meghir and Pistaferri, 2004; Guvenen, 2007; Blundell et al., 2008; Guvenen, 2009; Hryshko, 2012; Heathcote et al., 2014a; Blundell et al., 2016a). Abowd and Card (1989) were pioneers in analyzing the covariance structure of earnings and hours of work. They find that most of the idiosyncratic covariation of earnings and hours of work occurs at fixed wage rates.

In contrast, more recent papers have focused on insurance mechanisms rather than shock sources and restrict the source of risk to stem from wage shocks. In a rich model of family labor supply and consumption, Blundell et al. (2016a) estimate the Marshallian and Frisch consumption and labor supply elasticities using hours, income, asset, and consumption data. Similar to them, we allow for partial insurance of permanent wage shocks, but we depart from their approach by allowing for partially insured hours shocks and using hours and income data alone.

With a similar focus, Heathcote et al. (2014a) analyze the transmission of wage shocks to hours in a setting where shocks are either fully insurable or not insurable at all (island framework). They derive second hours-wage moments from which they identify variances of shocks, the Frisch elasticity of labor supply, and the coefficient of relative risk aversion. Our study differs in two important aspects: First, we assume that shocks are partially insurable as indicated by a consumption insurance parameter similar to Blundell et al. (2008, 2013, 2016a). This parameter may differ between individuals. Second, we introduce hours shocks and estimate their variance. While Heathcote et al. (2014a) allow for initial heterogeneity between agents in the disutility of work, they hold this parameter constant over the life-cycle.

There are some papers that do focus on shock sources more explicitly and for this purpose employ dynamic programming techniques. Low et al. (2010) quantify the contributions of productivity shocks, job losses, and job offers to overall earnings risk. As they point out, in order to disentangle shocks from the reaction to shocks, it is necessary to model consumer behavior. They model labor supply as a discrete decision with fixed hours of work and the possibility of job loss, while we focus on the intensive margin of work hours and allow for hours adjustment and permanent and transitory shocks to hours. Similarly, Kaplan (2012) models consumption and hours of work and allows for involuntary unemployment shocks. These shocks along with nonseparable hours preferences on the extensive and intensive margin aid in the modeling of the declining inequality in hours worked over the first half of the life-cycle.

On the other end of the spectrum, Altonji et al. (2013) allow for both i.i.d. wage and hours shocks in addition to employment and job changes, but they do not work with a fully structural model. They approximate economic decisions of agents in their account of the dynamics of earnings and wage profiles.

The next section outlines the life-cycle model of labor supply and consumption, section 3.3 describes how the magnitudes of shock variances and labor supply elasticities are estimated. In section 3.4 we present results for the parameters of wage and hours processes and the Frisch and Marshallian labor supply elasticities. Then we offer a decomposition of residual earnings variance, which spells out the the importance of wage and hours shocks. Further we calculate the influence of the two shock types on life-time earnings. In section 3.5 we give a characterization of permanent hours shocks, show results when varying the modeling assumptions, discuss the model fit and benchmark our results by relating them to consumption insurance. Section 3.7 concludes.

## 3.2 The Model

Individuals maximize the discounted sum of utilities over the lifetime running from $t_0$ to $T$:[2]

$$\max_{c_t,h_t} E_{t_0}\left[\sum_{t=t_0}^{T} \rho^{t-t_0} \mathrm{v}(c_t, h_t, b_t)\right], \tag{3.1}$$

where $c_t$ and $h_t$ denote annual consumption and hours of work, while $b_t$ contains taste shifters. $\rho$ denotes a discount factor and $\mathrm{v}(\cdot)$ an in-period utility function.

The budget constraint is

$$\frac{a_{t+1}}{(1+r_t)} = (a_t + w_t h_t + N_t - c_t), \tag{3.2}$$

where $a_t$ represents assets, $r_t$ the real interest rate, and $N_t$ non-labor income.

Instantaneous utility takes the additively-separable, constant relative risk aversion (CRRA) form

$$\mathrm{v}_t = \frac{c_t^{1-\vartheta}}{1-\vartheta} - b_t \frac{h_t^{1+\gamma}}{1+\gamma}, \qquad \vartheta \geq 0, \gamma \geq 0. \tag{3.3}$$

We specify $b_t = \exp(\varsigma \Xi_t - \upsilon_t)$. $\Xi_t$ as a set of personal characteristics. $\upsilon_t$ is an idiosyncratic disturbance with mean zero that captures shocks like unexpected changes in home production, e.g., childcare or spousal needs, sickness, and other unexpected changes in the disutility of labor supply.

[2]We omit individual-specific subscripts.

**Wage and hours shock processes** — Wage growth is determined by human capital related variables $X$, which contains $\Delta\Xi$, where $\Delta$ indicates first differences, and an idiosyncratic error $\omega$:

$$\Delta \ln w_t = \alpha X_t + \Delta\omega_t \tag{3.4}$$

Hours shocks ($\upsilon_t$) and wage shocks ($\omega_t$) consist of permanent and transitory components, $p_t$ and $\tau_t$, that follow a random walk and an MA(1)-process respectively. For $x \in \{\upsilon, \omega\}$:

$$\begin{aligned}
&x_t = p_t^x + \tau_t^x \\
&p_t^x = p_{t-1}^x + \zeta_t^x \\
&\tau_t^x = \theta_x \epsilon_{t-1}^x + \epsilon_t^x \\
&\zeta_t^x \sim N\left(0, \sigma^2_{\zeta,x}\right), \quad \epsilon_{it}^x \sim N\left(0, \sigma^2_{\epsilon,x}\right) \\
&E\left[\zeta_t^x \zeta_{t-l}^x\right] = 0, \quad E\left[\epsilon_t^x \epsilon_{t-l}^x\right] = 0 \quad \forall l \in \mathbb{Z}_{\neq 0}
\end{aligned}$$

Permanent and transitory hours and wage shocks are uncorrelated.

**Labor supply** — An approximation of the first order condition with respect to consumption yields the intertemporal labor supply equation (see MaCurdy, 1981; Altonji, 1986, and Appendix 3.8.1):

$$\Delta \ln h_t = \frac{1}{\gamma}\left[-\ln(1 + r_{t-1}) - \ln\rho + \Delta \ln w_t - \varsigma\Delta\Xi_t + \eta_t + \Delta\upsilon_t\right], \tag{3.5}$$

where $\frac{1}{\gamma}$ is the Frisch elasticity of labor supply, $\Xi_t$ contains taste shifters, $\upsilon_t$ is the associated error, and $\eta_t$ is a function of the expectation error in the marginal utility of wealth.[3] $\gamma$ is identified by estimating equation (3.5) using instrumental variables for $\Delta \ln w_t$.

Denote by $\widehat{\Delta x}$ idiosyncratic changes in $x$. The focus of this paper is on idiosyncratic changes in log earnings, $\widehat{\Delta lny_t}$, i.e., earnings changes that result from wage or hours shocks. It is useful to decompose these into temporary and permanent changes, distinguished by the superscripts *per* and *tem* respectively:

$$\widehat{\Delta lny_t} = \widehat{\Delta lnw_t^{per}} + \widehat{\Delta lnw_t^{tem}} + \widehat{\Delta lnh_t^{per}} + \widehat{\Delta lnh_t^{tem}} \tag{3.6}$$

[3] $\eta_t = \frac{\varepsilon_{\lambda t}}{\lambda_t} - \mathcal{O}\left(-1/2(\varepsilon_{\lambda_t}/\lambda_t)^2\right)$, i.e., it contains the expectation error of marginal utility of wealth and the approximation error.

The expressions for temporary and permanent wage changes in terms of shocks are obtained directly from the wage process:

$$\widehat{\Delta lnw_t^{tem}} = \epsilon_t^{\omega} + (\theta_{\omega} - 1)\epsilon_{t-1}^{\omega} - \theta_{\omega}\epsilon_{t-2}^{\omega} \tag{3.7}$$

$$\widehat{\Delta lnw_t^{per}} = \zeta_t^{\omega}. \tag{3.8}$$

Note that in the case of temporary wage changes, everything apart from $\epsilon_t^{\omega}$ is known to the agent at $t-1$. In contrast, the idiosyncratic wage change due to permanent shocks is entirely surprising. Write idiosyncratic hours growth as

$$\widehat{\Delta \ln h_t} = \frac{1}{\gamma}\left[\widehat{\Delta \ln w_t} + \eta_t + \Delta v_t\right]. \tag{3.9}$$

We make the simplifying assumption that transitory shocks do not impact $\eta$.[4] Thus, the expressions for temporary hours changes in terms of shocks follow immediately from the stochastic processes of transitory shock components and the Frisch labor supply equation (3.9):

$$\widehat{\Delta lnh_t^{tem}} = \frac{1}{\gamma}(\epsilon_t^{v} + (\theta_{v} - 1)\epsilon_{t-1}^{v} - \theta_{v}\epsilon_{t-2}^{v} + \epsilon_t^{\omega} + (\theta_{\omega} - 1)\epsilon_{t-1}^{\omega} - \theta_{\omega}\epsilon_{t-2}^{\omega}). \tag{3.10}$$

In our model the expectation error is a linear function of unexpected permanent changes to income. This is in line with models that approximate the life-time budget constraint like Blundell et al. (2016a). The expression is

$$\eta_t = -\phi_t^{\lambda}\left(\widehat{\Delta lnw_t^{per}} + \widehat{\Delta lnh_t^{per}}\right), \quad \ln\phi_t^{\lambda} \sim N\left(\mu_{\phi}, \sigma_{\phi}^2\right) \tag{3.11}$$

The parameter $\phi_t^{\lambda}$ measures how shocks to income transmit to $\eta_t$, which is in utility units. It is a measure of consumption insurance; perfectly insured individuals do not adjust their consumption as a response to a permanent shock and thus their marginal utility of consumption is unchanged. For instance, for individuals who have accumulated substantial assets, remaining life-time earnings only play a relatively small role in total life-time income. These individuals do not adjust their consumption by much in response to a wage shock. Blundell et al. (2016a) study in detail what governs the transmission of shocks to consumption and hours worked. The parameter is individual-specific since it depends - among other things

[4] This holds when one assumes, as we do, that transitory shocks do not affect the marginal utility of wealth. See the insurance parameters for transitory shocks in Blundell et al. (2008) for justification of that assumption.

- on the amount of assets currently held in relation to the total stock of human wealth (see Blundell et al., 2016a, p. 396, for the related consumption-insurance parameter). In the case of no insurance, a one percentage change in income leads to a one percentage change in consumption and $\phi_t^\lambda = \vartheta$.[5] In the case of full consumption insurance, $\phi_t^\lambda = 0$ and income changes do not translate into changes in consumption at all. In general it seems reasonable to expect that there is at least some degree of insurance, such that the estimate of $E[\phi_t^\lambda]$ is a lower bound for the average degree of relative risk aversion.

Positive income shocks lead to a decrease in the marginal utility of wealth, therefore $\phi_t^\lambda$ is positive and should follow a distribution with no support on negative values. Hence, we estimate the model under the assumption that $\phi_t^\lambda$ is lognormally distributed. An equivalent transmission parameter for income shocks to the marginal utility of wealth is estimated in Alan et al. (forthcoming).

Plugging equation (3.8) into (3.11) and subsequently (3.11) into (3.9) and solving for $\widehat{\Delta lnh_t}$ yields the expression for idiosyncratic permanent changes in hours of work:

$$\widehat{\Delta lnh_t^{per}} = \frac{1-\phi_t^\lambda}{\gamma+\phi_t^\lambda}\zeta_t^\omega + \frac{1}{\gamma+\phi_t^\lambda}\zeta_t^\upsilon \tag{3.12}$$

The term $\kappa = \frac{1-\phi_t^\lambda}{\gamma+\phi_t^\lambda}$ gives the uncompensated reaction to a permanent wage change, the Marshallian labor supply elasticity. If $\phi_t^\lambda = 0$, the case of perfect insurance, the Marshallian collapses to the Frisch elasticity, the reaction to a transitory shock. The transmission coefficient for a permanent hours shock, $\frac{1}{\gamma+\phi_t^\lambda}$ has the same property. The higher $\phi_t^\lambda$, the more are hours shocks cushioned. A further property of the hours shock transmission coefficient is that it equals the Hicksian elasticity for a permanent *wage* shock. This peculiarity arises from the way $b$ affects marginal trade-offs: in a static version of the model with no unearned income the marginal optimality condition (MRS = price ratio) is given by $b\frac{h^\gamma}{c^{-\vartheta}} = w$.[6] When we hold the level of consumption constant, a change in $w$ and a change in $b$ cause the same type of adjustment in $h$, although differently signed. However, when we let consumption adjust and derive the Marshallian demand for $h$, we find that $\ln h = \frac{1}{\gamma+\vartheta}[(1-\vartheta)\ln w - \ln b]$. Through the effect on the budget constraint, a change in $w$ causes an income effect of size $\frac{-\vartheta}{\gamma+\vartheta}$ in elasticity form. $b$ does not affect the budget constraint and therefore does not have the same effect.

[5] This can be seen by taking logs of the first derivative of equation (4.2) with respect to $c_t$. $\phi_t^\lambda$ might exceed $\vartheta$ if shocks to life-time income not captured by the model are positively correlated with permanent hours and wage shocks.

[6] In the static model, the consumer maximizes utility, $V = \frac{c^{1-\vartheta}}{1-\vartheta} - b\frac{h^{1+\gamma}}{1+\gamma}$, subject to the budget constraint, $c = wh$.

**Consumption** — The equation for consumption growth can be obtained analogously to equation (3.5) (see, e.g., Altonji, 1986):

$$\Delta \ln c_t = \frac{1}{\vartheta} [\ln(1 + r_{t-1}) + \ln \rho - \eta_t] \tag{3.13}$$

Thus income shocks are directly related to consumption growth by $-\eta_t/\vartheta$. The direct estimation of equation (3.13) using consumption data is beyond the scope of this study. Nonetheless, we benchmark our results by calculating the reaction of consumption to wage shocks by calibrating $\vartheta$.

Figures 3.1 and 3.2 show how each type of permanent shock propagates through the various quantities of interest. The major distinction for the two types is that wage shocks do not only have a direct effect on income, but also affect the choice of hours through the Marshallian elasticity.

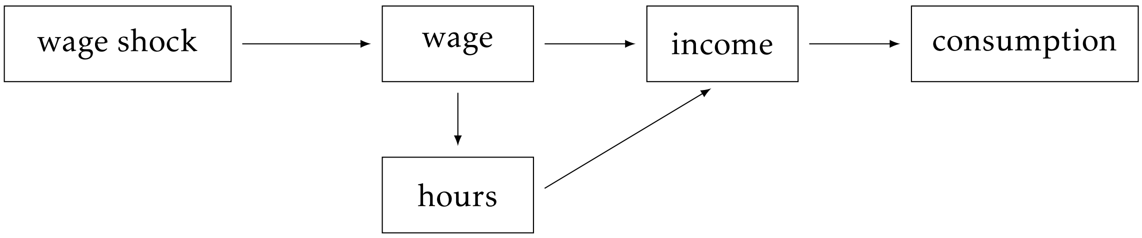



**Figure 3.1:** Transmission of Permanent Wage Shock

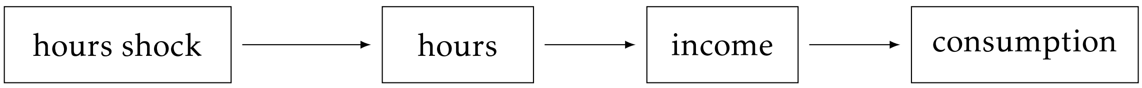



**Figure 3.2:** Transmission of Permanent Hours Shock

## 3.3 Recovering Labor Supply Elasticities, Wage Shocks, and Hours Shocks

In this section we detail how the labor supply elasticities as well as the standard deviations of permanent and transitory components of wage shocks, $\omega_t$, and hours shocks, $v_t$, are recovered in estimation. The estimation is carried out in three stages. First, we use OLS to obtain residuals from the wage equation and IV to obtain residuals from the hours equation as well as an estimate for the Frisch labor supply elasticity. Second, we estimate the variances of transitory and permanent shocks to wages by fitting three theoretical autocovariance moments of the wage residual to the data. Third, we estimate hours shock variances by fitting the corresponding autocovariance moments for the hours residual as well as the covariance of hours and wage residuals to the data.

We state the variance-covariance moments with measurement error. Measurement error is modeled as having no intertemporal correlation, but we do allow for correlation between the types of measurement error. Denote by

$$\ln \tilde{x}_t = \ln x_t + me_{x,t} \tag{3.14}$$

the observed value for the log of variable $x$, where $me_{x,t}$ is the mean zero measurement error with variance $\sigma^2_{me,x}$. The variances encountered in the moment conditions are $\sigma^2_{me,h}$, $\sigma^2_{me,w}$ and $\sigma^2_{me,h,w}$, which are the variance of measurement error in log hours, log wages and their covariance respectively. We state the precise way we calibrate the magnitudes of the measurement error variances at the end of the section.

**Frisch elasticity, hours residuals, and wage residuals** — The augmented empirical Frisch labor supply equation containing measurement errors is

$$\begin{aligned}\Delta \ln \tilde{h}_t =& \frac{1}{\gamma}\left[-\ln(1+r_{t-1}) - \ln\rho + \Delta \ln \tilde{w}_t - \varsigma \Delta \Xi_t + \eta_t + \Delta v_t\right] \\ & - \frac{1}{\gamma}\Delta me_{w,t} + \Delta me_{h,t}.\end{aligned} \tag{3.15}$$

The error term of equation (3.15) is correlated with differenced log wages because wage shocks impact the marginal utility of wealth and because of measurement error. To obtain the Frisch elasticity from equation (3.5) we use human capital related instrumental variables for $\Delta \ln \tilde{w}_t$ following MaCurdy (1981). Hours residuals $(\eta + \Delta\tilde{v}_t)/\gamma = (\eta + \Delta v_t - \Delta me_{w,t})/\gamma + \Delta me_{h,t}$ are obtained by running IV on differenced log hours using differenced year, child, disability and state dummies as covariates. The instruments for the differenced log wage are interactions of age and years of education, i.e., age, education, education$^2$, age $\times$ education, age $\times$ education$^2$, age$^2$ $\times$ education and age$^2$ $\times$ education$^2$. Wage residuals $\Delta\tilde{\omega}_t = \Delta\omega_t + \Delta me_{w,t}$ are obtained by estimating equation (3.4) augmented by an error term, i.e. regressing differenced log wages on the same exogenous regressors as in the hours equation as well as the excluded instruments.

**Wage shocks** — After recovering $\Delta\tilde{\omega}_t$, all parameters of the autoregressive process, $(\theta, \sigma^2_{\epsilon,\omega}, \sigma^2_{\zeta,\omega})$, are identified through combinations of the autocovariance moments. Label the $k$-th autocovariance moment by $\Lambda_{\tilde{\omega},k}$:

$$\Lambda_{\tilde{\omega},0} = E\left[(\Delta\tilde{\omega}_t)^2\right] = 2\left(1 - \theta_\omega + \theta_\omega^2\right)\sigma^2_{\epsilon,\omega} + \sigma^2_{\zeta,\omega} + 2\sigma^2_{me,w} \tag{3.16}$$

$$\Lambda_{\tilde{\omega},1} = E\left[\Delta\tilde{\omega}_t \Delta\tilde{\omega}_{t-1}\right] = -\left(\theta_\omega - 1\right)^2\sigma^2_{\epsilon,\omega} - \sigma^2_{me,w} \tag{3.17}$$

$$\Lambda_{\tilde{\omega},2} = E\left[\Delta\tilde{\omega}_t \Delta\tilde{\omega}_{t-2}\right] = -\theta_\omega\sigma^2_{\epsilon,\omega} \tag{3.18}$$

Net of $\sigma^2_{me,w}$, dividing $\Lambda_{\tilde{\omega},2}$ by $\Lambda_{\tilde{\omega},1}$ identifies the parameter $\theta_\omega$. Successively, the variance of the transitory shock is identified from $\Lambda_{\tilde{\omega},1}$ and the variance of the permanent shock from $\Lambda_{\tilde{\omega},0}$ (see Hryshko, 2012).

**Hours shocks** — The residual obtained from estimating the labor supply equation contains both hours shocks $v_t$ and a function of expectation errors, $\eta_t$. The variance of the residual of the labor supply equation contains both the mean and the variance of $\frac{\phi_t^\lambda}{\gamma+\phi_t^\lambda}$ and the variance of the permanent hours shocks, which causes an identification problem. The procedure for wage moments does not carry over. We use the contemporaneous covariance of hours and wage residuals to identify the mean of $1 - \frac{\gamma}{\gamma+\phi_t^\lambda}$, which is equivalent to $\frac{\phi_t^\lambda}{\gamma+\phi_t^\lambda}$. To arrive at the theoretical variance moment, use equations (3.8), (3.11), (3.12), and (3.15) to find the following expression for the hours residual

$$\begin{aligned} \frac{\eta + \Delta\tilde{v}_t}{\gamma} = & \frac{1}{\gamma}\Bigg[-\left(1 - \gamma\frac{1}{\gamma+\phi_t^\lambda}\right)\zeta_t^v - (1+\gamma)\left(1 - \gamma\frac{1}{\gamma+\phi_t^\lambda}\right)\zeta_t^\omega \\ & + \zeta_t^v + \epsilon_t^v + (\theta v - 1)\epsilon_{t-1}^v - \theta v \epsilon_{t-2}^v\Bigg] \\ & - \frac{1}{\gamma}\Delta me_{w,t} + \Delta me_{h,t}, \end{aligned} \tag{3.19}$$

where the first line on the right hand side equals $\eta_t/\gamma$. The variance can be written as

$$\Lambda_{\tilde{v},0} = E\left[\left(\frac{\eta_t + \Delta\tilde{v}_t}{\gamma}\right)^2\right] = \frac{1}{\gamma^2}\Big(\sigma^2_{\zeta,v} + 2\left(\theta_v^2 - \theta_v + 1\right)\sigma^2_{\epsilon,v} + (1+\gamma)^2(1 - 2\gamma\mathcal{M}_1 + \gamma^2\mathcal{M}_2)\sigma_{\zeta,\omega}\Big) + \mathcal{M}_2\sigma^2_{\zeta,v} + 2\sigma^2_{me,h} + \frac{2\sigma^2_{me,w}}{\gamma^2} - \frac{4\sigma^2_{me,h,w}}{\gamma}, \tag{3.20}$$

where $\mathcal{M}_1$ and $\mathcal{M}_2$ denote the first and second non-central moments of $1/(\gamma+\phi_t^\lambda)$, the random component in $1 - \frac{\gamma}{\gamma+\phi_t^\lambda}$. As no analytical expression exists for these moments, we find them numerically as described in Appendix 3.8.2.

The autocovariance moments of the hours residual $\Lambda_{\tilde{v},1}$ and $\Lambda_{\tilde{v},2}$ are analogous to their wage process counterparts:

$$\Lambda_{\tilde{v},1} = E\left[\frac{(\eta_t + \Delta\tilde{v}_t)(\eta_{t-1} + \Delta\tilde{v}_{t-1})}{\gamma^2}\right] = -\frac{(\theta_v - 1)^2\sigma^2_{\epsilon,v}}{\gamma^2} - \sigma^2_{me,h} - \frac{\sigma^2_{me,w}}{\gamma^2} + \frac{2\sigma^2_{me,h,w}}{\gamma} \tag{3.21}$$

$$\Lambda_{\tilde{v},2} = E\left[\frac{(\eta_t + \Delta\tilde{v}_t)(\eta_{t-2} + \Delta\tilde{v}_{t-2})}{\gamma^2}\right] = -\frac{\theta_v\sigma^2_{\epsilon,v}}{\gamma^2} \tag{3.22}$$

To estimate the variance of permanent hours shocks, we need to identify $\mathcal{M}_1$ using the contemporaneous covariance of hours and wage residuals:

$$\Lambda_{\tilde{\omega},\tilde{v},0} = E\left[\frac{(\eta_t + \Delta\tilde{v}_t)\Delta\tilde{\omega}_t}{\gamma}\right] = \frac{(\gamma+1)(\gamma\mathcal{M}_1 - 1)\sigma^2_{\zeta,\omega}}{\gamma} - \frac{2\sigma^2_{me,w}}{\gamma} + 2\sigma^2_{me,h,w}. \tag{3.23}$$

This covariance is larger in absolute value the smaller $\gamma$, which denotes resistance to intertemporal substitution of hours of work, and the larger $E[\phi_t^\lambda]$. When $\gamma$ goes to infinity, the effect of permanent wage shocks on income is only mechanical and not through labor supply reactions.

$\mathcal{M}_1$ and $\mathcal{M}_2$ contain both $\mu_\phi$ and $\sigma_\phi$. Theoretically, $\sigma_\phi$ is identified through the cokurtosis moments of the wage and hours residuals. However, cokurtosis moments are very noisy, hence $\sigma_\phi$ can only be estimated to a reasonable degree of reliability when using several million observations.[7] Therefore, we apply the alternative estimation strategy of calibrating $\sigma_\phi$ to 1.023 based on results in Alan et al. (forthcoming). Using this calibration, once $\mathcal{M}_1$ is estimated, the mean of $\phi_t^\lambda$, $E[\phi_t^\lambda] = e^{\mu_\phi + \frac{\sigma_\phi^2}{2}}$, can be recovered. In section 3.5 we show the robustness of our results to alternative values of this parameter.

**Marshallian elasticity** — The term multiplied with $\sigma_{\zeta,\omega}^2$ in equation (3.23) can be rewritten as $E\left[\frac{1-\phi_t^\lambda}{\gamma+\phi_t^\lambda}\right] - \frac{1}{\gamma}$, the average Marshallian minus the Frisch elasticity of labor supply. Thus, the Marshallian can directly be calculated from the parameter estimates. The Marshallian elasticity is the uncompensated reaction to a permanent wage shock.[8]

The Marshallian elasticity is the relevant concept for the evaluation of tax reforms, which are best described as unanticipated, permanent shifts in net-of-tax wages (Blundell and Macurdy, 1999). Using similar considerations as in our study, the Marshallian elasticity has been estimated using the covariance of earnings and wages, household sharing parameters, and the ratio of assets to total (human and non-human) wealth in Blundell et al. (2016a, eq. A2.23). Heathcote et al. (2014a) use the covariance of hours and consumption as well as of wages and consumption to estimate the Marshall elasticity. In contrast, we rely only on hours and wage data.[9]

**Estimation** — We estimate the parameters of the autoregressive processes and the transition of wage shocks by fitting the theoretical moments $\{\Lambda_{\omega,k}, \Lambda_{\upsilon,k}, \Lambda_{\omega,\upsilon,k}\}$ to those of the data. The vector of parameters, denoted $\Theta$, is estimated using the method of minimum distance and an identity matrix serves as the weighting matrix.[10] The distance function is given by

$$DF(\Theta) = [m(\Theta) - m^d]' I [m(\Theta) - m^d], \tag{3.24}$$

where $m(\Theta)$ indicates theoretical moments and $m^d$ empirical moments. An outline of the entire estimation procedure is detailed in Hryshko (2012). Standard errors are obtained by the block bootstrap with 200 replicates.

