Dynamic household decisions over the life cycle

DISSERTATION
of the University of St. Gallen,
School of Management,
Economics, Law, Social Sciences,
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Economics and Finance

submitted by
Janosch Brenzel-Weiss
from
Wallisellen (Zurich)

Approved on the application of
Prof. Dr. Winfried Koeniger
and
Prof. Dr. Stefan Niemann

Dissertation no. 5315
Difo-Druck GmbH, Untersiemau 2023

Dynamic household decisions over the life cycle

DISSERTATION
of the University of St. Gallen,
School of Management,
Economics, Law, Social Sciences,
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Economics and Finance

submitted by
Janosch Brenzel-Weiss
from
Wallisellen (Zurich)

Approved on the application of
Prof. Dr. Winfried Koeniger
and
Prof. Dr. Stefan Niemann
Prof. Dr. Reto Föllmi
Prof. Guido Cozzi, PhD

Dissertation no. 5315
Difo-Druck GmbH, Untersiemau 2023

The University of St.Gallen, School of Management, Economics, Law, Social Sciences, International Affairs and Computer Science, hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St. Gallen, November 15, 2022

The President:

Prof. Dr. Bernhard Ehrenzeller

To my parents

Acknowledgements

Throughout my PhD, I have been privileged to be surrounded by people who were tremendously supportive in my journey.

First and foremost, I would like to thank my supervisor, Winfried Koeniger. He offered me the opportunity to pursue a PhD at his chair and his support has been invaluable. His door was always open and thanks to his guidance and constructive feedback, I have been able to learn so much from him. I am really thankful I had the opportunity to have him as my mentor, and I am grateful for his time and effort. I would also like to express my gratitude to Thomas Hintermaier, who hosted my academic visit in Bonn and who was very welcoming and supportive. Likewise, I thank Stefan Niemann for being the external referee, as well as Reto Föllmi and Guido Cozzi for being the internal referees of this thesis.

I would like to give special thanks to everyone at the Swiss Institute for Empirical Economic Research for making my time in St. Gallen unforgettable. My thank goes to my co-author Arnau Valladares Esteban from whom I have learnt so much. Our discussions have been very stimulating and a great source of inspiration. Additionally, I would like to thank Conny David and Susanne Moser for their kind support and Marc-Antoine Ramelet, whose advice and support in my first two years have been indispensable. Most notably, I would like to thank Benedikt Lennartz for the time we shared and the engaging interactions we had.

I would like to acknowledge generous financial support by the Swiss National Science Foundation, which made my academic visit in Bonn possible. Moreover, I thank the Study Center Gerzensee for providing an extraordinary learning environment during my first year.

Above all, I am immensely privileged to have such a wonderful family. I would like to express my gratitude to my parents and my sister, who have supported me all along, and my wife, who stood by my side and always believed in me. Your love and encouragement mean everything to me.

Bonn / St. Gallen, June 2022 Janosch Brenzel-Weiss

Summary

This thesis consists of three essays on dynamic household decisions over the life cycle and the estimation of parameters for the analysis. The topics include the estimation of income processes and progressive tax and transfer schedules in Chapter I, the savings and amortisation behaviour of homeowners in Chapter II, and the education choices of students facing drop-out risk in Chapter III.

In Chapter I, I use panel data for Switzerland and Germany to provide estimates required for specifying quantitative life cycle models with idiosyncratic income risk. My estimates include life cycle income profiles, idiosyncratic income risk, and an approximated progressive income tax and transfer schedule. I find that households face income shocks which are sizeable and persistent. When comparing Germany to Switzerland, I find that the shocks are larger in Germany for gross labour income, but instead larger in Switzerland for disposable income. This finding is also reflected in the estimated tax schedules, which are more progressive in Germany than in Switzerland. Finally, I find that aggregate time effects rather than cohort effects explain the evolution of residual income inequality in Switzerland and Germany.

Chapter II is joint work with Winfried Koeniger and Arnau Valladares Esteban from the University of St. Gallen. Motivated by the observation that Swiss homeowners hold significantly more mortgage debt than homeowners in neighbouring countries, we analyse the effect of tax incentives on mortgage amortisation in Switzerland and Germany. For our analysis, we set up a dynamic portfolio-choice model of homeowners, with a focus on the tax incentives for mortgage amortisation and pension savings that determine household leverage. The model includes tax deductions for mortgage interest payments, which make mortgage amortisation relatively less attractive. Furthermore, households have the ability to save for retirement through an illiquid but highly tax-incentivised pension asset. Our results indicate that tax policy matters quantitatively for household leverage and may explain the observed differences. Furthermore, we find that a tax reform abolishing the tax-deductibility of mortgage interest payments in-

creases mortgage amortisation. However, such a tax reform might be offset if the scope for tax deductions of voluntary pension savings is increased.

In Chapter III, I analyse education choices in the presence of drop-out risk. I set up a life cycle model in which students differ by academic ability that determines the drop-out risk they face. I show that the education choices are governed by two key trade-offs. The binary choice between working and attending university generates the well-known trade-off between foregone earnings and returns to education. Introducing intermediate education degrees, such as a degree at a university of applied sciences, which differ in expected returns and drop-out risk, gives rise to a second trade-off. Because agents trade off higher drop-out risk against higher expected returns, they choose the optimal education degree on the risk-return frontier. This leads to a Pareto improvement, and thus improves sorting into education. Furthermore, I analyse how education choices are shaped by progressive income taxation. I find that progressive taxes might *increase* enrolment in tertiary education because it does not only reduce the expected return to education, but also its variance.

Zusammenfassung

Diese Dissertation umfasst drei Beträge, die sich mit den dynamischen Entscheidungen von Haushalten über den Lebenszyklus und der Schätzung von Parametern für die Analyse beschäftigen. Zu den behandelten Themen gehören die Schätzung von Einkommensprozessen und progressiven Steuersystemen im Kapitel I, das Spar- und Amortisationsverhalten von Hauseigentümern im Kapitel II und die Bildungsentscheidungen von Studierenden, die sich mit dem Risiko konfrontiert sehen, ihr Studium nicht zu bestehen (Kapitel III).

Im Kapitel I benutze ich Paneldaten aus der Schweiz und Deutschland, um Schätzungen bereitzustellen, welche als Grundlage zur Spezifizierung von quantitativen Lebenszyklusmodellen mit idiosynkratischem Einkommensrisiko benötigt werden. Meine Schätzungen umfassen Einkommensprofile, Parameter zu idiosynkratischem Einkommensrisiko und eine approximierte progressive Steuerfunktion. Meine Ergebnisse zeigen, dass das Einkommen von Haushalten beträchtlichen und persistenten Schocks ausgesetzt ist. Ein Vergleich zwischen der Schweiz und Deutschland zeigt, dass das Bruttoarbeitseinkommen in Deutschland von grösseren Schocks betroffen ist als in der Schweiz. Beim verfügbaren Einkommen ist die Grösse der Schocks hingegen umgekehrt. Dies zeigt sich auch in den Schätzungen der Steuerprogression, welche in Deutschland höher ist als in der Schweiz. Ausserdem ergibt meine Analyse, dass die Entwicklung der Einkommensungleichheit eher durch Zeiteffekte anstelle von Kohorteneffekten erklärt wird.

Das Kapitel II entstand in Zusammenarbeit mit Winfried Koeniger und Arnau Valladares Esteban von der Universität St. Gallen. Motiviert durch die Beobachtung, dass Schweizer Hauseigentümer deutlich höhere Hypothekenschulden haben als Hauseigentümer in Nachbarländern, analysieren wir den Effekt von Steueranreizen auf die Amortisation von Hypotheken in der Schweiz und Deutschland. Für unsere Analyse erstellen wir ein dynamisches Modell, in welchem Hauseigentümer über ihr Portfolio entscheiden. Der Fokus liegt dabei auf Steueranreizen für die Amortisation von Hypotheken und das Sparen für das Rentenalter, welche den Verschuldungsgrad von Haushalten bestimmen. Wir mo-

dellieren Steuerabzüge für Hypothekarzinsen, welche die Amortisation von Hypotheken relativ weniger attraktiv machen. Weiter besteht die Möglichkeit für das Rentenalter zu sparen, indem man in illiquide, aber stark steuerbegünstigte Vermögenswerte investiert. Unsere Resultate zeigen, dass die Steuerpolitik quantitativ beträchtlichen Einfluss auf den Verschuldungsgrad von Haushalten hat und die beobachteten Differenzen erklären kann. Ausserdem ergibt unsere Analyse, dass eine Steuerreform, welche Steuerabzüge für Hypothekarzinsen abschafft, zu mehr Amortisation führt. Allerdings hat eine Erhöhung der Maximalbeträge für steuerbegünstigtes Alterssparen den gegenteiligen Effekt, sodass die Auswirkungen einer solchen Steuerreform aufgehoben würden.

Im Kapitel III analysiere ich Bildungsentscheidungen, die das Risiko berücksichtigen, den Abschluss nicht zu schaffen. Dazu erstelle ich ein Lebenszyklusmodell mit Studierenden, die sich in ihren akademischen Fähigkeiten unterscheiden, welche die Wahrscheinlichkeit des Nichtbestehens beeinflusst. Ich zeige, dass Studierende zwei entscheidende Abwägungen machen müssen. Falls nur die binäre Entscheidung zwischen einem Universitätsstudium und dem direkten Eintritt in den Arbeitsmarkt zu treffen ist, entsteht die bekannte Abwägung zwischen entgangenem Lohn und zukünftigen erwarteten Bildungsrenditen. Durch die Einführung von zusätzlichen Bildungswegen, wie beispielsweise einem Fachhochschulstudium, die sich durch die erwarteten Bildungsrenditen und das Risiko den Abschluss nicht zu schaffen unterscheiden, ergibt sich eine zweite Abwägung. Weil Studierende das höhere Risiko gegen die höheren Renditen abwägen, treffen sie die optimale Bildungsentscheidung entlang der Risiko-Rendite-Grenze. Dies führt zu einer Paretoverbesserung und verbessert die Übereinstimmung von akademischen Fähigkeiten und Bildungsweg. Abschliessend analysiere ich, welchen Einfluss progressive Einkommensbesteuerung auf Bildungsentscheidungen hat. Meine Analyse ergibt, dass progressive Steuern die Anzahl Studierenden im tertiären Bildungsbereich *erhöhen* kann, da diese nicht nur den Erwartungswert der Bildungsrendite senkt, sondern auch deren Varianz.

Contents

I | Estimates of income dynamics for structural models: evidence for Switzerland and Germany

Abstract

I use panel data for Switzerland and Germany to provide estimates required for specifying quantitative life cycle models with idiosyncratic income risk. I estimate the life cycle profile of household and individual income, idiosyncratic income risk, the evolution of residual income inequality, and an approximated progressive income tax and transfer schedule. I find that life cycle income grows at an annual rate of 1.1% in Switzerland until it peaks at age 52 and 1.7% in Germany, peaking at age 54. The estimates of the stochastic income process reveal that income shocks are sizable and persistent. The shocks are higher in Germany for gross labour income, whereas they are higher in Switzerland for disposable income. This is also reflected in the estimated income tax and transfer schedule, which is more progressive in Germany than in Switzerland. Estimates for the U.S. in the literature are between my estimates for Switzerland and Germany. Similar as in the U.S., I find that aggregate time effects rather than cohort effects explain the evolution of residual income inequality in Switzerland and Germany.

Keywords: Income inequality, Income dynamics, Taxation.
JEL Codes: E25, H24, J31.

1 Introduction

Structural decision models with heterogeneity between households, for example because of uninsurable income risk, require estimates capturing various dimensions of heterogeneity. An important source of heterogeneity across households is income, which has deterministic and stochastic components. Income heterogeneity is also shaped by the redistribution implied by the tax and transfer system.

I use data from the Swiss Household Panel (SHP) and the German Socio-Economic Panel (SOEP) to provide estimates for life cycle income profiles, idiosyncratic income risk, and tax progressivity. To estimate income heterogeneity, I decompose income into a deterministic life cycle profile and a stochastic income process. I characterise the life cycle profiles in two ways, using age dummies and a fourth-order age polynomial. For the income process, I provide estimates for an AR(1) process and a process with a transitory and a permanent component following Blundell et al. (2008). A parsimonious and yet effective way to model progressive taxation is to use a simple tax function as used, for example, in Heathcote et al. (2017). I follow their approach and provide estimates for tax progressivity in Switzerland and Germany. Furthermore, I follow Heathcote et al. (2005) to investigate the relative importance of time vs. cohort effects when estimating the evolution of residual income inequality over the life cycle.

I find that life cycle income profiles are steeper in Germany, where gross household labour income has an average growth rate of around 1.7% per year until it peaks at age 54. In Switzerland, life cycle income grows at a rate of around 1.1% per year until it peaks at age 52. Looking at the difference between life cycle profiles on the household and the individual level, I find that life cycle profiles at the individual level are steeper at the beginning of the life cycle and do not fall towards the end of the life cycle. Comparing life cycle profiles between education groups, I find only a small difference in income at the beginning of the life cycle, but much steeper income profiles for individuals with a tertiary degree compared to individuals without tertiary education.

Estimating an AR(1) process for stochastic income, I find that income shocks are persistent with an income autocorrelation between 0.73 and 0.82 depending on the sample and income definition. Estimating a process with a transitory and a permanent component as in Blundell et al. (2008), I find that the variance of the permanent and transitory shocks is between 0.02 and 0.03 in Switzerland. In Germany, the variance of the permanent shock for gross household income is higher, at almost 0.04. In contrast, the shocks for disposable income are between 0.01 and 0.02 in Germany, and thus lower than in Switzerland. Furthermore, the shocks are higher for individual earnings, indicating the effect of insurance within the household.

The fact that income shocks are higher in Germany than in Switzerland for gross household labour income and vice versa for disposable income also shows in my estimates for the degree of tax progressivity. Using a constant progressivity tax function with progressivity τ as in Heathcote et al. (2017), I find a level of tax progressivity of $\tau = 0.287$ for Germany and $\tau = 0.092$ for Switzerland in my main estimation, thus showing a higher degree of progressivity in Germany. For comparison, Heathcote et al. (2017) report an estimate of $\tau = 0.181$ for the United States using a similar sample. Furthermore, I find that while progressivity in Switzerland is mainly driven by direct taxes, social security contributions and transfers also play an important role in Germany. Additionally, looking at the variance of gross labour income and (after-tax) disposable income in Switzerland and Germany, I find that while gross labour income is much more unequally distributed in Germany than in Switzerland, disposable income has a very similar variance in both countries, highlighting the role of tax progressivity in Germany.

Finally, I find that time effects rather than cohort effects explain the evolution of residual income inequality in Switzerland and Germany. This is consistent with the findings in Heathcote et al. (2005) who find similar results for the United States.

Related Literature. My contribution fits into the broader literature of inequality in macroeconomics as surveyed by Quadrini and Ríos-Rull (2015). The discussion on heterogeneity over the life cycle goes back to Deaton and Paxson (1994). Later contributions include, among others, Storesletten et al. (2004) and Huggett et al. (2006, 2011). A seminal contribution to the study of idiosyncratic earnings risk is Aiyagari (1994). I contribute to this literature using an income process with a permanent and transitory component as in Blundell et al. (2008). Furthermore, I contribute to the discussion of estimating income inequality over the life cycle with cohort vs. time effects as done by Heathcote et al. (2005) for the United States.

When estimating tax functions, my contribution builds on Heathcote et al. (2017), who provide an analytical framework to study tax progressivity. Additionally, Guner et al. (2014) provide parametric estimates of various tax functions for the United States.

Fuchs-Schündeln et al. (2010) provide cross-sectional facts for Germany in the special issue of the *Review of Economic Dynamics*, which also contains comparable evidence for eight other countries (Krueger et al., 2010).[1] Apart from providing new evidence for Switzerland, my contribution differs from theirs in three aspects. First, Fuchs-Schündeln et al. (2010) use data up to the year 2004, while I use data until 2018. Second, while there is some overlap in the estimated statistics, there are differences in evidence I cover in my analysis. Fuchs-Schündeln et al. (2010) analyse inequality trends between West Germany and East Germany. Furthermore, their focus is on trends over time, while my focus is on the evolution of income over the life cycle. Third, Fuchs-Schündeln et al. (2010) cover several additional topics, such as consumption or the German reunification, whereas I provide estimates for tax progressivity. Where my analysis is similar to theirs, I will compare my results to their findings.

Outline. I describe the data I use in Section 2. In Section 3, I investigate the relative importance of time and cohort effects. Section 4 provides stylised facts on income profiles in Switzerland and Germany. In Sections 5 and 6, I show my estimates of the parameters characterising the two income processes and the tax function for various samples. Section 7 concludes.

2 Data

I use data from the Swiss Household Panel (SHP) and the German Socio-Economic Panel (SOEP), which are unbalanced, annual panel data sets. The SHP consists of 20 waves from 1999 to 2018 of which I use the waves since 2000, as the first wave (1999) does not include information on social security pensions and total household taxes. The SOEP consists of 35 waves from 1984 to 2018. I use the waves since 1992, which is the first wave that includes both West Germany and East Germany. For both countries, I use the cross-national equivalent files (CNEF) which offer standardised variables from the SHP and SOEP core samples and thus improve comparability. For further details on the SHP and SOEP, I refer to Voorpostel et al. (2020) and Goebel et al. (2019).

[1] These countries are Canada, Italy, Mexico, Russia, Spain, Sweden, the United Kingdom, and the United States.

For each country, I construct two core samples, one sample at the household level and one sample at the individual level. Focussing on the employed population, I restrict the sample at the individual level to individuals between 26 and 60 years of age with annual working hours between 520 and 2600, i.e. up to 50 hours per week. At the household level, I restrict the sample to households where the household head is between 26 and 60 years of age and where the total annual working hours of the household is between 520 and 5200, i.e. up to 50 hours per week for two full-time equivalents per household. These sample restrictions are similar to Fuchs-Schündeln et al. (2010) and Heathcote et al. (2005).[2]

Additionally, I construct two extended samples, where I also include households and individuals with less than 520 annual working hours. The extended sample is of particular interest when studying income processes and tax progressivity, as it includes households and individuals with low income draws. Unless indicated otherwise, I will use the core sample going forward. At the individual level, my income variable of interest is gross individual labour earnings. At the household level, I construct five income variables:

- *Gross household labour income*: household labour earnings before social security contributions and transfers such as unemployment benefits.
- *Gross household income*: the sum of gross household labour income and capital income.
- *Net household labour income*: household labour earnings after social security contributions and transfers of all household members.
- *Net household income*: the sum of net household labour income and capital income.
- *Disposable income*: labour earnings and capital income after social security contributions and transfers net of tax payments.

I adjust income at the household level using the *square root equivalence scale* recently adopted by the OECD.[3] This equivalence scale is very similar to the scale reported in Fernández-Villaverde and Krueger (2007). Therefore, I divide income by the square root of the households' size. Furthermore, I adjust all

[2]See Appendix A for information on how many observations I drop because of these restrictions.

[3]See http://www.oecd.org/economy/growth/OECD-Note-EquivalenceScales.pdf

nominal variables by the consumer price index using the year 2015 as the base year. Further details of my data cleaning process and sample selection including summary statistics are reported in Appendix A.

3 Estimating life cycle profiles: time vs. cohort effects

3.1 Overview

The goal of estimating life cycle profiles is to assess how certain moments from the data vary along the age dimension. To extract the age effects, I proceed in two steps, as usually done in the literature.[4] In the first step, I calculate the moment of interest for each combination of age j and year t. In the second step, I assume that those moments vary along the three dimensions age j, time t and cohort c. Let $m_{j,t}$ denote the moment of interest for age group j at time t.

I estimate the age effect α_j using the following statistical model:[5]

$$m_{j,t} = \alpha_j + \beta_c + \gamma_t + \epsilon_{j,t}\,, \tag{3.1}$$

where β_c and γ_t are cohort and time fixed effects. As already mentioned by Deaton and Paxson (1994), the fact that age, time, and cohort effects are linearly dependent causes an identification problem. Given this problem, we have to take a stand on whether to include time or cohort effects in our estimation.

When estimating life cycle profiles for average income, it is rather hard to argue against the inclusion of time effects, as income usually follows a clear time trend in a growing economy such as Switzerland or Germany. However, for higher moments, e.g. the variance of income, this question is not as clear. Therefore, I will try to assess the relative importance of time and cohort effects in this section. For this, I proceed in three steps. In Section 3.2, I show life cycle profiles for the variance of several income variables of interest, providing two alternative measures of the age effects, which correspond to the two alternatives of including either time effects and assuming $\beta_c = 0$ or vice versa. This will highlight, that the choice between time and cohort effects is not an innocuous

[4] See for example Huggett et al. (2011).

[5] Note, that if we are interested only in the first moment, we can directly proceed with step two and replace $m_{j,t}$ by the variable of interest, denoting $y_{i,j,t}$ the observation for household i of age j at time t.

one. In Section 3.3, I illustrate graphically that for Switzerland and Germany, the variance of income is mainly driven by time effects. Although the graphical illustration lacks rigour, it will provide insight into how the cross-sectional variance of income has changed over time and by cohort. Finally, in Section 3.4, I treat the question in a more rigorous manner. For this, I follow Heathcote et al. (2005), who show using U.S. data that time effects play an important role, while there is no evidence for sizeable cohort effects. I replicate their approach with data from the SOEP and the SHP to show that the same holds for Switzerland and Germany.

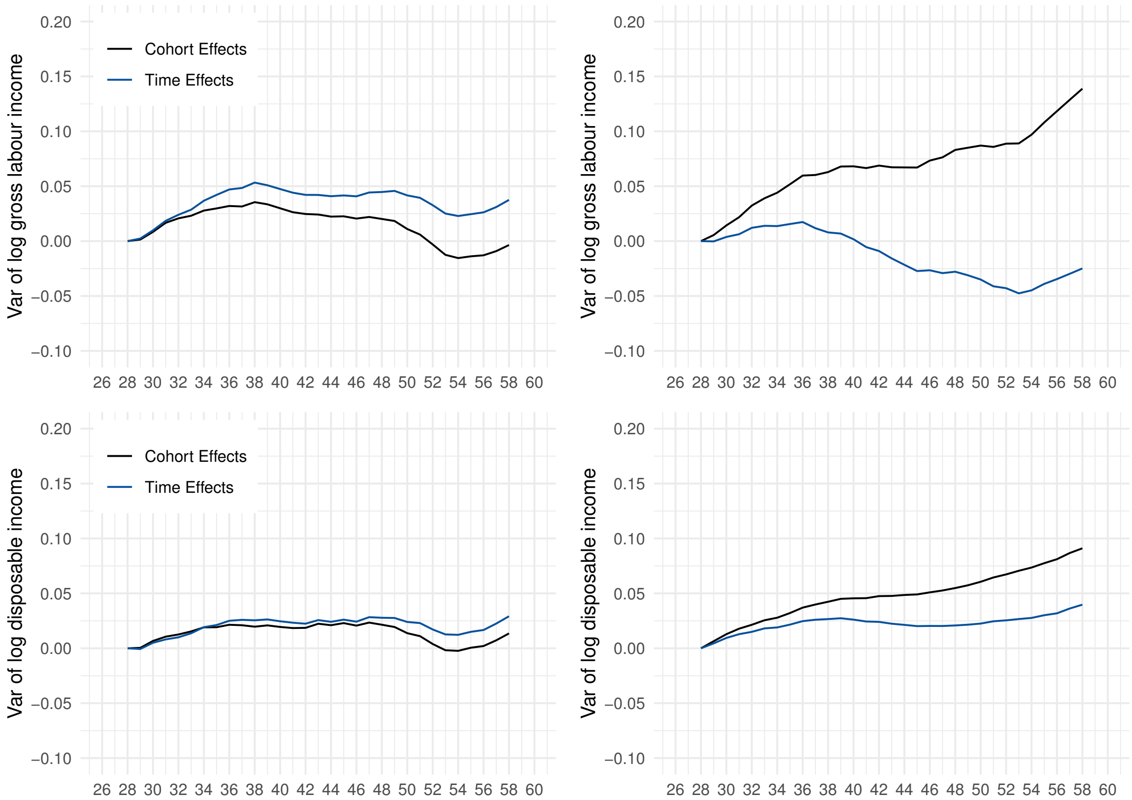
3.2 Cross-sectional variance over the life cycle

I estimate life cycle profiles for the variance of five variables. At the household level, I estimate the variance of the log of gross household labour income and disposable income as defined in Section 2. At the individual level, I estimate the variance of log earnings, the log wage rate, and log annual working hours. For all five variables, I proceed in two steps as described in the previous section.

As done by Heathcote et al. (2005), I apply a five-year rolling window for age. Thus, a household head or individual falls into age group j, if their actual age is in the interval $[j-2, j+2]$. In the first step, I calculate the variance for each age group in each year and estimate the age effect using the statistical model characterised in equation (3.1). I estimate the life cycle profiles in two different ways. First, I estimate them including cohort effects β_c and assuming the time effects $\gamma_t = 0$. Second, I estimate them including time effects γ_t and assuming the cohort effects $\beta_c = 0$. In Figures 3.1 and 3.2, I depict the resulting profiles in deviation from the values at the beginning of the life cycle. The magnitudes are stronger on the individual level, which is in line with the presence of intra-household insurance. Furthermore, the evidence for wages and hours reveals that the variance over the life cycle is driven by increased dispersion in wages rather than hours worked.

The life cycle profiles shown for Germany partly replicate Figure 18 from Fuchs-Schündeln et al. (2010) but with the most recent SOEP data available.[6] As they also point out, the figure illustrates that the choice between including time or cohort effects matters. Furthermore, two observations are worth mentioning. First, the life cycle profiles for Germany seem to be more sensitive to the choice between including time or cohort effects. As I show in Section 3.3, this is due

[6] Fuchs-Schündeln et al. (2010) include the years 1984–2004 and only use observations from West Germany. I use the years 1992–2018 and include both East and West Germany.



(a) Switzerland

(b) Germany

Figure 3.1: Household income inequality over the life cycle.

to the fact that the variance of income has increased in Germany and hence, including time effects has a strong impact on the resulting life cycle profiles. In contrast, the variance of income has been rather stable in Switzerland. However, Gazareth and Suter (2016) show for Switzerland, that survey data such as the SHP usually fail to fully represent the bottom and top incomes. Foellmi and Martínez (2017) show that top income shares in Switzerland have increased in Switzerland since the 1980s. Thus, the importance of time effects also applies to Switzerland. Second, when including cohort effects, the magnitude of the change in the variance of income in Germany is much smaller than reported in Fuchs-Schündeln et al. (2010). This can be explained by the fact that Fuchs-Schündeln et al. (2010) use data until 2004, whereas I use data until 2018. As I show in Section 3.3, when omitting time effects, the life cycle profiles are highly sensitive to the time period used.

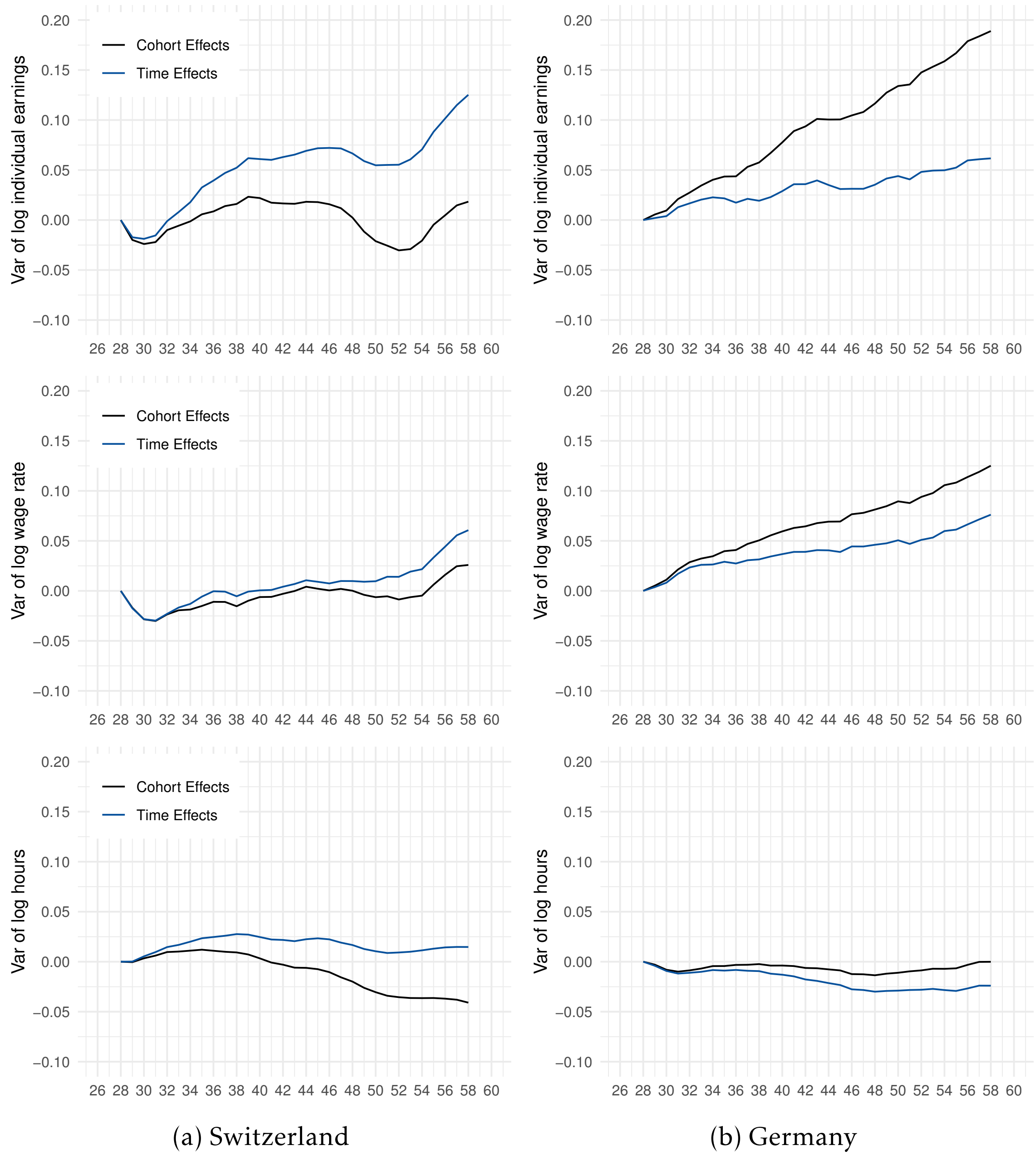


Figure 3.2: Individual income inequality over the life cycle.

3.3 Cross-sectional variance over time and by cohort

To assess the relative importance of time and cohort effects, we have to further investigate how the variance of the variable of interest evolves over time and how the evolution over time differs by cohort. In this section I treat this question graphically focusing on gross household labour income. This provides an intuitive understanding of the contribution of time and cohort effects to the estimated life cycle profile. In Section 3.4, I analyse the question more systematically following the approach of Heathcote et al. (2005).

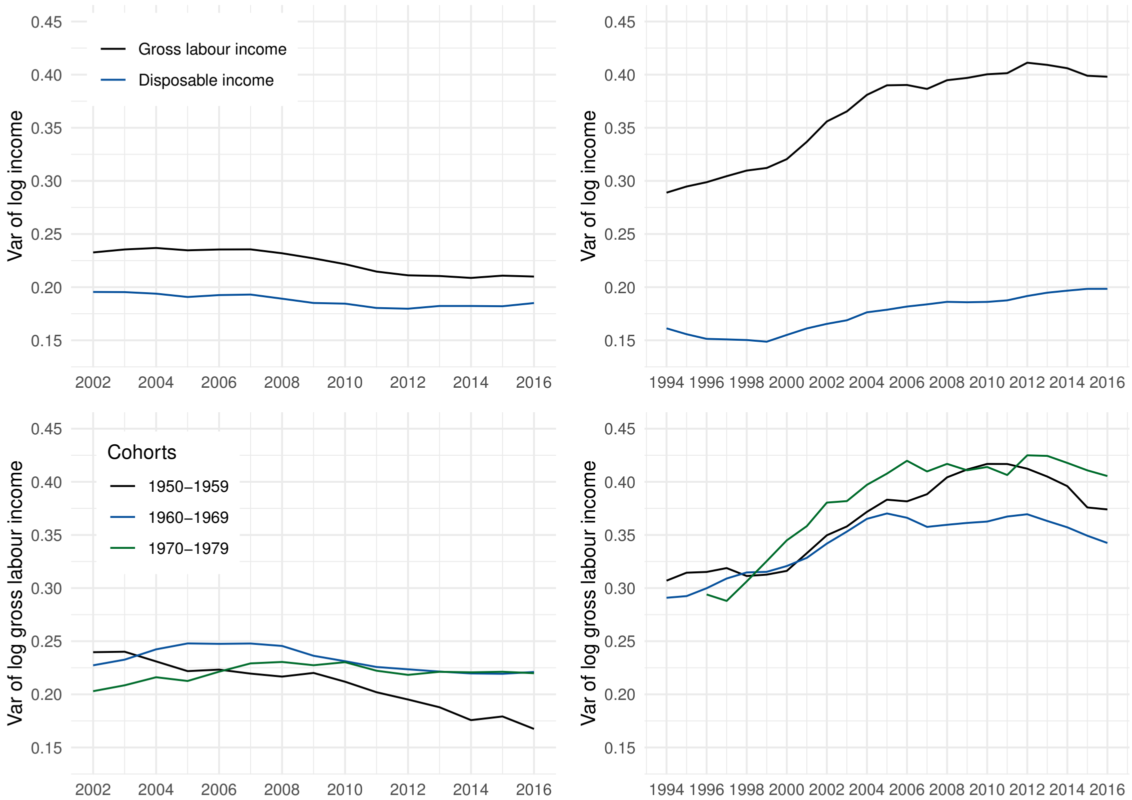
Figure 3.3 shows the variance of log income in two different ways. The top graphs show how the variance of gross household labour income and disposable income have evolved over time. Two things are worth pointing out. First, while the variance of disposable income is similar in Switzerland and Germany, the variance of gross labour income is much higher in Germany. As I show in Section 6, the German tax rates are both higher and more progressive than the tax rates in Switzerland, which is consistent with the findings in Figure 3.3. Second, Germany has experienced a substantial increase in the variance of income during the 90s and early 2000s.[7] The bottom graphs of Figure 3.3 show the evolution of the variance of gross labour income within three different cohorts.[8] As can be seen from the figure, all three cohorts closely follow the time trend observed in the full sample. While the time trend is rather striking for Germany, there seems to be little evidence that the evolution of the variance of income differs substantially between cohorts.

To further illustrate this point, I estimate the life cycle profile for the variance of gross labour income in Germany using samples from two different time periods. The estimates are shown in Figure 3.4. On the left side, I estimate the life cycle profile using the observations up to the year 2004, as done by Fuchs-Schündeln et al. (2010).[9] On the right side, I estimate the life cycle profile using the observations for the years 2005 to 2018. As discussed above, the crucial difference between these two time periods is that during the period up to 2004,

[7] See Fuchs-Schündeln et al. (2010) for a detailed analysis of the increase in market incomes after the German unification.

[8] The cohorts chosen correspond to the cohorts that are in the sample for (almost) the full sample period.

[9] Note that Fuchs-Schündeln et al. (2010) also include the data from before the German unification. However, the argument hinges on the fact that this period includes the 90s and early 2000s, where the variance increased the most. Therefore, this difference should not affect the argument.

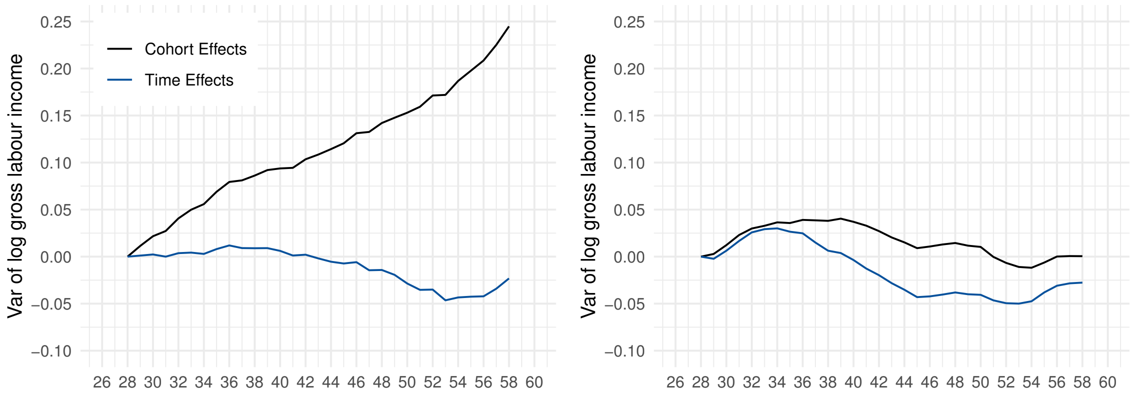


(a) Switzerland (b) Germany

Figure 3.3: Cross-sectional variance over time and by cohort.

the variance of gross labour income increased substantially, while in the period from 2005 to 2018 the increase was only moderate and even included periods of decreasing variance for some cohorts.[10] Observe that when including time effects, both graphs display similar profiles. However, when including cohort effects and omitting time effects, the profiles differ substantially. As the variance of income has increased during the first sample period, this increase is captured in the life cycle profile when time effects are omitted. This is different in the second sample period, for which the changes of the variance of log gross labour income over the life cycle remain within 0.05 in this case. This observation strongly suggests that failing to include time effects when estimating life cycle profiles distorts the profiles, as the increase in the variance is erroneously attributed to an age effect. I make this point more systematically in the next section.

[10]The choice of the sample split is done for comparison with Fuchs-Schündeln et al. (2010). Choosing another year in the early 2000s to split the sample yields very similar results.



