Unequal lives

Heterogeneity in income risk and household portfolios*

Christian Bayer† Sebastian Hildebrand‡ Thomas Hintermaier§
Moritz Kuhn¶ Gašper Ploj‖

March 20, 2025

Abstract

Income risk is the largest economic risk faced by households. We document large and persistent heterogeneity in this risk with a negative correlation between risk and income levels. We develop a Markov-switching state-space model for earnings risk that provides us with estimates of the time-varying risk type of a worker. Combining these estimates with wealth data, we show that low-risk workers accumulate more wealth, in less liquid and higher-return assets. A life-cycle model explains the empirical patterns of risk, wealth accumulation, and portfolios based on persistent risk types and incomplete, frictional asset markets with systematic return and liquidity differences.

JEL Classification: E23, J30, J64.
Keywords: risk heterogeneity, earnings dynamics, income risk, portfolio choice.

*We thank Tatjana Mika and Michael Stegmann for their help regarding questions on the data. We thank conference participants at the mini conference on heterogeneous agents at Zurich and Kurt Mitman, Xavier Ragot for helpful comments and remarks. Financial support by ADEMU - 649396 grant is gratefully acknowledged. Bayer thanks the DFG for support through CRC TR 224 EPoS (Project C05) and the Cluster of Excellence ECONtribute. Kuhn thanks the DFG for financial support (DFG No. 433368816). Ploj gratefully acknowledges financial support from the DAAD, the Bonn Graduate School of Economics, and the DFG through the Research Training Group "The Macroeconomic of Inequality". Hildebrand gratefully acknowledge funding and support by the DFG through the Research Training Group "The Macroeconomic of Inequality" and through CRC TR 224 EPoS (Project C05). All opinions expressed are personal and do not necessarily represent the views of the Banka Slovenije, the European Central Bank or the European System of Central Banks.

†Department of Economics, University of Bonn, CEPR, and IZA.
‡Department of Economics, University of Bonn.
§Department of Economics, University of Bonn.
¶Department of Economics, University of Mannheim, CEPR, IZA, and CESifo.
‖Banka Slovenije.

1 Introduction

Having a good income is nice, but knowing that your income is stable and will remain high is even nicer. In this paper, we document that labor income risks are very unevenly distributed. The majority of workers face very stable incomes, while some face highly volatile incomes. These high-risk workers also face lower average earnings. We also show that these differences in income and income risk have implications for the wealth these workers accumulate and the structure of their portfolios: High-risk workers accumulate less wealth relative to their income and invest in more liquid, lower-return assets. Conversely, some workers have an easy life with stable and steady income growth and high asset returns.

We use largely unexploited data from the German Social Security (pension) records on the employment histories of West German men to document novel facts about individual income risk. We combine these labor market data with information from individual balance sheets to document new facts about their saving and investment behavior. The strength of the social insurance data is that they contain the complete employment and earnings histories of a large representative sample of workers from age 14 to retirement. As a first finding, we confirm the findings of Guvenen et al. (2021) that income fluctuations have fat tails which can be modeled as mixtures of two normal distributions, one with a high and one with a low variance. Taking advantage of the fact that we have complete histories of workers, we can go beyond this finding and document a substantial and *persistent* heterogeneity in income risk across workers that standard Gaussian mixtures cannot capture. Most workers experience stable earnings, but a small fraction faces high earnings volatility. We therefore formulate a Markov-switching model where individual earnings risks fluctuates persistently over time. The comprehensive and complete coverage of earnings histories ("a large T setting") allows us to estimate this model of individual-level earnings dynamics because we can adapt and apply Markov-switching state-space estimators developed for time-series data (see in particular Kim, 1994). We find a significant persistence of the risk state. Conditional on entering the high risk state 10% of workers remain in high risk for more than five years. With a standard Gaussian mixture, this barely aver happens.

Having estimated the state-space model allows us to calculate for every worker at every point in time the likelihood of a worker's risk state. This allows us to relate risk empirically to other directly observable variables. First, we look at other direct measures of risk: health shocks and unemployment spells. We document that, consistent with heterogeneity in income risk, health and unemployment risks are concentrated among a minority of workers in the prime-age: 80 percent of unemployment (and serious illness) spells are concentrated among 20 (30) percent of workers. Exposure to high-risk earnings correlates with high unemployment and high health risks. A hypothetical worker who

would never be in the high-risk state between 30 and 50 would on average also never be unemployed at these ages and would virtually never have a long-term sickness spell. By contrast, a worker who hypothetically would be always in the high-risk state would also be more than 5 out of the 20 years unemployed and another close to two years in long-term sickness leave. In addition, we find that high-risk workers have roughly 30 percent lower labor earnings.