[7] Simulations evidencing this are available upon request from the authors.

[8] See Keane (2011, p.1008) for a discussion of why reactions to permanent shocks equal the Marshallian elasticity.

[9] Heathcote et al. (2014a) also estimate a variant that does not rely on consumption data. Their approach differs because their specific island framework implies that the marginal utility of wealth is constant across individuals in the same age-year cell.

[10] Altonji and Segal (1996) show that the identity weighting matrix is preferable for the estimation of autocovariance structures using micro data.

**The data** — We use annual data from the PSID for the survey years 1970 to 1997. After this point in time the PSID is bi-annual. Annual hours of work and earnings refer to the previous calendar year. Hours are constructed by the PSID by multiplying actual weeks worked with usual hours of work per week. Earnings include wages and salaries from all jobs and include tips, bonuses, and overtime. We calculate the hourly wage by dividing earnings through hours of work. As hours and earnings are measured with error, a negative correlation between measured hours and wages is induced, which we correct for as described in the final part of this section. Our sample consists of working, married males aged 28 to 60, who are the primary earners of their respective households. We focus on this group because the extensive labor supply margin plays a small role in their labor supply behavior and in order to allow comparisons to the extant literature. Table 3.1 shows summary statistics of the main sample. Monetary variables are adjusted to 2005 real values using the CPI-U.

**Table 3.1:** Descriptives

| | mean | s.d. |
|---|---|---|
| Age | 41.35 | 8.66 |
| Annual hours of work | 2220.28 | 530.11 |
| Hourly wage | 26.86 | 22.83 |
| Number of kids in household | 1.64 | 1.39 |
| $N$ | 46340 | |

*Note:* Own calculation based on the PSID. Monetary values inflated to 2005 real dollars.

**Measurement errors** — Wages and hours are measured with error, which we assume to be classical in levels, i.e. i.i.d. over time and individuals. As wages are calculated by dividing annual earnings through hours, the wage measure suffers from division bias, i.e., the measurement errors of wages and hours are negatively correlated. Following Meghir and Pistaferri (2004) and Blundell et al. (2016a) we use estimates from the validation study by Bound et al. (1994) for the signal-to-noise ratios of wages, hours, and earnings. As in Blundell et al. (2016a), we assume that the variance of the measurement error of hours is $\sigma^2_{me,h} = 0.23var(h)$, the variance of the measurement error of wages is $\sigma^2_{me,w} = 0.13var(w)$, and the variance of the measurement error of earnings is $\sigma^2_{me,y} = 0.04var(y)$, where $var(h)$, $var(w)$, and $var(y)$ denote the variances of the levels of log wages, log hours, and log earnings. The covariance of the measurement errors of log wages and hours is given by $\sigma^2_{me,h,w} = (\sigma^2_{me,y} - \sigma^2_{me,w} - \sigma^2_{me,h})/2$.

## 3.4 Main Results

**Standard deviations of wage shocks** — Table 3.2 reports the standard deviations of permanent and transitory wage shocks as well as the parameter of transitory shock persistence. Throughout the paper we show results for the full sample as well as three sub-samples, which might differ with respect to their labor supply behavior and exposure to shocks: a sample excluding young workers, one consisting of individuals with more than high school education and a sample of individuals without children less than seven years old in the household. First, while the magnitude of the standard deviation of permanent shocks is similar in the four samples, excluding young workers leads to a decline of this figure. This is in line with the finding of a u-shaped pattern of permanent wage shocks over the life-cycle as in Blundell et al. (2016a) and Meghir and Pistaferri (2004). Second, for all samples the standard deviation of transitory shocks is smaller than that of permanent shocks. Third, the highly educated face a substantially lower standard deviation of transitory shocks than the full sample. For those without children less than seven years permanent and transitory shocks are slightly lower than for the full sample.

**Table 3.2:** Wage Variances

| | I<br>Full sample | II<br>Age≥40 | III<br>High educ | IV<br>No children <7 |
|---|---|---|---|---|
| $\theta_\omega$ | 0.2701 | 0.3450 | 0.2737 | 0.1832 |
| | (0.0090) | (0.0241) | (0.0325) | (0.0212) |
| $\sigma_{\epsilon,\omega}$ | 0.1337 | 0.1382 | 0.0772 | 0.1166 |
| | (0.0017) | (0.0030) | (0.0015) | (0.0025) |
| $\sigma_{\zeta,\omega}$ | 0.1770 | 0.1554 | 0.1765 | 0.1639 |
| | (0.0009) | (0.0014) | (0.0007) | (0.0011) |
| $N$ | 46340 | 20607 | 19831 | 24547 |

*Note:* Own calculation based on the PSID. Bootstrapped standard errors in parentheses.

**Standard deviations of hours shocks** — The first three rows in Table 3.3 show the parameters of the process of shocks to the disutility of work. For ease of interpretation, the parameters are reported as they enter the hours equation (3.5), i.e. multiplied with $1/\gamma$. The size of these estimates are generally comparable to those of the wage process. The standard deviation of the permanent hours shocks drops when we consider the highly educated. This group also experiences less pronounced transitory shocks in comparison to the main sample. Otherwise, permanent shocks to the disutility of work are of a fairly consistent size across the samples. For all three subsamples the standard deviation of transitory shocks is lower than in the full sample. The magnitude of the standard deviation of permanent hours shocks is a first indicator that these shocks play a significant role for overall earnings risk.

However, as described in section 3.2, the effect of innovations in the marginal disutility of work on earnings depends on the degree of consumption insurance. At the end of this section we discuss the importance of wage shocks and hours shocks to overall earnings risk.

**Table 3.3:** Hours variances and labor supply elasticity

| | I | II | III | IV |
|---|---|---|---|---|
| | Full sample | Age>=40 | High educ | No children <7 |
| $\theta_v/\gamma$ | 0.1515 | 0.4013 | 0.1140 | 0.2463 |
| | (0.0039) | (0.0059) | (0.0065) | (0.0053) |
| $\sigma_{\epsilon,v}/\gamma$ | 0.1114 | 0.0730 | 0.0709 | 0.0790 |
| | (0.0011) | (0.0012) | (0.0014) | (0.0012) |
| $\sigma_{\zeta,v}/\gamma$ | 0.1990 | 0.2102 | 0.1648 | 0.1914 |
| | (0.0010) | (0.0327) | (0.0010) | (0.0058) |
| $1/\gamma$ | 0.3614 | 0.4020 | 0.2851 | 0.3148 |
| | (0.0856) | (0.3778) | (0.0975) | (0.1080) |
| $E[\phi_t^\lambda]$ | 1.8918 | 1.4084 | 0.5668 | 0.9565 |
| | (0.1117) | (4.0920) | (0.0436) | (1.2774) |
| $E[\kappa]$ | -0.0767 | -0.0023 | 0.1302 | 0.0631 |
| | (0.0105) | (0.0254) | (0.0078) | (0.0164) |

*Note:* Own calculation based on the PSID. Clustered standard errors for $1/\gamma$, bootstrapped standard errors for other coefficients in parentheses.

**Frisch elasticity** — Row 4 in Table 3.3 reports the estimates of the Frisch elasticity. In contrast to the most closely related papers (Blundell et al., 2016a; Heathcote et al., 2014a), we obtain the Frisch elasticity directly through IV estimation and not through covariance moments.[11] The estimated Frisch elasticity for the main sample is 0.36, which is in line with the literature (Keane, 2011). The point estimate of the Frisch elasticity increases when excluding younger individuals. This result is expected as younger individuals could be less willing to reduce their hours of work in the case of a decrease in the hourly wage because the accumulation of human capital impacts their opportunity costs of time (Imai and Keane, 2004). Similarly, human capital considerations are more important for the highly educated, where the Frisch elasticity is lower than that of the main sample. The estimate for those without young children is fairly close to that of the main sample, but slightly smaller.

**Transmission parameter** — Row 5 in Table 3.3 shows the estimated mean of the parameter that measures the transmission of shocks to the marginal utility of wealth, $E[\phi_t^\lambda]$. As this parameter falls, individuals become more insured against shocks. A

[11] Table 3.10 in the appendix additionally displays the Kleibergen and Paap (2006) F statistic, indicating that only sample II might suffer from substantial weak instrument bias and should therefore be interpreted with caution.

value of zero indicates that permanent shocks do not impact the marginal utility of wealth at all. We expect households with larger accumulation of assets relative to human wealth to exhibit smaller values of $E[\phi_t^\lambda]$. The point estimate drops only slightly relative to the full sample, when excluding young workers, but is substantially smaller when focusing on those without young children and especially on the highly educated. It is not surprising that more well-educated individuals are also insured to greater extent.

**Marshallian elasticity** —Table 3.3 reports the average of the Marshallian elasticity defined in equation (3.12) as the reaction to a permanent wage shock.[12] The wealth effect outweighs the substitution effect, leading to a negative (but small) estimate for the main sample, in line with the recent literature.[13] The negative Marshallian implies that hours move in the opposite direction of wages and thus function as a consumption smoothing device. When excluding younger workers, the estimate edges even closer to zero, signifying no long-term adjustment in hours for older workers. The smaller the average transmission parameter, the closer is the average Marshallian to the Frisch elasticity because the shock has a smaller effect on the marginal utility of wealth. The smaller wealth effect for older workers is expected because for individuals close to the end of their life-cycle transitory and permanent shocks have the same effect on the marginal utility of wealth. In the sample without young children in the household the estimate is positive, making the substitution effect the dominant force as the average transmission parameter is relatively small for this sample. The highly educated show the highest positive Marshallian elasticity driven by their very small transmission parameter.

**Importance of hours and wage shocks** — Using our estimates for the variances of hours and wage shocks allows us to quantify their contribution to the cross-sectional variance of overall earnings growth. The stochastic component of earnings net of measurement error is given by the sum of hours and wage residuals plus the Frisch reactions to idiosyncratic wage changes, which we have removed from the hours residual by detrending with wages, see equation (3.15). The variance of stochastic earnings growth is thus given by

[12]We calculate $\kappa$ as the numerical expectation $E\left[\frac{1-\phi_t^\lambda}{\gamma+\phi_t^\lambda}\right]$, which does not simply reduce to $\frac{1-E[\phi_t^\lambda]}{\gamma+E[\phi_t^\lambda]}$.

[13]Blundell et al. (2016a) and Heathcote et al. (2014a) find Marshallian elasticities for men of -0.08 and -0.16 respectively. The latter number is obtained by inserting the obtained parameter estimates in the formula for the labor supply reaction to an uninsurable shock. Altonji et al. (2013) report a coefficient that determines "the response to a relatively permanent wage change" of -0.08.

$$E\left[\left(\widehat{\Delta \ln y}\right)^2\right] = \frac{1}{\gamma^2}\Bigg[\gamma^2 \mathcal{M}_2\left(\sigma^2_{\zeta,v} + (\gamma+1)^2\sigma^2_{\zeta,\omega}\right) + 2(\gamma+1)^2((\theta_\omega - 1)\theta_\omega + 1)\sigma^2_{\epsilon,u} + 2((\theta_v - 1)\theta_v + 1)\sigma^2_{\epsilon,v}\Bigg] \tag{3.25}$$

$\mathcal{M}_2$ is obtained numerically using the estimates of the underlying parameters. Note that $\mathcal{M}_2$ depends on the variance of the transmission parameter $\phi^\lambda_t$, which is known to individuals. Additionally the realization of transitory components of wage and hours growth are partially known in advance, see equation (3.7). Therefore this overall variance is not a pure measure of risk. The first row of Table 3.4 shows the cross-sectional variance of the stochastic component of earnings growth for the four samples. Rows 2-5 show the contributions of shock components, i.e., the variance of earnings growth when the variances of all other shock components are set to zero. Earnings growth variances for the highly educated and for the sample excluding households with young children are substantially lower than that of the main sample. For all samples, except that of the highly educated, transitory wage shocks play the largest role in explaining earnings growth variance. The main source of the variance in income growth for the highly educated are permanent wage shocks. Their contribution is roughly double that of permanent hours shocks. For all samples the share of the cross-sectional hours variance due to permanent hours shocks is at minimum more than half the share explained by permanent wage shocks. For the sample excluding younger workers, the shares of earnings growth variance explained by permanent hours and wage shocks are of similar magnitude. For older workers the seniority dynamics of their wage become less relevant, which should drive down the contribution of permanent wage shocks to the variance of earnings. A similar dynamic does not necessarily follow for shocks to home production. The highly educated experience the bulk of their variation through permanent wage shocks, while all other sources of the variance lose relevance compared to the main sample. The earnings and wage profile of the highly educated is much steeper; human capital can still increase substantially through labor market experience (see Imai and Keane, 2004). These potential increases in productivity during working life leave room for a greater variation across individuals and over time.

**Table 3.4:** Decomposition of variance of earnings growth

| | I<br>Full sample | II<br>Age>=40 | III<br>High educ | IV<br>No children <7 |
|---|---|---|---|---|
| $V\left(\widehat{\Delta \ln y}\right)$ | 0.1464 | 0.1293 | 0.0857 | 0.1071 |
| $\sigma_{\epsilon,\omega}$ | 0.0532 | 0.0581 | 0.0158 | 0.0400 |
| $\sigma_{\zeta,\omega}$ | 0.0426 | 0.0326 | 0.0398 | 0.0319 |
| $\sigma_{\epsilon,\upsilon}$ | 0.0216 | 0.0082 | 0.0090 | 0.0100 |
| $\sigma_{\zeta,\upsilon}$ | 0.0290 | 0.0304 | 0.0210 | 0.0252 |

*Note:* Variance of $\widehat{\Delta \ln y}$ when all other shock variances are set to zero. First line: actual variance of $\widehat{\Delta \ln y}$ given by equation (3.25). Own calculation based on the PSID.

When evaluating *risk* of idiosyncratic earnings growth instead of its cross-sectional*variance*, everything that is known to an agent at $t-1$ must be excluded from (3.25) and $\phi_t^\lambda$ must be treated as non-stochastic. Denote by $I_{t-1}$ the agent's information set at $t-1$. At that point in time the agent knows $\phi_t^\lambda$ and the realization of shocks in $t-1$. Thus, $E\left[\widehat{\Delta \ln y_t}|I_{t-1}\right]$ includes the transitory components from the previous two periods. The resulting equation for earnings risk conditional on the information set in $t-1$ is

$$E\left[\left(\widehat{\Delta \ln y_t} - E\left[\widehat{\Delta \ln y_t}|I_{t-1}\right]\right)^2 \middle| I_{t-1}\right] = \frac{\sigma^2_{\zeta,\upsilon} + (\gamma+1)^2\sigma^2_{\zeta,\omega}}{(\gamma+\phi_t^\lambda)^2} + \frac{1}{\gamma^2}\left(\sigma^2_{\epsilon,\upsilon} + (\gamma+1)^2\sigma^2_{\epsilon,\omega}\right) \tag{3.26}$$

In Table 3.5 $\phi_t^\lambda$ is set to the sample mean. Thus, the cross-sectional variance of unexpected earnings growth can be interpreted as earnings risk for a typical individual in each sample. A comparison of the first lines in Tables 3.4 and 3.5 shows that for the full sample earnings growth risk at the mean is about 55% of the cross-sectional idiosyncratic earnings growth variance. The degree to which the size of contributions of transitory shocks decreases depends on the parameter $\theta$ of the respective MA(1) process. The smaller $\theta$, the larger is the share of risk in the total variance of idiosyncratic earnings growth. The importance of permanent shocks decreases relative to Table 3.4 because $\phi_t^\lambda$ is non-stochastic. The importance of permanent wage shocks decreases for all samples and to a large degree for the first two samples, where the Marshallian labor supply elasticity is negative at the mean of $\phi_t^\lambda$. Similarly, the importance of permanent hours shocks for total earnings risk is much smaller for these two samples as $\phi_t^\lambda$ cushions the reaction to innovations in the marginal disutility of work. For all samples, permanent hours shocks explain at least 17 percent of earnings risk. Nonetheless, wage risk is more important in all samples, although for older individuals the magnitudes are very close, as they were for the variance.

**Table 3.5:** Decomposition of earnings risk at mean

| | I | II | III | IV |
|---|---|---|---|---|
| | Full sample | Age>=40 | High educ | No children <7 |
| $V(\Delta \ln y)$ | 0.08 | 0.0803 | 0.0731 | 0.0788 |
| $\sigma_{\epsilon,\omega}$ | 0.0331 | 0.0375 | 0.0098 | 0.0235 |
| $\sigma_{\zeta,\omega}$ | 0.0205 | 0.0194 | 0.0381 | 0.0275 |
| $\sigma_{\epsilon,\upsilon}$ | 0.0124 | 0.0054 | 0.005 | 0.0062 |
| $\sigma_{\zeta,\upsilon}$ | 0.014 | 0.018 | 0.0201 | 0.0217 |

*Note:* Earnings growth risk with $\phi_t^\lambda$ set to sample mean. First line: Total earnings risk at mean given by equation (3.26).
Own calculation based on the PSID.

While transitory shocks are an important driver of cross-sectional earnings growth variance, only permanent shocks have a large impact on the present value of life-time earnings. A back-of-the-envelope calculation[14] using the average coefficients of the main sample shows that for an individual aged 30 and retiring at 65, a typical positive permanent wage shock of one standard deviation increases present value life-time earnings by about 150 000 Dollar, while a typical positive permanent hours shock increases life-time income by 124 000 Dollar.[15] Typical permanent wage and hours shocks at age 50 for the same individual increase life-time income by 92 000 and 76 000 Dollar respectively. Thus, both types of permanent shocks have a substantial impact on life-time earnings.

The impact of hours shocks depends largely on the degree of insurance. In the benchmark case of full insurance with $\phi_t^\lambda = 0$ individuals adjust their hours of work much more in response to a shock to the disutility of work. Then the impact of a typical permanent wage shock at age 30 is 252 000 Dollar because in this case the Frisch labor supply reaction amplifies the wage shock. The analogous impact of a typical hours shock is 208 000 Dollar. Clearly, the impact of a permanent shock on life-time income varies widely between individuals.

[14] The impact of a typical permanent wage shock at the mean of $\phi_t^\lambda$ is given by the geometric series $y\left(1+\left(1-\left(E[\phi_t^\lambda]\right)/\left(\gamma+E[\phi_t^\lambda]\right)\right)\sigma_{\zeta,\omega}\right)(1-1/r^{65-age})/(1-1/r)$ and the impact of a typical permanent hours shock is $y\,(1/(\gamma+E[\phi\lambda_t]))\,\sigma_{\zeta,\upsilon}(1-1/r^{65-age})/(1-1/r)$, where annual earnings $y$ are set to the sample mean of 57 267 Dollar and the real interest rate $r$ is 1.0448 based on World Bank figures for our period. This abstracts from deterministic earnings growth, i.e. it makes the simplifying assumption that earnings would remain constant without shocks.

[15] Note that the ratio of the impacts of typical permanent hours and wage shocks on lifetime earnings equals the square root of the corresponding ratio of contributions to earnings risk reported in Table 3.5.

## 3.5 Discussion

**Characterizing hours shocks** — In order to investigate and understand the sources of permanent hours shocks, we estimate their standard deviation in alternative samples. Column I in Table 3.6 reports the value for the full sample. Column II contains results for a sample of individuals in blue collar jobs. Individuals in advanced technical sectors, like electrical and mechanical engineering or skilled service jobs like legal or medical services are excluded. One could expect that the demand for these more regularized jobs only allows for very limited variation in hours. However, this does not seem to be the case, as the estimate of the permanent shocks hardly changes. In column III we exclude the years 1981 and 1982, when a global recession hit the US. The estimate of the standard deviation of permanent hours shocks hardly changes, which shows that the results are not driven by the crisis. Finally, in column IV the sample is restricted to individuals who have stayed in their current job for at least twelve months. This leads to a slight decrease in the estimate. However, it is safe to say that permanent hours shocks do not reflect changes in occupation or job instability. The upshot of these alternative estimations is that permanent hours shocks play an important role throughout all samples and are not restricted to very specific adjustments or at-risk groups. The fact that hours shocks do not seem to be driven by occupation changes or possibly unwanted changes in hours of work during crises suggests an interpretation of permanent hours shocks as shocks to home production.

**Table 3.6:** Permanent hours shock variances in alternative samples and models

| | I | II | III | IV |
|---|---|---|---|---|
| | Main | Blue collar | Exclude years 81-82 | Only stayers |
| $\sigma_{\zeta,v}/\gamma$ | 0.1990 | 0.2087 | 0.2066 | 0.1918 |
| | (0.0010) | (0.0018) | (0.0025) | (0.0016) |
| $N$ | 46340 | 38030 | 40999 | 35901 |

*Note:* Own calculation based on the PSID. Bootstrapped standard errors in parentheses.

**Hours shocks and transmission in alternative models** — In Table 3.7 we report the parameters of the hours shock process and the transmission parameter as well as the implied Marshallian elasticity for the main sample under various restrictions or alternative assumptions. Further we display a measure of overall fit of these alternative models, namely the value of the distance function $DF(\Theta)$, so that we may develop an idea about the value of the main model in describing the data. The estimates of the main model are repeated for comparison in column I. In column II, the variance of $\ln(\phi_t^\lambda)$ is calibrated to half the value of our main specification. All estimated coefficients except for the standard deviation of permanent hours shocks and the mean of the transmission parameter are unchanged. The standard deviation of the permanent hours shocks is slightly larger. The reason is that the variance of the transmission parameter interacts with the variance of hours shocks in explaining the variance of hours growth, see equation (3.20). The fit of this alternative model is just slightly worse, since the variance of permanent wage shocks can freely adjust. The exercise demonstrates that the results only depend to a small degree on this calibration. Columns III and IV illustrate the importance of allowing for hours shocks. In column III the variance of permanent hours shocks is set to zero. While the estimated variance of transitory hours shocks increases only slightly, the estimated mean of the transmission parameter increases to roughly 2.47. The fit of this model is substantially worse with an increase of the distance function by 6 orders of magnitude. The implied Marshallian elasticity doubles. In column IV both transitory and permanent hours shocks are restricted to zero. In this case the estimated average transmission parameter increases to roughly 20.8 and an implied Marshallian elasticity of about -0.7. These extreme estimates are explained by the fact that the transmission of wage shocks is now the only channel to explain hours variance. Naturally, the fit takes another hit from this restriction, although it is not as severe as the first jump. The order of magnitude of the distance function increases onefold. That the final change in fit is not as large as the one in model III further underlines the fact that permanent wage shocks are an important part of the picture in the attempt to explain the variance of the hours residual.

**Table 3.7:** AR Hours Estimation in Alternative Models

| | I | II | III | IV |
|---|---|---|---|---|
| | Main Model | $\sigma_\phi$ halved | $\sigma_{\zeta,v} = 0$ | $\sigma_{\zeta,v} = 0$ & $\sigma_{\epsilon,v} = 0$ |
| $\theta_v/\gamma$ | 0.1515 | 0.1515 | 0.1454 | |
| | (0.0039) | (0.0039) | (0.0013) | |
| $\sigma_{\epsilon,v}/\gamma$ | 0.1114 | 0.1114 | 0.1501 | |
| | (0.0011) | (0.0011) | (0.0005) | |
| $\sigma_{\zeta,v}/\gamma$ | 0.1990 | 0.2116 | | |
| | (0.0010) | (0.0026) | | |
| $E[\phi_t^\lambda]$ | 1.8918 | 1.4317 | 2.4705 | 20.7997 |
| | (0.1117) | (0.0691) | (0.1784) | (3.2642) |
| $E[\kappa]$ | -0.0767 | -0.0767 | -0.1450 | -0.6952 |
| | (0.0105) | (0.0105) | (0.0111) | (0.0093) |
| $DF(\Theta)$ | $1.9398 \times 10^{-11}$ | $7.0055 \times 10^{-11}$ | $3.8247 \times 10^{-05}$ | 0.0005 |

*Note:* Own calculation based on the PSID. Bootstrapped standard errors in parentheses.

**Model Fit** — The model attempts to fit the first three autocovariance moments of the hours and wage residuals and the covariance of the two. In Table 3.8 we show how the data estimates of these moments stack up against the values fit by the model. We use the main sample to evaluate the fit. As expected, the model fits the empirical moments very well.

**Table 3.8:** Model Fit

| | Var. wages | 1. AutoCov wages | 2. AutoCov wages | Var. hours | 1. AutoCov hours | 2. AutoCov hours | Cov. hours & wages |
|---|---|---|---|---|---|---|---|
| emp. | 0.1530 | -0.0560 | -0.0048 | 0.1443 | -0.0529 | -0.0052 | -0.1010 |
| mod. | 0.1530 | -0.0560 | -0.0048 | 0.1443 | -0.0529 | -0.0052 | -0.1010 |

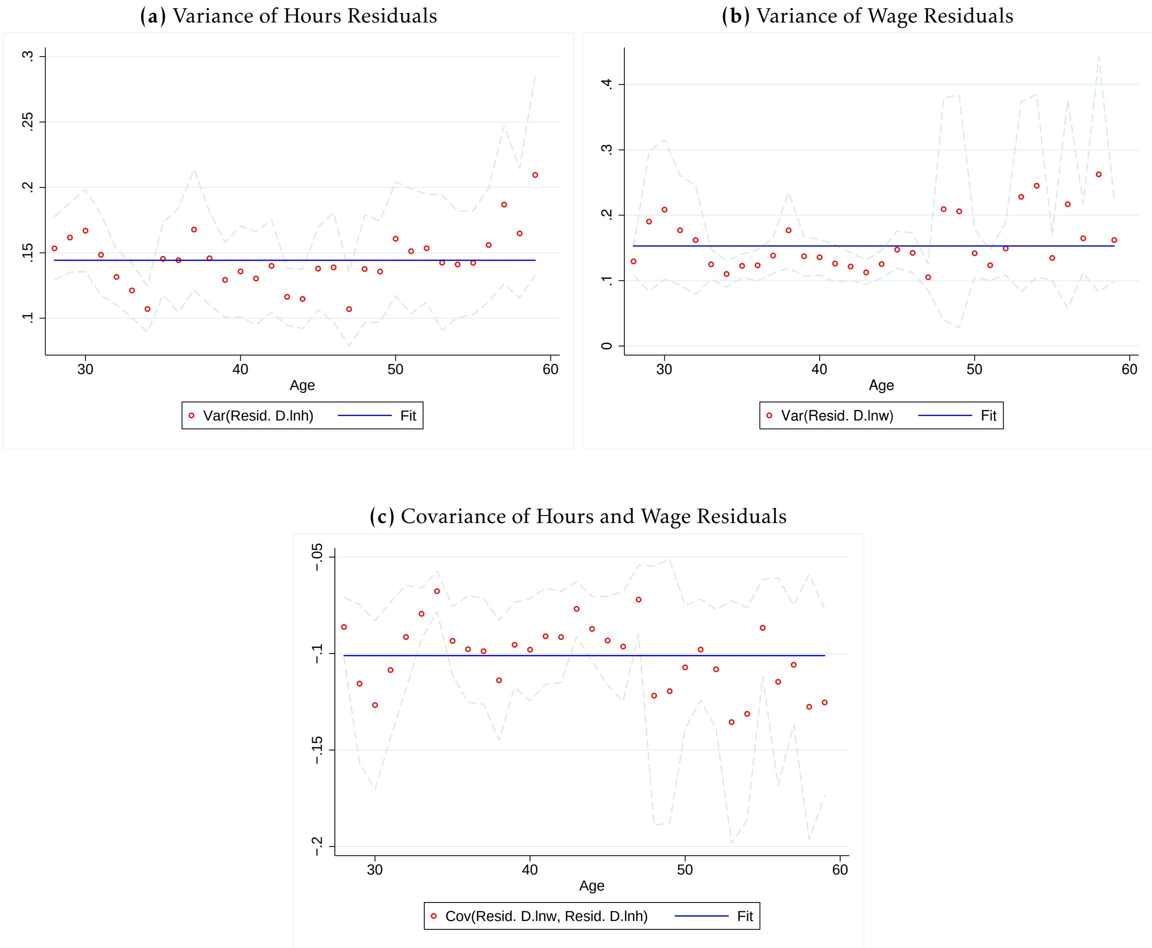
*Note:* Variance moments of residuals obtained from the regressions of equations (3.5) and (3.4) for the main sample.
Own calculation based on the PSID.