(a) Germany 1992–2004

(b) Germany 2005–2018

Figure 3.4: Life cycle profiles for two different sample periods.

3.4 Time vs. cohort effects

In this section, I replicate the test of Heathcote et al. (2005) for cohort or time effects for Switzerland and Germany. Using data from the Panel Study of Income Dynamics (PSID), they show that without including time effects, one cannot account for the change in inequality over the last decades observed in the PSID.

Heathcote et al. (2005) start with a general model for the moment $x(a,t,c)$, which is a function of age a, time t and cohort c. Furthermore, they assume separability between the effects and use $c = t - a$ to describe the model as follows.

$$x(a,t,t-a) = g_1(a) + g_2(t) + g_3(t-a),$$

where in the statistical model I describe above, this corresponds to a regression with dummies for age, time and cohorts, which, as discussed, is infeasible in practice. Now, Heathcote et al. (2005) define three measures which can be obtained from the data:

(i) The change of the moment x over time and within age group a:

$$\Delta x^a_{t,t+1} = \Delta g_2(t) + \Delta g_3(t-a).$$

(ii) The change of the moment x over time and within cohort $c = t - a$:

$$\Delta x^c_{t,t+1} = \Delta g_1(a) + \Delta g_2(t).$$

(iii) The change of the moment x within period and between age groups a and $a+1$:

$$\Delta x^{t}_{a,a+1} = \Delta g_1(a) + \Delta g_3(t-a).$$

Using these three measures Heathcote et al. (2005) test three hypotheses:

(i) If time effects are small, the within cohort change $\Delta x^{c}_{t,t+1}$ should be constant over time.

(ii) If cohort effects are small, the between age group change $\Delta x^{t}_{a,a+1}$ should be constant over time.

(iii) If time effects are sizeable, $\Delta x^{c}_{t,t+1}$ and $\Delta x^{a}_{t,t+1}$ should be highly correlated, as the time effect is their shared component.[11]

Tables 3.1 and 3.2 replicate Table 1 in Heathcote et al. (2005) for the variance of log gross labour income, individual earnings, wages and annual working hours in Switzerland and Germany. The tables show the average annual change of those variances within cohorts, within age group and between age groups as described above. The last column shows the correlation between $\Delta x^{c}_{t,t+1}$ and $\Delta x^{a}_{t,t+1}$, thus testing the hypotheses that those should be highly correlated in the presence of sizeable time effects.

Looking at Tables 3.1 and 3.2, we can assess which of the three hypotheses stated may be rejected. First, observe that the majority of within-cohort changes $\Delta x^{c}_{t,t+1}$ are significantly different from each other, which is counterfactual to what hypotheses (i) suggests and thus points to significant time effects in the data. Second, if cohort effects are small, then according to hypothesis (ii) the between age group changes $\Delta x^{t}_{a,a+1}$ should be constant. This cannot be rejected by the moments displayed. Third, we observe that $\Delta x^{c}_{t,t+1}$ and $\Delta x^{a}_{t,t+1}$ are highly correlated, which is in favour of sizeable time effects.

To make this analysis more precise, I calculate the within cohort variance changes $\Delta x^{c}_{t,t+1}$ and the between age group variance changes $\Delta x^{t}_{a,a+1}$ and run a regression on a full set of time dummies. If the resulting F-statistic is statistically significant, then the hypothesis that the effects are constant over time can be rejected.

[11] As Heathcote et al. (2005) point out, this has already been noted by Juhn et al. (1993).

Table 3.1: Average annual change within cohort, within age group and between age groups in Switzerland.

	2000–2003	2004–2008	2009–2013	2014–2018	Correlation
		Variance of log gross labour income			
Within cohort $\Delta x^c_{t,t+1}$	0.397	0.344	−0.640	0.015	
	(−0.096, 0.890)	(−0.162, 0.850)	(−1.033, −0.247)	(−0.336, 0.366)	0.994
Within age group $\Delta x^a_{t,t+1}$	0.262	0.171	−0.719	−0.022	
	(−0.219, 0.744)	(−0.364, 0.707)	(−1.166, −0.272)	(−0.393, 0.350)	
Between age group $\Delta x^t_{a,a+1}$	0.125	0.175	0.109	0.062	
	(−0.104, 0.354)	(−0.069, 0.419)	(−0.149, 0.367)	(−0.162, 0.285)	
		Variance of log individual earnings			
Within cohort $\Delta x^c_{t,t+1}$	1.508	1.048	−0.906	−0.620	
	(0.264, 2.751)	(0.108, 1.987)	(−1.815, 0.003)	(−1.363, 0.122)	0.997
Within age group $\Delta x^a_{t,t+1}$	0.890	0.525	−1.213	−0.774	
	(−0.345, 2.125)	(−0.405, 1.455)	(−2.165, −0.260)	(−1.576, 0.027)	
Between age group $\Delta x^t_{a,a+1}$	0.606	0.523	0.304	0.183	
	(0.142, 1.071)	(0.010, 1.035)	(−0.216, 0.825)	(−0.251, 0.617)	
		Variance of log wages			
Within cohort $\Delta x^c_{t,t+1}$	0.682	0.569	−0.637	0.126	
	(−0.473, 1.836)	(−0.312, 1.450)	(−1.321, 0.047)	(−0.566, 0.818)	0.982
Within age group $\Delta x^a_{t,t+1}$	0.294	0.248	−0.666	0.065	
	(−0.877, 1.465)	(−0.626, 1.121)	(−1.364, 0.031)	(−0.671, 0.801)	
Between age group $\Delta x^t_{a,a+1}$	0.397	0.277	0.027	0.091	
	(0.027, 0.767)	(−0.129, 0.684)	(−0.347, 0.401)	(−0.305, 0.488)	
		Variance of log hours			
Within cohort $\Delta x^c_{t,t+1}$	0.435	−0.108	−0.098	−0.534	
	(0.122, 0.747)	(−0.423, 0.208)	(−0.384, 0.187)	(−0.749, −0.320)	0.999
Within age group $\Delta x^a_{t,t+1}$	0.340	−0.172	−0.136	−0.526	
	(0.016, 0.664)	(−0.482, 0.139)	(−0.432, 0.160)	(−0.741, −0.312)	
Between age group $\Delta x^t_{a,a+1}$	0.090	0.088	0.019	−0.006	
	(−0.063, 0.243)	(−0.058, 0.235)	(−0.143, 0.182)	(−0.114, 0.102)	

Note: The table follows Heathcote et al. (2005). All values are multiplied by 100. The last column shows the correlation between the within cohort changes and the within age group changes. 95% confidence intervals in brackets.

Table 3.2: Average annual change within cohort, within age group and between age groups in Germany.

	1992–1997	1998–2003	2004–2008	2009–2013	2014–2018	Correlation
			Variance of log gross labour income			
Within cohort $\Delta x^c_{t,t+1}$	0.682	0.894	0.050	0.256	−0.454	
	(0.140, 1.223)	(0.364, 1.423)	(−0.674, 0.775)	(−0.389, 0.901)	(−1.048, 0.140)	0.991
Within age group $\Delta x^a_{t,t+1}$	0.701	1.012	0.160	0.257	−0.265	
	(0.129, 1.273)	(0.459, 1.565)	(−0.589, 0.908)	(−0.430, 0.944)	(−0.883, 0.354)	
Between age group $\Delta x^t_{a,a+1}$	−0.023	−0.112	−0.164	−0.027	−0.231	
	(−0.310, 0.265)	(−0.360, 0.137)	(−0.572, 0.244)	(−0.382, 0.329)	(−0.583, 0.120)	
			Variance of log individual earnings			
Within cohort $\Delta x^c_{t,t+1}$	0.333	1.928	0.448	0.513	−2.042	
	(−0.345, 1.011)	(1.258, 2.598)	(−0.263, 1.158)	(−0.318, 1.344)	(−2.704, −1.381)	0.998
Within age group $\Delta x^a_{t,t+1}$	0.107	1.761	0.267	0.141	−2.167	
	(−0.561, 0.775)	(1.063, 2.459)	(−0.486, 1.021)	(−0.719, 1.000)	(−2.860, −1.474)	
Between age group $\Delta x^t_{a,a+1}$	0.275	0.171	0.150	0.407	0.079	
	(0.010, 0.541)	(−0.142, 0.484)	(−0.280, 0.580)	(−0.010, 0.824)	(−0.287, 0.445)	
			Variance of log wages			
Within cohort $\Delta x^c_{t,t+1}$	−0.601	1.252	0.416	0.092	−0.526	
	(−1.074, −0.128)	(0.799, 1.706)	(−0.132, 0.964)	(−0.529, 0.713)	(−0.932, −0.119)	0.998
Within age group $\Delta x^a_{t,t+1}$	−0.850	1.059	0.112	−0.223	−0.745	
	(−1.317, −0.383)	(0.600, 1.519)	(−0.447, 0.670)	(−0.892, 0.446)	(−1.181, −0.308)	
Between age group $\Delta x^t_{a,a+1}$	0.252	0.215	0.267	0.332	0.163	
	(0.048, 0.456)	(0.007, 0.422)	(−0.028, 0.562)	(0.045, 0.619)	(−0.125, 0.452)	
			Variance of log hours			
Within cohort $\Delta x^c_{t,t+1}$	0.331	0.052	0.040	0.081	−0.591	
	(0.106, 0.556)	(−0.142, 0.246)	(−0.231, 0.310)	(−0.219, 0.381)	(−0.814, −0.368)	0.996
Within age group $\Delta x^a_{t,t+1}$	0.371	0.144	0.166	0.140	−0.466	
	(0.137, 0.605)	(−0.062, 0.351)	(−0.109, 0.440)	(−0.157, 0.438)	(−0.693, −0.240)	
Between age group $\Delta x^t_{a,a+1}$	−0.029	−0.087	−0.137	−0.069	−0.111	
	(−0.140, 0.081)	(−0.203, 0.029)	(−0.253, −0.021)	(−0.213, 0.074)	(−0.218, −0.004)	

Note: The table follows Heathcote et al. (2005). All values are multiplied by 100. The last column shows the correlation between the within cohort changes and the within age group changes. 95% confidence intervals in brackets.

Table 3.3 shows the p-values for the F-statistic of each regression. Furthermore, I display the correlation between $\Delta x^{c}_{t,t+1}$ and $\Delta x^{a}_{t,t+1}$.[12] Again, we can test the three hypotheses from above. First, we see that the F-statistic resulting from the regression of the within cohort variance changes $\Delta x^{c}_{t,t+1}$ on a full set of time dummies is highly significant. Therefore, we can reject the hypothesis that time effects are small. Second, the F-statistic resulting from the regression of the between age group variance changes $\Delta x^{t}_{a,a+1}$ on a full set of time dummies is not significant on any conventional significance level. Thus, we cannot reject the hypothesis that cohort effects are small. Third, the correlation between $\Delta x^{c}_{t,t+1}$ and $\Delta x^{a}_{t,t+1}$ is very high, which points towards sizeable time effects.

3.5 A case for the use of time over cohort effects

In this section, I have shown that when estimating life cycle profiles for the variance of income, wages and working hours, the choice between including time or cohort effects is an important one. The variance of income in Germany has increased substantially during the 90s and early 2000s. Therefore, omitting time effects would result in very steep life cycle profiles, whereas including time effects shows that the variance of income over the life cycle is more stable.

Replicating the work of Heathcote et al. (2005) with data from Switzerland and Germany, I show that the hypothesis that time effects are small can be rejected for both Switzerland and Germany, while the hypothesis that cohort effects are small cannot. Therefore, the evidence speaks for the use of time effects over cohort effects when estimating the evolution of income equality over the life cycle. Hence, the evidence in Figures 3.1 and 3.2 implies that the variance of household gross labour income over the life cycle starts decreasing beyond age forty. In Germany, it reaches even lower levels than initially. At the same time, the variance of individual earnings has an increasing trend over the whole life cycle in both countries. Quantitatively, the increase over the life cycle is still smaller than reported for the U.S. in Heathcote et al. (2005), where the variance of log earnings increases by around 0.2, whereas the increase is 0.10 – 0.15 in Switzerland and only just above 0.05 in Germany.

[12] Note that the correlation coefficients vary from those in Tables 3.1 and 3.2. In Table 3.3, I calculate the average within cohort and within age group changes per year, whereas in Tables 3.1 and 3.2, I calculate the average changes per time period displayed in the header of the tables.

Table 3.3: Test for significance of time-dummy coefficients.

	Switzerland	Germany
Variance of log gross labour income	*p-values*	
Within cohort $\Delta x^c_{t,t+1}$	$< 10^{-3}$	$< 10^{-3}$
Between age group $\Delta x^t_{a,a+1}$	1.000	1.000
$\mathrm{Cor}(\Delta x^c_{t,t+1}, \Delta x^a_{t,t+1})$	0.993	0.998
	(0.980, 0.997)	(0.994, 0.999)
Variance of log individual earnings	*p-values*	
Within cohort $\Delta x^c_{t,t+1}$	$< 10^{-3}$	$< 10^{-3}$
Between age group $\Delta x^t_{a,a+1}$	0.999	1.000
$\mathrm{Cor}(\Delta x^c_{t,t+1}, \Delta x^a_{t,t+1})$	0.998	0.997
	(0.994, 0.999)	(0.993, 0.999)
Variance of log wages	*p-values*	
Within cohort $\Delta x^c_{t,t+1}$	$< 10^{-3}$	$< 10^{-3}$
Between age group $\Delta x^t_{a,a+1}$	0.999	1.000
$\mathrm{Cor}(\Delta x^c_{t,t+1}, \Delta x^a_{t,t+1})$	0.998	0.998
	(0.994, 0.999)	(0.995, 0.999)
Variance of log hours	*p-values*	
Within cohort $\Delta x^c_{t,t+1}$	$< 10^{-3}$	$< 10^{-3}$
Between age group $\Delta x^t_{a,a+1}$	1.000	0.997
$\mathrm{Cor}(\Delta x^c_{t,t+1}, \Delta x^a_{t,t+1})$	0.998	0.997
	(0.996, 0.999)	(0.993, 0.999)

Note: The table shows the p-values from the F-statistic resulting from a regression of the within cohort and between age group variance changes on a full set of time dummies. The third row of each block shows the correlation between the within cohort changes and the within age group changes. 95% confidence intervals in brackets.

4 Life cycle income profiles: stylised facts

In this section, I estimate life cycle profiles for gross household labour income, disposable income and individual earnings. I estimate the profiles as discussed in the previous section, using the statistical model

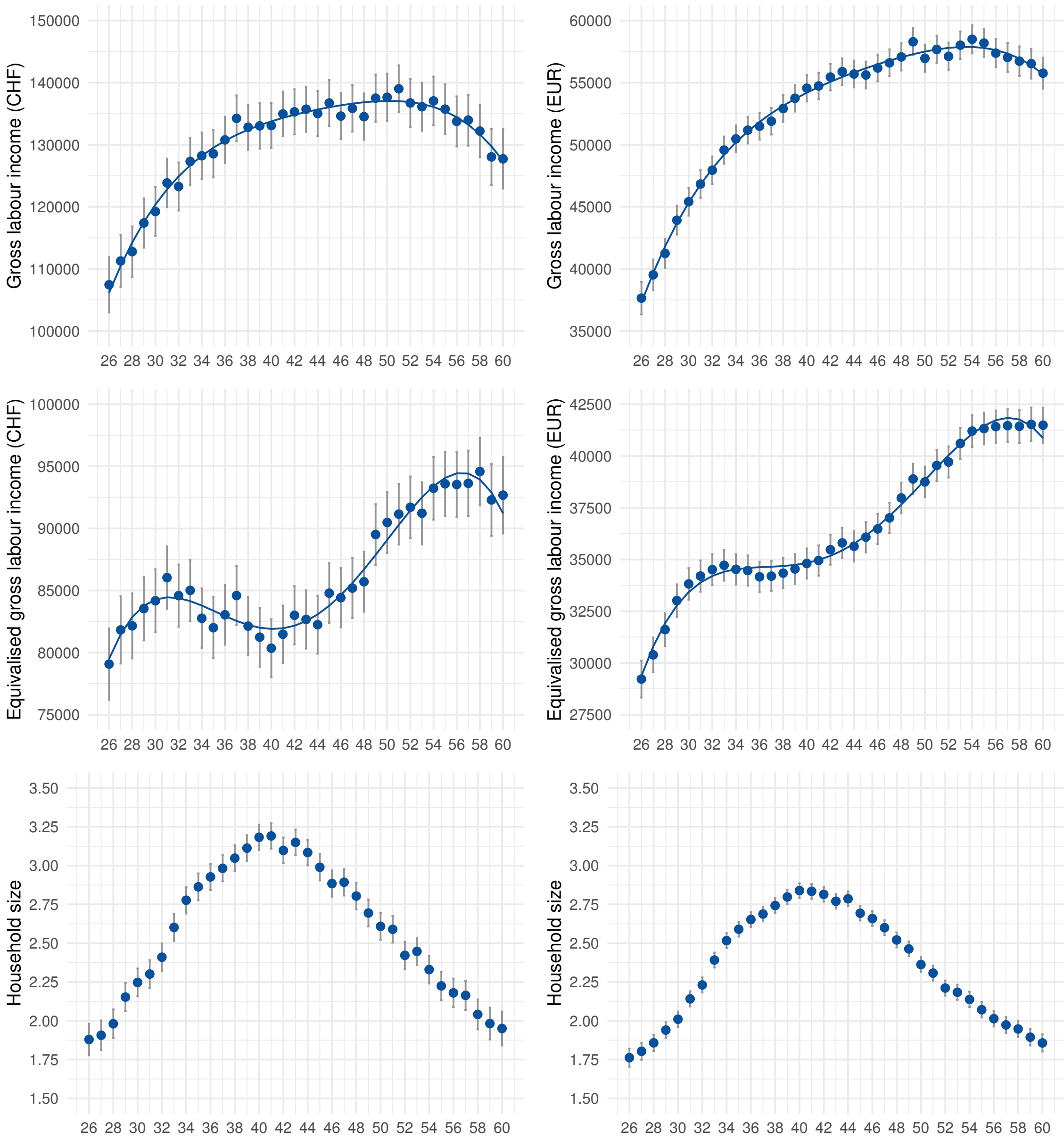
$$y_{i,j,t} = \alpha_j + \gamma_t + \epsilon_{i,j,t}\,, \tag{4.1}$$

where $y_{i,j,t}$ is income of household i of age j at time t, γ_t is a time effect and α_j is the age effect of interest. Thus, I assume the cohort effect to be zero. I estimate the age effect in two ways. First, I use a full set of age dummies to obtain the estimate for each age group. Second, I use a fourth-order age polynomial. This yields a smooth life cycle profile without discontinuities, which is an advantage when using the profiles in a quantitative model.

In Figure 4.1, I show the life cycle profiles for gross household labour income in Switzerland and Germany. In the top panel, I use income without applying an equivalence scale. In the middle panel, I apply an equivalence scale.[13] The dotted profile shows the estimates using age dummies with their respective 95% confidence sets, while the solid blue line shows the profile resulting from the estimated age polynomial. Two things are worth emphasising. First, the age polynomial follows the income profile quite accurately and remains within the confidence sets. Thus, using age polynomials in a quantitative model yields a good approximation.[14] Second, and more importantly, Figure 4.1 illustrates that using an equivalence scale has a significant impact on the shape of the life cycle profiles, which is due to the evolution of the household size over the life cycle. The bottom panel shows the average household size over the life cycle which follows an inverse U-shape. The increase in household members more than compensates the increase in household income, which yields a flat or even downward sloping income profile between age 30 and 40. In a life cycle model, one can either use equivalised income as model input directly or use non-equivalised income and account for changes in household size within the model. I thus provide inputs for researchers who use either method to account for changes in household size. Finally, it is worth noting that when estimating the life cycle

[13]See Section 2 for a description of the equivalence scale I use.

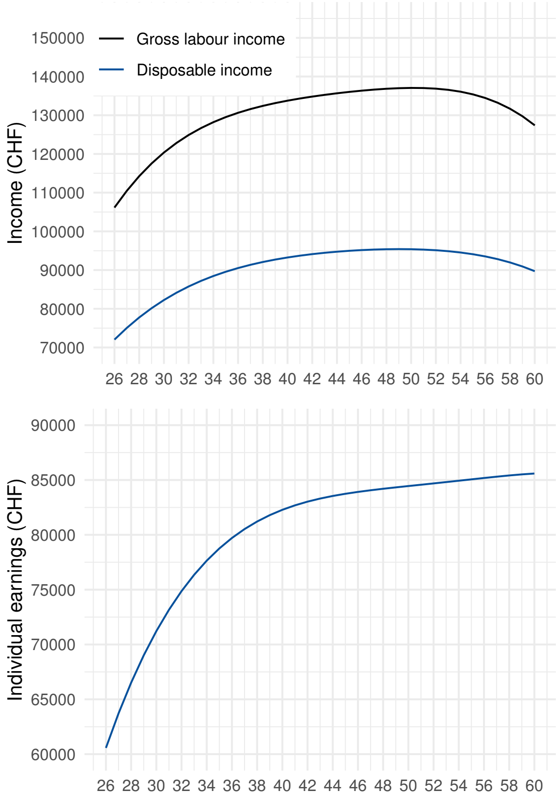
[14]Note however, that the polynomial might be less precise at the corners and when there are flat parts in the profile. This can be remedied by smoothing the profile before estimating the polynomial (e.g. with a moving average).



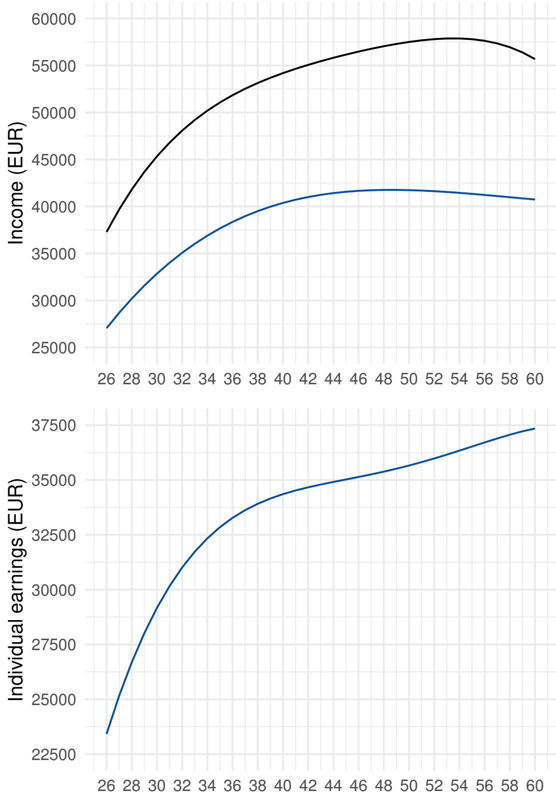
(a) Switzerland (b) Germany

Figure 4.1: Household income profiles with and without the use of an equivalence scale.

Notes: Income is in local currency. The base year is 2015.
In 2015, the Euro exchanged for approximately 1.00–1.10 CHF.



(a) Switzerland



(b) Germany

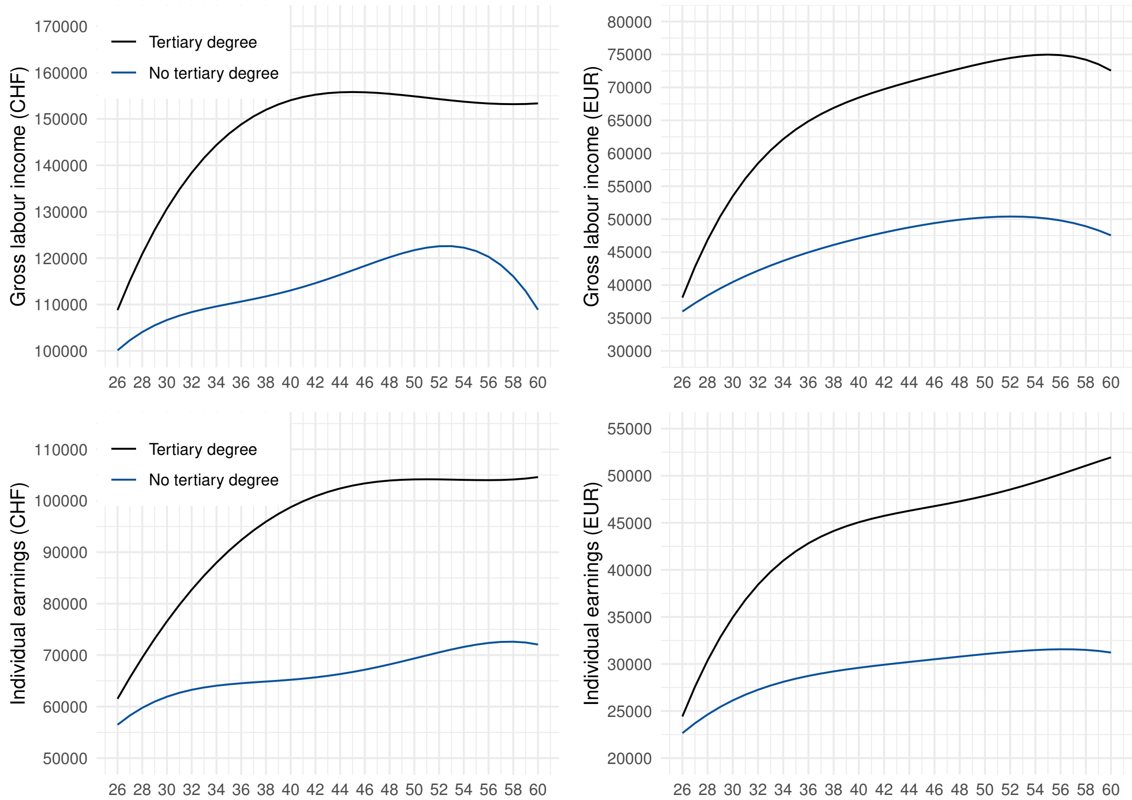
Figure 4.2: Income profiles by income definition.

Notes: Income is in local currency. The base year is 2015.
In 2015, the Euro exchanged for approximately 1.00–1.10 CHF.

profiles with a time effect, there is a separate intercept for each year and thus, one has to commit to a reference year, which I choose to be the year 2015.

Figures 4.2 and 4.3 show stylised facts for life cycle income profiles for Switzerland and Germany for the different income variables and across educational groups. In Figure 4.2, I show the age polynomials for gross household labour income, disposable income and individual earnings, without applying an equivalence scale.

When comparing individual income and household income, one can see that on the one hand, the life cycle profile for individual income is steeper than the profile for household income. On the other hand, household income is hump-shaped, whereas individual income is not. Therefore, the household composition and labour force participation of members within the household are strong driving forces of the household income profile.



(a) Switzerland (b) Germany

Figure 4.3: Income profiles by education.

Notes: Income is in local currency. The base year is 2015.
In 2015, the Euro exchanged for approximately 1.00–1.10 CHF.

This is further illustrated in Figure 4.3, where I plot income profiles by education. The black line shows the age polynomials for households and individuals with a tertiary degree and the blue line is the analogue for households and individuals without a tertiary degree.[15] From the profiles, we can see that at the beginning of the life cycle, a higher level of education does not necessarily translate into a higher level of income, but a much steeper income profile leading to a higher income dispersion between education groups over the life cycle. Furthermore, there is a significant difference in the shape of the household income profile for households without a tertiary degree between Switzerland and Germany. While in Germany the household income profile for this group is similar in shape to the profile of households with a tertiary degree, there is a significant

[15]Note that the Cross-National Equivalent File (CNEF) uses a rather broad definition of tertiary education as described in Appendix A. At the household level, I use the education level associated with the household head.

drop in income for the corresponding group in Switzerland. This difference, however, is not observed for individual income. As before, this illustrates that the labour force participation within the household plays a significant role for the evolution of household income over the life cycle.

5 Two stochastic income processes

In this section, I estimate two different stochastic income processes using income data from the SHP and the SOEP. I proceed in two steps. In the fist step, I decompose income into a deterministic component which varies over the life cycle and a stochastic component, controlling for time effects and other observables. In the second step, I estimate a stochastic income process from the stochastic component of income. In Section 5.1, I estimate the stochastic process as an AR(1) process and in Section 5.2, I estimate a process with a permanent and transitory component as in Blundell et al. (2008). I estimate these two processes for equivalised gross household labour income, equivalised disposable income, and individual earnings. Furthermore, I estimate the processes using the core sample and the extended sample. Using the core sample, I obtain estimates that abstract from the risk of a low income draw due to unemployment or other factors. Thus, using the extended sample provides insight into how the income process changes when including households and individuals with less than 520 annual working hours.

5.1 AR(1) income process

I estimate the income process in logs, and assume that the deterministic and stochastic income components are characterised by the following two equations:

$$\begin{aligned} \log y_{i,t,j} &= \alpha_j + \gamma_t + \mathbf{x}'\delta + \tilde{y}_{i,t} \\ \tilde{y}_{i,t} &= \rho \tilde{y}_{i,t-1} + \epsilon_{i,t}, \quad \text{with } \epsilon_{i,t} \sim \mathcal{N}(0, \sigma_\epsilon^2), \end{aligned} \tag{5.1}$$

where α_j is a full set of age dummies, γ_t controls for time effects, and $\mathbf{x}$ includes a vector of controls including a set of dummies for education and the area of residence. The stochastic income component $\tilde{y}_{i,t}$ has persistence given the assumption of an AR(1) process.

The persistence parameter ρ and the variance of innovations σ_ϵ^2 for the stochastic component of the income process are reported in Table 5.1. The estimated per-

Table 5.1: AR(1) income process

	Core sample		Extended sample	
	Switzerland	Germany	Switzerland	Germany
Gross household labour income				
Income autocorrelation ρ	0.779	0.795	0.774	0.788
	(0.772, 0.786)	(0.791, 0.798)	(0.767, 0.781)	(0.785, 0.791)
Variance of innovations σ_ϵ^2	0.070	0.096	0.081	0.132
Disposable income				
Income autocorrelation ρ	0.744	0.798	0.740	0.792
	(0.737, 0.752)	(0.795, 0.802)	(0.733, 0.748)	(0.789, 0.796)
Variance of innovations σ_ϵ^2	0.067	0.046	0.075	0.053
Individual Earnings				
Income autocorrelation ρ	0.800	0.815	0.819	0.797
	(0.795, 0.805)	(0.812, 0.818)	(0.814, 0.823)	(0.794, 0.800)
Variance of innovations σ_ϵ^2	0.095	0.122	0.135	0.206

Note: 95% confidence intervals in brackets.

sistence of the income process for household and disposable income is between 0.74 and 0.80 depending on the sample and income measure. The estimated variance is between 0.04 and 0.14. As one would expect, the estimates show two patterns. First, the variance of innovations σ_ϵ^2 is higher for the extended sample. This reflects the fact, that unemployment risk can usually not be fully insured away. Second, the variance of innovations σ_ϵ^2 is higher for gross household labour income than for disposable income and this effect is stronger in Germany than in Switzerland. This finding reflects the nature of the tax system, which provides some insurance through progressive taxation and which does so to a lesser extent in Switzerland than in Germany. I will return to this finding in Section 6.

Looking at the estimates for individual earnings, one can observe again, that the variance of innovations σ_ϵ^2 is higher for the extended sample. More importantly, the variance is much higher for individual earnings than for household income. This shows the effect of insurance within the household, translating a higher variance of individual earnings into a lower variance of household income.

5.2 Income process with permanent and transitory component

For this income process I follow Blundell et al. (2008) estimating the income process in logs, and assuming that the deterministic and stochastic components are characterised by the following three equations:

$$\begin{aligned}
\log y_{i,t,j} &= \alpha_j + \gamma_t + \mathbf{x}'\delta + \tilde{y}_{i,t} \\
\tilde{y}_{i,t} &= p_{i,t} + \eta_{i,t}, \qquad \text{with } \eta_{i,t} \sim \mathcal{N}(0, \sigma_\eta^2) \\
p_{i,t} &= p_{i,t-1} + \mu_{i,t}, \qquad \text{with } \mu_{i,t} \sim \mathcal{N}(0, \sigma_\mu^2)
\end{aligned} \tag{5.2}$$

As before, the first equation characterises the decomposition into the deterministic component, given by a full set of age dummies α_j, a time effect γ_t, a vector of controls $\mathbf{x}$, which includes a set of dummies for education and the area of residence, and the stochastic component $\tilde{y}_{i,t}$. The second equation shows the decomposition of the stochastic component $\tilde{y}_{i,t}$ into a transitory component $\eta_{i,t}$ and a permanent component $p_{i,t}$, which follows a random walk with innovation $\mu_{i,t}$. The transitory component $\eta_{i,t}$ and the innovation $\mu_{i,t}$ have variance σ_η^2 and σ_μ^2, which are assumed to be constant over time.[16]

To identify the variances σ_η^2 and σ_μ^2, write the first and second income difference as

$$\begin{aligned}
\Delta\tilde{y}_{i,t} &= \Delta\eta_{i,t} + \mu_{i,t} \\
\Delta^2\tilde{y}_{i,t} &= \Delta^2\eta_{i,t} + \mu_{i,t} + \mu_{i,t-1},
\end{aligned} \tag{5.3}$$

where I define $\Delta^2\tilde{y}_{i,t} = \tilde{y}_{i,t} - \tilde{y}_{i,t-2}$ and $\Delta^2\eta_{i,t}$ accordingly. Under the assumption that σ_η^2 and σ_μ^2 are constant over time we can write the variance of the first and second income difference as

$$\begin{aligned}
\mathrm{Var}[\Delta\tilde{y}_{i,t}] &= 2\sigma_\eta^2 + \sigma_\mu^2 \\
\mathrm{Var}[\Delta^2\tilde{y}_{i,t}] &= 2\sigma_\eta^2 + 2\sigma_\mu^2.
\end{aligned} \tag{5.4}$$

The restrictions in equation (5.4) thus identify the variances of the permanent and transitory shock through the first and second difference of $\tilde{y}_{i,t}$.

The estimates for the variance of the transitory shock σ_η^2 and the variance of the permanent shock σ_μ^2 are reported in Table 5.2. As before, one can see that the shocks are larger for individual earnings than for household income, reflecting

[16] I relax this assumption in Appendix B, where I estimate the income process with time varying variances $\sigma_{\eta,t}^2$ and $\sigma_{\mu,t}^2$.

Table 5.2: Income process with permanent and transitory component

	Core sample		Extended sample	
	Switzerland	Germany	Switzerland	Germany
Gross household labour income				
Variance of the transitory shock σ_η^2	0.023	0.026	0.027	0.035
Variance of the permanent shock σ_μ^2	0.027	0.044	0.031	0.061
Disposable income				
Variance of the transitory shock σ_η^2	0.025	0.014	0.029	0.017
Variance of the permanent shock σ_μ^2	0.022	0.018	0.024	0.020
Individual Earnings				
Variance of the transitory shock σ_η^2	0.026	0.035	0.038	0.059
Variance of the permanent shock σ_μ^2	0.034	0.043	0.050	0.076

household income insurance within the household. Furthermore, the variance of the permanent shock σ_μ^2 is larger for gross household labour income than for disposable income, showing that the tax system has an insurance component. Again, this effect is stronger in Germany. Furthermore, this is also true for the variance of the transitory shock σ_η^2 in Germany, while for Switzerland, the evidence is less clear, given that the estimates are rather close to each other. As with the AR(1) process, the estimates for the variances are smaller for the core sample than for the extended sample.

6 Tax progressivity

In this section, I estimate a non-linear tax function for Switzerland and Germany using income data from the SHP and the SOEP. A standard tax function used in the literature is the tax function with constant progressivity as described in Heathcote et al. (2017), given by

$$T(y_\tau) = y_\tau - \lambda y_\tau^{1-\tau}, \tag{6.1}$$

where y_τ is taxable income. This tax function is characterised by the parameter λ, which determines the tax level, and the parameter τ, which determines how progressive the tax system is. Furthermore, this tax function also captures transfers, as for a sufficiently low level of taxable income, the tax becomes negative. The threshold at which the tax is exactly zero, is given by

$$y_0 = \lambda^{1/\tau} . \tag{6.2}$$

The advantage of using this functional form is twofold. On the one hand, this tax function is very tractable when used in a model and might even allow for closed form solutions. On the other hand, it is straightforward to estimate the parameters λ and τ. Let y_d denote disposable income (i.e. the after-tax income). Then, the relationship between disposable income and taxable income is

$$y_\mathrm{d} = \lambda y_\tau^{1-\tau} , \tag{6.3}$$

which I estimate in logs:

$$\log y_\mathrm{d} = \log \lambda + (1-\tau) \log y_\tau . \tag{6.4}$$

I estimate the tax function using two different income definitions for taxable income, gross and net household income. When using gross household income, I estimate the degree of progressivity due to both taxes as well as social security contributions and transfers such as unemployment benefits. When using net household income, I estimate progressivity coming from taxes. Thus, we would expect a higher degree of progressivity when using gross household income.

The estimates for the tax function are reported in Table 6.1. Additionally, I report the resulting income threshold y_0 and the goodness of fit of the estimation. As in the previous section, I estimate the parameters using both the core sample and the extended sample.

Tax progressivity for the core sample when using gross household income is at $\tau = 0.092$ in Switzerland and $\tau = 0.287$ in Germany, reflecting the fact that taxes are more progressive in Germany. For comparison, Heathcote et al. (2017) report and estimate of $\tau = 0.181$ for the U.S. using a similar sample with only employed households. Therefore, the progressivity of the U.S. tax system is somewhere between the progressivity of the German and Swiss tax system.