Second, we ask how workers respond in their savings and portfolio choices to fluctuations in risk. For this purpose we impute data wealth and household balance sheet data to the pension data by exploiting detailed individual balance sheet data together with earnings histories, unemployment spell data, and health information that is available in the German Socio-Economic Panel (SOEP). We document that workers facing the highest risks accumulate significantly less wealth—our hypothetical worker always in high risk would accumulate close to three average annual incomes less wealth. High-risk workers also structure their portfolios significantly differently than low-risk workers. The hypothetical permanent high-risk worker invests a 19 percentage points larger share of their wealth in liquid assets, which earn low returns. As a result, the hypothetical high-risk individuals have significantly lower portfolio returns on average (-138 basis points).

Finally, we show that these patterns arise naturally in a model of portfolio choice. We model the life-cycle saving and portfolio decisions of individuals facing income risk and a second-order risk of switching between persistent risk types. As in the data, we model risks as correlated with workers' income histories. Negative income shocks precede more turbulent times with higher income risk. This implies that, on average, individuals draw down their wealth before facing a risk shock. After the shock, individuals structure their portfolios to provide more insurance against further negative income draws. As a result, the portfolio becomes safer and more liquid, but it earns lower returns. We model this portfolio choice in a novel and, for many individuals, more realistic way. Individuals choose between a low-yielding liquid asset, which can be accessed at no cost and in every period, and an illiquid savings plan. The illiquid savings plan is a commitment to add or withdraw a fixed amount each period to/from an illiquid account that holds assets with higher but also riskier returns. An example of such an account would be a mortgage with an amortization schedule. By paying an adjustment fee, the individual can change the per-period flow into or out of his or her illiquid account.

Quantitatively, this explains well the systematic differences in individual portfolios and their returns. This is true both in the cross-section as well as over the life cycle. We read this as novel and striking evidence that individuals structure their portfolios in light of the income risks they face. Other research has shown that this systematic behavior has important macroeconomic consequences (Bayer et al., 2019). Most of the time, individuals live in tranquil times with small income fluctuations. Rarely, however, and usually after a large negative income shock, individuals enter a state of labor market

turbulence. We calibrate the model such that, on average, this transition to turbulence happens once every 20 years. Once the worker enters high labor market turbulence, it persists for a considerable period of time. We estimate the average duration of turbulence to be about five years. We show that the effects of second-order risk are substantial, in the data and in the model, and affect individuals even in retirement because the changed portfolio structure in turbulent times implies lower wealth and thus lower consumption in retirement. In fact, individuals would be willing to give up a significant share of their lifetime consumption could the high-risk state were eliminated.

Related literature. This paper relates primarily to two strands of literature. First, it adds to the literature that investigates the heterogeneity in labor market risk. Going back at least to Hall (1982), economists have documented a large heterogeneity in labor market outcomes across workers over their lifetime. Recent empirical work has corroborated this finding and concluded that a representative earnings process will struggle to account for the regularities of earnings dynamics in the data (Arellano, Blundell, and Bonhomme, 2017; Guvenen et al., 2021; Jung and Kuhn, 2018; Karahan, Ozkan, and Song, 2019; Kuhn and Ploj, 2020; Morchio, 2020). The literature shows that labor income risk varies, among others, across age (Blundell, Graber, and Mogstad, 2015; Jung and Kuhn, 2018; Karahan and Ozkan, 2013; Kuhn and Ploj, 2020), educational and skill groups (Blundell, Graber, and Mogstad, 2015; Mukoyama and Şahin, 2006) and with the business cycle (Bayer et al., 2019; Gálvez and Paz-Pardo, 2022; Storesletten, Telmer, and Yaron, 2004).

The increased availability of large, detailed administrative data has also allowed for a more granular and less parametric analysis of heterogeneity in labor market risk, with an increased focus on the higher moments of the earnings growth rate distribution (Guvenen et al., 2021; Manuel, Blundell, and Stéphane, 2017). Our paper follows the spirit of these papers and provides additional evidence that workers differ systematically in terms of *lifetime* earnings risk, and that these differences are largely driven by differences in lifetime unemployment and health risk. In this sense, our work complements that of Schmillen and Möller (2012) and Morchio (2020), who report a high concentration of lifetime unemployment using German and US data.

Second, we also contribute to the literature that studies the implications of heterogeneity in labor market risk for household savings and portfolio choices. This literature predominantly finds that income risk reduces the demand for risky assets, although the estimated effects are often small and insignificant (Angerer and Lam, 2009; Guiso, Jappelli, and Terlizzese, 1996; Heaton and Lucas, 2000, 2001; Palia, Qi, and Wu, 2014). Recently, Fagereng, Guiso, and Pistaferri (2018) found much larger effects than previously documented after improving the identification of wage risk variability. In corroboration of these findings, Chang et al. (2020) also show that the empirically estimated response of the risk share to wage risk can be explained by a standard portfolio choice model aug-

mented with idiosyncratic, time-varying income volatility.[1] We complement these findings on the empirical side and provide additional evidence on the effect of risk heterogeneity on wealth accumulation, portfolio returns, and asset (market) participation. We also introduce a portfolio choice model with a frictional saving constraint, which we show can explain well the observed life-cycle patterns of heterogeneity in portfolio liquidity and wealth accumulation across households with different labor market risks.