The current model does not allow for variation in these targeted variances over age groups and thus imposes that their pattern is essentially flat over the life-cycle. Figures 3.4a, 3.4b and 3.4c show the two residual variance series and the covariance over age. The figures show that these variances do not vary substantially over the life-cycle.

**Partial consumption insurance** — The parameter $\phi_t^\lambda$ is directly related to consumption growth, see equation (3.13) and Alan et al. (forthcoming). In our model with endogenous labor supply permanent wage shocks translate into changes in consumption by $\phi_t^\lambda/\vartheta \times (1 + (1 - \phi_t^\lambda)/(\gamma + \phi_t^\lambda))$. We set $\vartheta = 2$, which is close to the estimates of related papers[16] and calculate the resulting pass-through at mean values of $\phi_t^\lambda$, reported in Table 3.9. For the full sample we find that on average a permanent wage shock of one percent leads to an increase in consumption by 0.76 percent. This figure can be compared to studies that use consumption data to obtain similar parameters. Blundell et al. (2016a) use 1999-2009 PISD data and find that the Marshallian response of consumption to male wage shocks is 0.58, when female labor supply is held constant. We obtain a slightly smaller pass-through parameter for the older sample than for the main sample, but find a substantially smaller pass-through of wage shocks to consumption for the highly educated, for whom a permanent wage increase by one percent leads to an increase in consumption of just 0.31 percent. Using a similar data set to ours, 1978-1992 PSID data, Blundell et al. (2008) estimate the pass-through of permanent *income* shocks to consumption, which is given by $\phi_t^\lambda/\vartheta$ in our model. With a Marshallian labor supply elasticity close to zero – as the one we have estimated – this parameter comes close to the pass-through of permanent wage shocks. Their estimate for the full sample is 0.64 and the estimate for their college sample is 0.42. We confirm the finding that the highly educated are much better insured against income shocks than is the case for the whole population.

[16] Blundell et al. (2016a) estimate a parameter of relative risk aversion of 2.4 and Heathcote et al. (2014a) estimate 1.7.

**Figure 3.3:** Fit of variance and covariance moments over the life-cycle




*Note:* Own calculation based on the PSID. Empirical and theoretical variance and covariance moments of residuals obtained from the estimation of equations (3.4) and (3.5) for the main sample with bootstrapped 95 percent confidence interval.

**Table 3.9:** Pass-through of permanent wage shocks to consumption

| I | II | III | IV |
|---|---|---|---|
| Full sample | Age>=40 | High educ | No young children |
| 0.7648 | 0.6304 | 0.3135 | 0.4820 |

*Note:* $E[\phi_t^{\lambda}]/\vartheta \times (1 - E[\phi_t^{\lambda}])/(\gamma + E[\phi_t^{\lambda}])$ with $\vartheta$=2.
Own calculation based on the PSID. Bootstrapped standard errors in parentheses.

Much of the literature on consumption insurance makes use of moment conditions involving consumption data. We obtain comparable estimates from labor supply and earnings data alone. Similarly, Heathcote et al. (2014a) estimate their model with and without moment conditions using consumption. The obtained estimates of the share of insurable wage dispersion are essentially the same. A simple back-of-the-envelope calculation based on our results for the pass-through parameter to the marginal utility of wealth yields consumption insurance parameters that are broadly comparable to those obtained in previous papers using consumption data. This adds to the notion that much can be learned about consumption insurance from earnings and labor supply data alone.

## 3.6 Qualifications and Extensions

The model chosen for this current analysis, although it is a workhorse for modern dynamic labor supply studies, is not without its drawbacks. Firstly, it is not a model of all of the individual's life-cycle. We do not consider the retirement phase of the individual and the influence this part of the life-cycle has on individual's behavior. Empirically, it seems that consumption strongly adjusts around the time of retirement, yet behavior before entry into retirement is quite consistent, which is also known as the retirement-consumption puzzle (Attanasio and Weber, 2010).[17] However, it is apparent that this should heavily bear on the labor supply decisions we seek to explain during the active working phase. Secondly, we abstract from the household context and abstract from joint decision-making or at least responsiveness to behavior of the partner in the household. Recent studies, for example Blundell et al. (2016a), allow for a joint consideration of the primary and secondary earner's decisions in the household. Since the paper is focused on the decomposition of earnings variance and risk, we are willing to omit this feature. However, it cannot be ruled out that some of the reactions of the primary earner to, for example,

[17] This puzzle has been resolved to some extent in recent papers. See for example Hurd and Rohwedder (2008)

the secondary earners wage process are picked up in our estimates. More on this below. Thirdly, we assume that agents know the type and parameters of their shock processes. In particular, they know whether shocks are permanent or transitory and the respective variance of the shocks. They simply do not know the realization of the shock. There has been some recent work to soften this assumption. The main contribution in this direction is Guvenen (2007), who develops a model of individuals applying Bayesian learning to discover the nature of their earnings process. While full knowledge of the shock process is certainly a very strong assumption, the generalization to a Bayesian learning framework brings with it other conceptual pitfalls. Firstly, individuals still need to know about the structure of the shock process. Secondly, there is the issue of individual's priors about the shock variances of a process. It is not obvious how to they should be modeled and, in fact, mirrors the issue of assuming the variances of the shocks as known. This consideration is beyond the scope of this paper.

A crucial assumption in the set-up of the model is that innovations to the disutility of labor arrive through the parameter $b_t$. This parameter can possibly also be associated with innovations to the utility from consumption, since $b_t$ appears as a factor in the marginal rate of substitution between labor supply and consumption. The following argument can be put forth to defend the interpretation in terms of the disutility from labor supply: shocks that shape utility from consumption may mainly be rooted in the innovation of new products, which are aggregate shocks not included in the idiosyncratic part of $b_t$. Otherwise, changes in the utility from consumption might largely be driven by age, which will also not be picked up in $v_t$. However, we cannot definitvely rule out that certain shocks to the utility from consumption do enter and affect our results.

A possible extension is to make the form of the shock process more flexible to capture not only dynamics in the variance of earnings, but higher moments and therefore be more precise about what type of risk individuals are facing. Recent work in this direction has been done by Guvenen and Smith (2014) and Guvenen et al. (2015) in a mainly descriptive fashion and documenting the nonlinear nature of earnings risk. Arellano et al. (2017) develop a method to estimate such nonlinear earnings processes using panel data and use the PSID and Norwegian register data to document nonlinear persistence of shocks and conditional skewness. They also trace the impact of these phenomena to differential consumption responses. However, they do not model labor supply.

## 3.7 Conclusion

At the outset we asked a simple question: What are the drivers of the riskiness of earnings? To get at the answer, we have decomposed idiosyncratic income uncertainty into contributions of transitory and permanent wage and hours shocks. This is a departure from extant work, where unexplained income volatility is entirely due to wage shocks. In order to separate hours shocks from labor supply reactions to wage shocks, we build on a life-cycle model of labor supply and consumption and estimate a transmission parameter that captures the impact of shocks on the marginal utility of wealth. This parameter captures consumption insurance and is allowed to vary between individuals. We find that both wages and hours are subject to permanent shocks. At the mean, permanent wage shocks have a stronger impact on life-time earnings. Using the mean of the transmission parameter and mean annual earnings, a positive permanent wage shock of one standard-deviation at age 30 increases life-time earnings by 150 000 Dollar, while the effect of a permanent hours shock of one standard-deviation is 124 000 Dollar. Both permanent hours and wage shocks are an important source of cross-sectional earnings growth volatility and earnings risk. Ergo, the data tell a story beyond wage risk.

Along the way to this result, we have shown an alternative way to calculate the Marshallian elasticity of labor supply, which we find to be negative, but small, at -0.08. There is more insurance against permanent wage shocks among the highly educated, for whom we estimate a small positive Marshallian elasticity. Setting the variance of both transitory and permanent hours shocks to zero, we estimate a Marshallian of -0.70 for the main sample, which demonstrates the importance of modeling hours shocks.

Our investigation of the sources of permanent hours shocks leads us to believe that they are best described as shocks to home production. We cannot rule out that other restrictions affect the variance of hours. However, we do rule out two potential major sources of hours shocks. Permanent hours shocks persist as a phenomenon when restricting the sample to individuals who stay in their respective jobs over time and when excluding the years 1981-82, when a major economic crisis hit the US. These tests, along with the results from our four main samples, strongly suggest that hours shocks are a phenomenon that is not restricted to very specific, one-off adjustments or only relevant for narrowly defined groups.

Calibrating the coefficient of relative risk aversion, we calculate the pass-through of permanent wage shocks to consumption and find reasonable figures in the same range as those reported in Blundell et al. (2008). These results are encouraging as they show that comparable estimates of consumption insurance can be obtained using consumption or labor supply data.

Natural extensions of our framework include modeling family labor supply and the extensive labor supply margin. Moreover, the sources of hours shocks merit further research. One promising avenue would be to explicitly model and then separate out shocks to home production from other restrictions to labor supply.

# 3.8 Appendix

## 3.8.1 Derivation of the Labor Supply Equation

The residual in the labor supply equation consists of in-period shocks and expectations corrections in the marginal utility of wealth due both to wage and hours shocks.

The first order condition of the consumer's problem w.r.t. $h_t$ is:

$$\frac{\partial \mathcal{L}}{\partial h_t} = E_t\left[\left(-b_t h_t^{\gamma}\right) + \lambda_t w_t\right] = 0, \tag{3.27}$$

where $\lambda_t = \frac{\partial u(c_t,h_t,b_t)}{\partial C_t}$ denotes the marginal utility of wealth. The Euler equation of consumption is given by

$$\frac{1}{\rho(1+r_t)}\lambda_t = E_t[\lambda_{t+1}]. \tag{3.28}$$

Expectations are rational, i.e., $\lambda_{t+1} = E_t[\lambda_{t+1}] + \varepsilon_{\lambda_{t+1}}$, where $\varepsilon_{\lambda_{t+1}}$ denotes the mean-zero expectation correction of $E_t[\lambda_{t+1}]$ performed in period $t+1$. Expectation errors are caused by innovations in the hourly wage residual $\omega_{t+1}$ and innovations in hours shocks $v_{t+1}$, which, as implied by rational expectations, are uncorrelated with $E_t[\lambda_{t+1}]$. Rational expectations imply that $\varepsilon_{\lambda_{t+1}}$ is uncorrelated over time, so that regardless of the autocorrelative structure of the shock terms, $\varepsilon_{\lambda_{t+1}}$ will only be correlated with the innovations of the shock processes.

Resolving the expectation operator in equation (3.27) yields

$$b_t h_t^{\gamma} = \lambda_t w_t. \tag{3.29}$$

Taking logs of both sides we arrive at the structural labor supply equation

$$\ln h_t = \frac{1}{\gamma}(\ln \lambda_t + \ln w_t - \ln b_t). \tag{3.30}$$

To find an estimable form for $\ln h_t$, we take logs of (3.28) and resolve the expectation:

$$\ln \lambda_t = \ln(1+r_t) + \ln \rho + \ln\left(\lambda_{t+1} - \varepsilon_{\lambda_{t+1}}\right)$$

A first order Taylor-expansion of $\ln\left(\lambda_{t+1} - \varepsilon_{\lambda_{t+1}}\right)$ gives $\ln(\lambda_{t+1}) - \frac{\varepsilon_{\lambda_{t+1}}}{\lambda_{t+1}}$, leading to the expression

$$\ln \lambda_t = \ln(1+r_t) + \ln \rho + \ln(\lambda_{t+1}) - \frac{\varepsilon_{\lambda_{t+1}}}{\lambda_{t+1}} + \mathcal{O}\left(-\frac{1}{2}\left(\frac{\varepsilon_{t+1}}{\lambda_{t+1}}\right)^2\right). \tag{3.31}$$

Accordingly, when we backdate (4.35), we can insert it in (3.30) and remove $\ln \lambda_t$ by first differencing.

### 3.8.2 Distribution of the Shock Pass-Through on Hours

The moments of the term $\frac{\phi_t^\lambda}{\gamma+\phi_t^\lambda}$ are not as tractable as the rest of the random variables in the variance moment estimation, since we assume $\ln \phi_t^\lambda \sim N(\mu_\phi, \sigma_\phi)$. We can refine the expression to find a more basic expression:

$$\frac{\phi_t^\lambda}{\gamma+\phi_t^\lambda} = 1 - \gamma \frac{1}{\gamma+\phi_t^\lambda}$$

The only random term in this expression is $\frac{1}{\gamma+\phi_t^\lambda}$. We can find its distribution by re-expressing its CDF in terms of the underlying normal distribution of $\ln \phi$. Let $\frac{1}{\gamma+\phi_t^\lambda} = Z$. Then

$$P(Z \le z) = P\left(\frac{1}{\gamma+\phi_t^\lambda} \le z\right) \tag{3.32}$$

$$P\left(\ln \phi_t^\lambda \le \ln\left(\frac{1}{z} - \gamma\right)\right) = \int_{-\infty}^{\ln\left(\frac{1}{z}-\gamma\right)} \frac{\exp\left(-\frac{(x-\mu_\phi)^2}{2\sigma_\phi^2}\right)}{\sqrt{2\pi\sigma_\phi^2}} dx \tag{3.33}$$

Integrating this CDF, we find the CDF for the random variable $Z$.

$$F(z) = 1/2\left(1 - \mathrm{Erf}\left(\frac{\ln\left(\frac{1}{z} - \gamma\right) - \mu_\phi}{(2\sigma_\phi^2)^{1/2}}\right)\right)$$

Here $\mathrm{Erf}(\cdot)$ is the Gaussian error function. To generate the first and second noncentral moments, we take the derivative to find the PDF of $Z$.

$$f(z) = -\frac{\exp\left(-\frac{\left(\ln\left(\frac{1}{z}-\gamma\right)-\mu_\phi\right)^2}{2\sigma_\phi^2}\right)}{\sqrt{2\pi\sigma_\phi^2} z(1+z\gamma)}$$

The first and second noncentral moments are $\mathcal{M}_1 = \int_0^{1/\gamma} z f(z) dz$ and $\mathcal{M}_2 = \int_0^{1/\gamma} z^2 f(z) dz$. These are calculated via numerical integration, as there is no closed form solution. We implement these formulas in our moment conditions. In estimation we restrict the values of $\mu_\phi$ not to exceed 5, as the moments of $\frac{\phi_t^\lambda}{\gamma+\phi_t^\lambda}$ asymptote beyond that point.

### 3.8.3 Tables

**Table 3.10:** Frisch Labor Supply Equation Estimation

| | I | II | III | IV |
|---|---|---|---|---|
| | Full sample | Age>=35 | High educ | With children |
| $\Delta \ln(wage)$ | 0.3614 | 0.4020 | 0.2851 | 0.3148 |
| | (0.0856) | (0.3778) | (0.0975) | (0.1080) |
| $N$ | 46340 | 20607 | 19831 | 24547 |
| Kleibergen and Paap (2006) F stat | 18.4680 | 1.2408 | 11.7317 | 11.6739 |

*Note:* Own calculation based on the PSID. Clustered standard errors in parentheses.

# 4 Earnings risk and tax policy

## 4.1 Introduction

Over their life-cycle individuals experience changes in the remuneration of their work. Some of these changes in earnings are expected, some are not. The unforeseen changes constitute risk that alters forward-looking working, saving, and consumption decisions. The class of models investigating these responses typically assume an incomplete financial market and frame saving with a non-contingent bond as the central device to insure against income risk. In this class of models the consumption reaction to a permanent change in income should be strong, even one-to-one if the utility function does not exhibit positive prudence.[1] However, as for example Blundell et al. (2008) document, pass-through of a permanent shock from income to consumption is not one-for-one empirically, which is generally referred to as *excess smoothness*. This phenomenon has lead researchers to explore other sources of insurance that mitigate the pass-through, such as precautionary saving, labor supply, labor supply of the secondary earner, and progressive taxation (Blundell et al., 2015, 2016a; Heathcote et al., 2014a).

The current paper sets out to answer two major questions related to the riskiness of earnings:

1. How has permanent earnings risk developed over the last decade in the US and to what extent has progressive taxation insured against this risk?

2. What is the insurance effect of the current and alternative tax schedules?

The analysis builds on a life-cycle model of labor supply, which features risk originating in hourly wages. The model is tractable and allows for the estimation of the relevant parameters giving the evolution of risk and the behavioral parameters of labor supply in two steps: 1. Estimation of growth in log wages and log hours, derived from the first order condition for labor supply, which yields residual wage and residual hours growth. 2. Method of moments estimation using the second moments of residual wage and residual hours growth, which yields the relevant parameters to measure risk and pass-through to earnings. This approach has successfully been employed by both Blundell et al. (2016a) and Heathcote et al. (2014a) for the purpose of exploring and comparing different insurance channels. The approach relies on the

[1] See Krueger (2007, pp.46-48) for an illustration with a specific income process and quadratic utility. Further, see Attanasio and Weber (2010) for a comprehensive survey on the issue of consumption responses to changes in the income process. Precautionary saving, which operates when prudence is positive, can smooth the consumption profile.

approximation of the labor supply equation and the life-time budget constraint.[2] Income taxation, like in previous contributions ( Blundell et al. (2016a) and Heathcote et al. (2014a)) is modeled by means of a power function approximation. Two novelties of the current paper are the integration of tax deductions and that I find a form of the approximation that is compatible with nonlabor income. In both cases the forms are chosen to maintain the tractability of the first order condition approach. On the empirical side, to quantify risk and insurance, I follow the standard approach of estimating shock variances and calculating their pass-through to earnings. I use a novel approach to assess how changes in the tax code alter the insurance provided by progressive taxation. First, I slightly modify the tax code of the tax calculator and compute the new distribution of incomes. Second, using this new distribution of incomes, I estimate the parameters of the tax function approximation and calculate the relative change to the base level.

I employ US data from the Panel Study of Income Dynamics (PSID) from 1998 to 2014, which means I can cover the influence of the US financial crisis of 2008 on my measure of risk. The PSID has been used extensively in other studies of this issue and enables me to embed my findings in the literature.

As concerns the research questions, I find a moderate rise in the permanent component of wage risk until 2006 (18% increase in the standard deviation) and a strong increase from 2006 to 2008 (14% increase in the standard deviation) that persists until 2010 and only partially reverts in 2012. In studying the pass-through of this "gross" risk onto earnings I find that progressive taxation played a minor mitigating role. On average about a 5% decrease in pass-through compared to the case with no progressive taxation. I cannot detect a major change in the estimated progressivity and, in turn, insurance around the 2008 crisis. The rise in gross risk after the crisis transfers to net earnings. Finally, I examine a counterfactual calculation, of how high the top tax rate needed to be to return earnings risk to pre-crisis levels. I conclude that an increase of about 56% of the actual rate would have achieved this goal.

As concerns the modeling of the tax function, I find that tax deductions are of minor importance in shaping how progressive the tax system is. However, they do have an influence on the parameters of the approximated function. Further, when investigating the approximation of net income, I find that the estimates are sensitive to whether one fits the relationship in levels or logs. The approximation in levels provides a better fit and implies less progressivity. Overall, I find that the fitted power function performs relatively well. I calculate a fiscal gap per tax unit, the difference between actual and fitted tax burden, of roughly -650$ implying that the model slightly underpredicts liabilities.

[2]The accuracy of these approximation methods has been explored in Domeij and Flodén (2006); Blundell et al. (2008, 2016a) among others. I will not examine the properties of these approximations.

In the next section I provide an overview of the related literature. In section 4.3 I relate the model set up and then focus on the approximation of the retention function in sections 4.4 to 4.6.2. I introduce the PSID data in section 4.5. Section 4.7 presents the results of estimating the life-cycle model and section 4.8 contextualizes these results in terms of the pass-through of wage risk to earnings risk and conducts the policy exercise regarding the top tax rate.

## 4.2 Taxation as Insurance

The idea that taxation can act as an insurance mechanism against risky income flows dates back to the 1980s with foundational contributions by Varian (1980) and Eaton and Rosen (1980). While Varian (1980) considers a dynamic model, in which an individual may self-insure against income risk via saving, Eaton and Rosen (1980) explicitly model labor supply with uncertainty in wages but neglect dynamics. Both Eaton and Rosen (1980) and Varian (1980) come to the conclusion that taxation can be desirable if an individual faces shocks and is risk averse. Varian (1980) considers an optimal nonlinear tax schedule with the finding that a more progressive tax schedule is optimal, when the Arrow-Pratt measure of absolute risk aversion is increasing. Under the imposition of a utility function that features constant relative risk aversion (CRRA) marginal tax rates increase in income, given that the coefficient of relative risk aversion is larger than 1.

Low and Maldoom (2004) make the connection between the two papers and examine optimal taxation when labor income is risky and both labor supply and savings are choice variables of the individual. The fundamental trade-off is between the incentive effect on labor supply stemming from income uncertainty and the benefit of social insurance that derives from lowering the variance of net income. They determine that the trade-off is parametrized by the ratio of prudence to risk aversion.

Several recent empirical studies like Blundell et al. (2016a), Blundell et al. (2015) ,Heathcote et al. (2014a), and Heathcote et al. (2017a) seek to estimate the degree of insurance stemming from sources like savings, (family) labor supply and taxes over the life-cycle. Blundell et al. (2016a) find that insurance via progressive taxation makes up a sizable contribution (~11%) to insurance of permanent wage shocks, but other forces, most prominently (family) labor supply, dominate. For Norwegian data analyzed in Blundell et al. (2015) the riskiness of earnings is strongly attenuated by the tax and transfer system, especially for those with lower education, who experience roughly 20% less impact of a permanent shock of one standard deviation on annual disposable compared to annual market income. Heathcote et al. (2014a) simply aim to calculate the overall insurance provided by the aforementioned mechanisms. They do, however, find that own labor supply dampens the effect persistent shocks have on consumption by roughly 15%. Heathcote et al. (2017a) go a step beyond the description of insurance mechanisms and provide a closed-form expression for social welfare, which crucially depends on the riskiness of earnings, and ultimately characterize the progressivity of the optimal income tax.

In the current paper I take a closer look at the descriptive side of insuring income fluctuations through progressive taxation. I evaluate how this insurance channel has been shaped by the policy maker. I do not provide a closed-form expression for social welfare, as Heathcote et al. (2017a) do, and am therefore silent on optimal policy. It is certainly the case that the policy maker should adopt some consistent stance on how to treat the trade-off between providing insurance and diminishing incentives for work. Therefore, I document the insurance effect of the chosen tax policy and its effect on net earnings risk.

## 4.3 The Model

The life-cycle behavior of individuals is described in the following. I give a detailed treatment of the way the tax and transfer system is modeled in sections 4.4-4.6.2.

**The Life-Cycle Problem** Life-cycle optimization by the individual proceeds by maximization of the sum of discounted in-period utilities from $t_0$ to $T$. I omit an individual specific subscript.

$$\max_{c_t,h_t} E_{t_0}\left[\sum_{t=t_0}^{T} \rho^{t-t_0} \mathrm{v}(c_t, h_t, b_t)\right], \tag{4.1}$$

where v is the in-period utility function taking consumption $c_t$, hours of work $h_t$, and taste- shifters $b_t$ as arguments. $\rho$ is the discount factor. I specify the taste shifter $b_t = \exp(\varsigma\Xi_t - \upsilon_t)$. $\varsigma\Xi_t$ is a linear combination of a set of personal characteristics. $\upsilon_t$ accounts for the non-systematic variation of the taste shifter, which is assumed to be normally distributed and uncorrelated over time. The functional form of the in-period utility function is given by

$$\mathrm{v}(c_t, h_t, b_t) = \frac{c_t^{1-\vartheta}}{1-\vartheta} - b_t \frac{h_t^{1+\gamma}}{1+\gamma}, \qquad \vartheta \geq 0, \gamma \geq 0, \tag{4.2}$$

where $\frac{1}{\vartheta}$ pins down the intertemporal elasticity of substitution with respect to consumption, while $\frac{1}{\gamma}$ gives the Frisch-elasticity of labor supply. Thus, in-period utility is additively-separable and conforms to constant relative risk aversion (CRRA).

The intertemporal budget constraint is

$$\frac{a_{t+1}}{(1+r_t)} = a_t + \mathcal{T}_t(w_t h_t + N_t) - c_t, \tag{4.3}$$

where $a_t$ represents assets, $r_t$ the real interest rate,[3] $N_t$ non-labor income, and $\mathcal{T}_t(\cdot)$ is the retention function that returns post-government/net income.

In the following I will examine interior solutions to this problem by estimating the associated labor supply equation derived from the first-order conditions.

**Uncertainty** Wages evolve according to the equation

$$\Delta \ln w_t = \alpha X_t + \Delta\omega_t, \tag{4.4}$$

where $X_t$ contains variables influencing human capital and $\Delta\omega_t$ contains the innovation in the idiosyncratic shock processes.