Table 6.1: Tax progressivity

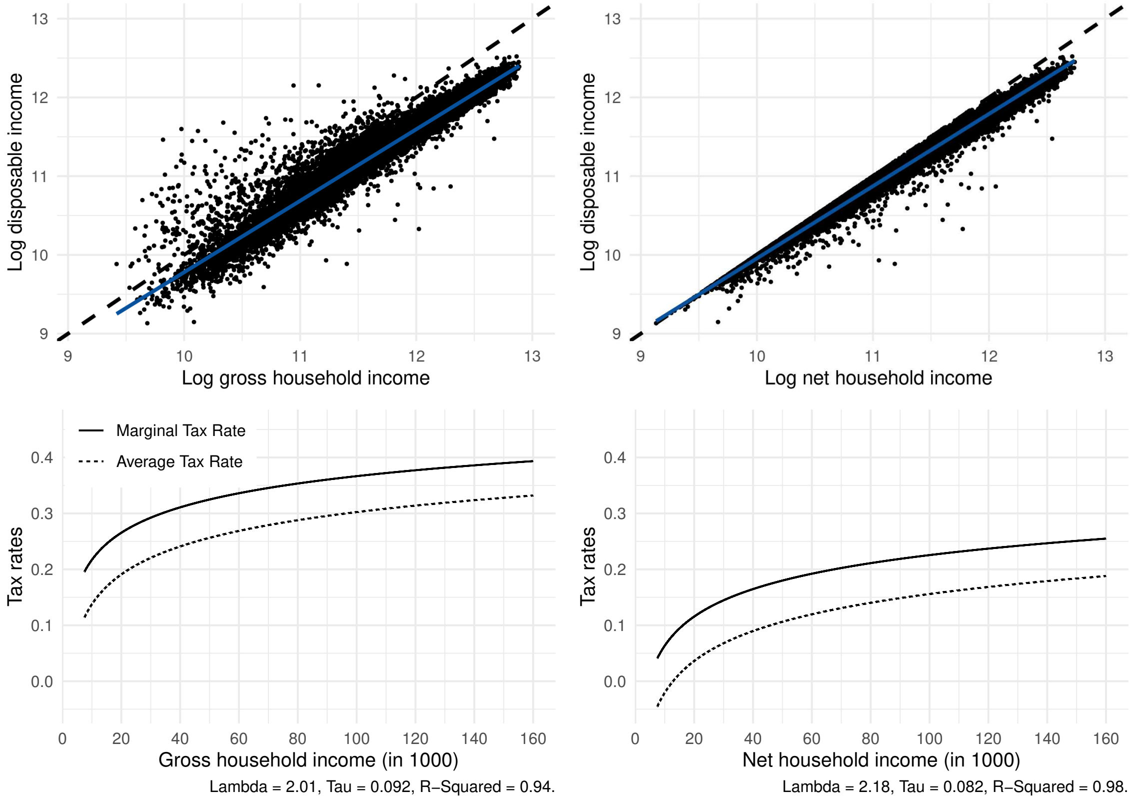
	Core sample		Extended sample	
	Switzerland	Germany	Switzerland	Germany
Gross household income – Disposable income				
Tax level parameter λ	2.01	16.37	2.21	41.13
	(1.96, 2.07)	(16.07, 16.67)	(2.15, 2.27)	(40.40, 41.86)
Tax progressivity parameter τ	0.092	0.287	0.100	0.372
	(0.090, 0.094)	(0.285, 0.289)	(0.098, 0.102)	(0.370, 0.373)
Income threshold y_0	2,006	17,045	2,770	21,940
R-Squared	0.94	0.82	0.93	0.77
Net household income – Disposable income				
Tax level parameter λ	2.18	3.25	2.18	3.29
	(2.15, 2.21)	(3.22, 3.27)	(2.16, 2.21)	(3.27, 3.32)
Tax progressivity parameter τ	0.082	0.128	0.082	0.129
	(0.081, 0.084)	(0.128, 0.129)	(0.081, 0.084)	(0.129, 0.130)
Income threshold y_0	12,791	9,697	12,983	9,977
R-Squared	0.98	0.97	0.98	0.97

Notes: Income in local currency. In 2015, the Euro exchanged for approximately 1.00–1.10 CHF. 95% confidence intervals in brackets.

Using net household income, the estimates for tax progressivity in Germany drop to more than half the value, while they only decrease slightly in Switzerland. Thus, in Switzerland, tax progressivity is mainly driven by taxes, whereas in Germany, both taxes as well as social security contributions and transfers play an important role for tax progressivity.

Figures 6.1 and 6.2 show the empirical fit of the estimation and the implied tax rates resulting from the estimated parameters using the core sample. The empirical fit is depicted by the solid blue line, while the dashed black line shows the 45 degree line, where taxable income and disposable income are equal and thus, taxes and transfers are zero. Therefore, the intersection of the two lines is exactly at y_0.

For Switzerland, the empirical fit is remarkable, as the goodness of fit is between 0.93 and 0.98. For comparison, the fit in Heathcote et al. (2017) is 0.91 and thus slightly lower. For Germany, the empirical fit is equally high when using net household income, where the goodness of fit is 0.97 for both samples.



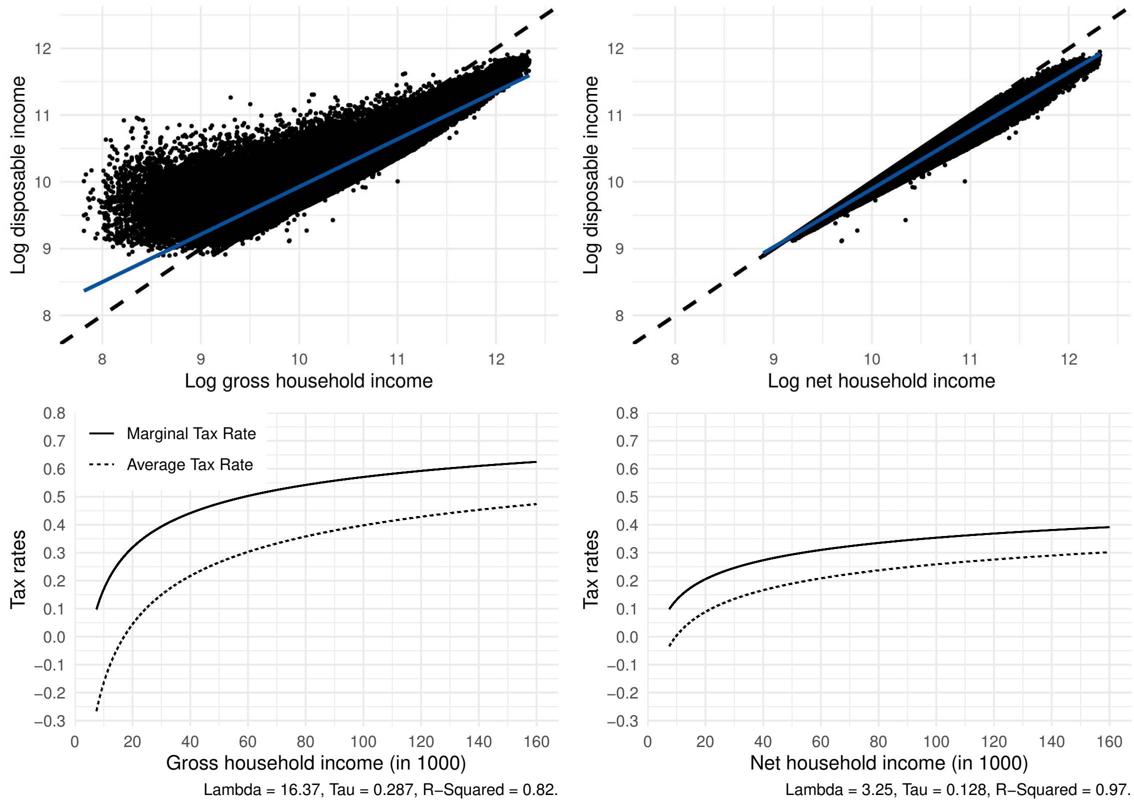
(a) Estimates using gross household income (b) Estimates using net household income

Figure 6.1: Estimated tax function and implied tax rates for Switzerland (core sample).

Notes: Income is equivalised and in local currency.
In 2015, the Euro exchanged for approximately 1.00–1.10 CHF.

When using gross household income, the empirical fit of 0.82 and 0.77 for the two samples is still high, but below the fit for the other estimates. Looking at the top left panel of Figure 6.2, one can see that the relationship between taxable income and disposable income tapers off for low incomes in Germany, implying that tax progressivity is not constant for lower income groups, which are taxed less. This is in line with the findings in Section 3.3, where I show that the German tax system is more redistributive than the Swiss tax system. This is not only reflected in the higher value for the tax progressivity τ, but also in the worse fit of the estimated tax function for lower incomes.

For economic models that focus on the role of taxes but do not explicitly model transfers and social security contribution, use results for redistribution through taxes only. For economic models that model redistribution through both



(a) Estimates using gross household income (b) Estimates using net household income

Figure 6.2: Estimated tax function and implied tax rates for Germany (core sample).

Notes: Income is equivalised and in local currency.
In 2015, the Euro exchanged for approximately 1.00–1.10 CHF.

channels explicitly, use results capturing redistribution through both channels but be aware that the simple redistribution function has a worse fit for low income households. Depending on the question at hand, more careful modeling of redistribution for these households is then required.

7 Conclusion

This paper provides estimates required for specifying quantitative life cycle models with income heterogeneity. Furthermore, I provide estimates for an approximated progressive income tax and transfer schedule. I have shown that income shocks are higher in Germany than in Switzerland for gross labour income and vice versa for disposable income, which is consistent with a more redistributive tax system in Germany. This is also reflected in my estimates for the tax and transfer schedule in Switzerland and Germany, which is more progressive in Germany. I have shown that tax progressivity in Switzerland arises mainly from direct taxes, whereas in Germany, social security contributions and public transfers also play an important role for tax progressivity. Furthermore, I have shown that the evolution of residual income inequality is consistent with the presence of time effects, while there is no evidence for sizeable cohort effects in Switzerland and Germany.

References

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *Quarterly Journal of Economics*, 109(3):659–684.

Blundell, R., Pistaferri, L., and Preston, I. (2008). Consumption inequality and partial insurance. *American Economic Review*, 98(5):1887–1921.

Deaton, A. and Paxson, C. (1994). Intertemporal choice and inequality. *Journal of Political Economy*, 102(3):437–467.

Fella, G., Frache, S., and Koeniger, W. (2020). Buffer-stock saving and households' response to income shocks. *International Economic Review*, 61(3):1359–1382.

Fernández-Villaverde, J. and Krueger, D. (2007). Consumption over the life cycle: facts from consumer expenditure survey data. *Review of Economics and Statistics*, 89(3):552–565.

Foellmi, R. and Martínez, I. Z. (2017). Volatile top income shares in Switzerland? Reassessing the evolution between 1981 and 2010. *Review of Economics and Statistics*, 99(5):793–809.

Fuchs-Schündeln, N., Krueger, D., and Sommer, M. (2010). Inequality trends for Germany in the last two decades: a tale of two countries. *Review of Economic Dynamics*, 13(1):103–132.

Gazareth, P. and Suter, C. (2016). Ungleichheit in der Schweiz – auch eine Frage der Daten. *Die Volkswirtschaft*, 2016(1–2):45–48.

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The German Socio-Economic Panel (SOEP). *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 239(2):345–360.

Guner, N., Kaygusuz, R., and Ventura, G. (2014). Income taxation of US households: facts and parametric estimates. *Review of Economic Dynamics*, 17(4):559–581.

Heathcote, J., Storesletten, K., and Violante, G. L. (2005). Two views of inequality over the life cycle. *Journal of the European Economic Association*, 3(2-3):765–775.

Heathcote, J., Storesletten, K., and Violante, G. L. (2017). Optimal tax progressivity: an analytical framework. *Quarterly Journal of Economics*, 132(4):1693–1754.

Huggett, M., Ventura, G., and Yaron, A. (2006). Human capital and earnings distribution dynamics. *Journal of Monetary Economics*, 53(2):265–290.

Huggett, M., Ventura, G., and Yaron, A. (2011). Sources of lifetime inequality. *American Economic Review*, 101(7):2923–54.

Juhn, C., Murphy, K. M., and Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, 101(3):410–442.

Krueger, D., Perri, F., Pistaferri, L., and Violante, G. L. (2010). Cross-sectional facts for macroeconomists. *Review of Economic Dynamics*, 13(1):1–14.

Quadrini, V. and Ríos-Rull, J.-V. (2015). Inequality in macroeconomics. In *Handbook of Income Distribution*, volume 2, pages 1229–1302. Elsevier.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Storesletten, K., Telmer, C. I., and Yaron, A. (2004). Consumption and risk sharing over the life cycle. *Journal of Monetary Economics*, 51(3):609–633.

Voorpostel, M., Tillmann, R., Lebert, F., Kuhn, U., Lipps, O., Ryser, V.-A., Antal, E., Monsch, G.-A., Dasoki, N., Klaas, H. S., and Boris, W. (2020). *Swiss Household Panel User Guide (1999–2018), Wave 20.* FORS.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

A Data description and sample selection

In this appendix, I discuss the construction of my samples from the Swiss Household Panel (SHP) and the German Socio-Economic Panel (SOEP). For the analysis of all the data sets, I use the infrastructure provided by the R Core Team (2020) and Wickham et al. (2019).

I use the cross-national equivalent files (CNEF) for the SHP and the SOEP, which offer standardised variables from their core samples and thus, I can apply the same data cleaning process to both data sets. For both countries, I construct a sample at the household level and a sample at the individual level. At the household level, I use the attributes of the household head. I use information on the age of the individual, the household size, education, annual working hours, the area of residence,[17] individual earnings, household labour income, household capital income, public transfers,[18] social security contributions, and taxes.

My data preparation process follows four steps, as shown in Table A.1. First, I remove all observations with missing values obtaining a cleaned sample. Second, I select all households, whose head is between 26 and 60 years of age. I do the same for all individuals in the individual data set. Third, I adjust all nominal variables with the consumer price index using the year 2015 as base year. I then construct all the income variables as described in Section 2. Afterwards, I only keep households with total annual working hours between 520 and 5200 and individuals with annual working hours between 520 and 2600. Furthermore, I remove households and individuals who receive social security pensions. This ensures there are no observations from early retirees. Finally, I only keep households and individuals with positive income according the constructed income variables and trim the top and bottom percent to remove outliers. This yields the core samples at the household and individual level. For the extended sample, I also keep the households and individuals with less than 520 annual working hours. The size of the sample after each cleaning step is summarised in Table A.1. Summary statistics are reported in Table A.2.

[17] The *Kanton* in Switzerland and the *Bundesland* in Germany.

[18] Public transfers include all transfers such as unemployment benefits, child benefits, social assistance or student assistance.

Table A.1: Number of observations by sample: SHP and SOEP.

	Household data		Individual data	
	Switzerland	Germany	Switzerland	Germany
Full sample	94,939	331,165	231,007	843,998
Cleaned sample	89,756	319,546	173,767	559,118
Sample age 26–60	57,362	213,414	87,282	365,412
Extended sample	44,766	168,252	61,892	238,216
Core sample	42,619	157,986	58,595	222,986

Definition of tertiary education. As mentioned in Section 4, the cross-national equivalent files (CNEF) use a rather broad definition of tertiary education. For Switzerland, the following education groups fall into this category:

- Vocational high school with master certificate, federal certificate. *(Berufsprüfung mit Meisterdiplom, Eidgenössischer Fachausweis)*
- Technical or vocational school. *(Techniker- oder Fachschule)*
- Vocational high school ETS, HTL. *(Höhere Fachschule, HTL)*
- University, ETH, PH, FH. *(Universität, ETH, PH, FH)*

For Germany the education groups are:

- School of health care. *(Schule des Gesundheitswesens)*
- Specialized college of higher education, post-secondary technical college. *(Fachhochschule)*
- Technical university usually requiring practical training as part of the studies. *(Technische Universität)*
- Civil service training. *(Beamtenausbildung)*

Table A.2: Summary statistics: SHP and SOEP (2015).

	Core sample		Extended sample	
	Switzerland	Germany	Switzerland	Germany
Gross household labour income				
Mean	131,191	54,388	129,053	51,855
Median	123,803	48,100	120,900	46,302
Mean (equivalised)	86,249	37,346	84,637	35,520
Median (equivalised)	80,360	33,654	78,808	32,263
Net household labour income				
Mean	105,745	47,950	104,029	46,129
Median	99,176	41,775	96,875	40,222
Mean (equivalised)	69,477	32,612	68,221	31,282
Median (equivalised)	65,091	28,698	63,466	27,625
Gross household income				
Mean	134,376	55,536	132,152	52,969
Median	126,481	49,027	124,179	46,800
Mean (equivalised)	88,449	38,134	86,774	36,288
Median (equivalised)	82,269	34,409	80,300	32,760
Net household income				
Mean	108,930	49,097	107,128	47,242
Median	101,089	42,559	99,720	40,745
Mean (equivalised)	71,677	33,400	70,358	32,049
Median (equivalised)	66,691	29,151	65,167	28,116
Disposable income				
Mean	91,044	39,695	89,683	38,412
Median	87,327	36,283	85,803	34,637
Mean (equivalised)	59,633	26,842	58,624	25,903
Median (equivalised)	56,283	24,535	55,383	23,660
Individual Earnings				
Mean	80,486	33,946	77,698	32,123
Median	77,612	30,800	75,600	29,518

Notes: Income in local currency. In 2015, the Euro exchanged for approximately 1.00–1.10 CHF.

B Alternative income process

In this section, I estimate the income process described in Section 5.2, which follows Blundell et al. (2008), but here, I allow the variances of the shocks to vary over time. Thus, I assume that the deterministic and stochastic components of the income process in logs are characterised by the following three equations:

$$\begin{aligned}
\log y_{i,t,j} &= \alpha_j + \gamma_t + \mathbf{x}'\delta + \tilde{y}_{i,t} \\
\tilde{y}_{i,t} &= p_{i,t} + \eta_{i,t}\,, && \text{with } \eta_{i,t} \sim \mathcal{N}(0, \sigma^2_{\eta,t}) \\
p_{i,t} &= p_{i,t-1} + \mu_{i,t}\,, && \text{with } \mu_{i,t} \sim \mathcal{N}(0, \sigma^2_{\mu,t})
\end{aligned} \tag{B.1}$$

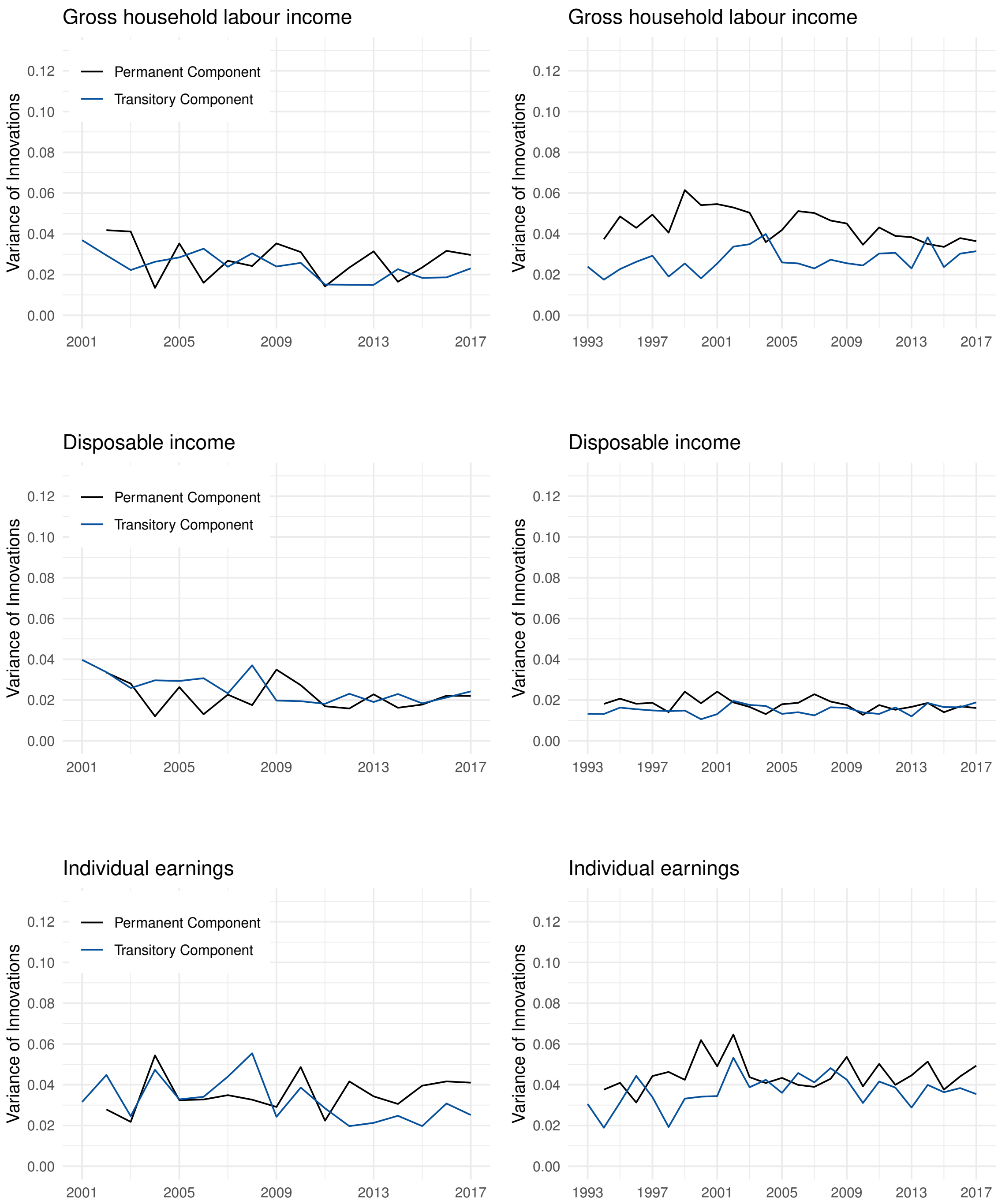
The notation is as before. The first equation characterises the decomposition into the deterministic component, given by a full set of age dummies α_j, a time effect γ_t, a vector of controls $\mathbf{x}$, which includes a set of dummies for education and the area of residence, and the stochastic component $\tilde{y}_{i,t}$. The second equation shows the decomposition of the stochastic component $\tilde{y}_{i,t}$ into a transitory component $\eta_{i,t}$ and a permanent component $p_{i,t}$, which follows a random walk with innovation $\mu_{i,t}$. In contrast to Section 5.2, the transitory component $\eta_{i,t}$ and the innovation $\mu_{i,t}$ have variance $\sigma^2_{\eta,t}$ and $\sigma^2_{\mu,t}$, which are allowed to vary over time. To identify the shocks I follow Fella et al. (2020). We can write the first income difference as

$$\Delta\tilde{y}_{i,t} = \Delta\eta_{i,t} + \mu_{i,t}\,. \tag{B.2}$$

Then, the variance of the first income difference and the first leading and lagging autocovariances are given by

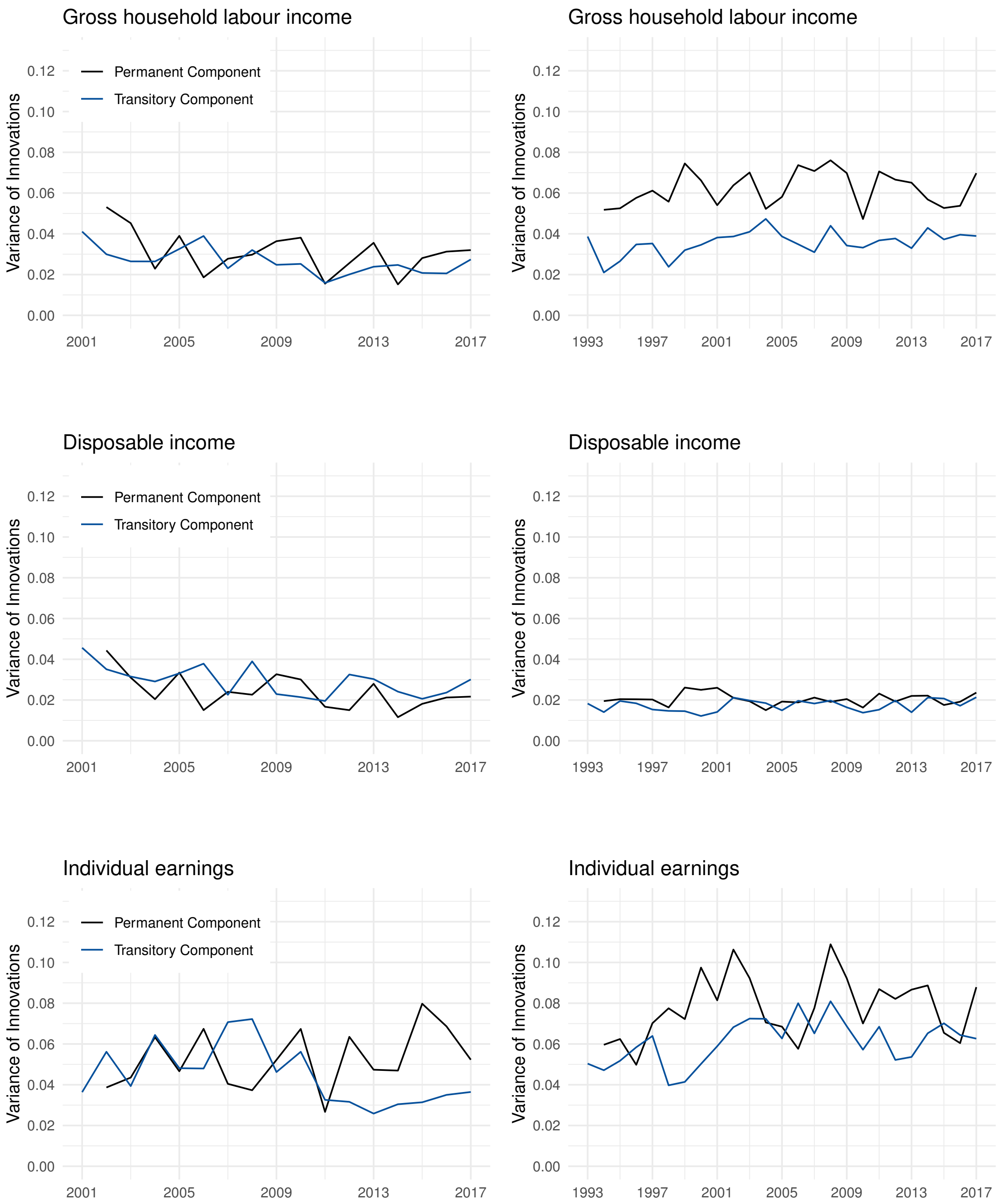
$$\begin{aligned}
\mathrm{Var}[\Delta\tilde{y}_{i,t}] &= \sigma^2_{\eta,t} + \sigma^2_{\eta,t-1} + \sigma^2_{\mu,t} \\
\mathrm{Cov}[\Delta\tilde{y}_{i,t}, \Delta\tilde{y}_{i,t+1}] &= -\sigma^2_{\eta,t} \\
\mathrm{Cov}[\Delta\tilde{y}_{i,t}, \Delta\tilde{y}_{i,t-1}] &= -\sigma^2_{\eta,t-1}\,.
\end{aligned} \tag{B.3}$$

Hence, the restrictions in equation (B.3) identify the variances of the permanent and transitory shock through the first difference and two autocovariance terms of $\tilde{y}_{i,t}$. The estimates for the variance of the transitory shock $\sigma^2_{\eta,t}$ and the variance of the permanent shock $\sigma^2_{\mu,t}$ are shown in Figure B.1 for the core sample and Figure B.2 for the extended sample. I estimate the shocks for gross household labour income, disposable income, and individual earnings as defined in Section 2.



(a) Switzerland (b) Germany

Figure B.1: Estimates for $\sigma^2_{\eta,t}$ and $\sigma^2_{\mu,t}$ (core sample).



(a) Switzerland

(b) Germany

Figure B.2: Estimates for $\sigma^2_{\eta,t}$ and $\sigma^2_{\mu,t}$ (extended sample).

II | Tax incentives, liquidity and mortgage amortisation

This chapter is joint work with Winfried Koeniger and Arnau Valladares Esteban.

Abstract

We analyse the effect of tax incentives on mortgage amortisation in Switzerland and Germany. Although these two countries are similar in many dimensions, homeowners in Switzerland hold significantly more mortgage debt than in Germany. We set up a dynamic portfolio-choice model of homeowners to investigate to which extent the tax deductibility of mortgage interest payments and pension savings contributes to explaining cross-country differences in household leverage. In our numerical illustration, differences in these tax incentives explain 50% of the observed differences in household leverage, and we show that the tax incentives are stronger if taxes are more progressive. Furthermore, we find that a tax reform abolishing the tax-deductibility of mortgage interest payments increases mortgage amortisation. However, this effect of the tax reform might be offset if the scope for tax deductions of voluntary pension savings is increased.

Keywords: Taxation, Mortgage amortisation, Pension savings, Portfolio choice.
JEL Codes: D15, E21, G11, H24.

1 Introduction

Household debt in developed countries mainly consists of mortgages to finance home purchases. Empirically, households in most developed countries amortise most of their mortgage until retirement and thus accumulate home equity as they approach retirement. Household leverage differs widely across developed countries (Bover et al., 2016), however, and this difference is particularly striking between Germany and Switzerland. While in Germany, more than half the homeowners own their house outright by the time they retire, only around ten percent do so in Switzerland (see Section 2). As a consequence, Switzerland is among the countries with the highest levels of household debt among developed countries with household debt per GDP of 134% in 2020, which is more than twice the value of 58% reported for Germany.[1]

We analyse how tax incentives for mortgage amortisation and retirement savings affect household leverage. Using a life cycle model with a portfolio choice between home equity and tax-incentivised pension savings, we analyse amortisation behaviour of homeowners in Germany and Switzerland. These countries are well suited for our analysis as they are very similar in many aspects but differ in the tax incentives for their portfolio choice.

On the one hand, household leverage in the two countries is constrained by relatively conservative rules which banks apply at mortgage origination (see Section 2). Thus, the risk of mortgage delinquency after a housing bust is similar in the two countries and much smaller, if not negligible for most households, than in the United States. This allows us to focus on tax incentives and optimal portfolio choice in our analysis. Moreover, the homeownership rate in the two countries is similar, which fits well with the focus on the portfolio choice of homeowners in our analysis.

On the other hand, tax incentives for portfolio choices differ between the two countries. In Switzerland, homeowners can deduct mortgage interest payments from their taxes, which is not the case in Germany. This makes mortgage amortisation relatively less attractive. At the same time, Swiss homeowners have to pay taxes on the so-called imputed rent, which is the theoretical rental income that could be earned by letting the house.

[1] See the global debt database of the IMF accessed in May 2022 at `https://www.imf.org/external/datamapper/datasets/GDD`.

Furthermore, the Swiss pension system allows for highly tax incentivised pension savings through the so-called second and third pillar. Although in Germany a similar pension scheme exists (the so-called *Riester-Rente*), the tax deductions offered are much less generous.

We find that tax policy matters quantitatively for household leverage. In a numerical illustration, we show that observed differences in tax incentives between Germany and Switzerland explain around 50% of the observed difference in median loan-to-value ratios at retirement. By exposing German homeowners to Swiss tax incentives, we further show how our results are affected by progressive taxes. Because the German tax system is more progressive, introducing tax incentives not to amortise would have an even stronger effect in Germany.

We use the model to illustrate the effects of a tax reform on amortisation behaviour in Switzerland. Abolishing the taxation of imputed rents and the tax deductibility of mortgage interest payments leads to a reduction of the median LTV ratio at retirement by 50% in our numerical illustration. However, if, at the same time, the tax deductions for voluntary pension savings are increased, the tax reform might be offset, because the after-tax returns of tax-incentivised pension savings dominate mortgage amortisation.

Our results inform the current policy debate about tax reforms that aim to change tax incentives and thus households' portfolios. For example, there is a continuing discussion in the U.S. on abandoning the tax deductibility of mortgage interest payments (Sommer and Sullivan, 2018), and the parliamentary initiative 17.400 in Switzerland proposes to abandon this tax deductibility together with the taxation of imputed rents.[2] Furthermore, the parliamentary initiative 20.494 in Switzerland proposes to more than double the amount of pension savings eligible for tax deductions. Our approach shows that these two proposals should not be discussed in isolation as there is an interaction between tax incentives on mortgage amortisation and pension savings.

Related Literature. This paper fits in the broad literature on household finance as surveyed by Campbell (2006) and Guiso and Sodini (2013) and the literature on housing in macroeconomics as surveyed by Piazzesi and Schneider (2016). We build on the literature that analyses how tax incentives shape households'

[2]Baumberger (1999) and Morger (2014) analyse the effects of a change in the taxation of imputed rents in the Swiss context, providing a qualitative discussion of the incentive effects, distributional effects and the effect on total tax revenues. In our dynamic model, such a reform also affects liquidity through intertemporal reallocation of resources.

portfolio choices and may affect households' welfare. See, for example, Poterba (2001), Amromin et al. (2007), Alan et al. (2010), Sommer and Sullivan (2018), Karlman et al. (2021) and Gruber et al. (2021). We also contribute to the emerging literature on how balance sheets of households affect the marginal propensity to consume (Kaplan and Violante, 2014; Cloyne et al., 2020) and portfolio choices (Chetty et al., 2017). Finally, our research relates to the literature on comparative household finance surveyed in Badarinza et al. (2016). Bringing together these strands of the literature allows us to show how tax incentives for mortgage amortisation and pension savings affect household debt and retirement wealth.

Outline. We proceed in the following steps. In Section 2, we document the differences in household leverage in Germany and Switzerland, and we describe the features of the respective institutional environment that are most relevant for our analysis. In Section 3, we present the dynamic portfolio choice model of homeowners. We discuss the calibration in Section 4, present the results of our quantitative analysis in Section 5, and conclude in Section 6.

2 Facts on mortgages, portfolios and tax incentives

Households in Germany and Switzerland differ in their portfolio choices. In this section, we present the facts and the institutional background that motivate why we focus on tax incentives in our analysis to explain these differences.

2.1 Household leverage and portfolios

Household leverage differs widely across developed countries (Bover et al., 2016). The difference is particularly striking between Germany and Switzerland. Household debt per GDP in 2020 has been 134% in Switzerland which is more than twice the value of 58% reported for Germany.[3] Given that most of the household debt is mortgage debt, it is remarkable that the very different mortgage debt levels are associated with similar homeownership rates in the two countries.

In Tables 2.1 and 2.2 we provide descriptive evidence on the leverage of homeowners and their portfolios based on household surveys. The evidence shows that the higher household debt in Switzerland compared to Germany is associated

[3] See the global debt database of the IMF accessed in May 2022 at `https://www.imf.org/external/datamapper/datasets/GDD`.

with less amortisation of mortgages by Swiss homeowners compared to German homeowners, as we now explain in further detail.

We present evidence based on three household surveys: the Household Finance and Consumption Survey (HFCS), which is available for Germany but not for Switzerland; the Survey of Health, Ageing and Retirement (SHARE), which is available for Germany and Switzerland but only for households with a head older than age 50; and the Household Budget Survey (HABE), which provides information on payment flows and thus allows us to infer the incidence of mortgagors in Switzerland also at younger ages. We provide further information on the surveys in Appendix A.

Because the main purpose of the HFCS is to provide accurate information on household wealth, the HFCS oversamples the wealthy because the wealth distribution is very skewed. The more accurate measurement of wealth in the upper tail of the wealth distribution increases the estimates for net worth compared to the sample in the SHARE. This is illustrated in column 1 of Table 2.2, which shows that the average net worth of German homeowners aged 56 to 65 is estimated to be much larger in the HCFS than in the SHARE, using the sampling weights so that the oversampling of the wealthy in the HFCS is taken into account. Columns 2 and 3 in Table 2.2 show that the portfolio shares of home equity for the main residence and other real estate, for which we have information for Germany both in the HFCS and the SHARE, are less different across the two surveys.[4] The statistics on the incidence of mortgages and the average LTV ratios, reported in columns 2 to 4 of Table 2.1 are also quite robust when such information exists for more than one survey.

Overall, the results reported in Table 2.1 show that the incidence and extent of leverage is higher in Switzerland than in Germany, and that the difference in the incidence of leverage is larger closer to retirement at ages 56 to 65. The different leverage is not associated with differences in the ownership of the first residence. The homeownership rate in Germany and Switzerland is similar at ages 36 to 65. It is 49% or 51%, depending on the survey data in the SOEP or HFCS for Germany; and 44% or 51%, respectively, in the survey data of the HABE and SHP for Switzerland.

Let us now comment in more detail on the differences in household leverage. In column 1 of Table 2.1, we see that most homeowners (95%) have a mortgage at

[4] The HFCS allows to compute the equity in other real estate for Germany, which we report in the notes of Table 2.2.

Table 2.1: Leverage of homeowners

	Fraction of homeowners with a mortgage		*Loan-to-value ratio of*	
			homeowners	*mortgagors*
	Ages		Ages	
	36 – 45	56 – 65	56 – 65	56 – 65
	(1)	(2)	(3)	(4)
Germany (*HFCS*)	0.71	0.41	0.12	0.33
Germany (*SHARE*)	-	0.35	0.09	0.27
Switzerland (*SHARE*)	-	0.90	0.38	0.42
Switzerland (*HABE*)	0.95	0.93	-	-

Sources: HFCS 2014 / SHARE 2015 / HABE 2015.
Notes: The SHARE samples households with a head above age 50. The HABE does not provide information on stocks such as mortgage amounts but only on payment flows. In the main text, we discuss further existing evidence on LTV ratios at mortgage origination for Germany and Switzerland.

ages 36 to 45 in Switzerland compared with 71% in Germany. Column 2 shows that most Swiss homeowners (90 – 93%) still have a mortgage at ages 56 to 65 whereas this is the case for only 35 – 41% of homeowners in Germany, indicating that Swiss households amortise less than German households.[5] This also shows at the intensive margin for the loan-to-value (LTV) ratios, as reported in columns 3 and 4 for households with heads aged 56 to 65 for which comparable data are available. In column 3, we report the LTV ratio of homeowners. This ratio is smaller than the LTV ratio of mortgagors in column 4 because some homeowners own their home outright and thus do not have a mortgage, as shown in column 2, because they have fully amortised their mortgage before retirement. In column 4, we report the LTV ratios for mortgagors, thus focusing on the intensive margin of mortgage debt.[6]

Comparable data on the size of leverage is scarce for young households in Germany and Switzerland because the SHARE interviews only households with

[5]These results are robust if we compute these statistics using data in the SHP and SOEP. 84% of homeowners in Switzerland in the SHP have a mortgage at ages 36 to 45 compared with 77% in Germany in the SOEP. At ages 56 to 65, 81% of homeowners in the SHP still have a mortgage relative to only 41% in the SOEP. We use the same survey year 2015 for comparison.

[6]The observed higher LTV ratios for older age groups in Switzerland compared to Germany cannot be explained by lower house price growth in Switzerland. Many households aged 56 to 65 in 2015 have purchased their homes in the 1990s. Given that house prices have increased *more* in Switzerland than in Germany in the 2000s (Schneider and Wagner, 2016), the mechanical effect of the stronger house price increase would imply *lower* LTV ratios for homeowners in Switzerland than in Germany.

Table 2.2: Net worth and portfolio shares of homeowners at retirement

		Portfolio shares of homeowners at ages 56 to 65		
	Net worth (1)	Home equity (main residence) / Net worth (2)	Other real estate / Net worth (3)	Voluntary pension savings / Net worth (4)
Germany *(HFCS)*	350,677	0.57	0.13†	0.08
Germany *(SHARE)*	265,568	0.64	0.09	-
Switzerland *(SHARE)*	821,321	0.52	0.10	0.17

Sources: HFCS 2014 / SHARE 2015.
Notes: Net worth is equivalised to account for differences in household size, and is denominated in units of local currency. In 2015, the euro exchanged aproximately for 1.0-1.1 CHF.
† *The HFCS provides information on the other real estate assets and the associated liabilities. Computing the ratio of other real estate equity to net worth for Germany, we obtain a ratio of* 0.11.

a head older than age 50, and Switzerland is not included in the HFCS. For Germany, data provided by the European Central Bank (2009), table 2, reveals that the typical LTV ratio for first-time buyers is 70%. Evidence reported in Basten and Koch (2015), table 3, shows that the LTV ratio of 74-75% at mortgage origination is very similar in Switzerland. The similar LTV ratios at mortgage origination in both countries, together with the lower LTV ratios at retirement in Germany compared to Switzerland, reported in Table 2.1, suggest that homeowners in Germany amortise their mortgage more until retirement than homeowners in Switzerland.

In Table 2.2 we provide descriptive evidence on portfolios of homeowners in Germany and Switzerland at, or close to, retirement. Column 1 of Table 2.2 shows that Swiss homeowners have more net worth than German homeowners. The higher LTV ratios of Swiss homeowners shown in Table 2.1 imply that they hold a smaller share of their wealth in home equity for the main residence (Table 2.2, column 2). The share of wealth in other real estate is more similar to German homeowners (Table 2.2, column 3). As we discuss below, the tax incentives differ for the amortisation of mortgages associated with the main residence. Swiss homeowners also have a larger share of their portfolio in voluntary pension savings (Table 2.2, column 4).

The descriptive evidence in Tables 2.1 and 2.2 raises the question about the causes of the different leverage, portfolio choices and implied portfolio diversification of German and Swiss homeowners. We now present the institutional background in both countries to motivate why we focus on tax incentives to answer this question, given that the supply side in the mortgage market in terms of lending criteria and mortgage types is similar across both countries. We also provide background on the pension system in both countries that is relevant for our analysis.

2.2 Mortgage markets

The description of the mortgage markets below draws on information by the European Central Bank (2009) for Germany and by the Swiss National Bank (2018) for Switzerland. The lending criteria applied in Germany and Switzerland at mortgage origination are similar. When buying a house, it is typically required that the future homeowner holds home equity of at least 20% of the home value. The remaining 80% can be financed through a mortgage. The implicit maximum LTV ratio is not legally binding but usually imposed by the bank because mortgages with higher LTV ratios trigger stricter capital requirements. The typical LTV ratios at mortgage origination are thus smaller than in the U.S. where the LTV ratio for the median purchased home is 89% on average.[7] The relatively conservative lending criteria imply that mortgage delinquency and negative home equity after house price corrections are less relevant for Germany and Switzerland than for the U.S.

In Switzerland, amortisation of the mortgage is required until the LTV ratio is $65-70\%$ (the ratio of the mortgage relative to the home evaluated at the purchase price) and this amortisation has to occur during 15 years or until the owner's retirement. In Germany banks typically offer mortgages with repayment plans that amortise the loan until retirement. Mortgage loans to households at older ages are a relatively recent phenomenon. To the best of our knowledge, there are no legal requirements that constrain amortisation of mortgages.

A further lending criterion applied in Switzerland at mortgage origination, which we do not explicitly model, is that the mortgage debt service payments cannot be larger than 30% of household labour earnings when applying an annual interest rate of 5%. Given that market mortgage interest rates have been much lower than 5% in the last decade, this criterion further illustrates the conservative lending practices in international comparison and motivates our focus on the effects of leverage for portfolio diversification.

[7] See the information on LTV ratios provided by CoreLogic at https://www.corelogic.com, accessed in September 2020.

In both countries the majority of mortgages are fixed-rate mortgages and refinancing or prepayment of mortgages before maturity is costly. Slicing of mortgages into more than two tranches to hedge the refinancing risk is becoming more common among newly originated mortgages but is less relevant for the stock of outstanding mortgages. We abstract from these financing details in our analysis because they would substantially complicate our analysis and they do not seem crucial to understand the differences in leverage between German and Swiss homeowners.

2.3 Tax incentives for portfolio choices

Tax incentives for mortgage amortisation and portfolio choices in Germany and Switzerland differ and we will gauge their quantitative importance for household portfolio choices in our analysis. Mortgage interest payments on the main residence can be tax deducted in Switzerland and the imputed rents for that main residence are taxed. This is not the case in Germany for the main residence but mortgage interest payments can be tax deducted for housing that generates taxable (rental) income, i.e. not for owner-occupied housing. Thus, mortgages on homes that are rented out can be attractive for tax reasons. For German households, home equity in real estate other than the main residence may thus be a substitute for tax-incentivised pension savings.[8]

The tax deductability of contributions to voluntary pension plans in Germany (within the so-called *Riester-Rente*) is capped at 2,100 euro, which is much less generous than what Swiss households can deduct for voluntary pension savings in the second or third pillar of the Swiss pension system.[9]

Homeowners can tax deduct interest payments on their main residence in Switzerland but, differently to the U.S. where mortgage interest payments also can be tax deducted, they have to pay taxes on the imputed rent which could be earned by renting their home. In theory, this tax treatment avoids distortions of the homeownership choice but in practice the calculation of the imputed rent is only based on a fraction of the housing value, which differs by canton and, by ruling of the federal court, cannot be lower than 60% (BGE 124 I 145 E.

[8] In the calibration for Germany we treat home equity in real estate other than the main residence as a substitute for tax-incentivised pension savings. We do not model it separately to economise on the number of state variables in our portfolio choice problem.

[9] In the calibration, we increase the deduction to 2,500 euro for Germany to capture subsidies to other illiquid saving instruments such as the *vermögenswirksame Leistungen*.

4d p. 156f.).[10] As long as the distortions are not too large, the approximate neutrality of the tax treatment for the home ownership choice justifies our focus on homeowners when we analyse the effect of the tax incentives on portfolio choices.

2.4 Pension systems

The German pension system mainly consists of a pay-as-you go component with net replacement rates between 45% and 60% depending on the income level, as documented in OECD (2007), p. 35. As mentioned previously additional voluntary pension contributions are only tax deductable up to 2,100 euro.

The Swiss pension system is composed of three pillars: a pay-as-you-go pension plan with more progressive replacement rates than in Germany, a funded pension plan and a private pension scheme, as documented in OECD (2007), p. 190, and Section 4. For employed households, it is compulsory to participate in the first two pillars. Additional voluntary contributions can be made to the second pillar, if there is a gap between the current pension wealth in that pillar and the contributions at the current income that would have accrued from age 24. Voluntary contributions to the third pillar can be tax deducted up to a maximum amount of CHF 6,826 (corresponding to approx. EUR 6,600 in May 2022) for households also contributing to the second pillar.

Pension savings are much less liquid than in the U.S. They remain illiquid until retirement but for the exception that the savings are withdrawn or used as collateral for the purchase of the first residence, or if the household becomes self employed. Once the savings are withdrawn from the third pillar before retirement, they are taxed separately from income at much lower tax rates of 3 – 6% depending on the region. The withdrawal can be spread over the last few years prior to retirement so that tax smoothing is possible.

[10] Furthermore, the valuations used by the tax authorities may not coincide with the current market valuation. More detailed information is in Eidgenössische Steuerverwaltung (2015). An example for the canton of Berne concerning the calculation of the imputed rent is provided by Steuerverwaltung des Kantons Bern (2016).

3 Model

3.1 Overview

We analyse the effect of the different tax incentives in Germany and Switzerland on the portfolio choice of homeowners. We focus on homeowners for at least two reasons. Firstly, there is no empirical evidence or theoretical prior that the different tax incentives across Germany and Switzerland should have strong effects on home ownership in our quantitative application. Empirically, Germany and Switzerland have very similar ownership rates, as discussed in subsection 2.1. Theoretically, it is not obvious that the different tax incentives in Switzerland, described in subsection 2.3, make home ownership more attractive. On the one hand, Swiss homeowners can deduct interest payments from taxable income, which makes home ownership more attractive. On the other hand, the imputed rent increases the taxable income of Swiss homeowners, which makes home ownership less attractive. In fact, this tax treatment is neutral for the home ownership choice but for the deviation of the housing value, which is applied in the computation of the imputed rent, from the current market value.[11]

Secondly, the focus on homeowners allows us to model their portfolio choices in more detail, by allowing wealth accumulation in two types of assets. We thus allocate the computational burden to capture those features which seem crucial for our analysis. See Campbell and Cocco (2003) or, more recently, Laibson et al. (2021) for a similar strategy.

In our dynamic model, homeowners with a mortgage have the choice between accumulating home equity and illiquid tax-incentivised pension assets. As we will illustrate in subsection 3.3, the possibility of accumulating a pension asset opens an important channel through which tax incentives affect mortgage amortisation. We model the degree of illiquidity by specifying at which point in the life cycle and at what cost pension assets can be withdrawn for consumption purposes.

We solve the model, which has two endogenous state variables for each of the assets in the portfolio and a stochastic state variable (labour income), by applying standard dynamic programming techniques. Stochastic labour income

[11]Gruber et al. (2021) provide empirical evidence that home ownership does not respond to changes of mortgage-interest tax deductions even if the imputed rent is smaller than the true rental income that would be generated by the home.

implies that agents value liquidity to smooth consumption after unexpected income changes.

We consider the problem of an owner of a house with size h at age j with mortgage debt $a < 0$. The homeowner has the portfolio choice between (i) reducing the debt by amortising the mortgage and (ii) accumulating pension wealth a_p by making a voluntary pension contribution.

The asset a denotes debt if $a < 0$ and assets if $a \geq 0$. We assume an interest spread $\zeta_\ell > 0$ for loans ($a < 0$). This allows us to model mortgage debt and liquid assets parsimoniously with one endogenous state variable. The benefit is thus that we can reduce the portfolio choice problem to two dimensions which reduces the comuptational burden substantially. The cost is that the position of (mortgage) debt is also influenced by liquidity considerations. Mortgage amortisation may be more attractive in our model than in reality because agents amortise also to have sufficient liquidity to self insure against shocks.

For agents close to the borrowing limit, amortisation for precautionary reasons makes it harder to detect the effect of tax incentives on amortisation behaviour. Precautionary savings will induce amortisation until the agent has enough liquidity buffer but then tax incentives shape the portfolio choices of agents. When interpreting the results, we thus have to keep in mind that liquidity considerations are important for the choice of the debt position $a < 0$ of consumers close to borrowing limit.

3.2 The portfolio choice problem

We solve the problem by backward induction. Although the focus of our analysis is on the effect of tax incentives on portfolio choices prior to retirement, we thus first characterise the continuation value at retirement.

The continuation value at retirement

The value at retirement depends on accumulated net worth and labour income prior to retirement. We provide a parsimonious characterisation in closed form for the continuation value assuming that the agent liquidates all assets at retirement and then consumes a fraction of the resources as implied by permanent income theory. Define net worth at retirement as

$$x \equiv h\phi + a_p + a \tag{3.1}$$

with ϕ denoting the relative house price at retirement in units of the consumption numeraire. Further assume that the agent receives income $\chi y_{J^{ret}}$ in each period during retirement through the public pension system, with $0 < \chi < 1$.

In the post-retirement period without uncertainty, a permanent income consumer with (approximately) infinite horizon, discounting the future at the interest rate ι, consumes the taxable income flow

$$y_{\tau}^{ret}(x, y(J^{ret})) \equiv \iota x + \chi y(J^{ret}), \tag{3.2}$$

of which taxes have to be deducted.[12] The value function at retirement is thus

$$V_{J^{ret}}(x, y(J^{ret})) = \frac{1}{1-\beta} u\left(y_{\tau}^{ret} - T(y_{\tau}^{ret})\right). \tag{3.3}$$

The optimisation problem prior to retirement

In the recursive problem prior to retirement, we have to distinguish the case $a' \geq 0$ and $a' < 0$ because of (i) the different tax treatment of interest paid on debt and on assets, and (ii) the interest spread which implies different prices for debt and assets. The recursive problem of the homeowner is:

[12]A finite horizon T at retirement could be accommodated with an additional factor $1 - \beta^{T+1}$ multiplying period utility and an additional factor $1/(1-(1+\iota)^{-(T+1)})$ multiplying net worth x. For our analysis the parsimonious version in the main text has enough degrees of freedom to capture a continuation value at retirement that depends on accumulated wealth.

$$V_j(a,a_p;y,h) = \max_{c,a',a'_p} u(c) + \beta \mathbb{E}_{y'} V_{j+1}(a',a'_p;y',h') \tag{3.4}$$

$$\text{s.t} \quad c^{a'\geq 0} = y - T(y_\tau^{a'\geq 0}) - (q_p a'_p - a_p) - (q_a a' - a) \tag{3.5}$$

$$c^{a'<0} = y - T(y_\tau^{a'<0}) - (q_p a'_p - a_p) - (q_\ell a' - a) \tag{3.6}$$

$$y_\tau^{a'\geq 0} = y + (\iota + \zeta_h)\omega h_0 \phi_0 - (q_p a'_p - a_p) + \iota q_a a' \tag{3.7}$$

$$y_\tau^{a'<0} = y + (\iota + \zeta_h)\omega h_0 \phi_0 - (q_p a'_p - a_p) + (\iota + \zeta_\ell) q_\ell a' \tag{3.8}$$

$$q_a = 1/(1+\iota); q_\ell = 1/(1+\iota+\zeta_\ell); q_p = 1/(1+\iota+\zeta_p) \tag{3.9}$$

$$q_\ell a' \geq -\underline{\mu} h \phi, \tag{3.10}$$

$$(\bar{p} + a_p)/q_p \geq a'_p \geq a_p/q_p \tag{3.11}$$

$$y(j) = \exp\left(\psi_0 + \sum_{i=0}^{4} \psi_i \cdot j^i + \hat{y}\right) \tag{3.12}$$

$$\hat{y}' = \rho \hat{y} + \epsilon, \text{ where } \epsilon \sim N(0, \sigma_\epsilon) \tag{3.13}$$

The problem has two endogenous state variables (a, a_p) and the stochastic state variable y. The portfolio choice also conditions on the house size h, which we keep the same for all homeowners for simplicity. Given that we analyse the portfolio choice conditional on home ownership, the value of h is thus exogenous.

The amount of debt held by the homeowner is limited by the collateral constraint (3.10). The term $\underline{\mu} h \phi$ denotes the maximum loan that households can hold.[13]

The stochastic realisation of the idiosyncratic income enters the problem as a stochastic state variable given that income shocks are persistent as shown in (3.13), with a persistence parameter ρ for which $0 \leq \rho \leq 1$. Equation (3.12) shows the common decomposition of income y into a stochastic component $\hat{y}$ and a deterministic hump-shaped life cycle profile, both of which we estimate using the household panel data from the SOEP for Germany and SHP for

[13] In the model, agents have some flexibility when to amortise, as may be achieved with indirect amortisation options available in mortgage markets. This type of amortisation allows households to accumulate pension assets instead of amortising each period. The accumulated pension assets are pledged as collateral to the bank and are used to amortise the loan amount at the point in time at which this is required, instead of amortising through payments each period over the life of the loan. This gives households more flexibility in their portfolio choices. The model accounts for this by not imposing a minimum amortisation constraint for payments each period.

Switzerland.[14] In equations (3.7) and (3.8) this income is converted into taxable income, taking into account how the portfolio positions of households may affect taxable income because of the imputed rent $(\iota+\zeta_h)\omega h_0\phi_0$ with housing evaluated at specified price ϕ_0 in regulation, interest income on the liquid assets $\iota q_a a'$ if $a' \geq 0$, deductions of mortgage interest payments $(\iota + q_\ell)q_\ell a'$ if $a' < 0$, and deductions of pension contributions $q_p a'_p - a_p$, with prices of the assets defined in (3.9) as for a zero-coupon bond given our timing assumption which we discuss next. In our quantitative application, the taxable income in equations (3.7) and (3.8) simplify for Germany because imputed rents are not taxed and mortgage interest payments of the main residence cannot be tax deducted, as we discussed in subsection 2.3.

As usual in discrete time, we have to make a timing assumption about the accrual of the interest rate. We choose the timing so that households can optimise their tax burden by choosing their portfolio after the income draw and reap the immediate gain of the tax optimisation in the current period before the next shock occurs. We assume that choices are made at the beginning of each period so that interest on the asset positions accrues until the end of the period. The constraints in the recursive problem reflect this timing assumption. Given the timing, the interest payment on debt $\iota+\zeta_\ell$ for taxable income in (3.8), for example, is multiplied by the value of the chosen debt position $q_\ell a'$.

The constraints in (3.11) make explicit that tax deductable pension contributions cannot be negative and are capped at $\bar{\mathrm{p}}$.

We allow for interest spreads in problem (3.4). The spread between the rate for liquid assets ι and the rate on pension wealth is $\zeta_p > 0$. The spread between the rate on liquid assets and mortgages is $\zeta_\ell > 0$. For the imputed rent in (3.7) and (3.8) we allow for a spread $\zeta_h > 0$. The parameter ω, with $1 \geq \omega > 0$, captures the fraction of the home value that is specified in the tax law to impute the rent.

Pension assets are illiquid before retirement. They cannot be consumed until retirement, directly or indirectly by collateralising them. Home equity is liquid instead. In a stylised way, this captures that housing and the associated mortgage liability is more liquid than pension assets in Germany and Switzerland. Empirically, some households tap into home equity in Germany and Switzerland although this is less common than in the U.S. (European Central Bank, 2009; Basten and Koch, 2015).

[14] We provide further information on the data in Appendix A.1, and on the estimation of the income process in Appendix A.2.

Optimal choices

We now characterise the optimal choices at interior optima and provide a simple example to show how tax incentives affect the portfolio choice of households. The first-order conditions for the choice of the portfolio (a', a_p') at interior optima are

$$a' : \quad \frac{\partial u(c)}{\partial c}\left(-q_k - \frac{\partial y_\tau}{\partial a'}\frac{\partial T(y_\tau)}{\partial y_\tau}\right) + \beta\frac{\partial \mathbb{E}_j V_{j+1}(a', a_p'; y', h')}{\partial a'} = 0 \ , \tag{3.14}$$

where $k = \ell$ if $a' < 0$ and $k = a$ if $a' \geq 0$, and

$$a_p' : \quad \frac{\partial u(c)}{\partial c}\left(-q_p - \frac{\partial y_\tau}{\partial a_p'}\frac{\partial T(y_\tau)}{\partial y_\tau}\right) + \beta\frac{\partial \mathbb{E}_j V_{j+1}(a', a_p'; y', h')}{\partial a_p'} = 0 \ , \tag{3.15}$$

with

$$\frac{\partial y_\tau^{a'\geq 0}}{\partial a'} = q_a \iota \ , \ \frac{\partial y_\tau^{a'<0}}{\partial a'} = q_\ell(\iota + \zeta_\ell) \ , \ \frac{\partial y_\tau}{\partial a_p'} = -q_p \ . \tag{3.16}$$

The envelope conditions at interior optima are

$$\frac{\partial V_j(a, a_p, h, y)}{\partial a} = \frac{\partial u(c)}{\partial c} \ , \tag{3.17}$$

and

$$\frac{\partial V_j(a, a_p, h, y)}{\partial a_p} = \frac{\partial u(c)}{\partial c}\left(1 - \frac{\partial T(y_\tau)}{\partial y_\tau}\right) \ , \tag{3.18}$$

with

$$\frac{\partial y_\tau}{\partial a} = 0 \ , \ \frac{\partial y_\tau}{\partial a_p} = 1 \ . \tag{3.19}$$

3.3 Tax incentives and portfolio choice: a simple example

Before we numerically solve and calibrate the life cycle model, it is instructive to consider a simple example to illustrate how tax incentives shape portfolio choices before retirement.

We illustrate key differences in the returns of the portfolio items by considering agents with linear utility who thus only care about the (expected) returns of the portfolio items. For simplicity, we abstract from uncertainty in this example.

Let us first consider the case without tax incentives in which portfolio choices are not distorted by taxes. The plausible assumption that the interest spread is higher for mortgage debt than for pension assets, i.e. $\zeta_\ell > \zeta_p > 0$, then implies that agents prefer to accumulate home equity by amortising the mortgage because $q_\ell < q_p < q_a$. Home equity has the highest return and thus the lowest price.

Let us now illustrate how tax incentives may change this portfolio choice. Agents prefer to accumulate pension assets rather than to amortise their mortgage if the after-tax prices for the assets, implied by the first-order conditions (3.14) and (3.15) given the assumption of linear utility and no uncertainty, are such that

$$q_p \frac{1 - \frac{\partial T(y_\tau)}{\partial y_\tau}}{1 - \frac{\partial T(y'_\tau)}{\partial y'_\tau}} < q_\ell \left(1 + (\iota + \zeta_\ell)\frac{\partial T(y_\tau)}{\partial y_\tau}\right) < q_a \left(1 + \iota \frac{\partial T(y_\tau)}{\partial y_\tau}\right) . \tag{3.20}$$

The two stylised cases without and with tax incentives illustrate how the different portfolio choices may occur that we documented for Germany and Switzerland in Section 2. We now provide further insights into the marginal tax rates at which agents prefer to accumulate pension assets rather than home equity. For this purpose, it is useful to consider proportional taxes with a tax rate τ, for which the inequalities in (3.20) simplify to

$$q_p < q_\ell (1 + (\iota + \zeta_\ell)\tau) < q_a (1 + \iota\tau) . \tag{3.21}$$

It then follows that agents prefer to accumulate pension assets rather than home equity if

$$\tau > \frac{\zeta_\ell - \zeta_p}{\left(1 + \iota + \zeta_p\right)(\iota + \zeta_\ell)} \equiv \tau^*_{\ell,a_p} \,. \tag{3.22}$$

We further comment on how the critical tax rate depends on the model parameters below.

Without uncertainty and linear utility, accumulating asset a once the debt is paid back is dominated by accumulating the pension asset because it earns a higher return, and consumption is a perfect substitute across time for generating utility (but for discounting). Agents then do not value higher liquidity because they do not have motives for consumption smoothing or precautionary saving. In this case, the portfolio choice thus reduces to a choice between home equity if $a < 0$ and the pension asset, depending on the tax incentives, and the pension asset dominates if the agent wants to accumulate further wealth once the debt has been paid off.

The previous discussion highlights the important role of tax-incentivised pension assets in the model. Pension assets change the portfolio choice in a crucial way for the model predictions about leverage. If we abstracted from pension assets, agents facing tax incentives as in Switzerland would prefer to accumulate home equity rather than to accumulate a risk-free asset for any tax rate below 100%. The model would then predict similar levels of leverage in Switzerland as in Germany, which is at odds with the stylised facts presented in Section 2.

To summarise, in the simple example with pension assets and tax incentives, agents amortise their mortgage and thus accumulate home equity if they face a tax rate $\tau < \tau^*_{\ell,a_p}$. Above that tax rate they accumulate pension assets. Only if this accumulation is capped and the mortgage is fully amortised, they accumulate risk-free assets.

The following numerical example shows at which marginal tax rates agents prefer accumulation of pension assets to home equity, in an economic environment with tax incentives as in Switzerland. Consider the plausible values for the spreads $\zeta_\ell = 0.015$, $\zeta_p = 0.005$, and the interest rate $\iota = 0.02$. The critical value in (3.22) is then $\tau^*_{\ell,a_p} = 0.28$. In this case, agents prefer to accumulate pension assets rather than home equity if they face a marginal tax rate larger than 0.28. If the spread for the mortgage interest rate decreases to $\zeta_\ell = 0.01$ so that the difference between the spreads $\zeta_\ell - \zeta_p$ is smaller, the critical value decreases to $\tau^*_{\ell,a_p} = 0.16$. Furthermore, (3.22) shows that amortisation becomes relatively more attractive at a lower interest rate ι. For example, if the interest rate decreases to $\iota = 0.01$, keeping the other parameters at their values $\zeta_\ell = 0.015, \zeta_p = 0.005$, the tax threshold increases to $\tau^*_{\ell,a_p} = 0.39$.

If taxation is progressive, then the factor

$$\phi_\tau \equiv \frac{1 - \frac{\partial T(y_\tau)}{\partial y_\tau}}{1 - \frac{\partial T(y'_\tau)}{\partial y'_\tau}} \tag{3.23}$$

in (3.20) makes accumulation in pension assets more attractive if $\phi_\tau < 1$, i.e. if the marginal change of the tax burden after an increase of taxable income in the current period is higher than in the future period. Hence, agents with currently high income who expect lower income in the future have an additional motive to use accumulation of pension assets for tax smoothing purposes.

Further effects become relevant at corners, i.e. when the amortisation choice is constrained or when the pension contributions of the agents reach the cap $\bar{p}$. Denoting with η_u the multiplier of the constraint in (3.11) that pension contributions are capped by $\bar{p}$, the optimality conditions imply that accumulation of the tax-incentivised pension asset today becomes more attractive if $\eta'_u > 0$, i.e. if the cap for pension contributions is binding in the next period. Intuitively, limiting the extent to which agents can take advantage of tax deductions for their pension contributions tomorrow, makes it more attractive to exploit the tax deductibility for pension contributions today.

The simple example in this subsection is based on the strong assumptions of linear utility and no uncertainty. With strictly concave utility and shocks to labour income, the liquid asset a plays an important role for agents to smooth consumption and to self insure. Therefore, we simulate the full model including concave utility, income shocks and progressive taxation. We proceed by presenting the model inputs used for the simulation.

4 Parameters and estimated model inputs

In this section, we describe the parametrisation of the model. Whenever we have values in local currency, we convert them into model units. We do so by dividing

Table 4.1: Parameter values

Common parameters		
Preferences		
Discount factor: β	0.986	
Risk aversion: σ	2	
Interest rates		
Liquid asset: ι	0.02	
Spread for mortgage: ζ_ℓ	0.01	
Spread for imputing rent: ζ_h	0.01	
Spread for pension asset: ζ_p	0.005	
Housing		
House size: h	5	
House price: ϕ	1	
Borrowing opportunities		
Borrowing constraint: $\underline{\mu}$	1.0	
Retirement		
Replacement rate: χ	0.6	
Country-specific parameters		
	Germany	Switzerland
Cap for tax-deductible pension contributions per year		
$\bar{\mathrm{p}}$	0.05	0.1
Taxes		
Income tax $T^y(y) = y - \lambda y^{1-t_y}$	$t_y = 0.174$, $\lambda = 5.020$	$t_y = 0.113$, $\lambda = 2.958$
Fraction used for imputing rent: ω	0	0.8
House size used for imputing rent: h_0	0	5
House price used for imputing rent: ϕ_0	0	1
Mortgage interest payments	Not tax-deductible	Tax-deductible
AR(1)-process for income: $\ln(y') = \rho \ln(y) + \epsilon$, $\ln y' \sim \mathcal{N}\left(-\frac{\sigma_\epsilon^2}{2(1-\rho^2)}, \frac{\sigma_\epsilon^2}{1-\rho^2}\right)$		
Income autocorrelation: ρ	0.806	0.771
Variance of innovations: σ_ϵ^2	0.078	0.086
Income age polynomial (4th order):		
	See graphical illustration in Appendix A.2	

the values from the data by the mean household income of homeowners prior to retirement obtained from the SHP and SOEP, which is around CHF 74,000 in Switzerland and EUR 37,000 in Germany in 2015.[15] Thus, model units can be interpreted as a fraction of average household income in the economy.

As mentioned in Section 2.1, data on young homeowners is scarce in Switzerland because the SHARE interviews only households with a head older than age 50, and Switzerland is not included in the HFCS. Thus, we use the HFCS to determine the initial distribution for Germany and then adjust this information for Switzerland, as detailed now. For the initial distribution of the pension asset a_p, we choose the empirical distribution in the data. For the initial distribution of the liquid asset a, we proceed in two steps. First, we assume that homeowners start with an average initial LTV of 80%. We set the house size to $h = 5$, as an illustration, and normalise its price to $\phi = 1$ for all agents in the economy. Thus, agents start with an average mortgage of 4 in model units. Second, we compute the median level of other assets held by young homeowners in the HFCS, which corresponds to approximately 1 model unit. Consolidating the mortgage and other assets to map them into asset a in our model, we obtain the average initial holding of asset a of -3, corresponding to 60% of the house value. For the initial distribution of asset a, we then draw from a uniform distribution between 20% and 100% of the house value, thus ensuring consistency with the average mentioned above.

We use the discount factor β for plausible spreads ζ_ℓ and ζ_p, as discussed further below, to match the median LTV and household net worth of Swiss homeowners prior to retirement. The median LTV ratio observed in the SHARE is roughly 40% and household net worth is approximately CHF 600,000.

Table 4.1 summarises the parameter values for Germany and Switzerland. The top panel of the table displays the parameters that are common across the two countries, and the bottom panel shows the parameters that are country specific. For the common preference parameters we choose what we consider to be uncontroversial values in the literature. The risk-free rate of 2% corresponds to the (approximately) risk-free return on long-term government bonds in Germany and Switzerland, net of inflation in the pre-crisis period in the 2000s. Compared to that rate, we set a spread of 1 pp for mortgages and the rates used for imputing rents, and we set the spread for pension assets to 0.5 pp.

[15] See Appendix A.1 for details.

For the purpose of the numerical illustration, we fix the house value ($h = 5$, $\phi = 1$) and allow for a borrowing limit $\underline{\mu}$ that reflects that homeowners might also borrow using instruments other than the mortgage. For the retirement period, we assume a replacement rate of $\chi = 0.6$.

Important for our quantitative analysis are the different tax incentives in Germany and Switzerland, reported in the bottom part of Table 4.1. These incentives are captured by the estimated income-tax functions, the different caps on tax deductible pension contributions, the different tax treatment of imputed rents for owned housing, and the differences in the tax-deductibility of mortgage interest payments. We set $\bar{p} = 0.1$ for Switzerland, which corresponds to approximately CHF 7,400. This is higher than what is deductible in the third pillar, but takes into account that tax-incentivised contributions to the second pillar are also possible.[16] For Germany, we set it to $\bar{p} = 0.05$. The estimates of the income-tax functions imply that these taxes are higher and more progressive in Germany than in Switzerland. Details on the estimations of the tax functions are provided in Appendix A.3. We refer to Section 2 for further background on the different tax incentives.

At the bottom of Table 4.1, we provide information on the country-specific age profiles of labour income and stochastic labour income processes. Further details on the estimation of these components are provided in Appendix A.2.

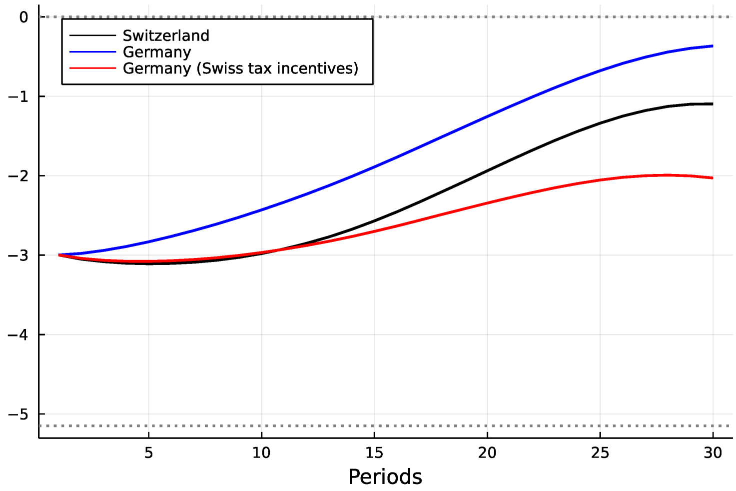
5 Results

We perform two counterfactual exercises. Section 5.1 compares the amortisation behaviour of homeowners in Switzerland and Germany. We first simulate both economies and then introduce Swiss tax incentives to German homeowners. In Section 5.2, we perform two experiments for Switzerland. First, we abolish the taxation of imputed rents and the tax deductions for mortgage interest payments. In a second step, we increase the cap for voluntary pension savings to show that the two reforms have opposite effects.

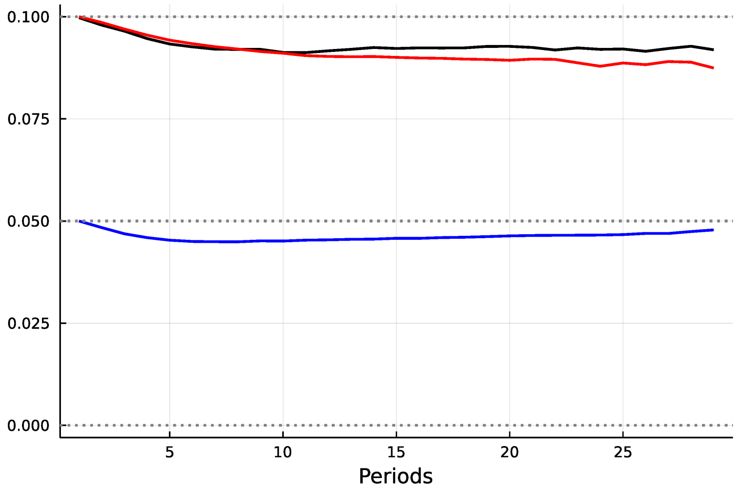
5.1 Comparison: Switzerland and Germany

We simulate the baseline economies in Switzerland and Germany, as described in Table 4.1, to assess how much of the difference in amortisation behaviour can

[16] cf. Section 2.4.



(a) Mean liquid equity.



(b) Mean pension contribution.

Figure 5.1: Comparison Switzerland–Germany.

be explained by the model. The tax incentives in Switzerland entail taxes on the imputed rent and tax deductions for mortgage interest payments and voluntary pension savings up to a threshold $\bar{p} = 0.1$. German homeowners do not face any tax incentives for their home, but can tax deduct voluntary pension savings up to $\bar{p} = 0.05$. Figure 5.1 depicts the mean of liquid equity a in Panel (a) and pension contributions, i.e. the savings in asset a_p, in Panel (b).

Comparing the two countries, we see that homeowners in Germany amortise more than in Switzerland as shown by the black and blue lines in Panel (a). This is in line with the empirical evidence presented in Section 2. Computing the LTV ratio of homeowners at retirement, we get a median of around 40% in Switzerland (as targeted) and around 20% in Germany.[17] Empirically, the median LTV ratio for homeowners in Germany at retirement is zero. Thus, our numerical illustration explains around 50% of the difference in LTV ratios between Switzerland and Germany.

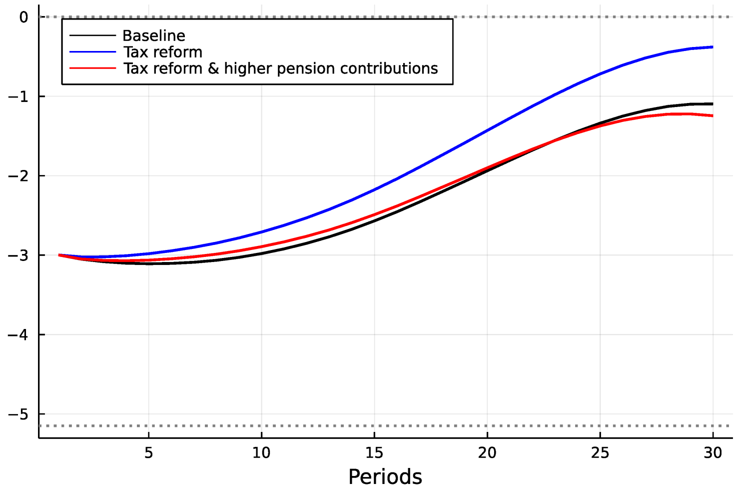
Panel (b) of Figure 5.1 depicts pension contributions over the life cycle. As before, the black and blue line represent pension contributions by Swiss and German homeowners. In both countries, the mean contribution is close to the cap $\bar{p}$. Once households have a high enough buffer stock, the after-tax returns of pension savings dominate mortgage amortisation. For each unit saved in the pension asset, agents get an *immediate* return corresponding to the marginal tax rate, whereas the gains of tax-deducting mortgage interest payments accrue over time and are discounted with discount factor β. Overall, the difference in amortisation behaviour is not only driven by the tax-deductibility of mortgage interest payments, but also by the difference in the cap $\bar{p}$.