It is assumed that the unobservable components determining wage growth can be decomposed into a permanent and a transitory process, which are chosen to be a random walk and a small-order moving average process (MA(1)).[4] Thus, the dynamics of the idiosyncratic component of wages are described by the following set of equations:

$$\begin{aligned}
&\omega_{it} = p_{it} + \tau_{it} \qquad (4.5)\\
&p_{it} = p_{it-1} + \zeta_{it}\\
&\tau_{it} = \theta\epsilon_{it-1} + \epsilon_{it}\\
&\zeta_{it} \sim N\left(0, \sigma^2_{\zeta,t}\right), \quad \epsilon_{it} \sim N\left(0, \sigma^2_{\epsilon,t}\right)\\
&E\left[\zeta_t\zeta_{t-l}\right] = 0, \quad E\left[\epsilon_t\epsilon_{t-l}\right] = 0 \quad \forall l \in \mathbb{Z}_{\neq 0}
\end{aligned}$$

$p_{it}$ is the permanent component and $\tau_{it}$ the transitory. While $p_{it}$ and $\tau_{it}$ exhibit serial correlation, their innovations $\zeta_{it}$ and $\epsilon_{it}$ do not. With this error structure I can proceed to estimate the shock variances using the autocovariance moments of wages. The identification of the individual shock variances from year to year is discussed in 4.7.2 and in appendix 4.11.6. Suffice it to say that I need to impose an assumption regarding the initial conditions of the transitory shocks variances. In the empirical implementation I choose the zeroth and first transitory variance to have the same value. I determine the propagation of the shocks to hours and ultimately earnings in section 4.6.2.

**Empirical Implementation** Using the first order condition for labor supply and the wage equation in log changes, I can estimate residuals for both quantities. I assume that the levels of wages, hours and earnings are measured with error. I give my treatment of measurement error of in appendix 4.11.1. Using the second

[3] Accordingly, the model features an incomplete capital market.

[4] Choosing this process is standard in the extant literature and has favorable properties as far as identification is concerned.

moments of the residuals from the labor supply and wage growth estimation I can recover the wage shock variances as well as the relevant parameters to determine shock pass-through to earnings. The roadblock in the way of the first step is how to model the function $\mathcal{T}(\cdot)$ in a way that makes the first step tractable. I will lay out the theoretical issues regarding this issue in section 4.4 and let the implementation follow in sections 4.5 and 4.6.

## 4.4 Approximation of the Tax and Transfer System

In brief, to proceed with the labor supply estimation, I need to approximate the retention function.[5] The retention function $\mathcal{T}(\cdot)$ takes as inputs gross income as well as characteristics of the tax unit and returns net income.[6] The function $\mathcal{T}(\cdot)$ is nonlinear, not continuous, and therefore non-differentiable. While it is not an issue to model the dependence on characteristics of the tax unit, it is not tractable to choose a non-continuous, non-differentiable retention function. When one proceeds with the first-order approach to labor supply estimation, while also approximating the life-time budget constraint, it turns out to be very advantageous to choose a power-function approximation of the tax function, as the resulting structural equations will be linear in log-space and the model becomes tractable.[7]

There are three objectives for this theoretical section and the empirical section 4.6: First, in this section I want to introduce the commonly chosen power function approximation of the retention function and its connection to a common measure of progressivity of the tax system. Further, I will show that, when one introduces an explicit distinction between gross and taxable income in the retention function, the deductions determining this relationship may or may not have an impact on labor supply decisions. This depends on whether the function giving taxable income is nonlinear.

Second, in section 4.6 I want to provide accuracy measures of this approximation along the distribution of gross income. It is common to refer to goodness of fit measures, like the $R^2$ from the regression of log net on log gross or log taxable income while imposing the power-function. Implicitly, the reported $R^2$ uses a linear model for the tax function to calculate the total sum of squares. In some situations this comparison is certainly enlightening, however, one can better assess the immediate performance by directly inspecting the deviations between fitted

[5] See Blundell and Macurdy (1999) and Keane (2011) for surveys on labor supply estimation and which approaches exist to implement taxation. Whether using an approximation is sensible, solely rests on how well the approximated post-government income corresponds to its observed counterpart.

[6] In this paper I will neither make an explicit distinction between pre-government and gross income nor between post-government and net income; I will define the terms in section 4.5.

[7] See Blundell et al. (2016a) and Blundell et al. (2016b).

and observed values. Therefore, I will generally rely on the root mean square error (RMSE) to evaluate fit. I will also calculate the implied fiscal gap arising from the approximation approach, so that I may gauge the effect on the government budget that the approximation will entail.

Third, I will explore the relationship between crucial parameters of the tax system, like the top tax rate and the size of the standard deduction, and the shape of the approximation function, in particular the parameter governing progressivity. This is crucial to detect the impact of a policy change on quantities like labor supply and the insurance provided through taxation.

Section 4.6 will also deal with the last steps toward tractability by extending the approximation for nonlabor income and allowing time-differencing.

### 4.4.1 The Power Function Approximation

Choosing a power function to represent the retention function was popularized by Feldstein (1969) and recent applications in the related literature include Kaplan (2012), Blundell et al. (2016a), Heathcote et al. (2014a), and Heathcote et al. (2017a). The relationship between gross and net income is given by

$$\mathcal{T}(y_{i,t}) \approx \chi y_{i,t}^{1-\tau}. \tag{4.6}$$

In this highly simplified version the parameters $\chi$ and $\tau$ are neither individual- nor time-specific. Accordingly such an approximation will miss out on much of the variation that is driven by i) the differences in the assessment criteria that apply to the particular tax unit, e.g. whether the tax unit consists of a couple filing jointly or a single, and ii) the differences in the relevant parameters of the tax code over time, e.g. the top tax rate.

For a single cross-section $t$ it is possible to introduce tax-unit specific variation in the parameters, by letting

$$\mathcal{T}(y_{i,t}) \approx \chi_i (y_i)^{1-\tau_i},$$

and permitting individual variation through a parametric form such that

$$\chi_i = f_\chi(Z_i),$$
$$\tau_i = f_\tau(Z_i),$$

where the functions $f_\chi$ and $f_\tau$ may, for example, be linear in the tax unit's characteristics $Z_i$. However, this type of modeling is incompatible with the structural labor supply estimation pursued later.[8] Further, the focus of the current paper is on

[8] Estimation of the Frisch elasticity would have to be performed per group leading to potential power issues in estimation.

how the policy maker shapes individuals' exposure to earnings risk. Hence, I will fix the parameters in the cross-section, but allow variation over time, so that

$$\mathcal{T}_t(y_{i,t}) \approx \chi_t y_{i,t}^{1-\tau_t}. \tag{4.7}$$

To investigate the relationship between changes in the tax system and changes in the parameters of the approximation, I will need to isolate the mechanical effect that changes in the tax system have on the parameters of the approximation. I pursue the following three-step process to derive the mechanical effect of a change in the tax code on $\tau$: 1) For a given cross-section $t$ I derive the baseline estimate $\tau_t$. 2) I change the tax system in one relevant variable in a microsimulation program, e.g. increase the top marginal tax rate by one percent for the year $t$, run the counterfactual tax simulation and find the new estimate $\tau_t^c$. 3) Finally, I determine the percentage change in the parameter $\frac{\tau_t^c-\tau_t}{\tau_t}$ to determine the elasticity of the approximation parameters with respect to the tax parameters, i.e. the mechanical effect. I use NBER's taxsim tax-calculator, which allows for only a couple of the tax code variations. I have chosen to restrict my attention to two parameters: the standard deduction and the top tax rate. The choice is motivated by the impact these two parameters have on the estimated tax function and their contested status in public policy debates. As described above, I will vary these parameters by one percentage point and calculate the impact on the parameters of the approximation. This exercise is detailed in section 4.6.1.

The above strategy does not come without drawbacks. Since I calculate the approximation using only the sample observed in this particular cross-section, there are bound to be characteristics of the sample that drive the approximation results.

### 4.4.2 The Measure of Progressivity

The choice of the power function to approximate the retention function entails a convenient link with a crucial metric discussed in the economics of taxation: the residual income progressivity or – in more intuitive terms – the elasticity of after-tax income with respect to gross income. Residual income progressivity can be expressed as $\frac{\partial(y-T(y))}{\partial y}\frac{y}{y-T(y)} = \frac{1-\partial T(y)/\partial y}{1-T(y)/y}$, where $T(\cdot)$ gives the tax liability. According to Jakobsson (1976), when comparing two tax schedules $T_1$ and $T_2$, one is more progressive than the other, when

$$\frac{1-\partial T_1(y)/\partial y}{1-T_1(y)/y} < \frac{1-\partial T_2(y)/\partial y}{1-T_2(y)/y} \quad \forall y. \tag{4.8}$$

One implication of Jakobsson's theorem is that for a progressive tax schedule $\partial T(y)/\partial y > T(y)/y \ \forall y$.[9] In the case of the power function approximation, $\frac{\partial(y-T(y))}{\partial y}\frac{y}{y-T(y)} = 1-\tau$. Further, the progressivity parameter $1-\tau$ directly impacts labor supply decisions, as shown in section 4.4.3, and the extent of insurance offered by taxation.

There is some ambiguity in the literature about how to measure the progressivity parameter. Specifically, it is possible to either measure progressivity by approximating net income based on gross income or net income minus deductions on taxable income; the latter being called statutory progressivity.[10] I am interested in the parameters of the tax-function that affect behavior, which, in my setting, are the parameters affecting the choice of hours of work. First, I will alter the notation slightly to examine the relationship between effective and statutory marginal tax rates. I denote the statutory tax function by $T_s(\cdot)$ and by $\tilde{y}(y)$ the function mapping from gross to taxable income. The tax liability is given by the composition of both functions $T_s(\tilde{y}(y))$. The statutory marginal tax rate is given by $T_s' = \frac{\partial T_s}{\partial \tilde{y}}$, while the effective marginal tax rate is $T_e' = \frac{\partial T_s}{\partial \tilde{y}}\frac{\partial \tilde{y}}{\partial y}$. This implies that statutory marginal rates bound effective marginal rates (weakly) from above, i.e.[11]

$$\frac{\partial T_s}{\partial \tilde{y}}\frac{\partial \tilde{y}}{\partial y} \leq \frac{\partial T_s}{\partial \tilde{y}}. \tag{4.9}$$

This is evident, because the above expression holds with equality only when $\frac{\partial \tilde{y}}{\partial y} = 1$. Whether this is the case is an empirical issue discussed in section 4.6.1. However, it is reasonable to expect that the effective tax function will be less progressive than the statutory one.

In the next section I use a static variant of the model in section 4.3 to fix ideas about how the retention function and its parameters affect labor supply decisions.

### 4.4.3 Modeling the Retention Function

Taking the model of section 4.3 while suppressing the time indices and setting non-labor income to zero, the first order condition for hours of work will take the form,

[9] Accordingly, when we approximate progressive tax schedules with continuous and differentiable functions, this implies that the approximation function has to be strictly convex. For intuition why this must be the case, construct the limiting case where $T(\cdot)$ is linear and therefore $\partial T(y)/\partial y = T(y)/y \ \forall y$.

[10] See Blundell et al. (2016a) and Heathcote et al. (2017a) for illustrations of the two competing methods.

[11] This statement only holds when $\frac{\partial \tilde{y}}{\partial y} \leq 1 \ \forall y$, as one would reasonably anticipate.

$$h^{\gamma} = \frac{1}{b}\lambda \mathcal{T}'(wh)w, \tag{4.10}$$

$\mathcal{T}(\cdot)$ is the function giving net income, so I can replace it with $y - T_s(\tilde{y}(y))$. Notice that, depending on whether the function $T_s(\cdot)$ is nonlinear deductions will enter (4.10) and therefore influence the individuals' decision on labor supply. This is clearly the case when we consider a progressive tax system. Accordingly, in this setting I find it to be sensible to let the function giving taxable income also directly depend on gross income to retain tractability. A straightforward way to model both the tax and taxable income based on gross income is,

$$\begin{aligned} T_s(\tilde{y}) &= y - \tilde{\chi}\tilde{y}^{1-\tilde{\tau}}, \\ \tilde{y}(y) &= \kappa y^{1-\iota}. \end{aligned} \tag{4.11}$$

This choice also conforms with the structure of the model of section 4.3 regarding consumption, since there is only one composite consumption good and no other good relevant for deductions. I illustrate more generally how deductions affect labor supply decisions in appendix 4.11.4, where I introduce a separate composite consumption good which can be deducted from gross income.[12] The indication of that model, however, is the same as the one of equation (4.10): as long as the retention function possesses some nonlinearity, deductions do influence the first order condition for labor supply. With the approximation chosen, I resolve $\mathcal{T}'$ in (4.10) to obtain

$$h^{\gamma} = \frac{1}{b}\lambda(1-\tilde{\tau})\tilde{\chi}\left(\kappa(wh)^{1-\iota}\right)^{-\tilde{\tau}}\kappa(1-\iota)(wh)^{-\iota}w. \tag{4.12}$$

Applying logs and rearranging terms,

$$\begin{aligned} \gamma \ln h &= \ln((1-\tilde{\tau})\tilde{\chi}) - \tilde{\tau}(\ln\kappa + (1-\iota)(\ln w + \ln h)) + \ln(\kappa(1-\iota)) - \iota(\ln w + \ln h) + \ln w \\ \ln h &= \frac{[(1-\tilde{\tau})(1-\iota)\ln w - \ln b + \ln\lambda + \ln((1-\tilde{\tau})\tilde{\chi}) + \ln(\kappa(1-\iota)) - \tilde{\tau}\ln\kappa]}{\gamma + \tilde{\tau}(1-\iota) + \iota}. \end{aligned} \tag{4.13}$$

Which implies that the relevant tax-modified $\lambda$-constant elasticity, the analogue to the Frisch, is given by $\frac{(1-\tilde{\tau})(1-\iota)}{\gamma+\tilde{\tau}(1-\iota)+\iota}$. Accordingly, with this specification for the retention function, not only the parameter of the statutory system $\tilde{\tau}$, but also the

[12] This approximation of net income and the tax system is different from the model of net income in Heathcote et al. (2017b), where deductions would appear as an extra additive term in the labor supply equation. This also implies that my statutory tax function is not comparable with their statutory tax function.

parameter of the taxable income function $\iota$ influence the individual's hours-response to changes in the wage.

With these results, I can make some further observations. First, it follows from (4.11) that there exists a direct mapping from gross to net income, which is also a power function. Accordingly, whether I estimate both equations or just one power function approximation from gross to net income results in the same predictions for net income. I define $(1-\iota)(1-\tilde{\tau}) := 1-\tau$ and $\tilde{\chi}\kappa^{1-\tilde{\tau}} := \chi$, so that the tax is given by

$$T(y) = y - \chi y^{1-\tau}. \tag{4.14}$$

This is very convenient if only observations on net and gross income are available in the data or microsimulation of the tax and benefit system is not feasible.

Second, this also has an implication for how progressivity is measured. It becomes clear that progressivity of the entire system can be summarized as,

$$\begin{aligned}\frac{\partial(y - T_s(\tilde{y}(y)))}{\partial y}\frac{y}{y - T_s(\tilde{y}(y))} &= \frac{\tilde{\chi}(1-\tilde{\tau})\left(\kappa y^{1-\iota}\right)^{-\tilde{\tau}}\kappa(1-\iota)y^{-\iota}}{\tilde{\chi}\left(\kappa y^{1-\iota}\right)^{1-\tilde{\tau}}y^{-1}} \\ &= \frac{(1-\tilde{\tau})(1-\iota)\tilde{\chi}\kappa^{1-\tilde{\tau}}y^{-\iota-\tilde{\tau}+\iota\tilde{\tau}}}{\tilde{\chi}\kappa^{1-\tilde{\tau}}y^{-\iota-\tilde{\tau}+\iota\tilde{\tau}}} = (1-\tilde{\tau})(1-\iota).\end{aligned}$$

Accordingly, when $\iota$ increases, progressivity of the tax system increases.

Third, it is important to note that the chosen model implies no impact on progressivity when $\iota = 0$. In this case the overall progressivity $\tau$ could be recovered both from the mapping between gross and net income, but also taxable and net income.

This final issue of whether a power or an affine function best represents the mapping from gross to taxable income is an empirical issue, which I will tackle in section 4.6 after introducing the data source in the next section.

## 4.5 Data

For all empirical exercises I use 9 waves of the Panel Study of Income Dynamics (PSID) coming from the survey years 1999 to 2015.[13] The data are collected biennially and with reference to the previous year, so that the respective observed years are 1998 to 2014. I restrict the sample to households that either have a married or single head of household. Further, I follow Heathcote et al. (2017a) in restricting the sample to households with heads in prime working age, namely 25 to 60, and

[13] PSID (2015) marks the data and see PSID (2017) for an introduction to the PSID.

only keeping households with the head earning at least the equivalent of part-time work at the minimum wage.[14] Finally, I drop observations with less than 260 and more than 4000 yearly hours. These restrictions allow me to focus on the population unambiguously participating in the labor market. Monetary variables are deflated to the base year 1998 using the CPI-U. Sample statistics of relevant variables for the final panel are presented in the reference year 2000 are presented in table 4.1.[15] Hours are on a yearly scale and wages are hourly wages. The underlying income variable is the labor income of the head of the household.

**Table 4.1:** Sample Statistics in Year 2000

| | age | hours | wage | years of education | num. of children |
|---|---|---|---|---|---|
| mean | 40.80 | 2164.25 | 19.88 | 13.22 | 1.08 |
| sd | 9.16 | 554.98 | 21.39 | 2.47 | 1.19 |
| min | 25 | 260 | 2.15 | 0 | 0 |
| max | 60 | 4000 | 447.29 | 17 | 8 |
| taxation variables | gross income | tax liability | net income | | |
| mean | 64508 | 18921 | 50411 | | |
| sd | 71911 | 29591 | 46437 | | |
| min | 5648 | -3481 | 5231 | | |
| max | 2026291 | 905915 | 1145453 | | |
| Obs. | 4878 | | | | |

*Note*: Own calculation based on PSID (2015). All statistics are unweighted.

**Tax Variables** I calculate all taxation variables using the *taxsim* tax calculator provided by the National Bureau of Economic Research (see Feenberg and Coutts (1993)). The preparation of the data before applying the tax calculator is done by adapting files provided by Kimberlin et al. (2016) accessible on the NBER taxsim webpage. Gross income includes all labor and non-labor income of the household, like interest, dividend, and rent income, plus half of FICA.[16] Deductions are calculated using PSID data on mortgage interest, medical expense and charitable giving deductions, while deductions stemming from state taxes are calculated by taxsim. The tax liability contains federal, state and FICA tax. Postgovernment or net income is gross income minus the tax liability plus transfers and half of FICA

[14] This amounts to an hourly wage of $5.15 deflated to 1998 times 1000 yearly hours.

[15] Complete sample statistics for every year are presented in appendix 4.11.2. The year 2000 is fairly representative, although the sample size increases in 2002 and 2008 to a level slightly above 5100 and 5400 observations respectively.

[16] FICA or FICA tax is a payroll tax to fund social security and medicare payments.

(employer's share). Transfers, except for the EITC, which is calculated by taxsim, are recorded in the PSID data, i.e. TANF, social security, unemployment benefits, workers' compensation and veterans' pensions.

## 4.6 Estimating the Tax Function Approximation

In this section I address two broad sets of issues: First, I cover the questions raised in the previous section and explore the fit of the approximation, how relevant deductions are in shaping progressivity, and the policy maker's influence over measured progressivity. Results on these issues apply across the board when a power function is used to approximate the retention function and are not specific to my application. Second, I implement a further modification to the approximation to allow for nonlabor income and I examine the issue of time-differencing with the power function present in the labor supply equation. These are tractability issues that are specific to this paper only.

### 4.6.1 General Results

#### Estimating the Taxable Income Function

I estimate both a linear ($\iota = 0$) and a power function for the relationship between gross and taxable income. If the relationship is linear, then deductions do not affect progressivity. Table 4.2 shows the estimated values for the parameters $\iota$ and $\kappa$ over time in the unrestricted model and in the restricted model ($\iota = 0$). Table 4.2 further shows a likelihood ratio test of the restricted model and a comparison of the root mean square error (RMSE) of the two models. $\iota$ is slightly below zero and so the exponent in the model is greater than one, making the estimated function convex. This indicates that retention of taxable from gross income rises over the distribution of gross income. However, $\frac{\partial \tilde{y}(y)}{\partial y} < 0$ over the relevant range, so that overall progressivity of the tax system is lower due to deductions, as hypothesized in section 4.4.2. The values of $\iota$ exhibit some variation over time, but group pretty tightly in the range of -0.05 to -0.09. In the restricted model, with $\iota$ set to zero, the values for $\kappa$ lie in the range of 0.8 to 0.9 and are much larger than in the unrestricted model. The simple linear model implies that roughly 85 percent of a tax unit's income is taxable. The likelihood ratio tests, however, show that the null hypothesis of model equivalence is easily rejected in all years, indicating that the taxable income function is nonlinear. This is also evidenced by the consistently lower RMSE for the unrestricted model. In relative differences the unrestricted model delivers a roughly 35% lower RMSE over the range of years when the restricted model is the base. Therefore, I reject the linear model of taxable income.

To visualize the fit I plot predicted versus actual values of both models for the year 2000 in figure 4.1. It is clear that the unrestricted model fits the mass at the lower end of the gross income distribution better than the restricted model, while also being able to curve up and fit values at the tail of the distribution. However, both functions have to start in the origin and fail to fit values with positive gross and zero taxable income well.

**Table 4.2:** Taxable Income Function

| | unres. | | restr. | LR-Test | | RMSE | | | |
|---|---|---|---|---|---|---|---|---|---|
| Year | $\kappa$ | $1-\iota$ | $\kappa$ | $\chi^2$ | p | unres. | restr. | rel. diff. | Obs. |
| 1998 | 0.263 | 1.091 | 0.813 | 76401 | 0 | 7773 | 11652 | -0.333 | 4535 |
| 2000 | 0.245 | 1.096 | 0.827 | 82065 | 0 | 8342 | 12794 | -0.348 | 4878 |
| 2002 | 0.373 | 1.063 | 0.854 | 83267 | 0 | 8283 | 12642 | -0.345 | 5104 |
| 2004 | 0.339 | 1.069 | 0.903 | 109172 | 0 | 10427 | 18149 | -0.425 | 5133 |
| 2006 | 0.310 | 1.076 | 0.829 | 80580 | 0 | 8605 | 12836 | -0.330 | 5191 |
| 2008 | 0.388 | 1.058 | 0.846 | 86231 | 0 | 8819 | 13473 | -0.345 | 5415 |
| 2010 | 0.317 | 1.073 | 0.791 | 50267 | 0 | 8313 | 10673 | -0.221 | 5141 |
| 2012 | 0.388 | 1.058 | 0.854 | 91930 | 0 | 7595 | 12227 | -0.379 | 5225 |
| 2014 | 0.325 | 1.073 | 0.836 | 99186 | 0 | 7613 | 12280 | -0.380 | 5346 |

*Note*: Own calculation based on PSID (2015). Restricted ($\iota = 0$) and unrestricted model estimated using nonlinear least squares. Estimation was performed using cross-sectional frequency weights. The relative difference of RMSEs is calculated in the following way: $\frac{RMSE^{unres.}-RMSE^{res.}}{RMSE^{res.}}$

## Estimating the Retention Function

I now estimate the power function mapping for the retention function in two different ways: First, I estimate the mapping from taxable to net income (partial retention function) and then the combined relationship, i.e. the mapping from gross to net income (complete retention function). A visual illustration of the two approaches is given by figure 4.2. Qualitatively, both models fit a globally concave function, but the implied estimates for the progressivity parameter are clearly different. This can further be verified from table 4.3, where in the first four columns the parameter estimates for both models are displayed. That the estimated functions are concave is expected considering the results in the previous section.

I find that the that the progressivity parameter of the partial retention function $1-\tilde{\tau}$ is roughly of the size 0.8 to 0.88. However, here I need to stress again that these estimates are not directly comparable with those in Heathcote et al. (2017a), since the authors model the relationship between net income minus deductions and taxable income.

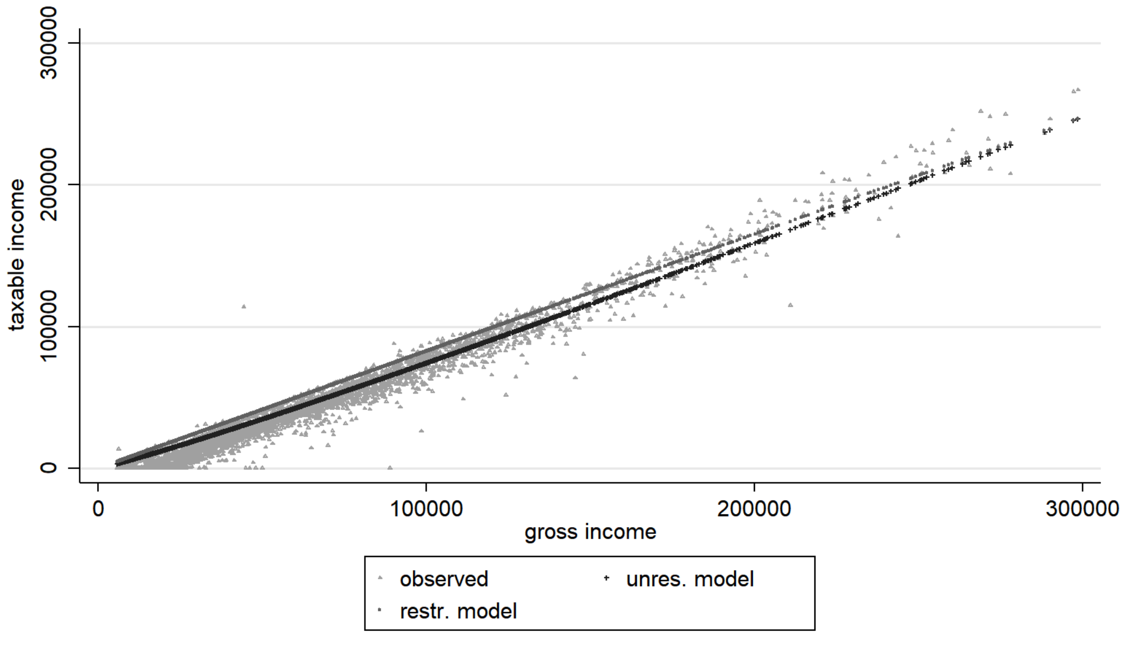


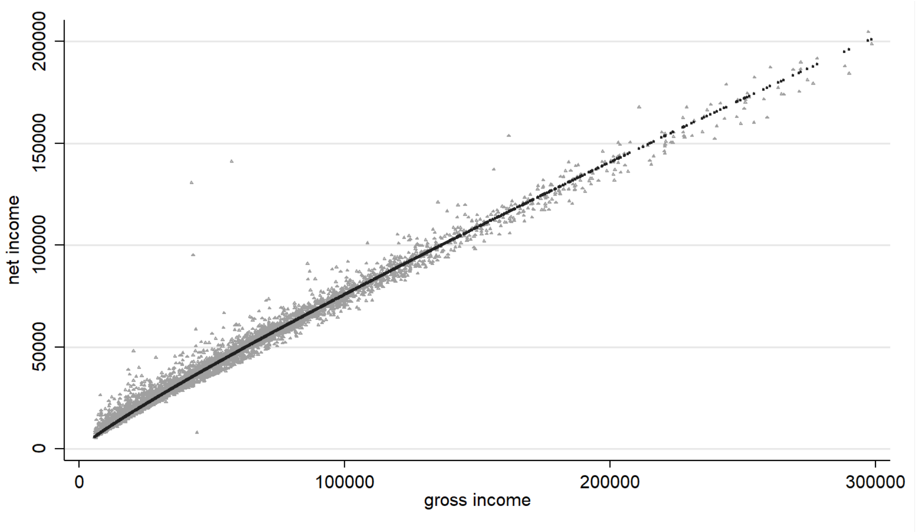

**Figure 4.1:** Taxable Income Model Fit in 2000

*Note*: Own calculation based on PSID (2015). The graph plots gross income against taxable income. Light gray dots indicate observed values, black dots are predictions based on the unrestricted model and gray dots are predictions from the restricted model. Gross income above $300000 not shown.