The red lines in Figure 5.1 show the result of a counterfactual experiment, in which we simulate German homeowners facing Swiss tax incentives. This means they can tax deduct mortgage interest payments but have to pay taxes on their imputed rent. Furthermore, the cap $\bar{p}$ doubles from $\bar{p} = 0.05$ to $\bar{p} = 0.1$. The results show that exposing German homeowners to Swiss tax incentives results in much less mortgage amortisation over the life cycle. In fact, the resulting median LTV ratio at retirement would be approximately 60% and thus, 50% higher than in Switzerland. This is due to the fact that taxes in Germany are more progressive than in Switzerland and therefore, the incentives to reduce taxable income are stronger. Introducing tax-deductibility of mortgage interest payments in Germany would thus not only reduce amortisation, but would do so to a greater extent than in Switzerland.

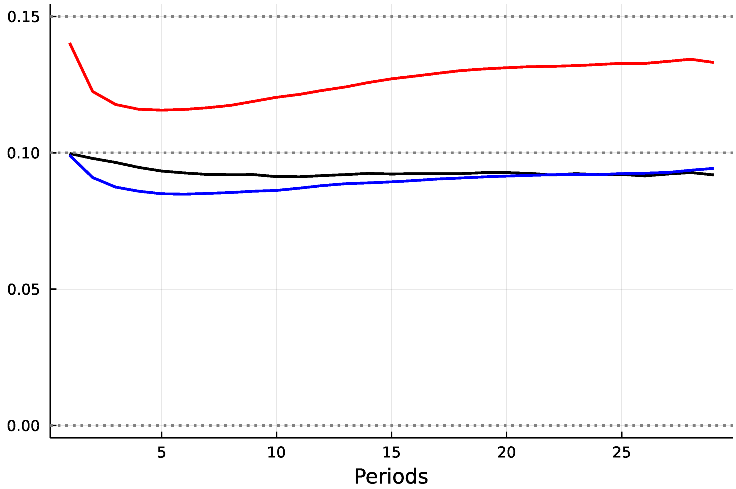
5.2 Tax reform in Switzerland

We perform two counterfactual experiments for Switzerland, which inform the current policy debate on parliamentary initiatives 17.400 and 20.494. First, we introduce a tax reform by abolishing the taxation of imputed rents and the tax deductions for mortgage interest payments (related to discussions within initiative 17.400). In a second step, we increase the cap for voluntary pension

[17] We compute the LTV ratio by setting it to zero for homeowners with positive liquid equity, i.e. $a \geq 0$, and to $-a/h_0$ for homeowners with $a < 0$.



(a) Mean liquid equity.



(b) Mean pension contribution.

Figure 5.2: Tax reform and higher pension contributions.

savings by 50% to $\bar{p} = 0.15$ (related to discussions within initiative 20.494). Thus, homeowners cannot tax deduct mortgage interest payments but can tax deduct a higher amount of voluntary pension savings.

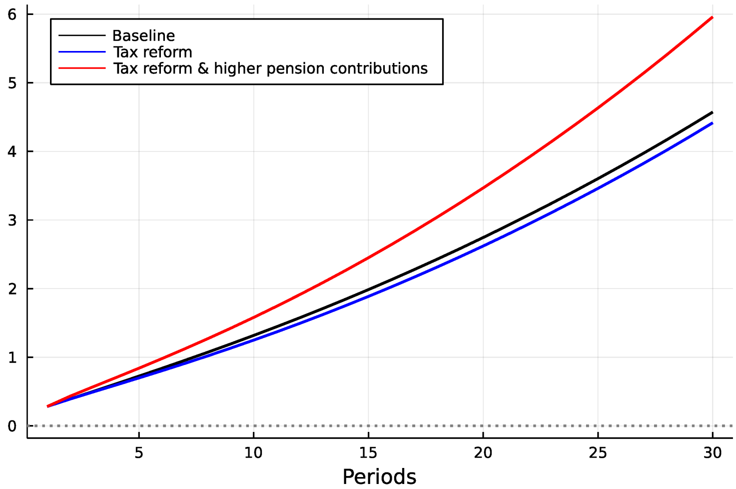
We show our results in Figure 5.2. As before, Panel (a) shows the mean holdings of asset a and Panel (b) depicts pension contributions. The black lines depict the baseline case for Switzerland as in the previous subsection. Although homeowners reduce their mortgage over the life cycle, they still hold a significant amount of mortgage debt at retirement.

We abolish the taxation of imputed rents and the deduction of mortgage interest payments from taxable income. The blue line represents the tax reform proposed in parliament. As one can see in Panel (a), this leads to higher mortgage amortisation over the life cycle. Computing the LTV ratio at retirement yields a median LTV ratio of approximately 20% compared to 40% observed in Switzerland. Thus, the tax reform leads to a reduction of the median LTV ratio at retirement by 50% in our numerical illustration. In Panel (b), we see that in both cases, average pension contributions are close to $\bar{p}$ and thus, mortgage amortisation is driven by homeowners saving more than implied by the maximum amount of possible voluntary pension savings.

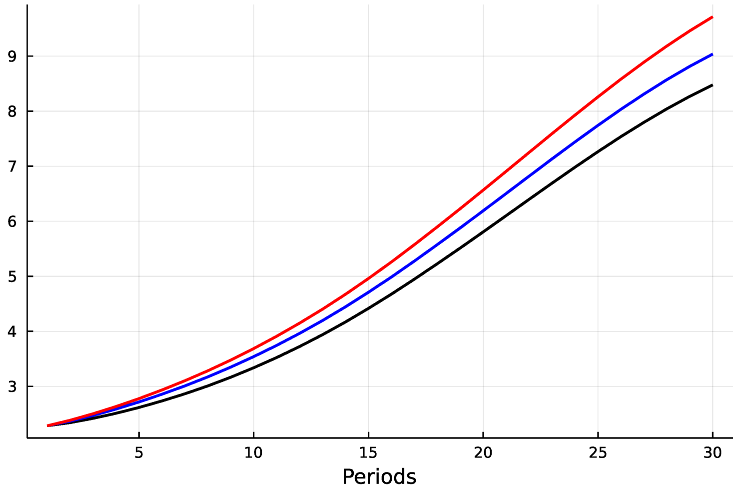
In addition to the tax reform, we then increase the cap for the tax-deductibility of voluntary pension contributions by 50% to $\bar{p} = 0.15$. The resulting amortisation behaviour is shown by the red line in Panel (a) of Figure 5.2. Given that the after-tax return of the pension asset dominates mortgage amortisation, increasing the cap by 50% directs savings away from home equity to pension savings. As a result, the increase in amortisation caused by the tax reform is offset by the increase of the cap $\bar{p}$ and in our numerical illustration. The two effects practically cancel out.

While abolishing the tax-deductibility of mortgage interest payments and the increase of the cap for voluntary pension savings are discussed separately in parliament, our numerical illustration highlights the interaction between the two proposals. Abolishing the taxation of imputed rents and the tax-deductibility of mortgage interest payments may indeed lead to more mortgage amortisation, although our numerical example suggests that the effect is not as strong is commonly believed. However, increasing the cap for voluntary pension savings has an opposing effect and is very likely to offset the effect of the tax reform. Although fostering mortgage amortisation is not the only goal of the tax reform, it is talked about very prominently and might not be achieved if the cap for voluntary pension savings is increased at the same time. Thus, the two proposals should not be discussed in isolation.

Moreover, the tax incentives also have an impact on the accumulation of wealth, as shown in Figure 5.3. Panel (a) shows the mean holdings of the pension asset a_p over the life cycle while Panel (b) shows household net worth. The tax reform leads to higher mortgage amortisation, which results in slightly less pension savings. The net effect is positive, which can be seen by higher household net worth in Panel (b). Introducing a higher cap for voluntary pension savings leads to even more pension wealth and higher household net worth, as shown by the red line. This is the result of two effects at play. The first effect is a mechanical



(a) Mean pension fund.



(b) Mean net worth.

Figure 5.3: Household wealth.

wealth effect arising due to the tax deductions for homeowners investing in the pension asset. The second effect is due to the incentive to save through the illiquid pension asset itself, as it encourages more homeowners to invest part of their income long-term. Given that more homeowners choose to save through an illiquid asset and reap the benefit of compounded interest, the resulting net worth is higher.

6 Conclusion

Amortisation patterns differ strongly between Switzerland and other European countries. While homeowners usually amortise their mortgage until retirement, this is not the case in Switzerland, where around 90 percent of homeowners still hold mortgage debt at retirement. Although saving in home equity is often seen as a means to save for retirement, this is less common for Swiss homeowners who rather accumulate voluntary pension savings that are strongly tax-incentivised.

We have analysed how amortisation behaviour is affected by tax incentives for mortgage amortisation and retirement savings in Switzerland and Germany. We find that tax policy matters quantitatively for household leverage and may explain differences in amortisation behaviour between Switzerland and Germany. By exposing German homeowners to Swiss tax incentives, we have further illustrated that tax incentives are stronger if tax progressivity is higher as in Germany.

We have applied our model to inform the current policy debate in Switzerland. We find that a tax reform which abolishes the tax-deductibility of mortgage interest payments leads to more amortisation. If the scope for tax deductions of voluntary pension savings is increased at the same time, as currently debated in the Swiss parliament, the effects of the tax reform might be offset.

Our work presented here offers several directions for future research. Mian et al. (2013) show that household leverage poses a risk to financial and economic stability because of mortgage delinquency after housing busts and the deleveraging of households that results in persistent consumption slumps. This risk has to be weighed against the opportunity for portfolio diversification which leverage offers. Higher leverage, for example due to slower amortisation of mortgages, is not necessarily associated with more current household consumption. Higher leverage may allow households to invest more into other assets to strengthen their portfolio diversification.

Our model lends itself to the analysis of the consumption response to house price shocks and the associated impact on economic stability. For such an analysis, a natural extension of the model is to include the decision to be a renter or homeowner. If circumstances change, for example due to unemployment or children leaving the household, homeowners might decide to sell their property to become renters or buy a smaller house.

References

Alan, S., Atalay, K., Crossley, T. F., and Jeon, S.-H. (2010). New evidence on taxes and portfolio choice. *Journal of Public Economics*, 94(11-12):813–823.

Amromin, G., Huang, J., and Sialm, C. (2007). The tradeoff between mortgage prepayments and tax-deferred retirement savings. *Journal of Public Economics*, 91:2014–040.

Badarinza, C., Campbell, J. Y., and Ramadorai, T. (2016). International comparative household finance. *Annual Review of Economics*, 8:111–144.

Basten, C. and Koch, C. (2015). The causal effect of house prices on mortgage demand and mortgage supply: Evidence from Switzerland. *Journal of Housing Economics*, 30:1–22.

Baumberger, J. (1999). Die volkswirtschaftlichen Auswirkungen bei einem Systemwechsel der Eigenmietwertbesteuerung. *Gutachten zuhanden der Eidgenössischen Kommission Eigenmietwert/Systemwechsel (KES)*.

Börsch-Supan, A., Bristle, J., Andersen-Ranberg, K., Brugiavini, A., Jusot, F., Litwin, H., and Weber, G., editors (2019). *Health and Socio-Economic Status over the Life Course. First Results from SHARE Waves 6 and 7*. De Gruyter.

Bover, O., Casado, J. M., Costa, S., Caju, P. D., McCarthy, Y., Sierminska, E., Tzamourani, P., Villanueva, E., and Zavadil, T. (2016). The distribution of debt across euro-area countries: the role of individual characteristics, institutions, and credit conditions. *International Journal of Central Banking*, 12(2):71–128.

Bundesamt für Statistik (2013). Haushaltsbudgeterhebung 2011: Kommentierte Ergebnisse und Tabellen. *BFS Aktuell, July*.

Campbell, J. Y. (2006). Household finance. *Journal of Finance*, 61(4):1553–1604.

Campbell, J. Y. and Cocco, J. F. (2003). Household risk management and optimal mortgage choice. *Quarterly Journal of Economics*, 118(4):1449–1494.

Chetty, R., Sándor, L., and Szeidl, A. (2017). The effect of housing on portfolio choice. *Journal of Finance*, 72(3):1171–1212.

Cloyne, J., Ferreira, C., and Surico, P. (2020). Monetary policy when households have debt: New evidence on the transmission mechanism. *Review of Economic Studies*, 87(1):102–129.

Eidgenössische Steuerverwaltung (2015). Die Besteuerung der Eigenmietwerte. *Steuerinformationen Schweizerische Steuerkonferenz SSK*.

European Central Bank (2009). Housing finance in the Euro area. *European Central Bank, Occasional Paper Series, No. 101*.

Eurosystem Household Finance and Consumption Network (2013). The Eurosystem Household Finance and Consumption Survey. Results from the first wave. *ECB statistics paper*.

Fernández-Villaverde, J. and Krueger, D. (2007). Consumption over the life cycle: Facts from consumer expenditure survey data. *Review of Economics and Statistics*, 89(3):552–565.

Fuchs-Schündeln, N., Krueger, D., and Sommer, M. (2010). Inequality trends for Germany in the last two decades: A tale of two countries. *Review of Economic Dynamics*, 13(1):103–132.

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2019). The German Socio-Economic Panel (SOEP). *Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik)*, 239(2):345–360.

Gruber, J., Jensen, A., and Kleven, H. (2021). Do people respond to the mortgage interest deduction? Quasi-experimental evidence from denmark. *American Economic Journal: Economic Policy*, 94(11-12):813–823.

Guiso, L. and Sodini, P. (2013). Household finance: an emerging field. In *Handbook of the Economics of Finance*, volume 2, pages 1397–1532. Elsevier.

Heathcote, J., Storesletten, K., and Violante, G. L. (2017). Optimal tax progressivity: An analytical framework. *The Quarterly Journal of Economics*, 132(4):1693–1754.

Hlavac, M. (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Central European Labour Studies Institute (CELSI), Bratislava, Slovakia. R package version 5.2.2.

Kaplan, G. and Violante, G. L. (2014). A model of the consumption response to fiscal stimulus payments. *Econometrica*, 82(4):1199–1239.

Karlman, M., Kinnerud, K., and Kragh-Sørensen, K. (2021). Costly reversals of bad policies: the case of the mortgage interest deduction. *Review of Economic Dynamics*, 40:85–107.

Laibson, D., Maxsted, P., and Moll, B. (2021). Present bias amplifies the household balance-sheet channels of macroeconomic policy. *NBER Working Paper 29094*.

Mian, A., Rao, K., and Sufi, A. (2013). Household balance sheets, consumption, and the economic slump. *The Quarterly Journal of Economics*, 128 (4):1687–1726.

Morger, M. (2014). Anreizmechanismen, Verteilungseffekte und finanzielle Auswirkungen verschiedener Reformoptionen. *Gutachten zuhanden der Eidgenössischen Steuerverwaltung (ESTV)*.

OECD (2007). *Pensions at a Glance*. OECD.

Piazzesi, M. and Schneider, M. (2016). Housing and macroeconomics. In *Handbook of Macroeconomics*, volume 2, pages 1547–1640. Elsevier.

Poterba, J. M. (2001). Taxation and portfolio structure: issues and implications. *NBER Working Paper 24196*.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Schneider, M. and Wagner, K. (2016). Housing markets in Austria, Germany and Switzerland. In Augustyniak, H., Łaszek, J., Olszewski, K., and Waszczuk, J., editors, *Papers presented during the Narodowy Bank Polski Workshop: Recent trends in the real estate market and its analysis*, NBP Conference Publications, chapter 6. Narodowy Bank Polski, Economic Research Department.

Sommer, K. and Sullivan, P. (2018). Implications of US tax policy for house prices, rents and homeownership. *American Economic Review*, 108(2):241—-274.

Steuerverwaltung des Kantons Bern (2016). Erläuterungen zum steuerlichen Bewertungssystem von Grundstücken und Liegenschaften im Kanton Bern. *Amtliche Bewertung*.

Swiss National Bank (2018). *Financial Stability Report 2018*.

Voorpostel, M., Tillmann, R., Lebert, F., Kuhn, U., Lipps, O., Ryser, V.-A., Antal, E., Monsch, G.-A., Dasoki, N., Klaas, H. S., and Wernli, B. (2020). *Swiss Household Panel User Guide (1999-2018), Wave 20.* FORS.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686.

A Appendix on data sources and calibration

We document our data sources and the construction of the sample in Appendix A.1. In Appendix A.2 we provide details on the estimation of the income process. Appendix A.3 explains how we estimate the tax function. For the analysis of all the data sets, we use the infrastructure provided by the R Core Team (2020), Wickham et al. (2019) and Hlavac (2018, for tables with summary statistics).

A.1 Data description

Panel data on income

The Swiss Household Panel (SHP) and the German Socio-Economic Panel (SOEP) are unbalanced, annual panel data sets, consisting of 20 waves from 1999 to 2018 for the SHP and 35 waves from 1984 to 2018 for the SOEP. We use the waves for the sample years since 2000,[18] for which we have information for both countries and which coincide with the sample period after the introduction of the euro. For further description of the two data sets, we refer to Voorpostel et al. (2020) and Goebel et al. (2019) for the SHP and SOEP.

For both the SHP and the SOEP, we use the cross-national equivalent files (CNEF) which offer standardised variables for their core samples. We use information from the respective core sample on the age of individual, the number of persons and the number of children in the household, education, the employment level of household members, the area of residence,[19] household labour income, household asset income, public transfers received by the household, social security pensions, tax payments and social security taxes, and housing tenure.

The SOEP and SHP consist of $275,340$ and $94,939$ observations for the household reference person in the considered sample period. We eliminate observations with missing information in any of the variables, and we trim the top and bottom percent of households according to the three income definitions defined below to remove outliers. This reduces the sample to $170,542$ and $54,535$ observations for the SOEP and SHP. Based on these observations, we consider household reference persons between the ages of 36 and 65 who are working[20] given

[18] In the first wave of the SHP (1999), information on social security pensions and total household taxes is not included.

[19] This is the *Kanton* in Switzerland and the *Bundesland* in Germany, which are geographical units comparable to the states in the U.S.

[20] We remove all households receiving pension benefits.

the focus on homeowners with labour earnings prior to retirement in our analysis. After removing households outside this age group, our sample consists of 125,605 and 41,442 observations for the SOEP and SHP. For the calibration of our model, we focus on homeowners who account for 64,212 and 23,550 observations in the SOEP and SHP.

For our analysis, we construct three income variables:

- *Household labour income*: household labour income after social security contributions and transfers such as unemployment benefits.
- *Taxable income*: the sum of household labour income and asset income.[21]
- *Disposable income*: taxable income net of tax payments.[22]

We use *household labour income* to estimate the life cycle income profile and the income process, as further explained in Appendix A.2. We estimate the tax function using *disposable income* and *taxable income*, as described in Appendix A.3.

We adjust household labour income using the *square root equivalence scale* recently adopted by the OECD.[23] This equivalence scale is very similar to scales previously proposed in the literature, as for example the benchmark scale reported in Fernández-Villaverde and Krueger (2007), Table 1, last column. We thus divide household labour income by the square root of the households' size.

Tables A.1 and A.2 provide summary statistics for the most important variables in our sample. For comparability, we show the statistics for the sample year 2015, given that we use the waves of the SHARE and HFCS for the years 2015 and 2014, respectively.

Tables A.1 and A.2 show that the number of children and the household size is slightly larger in Switzerland than in Germany in the population and also in the sample of homeowners. Educational attainment is also somewhat higher in the Swiss than in the German sample. As is well known, (equivalised) household labour income is much higher in Switzerland than in Germany, where one euro exchanged approximately for 1.0 – 1.1 CHF in 2015. Within each of the two countries, the household size is larger and income is higher on average for

[21] Household labour income includes unemployment benefits or other welfare transfers which are taxed in Switzerland and determine the marginal tax rate in Germany.

[22] Total taxes paid by households include social security taxes and direct taxes. Because social security taxes have been deducted already when computing household labour income above, only direct taxes are deducted in this step.

[23] See http://www.oecd.org/economy/growth/OECD-Note-EquivalenceScales.pdf

Table A.1: Summary statistics for the SHP.

	All	Ages 36–65	Homeowners (36–65)
Average household size	2.64	2.75	3.09
Average number of kids	0.68	0.74	0.85
Fraction with tertiary degree	0.46	0.46	0.48
Mean household income	67,713	70,330	74,146
Median household income	63,119	65,536	68,390
Mean disposable income	58,293	60,627	64,578
Median disposable income	55,304	56,935	60,366

Notes: Income equivalized and in CHF. The tertiary education degree is associated to the household reference person.
In 2015, the euro exchanged approximately for 1.0–1.1 CHF.
Source: Swiss Household Panel (2015).

homeowners than in the rest of the sample. In Germany, homeowners also have higher education attainment relative to the rest of the sample.

Data on wealth

We use three different data sets to obtain information on wealth: the Survey of Health Ageing and Retirement in Europe (SHARE), the Household Finance and Consumption Survey (HFCS) and the Swiss Household Budget Survey (*Haushaltsbudgeterhebung* or HABE). For comparability, we use data for the years 2014 (HFCS) and 2015 (SHARE, HABE).

The SHARE is an unbalanced panel data set collecting data from around 140,000 individuals at age 50 or above in 28 European countries and Israel. It is harmonised to be comparable with its data-set counterparts in England (ELSA) and the U.S. (HRS). Currently, seven waves are available between 2004 and 2017. We use the data for Germany and Switzerland. Börsch-Supan et al. (2019) provide further description of the data set.

The SHARE consists of 2,704 German and 1,855 Swiss households in the considered sample period. We eliminate households with a measured LTV ratio above one to remove outliers. This leaves us with a sample of 2,694 German and 1,816 Swiss households. The age group 56–65 contains 911 and 635 households for Germany and Switzerland.

The HFCS is a repeated survey collecting detailed data on wealth for the euro area. Currently, three waves are available for 2010, 2014 and 2017. We use

Table A.2: Summary statistics for the GSOEP.

	All	Ages 36–65	Homeowners (36–65)
Average household size	2.30	2.36	2.72
Average number of kids	0.51	0.51	0.60
Fraction with tertiary degree	0.30	0.29	0.34
Mean household income	29,426	31,130	36,704
Median household income	25,931	27,932	33,032
Mean disposable income	24,543	25,839	30,178
Median disposable income	22,590	24,035	27,808

Notes: Income equivalized and in EUR. The tertiary education degree is associated to the household reference person.
Source: German Socio-Economic Panel (2015).

the data for Germany given that Switzerland does not take part in this survey. For further information see Eurosystem Household Finance and Consumption Network (2013).

The HFCS consists of 4,461 German households in the considered sample period. We eliminate households with a measured LTV ratio above one to remove outliers. This leaves us with a sample of 4,400 German households, of which 582 are in the age group 36–45 and 955 are in the age group 56–65.

To estimate the homeownership rate and the incidence of mortgages for Swiss households below age 50, we use the HABE. This data set contains information on household expenditure and income. We use the information on housing expenditure and mortgage interest payments to infer the homeownership rate and the incidence of mortgages by age group.

The HABE is a repeated cross-sectional data set with survey information on household expenditures and income since 2000. The survey wave in 2015 has 3,469 observations. 651 of these observations are in the age group 36–45 and 672 are in the age group 56–65. For further information on the data set, see Bundesamt für Statistik (2013).

A.2 Income process

We use the panel data on income from the SHP and the SOEP to decompose life cycle income into a deterministic age profile, time effects, and a stochastic, idiosyncratic AR(1) income process. As described in Appendix A.1, household labour income is defined as labour income after social security contributions and transfers, and it is equivalised to account for differences in household size. We estimate the process for households with a reference person in the labour force between ages 36 and 65, and thus for the part of the life cycle before retirement.

We estimate the income process in logs, and assume that the deterministic and stochastic income components are characterised by the following two equations:

$$\begin{aligned} \log y_{i,t,j} &= \alpha_j + \gamma_t + \mathbf{x}'\delta + \tilde{y}_{i,t} \\ \tilde{y}_{i,t} &= \rho \tilde{y}_{i,t-1} + \epsilon_{i,t}, \quad \text{with } \epsilon_{i,t} \sim \mathcal{N}(0, \sigma_\epsilon^2), \end{aligned} \tag{A.1}$$

where α_j denotes age effects that we approximate by a fourth-order polynomial, γ_t controls for time effects that are common across households, and $\mathbf{x}$ includes a vector of controls including a set of dummies for education and the area of residence. The stochastic income component $\tilde{y}_{i,t}$ has persistence given the assumption of an AR(1) process.

Figure A.1 shows the deterministic age-dependent income profile for Switzerland and Germany. The figure displays the fourth-order age polynomial estimated on age-year moving averages of income using bins with a width of three age years. The figure shows that equivalised household labour income reaches its peak at ages in the mid fifties both in Germany and in Switzerland. The respective cumulative income growth during the preceding twenty age years is 16% and 23%, corresponding to 0.75% and 1% annualised income growth between ages 36 and 56.

For the income process in our stationary model economy, we retrieve the estimates of six parameters – the four coefficients for the age polynomial, which we have illustrated in Figure A.1, and the persistence parameter ρ and the variance of innovations σ_ϵ^2 for the stochastic component of the income process, which are reported in Table A.3. The estimated persistence of the income process is 0.77 for Switzerland and 0.81 for Germany, and the variance of the innovations is 0.09 and 0.08. For Germany, we can benchmark these results to Fuchs-Schündeln et al. (2010, pp. 122–123), who estimate a slightly different income process, with a permanent and transitory component, for the time period between 1984 and

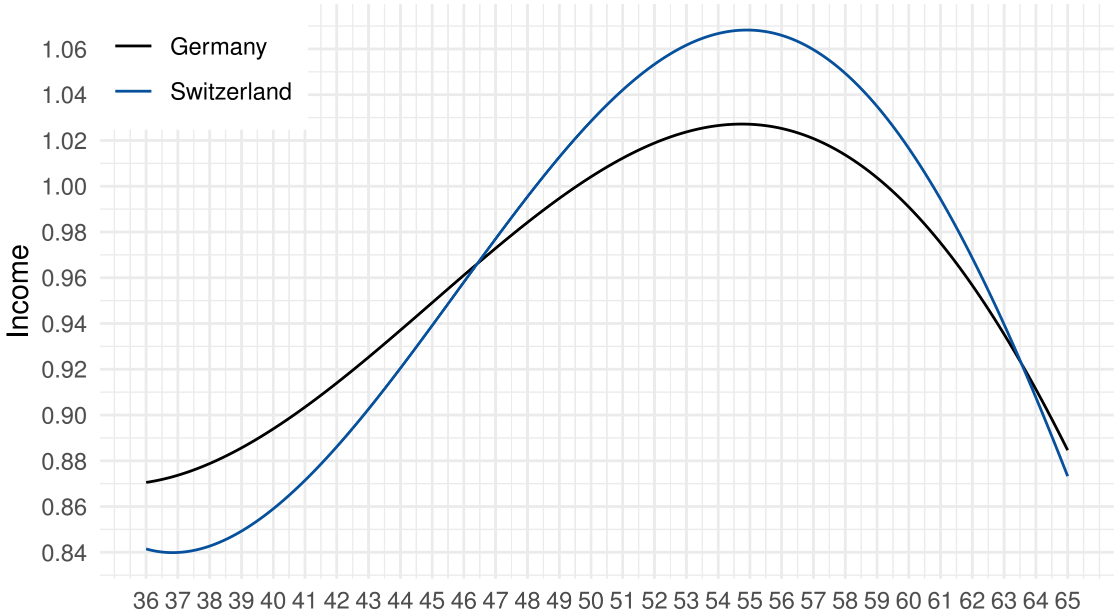


Figure A.1: Labour income profiles.

Notes: Age on the horizontal axis in years; labour income in model units of average equivalised household labour income, as defined in Appendix A.1.
Source: SHP / SOEP 2000–2018.

2003. Our estimate for the variance of the innovation is above their estimate of 0.05 for the variance of the permanent component at the beginning of the 2000s, and below their estimate of 0.15 for the variance of the transitory component, as one would expect for our assumption of an AR(1) income process with less than full persistence.

Table A.3: Income Process

	Switzerland	Germany
Income autocorrelation ρ	0.771	0.806
	(0.761, 0.780)	(0.800, 0.811)
Variance of innovations σ_ϵ^2	0.086	0.078

Note: 95% confidence intervals in brackets.
Sources: SHP / SOEP.

A.3 Estimation of the tax function

We estimate a tax function with constant progressivity as, for example, in Heathcote et al. (2017):

$$T(y_\tau) = y_\tau - \lambda y_\tau^{1-t_y}. \tag{A.2}$$

The tax function is characterised by two parameters, λ and t_y, where λ determines the tax level and t_y the degree of progressivity. This function includes both tax and transfer policies because the tax function evaluates to negative values at sufficiently low levels of income. The income threshold which distinguishes taxpayers from receivers of transfers is

$$y_0 = \lambda^{1/t_y}. \tag{A.3}$$

We define taxable income y_τ and disposable income y_d as described in Appendix A.1. For the tax function above, the relationship between taxable income and disposable income is

$$y_\mathrm{d} = \lambda y_\tau^{1-t_y}, \tag{A.4}$$

which we estimate in logs:

$$\log y_\mathrm{d} = \log \lambda + (1 - t_y) \log y_\tau. \tag{A.5}$$

Table A.4 displays our OLS-estimates for Switzerland and Germany. The estimated tax progressivity is significantly higher in Germany, in line with the statutory tax rates in the two countries.[24]

[24]For comparison, the estimate $t_y = 0.181$ by Heathcote et al. (2017) for the U.S. is based on a sample of households with at least part-time employment. If we imposed the same sample restrictions, the U.S. estimate would lie between the estimate for Germany and Switzerland. We use a different sample for our analysis because the income risk which we approximate with the

Table A.4: Tax parameters

	Switzerland	Germany
Tax level parameter λ	2.96	5.02
	$(2.91, 3.01)$	$(4.98, 5.06)$
Tax progressivity parameter t_y	0.113	0.174
	$(0.112, 0.115)$	$(0.173, 0.174)$
Income threshold y_0	$14,596$	$10,867$
R-Squared	0.98	0.99

Note: 95% confidence intervals in brackets. Income in local currency.
In 2015, one euro exchanged approximately for 1.0–1.1 CHF.
Sources: SHP / SOEP.

The top panel of figure A.2 illustrates the empirical fit for Switzerland and Germany, as depicted by the respective blue line. The intersection with the dashed 45-degree line depicts the income threshold at which net transfer receivers become net taxpayers. The values of the threshold are displayed in the third row of Table A.4. They are broadly in line with the statutory income thresholds of $8,472$ euro and $14,500$ CHF reported in the OECD tax database for both countries in 2015 (see table i1). The empirical fit of the tax function is remarkable, illustrated by the R^2-statistic of 0.98 and 0.99 for Switzerland and Germany. The fit is slightly better than in Heathcote et al. (2017) for the U.S., who report a R^2-statistic of 0.91 for a sample of households with at least part-time employment. The bottom panel of figure A.2 plots the average and marginal tax rates implied by our estimates.

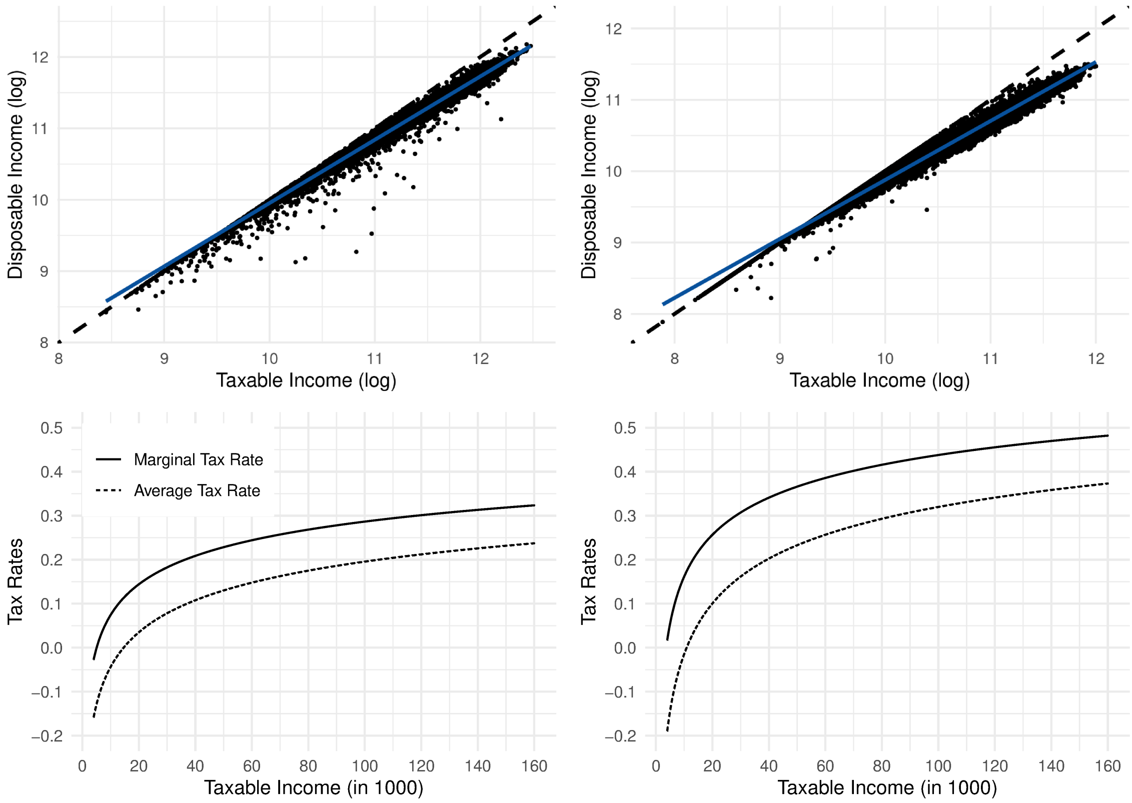
When we normalise income, we adjust the tax function so that the tax rate is invariant to the normalisation for a given progressivity. Thus, when normalising income by some value $\overline{y}$, we require

$$1 - \lambda y_\tau^{-t_y} = 1 - \tilde{\lambda} \left(\frac{y_\tau}{\overline{y}} \right)^{-t_y}, \tag{A.6}$$

where $\tilde{\lambda}$ is the parameter for the adjusted tax function. Thus,

$$\tilde{\lambda} = \lambda \overline{y}^{-t_y}. \tag{A.7}$$

stochastic income process described in Appendix A.2 includes the risk of unemployment. Hence, household labour income, defined in Appendix A.1, includes transfers such as unemployment benefits. These benefits are taxed in Switzerland and determine the marginal tax in Germany.



(a) Switzerland

(b) Germany

Figure A.2: Estimation of the tax function and implied tax rates.

Notes: Income is equivalised and in local currency. The base year is 2015.
In 2015, one euro exchanged approximately for 1.0–1.1 CHF.
Source: SHP / SOEP 2000–2018.

III | Education choices and drop-out risk: the role of the education system

Abstract

I analyse the idiosyncratic drop-out risk associated with education choices over the life cycle. The binary choice between working and attending university generates the well-known trade-off between foregone earnings and returns to education. I show that education systems which enrich education choices beyond this binary choice, offering choices which differ in expected returns and drop-out risk, improve sorting into education, which yields a Pareto improvement. Because drop-out risk depends on innate ability, a second trade-off then arises between higher expected returns and higher drop-out risk, which is shaped by the design of the education system. I characterise this trade-off and show that progressive taxation might *increase* enrolment in tertiary education because it does not only reduce the expected return to education, but also its variance.

Keywords: University drop-out risk, Returns to education, Taxation.
JEL Codes: H24, I21, J24.

1 Introduction

Going to university is a risky endeavour, as dropping out of it may completely change the course of a young student's career. Yet, whether students succeed in their studies not only depends on their effort and ability but also on *where* they study, as there is a lot of heterogeneity in drop-out rates between countries, but also within countries.[1]

I study how the design of the education system affects education choices along the academic ability distribution. More specifically, I look at how the strictness of examinations both at entry and exit of tertiary education, i.e. admission and final exams, affects enrolment and attainment rates in tertiary education. Furthermore, I show that education systems enriching education choices beyond the binary choice between working and attending university may improve sorting into education, yielding a Pareto improvement. This is done by providing options which differ in expected drop-out risk and returns to education, hence offering more risk-return profiles to choose from.

I model education choices in a life cycle model where agents differ in academic ability, which determines their drop-out probabilities while in education. The binary choice between working and going to university generates the well-known trade-off between foregone earnings and returns to education, and highlights how drop-out risk affects expected returns to education. I then extend the model to allow for a richer education system offering a third choice, think, for example, of a shorter degree at a community college, which is characterised by lower drop-out risk and a lower expected education premium compared to university. This adds a second trade-off: given the choice to get tertiary education, agents have to weigh higher expected returns to education against higher drop-out risk.

I show that the education system with the richer choice set improves sorting into education. Introducing an additional option improves the outcome for two groups. On the one hand, among those who do not go to university, individuals at the top end of the academic ability distribution profit from an education option with lower drop-out risk and a lower expected return. Those are the marginal agents, for whom going to university would be too risky. The additional option offers these agents a better risk-return profile. On the other hand, among those who go to university, individuals at the bottom end of the academic ability dis-

[1] See, e.g. OECD (2008, 2019) for an international comparison and BFS (2018) for evidence on the heterogeneity within Switzerland.

tribution prefer an option with lower drop-out risk as well. For them, the risk of going to university is worth taking, but they are among those agents with the highest drop-out risk and thus, the additional option offers them a better risk-return profile as well.