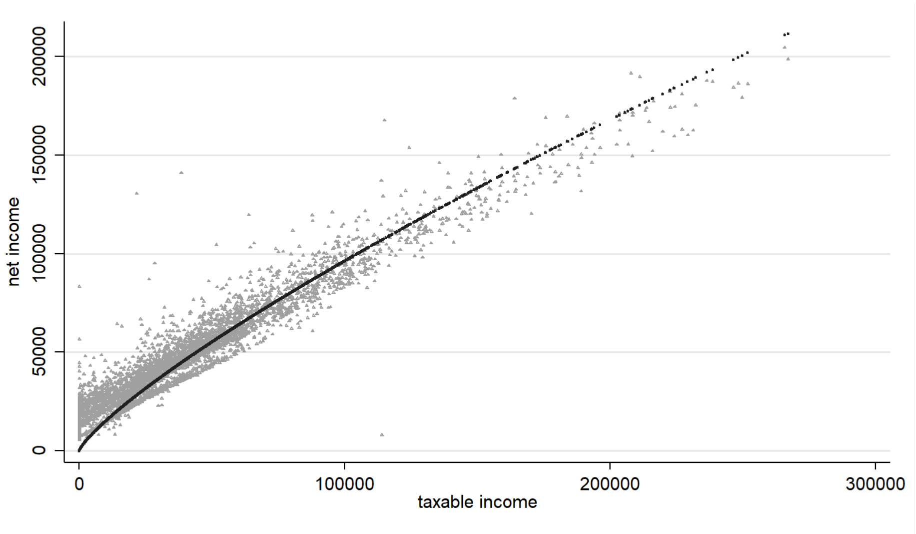
The central result of this section is that the progressivity parameter of the complete retention function is in the range of 0.89 to 0.95. These estimates are reassuringly close to the estimates found in Blundell et al. (2016a) and to estimates in Gouveia and Strauss (1994). Blundell et al. (2016a) estimate a progressivity parameter of roughly 0.9 and Gouveia and Strauss (1994) arrive at values between 0.92 and 0.95 for the period from 1979 to 1989.[17] These estimates diverge from the one provided by Heathcote et al. (2017a). Their estimate suggests more progressivity, with $1-\tau = 0.82$. Since this is ultimately the parameter that will adjust the Frisch-elasticity of labor supply the comparison is appropriate. I comment on the discrepancy below.

I give a more intuitive description of the fit for the complete model by calculating the fiscal gap implied by the gross-to-net approximation. This fiscal gap is the difference between the predicted and the observed tax liability based on taxsim $(T-\hat{T})$. I plot the fiscal gap over the distribution of gross income in figure 4.3 for the year 2000. As one would expect, there are a lot of observations in the left tail of the income distribution up to roughly $100000 of gross income that achieve an

[17] Blundell et al. (2016a) do not use the taxsim tax-calculator, but have rather programmed their own.




**(a)** Complete Retention Function Fit in 2001




**(b)** Partial Retention Function Fit in 2001

**Figure 4.2:** Partial and Complete Retention Function Fit in 2000

*Note*: Own calculation based on PSID (2015). Complete model shown in the upper and partial in the lower panel. The graphs plot gross/taxable income against net income. Light gray dots indicate observed values, black dots are model predictions. Gross income above $300000 not shown.

acceptable fit. This is also evidenced by the fit of the LOWESS-line[18] centering around zero. However, many observations do have overpredicted tax liabilities, which is clear from a somewhat strong spread into negative values at the very low

[18]The acronym LOWESS stands for locally weighted scatterplot smoothing, a local linear regression method.

end of the income distribution. Up to roughly $200000 residuals are centered on zero. Beyond this threshold the LOWESS-line bends up, which indicates that tax liabilities are underpredicted.

Further, to illustrate the impact on the budget of the state, I also calculate the cumulative and the average fiscal gap in every year. They are shown in the third- and second-to-last columns of table 4.3. Overall, the approximation imposes higher tax liabilities in every year compared to the taxsim values. The pattern over the years is not uniform. There are years with larger fiscal gaps, like 2004 and 2008, but the average overestimation of the fiscal gap is roughly $64 million. This number can be more easily interpreted when one considers the adjacent column, which contains the average fiscal burden. On average each tax unit must pay $650 more than under the actual tax schedule. The average tax liability over all years is roughly $17000, so that the approximation would impose an increase of 3.82% of the average tax liability.

**Table 4.3:** Retention Function

| | Partial Model | | Complete Model | | Absolute Fiscal Gap | | |
|---|---|---|---|---|---|---|---|
| Year | $\tilde{\chi}$ | $1-\tilde{\tau}$ | $\chi$ | $1-\tau$ | cumul. | avg. | Obs. |
| 1998 | 8.905 | 0.806 | 2.456 | 0.897 | -29588838 | -314 | 4535 |
| 2000 | 9.384 | 0.802 | 2.610 | 0.893 | -22463340 | -234 | 4878 |
| 2002 | 5.854 | 0.842 | 2.298 | 0.903 | -54895752 | -558 | 5104 |
| 2004 | 3.519 | 0.888 | 1.373 | 0.949 | -164579872 | -1671 | 5133 |
| 2006 | 6.526 | 0.836 | 2.121 | 0.912 | -49338128 | -490 | 5191 |
| 2008 | 4.086 | 0.876 | 1.692 | 0.931 | -97131440 | -955 | 5415 |
| 2010 | 6.892 | 0.833 | 2.318 | 0.905 | -36420912 | -362 | 5141 |
| 2012 | 4.128 | 0.876 | 1.775 | 0.928 | -80230176 | -831 | 5225 |
| 2014 | 6.973 | 0.829 | 2.401 | 0.901 | -45120448 | -435 | 5346 |
| mean | | | | | -64418767 | -650 | |

*Note*: Own calculation based on PSID (2015). Both models estimated using nonlinear least squares. The fiscal gap is calculated using predicted tax liabilities from taxsim and the complete model. Estimation and ancillary calculations were performed using cross-sectional frequency weights.

**Relationship to the Estimates of Heathcote et al. (2017a)** Finally, I want to explain why Heathcote et al. (2017a) find a different estimate for the progressivity parameter $1-\tau$. The clearest conceptual distinction to point out is that Heathcote et al. (2017a) measure statutory progressivity by regressing the log of net income minus deductions on the log of taxable income.[19] Even though the conceptual

[19] In their model, deductions, at least at the margin, are fixed and do not react to income changes.

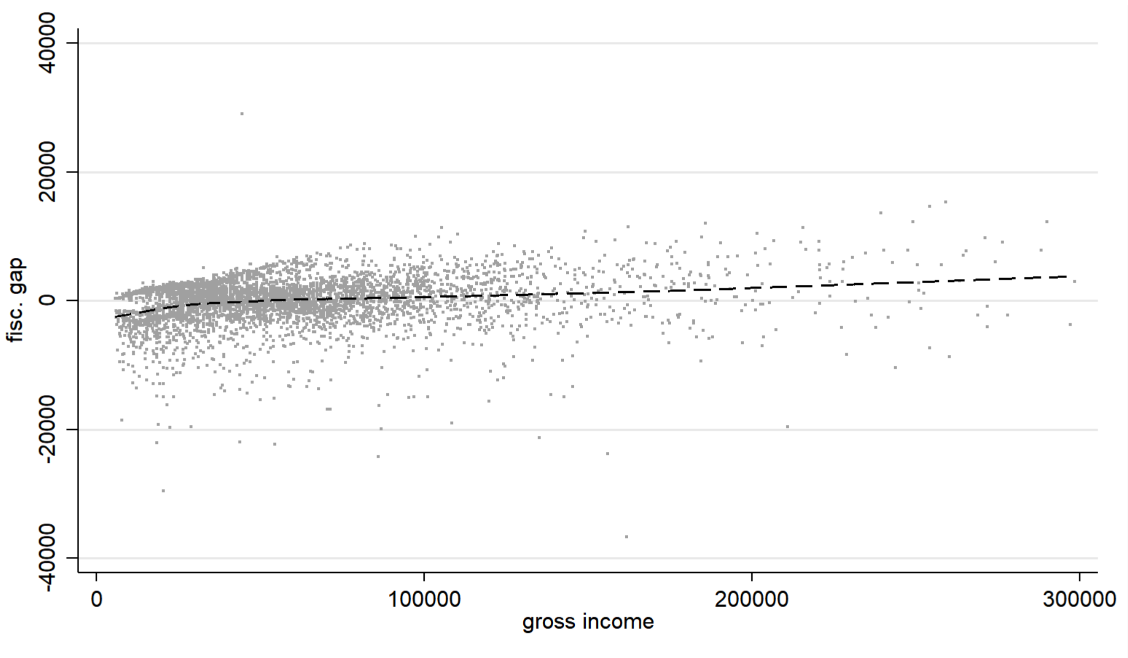



**Figure 4.3:** Fiscal Gap in 2000

*Note*: Own calculation based on PSID (2015). The graph plots the difference between approximated and actual tax liabilities over the distribution of gross incomes. Light gray dots are observed fiscal gaps, the black dashed line is LOWESS fit of the data. Gross income above $300000 not shown.

difference is large, empirically it appears to matter little. Rather, it is of crucial importance whether one estimates the model using nonlinear least-squares or with OLS after a log-transformation.

In appendix 4.11.5 I have run the log-specification and the equivalent nonlinear least-squares specification using the data that Heathcote et al. (2017a) provide for replication. I estimate both models pooled over the period 2000-2006. When using nonlinear least-squares I arrive at a much larger value of about 0.93 instead of 0.82 for the progressivity parameter. When comparing the RMSE of the residuals in levels, the log specification performs worse than the nonlinear one; the RMSE is slightly more than twice as large. This is because the log specification implicitly puts less weight on observations at the top end of the taxable income distribution and overpredicts their tax liabilities.

The reason why the estimates differ lies in the specification and not in the conceptual difference between effective or statutory progressivity, as my "effective" estimate of $\tau$ is very close to the "statutory" $\tau$ found using nonlinear least-squares using their data.[20] So, although I am convinced that it is appropriate to model the

[20] It can further be argued that their approximation function only partly captures statutory progressivity. Heathcote et al. (2017a) do include transfers in their measure of net income and therefore

effective progressivity following the arguments in section 4.4.3 and appendix 4.11.4, the importance of whether the effective or the statutory progressivity is modeled is ultimately of minor importance for the empirics.

### The Policy Maker's Influence over Progressivity

As discussed above, it is vital for policy making to have a measure of the impact a change in tax policy has on the progressivity parameters of the approximation. Therefore, I calculate the mechanical effect of policy changes on the parameters of the approximation determining progressivity.

I conduct two experiments using the non-standard options of the taxsim tax-calculator: 1) I increase the standard deduction available to the tax units by one percent. 2) I increase the top tax rate by one percent.

Following that, I calculate taxable and net income, and run the approximation procedures. I recover the new, counterfactual approximation parameters $\iota_t^c$ and $\tau_t^c$, and compute relative differences from the baseline in each year. These relative differences give the elasticity of the approximation parameters with respect to a change in the tax function, i.e. the desired mechanical effect of a change in the tax function. Table 4.4 lists the effects on $\iota$ and on $\tau$ of the two scenarios.

**Table 4.4:** Elasticities of the Approximation Parameters

| scenario | | 1998 | 2000 | 2002 | 2004 | 2006 |
|---|---|---|---|---|---|---|
| deduction | $\iota$ | 0.485 | 0.419 | 0.465 | 0.386 | 0.546 |
| | $\tau$ | 0.117 | 0.101 | 0.084 | 0.122 | 0.118 |
| top tax rate | $\iota$ | 0 | 0 | 0 | 0 | 0 |
| | $\tau$ | 1.380 | 1.424 | 1.725 | 2.730 | 1.558 |

| | | 2008 | 2010 | 2012 | 2014 | mean |
|---|---|---|---|---|---|---|
| deduction | $\iota$ | 0.499 | 0.815 | 0.579 | 0.564 | 0.529 |
| | $\tau$ | 0.103 | 0.135 | 0.113 | 0.107 | 0.111 |
| top tax rate | $\iota$ | 0 | 0 | 0 | 0 | 0 |
| | $\tau$ | 1.954 | 1.164 | 1.896 | 1.577 | 1.712 |

*Note*: Own calculation based on PSID (2015). Approximation models computed in the same way as in sections 4.6.1 and 4.6.1. Calculations performed using cross-sectional frequency weights. Scenario *deduction* corresponds to an increase of the standard deduction by one percent. Scenario *top tax rate* corresponds to an increase of the top tax rate by one percent.

count transfers as part of the statutory system. However, by regressing the log of of net income minus deductions on the log of taxable income, individuals with zero taxable income with positive net income achieved with transfers cannot affect the progressivity estimate as they are excluded from the regression. The same would hold if they estimated the relationship using nonlinear least-squares, since the function has to start in the origin. This problem does not arise when one directly maps from gross to net income, provided the sample is restricted to the active population.

The most noticeable feature of the results is that a change in the standard deduction affects both $\iota$ and $\tau$, but a change in the top tax rate does not affect $\iota$. But this is expected because a change in the top tax rate should not affect the mapping from gross to taxable.

In terms of the magnitude of the reactions, the picture is relatively uniform. A one percent increase in the standard deduction increases $\iota$ by half a percent. Since an increase in $\iota$ necessitates an increase in overall progressivity, $\tau$ rises by roughly a tenth of a percent. The increase of the top tax rate by one percent shows a much greater effect on overall progressivity. The change in $\tau$ in the mean is roughly 1.7 percent. However, the variability of the estimates is a lot higher in this scenario, which is likely driven by fluctuations in the size and the respective incomes of the group paying the top tax rate.

### 4.6.2 Tractability Results

#### Deriving the Labor Supply Equation

In the previous sections I have introduced the power function as a convenient approximation of the retention function. However, to make the approximation fully tractable with the first order approach of estimating labor supply, I need to make a further adjustment to allow for nonlabor income and time-differencing.

To allow for non-labor income $N_t$, I will adjust the approximation slightly and show that in terms of tractability of the first-order approach the adjustment has very desirable properties, while the cost in terms of fit is acceptable.

The life-cycle model is the same as described in section 4.3. I choose the following functional form for $\mathcal{T}_t(\cdot)$:

$$\begin{aligned} \mathcal{T}_t(w_t h_t, N_t) &\approx \tilde{\chi}_t \left(\kappa_t (w_t h_t)^{1-\iota_t}\right)^{1-\tilde{\tau}_t} + \tilde{\chi}_t \left(\kappa_t (N_t)^{1-\iota_t}\right)^{1-\tilde{\tau}_t} \\ &= \chi_t (w_t h_t)^{1-\tau_t} + \chi_t (N_t)^{1-\tau_t}, \end{aligned} \tag{4.15}$$

so that the budget constraint takes the form,

$$\frac{a_{t+1}}{(1+r_t)} = a_t + \chi_t (w_t h_t)^{1-\tau_t} + \chi_t (N_t)^{1-\tau_t} - c_t. \tag{4.16}$$

The additional assumption imposed in contrast to the approximations in 4.4.1 is that one can separately apply the power function approximation to labor earnings and non-labor income and arrive at the same net income. I call this the *additive approximation* and will discuss how appropriate this alternate choice is in terms of fit in section 4.6.2. The first order condition for choosing labor supply in this setting is

$$h_t^{\gamma} = \frac{1}{b_t}\lambda_t \chi_t (1-\tau_t)(w_t h_t)^{-\tau_t} w_t. \tag{4.17}$$

To make the estimation in growth rates, which eliminates $\lambda_t$, tractable, I posit the following lagged approximation:

$$\mathcal{T}_{t-1}(w_{t-1}h_{t-1}, N_{t-1}) \approx \hat{\chi}_{t-1}(w_{t-1}h_{t-1})^{1-\tau_t} + \hat{\chi}_{t-1}(N_{t-1})^{1-\tau_t}. \tag{4.18}$$

Here the degree of progressivity is set to the value of the upcoming period and the parameter $\hat{\chi}$ can freely adjust to fit the period $t-1$ distribution of post-government income. Again, the question whether it is possible to make this substitution depends on the fit of the approximation, which I discuss in section 4.6.2.

Divide (4.17) by $h_{t-1}^{\gamma}$, so that

$$\left(\frac{h_t}{h_{t-1}}\right)^{\gamma} = \frac{b_{t-1}}{b_t}\frac{\lambda_t}{\lambda_{t-1}}\frac{\chi_t}{\tilde{\chi}_{t-1}}\frac{(1-\tau_t)}{(1-\tau_t)}\left(\frac{w_t}{w_{t-1}}\frac{h_t}{h_{t-1}}\right)^{-\tau_t}\frac{w_t}{w_{t-1}}. \tag{4.19}$$

Taking logs I find that

$$\begin{aligned}
\gamma\Delta\ln h_t &= \Delta\ln\lambda_t + (\ln\chi_t - \ln\hat{\chi}_{t-1}) - \tau_t\Delta\ln h_t + (1-\tau_t)\Delta\ln w_t - \Delta\ln b_t \\
\Delta\ln h_t &= \frac{1}{\gamma+\tau_t}\left[\Delta\ln\lambda_t + (\ln\chi_t - \ln\hat{\chi}_{t-1}) + (1-\tau_t)\Delta\ln w_t - \Delta\ln b_t\right] \\
\Delta\ln h_t &= \frac{1}{\gamma+\tau_t}\left[cons_t + \Delta\ln\lambda_t + (1-\tau_t)\Delta\ln w_t - \varsigma\Delta\Xi_t + \Delta v_t\right].
\end{aligned} \tag{4.20}$$

The term $cons_t$ contains all the terms not varying in the cross-section. Finally, I obtain the estimating equation for labor supply by resolving the expression for the intertemporal difference in the log marginal utility of wealth. Everything in that term, except for the innovations $\eta_t$, are absorbed into the constant. I derive the approximation of the Euler equation that underlies this substitution in appendix 4.11.3. Then,

$$\Delta\ln h_t \approx \frac{1}{\gamma+\tau_t}\left[cons_t + (1-\tau_t)\Delta\ln w_t - \varsigma\Delta\Xi_t + \Delta v_t + \eta_t\right]. \tag{4.21}$$

This tax-modified Frisch elasticity can be estimated by IV techniques. Its estimation is the objective in section 4.7.1.

The innovations $\eta_t$ are purely a function of the permanent wage shocks $\zeta_{it}$. An approximation of the life-time budget constraint as described in Blundell et al. (2016a) reveals

$$\eta_{it} = -\varphi_{it}(1-\tau)\left(1 + \frac{1-\tau_t}{\gamma+\tau_t}\right)\zeta_{it}, \quad \varphi_{it} \sim LN\left(\mu_{\phi}, \sigma_{\phi}^2\right). \tag{4.22}$$

$\varphi_{it}$ is a transmission parameter measuring how permanent wage shocks affect $\eta_{it}$. It is individual specific because it is mainly driven by the ratio of wealth to total wealth including human wealth (Jessen and König, 2018). Further, the parameter is necessary to compute the Marshallian elasticity of labor supply. Since I do not use consumption or asset data, I choose a log-normal distribution with the underlying parameters $\mu_\varphi$ and $\sigma_\varphi$ to model its distribution, as $\varphi_{it}$ can only take on non-negative values.

### Additive Approximation

The main concern regarding the use of the additive formulation of the tax function approximation is whether it restricts the goodness of fit or whether it gravely alters the parameter estimates relevant for the calculation of progressivity. This is not the case. The additive formulation neither has a major impact on the parameter estimates for $\iota$ and $\tau$, nor does it gravely change the goodness of fit. Unfortunately it is not possible to conduct a direct, formal test because the models are not nested. However, it is obvious from the results in table 4.5 that the two approximations deliver very similar results in terms of coefficients and goodness of fit.

The table's contents are reassuring: The parameters that determine progressivity, $\iota$ and $\tau$, are exceptionally close in almost every year. Despite the major alteration of the approximation, my results in terms of the effect of taxation on labor supply are bound to be comparable to the rest of the literature. The largest differences occur in 2014, where $1-\iota$ is roughly one percent smaller and $1-\tau$ roughly one percent larger in the additive model. The multiplicative parameters, $\kappa$ and $\chi$, show larger differences and more variability, but this will not impact the estimation of the Frisch elasticity.

Although goodness of fit in terms of the RMSE does not have a uniform pattern, it is very close in almost all years. In some years, like 1998 and 2008, the additive deduction model even dominates the original, while in 2004 and 2008 the additive, complete model dominates the original. The overall impression is that the original model has better fit in most years. However, the difference is slight: the biggest relative difference in the deduction model is recorded in 2002, where the RMSE is 2.6 percent larger in the additive model. The largest relative difference in the complete model also occurs in 2002; a difference of 6.7 percent. Generally, one can observe that relative differences in the RMSE in the complete model are larger, clustering in the range of 0.3 to 6.2 percent, excluding 2004 and 2008. In sum, I can conclude that the additive model does almost as good of a job of fitting taxable and net income as the original model.

**Table 4.5:** Additive Tax Function

Taxable Income Model

| | Additive Model | | Original Model | | RMSE | | | |
|---|---|---|---|---|---|---|---|---|
| Year | $1-\iota$ | $\kappa$ | $1-\iota$ | $\kappa$ | add. | orig. | rel. diff. | Obs. |
| 1998 | 1.090 | 0.275 | 1.091 | 0.263 | 7692 | 7773 | -0.010 | 4535 |
| 2000 | 1.094 | 0.263 | 1.096 | 0.245 | 8407 | 8342 | 0.008 | 4878 |
| 2002 | 1.056 | 0.414 | 1.062 | 0.373 | 8496 | 8283 | 0.026 | 5104 |
| 2004 | 1.071 | 0.344 | 1.069 | 0.339 | 10676 | 10427 | 0.024 | 5133 |
| 2006 | 1.069 | 0.351 | 1.076 | 0.310 | 8786 | 8605 | 0.021 | 5191 |
| 2008 | 1.057 | 0.400 | 1.057 | 0.388 | 8717 | 8819 | -0.012 | 5415 |
| 2010 | 1.066 | 0.353 | 1.073 | 0.317 | 8329 | 8313 | 0.002 | 5141 |
| 2012 | 1.051 | 0.430 | 1.058 | 0.388 | 7652 | 7595 | 0.008 | 5225 |
| 2014 | 1.064 | 0.372 | 1.073 | 0.325 | 7783 | 7613 | 0.022 | 5346 |

Retention Function Model

| | Additive Model | | Original Model | | RMSE | | | |
|---|---|---|---|---|---|---|---|---|
| Year | $1-\tau$ | $\chi$ | $1-\tau$ | $\chi$ | add. | orig. | rel. diff. | Obs. |
| 1998 | 0.899 | 2.277 | 0.897 | 2.456 | 6587 | 6582 | 0.001 | 4535 |
| 2000 | 0.895 | 2.402 | 0.893 | 2.610 | 13230 | 13222 | 0.001 | 4878 |
| 2002 | 0.911 | 2.007 | 0.903 | 2.298 | 4362 | 4089 | 0.067 | 5104 |
| 2004 | 0.946 | 1.379 | 0.949 | 1.373 | 6925 | 7177 | -0.035 | 5133 |
| 2006 | 0.919 | 1.868 | 0.912 | 2.121 | 4754 | 4505 | 0.055 | 5191 |
| 2008 | 0.933 | 1.609 | 0.931 | 1.692 | 4589 | 4626 | -0.008 | 5415 |
| 2010 | 0.913 | 2.031 | 0.905 | 2.318 | 4369 | 4245 | 0.029 | 5141 |
| 2012 | 0.936 | 1.566 | 0.928 | 1.775 | 4639 | 4485 | 0.034 | 5225 |
| 2014 | 0.911 | 2.048 | 0.901 | 2.401 | 4544 | 4391 | 0.035 | 5346 |

*Note*: Own calculation based on PSID (2015). Models estimated using nonlinear least squares. Estimation and ancillary calculations were performed using cross-sectional frequency weights. Relative difference between RMSEs calculated using this expression: $\frac{RMSE^{add}-RMSE^{orig}}{RMSE^{orig}}$.

#### Approximation of Lagged Net Income

The final issue to resolve is the evaluation of the fit of the tax approximation when restricting the progressivity parameter $1-\tau$ in the approximation of the $t-1$-tax-system to the value estimated in $t$. In this restricted model only $\chi$ can adjust. The corresponding unrestricted model is the additive, effective model shown in table 4.5. I present the RMSE of both the restricted and the unrestricted model in table 4.6 along with a linear one for comparison. The linear model is a natural benchmark, since it restricts the effect of taxation on labor supply to be nil.

At first glance table 4.6 shows that the restricted model generates noticeable differences in terms of fit in years like 2002 and 2012 compared to the unrestricted model. However, since changes in progressivity between years cannot be fully accounted for by letting only $\chi$ adjust, this loss in fit is expected. When compared to the linear model, in which $\tau$ is set to zero, the restricted still compares favorably. The loss of fit is primarily an issue in the upper tail of the distribution. In figure 4.4 I plot both the restricted and the unrestricted model predictions against the observed values for the year 2002, which has the worst relative gap in RMSE. It is clear from the figure that the most severe deviations from the unrestricted model only occur after levels of roughly \$200000 gross income. Below this threshold both models make very similar predictions. While the approximation with the imposed leading progressivity parameter is certainly not preferable to the unrestricted model, it does not lead to a major reduction in terms of goodness of fit. Further, it still handily outperforms the linear model. Accordingly, imposing this relationship implicitly when I estimate the labor supply equation (4.21) is a justifiable sacrifice to attain tractability.

## 4.7 Results

I now turn to the empirical implementation: First, I estimate both the tax-modified and the unmodified Frisch elasticity of labor supply and the residuals of the wage and hours equations. Second, I estimate the permanent and transitory shock variances of the wage process and finally the Marshallian elasticity of labor supply.

### 4.7.1 First Stage - Labor Supply

To obtain the residuals of the wage equation (4.4), I perform a first-differenced regression of the log of hourly wages on indicator variables for calendar years, states, industries, occupations, number of children in the household, and race. Further, I include a second order polynomial of years of education and its interaction with age. This last set of variables is also used as the set of excluded instruments for wages in the labor supply estimation.

**Table 4.6:** Fit of the Retention Function with Lead of Progressivity Parameter

| | RMSE | | | rel. diff. of RMSE | | | |
|---|---|---|---|---|---|---|---|
| Year | restr. | unres. | linear | restr. vs. unres. | unres. vs. lin. | restr. vs. lin. | Obs. |
| 1998 | 6587 | 6595 | 10159 | 0.001 | -0.352 | -0.351 | 4535 |
| 2000 | 13230 | 13293 | 15791 | 0.005 | -0.162 | -0.158 | 4878 |
| 2002 | 4362 | 6148 | 11837 | 0.409 | -0.631 | -0.481 | 5104 |
| 2004 | 6925 | 8300 | 11260 | 0.199 | -0.385 | -0.263 | 5133 |
| 2006 | 4754 | 4950 | 9845 | 0.041 | -0.517 | -0.497 | 5191 |
| 2008 | 4589 | 5359 | 10603 | 0.168 | -0.567 | -0.495 | 5415 |
| 2010 | 4369 | 4779 | 8699 | 0.094 | -0.498 | -0.451 | 5141 |
| 2012 | 4639 | 5995 | 10671 | 0.292 | -0.565 | -0.438 | 5225 |

*Note*: Own calculation based on PSID (2015). Models estimated using nonlinear least squares. Estimation and ancillary calculations were performed using cross-sectional frequency weights. Relative difference between RMSEs calculated using these expression from left to right: $\frac{RMSE^{restr.}-RMSE^{unres.}}{RMSE^{unres.}}$, $\frac{RMSE^{restr.}-RMSE^{lin.}}{RMSE^{lin.}}$, $\frac{RMSE^{unres.}-RMSE^{lin.}}{RMSE^{lin.}}$.

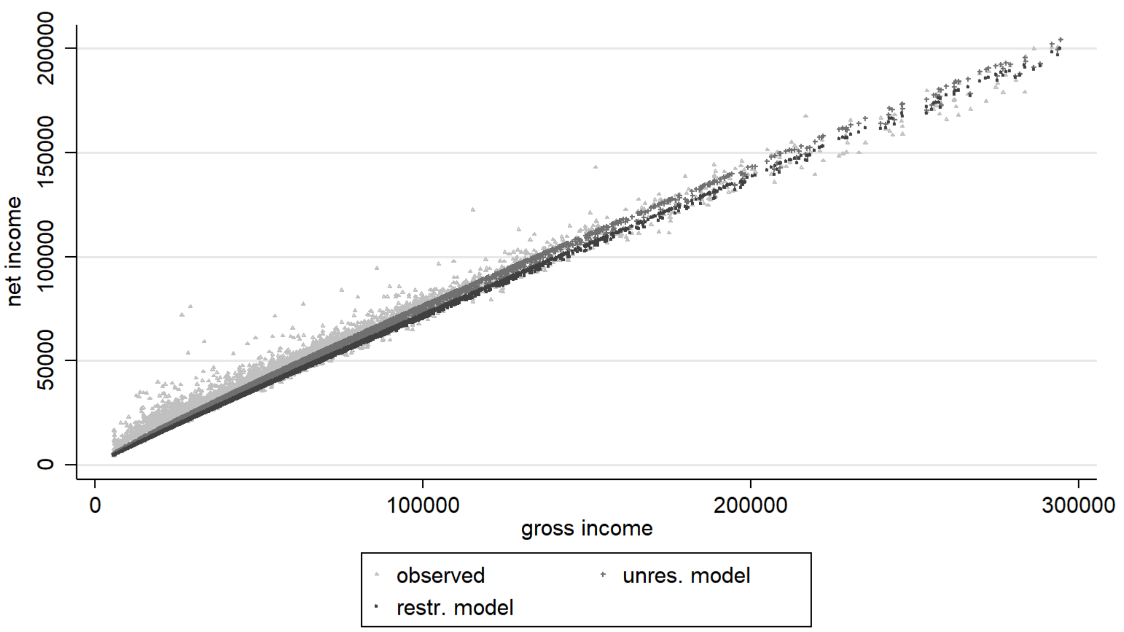



**Figure 4.4:** Fit of Restricted and Unrestricted Model in 2002

*Note*: Own calculation based on PSID (2015). The graph plots gross income against net income. Light gray dots indicate observed values, black dots are predictions based on the unrestricted model and gray dots are predictions from the restricted model. Gross income above $300000 not shown.

Now I specify the observable portion of the taste-shifter $\Xi_t$ in the labor supply equation (4.21). Again, I include indicators for years, states, industries, occupations, number of children in the household, and race. I first estimate equation (4.21) pooled over all years to find an estimate of the average tax-modified Firsch elasticity. It is an average over the years because the tax-modified elasticity varies with $\tau$, which in turn changes year to year. I use the average progressivity parameter over the years to back out the unmodified Frisch elasticity $1/\gamma$, as it does not vary over the years. The average of $1-\tau$ over the years relevant for estimation (2000-2014) can be determined from the pooled estimation of the additive model of equation (4.15), which is $1-\bar{\tau} = 0.925$ .

I display the estimated tax-modified and the unmodified Frisch elasticity along with the average progressivity parameter in table 4.7.

**Table 4.7:** Regression Results

| | $1-\bar{\tau}$ | $(1-\bar{\tau})/(\gamma+\bar{\tau})$ | $1/\gamma$ |
|---|---|---|---|
| | 0.925 | 0.469 | 0.528 |
| | | (0.061) | (0.071) |
| Kleibergen Paap F Stat. | | 64.56 | |
| Obs. | | 30558 | 30558 |

*Note:* Own calculation based on PSID (2015). Robust standard errors in parentheses. $(1-\bar{\tau})/\gamma+\bar{\tau}$ is the tax-modified Frisch elasticity and $1/\gamma$ is the regular Frisch-elasticity. Observations are person-years.

The estimated tax-modified Frisch elasticity of labor supply is about 0.47 and the unmodified elasticity is slightly larger with a point estimate of 0.528. Both are statistically significant at conventional levels and the F statistic according to Kleibergen and Paap (2006) does not indicate a weak-instrument issue. That unmodified Frisch elasticities are larger than tax-modified ones is known in the literature and expected. The tax-modified elasticity gives the response to a transitory wage change before taxes, while the unmodified Frisch gives the response after-tax. The tax-modified elasticity is adjusted for the disincentive effect of the progressive tax system, which lowers individuals' responsiveness to a change in the wage. My result for the unmodified Frisch elasticity is in line with the estimates presented in Heathcote et al. (2014a) (0.462) and in Blundell et al. (2016a) (0.681).[21] That the estimates agree fairly well is encouraging as both studies employed the same data source and similar portions of the data, but they did not use the same method to estimate the unmodified Frisch elasticity. Rather, both use the method of moments applied to residual variances to estimate it. Note, however, that in Blundell et al. (2016a) the second earner is explicitly considered and, therefore, the structural equations on which the estimation rests are quite distinct from mine.

[21] Heathcote et al. (2017a) use the unmodified Frisch elasticity estimated in Heathcote et al. (2014a) for the calibration of their model.

The difference between the tax-modified and the unmodified elasticity is rather small and most likely not statistically significant in spite of the small standard errors. The unmodified elasticity is roughly four percent larger than the tax-modified counterpart. In comparison, it is roughly 17 percent larger for Blundell et al. (2016a). But even in that study the increase is most likely not statistically significant.

The residuals for the second stage of the estimation are obtained from estimating equation (4.21) year by year to account for the time-dependence of the tax-modified Frisch elasticity.

### 4.7.2 Second Stage - Wage Variance Process

I estimate the stochastic process for wages described in 4.5. I provide an example of the identification for the process with t=3 in appendix 4.11.6. As I discussed in section 4.3, it is necessary to set an initial condition for the transitory process in period zero, so I restrict the innovation variance for the transitory shock in period 0 to have the same value as in period 1, $\sigma^2_{\epsilon,0} = \sigma^2_{\epsilon,1}$. Further, as shown in appendix 4.11.6, I can only identify all the parameters of interest up to period $t-1$ if $t$ periods are available.

**Method of Moments** I estimate the wage process using the method of moments with a unit weighting matrix. Let the set of parameters of interest be denoted by $\Sigma$, so that it contains all the 7 permanent shock variances $\sigma^2_\zeta$, 7 transitory shock variances $\sigma^2_\epsilon$, and the persistence parameter $\theta$. Then the minimization program for the method of moments is given by

$$\underset{\Sigma}{\text{minimize}}\ [\mathbf{m}(\Sigma)-\mathbf{m}^e]'\mathbf{I}[\mathbf{m}(\Sigma)-\mathbf{m}^e], \tag{4.23}$$

where $\mathbf{m}(\Sigma)$ is the vector of theoretical autocovariance moments of $\Delta\omega_{it}$ and $\mathbf{m}^e$ is the observed counterpart. By choosing the identity matrix $\mathbf{I}$ as the weighting matrix, I minimize the squared sum of deviations between the observed and theoretical moments.[22] I calculate standard errors for the parameters using the block bootstrap method. I draw 200 boostrap replications of the data.[23]

I present the estimates of the persistence parameter and the standard deviations of the two shock types in table 4.8.

The time trend of the standard deviations of the two shock types is striking. The size of permanent shocks to the wage process increases steadily until the financial crisis of 2008 hits. In 2008 the size of permanent shocks is extraordinarily large and

[22] Altonji and Segal (1996) show that the identity weighting matrix is generally preferable for the estimation of autocovariance structures using panel data.

[23] This accounts for various issues that would otherwise affect the more conventional Delta-method standard errors, which include the use of estimates for the variance of the measurement error, heteroskedasticity and serial correlation.

**Table 4.8:** Wage Process

| | 2000 | 2002 | 2004 | 2006 | 2008 | 2010 | 2012 | all |
|---|---|---|---|---|---|---|---|---|
| $\sigma_{\zeta,t}$ | 0.1853 | 0.2011 | 0.1970 | 0.2194 | 0.2503 | 0.2503 | 0.2374 | |
| | (0.0028) | (0.0018) | (0.0014) | (0.0014) | (0.0011) | (0.0011) | (0.0009) | |
| $\sigma_{\epsilon,t}$ | 0.1728 | 0.2688 | 0.1900 | 0.1055 | 0.1041 | 0.1372 | 0.0615 | |
| | (0.0019) | (0.0016) | (0.0025) | (0.0023) | (0.0022) | (0.0018) | (0.0030) | |
| $\theta$ | | | | | | | | -0.0329 |
| | | | | | | | | (0.0076) |

*Note:* Own calculation based on PSID (2015). Bootstrap standard errors based on 200 replications in parentheses.

remains at this level in 2010 with a slight recovery in 2012. To contrast, the increase in the standard deviation of permanent shocks from 2000 to 2006 is roughly 14 percent, while the increase from 2006 to 2008 is, again, 14 percent.

Transitory shocks show a completely different intertemporal pattern. While the size of the standard deviation is comparable to the permanent counterpart in the first couple of years, transitory shock size almost halves in 2006. This downtrend is not heavily impacted by the financial crisis of 2008. Even in 2012 transitory shock size appears to decrease further rather than tend back to pre-crisis levels.

This pattern is interesting on its own as it suggests that permanent, partially uninsurable wage risk has increased for the active population and not just due to the impact of the financial crisis. The pattern of rising partially uninsurable wage risk is also documented in Table E1 of Heathcote et al. (2014b). Even though they do not estimate the process I have chosen, they do report close analogues, namely the uninsurable, island-level shock variances. They estimate their process until 2006 and report a standard deviation of 0.1236 in 2000 for these shocks and 0.1378 in 2006 when frequency-adjusted.[24] This implies an increase of the standard deviation from 2000 to 2006 of about 11 percent, showing that the relative trend lines up in both sets of results.

The results in Blundell et al. (2016a) are harder to compare because there is explicit consideration of a secondary earner, shocks are allowed to be correlated across primary and secondary earner, variances are calculated with respect to the age group and not the calendar year, and their sample period runs from 1998 to 2008. Still, their average figure for the permanent shock standard deviation primary earners face is 0.1741 and, considering all these caveats, fairly close to my estimates for the pre-crisis period.

[24]Heathcote et al. (2014a) assume a distinction between insurable and uninsurable shocks from the outset and rationalize it using an island-structure for the shock process that agents face (Attanasio and Rios-Rull, 2000). I add the two shock variances reported in Heathcote et al. (2014b) for the two years that are pooled in my dataset to account for the frequency difference.

### 4.7.3 Marshallian Elasticity

Finally, I use the autocovariance moments of the hours residuals and the covariance moments with the wage residuals to estimate the parameters of $\varphi$, $\mu_\varphi$ and $\sigma^2_\varphi$.[25] Consistent with the estimation procedure above, this is done by the method of moments choosing the identity matrix as the weighting matrix. The reader should note that the parameter $\varphi$ does depend on $\tau_t$ and so the estimation delivers values for the parameters $\mu_\varphi$ and $\sigma^2_\varphi$ that let me calculate the mean of $\varphi$ for the average degree of progressivity over the years. In section 4.8 I recalculate the mean of $\varphi$ under the assumption that $\tau$ is zero, which then in turn enables me to calculate the mean of $\varphi$ at all different $\tau_t$. I show the results of the estimation in table 4.9.

**Table 4.9:** Hours Process

| $\mu_\varphi$ | $\sigma_\varphi$ | $E[\varphi]$ | $\bar{\kappa}$ |
|---|---|---|---|
| -0.2850 | 0.7776 | 1.0338 | -0.2405 |
| (0.0148) | (0.008) | (0.0099) | (0.0068) |

*Note:* Own calculation based on PSID (2015). Bootstrap standard errors based on 200 replications in parentheses. $E[\varphi]$ denotes the average pass-through parameter for permanent shocks over time. $\bar{\kappa}$ is the average Marshall elasticity of labor supply calculated using $E[\varphi]$ and $\bar{\tau}$.

The implied estimate for the average Marshallian elasticity across the years can be calculated from the following formula implied by (4.21) and (4.22),

$$\bar{\kappa} = \frac{1-\bar{\tau}}{\gamma+\bar{\tau}}\left(1 - E[\varphi]\left(1 + \frac{1-\bar{\tau}}{\gamma+\bar{\tau}}\right)\right). \tag{4.24}$$

The estimate in table 4.9 is negative and statistically significant with a point estimate of -0.2405. This estimate is larger in absolute terms than the one shown in Blundell et al. (2016a); that being -0.08. However, their 95-percent confidence band does overlap with mine and their result is most likely driven by the more comprehensive model with a second earner and non-separable preferences. This follows, because I get a very similar result for the Marshallian elasticity compared to mine if I plug their baseline estimates into my formula for the Marshallian elasticity (about -0.27). Accordingly, the difference between the two estimates is not much of a surprise, as some of the neglected issues, like correlated shocks between primary and secondary earner, are bound to be picked up by my estimate.

[25] Further, I estimate the variances of the innovations to the taste-shifter $b$ that moderates the disutility from work. Again, an initial condition needs to be chosen: I assume that the zeroth and first innovation variance are of the same magnitude. I display these in Appendix 4.11.7.

## 4.8 Insurance of Earnings Risk

### 4.8.1 Calculating Pass-Through

In the following I provide a calculation of the amount of insurance offered by progressive taxation against the risk stemming from the stochastic process underlying wages. In particular, I quantify how much of a given permanent wage shock transfers onto hours and subsequently earnings.[26] The following decomposition of the pass-through coefficient of a permanent shock to earnings is:

$$\frac{\partial \Delta \ln y_t}{\partial \zeta} \approx (1-\tau_t)\left(1+\frac{\partial \Delta \ln h_t}{\partial \zeta}\right) \tag{4.25}$$

Using the structural equations (4.4),(4.21) and (4.22) this expands to,

$$\frac{\partial \Delta \ln y_t}{\partial \zeta} \approx (1-\tau_t)\left(1+\frac{1-\tau_t}{\gamma+\tau_t}\left(1-E[\varphi]\left(1+\frac{1-\tau_t}{\gamma+\tau_t}\right)\right)\right). \tag{4.26}$$

Earnings react with one plus the Marshall elasticity to a given shock and the total response is dampened by the progressivity parameter. If the tax and transfer system did not feature progressivity, the impact on earnings would be

$$\left.\frac{\partial \Delta \ln y_t}{\partial \zeta}\right|_{\tau=0} \approx \left(1+\frac{1}{\gamma}\left(1-E\left[\varphi|_{\tau=0}\right]\left(1+\frac{1}{\gamma}\right)\right)\right). \tag{4.27}$$

$\varphi|_{\tau=0}$ is the transmission parameter, that depends on the unmodified Frisch elasticity $1/\gamma$.[27]

I calculate the insurance due to the progressive tax system by relating the impact of a given shock with and without the progressive tax system. A natural shock-size to pick is the standard deviation of the permanent shock in that period because it communicates how risky a given period is for someone experiencing the average

[26] Authors often quantify the consumption response as well. However, since I don't use consumption data, this is not possible. The consumption response is

$$\frac{\partial \Delta \ln c_t}{\partial \zeta} \approx \frac{\partial \Delta \ln y_t}{\partial \zeta} - \frac{\partial \Delta s/y}{\partial \zeta},$$

where $\frac{\partial \Delta s/y}{\partial \zeta}$ is the savings response and $s/y$ is the average propensity to save.

[27] The conversion is straightforward.

$$E[\varphi] = \frac{1-\pi}{1/\vartheta + (1-\pi)\frac{1-\bar{\tau}}{\gamma+\bar{\tau}}}, \tag{4.28}$$

shock in absolute terms. The average insurance value over all years, shown in table 4.10, can be calculated by using the average progressivity parameter and the average standard deviation of the permanent shock.

**Table 4.10:** Average Insurance of Earnings

| $1-\bar{\tau}$ | $\bar{\sigma}_\zeta$ | $\frac{\partial \Delta \ln y_t}{\partial \zeta}\bar{\sigma}_\zeta$ | $\left.\frac{\partial \Delta \ln y_t}{\partial \zeta}\right\vert_{\tau=0}\bar{\sigma}_\zeta$ | %-reduction |
|---|---|---|---|---|
| 0.925 | 0.22 | 0.1538 | 0.1631 | 5.7 |

*Note:* Own calculation based on PSID (2015).

A shock of size 0.22 is attenuated by roughly 30 percent, through both the tax system and the labor supply reaction because the Marshallian is negative. When the tax system offers no insurance, the attenuation is only about 25 percent. Therefore, insurance offered through progressive taxation is roughly 5.7 percent. When I shut down the labor supply reaction ($\frac{\partial \Delta \ln h_t}{\partial \zeta} = 0$), progressive taxation is the only source of insurance, and the percentage reduction of the shock equals $\bar{\tau}$, so 7.5 percent. The degree of insurance offered by progressive taxation is attenuated by the labor supply reaction.

I can now calculate the year-specific impact of a permanent wage shock $\left.\frac{\partial \Delta \ln y_t}{\partial \zeta}\right\vert_{\tau_t}$ and the year-specific degree of insurance. I show these values in table 4.11.

Table 4.11 relates two important facts. First, the amount of pass-through from permanent wage shocks to income has grown over time and therefore the amount of insurance offered by the tax system has waned. However, the change in the pass-through coefficient is relatively small. The highest pass-through is recorded in 2004 and the lowest in 2000, while overall growth from 2000 to 2012 was about 3 percent. This leads me to conclude that the pass-through coefficient has been rather stable over time even through the crisis. Second, and in stark contrast to the time series of the pass-through coefficient: risk, measures by the standard deviation of permanent wage shocks, has grown quite substantially. $\sigma_{\zeta,t}$ started at a level of 0.18, grew to a peak of 0.25 during the crisis and fell to 0.23 in 2012; total growth from 2000 to 2012 is about 28 percent. Since the pass-through parameter remained roughly constant, shocks passed on to earnings follow the trend in wage shocks. Pass-through adjusted risk grew from 0.12 in 2000 to 0.16 in 2012, a 33% increase. In sum, permanent earnings risk rose quite drastically, but the tax system, as captured by $\tau$, did not undergo major alterations after that.

where $\pi$ is the mean of the ratio of assets to total wealth, which is the sum of assets and human wealth. Then

$$E\left[\varphi|_{\tau=0}\right] = 1/\left(1/\varphi + \frac{(1+\gamma)\bar{\tau}}{\gamma(\gamma+\bar{\tau})}\right). \tag{4.29}$$

**Table 4.11:** Earnings Pass-Through Values and Insurance by Year

| | 2000 | 2002 | 2004 | 2006 | 2008 | 2010 | 2012 |
|---|---|---|---|---|---|---|---|
| $\frac{\partial \Delta \ln y_t}{\partial \zeta}\Big\vert_{\tau_t}$ | 0.6820 | 0.6912 | 0.7113 | 0.6962 | 0.7037 | 0.6924 | 0.7054 |
| $\sigma_{\zeta,t}$ | 0.1853 | 0.2011 | 0.1970 | 0.2194 | 0.2503 | 0.2503 | 0.2374 |
| $\frac{\partial \Delta \ln y_t}{\partial \zeta}\Big\vert_{\tau_t} \sigma_{\zeta,t}$ | 0.1264 | 0.1390 | 0.1401 | 0.1527 | 0.1761 | 0.1733 | 0.1675 |
| %-reduction from $\tau = 0$ | 7.3 | 6.0 | 3.3 | 5.4 | 4.4 | 5.9 | 4.1 |

*Note:* Own calculation based on PSID (2015). The last line represents the additional insurance above the level of $\frac{\partial \Delta \ln y_t}{\partial \zeta}\Big\vert_{\tau=0}$.

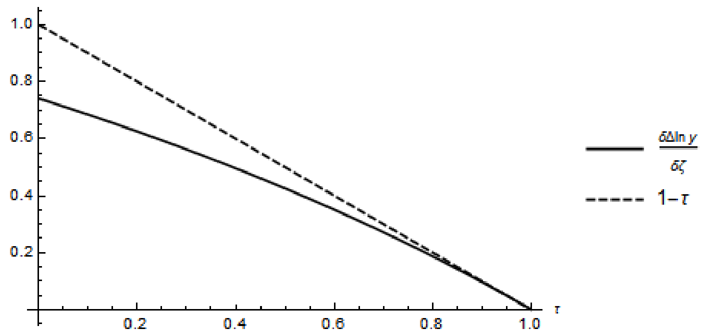
To characterize the relationship between $\tau$ and the pass-through coefficient to earnings, I display figure 4.5.

The upper panel compares the pass-through to earnings as calculated above and the pass-through when the labor supply reaction is set to zero, while the lower panel characterizes the labor supply reaction by graphing the Marshallian elasticity over the range of $\tau$. At the origin $\tau = 0$, so that the pure labor supply reaction can be seen on the abscissa. As $\tau$ and therefore progressivity increase, pass-through is diminished and insurance increases. The initial rise is slow, as the labor supply reaction runs counter to the new insurance offered through the tax system. However, at very high levels of progressivity, the labor supply reaction becomes less and less important, as the margin to respond becomes thinner and thinner. This leads to the convergence of the two curves at $\tau = 1$. This can also be verified in the lower panel. The labor supply reaction, i.e. the Marshallian elasticity, increases with $\tau$, so that it becomes more muted. Finally the reaction is zero at $\tau = 1$.

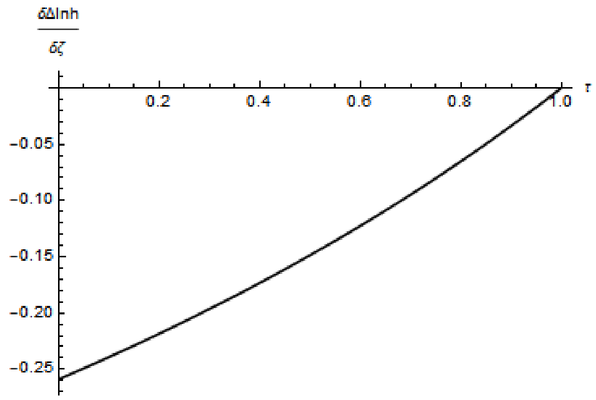
## 4.8.2 Stabilizing Earnings Risk

I have shown the rise in permanent earnings risk over the early 2000s and that progressive taxation in terms of the progressivity parameter $1 - \tau$ played a minor role in shaping how this risk transferred onto earnings. Rather, I find that the labor supply response was of primary importance for the pass-through.

The following is an illustration of the counterfactual approach to tax policy evaluation using the methods of section 4.6.1. I calculate the levels of the progressivity parameter $1 - \tau$ and the adjustment to the top tax rate in each year that would have resulted in holding the level of earnings risk ($\frac{\partial \Delta \ln y_t}{\partial \zeta}\Big\vert_{\tau_t} \sigma_{\zeta,t}$) constant at the average




(**a**) Earnings Pass-Through




(**b**) Marshallian Elasticity

**Figure 4.5:** Earnings Pass-Through and Marshallian Elasticity as Functions of $\tau$

*Note*: Own calculation based on PSID (2015). Shows the pass-through coefficient $\frac{\partial \Delta \ln y_t}{\partial \zeta}$ at different levels of $\tau$ and the pass-through if there were no labor supply response, i.e. $1-\tau$. Also shows the and the Marshallian elasticity $\kappa$ as a function of $\tau$.

level over the sample period, which is about 0.1536.[28] The top tax rate is paid by a small fraction of tax units in the dataset, ranging from 655 to 1990 cases (frequency weights applied) in the sample period. This exercise is supposed to illustrate the influence of the policy maker in shaping the risk experienced by individuals. I show the calculations in table 4.12.

At an average passed-on earnings risk of 0.1536, the period most closely resembling this level of risk is 2006, which – necessarily – is the period with the smallest implied change to the top tax rate. All previous periods had lower values of risk and therefore, to achieve stabilization, progressivity has to be decreased by cutting rates quite substantially. In contrast, the periods after 2006 imply changes toward higher progressivity than is observed. The implied progressivity parameter in this period hovers around 0.8, which comes with a strong increase in the top tax rate. However,

[28]Note that this exercise is not budget neutral

**Table 4.12:** Calculation of Stabilizing Progressivity and Tax Change

| | 2000 | 2002 | 2004 | 2006 | 2008 | 2010 | 2012 |
|---|---|---|---|---|---|---|---|
| $\frac{\partial \Delta \ln y_t}{\partial \zeta}\Big\rvert_{\tau_t} \sigma_{\zeta,t}$ | 0.1264 | 0.1390 | 0.1401 | 0.1527 | 0.1761 | 0.1733 | 0.1675 |
| $1-\tau_t^{\text{avg. risk}}$ | 1.1661 | 1.0412 | 1.0709 | 0.9263 | 0.7808 | 0.7808 | 0.8356 |
| %-change of top tax rate to reach $\tau_t^{\text{avg. risk}}$ | -47.81 | -29.39 | -45.6 | -1.84 | 56.38 | 38.02 | 36.91 |

*Note:* Own calculation based on PSID (2015). Last line calculated using the year-specific percentage changes that a 1% change in the top tax rate induces.

all the implied changes are to be taken with a grain of salt, as I have calculated only what would be implied for the top tax rate holding all other aspects of the tax system fixed. Certainly, decreases of the top tax rate would be tied to lowering some or all the other rates as well, if a certain change resulted in the top rate falling below the second-highest or other lower rates. However, I cannot take this into account.[29]

However, the broad picture in considering the changes around the crisis is clear. The financial crisis of 2008 was accompanied by a substantial rise in permanent wage and earnings risk. To mitigate this risk and keep it at the level of 2006, the state would have had to increase progressivity drastically during the crisis. For example, the relative difference between the $1-\tau_{2008}$ and $1-\tau_{2008}^{\text{avg. risk}}$ is roughly -16.3%. However, the observed pattern of progressivity during the crisis is the exact opposite. Progressivity generally decreased or at least stayed above the level of 2006, implying that the state did not substantially react to this rise in the riskiness of earnings by altering the tax system.