To assess how the education system affects sorting into education, I compare the outcomes with the solution of the maximisation problem of a planner who perfectly observes academic ability and offers full insurance against drop-out risk. The analysis shows that in a laissez-faire environment, the enrolment rates in tertiary education are too low and the planner would choose that more people go to university. I then show that an education system which offers more choices along the risk-return frontier allows to implement higher enrolment in tertiary education without requiring information on ability as in the planner problem. The introduction of an alternative education choice can be compared to offering partial insurance for the otherwise uninsurable drop-out risk. Agents forego a part of their education premium, which is the analogue of paying an insurance premium. In return, their risk of dropping out is reduced. Thus, enriching the education system by offering various risk-return profiles improves sorting into education.

In my analysis of the effect of progressive income taxation on education choices, I shed new light on the *Siamese twin* result by Bovenberg and Jacobs (2005), who argue that progressive taxation decreases investment in human capital as it diminishes the return to education. I show that in addition to decreasing the expected return to education, progressive taxation also decreases its variance and thus the risk associated with going to university. With CRRA preferences and risk aversion $\gamma > 1$, as typically assumed in the literature, progressive taxation might even *increase* enrolment in university. Hence, in contrast to the *Siamese twin* result, progressive taxation might bring the outcome closer to the planner's solution and improve sorting into education through risk reduction.

Finally, I use the model to analyse cross-sectional heterogeneity in enrolment and drop-out rates between cantons in Switzerland. Data from the Swiss Federal Statistical Office (BFS, 2018) reveal three key empirical relationships. There is a negative correlation between cantonal enrolment rates and success rates at the *Gymnasium* level, a negative correlation between cantonal *Matura* rates and success rates at university and a negative correlation between cantonal *Matura* rates and the share of graduates from universities of applied sciences (FH) among all graduates from tertiary education. In a numerical illustration, I simulate the model and show that both heterogeneity in the strictness of admission to *Gymnasium* and the difficulty of the *Matura* exams may explain the latter two relationships. However, a negative correlation between enrolment rates and success rates at the *Gymnasium* level can only be generated by heterogeneity in the strictness of admission to *Gymnasium*, because differences in the difficulty of *Matura* exams alone would yield a positive correlation instead.

Studying the case of Switzerland is interesting beyond the Swiss education policy debate. Given the strong focus on vocational training and alternative tertiary degrees, the Swiss education system has drawn attention from many other developed countries. In the past, ministers from the U.S. and other countries have been interested in learning about the Swiss education system.[2] Furthermore, there is substantial heterogeneity in the design of the education system across Swiss cantons, which can be exploited in the analysis.

My analysis contributes to the existing literature in three ways. First, I show how the design of the education system affects sorting into education. While the literature analyses various questions using a model of college investment with drop-out risk, none of the analyses have, to the best of my knowledge, investigated how changing drop-out probabilities affect sorting. Second, I contribute to the literature on taxation and human capital. As discussed before, the interaction of reducing both the mean and the variance of expected education returns qualifies well-known results in the literature. Third, I illustrate how my conceptual framework can be applied to interpret education outcomes in a country such as Switzerland, where the young can choose between a vocational track and between tertiary degrees with different risks and returns. Specifically, I discuss a potential mechanism explaining the observed heterogeneity between Swiss cantons.

Outline. Section 2 provides an overview of the related literature. In Section 3, I describe the Swiss education system and show stylised facts on cantonal differences in drop-out and graduation rates. Afterwards, I begin by illustrating the core mechanism of the model in Section 4, highlighting the central trade-offs. Section 5 then presents the model in full detail, followed by an analysis of the key implications of the model in Section 6. Section 7 concludes.

[2]See, e.g. **here** or **here**.

2 Related literature

This paper fits in the broad literature studying the determinants of educational attainment (e.g. Castro and Coen-Pirani, 2016; Donovan and Herrington, 2019). More specifically, I study the relation between drop-outs and educational attainment as surveyed by Bound and Turner (2011). Many authors have documented evidence on drop-outs and on the main factors associated with them.

Larsen et al. (2013) provide a systematic review of the empirical literature on drop-out phenomena. A very recent contribution is Behr et al. (2021) who study drop-out motives using data from bachelor students in Germany. Their findings indicate that drop-outs happen mainly because of student performance as well as expectations and interest about the courses they enrol in.

For Switzerland, Wolter et al. (2014) study drop-outs from universities with data covering all students between 1975 and 2008. They find a positive and significant association between cantonal *Matura* rates and the drop-out probabilities of students from those cantons. As pointed out by Wolter et al. (2014), this finding is consistent with the study of Eberle et al. (2008) who find a negative correlation between cantonal *Matura* rates and academic ability measured when taking the *Matura* exams. Thus, Wolter et al. (2014) conclude that academic ability and performance prior to entering university are the main reasons for dropping out of university.

Given the evidence, I analyse educational attainment in an environment where the drop-out probabilities are a function of academic ability. This approach follows recent contributions of several other authors.

Athreya and Eberly (2021) are most closely related to my work as they also analyse college enrolment decisions under drop-out risk.[3] The focus of their analysis is how college enrolment and attainment are affected by increasing college premia. Their main finding is that college enrolment and, to an even greater extent, college attainment react much less to a rise in college premia than one might expect because the students at the margin of enrolling have the highest drop-out risk. Thus, they conclude that higher college premia are more likely to increase income inequality rather than college attainment.

Similarly, Castex (2017) investigates how much of the college premium can be explained by drop-out risk. Using a life cycle model with drop-out risk, he finds

[3]See also Akyol and Athreya (2005) for an earlier contribution.

that the drop-out risk premium explains around 14 to 18 percent of the college premium.

My work is also related to Hendricks and Leukhina (2017, 2018). In both papers, the authors use a very similar model framework to study two different questions. The goal of Hendricks and Leukhina (2017) is to assess, how risky the investment into college education is. In contrast, I take drop-out probabilities as given and assess the effect of the education system on education choices. Hendricks and Leukhina (2018) aim to measure how much of the observed income differences between high school and college graduates can be explained by differences in ability. They find a large gap in ability between graduates from college and high school, which explains approximately 60 percent of the college premium. The authors model the process towards graduation in more detail than I do and include imperfect information about ability. However, they consider only one type of college, so there is no choice between different types of tertiary education.

Furthermore, I also shed new light on the *Siamese twin* result by Bovenberg and Jacobs (2005).[4] Redistribution through progressive taxation reduces investment into human capital as part of the returns are taxed away. However, in an environment with drop-out risk, progressive taxation also reduces the variance of mean earnings and thus the risk associated with human capital investment. The net effect depends on the level of risk aversion and I show that for common parameter values, progressive taxation might even *increase* enrolment in tertiary education.

My analysis also contributes to the literature on vocational education and training (VET) as surveyed by Wolter and Ryan (2011). Vocational education and training offers more specific human capital which facilitates a smoother school-to-work transition as documented in Ryan (2001) and Bradley and Nguyen (2004). The downside, however, is that specific human capital can become obsolete much faster as pointed out by Hanushek et al. (2017). This reverses the advantage of vocational education and training later in the life cycle. Yet, more recent work by Schweri et al. (2020) shows that in Switzerland, the skills acquired in vocational education and training are more transferable than conjectured.[5] Hence, VET offers a good alternative to general education in Switzerland not only in terms of

[4]Jacobs and Bovenberg (2011) expands on the result.
[5]See also Mueller and Schweri (2015).

the school-to-work transition but also when considering income over the full life cycle, as documented by Korber and Oesch (2019) and Aepli et al. (2021).

3 Institutional background

In this section, I outline the core aspects of the Swiss institutional setting. Section 3.1 gives a brief overview of the Swiss education system and introduces the necessary taxonomy. In Section 3.2, I show stylised facts on cantonal differences in drop-out rates and graduation rates. Readers who are interested in a more detailed breakdown of the Swiss education system are referred to SKBF (2018).

3.1 The Swiss education system

Swiss pupils have to complete nine years of compulsory schooling, starting at the age of six. After that, they can choose three more years of schooling, called *Gymnasium*, which ends with the *Matura* exam, which corresponds to the level 3A of the International Standard Classification of Educations (ISCED). Passing the *Matura* exams is the entry requirement to study at university.[6] However, only around 25 percent of all pupils choose to go to *Gymnasium*. The majority decides to get vocational education and training (VET), which corresponds to the ISCED 3B level. Vocational education and training is usually offered by firms, where apprentices work for three to four days a week, and attend school for the rest of the week to deepen their knowledge. Given the low *Matura* rates, access to university used to be restricted to a very small fraction of the population. This led to the introduction of the university of applied sciences, *Fachhochschule (FH)*, in the mid-90s, which offered access to tertiary degree programmes for VET graduates, corresponding to the ISCED 5B level (with university corresponding to level 5A). The programmes at universities of applied sciences (FH) typically offer degree courses, which are designed for students who already have experience in their field from their vocational programmes. They are often taught by practitioners and thus, have a strong focus on direct applicability to the workplace. This option makes vocational education and training a viable alternative to *Gymnasium*,

[6]There are a few other options, but those are only chosen by a small proportion of pupils. For more details, see SKBF (2018).

which is reflected in the fact that around half of all the graduates from tertiary degree programmes graduate from a university of applied sciences (FH).[7]

Taxonomy. The names of upper-secondary and tertiary education vary strongly between countries. I briefly clarify the terminology used here. The term *Gymnasium* is used for schools offering general education at the upper-secondary level. It is used both in Switzerland and Germany and is comparable to the *Lycée* in France or the *Sixth Form College* in the United Kingdom. *Gymnasium* ends with the *Matura* exams and pupils who passed them are said to *have a Matura*. The corresponding term in Germany is *Abitur*, which is comparable to the *Baccalaureate* in France and the *A-Levels* in the United Kingdom. Pupils choosing to get vocational education and training (VET) may get a *Berufsmatura*, which is comparable to the *Fachabitur* in Germany. The *Berufsmatura* enables them to enrol in a degree course at a university of applied sciences, *Fachhochschule (FH)*, in a field related to their VET degree. The term *Fachhochschule (FH)* is also used in Germany and is comparable to the former polytechnic in the United Kingdom or community colleges in the United States. For the model, I will be using the subscripts m, f and u for variables related to *Matura*, *Fachhochschule (FH)* and university.

3.2 Stylised facts

There is a lot of heterogeneity in university drop-out rates between countries.[8] The same is true if we look at the sub-national level in Switzerland both for drop-outs on the tertiary level and in upper-secondary education (*Gymnasium*). Figure 3.1 shows the enrolment and success rates on the *Gymnasium* level for each canton. One can see that there are large differences in enrolment rates, which range from less than 15 percent to almost 50 percent. The national average is around 25 percent. Furthermore, there is a negative correlation between enrolment and success rates. Assuming that the academic ability distribution is the same in all cantons, pupils attending *Gymnasium* in a canton with higher enrolment rates thus have, on average, a lower level of academic ability resulting in lower average success rates. This is in line with Eberle et al. (2008) and Wolter et al. (2014) who discuss this in more detail.

[7] See BFS (2018) for details.
[8] See, e.g. OECD (2008, 2019).

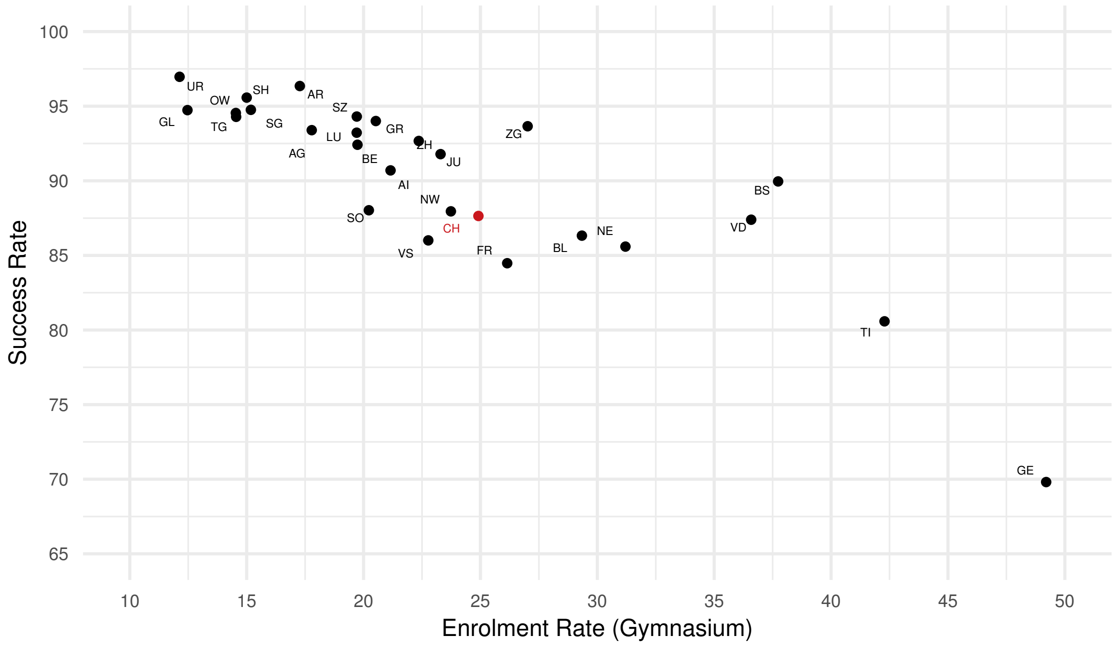


Figure 3.1: Enrolment and success rates by canton (*Gymnasium*).

Source: BFS (2018).

The heterogeneity in enrolment rates also translates into heterogeneity in cantonal *Matura* rates as shown in Figure 3.2. Given the negative correlation between enrolment and success rates, the heterogeneity in cantonal *Matura* rates is lower than the heterogeneity in enrolment rates, but the range between around 12 percent and almost 35 percent is still sizeable. With a national average between 20 and 25 percent, Switzerland has a very low *Matura* rate compared to other OECD countries.[9] The heterogeneity in cantonal *Matura* rates also affects drop-out rates at university. Figure 3.2 depicts university drop-out rates for students based on the canton in which they passed their *Matura* exams. As before, higher cantonal *Matura* rates are associated with higher drop-out rates, which points towards lower average academic ability for students from cantons with higher *Matura* rates.

However, whether or not to go to university is not a binary choice, as the Swiss education system also offers the option to go to a university of applied sciences (FH) after graduating from vocational education and training (VET). Thus, one can hypothesise that cantons with a lower share of university graduates have a higher share of graduates from a university of applied sciences. The reasoning

[9]See OECD (2008) and SKBF (2014, 2018) for a discussion.

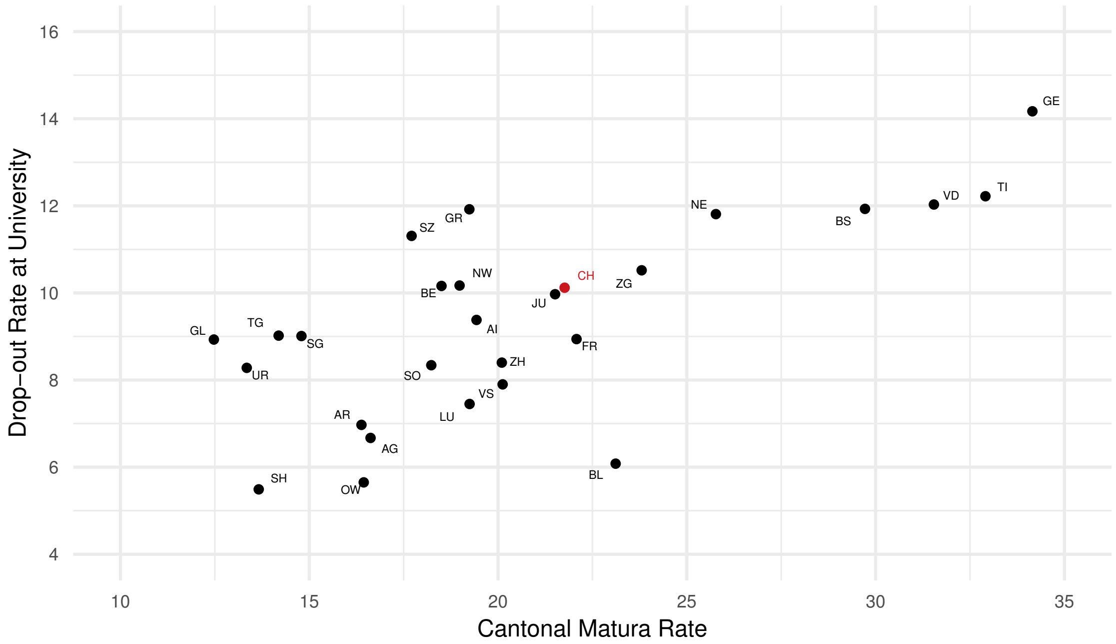


Figure 3.2: *Matura* and university drop-out rates by canton.

Source: BFS (2018).

behind this is that among the pupils who do not go to *Gymnasium*, those at the upper end of the academic ability distribution would still choose to study at a university of applied sciences. This group is larger in a canton with a low *Matura* rate, as a smaller share of the academic ability distribution goes to *Gymnasium*. This relation is shown in Figure 3.3. The y-axis depicts the share of graduates from universities of applied sciences (FH) and universities of teacher training (PH) as a fraction of all graduates from FH/PH and university.[10] As before, the cantonal *Matura* rate is depicted on the x-axis. The graph shows a negative correlation between the cantonal *Matura* rate and the share of students graduating from universities of applied sciences. This shows that in cantons where fewer pupils go to *Gymnasium*, more of them choose to study at a university of applied sciences (FH).

Summing up, students face drop-out risk both at the upper-secondary and tertiary level. In cantons with higher *Gymnasium* enrolment rates, pupils have a higher drop-out risk, on average. The same effect is visible at university, where students from cantons with higher *Matura* rates face higher drop-out risk. Fur-

[10]Note that in Switzerland, universities of teacher education (PH) are usually put into the same category as universities of applied sciences (FH).

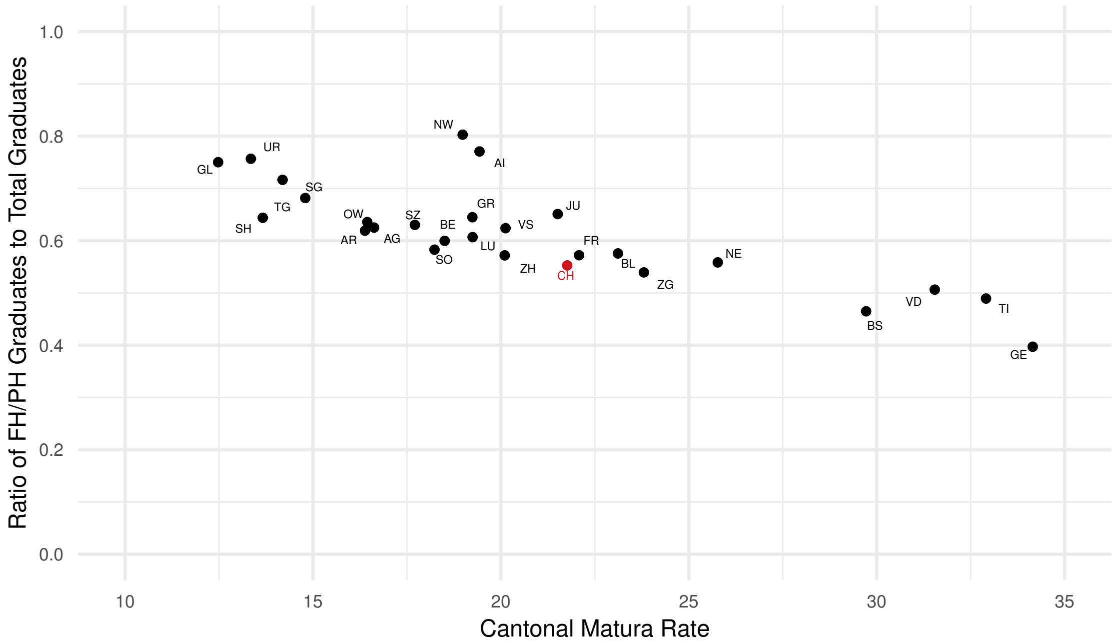


Figure 3.3: Fraction of FH graduates by canton.

Note: The graph shows the ratio of graduates from university of applied sciences (FH) and university of teacher training (PH) to total graduates from FH/PH and university.
Source: BFS (2018).

thermore, students from cantons with high *Matura* rates are less likely to go to a university of applied sciences (FH). They go to university instead, which contributes to the increased drop-out rates at the tertiary level.

4 Model mechanism

I begin by describing the model framework in the most general fashion. When choosing whether or not to go to university, young adults face a trade-off between going to work, which yields income now, and attending university, expecting to earn an education premium later. However, this is not a binary choice as there are many different options to choose from in tertiary education. Apart from choosing between various academic fields, these options differ by their expected drop-out probabilities and education premia. Thus, there is a second trade-off between maximising the expected education premium and minimising the probability to fail to graduate. I model these two trade-offs.

The economy is composed of a continuum of agents with an initial level of human capital h_0. Agents differ ex ante by their academic ability θ. They live for two periods and choose between working and getting tertiary education.

In the first period, agents who decide to get tertiary education choose their level of human capital h, which they attain with probability $p(h,\theta)$. With probability $1 - p(h,\theta)$, they drop out and their level of human capital remains at $h = h_0$. Hence, there is an uninsurable drop-out risk. While in education, agents get a subsistence income y_e. Agents choosing to work get income y_1. In the second period, agents get income $R(h,\theta)y_2$, with $R(h,\theta) \geq 1$. Thus, $R(h,\theta)$ is the education premium, which I normalise to $R(h_0,\theta) = 1$ for agents without a degree. This captures in a stylised way that, on average, drop-outs earn less than graduates. The assumption that they earn the same income as individuals that have not gone to university is made for simplicity. An intermediate income level between y_2 and $R(h,\theta)y_2$ could be specified without changing the main insights.

Each level of human capital h is therefore characterised by its corresponding success probability $p(h,\theta)$ and education premium $R(h,\theta)$, which can be described by their first derivatives. The success probability $p(h,\theta)$ is decreasing in the desired level of human capital h and increasing in academic ability θ, i.e.

$$\frac{\partial p(h,\theta)}{\partial h} \leq 0 \quad \text{and} \quad \frac{\partial p(h,\theta)}{\partial \theta} \geq 0\,.$$

The education premium $R(h,\theta)y$ is is increasing in both human capital and academic ability, i.e.

$$\frac{\partial R(h,\theta)}{\partial h} \geq 0 \quad \text{and} \quad \frac{\partial R(h,\theta)}{\partial \theta} \geq 0\,.$$

There are no assets, so agents consume their full income in each period and get utility $u(y)$. This means that I make three implicit assumptions. First, the income profile for agents in the first period is steep enough so that they do not choose to save in order to smooth consumption. Second, there is no precautionary savings motive. While the first assumption is likely to be a result of individual optimisation in a model with assets, there will still be precautionary savings when using standard preferences. However, since precautionary savings are not key for the main model mechanism, I assume for simplicity that utility in each period is $u(y)$. The third assumption is that agents cannot borrow against their future income. While student debt is very common in the United States, it is much less prevalent in Switzerland, which is why I abstract from it.[11] The model is characterised recursively as follows:

$$
\begin{aligned}
V(\theta) &= \max\{V_w(\theta), V_e(\theta)\}, \\
V_w(\theta) &= u(y_1) + \beta \mathbb{E}[u(y_2)], \\
V_e(\theta) &= u(y_e) + \beta \mathbb{E}\left[V_e'(\theta)\right], \\
V_e'(\theta) &= \max_h \Big\{ p(h,\theta)\mathbb{E}\Big[u\Big(R(h,\theta)y_2\Big)\Big] + \Big(1 - p(h,\theta)\Big)\mathbb{E}\Big[u(y_2)\Big]\Big\}.
\end{aligned}
$$

The value $V(\theta)$ for an individual with academic ability θ is given by the maximum of the value when working $V_w(\theta)$ and the value when getting education $V_e(\theta)$. Choosing to work means that the agents gets income y_1 and stochastic income y_2 over the course of the two periods. Agents in education get income $y_e < y_1$. With probability $p(h,\theta)$ they get income $R(h,\theta)y_2$ and income y_2 otherwise.

The agent chooses the level of human capital h to maximise the value $V_e(\theta)$. Rearranging terms, the model can be reduced to three equations:

$$
\begin{aligned}
V(\theta) &= \max\{V_w(\theta), V_e(\theta)\}, \\
V_w(\theta) &= u(y_1) + \beta \mathbb{E}[u(y_2)], \\
V_e(\theta) &= u(y_e) + \beta \mathbb{E}[u(y_2)] + \max_h \Big\{ p(h,\theta)\beta\mathbb{E}\Big[u\Big(R(h,\theta)y_2\Big) - u(y_2)\Big]\Big\}.
\end{aligned}
$$

Now, the dynamic trade-off between working and getting tertiary education becomes apparent. To show this clearly, let h^* denote the level of human capital which maximises the expression in brackets. Thus, the value functions are given by:

$$
\begin{aligned}
V(\theta) &= \max\{V_w(\theta), V_e(\theta)\}, \\
V_w(\theta) &= u(y_1) + \beta \mathbb{E}[u(y_2)], \\
V_e(\theta) &= u(y_e) + \beta \mathbb{E}[u(y_2)] + p(h^*,\theta)\beta\mathbb{E}\Big[u\Big(R(h^*,\theta)y_2\Big) - u(y_2)\Big].
\end{aligned}
$$

Agents choose to get education if $V_e(\theta) > V_w(\theta)$, which holds if

$$
\underbrace{p(h^*,\theta)\beta\mathbb{E}\Big[u\Big(R(h^*,\theta)y_2\Big) - u(y_2)\Big]}_{\text{expected increase in utility}} > \underbrace{u(y_1) - u(y_e)}_{\text{foregone utility}}.
$$

[11] This constraint is implied by the natural borrowing limit if y_2 is stochastic and its lowest realisation is zero.

This illustrates the dynamic trade-off between foregone utility from lower income during education and the expected increase in utility from future education premia. Furthermore, it highlights the importance of modeling drop-out risk as the increase in expected utility is significantly affected by the success probability $p(h^*, \theta)$, which can be substantially below one as shown in Section 3. An additional static trade-off is illustrated by the maximisation problem that agents solve to find the optimal level of human capital h^*:

$$\max_{h} \quad p(h,\theta)\beta\mathbb{E}\Big[u\Big(R(h,\theta)y_2\Big) - u(y_2)\Big].$$

The first-order condition is

$$p(h,\theta)\mathbb{E}\left[u'\Big(R(h,\theta)y_2\Big)\frac{\partial R(h,\theta)}{\partial h}y_2\right] = -\frac{\partial p(h,\theta)}{\partial h}\mathbb{E}\Big[u\Big(R(h,\theta)y_2\Big) - u(y_2)\Big].$$

Agents trade off the marginal increase in the education premium when choosing a higher level of human capital and the marginal decrease in the success probability, which determines whether the tertiary degree can be attained and the education premium realised. The dynamic and static trade-off are the core features of the model I present in the next section.

5 Model

I begin by describing a model for tertiary education in Section 5.1, which will include the dynamic and static trade-off. In Section 5.2, I show the key insights for the education choices along the ability distribution. Section 5.3 then introduces the phase of upper-secondary education to highlight a mechanism for controlling admission to tertiary education.

5.1 Tertiary education

The economy is composed of a continuum of agents with an initial level of human capital $h = s$ corresponding to an upper-secondary degree (VET or *Matura*). Agents differ ex ante by their academic ability θ. They live for two periods and choose between working and getting tertiary education. Agents deciding to get tertiary education can choose between going to university or university of applied sciences (FH).

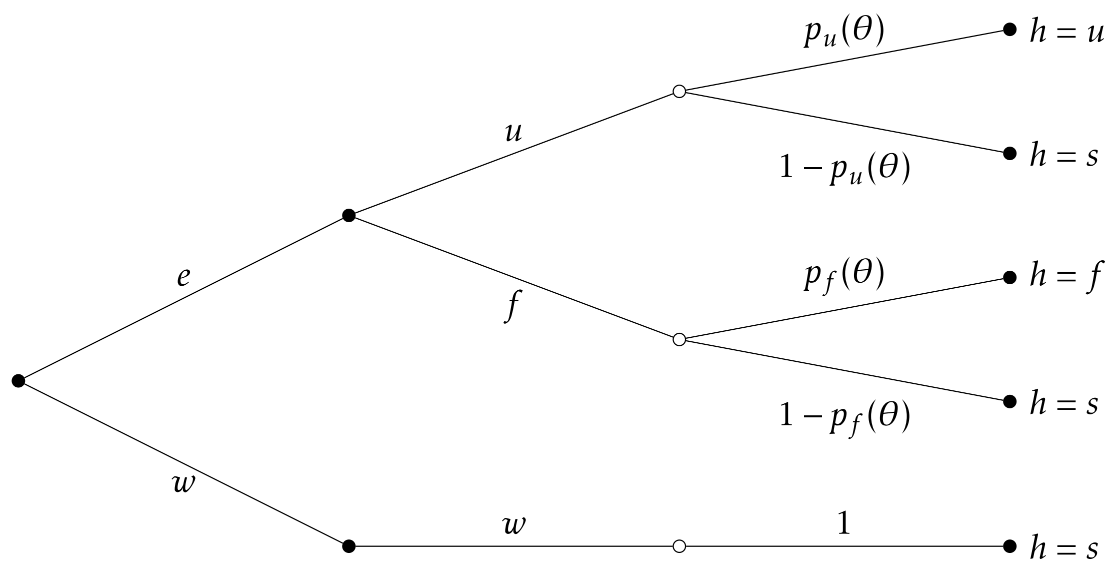


Figure 5.1: Decision tree for the dynamic and static trade-off.

In the first period, agents who decide to get tertiary education will attain a degree with probability $p_e(\theta)$, with $e \in \{f, u\}$, where $p_u(\theta)$ and $p_f(\theta)$ are the success probabilities for university and university of applied sciences (FH). With probability $1 - p_e(\theta)$ they drop out. The level of human capital for agents who drop out is $h = s$. For agents who attain tertiary degree, it is $h = u$ and $h = f$ for graduates of the respective tertiary institution.

Figure 5.1 depicts the decision tree. The black nodes represent points where the agents choose between options. The white nodes represent points where nature chooses based on a given probability distribution. First, the agents choose between work (w) and education (e). This choice features the dynamic trade-off between foregone earnings and expected education premia. Agents choosing education then choose between going to university (u) or university of applied sciences (f), from which the static trade-off between the drop-out probabilities and the expected education premia arises. The resulting levels for human capital are $h \in \{s, f, u\}$, corresponding to an upper-secondary degree and a degree from university of applied sciences and university.

While studying, agents get a subsistence income y_e. Agents choosing to work get income y_1. Both y_e and y_1 are deterministic. In the second period, agents get income $R_h y_2$, with $h \in \{s, f, u\}$ and R_s normalised to $R_s = 1$. Income y_2 is lognormally distributed, i.e. $\log y_2 \sim \mathcal{N}(\mu_y, \sigma_y^2)$. As described in the previous section, there are no assets in the economy, which means agents get utility from income $u(y)$ in each period. The model is characterised recursively as follows:

$$
\begin{aligned}
V(\theta) &= \max\{V_w(\theta), V_e(\theta)\}, \\
\\
V_w(\theta) &= u(y_1) + \beta\mathbb{E}[u(y_2)], \\
V_e(\theta) &= \max\Big\{V_f(\theta), V_u(\theta)\Big\}, \\
\\
V_f(\theta) &= u(y_e) + \beta\mathbb{E}\Big[p_f(\theta)u(R_f y_2) + \Big(1 - p_f(\theta)\Big)u(y_2)\Big], \\
V_u(\theta) &= u(y_e) + \beta\mathbb{E}\Big[p_u(\theta)u(R_u y_2) + \Big(1 - p_u(\theta)\Big)u(y_2)\Big].
\end{aligned}
$$

The central assumption is that university offers a higher education premium, i.e. $R_u > R_f$ but has a lower success probability, i.e. $p_u(\theta) < p_f(\theta)$, for a given level of θ. The success probabilities are characterised as defined in the following way:

Definition 5.1. *The success probabilities $p_f(\theta)$ and $p_u(\theta)$ are characterised by the following three properties:*

(CDF) *The success probabilities $p_f(\theta)$ and $p_u(\theta)$ are both between zero and one and increasing in θ. Thus, they can be characterised by two cumulative distribution functions $F_f(\theta)$ and $F_u(\theta)$.*

(FOSD) *The success probability for university is lower than the success probability for FH for all levels of θ, i.e. $F_u(\theta) \leq F_f(\theta)\ \forall\theta$. Hence, $F_u(\theta)$ first-order stochastically dominates $F_f(\theta)$.*

(Regularity condition) *We require*

$$
\frac{\partial}{\partial\theta}\frac{p_u(\theta)}{p_f(\theta)} > 0
$$

to ensure monotonicity in the choice of human capital, i.e. for two agents A and B with $\theta_A > \theta_B$, it cannot be optimal for A to choose FH, while B optimally chooses university. Applying the quotient rule, this implies

$$
\frac{p_u'(\theta)}{p_u(\theta)} > \frac{p_f'(\theta)}{p_f(\theta)} \Longleftrightarrow \frac{f_u(\theta)}{F_u(\theta)} > \frac{f_f(\theta)}{F_f(\theta)},
$$

where $f(\theta)$ is the probability density function. Hence, when increasing θ, $p_u(\theta)$ grows at a higher rate than $p_f(\theta)$.

The dynamic and static trade-off give rise to two optimality conditions. Agents choose to get tertiary education if $V_e(\theta) > V_w(\theta)$, which holds if

$$\underbrace{p_e^*(\theta)\beta\mathbb{E}\left[u(R_e^* y_2) - u(y_2)\right]}_{\text{increase in expected utility}} > \underbrace{u(y_1) - u(y_e)}_{\text{foregone utility}},$$

where I denote $p_e^*(\theta)$ and R_e^* the success probability and education return for the optimal education choice, which is governed by the static trade-off. Agents prefer university over university of applied sciences (FH) if $V_u(\theta) > V_f(\theta)$, which holds if

$$\underbrace{p_u(\theta)\mathbb{E}\left[u(R_u y_2) - u(y_2)\right]}_{\text{increase in expected utility } (u)} > \underbrace{p_f(\theta)\mathbb{E}\Big[u(R_f y_2) - u(y_2)\Big]}_{\text{increase in expected utility } (f)}.$$

This can be rewritten as

$$\underbrace{p_u(\theta)\mathbb{E}\Big[u(R_u y_2) - u(R_f y_2)\Big]}_{\text{increase in expected utility } (u \text{ vs. } f)} > \underbrace{\Big(p_f(\theta) - p_u(\theta)\Big)}_{\text{change in success probability}} \underbrace{\mathbb{E}\Big[u(R_f y_2) - u(y_2)\Big]}_{\text{increase in expected utility } (f \text{ vs. } s)},$$

which highlights the trade-off between the higher education premium for university on the left-hand side and the lower success probability at university on the right-hand side.

5.2 Education choices along the ability distribution

Given the conditions describing the success probabilities and the optimality conditions in the previous subsection, the education choices along the ability distribution are governed by three critical values for ability θ. Let θ_{uw}^* denote the level of θ, where agents are indifferent between studying at university and working, i.e. $V_u(\theta_{uw}^*) = V_w(\theta_{uw}^*)$ and define θ_{fw}^* and θ_{uf}^* accordingly.

These values are characterised by the following conditions:

$$p_u(\theta^*_{uw}) = \frac{u(y_1) - u(y_e)}{\beta \mathbb{E}\left[u(R_u y_2) - u(y_2)\right]},$$

$$p_f(\theta^*_{fw}) = \frac{u(y_1) - u(y_e)}{\beta \mathbb{E}\Big[u(R_f y_2) - u(y_2)\Big]},$$

$$\frac{p_u(\theta^*_{uf})}{p_f(\theta^*_{uf})} = \frac{\mathbb{E}\Big[u(R_f y_2) - u(y_2)\Big]}{\mathbb{E}\Big[u(R_u y_2) - u(y_2)\Big]}.$$

The following results show that these critical values split the ability distribution. Furthermore, they are unique.

Proposition 5.1. *Let θ^*_{uw} be the level of θ where the agent is indifferent between working and studying at university. Then, for all values $\theta > \theta^*_{uw}$, studying at university is preferred over working.*

Proof. See Appendix C. □

Corollary 5.1.1. *By the same argument, it follows that for all values $\theta < \theta^*_{uw}$, choosing university is dominated and thus, θ^*_{uw} is unique.*

Corollary 5.1.2. *Using the same line of reasoning, Proposition 5.1 and Corollary 5.1.1 also hold for θ^*_{fw} and thus, θ^*_{fw} is unique.*

Proposition 5.2. *Let θ^*_{uf} be the level of θ where the agent is indifferent between the two education choices. Then, for all values $\theta > \theta^*_{uf}$, studying at university is preferred over studying at a university of applied sciences (FH).*

Proof. See Appendix C. □

Corollary 5.2.1. *By the same argument, it follows that for all values $\theta < \theta^*_{uf}$, choosing university is dominated and thus, θ^*_{uf} is unique.*

Although the three critical values fully describe the education choices along the ability distribution, it is their order which defines how partitions of the ability

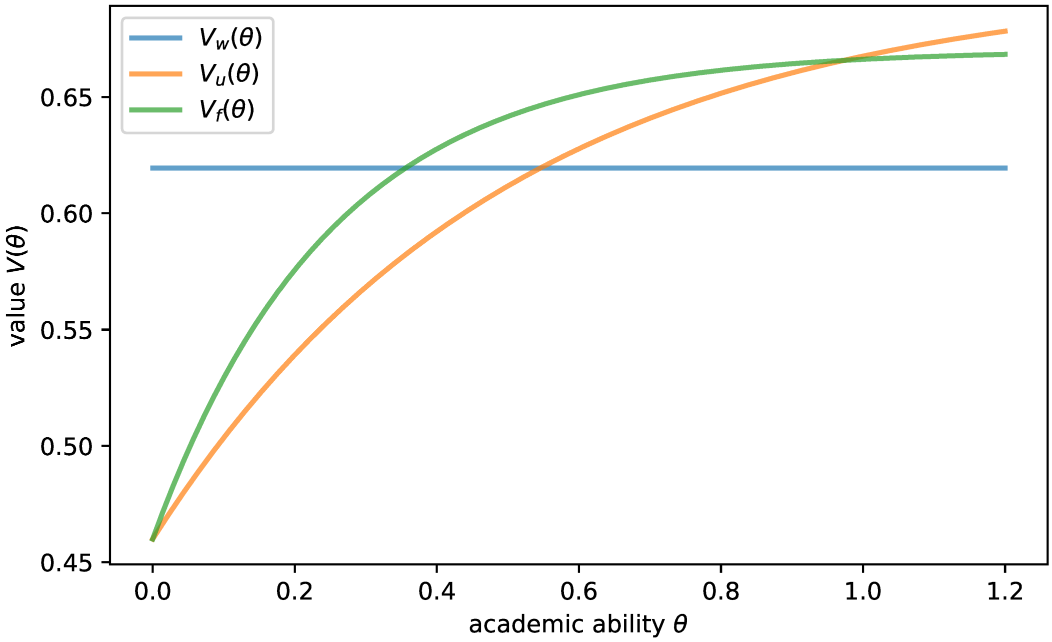


Figure 5.2: Value functions for work, university of applied sciences (FH) and university.

distribution are mapped into education choices. The economically interesting case is given by $\theta^*_{fw} < \theta^*_{uw} < \theta^*_{uf}$.[12]

Figure 5.2 depicts the value functions for this case. The value function for work $V_w(\theta)$ is constant in θ, as there is no drop-out risk.[13] The value $V_f(\theta)$ is higher than the value $V_u(\theta)$ at the lower end of the ability distribution. For these agents, the higher success probability at a university of applied sciences (FH) dominates the lower education premium resulting from it. At the upper end of the ability distribution, the higher education premium resulting from a university degree dominates the fact that agents face a higher drop-out risk and thus $V_u(\theta)$ is larger than $V_f(\theta)$. As a result, the distribution of academic ability θ is split into three parts: agents with $\theta < \theta^*_{fw}$ choose to work, those with

[12] I discuss alternative orderings in Appendix A. As I show, only three cases are possible of which the case discussed here is the only case where at least some students attend a university of applied sciences (FH).

[13] This is because income y_2 is independent of academic ability θ. If income were increasing in academic ability θ, this would imply an increasing value function for work $V_w(\theta)$, but also steeper value functions $V_f(\theta)$ and $V_u(\theta)$. Such a change would merely tilt the value functions in the same direction and would thus not change the result qualitatively.

$\theta^*_{fw} < \theta < \theta^*_{uf}$ go to a university of applied science (FH) and agents with $\theta > \theta^*_{uf}$ go to university.

Prior to the introduction of the university of applied sciences, the ability distribution was simply split into two parts, with agents having ability $\theta < \theta^*_{uw}$ working and those with $\theta > \theta^*_{uw}$ attending university. A closer look at the agents choosing to go to a university of applied sciences (FH) reveals what the addition of a static trade-off by introducing another option changes. These agents can be split into two groups. On the one hand, there are the agents with $\theta^*_{fw} < \theta < \theta^*_{uw}$, i.e. $V_f(\theta) > V_w(\theta) > V_u(\theta)$. This group would prefer working over studying in an environment without the option to go to FH. However, among the working agents, they are the ones with the highest level of ability θ. Thus, they are the marginal agents, for whom university is too risky and FH offers a less risky option to get a tertiary degree. On the other hand, there are those for which $\theta^*_{uw} < \theta < \theta^*_{uf}$, i.e. $V_f(\theta) > V_u(\theta) > V_w(\theta)$. This group would prefer to study at university over working in an environment without the option to go to FH. Yet, among the university students, they are those with the lowest level of academic ability and thus face the highest drop-out risk. Hence, they are the marginal agents, for whom the risk of going to university would be just low enough to choose it, but now FH offers a less risky choice. Thus, introducing the possibility to go to a university of applied sciences, adds a convex option to the lottery of choosing the right education path. This possibility will be taken by those with the highest level of academic ability among the working agents and those with the lowest level of academic ability among those agents going to university. As one can see from Figure 5.2, the introduction of another option thus improves sorting into education by yielding a Pareto improvement.

5.3 Introducing upper-secondary education

I extend the model and introduce two education choices on the upper-secondary level: vocational education and training (VET) and *Gymnasium*. This offers a possibility to control access to university within the model. Agents who successfully graduate from *Gymnasium* get a *Matura*, which is needed to enrol in a university degree programme.

As before, the economy is composed of a continuum of agents who differ ex ante by their academic ability θ. To model upper-secondary education, I add a

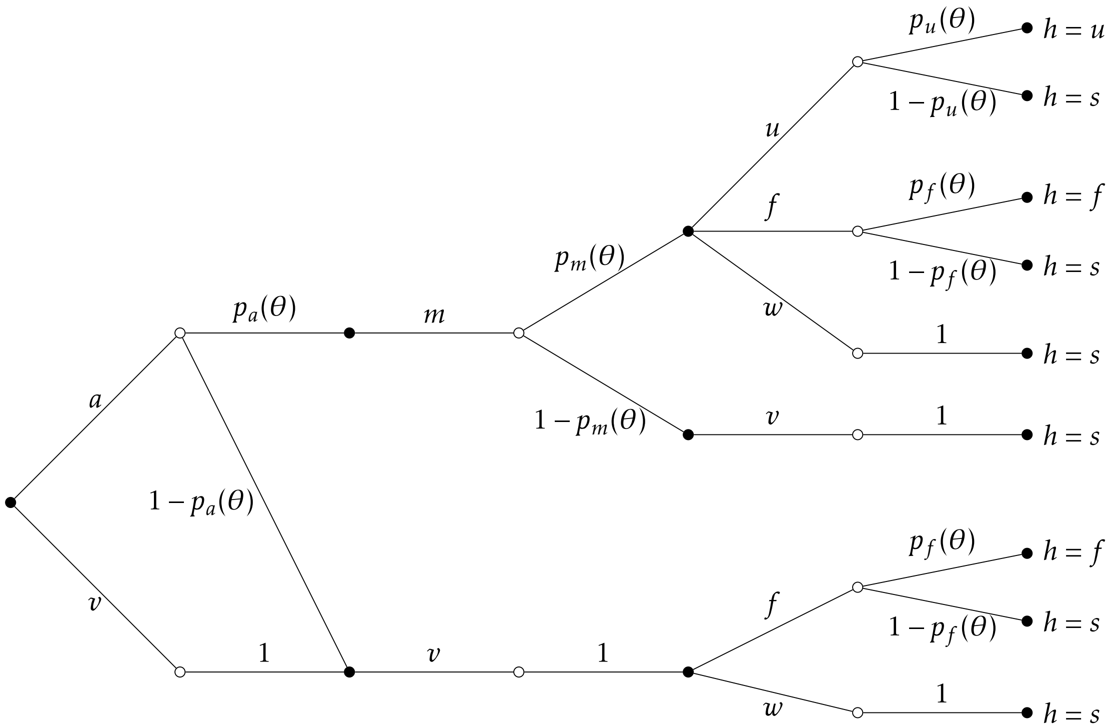


Figure 5.3: Decision tree including upper-secondary education.

third period prior to tertiary education. Agents enter the model with an initial level of human capital $h = ls$, corresponding to a lower secondary degree.

In the first period, agents decide to go to *Gymnasium* or get vocational education and training (VET). In this period, all agents get subsistence income y_e. At the end of the period, agents in VET get human capital $h = s$, corresponding to an upper-secondary degree. Agents who go to *Gymnasium* successfully graduate and get a *Matura* with probability $p_m(\theta)$ and their level of human capital is $h = s$. With probability $1 - p_m(\theta)$ they drop out and their level of human capital remains at $h = ls$. In the second period, agents with a *Matura* get to choose between working, studying at a university of applied sciences (FH), and university. Thus, their decision problem is identical to the one described in the previous part. Agents who got vocational education and training (VET) are not eligible to go to university, but can still choose to either work or study at a university of applied sciences (FH). Finally, agents who dropped out of *Gymnasium* get vocational education and training in the second period and attain human capital $h = s$. As before, agents work in the third period and earn income according to their level of human capital.

In this extended model, the number of university graduates is controlled by two factors: admission to university and success at university. Admission is given by the requirement for students to have obtained a *Matura* after they went to *Gymnasium*, and is thus governed by $p_m(\theta)$. Success at university is governed by $p_u(\theta)$. In a final step, I extend the model to control access to Gymnasium. I do this by introduction an admission exam to *Gymnasium*. All agents choosing to go to *Gymnasium* first have to take an admission exam, which they pass with probability $p_a(\theta)$. If they pass the exam, they will go to *Gymnasium* and if not, they are getting vocational education and training. Note, that once agents have passed the admission exam, they cannot choose to get VET but have to go to *Gymnasium* which entails the risk of failure. This ensures that agents do not take the admissions exam without having to face any downside risk.

Figure 5.3 depicts the decision tree. As before, the black nodes represent points where the agents choose between options and the white nodes represent points where nature chooses based on a given probability distribution. The model is characterised recursively as follows:

$$
\begin{aligned}
V(\theta) &= \max\{V_v(\theta), V_a(\theta)\}, \\
\\
V_v(\theta) &= u(y_e) + \beta\mathbb{E}\left[V_v'(\theta)\right], \\
V_v'(\theta) &= \max\Big\{V_w(\theta), V_f(\theta)\Big\}, \\
\\
V_a(\theta) &= p_a(\theta)V_m(\theta) + \Big(1 - p_a(\theta)\Big)V_v(\theta), \\
V_m(\theta) &= u(y_e) + \beta\mathbb{E}\Big[p_m(\theta)V_m'(\theta) + \Big(1 - p_m(\theta)\Big)V_{ls}(\theta)\Big], \\
V_m'(\theta) &= \max\Big\{V_w(\theta), V_f(\theta), V_u(\theta)\Big\}, \\
\\
V_{ls}(\theta) &= u(y_e) + \beta\mathbb{E}\left[u(y_3)\right], \\
V_w(\theta) &= u(y_2) + \beta\mathbb{E}\left[u(y_3)\right], \\
V_f(\theta) &= u(y_e) + \beta\mathbb{E}\Big[p_f(\theta)u(R_f y_3) + \Big(1 - p_f(\theta)\Big)u(y_3)\Big], \\
V_u(\theta) &= u(y_e) + \beta\mathbb{E}\Big[p_u(\theta)u(R_u y_3) + \Big(1 - p_u(\theta)\Big)u(y_3)\Big].
\end{aligned}
$$

At this point, it is possible to control both access and success at the *Gymnasium* and the university level through the probabilities $p_a(\theta)$, $p_m(\theta)$ and $p_u(\theta)$.

Agents now have various choices, but some choices are dominated by others. Specifically, it is never optimal to go to *Gymnasium* and study at a university of applied sciences afterwards. Intuitively, if an agent would like to study at a university of applied sciences, then going through vocational education and training (VET) is the better option. Similarly, it is never optional to go to *Gymnasium* and work afterwards. All the corresponding optimality conditions are discussed in Appendix B.

Although the model predicts that no agent with a *Matura* goes to a university of applied sciences (FH), around 10 percent of them do so in Switzerland.[14] On the one hand, some students prefer the more practical approach offered at a university of applied sciences (FH). On the other hand, some students only realise during *Gymnasium* that their academic ability is not as high as they thought it would be. These are aspects I do not model. To resolve this discrepancy, one could modify the model by either introducing preference shocks or by relaxing the assumption that agents perfectly observe their own ability type. However, since my model mechanism does not hinge on these particular points, I abstract from them.

The agents thus choose between three options: going to *Gymnasium* with the intent to study at university, getting VET with the intent to study at a university of applied sciences (FH) or getting VET and work afterwards. The value functions for these three options yield a very similar result as shown in Figure 5.2 on page 99.[15] Introducing the option to go to a university of applied sciences (FH) offers a better alternative for those with the highest level of academic ability among the working agents and those with the lowest level of academic ability among those agents going to university. Although this result is the same as in the model without upper-secondary education, the new model now allows to shed light on the cross-sectional variation in enrolment rates, *Matura* rates, and drop-out rates between cantons. This analysis is done in Section 6.3. Furthermore, the inclusion of upper-secondary education into the model allows to analyse how the education choices along the ability distribution are affected by $p_a(\theta)$ and $p_m(\theta)$.

[14]See BFS (2018) for details.
[15]See Appendix B for details.

As in the previous part, these choices are governed by the critical values θ^*_{uw}, θ^*_{fw} and θ^*_{uf}, which are characterised by[16]

$$p_m(\theta^*_{uw})p_u(\theta^*_{uw}) = \frac{u(y_2) - u(y_e)}{\beta\mathbb{E}[u(R_u y_3) - u(y_3)]},$$

$$p_f(\theta^*_{fw}) = \frac{u(y_2) - u(y_e)}{\beta\mathbb{E}\Big[u(R_f y_3) - u(y_3)\Big]},$$

$$\frac{p_m(\theta^*_{uf})p_u(\theta^*_{uf})}{p_f(\theta^*_{uf})} = \frac{\mathbb{E}\Big[u(R_f y_3) - u(y_3)\Big]}{\mathbb{E}\Big[u(R_u y_3) - u(y_3)\Big]}.$$

The critical values are very similar to those in Section 5.2, but include the success probability $p_m(\theta)$. However, note that the critical values are independent of the admission probability $p_a(\theta)$. This implies, that when choosing whether or not to go to *Gymnasium*, agents internalise how difficult it is to successfully get a *Matura*, but their choice is unaffected by how hard it is to get admitted to *Gymnasium*. Although this seems counterintuitive at first, it becomes clear when looking at the special case of $p_a(\theta) = 1$ for all levels of θ. In this case, the critical values are solely dependent on the success probabilities $p_m(\theta)$, $p_f(\theta)$ and $p_u(\theta)$. Agents with academic ability $\theta < \theta^*_{uf}$ choose to get vocational education and training (VET), and those with $\theta \geq \theta^*_{uf}$ go to *Gymnasium*, to which they get admitted with probability one. Introducing an admission exam, and thus $p_a(\theta) \leq 1$, does not affect the decision of those choosing to get VET, as the admission exam does not make *Gymnasium* any more attractive. Furthermore, those agents choosing to go to *Gymnasium* will also not choose differently either, since trying to pass the admission exam does not come at any cost. Thus, the education choices are independent of $p_a(\theta)$. Yet, the level of difficulty to get admitted to *Gymnasium* still affects the effective enrolment rates and thus also the average drop-out risk of those agents who attend *Gymnasium*. Having laid out the model and its key features, I will now turn to the analysis of its key implications.

[16] See Appendix B for a derivation of these values.

6 Key implications of the model

6.1 Additional choice improves sorting through risk reduction

In this section, I analyse the implication of adding an additional choice to an education system with a binary choice between working and studying. This alternative offers lower drop-out risk, but comes at the cost of a lower education premium. For this analysis, I return to the simpler model with just three choices: working, studying at university and the additional choice of studying at a university of applied sciences (FH). Recall that the model is given by the following value functions:

$$
\begin{aligned}
V(\theta) \quad &= \max\Big\{V_w(\theta), V_f(\theta), V_u(\theta)\Big\}\,, \\
V_w(\theta) \quad &= u(y_1) + \beta\mathbb{E}\left[u(y_2)\right]\,, \\
V_f(\theta) \quad &= u(y_e) + \beta\mathbb{E}\Big[p_f(\theta)u(R_f y_2) + \big(1 - p_f(\theta)\big)u(y_2)\Big]\,, \\
V_u(\theta) \quad &= u(y_e) + \beta\mathbb{E}\Big[p_u(\theta)u(R_u y_2) + \big(1 - p_u(\theta)\big)u(y_2)\Big]\,.
\end{aligned}
$$

For the analysis, I compare the outcomes of the model with the planner's solution. Assume the planner perfectly observes each individual θ and offers perfect insurance against drop-out risk. The value functions for the planner are given by

$$
\begin{aligned}
V(\theta) \quad &= \max\Big\{V_w(\theta), V_f(\theta), V_u(\theta)\Big\}\,, \\
V_w^p(\theta) \quad &= u(y_1) + \beta\mathbb{E}\left[u(y_2)\right]\,, \\
V_f^p(\theta) \quad &= u(y_e) + \beta\mathbb{E}\Big[u\Big(p_f(\theta)R_f y_2 + \big(1 - p_f(\theta)\big)y_2\Big)\Big]\,, \\
V_u^p(\theta) \quad &= u(y_e) + \beta\mathbb{E}\Big[u\Big(p_u(\theta)R_u y_2 + \big(1 - p_u(\theta)\big)y_2\Big)\Big]\,.
\end{aligned}
$$

Note that the planner insures the agents against the risk of dropping out, but not against any income risk so that the expectation operator appears outside the

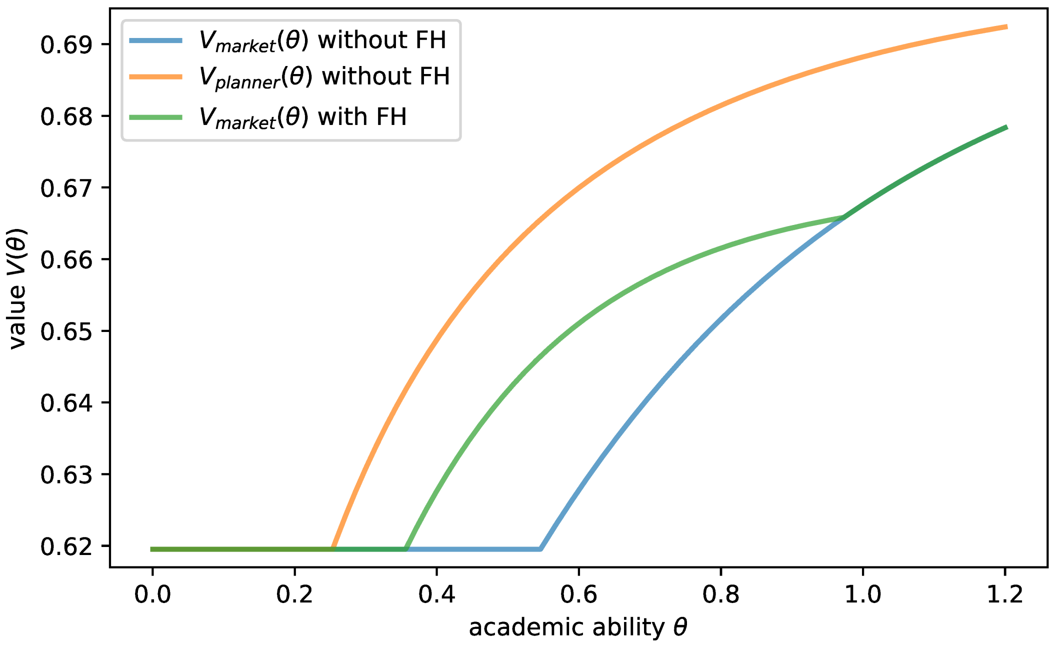


Figure 6.1: Market and planner solutions.

utility function. Denoting $r = R - 1$, the value functions simplify to

$$\begin{aligned}
V(\theta) &= \max\left\{V_w(\theta), V_f(\theta), V_u(\theta)\right\}, \\
V_w^p(\theta) &= u(y_1) + \beta\mathbb{E}\left[u(y_2)\right], \\
V_f^p(\theta) &= u(y_e) + \beta\mathbb{E}\Big[u\Big(\Big(1 + p_f(\theta)r_f\Big)y_2\Big)\Big], \\
V_u^p(\theta) &= u(y_e) + \beta\mathbb{E}\Big[u\Big(\Big(1 + p_u(\theta)r_u\Big)y_2\Big)\Big].
\end{aligned}$$

The agents obtain utility from the expected return $1 + p(\theta)r$ in the planner's solution, instead of the expected utility across different states with different return realisations. Thus, by the concavity of the utility function and Jensen's inequality, the planner improves the outcomes of studying risk-averse agents. Note however, that the utility for those choosing to work remains unchanged as they do not face any drop-out risk and hence do not profit from any insurance offered.

Figure 6.1 illustrates the market and planner's solution in an education system without the additional option of studying at a university of applied sciences (FH). The figure depicts the envelope of the value function for work, $V_w(\theta)$, and the value function for going to university, $V_u(\theta)$, i.e. it depicts $\max\{V_w(\theta), V_u(\theta)\}$. The kink in the function is where $V_w(\theta)$ and $V_u(\theta)$ intersect and is thus at θ^*_{uw}. This is the critical level of θ where the choice between working and going to university changes.

One can see that this critical value is much lower for the planner's solution, as with full insurance against drop-out risk, more agents would choose to go to university. Thus, in an economy with drop-out risk, less agents choose to go to university than would be desirable. This is where the introduction of a third option improves sorting into education. The third value function depicted in Figure 6.1 is the market solution in an education system with the option of studying at a university of applied sciences (FH), i.e. $V_f(\theta)$. The first kink is at the value of θ, where the agents are indifferent between working and going to FH. The second kink is where the agents are indifferent between going to FH and going to university.

The figure shows that the introduction of the university of applied science (FH) is a Pareto improvement as it makes no one worse of, but offers a choice with a better risk-return combination for those who choose to study at a university of applied sciences (FH). Furthermore, it shows that without the intervention of a planner, the number of agents enrolling in a tertiary degree programme increases, thus bringing the education system closer to the planner's solution. Admittedly, once the university of applied sciences (FH) is introduced into the education system, the planner's solution also improves the outcome, as shown in Figure 6.2. Yet, since the FH drop-out rates are lower than the university drop-out rates, there is less room for the planner to improve.

This example illustrates the role of the education system. Offering alternatives with better risk-return profiles can improve sorting into education without the need for a planner to possess information on the agents' academic ability. The introduction of an alternative education choice can be compared to offering partial insurance for the otherwise uninsurable drop-out risk. Agents forego a part of their education premium, which is the analogue of paying an insurance premium. In return, their risk of dropping out is reduced. This improves welfare as illustrated by the smaller distance between the first best and the decentralized implementation.

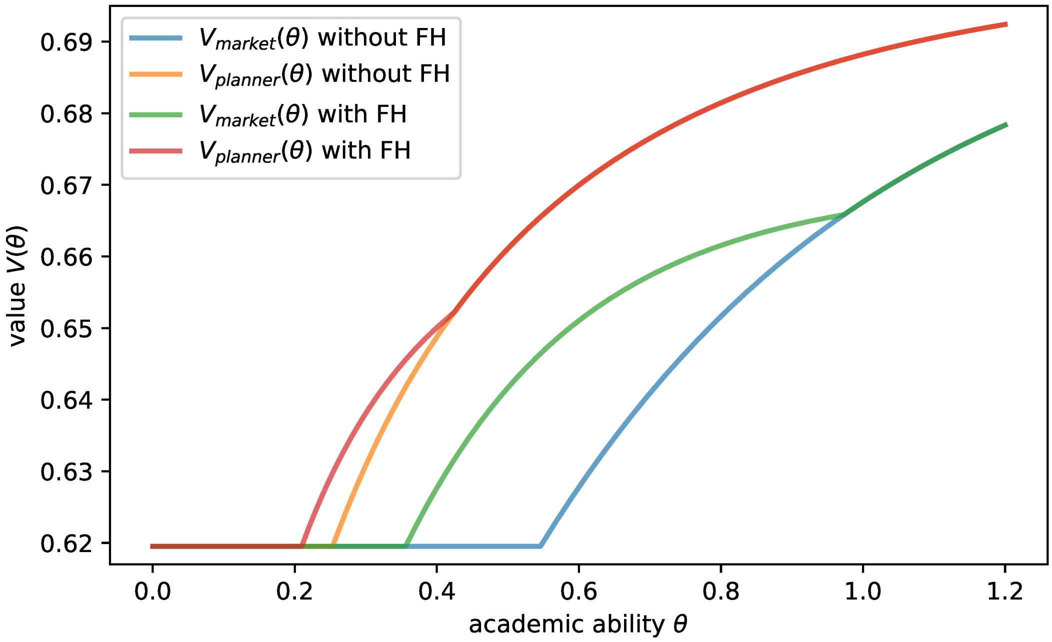


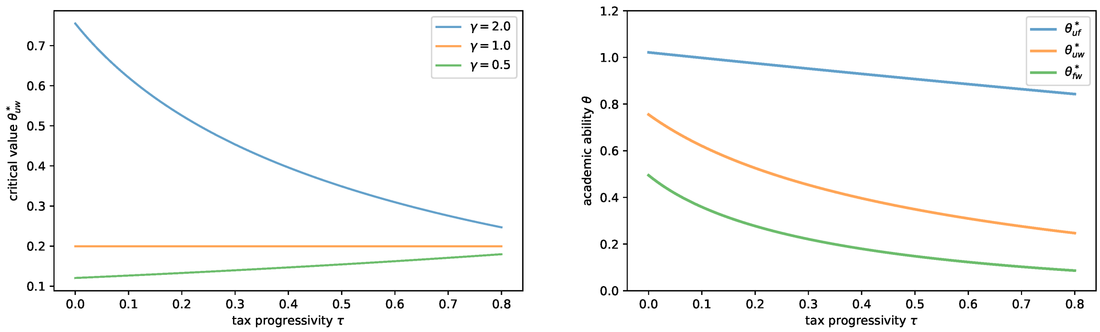
Figure 6.2: Market and planner solutions with FH.

6.2 Progressive taxation improves sorting through risk reduction

I now turn to the effect of income taxation on the education choices along the ability distribution. I use the tax function described in Heathcote et al. (2017), which exhibits constant progressivity and is given by

$$T(y) = y - \lambda y^{1-\tau},$$

so that after-tax income is given by $\lambda y^{1-\tau}$. It is characterised by two parameters, λ and τ, determining the tax level and the tax progressivity. As a special case, the function contains linear taxes when $\tau = 0$. In that case, the linear tax rate is given by $1 - \lambda$. The education choices are characterised by the three critical values θ^*_{uw}, θ^*_{fw} and θ^*_{uf} described in Section 5.3.



(a) Progressive taxation without FH. (b) Progressive taxation with FH ($\gamma = 2$).

Figure 6.3: Progressive taxation.

Introducing taxes and assuming CRRA preferences with risk aversion γ, these critical values become

$$p_m(\theta^*_{uw})p_u(\theta^*_{uw}) = \frac{y_2^{1-\tilde{\gamma}} - y_e^{1-\tilde{\gamma}}}{\beta\mathbb{E}\Big[[R_u y_3]^{1-\tilde{\gamma}} - y_3^{1-\tilde{\gamma}}\Big]},$$

$$p_f(\theta^*_{fw}) = \frac{y_2^{1-\tilde{\gamma}} - y_e^{1-\tilde{\gamma}}}{\beta\mathbb{E}\Big[[R_f y_3]^{1-\tilde{\gamma}} - y_3^{1-\tilde{\gamma}}\Big]},$$

$$\frac{p_m(\theta^*_{uf})p_u(\theta^*_{uf})}{p_f(\theta^*_{uf})} = \frac{\mathbb{E}\Big[[R_f y_3]^{1-\tilde{\gamma}} - y_3^{1-\tilde{\gamma}}\Big]}{\mathbb{E}\Big[[R_u y_3]^{1-\tilde{\gamma}} - y_3^{1-\tilde{\gamma}}\Big]},$$

where I define $\tilde{\gamma} = \gamma + \tau(1-\gamma)$ to be the tax-modified RRA coefficient. First, note that the linear parameter λ cancels out and with $\tau = 0$, we get $\tilde{\gamma} = \gamma$. Thus, linear taxes have no effect on the education choices. However, with progressive taxes, i.e. $\tau > 0$, a change in progressivity has the same as the effect as changing the parameter for risk aversion γ. The change of $\tilde{\gamma}$ when changing τ is given by

$$\frac{\partial\tilde{\gamma}}{\partial\tau} = 1 - \gamma .$$

This implies that with $\gamma = 1$, i.e. with log-preferences, even progressive taxes have no effect on the education choices along the ability distribution. However, with $\gamma > 1$, an increase in the tax progressivity implies a decrease in $\tilde{\gamma}$ and thus, this is equivalent to agents being less risk averse and willing to take higher risk. With $\gamma < 1$, the opposite holds.

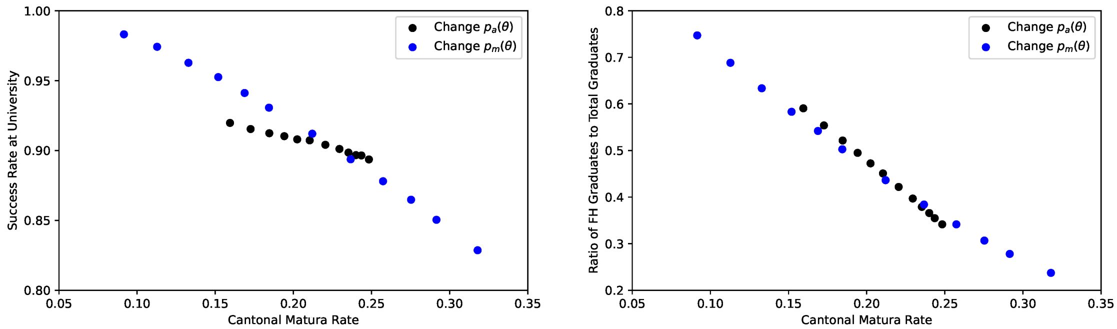
This result is illustrated in Panel (a) of Figure 6.3, where I only look at the simple model with no FH choice. The figure depicts how the critical value θ^*_{uw} changes when taxes get more progressive using three different values for γ: 0.5, 1.0, and 2.0. Panel (b) of Figure 6.3 shows how the critical values react to an increase in tax progressivity for $\gamma = 2$, which is commonly used in the literature. A value of $\gamma = 2$ implies that an increase in tax-progressivity actually yields an *increase* in enrolment rates in tertiary education.

This is somewhat counterintuitive and qualifies the *Siamese twin* result by Bovenberg and Jacobs (2005), who argue that progressive taxation *decreases* investment in human capital as it diminishes the return to education. In an environment with drop-out and income risk, progressive taxation has a second effect: it decreases the after-tax income variance. Hence, there are two counteracting effects. On the one hand, progressive taxation decreases the expected return to education which disincentivises investment into education. On the other hand, it also decreases its variance and thus makes going to university less risky. The net effect depends on the level of risk aversion and one can see in Figure 6.3 that for $\gamma > 1$ the net effect yields higher enrolment rates in tertiary education. This finding is similar to the income and substitution effect in standard economic theory and is in line with a set of results in the literature on decisions under uncertainty and portfolio theory. I refer the reader to Gollier (2001) for an in depth treatment of the topic.

In light of the planner's solution discussed in the previous subsection, I conclude that, contrary to the standard *Siamese twin* result, progressive taxation improves sorting into education through risk reduction for quantitatively plausible values of risk aversion.

6.3 Heterogeneity in admission probabilities drives heterogeneity in enrolment rates

I now return to the stylised facts for Switzerland shown in Section 3 and look at them through the lens of the model. Recall the three empirical relationships uncovered in the data: there is a negative correlation between cantonal enrolment rates and success rates at the *Gymnasium* level, a negative correlation between cantonal *Matura* rates and success rates at university and a negative correlation



(a) University success rates. (b) Fraction of FH graduates.

Figure 6.4: Simulated cantons.

between cantonal *Matura* rates and the share of graduates from universities of applied sciences (FH) among all graduates from tertiary education.

As a numerical illustration, I simulate the model for different cantons to see how changes in success probabilities affect the cross-sectional variation of enrolment rates, attainment rates, and drop-out rates between cantons.[17] I hereby assume, that cantons can set both the strictness of admission to *Gymnasium* $p_a(\theta)$ and the *Matura* exams $p_m(\theta)$, but they do not have control over the success probabilities at the tertiary level, i.e. $p_f(\theta)$ and $p_u(\theta)$.

I perform two experiments. First, I simulate cantons which are identical in all dimensions but differ in the strictness of admission to *Gymnasium*, i.e. I change $p_a(\theta)$. Second, I vary the strictness of the *Matura* exams between cantons, i.e. I change $p_m(\theta)$, while keeping all other dimensions fixed. I then generate the same graphs as in Section 3 with the simulated data to see whether the model can reproduce the empirical relationships observed in the data.

Figure 6.4 depicts the simulated relationships between *Matura* rates and university success rates in Panel (a) and between *Matura* rates and the fraction of FH graduates in Panel (b). These relationships are negative both in the data and the simulation results and hence, heterogeneity in the strictness in admission to *Gymnasium* and in the difficulty of the *Matura* exams might both explain the stylised facts.

It is worth noting how the two experiments differ. As explained before, changes in $p_a(\theta)$ do not affect the agents' choices, whereas changes in $p_m(\theta)$ do. This means that in the experiment where $p_a(\theta)$ is varied, the same amount of agents

[17] See Appendix D for a description of the functional form assumptions and parameter choices.

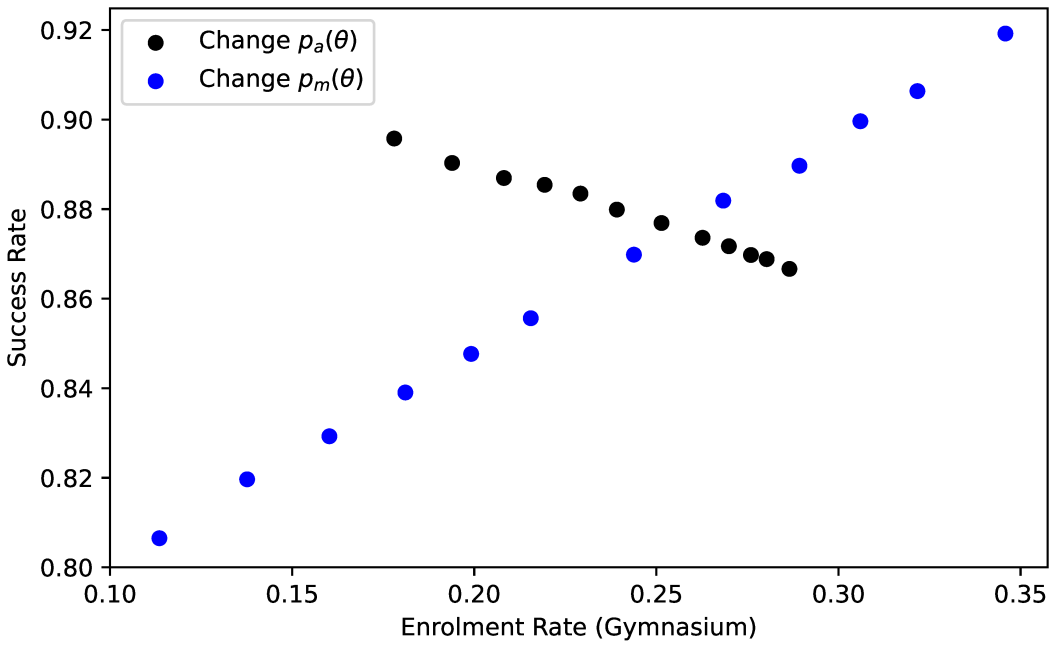


Figure 6.5: Enrolment and success rates (*Gymnasium*).

have chosen to go to *Gymnasium*, but in cantons with stricter admission to *Gymnasium*, less of them have been admitted. Thus, the average agent in these cantons has a higher level of academic ability, which leads to higher success rates at university. In contrast, in the experiment where I vary $p_m(\theta)$, more agents choose to go to *Gymnasium* in cantons, where it is less difficult to pass the *Matura* exams. Again, in cantons with more difficult *Matura* exams, the average agent has a higher level of academic ability leading to higher success rates at university.

In Figure 6.5, I depict the relationship between enrolment and success rates at the *Gymnasium* level, as generated by the simulated data. When varying $p_a(\theta)$, a negative relationship arises, as in cantons with stricter admission to *Gymnasium*, those who get admitted are more likely to successfully pass the *Matura* exams. This is in line with the stylised facts.

On the other hand, in the experiment where I change $p_m(\theta)$, more agents will go to *Gymnasium* in cantons with less difficult *Matura* exams leading to higher enrolment rates in those cantons. And because the exams are less difficult, the success rates will also be higher, yielding a positive relationship between enrolment and success rates. This, however, is counterfactual to the data from Swiss cantons.

When looking at the empirical relationships through the lens of the model, I can thus conclude that heterogeneity in the difficulty to pass the *Matura* exams alone cannot account for the negative relationship of enrolment and success rates observed. Hence, how strict the Swiss cantons are in admitting students to *Gymnasium* plays a crucial role in explaining the drop-out risk they face, both at the upper-secondary and the tertiary level. Additionally, at the tertiary level, students coming from cantons with higher *Matura* rates not only face higher drop-out rates because they have, on average, a lower level of academic ability, but also because they are more likely go to university instead of a university of applied sciences, which again leads to higher drop-out rates.

7 Conclusion

Drop-out risk is a central factor in a young student's education choice. Yet, there is a lot of heterogeneity in drop-out risk between countries and also within countries. In this paper, I have shown how the interaction between drop-out risk and education premia shape education choices along the ability distribution. Modelling an education system which goes beyond the binary choice between working and attending university, I have shown that the possibility for students to choose between more than one risk-return combination improves sorting into education.