## 4.9 Qualifications and Extensions

The current model features many of the issues as the one of chapter 3. Namely, it does not consider the retirement-phase of the life-cycle, there is no joint decision-making in the household and the second earner is not explicitly modeled, and agents know the type and parametrization of their shock processes. The last issue has now become more pressing as I am assuming that the agents know the evolution of the variances over time. A fundamental question to ask in that regard is whether agents

[29] However, as the parameter $\tau$ shapes the retention function globally, the new system after the policy change does generally imply a change for individuals other than those paying the top tax rate. So in this sense the new approximated system corresponds to a certain tax schedule with lowered tax liabilities for tax units at the lower end of the income distribution.

can use signals to inform themselves about the moments of the stochastic processes. To make the statement slightly more piqued: can they distinguish times of high and low uncertainty? Nieuwerburgh and Veldkamp (2006), for example, gives an account of agents learning about an uncertain economic environment before and after booms. Decisive for learning about the state of the world is the frequency of economic decision-making, which slows down past a boom. Potentially, this entails that agents might be far more off in their estimates of the variance parameters of their processes during the crisis of 2008.

The current model makes many concessions to tractability. The most prominent in this paper is the power function approximation of the tax system. As discussed above, the power function approximation is both a blessing and a curse. While it is both convenient for tractability of the model and the link to progressivity, it is restrictive in terms of the tax schedules it can represent. A possible extension is to switch to a fully specified structural model that uses dynamic programming as the solution method. In that case, the tax function can be arbitrary, even though using a tax calculator like taxsim is likely still infeasible due to the computational intensity. However, being able to introduce a discontinuous function that can fit observations with no gross and positive net income would be a marked improvement.

Finally, the tax experiment is somewhat limited in its scope: When I change the tax schedule and run through microsimulation, I change the estimate of the progressivity parameter. However, a change in the progressivity parameter does imply a change for the net incomes across the whole range of gross incomes according to the power function. Thus, simulating a reform that, for example, only affected the bottom of the distribution, would imply a change in the progressivity parameter that would also affect the top. Therefore, to take account of this fact, it would be appropriate to consider only reforms, where at least changes at the top and the bottom of the tax schedule take place. This is certainly an extension that I will consider. However, this requires custom reform options in the tax calculator taxsim, which have to be implemented by NBER based on request.

## 4.10 Conclusion

In this paper I document rising permanent earnings risk from 2000 to 2012 in a model of life-cycle labor supply. The increase of the permanent earnings risk is steady over the first half of this period and punctuated by the crisis in 2008. Namely, I document a 14% increase of the standard deviation of the permanent risk component in 2008 compared to the previous period. Further, I shed light on the role that progressive taxation played in the mitigation of this risk, which is minor.

The tax and transfer system is approximated by way of a power function to make the labor supply estimation tractable. Deductions have a minor role in shaping the progressivity of the system, but they do make it less progressive. An intriguing finding is that the relevant parameters of the approximation, especially the parameter $\tau$, which determines progressivity, are sensitive to the estimation method. Estimating the approximation using nonlinear least-squares implies smaller values of $1-\tau$; 0.93 instead of 0.82 found with the log specification. The fit of the model estimated using nonlinear least-squares performs about two times better in terms of root mean square error compared to the log-specification. In general, the power-function fits the data quite well, with an implied tax liability that is on average 650$ higher than the values derived from the tax-simulation model taxsim.

The estimation of the life-cycle labor supply model mostly confirms findings in the related literature. I find a tax-modified Frisch-elasticity of labor supply of 0.469 and an unmodified, after-tax Frisch-elasticity of 0.528. These values locate in the middle of the estimates presented in Blundell et al. (2016a) and Heathcote et al. (2014a). The Marshallian elasticity of labor-supply is negative and larger in absolute value compared to Blundell et al. (2016a).

Finally, I find that the pass-through of permanent wage-risk to earnings is roughly constant over time, which implies that earnings risk is mainly driven by the rise of permanent wage risk. From a counterfactual calculation I determine that to drive down earnings risk to pre-crisis levels, the progressivity parameter should have been lowered to 0.78 instead of the observed 0.93. This could have been achieved by raising the top tax rate by 56 percent.

In sum, I find that permanent earnings risk has been increasing at a steady clip over the early 2000s and taken a significant jump after the crisis of 2008 hit. The government, however, has not exercised much influence over this rise and has hardly varied the progressivity of the tax and transfer system.

# 4.11 Appendix

## 4.11.1 Measurement Error in Hours, Wages and Earnings

Following Blundell et al. (2016a), I correct the measurement error in log earnings, hours and wages using the estimates from the validation study Bound et al. (1994) to determine the proportion of the overall variance that is due to measurement error. Let $\tilde{y}, \tilde{h}$ and $\tilde{w}$ denote observed log earnings, hours and wages respectively. Then for any of these the following relationship holds,

$$\tilde{x} = x + me^{x}, x \in \{y, h, w\},$$

where $me^{x}$ denotes the measurement error and $x$ the true value. From Blundell et al. (2016a) I adopt the following relationships,

$$\begin{aligned} Var(me^{y}) &= 0.04 Var(\tilde{y}), \\ Var(me^{h}) &= 0.23 Var(\tilde{h}), \\ Var(me^{w}) &= 0.13 Var(\tilde{w}). \end{aligned}$$

It follows that the covariance between measurement error in wages and hours is given by

$$Cov\left(me^{w}, me^{h}\right) = \frac{1}{2}\left(Var(me^{y}) - Var\left(me^{h}\right) - Var(me^{w})\right) \tag{4.30}$$

**With Differenced Variables** The estimation of the stochastic processes is defined in terms of differenced variables, hence I need to account for the differenced and not the contemporary measurement error. In analogue to the above definitions, let $\Delta\tilde{x} = \tilde{x}_t - \tilde{x}_{t-1}$. Thus,

$$\Delta\tilde{x} = \Delta x + \Delta me^{x}.$$

The variance of the measurement error of differenced earnings is thus given by,

$$Var(\Delta me^{y}) = Var\left(me_t^{y}\right) - 2Cov\left(me_t^{y}, me_{t-1}^{y}\right) + Var\left(me_{t-1}^{y}\right) \tag{4.31}$$

$$Var(\Delta me^{y}) = Var\left(me_t^{y}\right) + Var\left(me_{t-1}^{y}\right) \tag{4.32}$$

where the second line follows from the assumption that measurement error is not correlated over time. Since the information about measurement error variances available from the validation study only covers relationships in levels, this assumption is required for the correction to be generalizable to temporal differences.

Again, I need to know the covariance of measurement errors in hours and wages to proceed with the estimation. By directly evaluating the covariance I find that,

$$Cov\left(\Delta me^{w}, \Delta me^{y}\right) = Cov\left(me_t^{w}, me_t^{h}\right) + Cov\left(me_{t-1}^{w}, me_{t-1}^{h}\right). \tag{4.33}$$

## 4.11.2 Sample Statistics by Year

**Table 4.13:** Sample Statistics

| | | 1998 | 2000 | 2002 | 2004 | 2006 | 2008 | 2010 | 2012 | 2014 |
|---|---|---|---|---|---|---|---|---|---|---|
| age | mean | 40.25 | 40.80 | 41.04 | 41.01 | 41.14 | 41.02 | 40.75 | 40.44 | 40.15 |
| | sd | 8.86 | 9.16 | 9.51 | 9.89 | 10.16 | 10.29 | 10.31 | 10.40 | 10.19 |
| | min | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 | 25 |
| | max | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 |
| hours | mean | 2220.35 | 2164.25 | 2175.95 | 2170.55 | 2187.35 | 2082.21 | 2092.34 | 2115.17 | 2099.96 |
| | sd | 568.53 | 554.98 | 585.47 | 595.44 | 572.71 | 561.58 | 597.68 | 582.91 | 572.19 |
| | min | 260 | 260 | 280 | 284 | 301 | 288 | 260 | 260 | 260 |
| | max | 4000 | 4000 | 4000 | 4000 | 4000 | 4000 | 4000 | 4000 | 4000 |
| wage | mean | 18.72 | 19.88 | 19.50 | 20.13 | 19.42 | 20.13 | 19.35 | 18.76 | 18.28 |
| | sd | 17.78 | 21.39 | 27.32 | 39.07 | 25.51 | 31.59 | 22.43 | 34.73 | 26.31 |
| | min | 1.84 | 2.15 | 1.41 | 1.35 | 1.77 | 1.75 | 1.68 | 1.75 | 1.72 |
| | max | 369.65 | 447.29 | 1079.63 | 1962.33 | 1078.91 | 1655.29 | 541.02 | 1695.15 | 1404.01 |
| years of education | mean | 13.18 | 13.22 | 13.20 | 13.33 | 13.34 | 13.64 | 13.74 | 13.76 | 13.67 |
| | sd | 2.60 | 2.47 | 2.58 | 2.41 | 2.49 | 2.43 | 2.33 | 2.40 | 2.69 |
| | min | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | max | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 | 17 |
| num. of children | mean | 1.14 | 1.08 | 1.02 | 1.02 | 1.02 | 1.00 | 0.99 | 0.99 | 0.99 |
| | sd | 1.21 | 1.19 | 1.16 | 1.16 | 1.19 | 1.20 | 1.21 | 1.20 | 1.22 |
| | min | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | max | 8 | 8 | 8 | 8 | 7 | 9 | 11 | 9 | 10 |
| gross income | mean | 62033 | 64508 | 64011 | 64829 | 64348 | 62935 | 59815 | 57965 | 56763 |
| | sd | 63859 | 71911 | 88631 | 107774 | 77425 | 93988 | 69498 | 98594 | 75944 |
| | min | 5598 | 5648 | 5585 | 5558 | 5575 | 5547 | 5555 | 5515 | 5562 |
| | max | 2021635 | 2026291 | 3370736 | 4822393 | 2759532 | 4850993 | 1605779 | 4544498 | 3689743 |
| tax liability | mean | 17808 | 18921 | 18400 | 17605 | 17382 | 17094 | 15441 | 14712 | 15194 |
| | sd | 25820 | 29591 | 38444 | 41616 | 30133 | 37067 | 26351 | 38906 | 32249 |
| | min | -2778 | -3481 | -4656 | -3665 | -7157 | -4715 | -7560 | -5838 | -6969 |
| | max | 895765 | 905915 | 1565206 | 1790800 | 1148302 | 2002816 | 666962 | 1842388 | 1705600 |
| net income | mean | 48789 | 50411 | 50142 | 51804 | 51440 | 50232 | 48620 | 46886 | 45738 |
| | sd | 41082 | 46437 | 52933 | 69025 | 50053 | 59587 | 45842 | 62244 | 46860 |
| | min | 5541 | 5231 | 4176 | 5019 | 5900 | 3616 | 5551 | 4035 | 4752 |
| | max | 1158005 | 1145453 | 1856328 | 3105156 | 1659988 | 2915430 | 968201 | 2771067 | 2057084 |
| Obs. | | 4535 | 4878 | 5104 | 5133 | 5191 | 5415 | 5141 | 5225 | 5346 |

*Note*: Own calculation based on PSID (2015). All statistics are unweighted.

## 4.11.3 Approximation of the Euler Equation

The Euler equation of consumption is given by

$$\frac{1}{\rho(1+r_t)}\lambda_t = E_t[\lambda_{t+1}]. \tag{4.34}$$

Expectations are rational, i.e., $\lambda_{t+1} = E[\lambda_{t+1}] + \varepsilon_{\lambda_{t+1}}$, where $\varepsilon_{\lambda_{t+1}}$ denotes the mean-zero expectation correction of $E[\lambda_{t+1}]$ performed in period $t+1$. Expectation errors are caused by innovations in the hourly wage residual $\omega_{t+1}$, which, as implied by rational expectations, are uncorrelated with $E_t[\lambda_{t+1}]$. Rational expectations imply that $\varepsilon_{\lambda_{t+1}}$ is uncorrelated over time, so that regardless of the autocorrelative structure of the shock terms, $\varepsilon_{\lambda_{t+1}}$ will only be correlated with the innovations of the shock processes.

To find an estimable form for $\Delta \ln h_t$, we take logs of (4.34) and resolve the expectation:

$$\ln \lambda_t = \ln(1+r_t) + \ln \rho + \ln(\lambda_{t+1} - \varepsilon_{t+1})$$

A first order Taylor-expansion of $\ln(\lambda_{t+1} - \varepsilon_{t+1})$ gives $\ln(\lambda_{t+1}) + \frac{\varepsilon_{\lambda_{t+1}}}{\lambda_{t+1}}$, leading to the expression

$$\ln \lambda_t = \ln(1+r_t) + \ln \rho + \ln(\lambda_{t+1}) + \frac{\varepsilon_{\lambda_{t+1}}}{\lambda_{t+1}} + \mathcal{O}\bigg(-1/2(\varepsilon_{t+1}/\lambda_{t+1})^2\bigg). \tag{4.35}$$

Accordingly, when we backdate 4.35, I can remove $\ln \lambda_t$ in the first difference formulation of the labor supply equation 4.20.

### 4.11.4 A Model with Explicit Expenditure for Deductions

#### Two-Stage Budgeting

In the following I explore the within-period leisure-consumption-allocations of individuals when they have the ability to deduct from their taxable income by making purchases of deductible goods.

In the case of perfect foresight and with intertemporally additively-separable utilities over the life-cycle one can decompose the standard optimization of the consumer into two separate optimization procedures (see Blundell and Walker (1986) and Keane (2011)). First, the consumer allocates full income $F_t$ into each period out of life-time wealth $W_t$ so as to equate the appropriately discounted values of the marginal utility of income. Second, the consumer chooses in-period consumption-leisure bundles to maximize in-period utility. I derive in an expression for the optimal choice of hours and whether a change in the wage $w_t$ influences hours also through the parameters that relate expenditure for deductible goods to deductions.

### In-Period Allocation

The optimization of the in-period utility function is the same as in any static problem with the exception that it proceeds with full income given, so that $F_t$ is fixed. The optimization program is

$$\max_{c_t, c_t^d, h_t} \mathrm{U}(c_t, c_t^d, h_t), \tag{4.36}$$

$$s.t. \quad F_t = \chi\left(w_t T - \mathcal{D}\left(p_{d,t} c_t^d\right)\right)^{1-\tau} - \chi\left(w_t h_t - \mathcal{D}\left(p_{d,t} c_t^d\right)\right)^{1-\tau} + c_t + p_{d,t} c_t^d \tag{4.37}$$

where, for simplicity's sake, I use the parametric form

$$\mathrm{U}(c_t, c_t^d, h_t) = u\left(c_t^d, c_t\right) - \frac{h_t^{1+\gamma}}{1+\gamma}, \qquad \gamma \geq 0. \tag{4.38}$$

Here I divide consumption into two types of goods, non-tax-deductible goods $c_t$, with its price normalized to 1, and tax-deductible goods $c_t^d$ and price $p_{d,t}$. Utility from consumption is given by the concave and twice continuously differentiable function $u(\cdot)$. The function $\mathcal{D}(\cdot)$ gives the deductions from gross income and is increasing in $p_{d,t} c_t^d$. Further, $T$ is the total time endowment in the period. To find an indication of the role that deductible goods play in determining labor supply, I have to inspect the first-order condition for $h$:

$$h_t^{\gamma} = \lambda\left(\chi(1-\tau)\left(w_t h_t - \mathcal{D}\left(p_{d,t} c_t^d\right)\right)^{-\tau} w_t\right). \tag{4.39}$$

This expression reveals that, the choice of hours depends nonlinearly on deductions.

Building on this result, I make the pragmatic choice in section 4.4.3 of approximating taxable income as a power function of gross income. Otherwise there would be no way of proceeding with the first-order approach or the impact of deductions would have be neglected.

## 4.11.5 Replication Estimation of the Tax Function Approximation in Heathcote et al. (2017a)

I estimate the model for statutory progressivity shown in eq. A2 of Heathcote et al. (2017b) using the data provided in the replication files for Heathcote et al. (2017a). This means, I model their net income variable minus deductions based on taxable income. I do it once in logs and once using nonlinear least-squares with the quantities in levels. In line with their estimation procedure I pool all observations in their panel from 2000 to 2006. Each cross-section contains between 3000 to 3500 observations. The results are shown in table 4.14. I use their notation, where $1-\tau$ is the progressivity parameter and $\lambda$ is the coefficient of the function equivalent to $\chi$ in my notation.

**Table 4.14:** Progressivity Estimates

| | log spec. | nonlin. spec. |
|---|---|---|
| $\lambda$ | 5.568 | 1.449 |
| $1-\tau$ | 0.819 | 0.937 |
| RMSE | 18714 | 8932 |
| Obs. | 12875 | 12875 |

The table shows that the nonlinear specification and the log specification imply very different progressivity estimates. The nonlinear model implies much lower progressivity. Further, the fit, assessed by computing the RMSE of the predicted residuals in levels, is slightly more than twice as large when computed based on the log specification.

In table 4.15 I show the estimates year by year along with my own estimates for the total progressivity. The take-away is that the estimates of the statutory progressivity parameter, when measured with nonlinear least-squares, are very close to the estimates of total progressivity that I calculate. Hence, I find that the quantitative importance of the distinction between statutory and total progressivity is quite small. However, the table does highlight the inferior fit of the log specification in terms of the RMSE. In every year the log specification fits worse than the equivalent nonlinear specification.

**Table 4.15:** Progressivity Estimates by Year

| | log specification | | | | nonlinear specification | | | | complete model | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | $\lambda$ | $1-\tau$ | RMSE | Obs. | $\lambda$ | $1-\tau$ | RMSE | Obs. | $\chi$ | $1-\tau$ | Obs. |
| 2000 | 4.284 | 0.840 | 7229 | 3198 | 2.286 | 0.895 | 5081 | 3198 | 2.610 | 0.893 | 4878 |
| 2002 | 6.05 | 0.811 | 12953 | 3204 | 2.336 | 0.896 | 6370 | 3204 | 2.298 | 0.903 | 5104 |
| 2004 | 6.221 | 0.811 | 32099 | 3266 | 1.097 | 0.960 | 9813 | 3266 | 1.373 | 0.949 | 5133 |
| 2006 | 5.789 | 0.818 | 13402 | 3207 | 2.037 | 0.910 | 7903 | 3207 | 2.121 | 0.912 | 5191 |

*Note*: The first two models are the same as the two displayed in table 4.14, but are disaggregated by year. The last model is the complete model of table 4.3 displayed for comparison.

### 4.11.6 Identification Example of the Stochastic Process for Wages

Let the stochastic process for wages be

$$\begin{aligned}
\omega_{it} &= p_{it} + \tau_{it} + \text{me}_{it} \\
p_{it} &= p_{it-1} + \zeta_{it} \\
\tau_{it} &= \theta\epsilon_{it-1} + \epsilon_{it} \\
\zeta_{it} &\sim N\left(0, \sigma^2_{\zeta,t}\right), \quad \epsilon_{it} \sim N\left(0, \sigma^2_{\epsilon,t}\right) \\
E\left[\zeta_t\zeta_{t-l}\right] &= 0, \quad E\left[\epsilon_t\epsilon_{t-l}\right] = 0 \quad \forall l \in \mathbb{Z}_{\neq 0}
\end{aligned} \tag{4.40}$$

The equations giving the value of the time-differenced innovation $\Delta\omega_{it}$ are

$$\Delta\omega_{it} = \begin{cases} (\theta-1)\epsilon_{it-1}+\zeta_{it}+\epsilon_{it}+\Delta\text{me}_{it} & \text{if } t=1 \\ (\theta-1)\epsilon_{it-1}-\theta\epsilon_{it-2}+\zeta_{it}+\epsilon_{it}+\Delta\text{me}_{it} & \text{if } t>1 \end{cases} \tag{4.41}$$

Let there be three periods, such that $t \in \{1,2,3\}$. Then I obtain the following matrix of autocovariance moments:

$$VCV_{\Delta\omega} = \tag{4.42}$$

$$\begin{pmatrix} 2\sigma^2_{\text{me}}+(\theta-1)^2\sigma^2_{\epsilon,0}+\sigma^2_{\epsilon,1}+\sigma^2_{\zeta,1} & -\sigma^2_{\text{me}}-\theta^2\sigma^2_{\epsilon,0}-\sigma^2_{\epsilon,1}+\theta(\sigma^2_{\epsilon,0}+\sigma^2_{\epsilon,1}) & -\theta\sigma^2_{\epsilon,1} \\ -\sigma^2_{\text{me}}-\theta^2\sigma^2_{\epsilon,0}-\sigma^2_{\epsilon,1}+\theta(\sigma^2_{\epsilon,0}+\sigma^2_{\epsilon,1}) & 2\sigma^2_{\text{me}}+\sigma^2_{\epsilon,1}-2\theta\sigma^2_{\epsilon,1}+\theta^2(\sigma^2_{\epsilon,0}+\sigma^2_{\epsilon,1})+\sigma^2_{\epsilon,2}+\sigma^2_{\zeta,2} & -\sigma^2_{\text{me}}-\theta^2\sigma^2_{\epsilon,1}-\sigma^2_{\epsilon,2}+\theta(\sigma^2_{\epsilon,1}+\sigma^2_{\epsilon,2}) \\ -\theta\sigma^2_{\epsilon,1} & -\sigma^2_{\text{me}}-\theta^2\sigma^2_{\epsilon,1}-\sigma^2_{\epsilon,2}+\theta(\sigma^2_{\epsilon,1}+\sigma^2_{\epsilon,2}) & 2\sigma^2_{\text{me}}+\sigma^2_{\epsilon,2}-2\theta\sigma^2_{\epsilon,2}+\theta^2(\sigma^2_{\epsilon,1}+\sigma^2_{\epsilon,2})+\sigma^2_{\epsilon,3}+\sigma^2_{\zeta,3} \end{pmatrix}$$

There are six unique moments in the above matrix that are used for identification. Let every element of $VCV_{\Delta\omega}$ be denoted by the symbols $\Gamma_{k,j}$, where $k,j \in \{1,2,3\}$, so that for example the variance in period 1 is $\Gamma_{1,1} = 2\sigma^2_{\text{me}} + (\theta-1)^2\sigma^2_{\epsilon,0} + \sigma^2_{\epsilon,1} + \sigma^2_{\zeta,1}$. The easiest way to proceed is to set the innovation variance of the transitory process in $t = 0$ to zero. Further, as laid out in appendix 4.11.1, I can treat the variance of the measurement error as known. Then the set of moments used for identification becomes:

$$\begin{aligned}
\Gamma_{1,1} &= \sigma^2_{\epsilon,1} + \sigma^2_{\zeta,1} \\
\Gamma_{1,2} &= -\sigma^2_{\epsilon,1} + \theta\sigma^2_{\epsilon,1} \\
\Gamma_{1,3} &= -\theta\sigma^2_{\epsilon,1} \\
\Gamma_{2,2} &= \sigma^2_{\epsilon,1} - 2\theta\sigma^2_{\epsilon,1} + \theta^2\sigma^2_{\epsilon,1} + \sigma^2_{\epsilon,2} + \sigma^2_{\zeta,2} \\
\Gamma_{2,3} &= -\theta^2\sigma^2_{\epsilon,1} - \sigma^2_{\epsilon,2} + \theta(\sigma^2_{\epsilon,1} + \sigma^2_{\epsilon,2}) \\
\Gamma_{3,3} &= \sigma^2_{\epsilon,2} - 2\theta\sigma^2_{\epsilon,2} + \theta^2(\sigma^2_{\epsilon,1} + \sigma^2_{\epsilon,2}) + \sigma^2_{\epsilon,3} + \sigma^2_{\zeta,3}
\end{aligned}$$

Accordingly, the identification proceeds by solving for the variances and the persistence parameter

$$\sigma^2_{\epsilon,1} = -(\Gamma_{1,2} + \Gamma_{1,3})$$
$$\theta = \Gamma_{1,3}/(\Gamma_{1,2} + \Gamma_{1,3})$$
$$\sigma^2_{\zeta,1} = \Gamma_{1,1} + (\Gamma_{1,2} + \Gamma_{1,3})$$
$$\sigma^2_{\epsilon,2} = \frac{1}{\frac{\Gamma_{1,3}}{\Gamma_{1,2}+\Gamma_{1,3}} - 1}\left[\Gamma_{2,3} + \left(\left(\frac{\Gamma_{1,3}}{\Gamma_{1,2}+\Gamma_{1,3}}\right)^2 - \frac{\Gamma_{1,3}}{\Gamma_{1,2}+\Gamma_{1,3}}\right) \times (-(\Gamma_{1,2}+\Gamma_{1,3}))\right]$$
$$\sigma^2_{\zeta,2} = \Gamma_{2,2} - \left(\frac{\Gamma_{1,3}}{\Gamma_{1,2}+\Gamma_{1,3}} - 1\right)^2 \times (-(\Gamma_{1,2}+\Gamma_{1,3})) - \frac{1}{\frac{\Gamma_{1,3}}{\Gamma_{1,2}+\Gamma_{1,3}} - 1}\left[\Gamma_{2,3} + \left(\left(\frac{\Gamma_{1,3}}{\Gamma_{1,2}+\Gamma_{1,3}}\right)^2 - \frac{\Gamma_{1,3}}{\Gamma_{1,2}+\Gamma_{1,3}}\right) \times (-(\Gamma_{1,2}+\Gamma_{1,3}))\right]$$

Each further period delivers two more moments that can be used for identification, so that the next set of permanent and transitory variances can be identified.

Another possibility to identify the process is to set the first two transitory shock variances equal to each other, $\sigma^2_{\epsilon,0} = \sigma^2_{\epsilon,1}$. The identification proceeds analogously, except that instead of $\theta$ being identified directly, the ratio $\theta/(1-\theta)$ is identified. However, this makes no difference in practice.

### 4.11.7 Innovations to Taste-Shifter $b$

**Table 4.16:** Taste-Shifter Innovations

| | 2000 | 2002 | 2004 | 2006 | 2008 | 2010 | 2012 |
|---|---|---|---|---|---|---|---|
| $\sigma_{\nu,t}$ | 0.3188 | 0.0370 | 0.0009 | 0.4594 | 0.4237 | 0.3263 | 0.5023 |
| | (0.0033) | (0.0044) | (0.0006) | (0.0025) | (0.0027) | (0.0032) | (0.0021) |

*Note:* Own calculation based on PSID (2015). Bootstrap standard errors based on 200 replications in parentheses.