Furthermore, the model uncovered a new channel through which progressive taxation affects education choices. While typically, progressive taxation is seen to diminish the return to education and thus reduces investment into human capital, I show that progressive taxation also reduces the variance of the returns. Thus, there is a counteracting effect which reduces risk and I show that for common parameter values, this effect dominates. Thus, progressive taxation might improve sorting into education through risk reduction.

I have also used the model to analyse cross-sectional variation in enrolment, attainment and success rates between different Swiss cantons. Generating simulated data from the model in a numerical example, I have shown that heterogeneity in the strictness of admission to upper-secondary education plays a central role in explaining cross-cantonal variation. In contrast, heterogeneity in the difficulty to successfully graduate alone cannot account for empirical relationships observed in Swiss data.

My work presented here offers several directions for future research. The model can be refined to include learning about ability as done in Hendricks and Leukhina (2017, 2018). As discussed in Section 5.3, this would result in some graduates from *Gymnasium* to study at a university of applied sciences (FH), which is never optimal in the model presented here.

Furthermore, modelling the education period over more than one period, two model extensions are possible. First, this makes education choices with differing lengths possible, reflecting the fact that students from a university of applied sciences often choose to enter the job market after graduating with a bachelor's degree, whereas at university, it is very common to pursue a master's degree. Second, agents might choose to drop-out voluntarily after each education period in case they get a favorable income draw as modelled in Athreya and Eberly (2021).

The model also allows for an analysis of optimal policies. Optimal taxation can be studied in a setting where the planner can only infer the individual levels of ability by their education choices and income draws. The optimal design of the education system can by studied by analysing how the functions describing the drop-out probabilities $p(\theta)$ should optimally look like. Finally, calibrating the model would allow for a quantitative analysis of the heterogeneity in enrolment and drop-out rates between Swiss cantons beyond the numerical illustration in Section 6.3.

References

Aepli, M., Kuhn, A., and Schweri, J. (2021). *Der Wert von Ausbildungen auf dem Schweizer Arbeitsmarkt*. Grundlagen für die Wirtschaftspolitik Nr. 31. Staatssekretariat für Wirtschaft SECO, Bern, Schweiz.

Akyol, A. and Athreya, K. (2005). Risky higher education and subsidies. *Journal of Economic Dynamics and Control*, 29(6):979–1023.

Athreya, K. and Eberly, J. (2021). Risk, the college premium, and aggregate human capital investment. *American Economic Journal: Macroeconomics*, 13(2):168–213.

Behr, A., Giese, M., Teguim Kamdjou, H. D., and Theune, K. (2021). Motives for dropping out from higher education – An analysis of bachelor's degree students in Germany. *European Journal of Education*, 56(2):325–343.

BFS (2018). *Längsschnittanalysen im Bildungsbereich (LABB)*. Bundesamt für Statistik (BFS).

Bound, J. and Turner, S. (2011). Dropouts and diplomas: the divergence in collegiate outcomes. In *Handbook of the Economics of Education*, volume 4, pages 573–613. Elsevier.

Bovenberg, A. L. and Jacobs, B. (2005). Redistribution and education subsidies are Siamese twins. *Journal of Public Economics*, 89(11–12):2005–2035.

Bradley, S. and Nguyen, A. N. (2004). The school-to-work transition. In Johnes, G. and Jones, J., editors, *International Handbook on the Economics of Education*, chapter 13, pages 484–521. Edward Elgar.

Castex, G. (2017). College risk and return. *Review of Economic Dynamics*, 26:91–112.

Castro, R. and Coen-Pirani, D. (2016). Explaining the evolution of educational attainment in the United States. *American Economic Journal: Macroeconomics*, 8(3):77–112.

Donovan, K. and Herrington, C. (2019). Factors affecting college attainment and student ability in the us since 1900. *Review of Economic Dynamics*, 31:224–244.

Eberle, F., Gehrer, K., Jaggi, B., Kottonau, J., Oepke, M., and Pflüger, M. (2008). *Evaluation der Maturitätsreform 1995 (EVAMAR). Schlussbericht zur Phase II*. Staatssekretariat für Bildung und Forschung SBF.

Gollier, C. (2001). *The economics of risk and time*. MIT press.

Hanushek, E. A., Schwerdt, G., Woessmann, L., and Zhang, L. (2017). General education, vocational education, and labor-market outcomes over the lifecycle. *Journal of Human Resources*, 52(1):48–87.

Heathcote, J., Storesletten, K., and Violante, G. L. (2017). Optimal tax progressivity: an analytical framework. *Quarterly Journal of Economics*, 132(4):1693–1754.

Hendricks, L. and Leukhina, O. (2017). How risky is college investment? *Review of Economic Dynamics*, 26:140–163.

Hendricks, L. and Leukhina, O. (2018). The return to college: selection and dropout risk. *International Economic Review*, 59(3):1077–1102.

Jacobs, B. and Bovenberg, A. L. (2011). Optimal taxation of human capital and the earnings function. *Journal of Public Economic Theory*, 13(6):957–971.

Korber, M. and Oesch, D. (2019). Vocational versus general education: employment and earnings over the life course in Switzerland. *Advances in Life Course Research*, 40:1–13.

Larsen, M. S., Kornbeck, K. P., Kristensen, R. M., Larsen, M. R., and Sommersel, H. B. (2013). *Dropout Phenomena at Universities: What is Dropout? Why does Dropout Occur? What Can be Done by the Universities to Prevent or Reduce it? A systematic review*, volume 45. Danish Clearinghouse for Educational Research, Department of Education, Aarhus University.

Mueller, B. and Schweri, J. (2015). How specific is apprenticeship training? Evidence from inter-firm and occupational mobility after graduation. *Oxford Economic Papers*, 67(4):1057–1077.

OECD (2008). *Education at a glance*. Organisation for Economic Co-operation and Development (OECD).

OECD (2019). *Education at a glance*. Organisation for Economic Co-operation and Development (OECD).

Ryan, P. (2001). The school-to-work transition: a cross-national perspective. *Journal of Economic Literature*, 39(1):34–92.

Schweri, J., Eymann, A., and Aepli, M. (2020). Horizontal mismatch and vocational education. *Applied Economics*, 52(32):3464–3478.

SKBF (2014). *Bildungsbericht Schweiz*. Schweizerische Koordinationsstelle für Bildungsforschung (SKBF).

SKBF (2018). *Bildungsbericht Schweiz*. Schweizerische Koordinationsstelle für Bildungsforschung (SKBF).

Wolter, S. C., Diem, A., and Messer, D. (2014). Drop-outs from Swiss universities: an empirical analysis of data on all students between 1975 and 2008. *European Journal of Education*, 49(4):471–483.

Wolter, S. C. and Ryan, P. (2011). Apprenticeship. In *Handbook of the Economics of Education*, volume 3, pages 521–576. Elsevier.

A Alternative education choices

As mentioned in Section 5.2, the education choices along the ability distribution are governed by the three critical values for ability θ^*_{uw}, θ^*_{fw}, and θ^*_{uf}. Crucially, their order defines how partitions of the ability distribution are mapped into education choices. Yet, depending on the model parameters, this order can potentially be any of the six permutations with and without (strict) inequality. In this appendix, I show that only three cases are possible.

Proposition A.1. *The values θ^*_{uw}, θ^*_{fw}, and θ^*_{uf} can only be ordered in three ways:*

$$\begin{aligned} &\textit{Case 1:} \quad \theta^*_{uf} = \theta^*_{uw} = \theta^*_{fw}, \\ &\textit{Case 2:} \quad \theta^*_{uf} < \theta^*_{uw} < \theta^*_{fw}, \\ &\textit{Case 3:} \quad \theta^*_{fw} < \theta^*_{uw} < \theta^*_{uf}. \end{aligned}$$

Proof. See Appendix C. □

Of the three cases given above, only the third case is plausible, as it is the only case where anyone chooses to study at a university of applied sciences (FH). In the first case, the values $V_u(\theta)$, $V_f(\theta)$, and $V_w(\theta)$ all coincide at the same value θ^*. All agents with $\theta < \theta^*$ would then choose to work and all agents with $\theta > \theta^*$ would go to university and no agent chooses university of applied sciences (FH). As such a case could only happen by coincidence, I discard it.

Figure A.1 depicts the second case. As can be seen in the graph, going to a university of applied sciences (FH) is always dominated by either work or university. Thus, as in the first case, the ability distribution is split into two groups: those with academic ability $\theta < \theta^*_{uw}$ choose to work, and those with academic ability $\theta > \theta^*_{uw}$ go to university. This case can happen through three different channels. First, if the success probability $p_f(\theta)$ is too low relative to the success probability $p_u(\theta)$, agents from the upper end of the ability distribution rather choose to go to university directly instead of choosing FH. Second, if the education premium R_f is too low relative to R_u, the same is true. Third, if the education premium R_f is too low relative to R_s, then agents rather choose to work instead of going to FH. However, since we observe students going to FH, I abstract from this case and always assume the parameter values to be such that the third case holds.

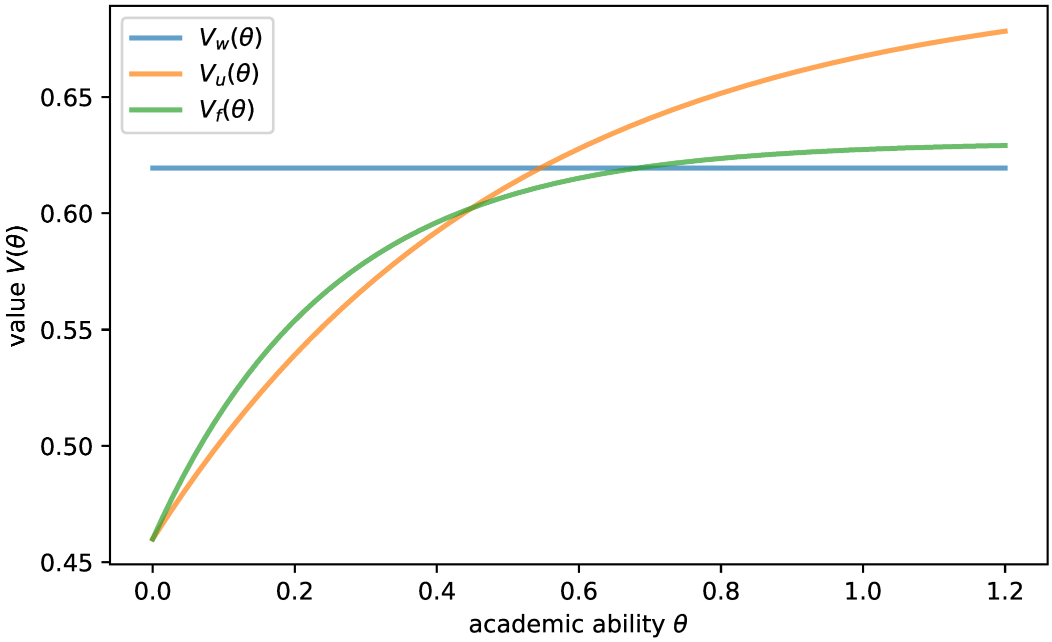


Figure A.1: Value functions with university of applied sciences (FH) never optimal.

B Optimal choices in the model with upper-secondary education

In this appendix, I characterise optimal education choices in the model with upper-secondary education. Furthermore, I derive the critical values for academic ability θ^*_{uw}, θ^*_{fw} and θ^*_{uf}. All agents have to take several decisions sequentially. In order to structure the analysis, I define the set of available education choices as follows.

Definition B.1. *Let a and v denote the choices for attempting to get admission to Gymnasium and VET, and let w, f and u denote the choices to work, go to a university of applied sciences (FH) or go to university. The set E is defined as the set of all education choices available to an individual. It is given by*

$$E = \{(a, uf), (a, uw), (a, ff), (a, fw), (a, wf), (a, ww), (v, f), (v, w)\}.$$

where (a, uf) means choosing to go to university, if Gymnasium is successfully completed and choosing to go to FH, if the agent fails to get admission to Gymnasium and gets VET instead. The other elements of E are defined accordingly.

The set E thus describes the full set of available education choices. However, some of the education choices are never optimal.

Proposition B.1. *It is never optimal to choose go to Gymnasium and work afterwards, i.e. (a, wf) and (a, ww) are never optimal.*

Proof. See Appendix C. □

Proposition B.2. *It is never optimal to choose go to Gymnasium and study at a university of applied sciences (FH) afterwards, i.e. (a, fw) and (a, ff) are never optimal.*

Proof. See Appendix C. □

The set of education choices E therefore reduces to

$$E' = \{(a, uf), (a, uw), (v, f), (v, w)\}.$$

I start by analysing the education system without university of applied sciences (FH). In this case, the only two choices available are (a, uw) and (v, w).

The corresponding value functions are given by

$$
\begin{aligned}
V_{vw}(\theta) &= u(y_e) + \beta u(y_2) + \beta^2 \mathbb{E}\left[u(y_3)\right], \\
V_{auw}(\theta) &= u(y_e) + \beta\Big[p_a(\theta)u(y_e) + \big(1 - p_a(\theta)\big)u(y_2)\Big] + \\
&\quad \beta^2 \mathbb{E}\Big[p_a(\theta)p_m(\theta)p_u(\theta)u(R_u y_3) + \big(1 - p_a(\theta)p_m(\theta)p_u(\theta)\big)u(y_3)\Big].
\end{aligned}
$$

Agents choosing to get vocational education and training (VET) and work afterwards get an income stream of y_e, y_2, and y_3. Agents choosing to attempt getting admitted to *Gymnasium* with the goal of studying at university succeed to get admission with probability $p_a(\theta)$. In the first period, they get subsistence income y_e irrespective of whether they get VET or attend *Gymnasium*. Agents who get into *Gymnasium* succeed with probability $p_m(\theta)$ and go to university in the second period. With probability $1 - p_m(\theta)$, they drop out of *Gymnasium* and get VET in period 2. In either case, they earn subsistence income y_e, which is why $u(y_e)$ enters the value function with probability $p_a(\theta)$ in the second period. With probability $1-p_a(\theta)$, they work in the second period earning income y_2 after they got vocational education and training in the first period. In the third period, agents have obtained a university degree with probability $p_a(\theta)p_m(\theta)p_u(\theta)$ and earn income $R_u y_3$. With probability $1 - p_a(\theta)p_m(\theta)p_u(\theta)$, they have either failed to get admitted to *Gymnasium* or have dropped out at *Gymnasium* or university. They earn income y_3 in the third period. The critical value θ^*_{uw} satisfies $V_{vw}(\theta^*_{uw}) = V_{auw}(\theta^*_{uw})$. It is given by

$$
p_m(\theta^*_{uw})p_u(\theta^*_{uw}) = \frac{u(y_2) - u(y_e)}{\beta \mathbb{E}\left[u(R_u y_3) - u(y_3)\right]}.
$$

I turn to the case where going to a university of applied sciences (FH) is an available option. In this case, agents can choose between the four options given in the choice set E'. As before, let θ^*_{fw} denote the critical value, where the agents are indifferent between working and going to FH, i.e. $V_{vw}(\theta^*_{fw}) = V_{vf}(\theta^*_{fw})$ and define θ^*_{uw} accordingly. In the same way as discussed in Proposition A.1, I assume the parameters to be such that at least some agents optimally choose to go to FH (the standard case). This case implies $\theta^*_{fw} < \theta^*_{uw}$, i.e. when moving up along the ability distribution, some agents prefer going to FH over working before going to university is preferred. Assuming this case reduces the choice set E' even further:

Proposition B.3. *If $\theta^*_{fw} < \theta^*_{uw}$, then the choice (a, uw) is never optimal, i.e. it is never optimal to choose to go to Gymnasium and study at university and choose to work in case of failing to get admitted to Gymnasium.*

Proof. See Appendix C. □

Proposition B.3 implies that agents who fail to get admitted to *Gymnasium* will choose to go to a university of applied sciences (FH) after they finish their VET degree. Thus, the choice set E' reduces to

$$E^* = \{(a, uf), (v, f), (v, w)\}.$$

The corresponding value functions are depicted in Figure B.1 and are given by

$$\begin{aligned}
V_{vw}(\theta) &= u(y_e) + \beta u(y_2) + \beta^2 \mathbb{E}[u(y_3)], \\
V_{vf}(\theta) &= u(y_e) + \beta u(y_e) + \beta^2 \mathbb{E}\Big[p_f u(R_f y_3) + (1 - p_f) u(y_3)\Big], \\
V_{auf}(\theta) &= u(y_e) + \beta u(y_e) + \beta^2 \mathbb{E}\Big[p_a p_m p_u u(R_u y_3) + (1 - p_a) p_f u(R_f y_3) + \\
& \Big[1 - p_a p_m p_u - (1 - p_a) p_f u(y_3)\Big]\Big].
\end{aligned}$$

Note that all probabilities are a function of academic ability θ, which I omit for better readability. Agents choosing to get vocational education and training (VET) and work afterwards get an income stream of y_e, y_2, and y_3. Agents choosing to go to a university of applied sciences (FH) after VET get subsistence income y_e in the first two periods. With probability $p_f(\theta)$, they graduate from FH and earn income $R_f y_3$ in the third period. With probability $1 - p_f(\theta)$, they drop out and earn income y_3. Finally, agents choosing to attempt getting admitted to *Gymnasium* with the goal of studying at university succeed to get admission with probability $p_a(\theta)$. As discussed before, they get subsistence income y_e irrespective of whether they get VET or attend *Gymnasium*. In the third period, agents have obtained a university degree with probability $p_a(\theta)p_m(\theta)p_u(\theta)$ and earn income $R_u y_3$. With probability $(1 - p_a(\theta))p_f(\theta)$, they have failed to get admitted to *Gymnasium* but successfully graduate from a university of applied sciences (FH). In this case, they earn $R_f y_3$. With the remaining probability, i.e. $1 - p_a(\theta)p_m(\theta)p_u(\theta) - (1 - p_a(\theta))p_f(\theta)$, they have dropped out at either *Gymnasium*,

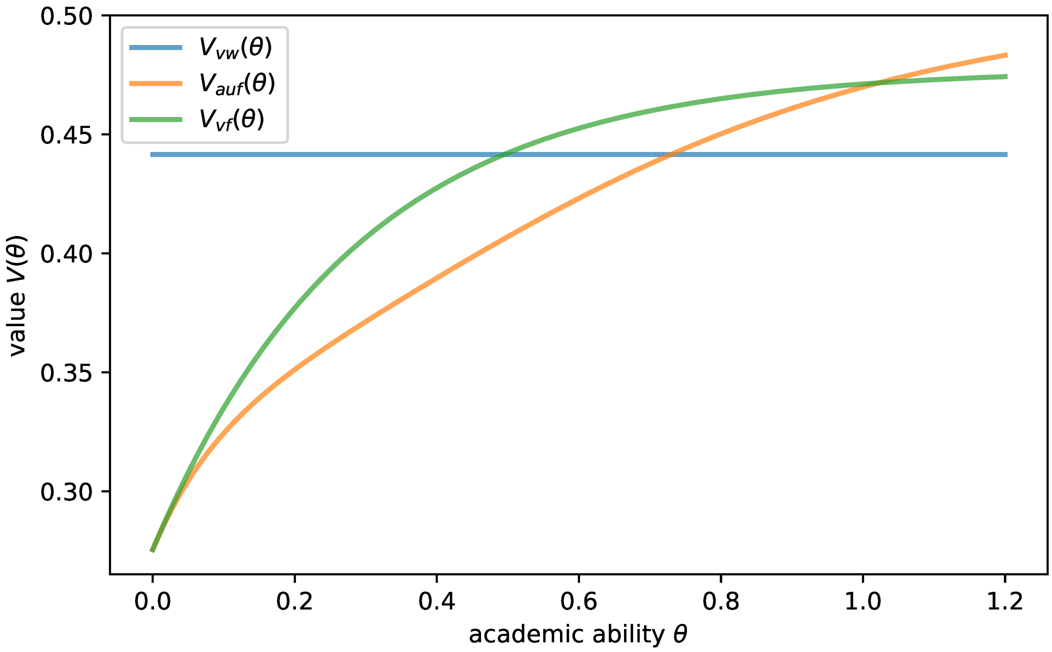


Figure B.1: Value functions for work, university and FH.

university of applied sciences (FH) or at university. They earn income y_3 in the third period. The critical values θ^*_{fw} and θ^*_{uf} satisfy $V_{vw}(\theta^*_{fw}) = V_{vf}(\theta^*_{fw})$ and $V_{vf}(\theta^*_{uf}) = V_{auf}(\theta^*_{uf})$. They are given by

$$p_f(\theta^*_{fw}) = \frac{u(y_2) - u(y_e)}{\beta \mathbb{E}\Big[u(R_f y_3) - u(y_3)\Big]},$$

$$\frac{p_m(\theta^*_{uf}) p_u(\theta^*_{uf})}{p_f(\theta^*_{uf})} = \frac{\mathbb{E}\Big[u(R_f y_3) - u(y_3)\Big]}{\mathbb{E}\Big[u(R_u y_3) - u(y_3)\Big]}.$$

C Proofs

Proposition 5.1. *Let θ^*_{uw} be the level of θ where the agent is indifferent between working and studying at university. Then, for all values $\theta > \theta^*_{uw}$, studying at university is preferred over working.*

Proof. The threshold θ^*_{uw} is characterised by

$$p_u(\theta^*_{uw})\beta\mathbb{E}[u(R_u y_2) - u(y_2)] = u(y_1) - u(y_e).$$

This can be rewritten as

$$p_u(\theta^*_{uw}) = \frac{u(y_1) - u(y_e)}{\beta\mathbb{E}[u(R_u y_2) - u(y_2)]}.$$

As $p_u(\theta)$ increases monotonically in θ, it follows that for all values $\theta > \theta^*_{uw}$,

$$p_u(\theta) \geq \frac{u(y_1) - u(y_e)}{\beta\mathbb{E}[u(R_u y_2) - u(y_2)]}.$$

Hence, for all values $\theta > \theta^*_{uw}$, choosing to work is dominated. □

Corollary 5.1.1. *By the same argument, it follows that for all values $\theta < \theta^*_{uw}$, choosing university is dominated and thus, θ^*_{uw} is unique.*

Corollary 5.1.2. *Using the same line of reasoning, Proposition 5.1 and Corollary 5.1.1 also hold for θ^*_{fw} and thus, θ^*_{fw} is unique.*

Proposition 5.2. *Let θ^*_{uf} be the level of θ where the agent is indifferent between the two education choices. Then, for all values $\theta > \theta^*_{uf}$, studying at university is preferred over studying at a university of applied sciences (FH).*

Proof. The threshold θ^*_{uf} is characterised by

$$p_u(\theta^*_{uf})\mathbb{E}\left[u(R_u y_2) - u(y_2)\right] = p_f(\theta^*_{uf})\mathbb{E}\Big[u(R_f y_2) - u(y_2)\Big].$$

This can be rewritten as

$$\frac{p_u(\theta^*_{uf})}{p_f(\theta^*_{uf})} = \frac{\mathbb{E}\Big[u(R_f y_2) - u(y_2)\Big]}{\mathbb{E}\Big[u(R_u y_2) - u(y_2)\Big]}.$$

As p_u/p_f increases monotonically in θ by the regularity condition imposed in Definition 5.1, it follows that for all values $\theta > \theta^*_{uf}$,

$$\frac{p_u(\theta)}{p_f(\theta)} > \frac{\mathbb{E}\Big[u(R_f y_2) - u(y_2)\Big]}{\mathbb{E}\Big[u(R_u y_2) - u(y_2)\Big]}.$$

Hence, for all values $\theta > \theta^*_{uf}$, choosing FH is dominated. □

Corollary 5.2.1. *By the same argument, it follows that for all values $\theta < \theta^*_{uf}$, choosing university is dominated and thus, θ^*_{uf} is unique.*

Proposition A.1. *The values θ^*_{uw}, θ^*_{fw}, and θ^*_{uf} can only be ordered in three ways:*

$$\begin{aligned}
&\textit{Case 1:} \quad \theta^*_{uf} = \theta^*_{uw} = \theta^*_{fw}\,,\\
&\textit{Case 2:} \quad \theta^*_{uf} < \theta^*_{uw} < \theta^*_{fw}\,,\\
&\textit{Case 3:} \quad \theta^*_{fw} < \theta^*_{uw} < \theta^*_{uf}\,.
\end{aligned}$$

Proof. The proof consists of three parts. In Part 1, I show that if $\theta^*_{uf} = \theta^*_{uw}$, then the first case must follow. In Part 2, I show that if $\theta^*_{uf} < \theta^*_{uw}$, then the second case must follow. In Part 3, I show that if $\theta^*_{uf} > \theta^*_{uw}$, then the third case must follow.

Part 1. *I show that $\theta^*_{uf} = \theta^*_{uw}$ implies $\theta^*_{uf} = \theta^*_{uw} = \theta^*_{fw}$.*
The value θ^*_{uf} is defined to satisfy $V_u(\theta^*_{uf}) = V_f(\theta^*_{uf})$. Likewise, the value θ^*_{uw} is defined to satisfy $V_u(\theta^*_{uw}) = V_w(\theta^*_{uw})$. If $\theta^*_{uf} = \theta^*_{uw}$, it follows that $V_f(\theta^*_{uw}) = V_w(\theta^*_{uw})$. As θ^*_{fw} is defined to satisfy $V_f(\theta^*_{fw}) = V_w(\theta^*_{fw})$, it must follow that $\theta^*_{uw} = \theta^*_{uf}$ by uniqueness of θ^*_{fw} (Corollary 5.1.2). Hence, it follows that $\theta^*_{uf} = \theta^*_{uw} = \theta^*_{fw}$.

Part 2. *I show that $\theta^*_{uf} < \theta^*_{uw}$ implies $\theta^*_{uf} < \theta^*_{uw} < \theta^*_{fw}$ (by contradiction).*
Given $\theta^*_{uf} < \theta^*_{uw}$, it follows from Proposition 5.2 that $V_u(\theta^*_{uw}) > V_f(\theta^*_{uw})$. As the value θ^*_{uw} is defined to satisfy $V_u(\theta^*_{uw}) = V_w(\theta^*_{uw})$ it follows that $V_w(\theta^*_{uw}) > V_f(\theta^*_{uw})$. Now assume $\theta^*_{fw} < \theta^*_{uw}$. From Proposition 5.1 it follows that $V_w(\theta^*_{uw}) < V_f(\theta^*_{uw})$, which is a contradiction. Thus, it must be that $\theta^*_{uw} \leq \theta^*_{fw}$. However, if $\theta^*_{fw} = \theta^*_{uw}$, then this would imply $\theta^*_{uf} = \theta^*_{uw}$, by the same argument as in Part 1, which is a contradiction. Hence, it follows that $\theta^*_{uf} < \theta^*_{uw} < \theta^*_{fw}$.

Part 3. *I show that $\theta^*_{uw} < \theta^*_{uf}$ implies $\theta^*_{fw} < \theta^*_{uw} < \theta^*_{uf}$ (by contradiction).*
The argument is the same is in Part 2 but using Corollary 5.1.1 and Corollary 5.2.1 instead of Proposition 5.1 and Proposition 5.2. □

Proposition B.1. *It is never optimal to choose go to Gymnasium and work afterwards, i.e.* (a, wf) *and* (a, ww) *are never optimal.*

Proof. The proof is in two parts. In the first part, I show that (a, wf) is never optimal. In the second part, I show that (a, ww) is always dominated by (v, w).

Part 1. Recall the following value functions:

$$\begin{aligned}
V_m(\theta) &= u(y_e) + \beta \mathbb{E}\Big[p_m(\theta) V_m'(\theta) + \big(1 - p_m(\theta)\big) V_{ls}(\theta)\Big], \\
V_v(\theta) &= u(y_e) + \beta \mathbb{E}\left[V_v'(\theta)\right], \\
V_m'(\theta) &= \max\Big\{V_w(\theta), V_f(\theta), V_u(\theta)\Big\}, \\
V_v'(\theta) &= \max\Big\{V_w(\theta), V_f(\theta)\Big\}.
\end{aligned}$$

Assume that for a given level of academic ability $\hat{\theta}$, $V_w(\hat{\theta})$ maximises $V_m'(\theta)$. Then it must be that $V_w(\hat{\theta}) > V_f(\hat{\theta})$ and which implies that $V_w(\hat{\theta})$ also maximises $V_v'(\theta)$. Hence, (a, wf) is never optimal.

Part 2. Let $V_{vw}(\theta)$ denote the value of someone getting VET and going to work afterwards. Similarly, let $V_{mw}(\theta)$ denote the value of someone going to *Gymnasium* and going to work afterwards. The values $V_{vw}(\theta)$ and $V_{mw}(\theta)$ are given by

$$\begin{aligned}
V_{vw}(\theta) &= u(y_e) + \beta \mathbb{E}\left[V_w(\theta)\right], \\
V_{mw}(\theta) &= u(y_e) + \beta \mathbb{E}\Big[p_m(\theta) V_w(\theta) + \big(1 - p_m(\theta)\big)\big(u(y_e) + \beta u(y_3)\big)\Big], \\
V_w(\theta) &= u(y_2) + \beta \mathbb{E}\left[u(y_3)\right].
\end{aligned}$$

Plugging in $V_w(\theta)$ and subtracting $V_{mw}(\theta)$ from $V_{vw}(\theta)$, we get

$$V_{vw}(\theta) - V_{mw}(\theta) = \beta\big(1 - p_m(\theta)\big)\mathbb{E}\Big[u(y_2) - u(y_e)\Big],$$

which is always positive. Thus, $V_{vw}(\theta) > V_{mw}(\theta)$ for all levels of θ. Furthermore, by choosing (a, ww), the value $V_a(\theta)$ is given by

$$V_a(\theta) = p_a(\theta) V_{mw}(\theta) + \big(1 - p_a(\theta)\big) V_{vw}(\theta).$$

Therefore, $V_a(\theta)$ is a linear combination of $V_{vw}(\theta)$ and $V_{mw}(\theta)$ and is thus never greater than $V_{vw}(\theta)$. Hence, (v, w) is always (weakly) preferred to (a, ww). □

Proposition B.2. *It is never optimal to choose go to Gymnasium and study at a university of applied sciences (FH) afterwards, i.e.* (a, fw) *and* (a, ff) *are never optimal.*

Proof. The proof is in two parts. In the first part, I show that (a, fw) is never optimal. In the second part, I show that (a, ff) is always dominated by (v, f).

Part 1. Showing that (a, fw) is never optimal is achieved by the same reasoning as before. If $V_f(\theta)$ maximises $V_m'(\theta)$ then it cannot be that $V_w(\theta)$ maximises $V_v'(\theta)$. Hence, (a, fw) is never optimal.

Part 2. Let $V_{vf}(\theta)$ denote the value of someone getting VET and going to work afterwards. Similarly, let $V_{mf}(\theta)$ denote the value of someone going to *Gymnasium* and study at a university of applied science (FH) afterwards. The values $V_{vf}(\theta)$ and $V_{mf}(\theta)$ are given by

$$\begin{aligned} V_{vf}(\theta) &= u(y_e) + \beta \mathbb{E}\Big[V_f(\theta)\Big], \\ V_{mf}(\theta) &= u(y_e) + \beta \mathbb{E}\Big[p_m(\theta) V_f(\theta) + \big(1 - p_m(\theta)\big)\big(u(y_e) + \beta u(y_3)\big)\Big], \\ V_f(\theta) &= u(y_e) + \beta \mathbb{E}\Big[p_f(\theta) u(R_f y_3) + \big(1 - p_f(\theta)\big) u(y_3)\Big]. \end{aligned}$$

Plugging in $V_f(\theta)$ and subtracting $V_{mf}(\theta)$ from $V_{vf}(\theta)$, we get

$$V_{vf}(\theta) - V_{mf}(\theta) = \beta^2 \big(1 - p_m(\theta)\big) p_f(\theta) \mathbb{E}\Big[u(Ry_3) - u(y_3)\Big],$$

which is always positive. As before, by choosing (a, ff), the value $V_a(\theta)$ is a linear combination of $V_{vf}(\theta)$ and $V_{mf}(\theta)$ and is thus never greater than $V_{vf}(\theta)$. Hence, (v, f) is always (weakly) preferred to (a, ff). □

Proposition B.3. *If* $\theta^*_{fw} < \theta^*_{uw}$*, then the choice* (a, uw) *is never optimal, i.e. it is never optimal to choose to go to Gymnasium and study at university and choose to work in case of failing to get admitted to Gymnasium.*

Proof. For all values $\theta < \theta^*_{uw}$, we have $V_{auw}(\theta) < V_{vw}(\theta)$, and thus, the choice (a, uw) is dominated by (v, w). Furthermore, for all values $\theta > \theta^*_{fw}$, we have $V_{auw}(\theta) < V_{auf}(\theta)$, and thus, the choice (a, uw) is dominated by (a, uf). Given that $\theta^*_{fw} < \theta^*_{uw}$, it follows that for all levels of academic ability θ, the choice (a, uw) is either dominated by (v, w), (a, uf) or both. $\square$

D Parameters for the numerical illustration

In this appendix, I describe the functional form assumptions and parameter values used for the numerical illustration in Section 6.3. Note that this is not a calibration exercise but merely a numerical illustration and thus, the choice of functional forms and parameter values is somewhat arbitrary. I have chosen the parameter values so that the resulting enrolment and success rate at *Gymnasium*, the *Matura* rate, the success rate at university, and the share of tertiary graduates are similar to the Swiss average.

The parameter choices are summarised in Table D.1. The discount rate is $\beta = 0.96$. Preferences are CRRA with risk aversion $\gamma = 2$. I normalise the working income in period two to $y_2 = 1$.[18] Income while in education is $y_e = 0.9$. Income in period three is lognormally distributed, i.e. $\log y_3 \sim \mathcal{N}(\mu_y, \sigma_y^2)$, with parameters $\mu_y = 1$ and $\sigma_y = 0.5$. The education premia are $R_u = 6$ and $R_f = 2$. I model the functions describing the success probabilities $p(\theta)$ as CDFs of an exponential distribution with parameter λ. With this functional form assumption, the conditions specified in Definition 5.1 are satisfied if $\lambda_u < \lambda_f$. I choose $\lambda_u = 1.4$ and $\lambda_f = 1.7$. For the numerical illustration of the heterogeneity between cantons, I vary λ_a and λ_m between 0.5 and 2.4. While varying one, I fix the other at $\lambda = 1.3$. Finally, I assume the ability distribution to be a Pareto distribution with support $\mathbb{R}_{\geq 0}$ and mean 1.

I also want to quickly discuss the relationship between the parameter values and the observed success rates. As an example, I use the success rates on the university level. Let $\overline{p}_u$ denote the average success rate observed in a given canton, i.e. one minus the drop-out rate. While $\overline{p}_u$ is just a scalar, the success rate $p_u(\theta)$ is a function that maps each level of ability θ into an individual success rate. The average success rate resulting in the model is then the conditional mean of $p_u(\theta)$ for all levels of ability θ choosing to attend university. Let Θ_u denote the set ability levels of all agents attending university. Then, we have $\overline{p}_u = \mathbb{E}[p_u(\theta) \mid \theta \in \Theta_u]$. The set Θ_u itself is endogenous and depends on all the parameter values, income levels and education premia. Furthermore, to obtain $\overline{p}_u$ in the model, one needs to integrate over the ability distribution, which itself depends on the distribution and parameters chosen. In the model which only includes tertiary education, the set Θ_u is given by $\Theta_u = \{\theta \mid \theta > \theta^*_{uf}\}$. Once we

[18] In period one, everyone is getting education and hence, a value for working income is only needed for periods two and three.

introduce upper-secondary education, not all of the agents with $\theta > \theta^*_{uf}$ attend university as some of them fail to get admitted to *Gymnasium* and some of them fail to successfully complete *Gymnasium*. I obtain the values for $\overline{p}_u$ in Section 6.3 by simulating one million agents for each canton and then calculating the average success rate of all agents attending university.

Even though for a given level of θ, we assume $p_f(\theta) > p_u(\theta)$, the same is not necessarily true for the average success rates $\overline{p}_f$ and $\overline{p}_u$. This is due to a selection effect. Although the success rate is lower at university for a given ability level, it is the high ability types that select this education path. Thus, depending on the levels of ability choosing each education path, we might have $\overline{p}_f > \overline{p}_u$ or vice versa. This is also why the assumption that $p_f(\theta) > p_u(\theta)$ for a given ability level θ cannot be tested empirically using only data on drop-outs without having a measure for ability.

Table D.1: Parameter values for numerical example

Preferences	
Discount factor β	0.96
Risk aversion γ	2
Income	
Work income y_2	1
Education income y_e	0.9
Location parameter μ_y	1
Scale parameter σ_y	0.5
Education premia	
Premium for university R_u	6
Premium for FH R_f	2
Success probabilities	
Parameter for university λ_u	1.4
Parameter for FH λ_f	1.7
Parameter for *Gymnasium* λ_m	$[0.5, 2.4]$
Parameter for admission to *Gymnasium* λ_a	$[0.5, 2.4]$
Ability distribution (Pareto)	
Support	$\mathbb{R}_{\geq 0}$
Mean	1

Curriculum Vitae

Education

2017 – 2022	**University of St. Gallen** Doctoral student in Economics and Finance Advisor: Prof. Dr. Winfried Koeniger
2012 – 2015	**University of Zurich** Master of Arts in Economics
2009 – 2012	**University of St. Gallen** Bachelor of Arts in Economics

Professional experience

2017 – 2021	**University of St. Gallen** Research and Teaching Assistant Swiss Institute for Empirical Economic Research (SEW)
2016 – 2017	**Ecoplan, Economic Research and Policy Consultancy** Internship
2015 – 2017	**Swiss Federal Railways SBB CFF FFS** Graduate IT-Trainee Programme
2015	**Fahrländer Partner, Research in Spatial Development** Internship