The standard deviations for the innovations of the taste-shifter are roughly two times as large as the permanent shock variances of the wage process. There is no very clear-cut trend across the years. In 2004 the estimate is so small that it turns insignificant. This points to a large degree of instability in the evolution of the variance of the taste-shifter. In the current model this component of the hours variance is purely transitory.

# List of Tables

# List of Figures

# Bibliography


Abowd, John M and David Card (1989), "On the covariance structure of earnings and hours changes." *Econometrica*, 57, 411–445, URL http://ideas.repec.org/a/ecm/emetrp/v57y1989i2p411-45.html.

Alan, Sule, Martin Browning, and Mette Ejrnæs (forthcoming), "Income and consumption: a micro semistructural analysis with pervasive heterogeneity." *Journal of Political Economy*.

Altonji, Joseph G (1986), "Intertemporal substitution in labor supply: evidence from micro data." *Journal of Political Economy*, 94, S176–S215, URL http://ideas.repec.org/a/ucp/jpolec/v94y1986i3ps176-s215.html.

Altonji, Joseph G and Lewis M Segal (1996), "Small-sample bias in GMM estimation of covariance structures." *Journal of Business & Economic Statistics*, 14, 353–366, URL https://ideas.repec.org/a/bes/jnlbes/v14y1996i3p353-66.html.

Altonji, Joseph G., Anthony A. Smith Jr., and Ivan Vidangos (2013), "Modeling earnings dynamics." *Econometrica*, 81, 1395–1454, URL https://ideas.repec.org/a/ecm/emetrp/v81y2013i4p1395-1454.html.

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme (2017), "Earnings and consumption dynamics: a nonlinear panel data framework." *Econometrica*, 85, 693–734, URL https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA13795.

Arrow, Kenneth J. and Alain C. Enthoven (1961), "Quasi-concave programming." *Econometrica*, 29, 779–800.

Atkinson, Anthony B. (1970), "On the measurement of inequality." *Journal of Economic Theory*, 2, 244–263.

Attanasio, Orazio and José-Vıctor Rıos-Rull (2000), "Consumption smoothing in island economies: can public insurance reduce welfare?" *European economic review*, 44, 1225–1258.

Attanasio, Orazio P. and Luigi Pistaferri (2016), "Consumption inequality." *Journal of Economic Perspectives*, 30, 3–28, URL http://www.aeaweb.org/articles?id=10.1257/jep.30.2.3.

Attanasio, Orazio P and Guglielmo Weber (2010), "Consumption and saving: models of intertemporal allocation and their implications for public policy." *Journal of Economic literature*, 48, 693–751.

Blackorby, Charles and David Donaldson (1978), "Measures of relative equality and their meaning in terms of social welfare." *Journal of Economic Theory*, 18, 59–80.

Blundell, Richard, Michael Graber, and Magne Mogstad (2015), "Labor income dynamics and the insurance from taxes, transfers, and the family." *Journal of Public Economics*, 127, 58 – 73, URL http://www.sciencedirect.com/science/article/pii/S0047272714000930. The Nordic Model.

Blundell, Richard, Hamish Low, and Ian Preston (2013), "Decomposing changes in income risk using consumption data." *Quantitative Economics*, 4, 1–37, URL https://ideas.repec.org/a/ecm/quante/v4y2013i1p1-37.html.

Blundell, Richard and Thomas Macurdy (1999), "Labor supply: a review of alternative approaches." In *Handbook of Labor Economics* (O. Ashenfelter and D. Card, eds.), volume 3 of *Handbook of Labor Economics*, chapter 27, 1559–1695, Elsevier, URL http://ideas.repec.org/h/eee/labchp/3-27.html.

Blundell, Richard, Luigi Pistaferri, and Ian Preston (2008), "Consumption inequality and partial insurance." *American Economic Review*, 98, 1887–1921.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten (2016a), "Consumption inequality and family labor supply." *The American Economic Review*, 106, 387–435.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten (2016b), "Consumption inequality and family labor supply: Online appendix." URL https://www.aeaweb.org/aer/app/10602/20121549_app.pdf.

Blundell, Richard and Ian Walker (1986), "A life-cycle consistent empirical model of family labour supply using cross-scection data." *The Review of Economic Studies*, 53, 539–558, URL http://restud.oxfordjournals.org/content/53/4/539.abstract.

Börsch-Supan, Axel H., Michela Coppola, and Anette Reil-Held (2012), "Riester pensions in germany: design, dynamics, targetting success and crowding-in." *NBER Working Paper No. 18014*.

Börsch-Supan, Axel H., Anette Reil-Held, and Daniel Schunk (2008), "Saving incentives, old-age provision and displacement effects: evidence from the recent german pension reform." *Journal of Pension Economics and Finance*, 7, 295–319.

Bound, John, Charles Brown, Greg J Duncan, and Willard L Rodgers (1994), "Evidence on the validity of cross-sectional and longitudinal labor market data." *Journal of Labor Economics*, 12, 345–368.

Boyd, Stephen and Lieven Vandenberghe (2004), *Convex optimization*. Cambridge University Press.

Burman, Leonard E., William G. Gale, Matthew Hall, and Peter R. Orszag (2004), "Distributional effects of defined contribution plans and individual retirement arrangements." *National Tax Journal*, 57, 671–701.

Carbonnier, Clément, Alexis Direr, and Ihssane Slimani Houti (2014), "Do savers respond to tax incentives? the case of retirement savings." *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, 225–256.

Charnes, Abraham and William W. Cooper (1962), "Programming with linear fractional functionals." *Naval Research Logistics Quarterly*, 9, 181–186.

Chernozhukov, Victor and Christian Hansen (2004), "The effects of 401 (k) participation on the wealth distribution: an instrumental quantile regression analysis." *Review of Economics and Statistics*, 86, 735–751.

Chetty, Raj, John N. Friedman, Soeren Leth-Petersen, Torben Heien Nielsen, and Tore Olsen (2014), "Active vs. passive decisions and crowd-out in retirement savings accounts: evidence from denmark." *The Quarterly Journal of Economics*, 129, 1141–1219.

Chiappori, Pierre-Andre (2016), "Equivalence versus indifference scales." *The Economic Journal*, 126, 523–545.

Coppola, Michela and Anette Reil-Held (2009), *Dynamik der Riester-Rente: Ergebnisse aus SAVE 2003 bis 2008*. Mannheim Research Institute for the Economics of Aging.

Corneo, Giacomo, Matthias Keese, and Carsten Schröder (2009), "The riester scheme and private savings: an empirical analysis based on the german soep." *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 129, 321–332, URL http://ideas.repec.org/a/aeq/aeqsjb/v129_y2009_i1_q1_p321-332.html.

Corneo, Giacomo, Matthias Keese, and Carsten Schröder (2010), "The effect of saving subsidies on household saving–evidence from germany." *Ruhr Economic Paper*.

Corneo, Giacomo, Johannes König, and Carsten Schröder (2018), "Distributional effects of subsidizing retirement savings accounts: Evidence from germany." *Finanzarchiv / Public Finance Analysis (forthcoming)*.

Domeij, David and Martin Flodén (2006), "The labor-supply elasticity and borrowing constraints: why estimates are biased." *Review of Economic Dynamics*, 9, 242–262, URL http://ideas.repec.org/a/red/issued/v9y2006i2p242-262.html.

D'Orazio, Marcello, Marco Di Zio, and Mauro Scanu (2006), *Statistical matching: theory and practice*. John Wiley & Sons.

Eaton, Jonathan and Harvey S Rosen (1980), "Labor supply, uncertainty, and efficient taxation." *Journal of Public Economics*, 14, 365–374.

Ebert, Udo (1999), "Using equivalent income of equivalent adults to rank income distributions." *Social Choice and Welfare*, 16, 233–258.

Ebert, Udo and Patrick Moyes (2003), "Equivalence scales reconsidered." *Econometrica*, 71, 319–343.

Engelhardt, Gary V and Anil Kumar (2011), "Pensions and household wealth accumulation." *Journal of Human Resources*, 46, 203–236.

Engen, Eric M., William G. Gale, and John K. Scholz (1996), "The illusory effects of saving incentives on saving." *The Journal of Economic Perspectives*, 10, 113–138.

Even, William and David Macpherson (2007), "Defined contribution plans and the distribution of pension wealth." *Industrial Relations: A Journal of Economy and Society*, 46, 551–581.

Fasshauer, Stephan and Nora Toutaoui (2009), "Die anzahl des förderberechtigten personenkreises der riester-rente - eine annäherung." *Deutsche Rentenversicherung*.

Feenberg, Daniel and Elisabeth Coutts (1993), "An introduction to the taxsim model." *Journal of Policy Analysis and management*, 12, 189–194.

Feldstein, Martin S. (1969), "The effects of taxation on risk taking." *Journal of Political Economy*, 77, 755–764, URL http://www.jstor.org/stable/1829965.

Friedman, Milton (1957), *A theory of the consumption function*. NBER Book, National Bureau of Economic Research, URL http://ideas.repec.org/b/nbr/nberbk/frie57-1.html.

Glewwe, Paul (1991), "Household equivalence scales and the measurement of inequality: transfers from the poor to the rich could decrease inequality." *Journal of Public Economics*, 44, 211–216.

Gouveia, Miguel and Robert P Strauss (1994), "Effective federal individual income tax functions: An exploratory empirical analysis." *National Tax Journal*, 317–339.

Guvenen, Fatih (2007), "Learning your earning: are labor income shocks really very persistent?" *American Economic Review*, 97, 687–712, URL http://www.aeaweb.org/articles?id=10.1257/aer.97.3.687.

Guvenen, Fatih (2009), "An empirical investigation of labor income processes." *Review of Economic Dynamics*, 12, 58–79, URL https://ideas.repec.org/a/red/issued/06-15.html.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song (2015), "What do data on millions of u.s. workers reveal about life-cycle earnings risk?" URL http://www.nber.org/papers/w20913.

Guvenen, Fatih and Anthony A. Smith (2014), "Inferring labor income risk and partial insurance from economic choices." *Econometrica*, 82, 2085–2129, URL http://dx.doi.org/10.3982/ECTA9446.

Hall, Robert E. (1978), "Stochastic implications of the life cycle permanent income hypothesis: theory and evidence." *Journal of Political Economy*, 86, 971–987, URL http://dx.doi.org/10.1086/260724.

Hastings Jr., Cecil, Frederick Mosteller, John W. Tukey, and Charles P. Winsor (1947), "Low moments for small samples: a comparative study of order statistics." *The Annals of Mathematical Statistics*, 18, 413–426.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2014a), "Consumption and labor supply with partial insurance: an analytical framework." *American Economic Review*, 104, 2075–2126, URL http://www.aeaweb.org/articles?id=10.1257/aer.104.7.2075.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2014b), "Consumption and labor supply with partial insurance: an analytical framework - online appendix." URL https://www.aeaweb.org/aer/app/10407/20090634_app.pdf.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L Violante (2017a), "Optimal tax progressivity: An analytical framework." *Quarterly Journal of Economics*, forthcoming.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante (2017b), "Optimal tax progressivity: An analytical framework: Online appendix."

HFCN, (Household Finance and Consumption Network) (2013), "The eurosystem household finance and consumption survey - methodological report for the first wave." *Statistics Paper Series No. 1 / April 2013*.

Hryshko, Dmytro (2012), "Labor income profiles are not heterogeneous: evidence from income growth rates." *Quantitative Economics*, 3, 177–209.

Hurd, Michael D and Susann Rohwedder (2008), "The retirement consumption puzzle: actual spending change in panel data." URL http://www.nber.org/papers/w13929.

Imai, Susumu and Michael P. Keane (2004), "Intertemporal labor supply and human capital accumulation." *International Economic Review*, 45, 601–641, URL `http://dx.doi.org/10.1111/j.1468-2354.2004.00138.x`.

Jakobsson, Ulf (1976), "On the measurement of the degree of progression." *Journal of Public Economics*, 5, 161 – 168.

Jessen, Robin and Johannes König (2018), "Hours risk, wage risk, and life-cycle labor supply." *mimeo*.

Joulfaian, David and David P. Richardson (2001), "Who takes advantage of tax-deferred saving programs? evidence from federal income tax data." *National Tax Journal*, 54, 669–88.

Joumard, Isabelle, Mauro Pisu, and Debbie Bloch (2012), "Tackling income inequality: The role of taxes and transfers." *OECD Journal: Economic Studies*, 2012.

Judd, Kenneth L. (1998), *Numerical methods in economics*. MIT press.

Kaplan, Greg (2012), "Inequality and the life cycle." *Quantitative Economics*, 3, 471–525, URL `http://dx.doi.org/10.3982/QE200`.

Keane, Michael P (2011), "Labor supply and taxes: a survey." *Journal of Economic Literature*, 49, 961–1075.

Kimberlin, Sara, H Luke Shaefer, and Jiyoon Kim (2016), "Measuring poverty using the supplemental poverty measure in the panel study of income dynamics, 1998 to 2010." *Journal of Economic and Social Measurement*, 41, 17–47.

Kleibergen, Frank and Richard Paap (2006), "Generalized reduced rank tests using the singular value decomposition." *Journal of Econometrics*, 133, 97–126.

König, Johannes (2018), "Earnings risk and tax policy." *mimeo*.

König, Johannes and Carsten Schröder (2018), "Inequality-minimization with a given public budget." *Journal of Economic Inequality*.

Krueger, Dirk (2007), "Consumption and saving: theory and evidence." *Manuscript, University of Pennsylvania*.

Lambert, Peter (2001), *The distribution and redistribution of income*. Manchester University Press.

Low, Hamish and Daniel Maldoom (2004), "Optimal taxation, prudence and risk-sharing." *Journal of Public Economics*, 88, 443 – 464.

Low, Hamish, Costas Meghir, and Luigi Pistaferri (2010), "Wage risk and employment risk over the life cycle." *American Economic Review*, 100, 1432–67, URL http://www.aeaweb.org/articles?id=10.1257/aer.100.4.1432.

Luenberger, David G. (1968), "Quasi-convex programming." *SIAM Journal on Applied Mathematics*, 16, 1090–1095.

MaCurdy, Thomas E (1981), "An empirical model of labor supply in a life-cycle setting." *Journal of Political Economy*, 89, 1059–85, URL http://ideas.repec.org/a/ucp/jpolec/v89y1981i6p1059-85.html.

MaCurdy, Thomas E. (1982), "The use of time series processes to model the error structure of earnings in a longitudinal data analysis." *Journal of Econometrics*, 18, 83–114, URL http://ideas.repec.org/a/eee/econom/v18y1982i1p83-114.html.

Meghir, Costas and Luigi Pistaferri (2004), "Income variance dynamics and heterogeneity." *Econometrica*, 72, 1–32, URL https://ideas.repec.org/a/ecm/emetrp/v72y2004i1p1-32.html.

Mirrlees, J. A. (1971), "An exploration in the theory of optimum income taxation." *The Review of Economic Studies*, 38, 175–208, URL http://www.jstor.org/stable/2296779.

Nieuwerburgh, Stijn Van and Laura Veldkamp (2006), "Learning asymmetries in real business cycles." *Journal of Monetary Economics*, 53, 753 – 772, URL http://www.sciencedirect.com/science/article/pii/S0304393206000390.

Pfarr, Christian and Udo Schneider (2011), "Anreizeffekte und angebotsinduzierung im rahmen der riester-rente: Eine empirische analyse geschlechts-und sozialisationsbedingter unterschiede." *Perspektiven der Wirtschaftspolitik*, 12, 27–46.

Prete, Vincenzo, Alessandro Sommacal, and Claudio Zoli (2016), "Optimal non-welfarist income taxation for inequality and polarization reduction." *University of Verona, Department of Economics Working Papers*, 23.

PSID (2015), "PSID public use file." *Produced and distributed by the Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI*. Year of access: 2017.

PSID (2017), "PSID main interview user manual: release 2017." *Institute for Social Research, University of Michigan*.

Rubin, Donald B. (2004), *Multiple imputation for nonresponse in surveys*, volume 81. John Wiley & Sons.

Shalit, Haim and Shlomo Yitzhaki (2005), "The mean-gini efficient portfolio frontier." *Journal of Financial Research*, 28, 59–75.

Shorrocks, Anthony (2004), "Inequality and welfare evaluation of heterogeneous income distributions." *The Journal of Economic Inequality*, 2, 193–218.

Statistisches Bundesamt (2008), "Jährliche einkommensstatistik sonderthema werbungskosten." *Fachserie 14 Reihe 7.1.1 Finanzen und Steuern*.

Steiner, Viktor, Katharina Wrohlich, Peter Haan, and Johannes Geyer (2008), "Documentation of the tax-benefit microsimulation model stsm - version 2008." Technical report, DIW Berlin, German Institute for Economic Research.

Stolz, Ulrich and Christian Rieckhoff (2013), "Die riester-rente im beitragsjahr 2010: Zulageförderung erstmals für mehr als 10 millionen berechtigte." *RVaktuell 12/2013*.

Varian, Hal R (1980), "Redistributive taxation as social insurance." *Journal of Public Economics*, 14, 49–68.

von Kalckreuth, Ulf, Martin Eisele, Julia Le Blanc, Tobias Schmidt, and Junyi Zhu (2012), "The phf: a comprehensive panel survey on household finances and wealth in germany." *Discussion Paper Deutsche Bundesbank No. 13/2012*.

von Kalckreuth, Ulf, Tobias Schmidt, Martin Eisele, Julia Le Blanc, and Junyi Zhu (2013), "Panel on household finances." URL `http://dx.doi.org/10.12757/PHF.01.01.01.STATA`.

Wagner, Gert G., Joachim R. Frick, and Jürgen Schupp (2007), "The german socio-economic panel study (soep)-evolution, scope and enhancements." *Schmollers Jahrbuch: Journal of Applied Social Science Studies/Zeitschrift für Wirtschafts-und Sozialwissenschaften*, 127, 139–169.

Wodon, Quentin T. and Shlomo Yitzhaki (2005), "Inequality and social welfare when using equivalence scales." *SSRN Working Paper*.

Yitzhaki, Shlomo (1982), "A tax programming model." *Journal of Public Economics*, 19, 107–120.

Yitzhaki, Shlomo (1983), "On an extension of the Gini inequality index." *International Economic Review*, 617–628.

Yitzhaki, Shlomo and Peter J. Lambert (2013), "The relationship between the absolute deviation from a quantile and Gini's mean difference." *Metron*, 71, 97–104.

Yitzhaki, Shlomo and Edna Schechtman (2012), *The Gini methodology: a primer on a statistical methodology*. Springer Science & Business Media.

Zhu, Junyi and Martin Eisele (2013), "Multiple imputation in a complex household survey - the german panel on household finances (phf): challenges and solutions." *PHF User Guide*, URL http://www.bundesbank.de/Redaktion/EN/Downloads/Bundesbank/Research_Centre/phf_imputation.pdf?__blob=publicationFile.

# English Summary

This dissertation consists of four empirical chapters with the second also making a methodological contribution.

The first chapter empirically investigates the distributional consequences of the Riester scheme, the main private pension subsidization program in Germany. 38% of the aggregate subsidy accrues to the top two deciles of the income distribution, but only 7.3% to the bottom two. Nonetheless the Riester scheme is almost distributionally neutral in terms of standard inequality measures. Two effects offset each other: a progressive one stemming from the subsidy schedule and a regressive one due to voluntary participation. Participation is associated not only with high income but also with high household wealth.

The second chapter solves the problem of a social planner who seeks to minimize inequality via transfers with a fixed public budget in a distribution of exogenously given incomes. The appropriate solution method depends on the objective function: If it is convex, it can be solved by an interior-point algorithm. If it is quasiconvex, the bisection method can be used. Using artificial and real-world data, an implementation the procedures shows that the optimal transfer scheme need not comply with a transfer scheme that perfectly equalizes incomes at the bottom of the distribution.

The third chapter investigates the nature of earnings risk in a model of life-cycle labor supply. The recent literature on life-cycle consumption, saving and labor supply focuses on wage shocks as the central source of risk. In the paper we propose a life-cycle labor supply model that features risk in both wages and hours and disentangle their effects on earnings risk. To this end we estimate a transmission parameter that measures how permanent wage shocks impact the marginal utility of wealth. Estimating our model with the Panel Study of Income Dynamics (PSID) shows that both permanent wage and hours shocks play an important role in explaining the cross-sectional variance in earnings growth. Still, permanent wage shocks have a quantitatively larger impact on life-time earnings. Allowing for hours shocks improves the model fit considerably. The empirical strategy allows for estimating the Marshallian labor supply elasticity without the use of consumption or asset data. We find this elasticity on average to be negative, but small. Finally, we link our estimate of the transmission parameter to consumption insurance and show that the sensitivity of consumption to wage shocks implied by our estimate is in line with recent estimates in the literature.

The fourth chapter is another application of a life-cycle labor supply model, but this time with a focus on progressive taxation. I quantify individuals' exposure to permanent earnings risk and find that permanent earnings risk in the US has been on the rise since the early 2000s. Most importantly, it has taken a marked hike during the financial crisis of 2008. In contrast, the insurance effect of the progressive tax and transfer system, which mitigates this risk, has remained flat. I estimate the progressivity of the tax and transfer system using a power function approximation and evaluate its properties in representing the tax system. This progressivity parameter is sufficient to identify the insurance effect of the tax and transfer system. When progressivity is shut down, the model features 5% less insurance. Earnings risk could have been reduced to pre-crisis levels by increasing progressivity substantially, lowering the progressivity parameter from the observed level of 0.93 to 0.78.

# Deutsche Zusammenfassung

Diese Dissertation besteht aus vier empirischen Kapiteln, wobei das zweite auch einen methodologischen Beitrag leistet.

Das erste Kapitel untersucht empirisch die Auswirkungen der Riesterförderung auf die Einkommensverteilung. 38% der Gesamtsubventionen kommen den oberen zwei Dezilen der Einkommensverteilung zugute, aber nur 7,3% in die unteren zwei. Die Riesterförderung ist trotzdem nahezu verteilungsneutral in Termini von Standardungleichheitsmaßen. Die Ursache hierfür liegt in zwei Effekte, die sich gegenseitig ausgleichen: Eine Progressiver, der aus der Art der Subventionierung entspringt, und ein Regressiver, der aus den Teilnahmemustern folgt. Partizipation an der Riesterförderung ist nicht nur mit hohem Einkommen, sondern auch mit hohem Haushaltsvermögen korreliert.

Das zweite Kapitel löst das Problem eines sozialen Planers, der versucht, die Ungleichheit durch Transfers mit einem festen öffentlichen Budget in einer Verteilung von exogen gegebenen Einkommen zu minimieren. Die geeignete Lösungsmethode hängt von der Zielfunktion ab: Wenn die Zielfunktion konvex ist, kann sie durch einen Interior-Point-Algorithmus gelöst werden. Wenn sie quasikonvex ist, kann die Bisection-Methode verwendet werden. Mit künstlichen und realen Daten zeigen wir, dass die Transfers nicht mit Transfers übereinstimmen müssen, welche die Einkommen am unteren Ende der Verteilung ausgleichen.

Das dritte Kapitel untersucht die Art das Risiko der Erwerbseinkommen in einem Modell des Arbeitsangebots über den Lebenszyklus. Die aktuelle Literatur über den Lebenszykluskonsum, sowie das Spar- und Arbeitsangebotsverhalten konzentriert sich auf Lohnschocks als zentrale Quelle des Risikos. In vorliegenden Papier schlagen wir ein Lebenszyklusmodell vor, das Risiko sowohl in Löhnen als auch in Arbeitsstunden berücksichtigt und mithilfe dessen man die zwei Risikoarten voneinander trennen kann. Zu diesem Zweck schätzen wir einen Parameter, der misst wie permanente Lohnschocks auf den Grenznutzen des Vermögens übertragen werden. Die Schätzung des Modells mit der Panel Study of Income Dynamics (PSID) zeigt, dass sowohl permanente Lohn- als auch Stundenschocks eine wichtige Rolle bei der Erklärung der Querschnittsvarianz des Gewinnwachstums spielen. Dennoch haben permanente Lohnschocks einen quantitativ größeren Einfluss auf die Lebensverdienst. Stundenschocks verbessern die Modellfit erheblich. Die empirische Strategie erlaubt es uns, die Marshallische Elastizität Arbeitsangebots zu schätzen, ohne dass die Verwendung von Konsum- oder Vermögensdaten notwendig ist. Die Marshallische Elastizität ist im Durchschnitt negativ, aber klein. Schließlich zeigen wir die Verbindung zwischen unserem Übertragungsparameter und dem Versicherungsparameter für Konsum bekannt aus Blundell et al. (2008). Wir zeigen, dass die durch unsere Schätzung bestimmte Sensitivität des Konsums hinsichtlich permanenten Lohnschocks mit den relevanten Schätzungen aus der Literatur übereinstimmt.

Das vierte Kapitel ist eine weitere Anwendung eines Lebenszklusmodells des Arbeitsangebots. Diesmal liegt der Schwerpunkt auf der Modellierung des progressiven Steuersystems. Ich quantifiziere den Grad zum dem Individuen permanentem Lohnrisiko ausgesetzt sind und stelle fest, dass das permanente Risiko des Erwerbseinkommens in den USA seit den frühen 2000ern zugenommen hat. Vor allem stieg es während der Finanzkrise in 2008 deutlich an. Eine entsprechende Ausweitung des Versicherungseffekt des progressiven Steuer- und Transfersystems, der diesen Anstieg hätte bremsen können, blieb aus. Ich schätze die Progressivität des Steuer- und Transfersystems unter Verwendung einer Approximation mittels einer Potenzfunktion und werte die Eigenschaften der Approximation hinsichtlich der Darstellung des Steuersystems aus. Der geschätzte Progressionsparameter reicht aus, um den Versicherungseffekt des Steuer- und Transfersystems zu ermitteln. Berechnet man die Transmission des Risikos bei einer Progressivität von null, dann ergeben sich 5% weniger Versicherung der Schocks. Das Risiko der Erwerbseinkommen hätte auf das Niveau vor der Krise reduziert werden können, wenn die Progressivität wesentlich erhöht und der Progressivitätsparameter von 0,93 auf 0,78 gesenkt wurde worden wäre.

# Erklärung

***Erklärung gemäß §4 Abs. 2***

Hiermit erkläre ich, dass ich mich noch keinem Promotionsverfahren unterzogen oder um Zulassung zu einem solchen beworben habe, und die Dissertation in der gleichen oder einer anderen Fassung bzw. Überarbeitung einer anderen Fakultät, einem Prüfungsausschuss oder einem Fachvertreter an einer anderen Hochschule nicht bereits zur Überprüfung vorgelegen hat.

(Unterschrift, Ort, Datum)

***Erklärung gemäß §10 Abs. 3***

Ich habe meine Dissertation soweit im Folgenden nicht anders vermerkt selbständig verfasst habe.

Folgende Hilfsmittel wurde benutzt

- Statistik und Mathematik: Stata, Matlab, Mathematica
- Schriftsatz und Formatierung: LaTeX, Sublime

(Unterschrift, Ort, Datum)