The remainder of the paper is structured as follows. In Section 2, we introduce the data from the German social security records and the German Socio-Economic Panel that we use. Section 3 then gives our main empirical results: earnings and wage risk are heterogeneous across workers and highly concentrated, they are negatively correlated with income growth and positively correlated with unemployment risk and health risk. Finally, we show that labor market risk heterogeneity leaves a clear mark on household savings decisions and the composition of their portfolios. In Section 4, we develop a life-cycle model with risk heterogeneity and portfolio choice to study the consequences of labor market risk through the lens of models of incomplete markets. Section 5 concludes.

2 Data

Our empirical analysis revisits the question of inequality in working life using little explored administrative data from German social security (pension) records, the *Versichertenkontenstichprobe* (VSKT). These administrative data are particularly well suited to address this question because they cover the entire employment history of workers starting at 14 years of age and also provide a high resolution of the earnings dynamics with monthly earnings and labor market observations. Compared to other social security data, such as those used in Guvenen et al. (2021), these data also report life events such as unemployment or health shocks that we can relate to the observed earnings dynamics. The VSKT data come as 2% annual samples of employment histories for workers with a social security record and from birth cohorts that are at the time of sampling between 15 and 67 years old.[2] We use the 18 scientific use files of the VSKT samples from 2002 to 2020 (Rentenversicherung (FDZ-RV), 2020).[3] The scientific use file provides a 25% subsample and restricts birth cohorts so that the youngest birth cohort is born in 1990 and the oldest birth cohort in 1935. The 2002 file has data on 57,832 individuals, the 2012 file has data on 64,295 individuals, and the 2019 file has data on 71,558 individuals.[4] The surveys contain weights to preserve representativeness for the cohorts.

[1] From a macroeconomic perspective, Bayer et al. (2019) show that business cycle fluctuations in income risk induce aggregate portfolio rebalancing with non-negligible effects on aggregate consumption and output. Larkin (2023) shows the consequences for the decline in consumption depending on the stability of jobs in the economy.

[2] The VSKT2020 samples workers up to age 70 in 2020.

[3] Himmelreicher and Stegmann (2008) provide a detailed description of the data.

[4] The sample for the 2020 file is slightly different so that this file contains data on 187,113 individuals.

We combine these data with data from the *German Socio-Economic Panel* (SOEP, Goebel et al., 2018). The SOEP data is a longitudinal panel data set of the population in Germany. The survey collects data on labor market experiences as well as on the financial situation of households and its members. Wealth data are available in the SOEP for every fifth year since 2002. Since the Social Security data end in 2020, we use four waves of wealth data: 2002, 2007, 2012, and 2017. We exploit the fact that the SOEP contains detailed information on labor market histories, including earnings, unemployment spells, and long-term sickness, to impute wealth in the Social Security data. Concretely we impute asset holdings in the *VSKT* by asset class based on earnings levels and lags thereof, the volatility of earnings in the last ten years, the number and length of unemployment spells, and sickness spells, age and place of residence (state). This information is available in both datasets. We rely on the social security data for the estimation of the earnings dynamics as the SOEP data lacks the sample size and we do only observe long labor market histories, but not the entire history of a worker. By combining data from the two data sources, we will link heterogeneity in earnings risks and portfolio choices.

2.1 Wages and earnings

The German pension data cover the entire employment history of individuals starting at age 14. For each months, the social security agency reports an employment status for each months based on the main activity in a month and the associated pension points that are a transformation of earnings for that month for all employed workers. In addition to this exceptional coverage, the data are highly accurate. The pension points determine old-age and disability benefits and are taken from the administrative records of employers' reports to the social security system and are sent to individuals to verify their recorded employment histories.[5] The data include typical demographic information such as age, gender, and education. For our analysis, we restrict the sample to male workers with all of their employment records in West Germany. We exclude workers with any employment information from the GDR period as these data are not comparable.[6] We restrict the analysis to men, as the labor market experiences of those women for whom we observe complete labor market histories in the data were still largely determined by a strongly changing institutional and cultural environment, which makes the study of their working lives an important and interesting research topic of independent interest. Finally, we

[5]In contrast to the widely used German social security data from the unemployment insurance records (Antoni, Ganzer, and Vom Berge, 2019), the retirement records are constructed by biography time rather than calendar time as the entire employment history is relevant to determine benefits. For our analysis this is important as it allows us to observe individuals' entire employment histories.

[6]Later, when we combine the pension records with SOEP data we cannot identify the place of work but only the place of residence. Therefore, we restrict the SOEP sample to workers who do not live in East Germany. This additional restriction affects very few workers. We also exclude workers with "Fremdrenten" points, i.e., workers who spent part of their working life in another country (mostly in the EU). The latter restriction does not affect many workers.

exclude (sworn-in) civil servants ("Beamte") and workers who were mainly self-employed throughout their lives.[7]

The data are representative of the universe of all workers covered by social security legislation, and earnings represent labor market income subject to social security taxes. From this group, we select workers in cohorts born between 1937 and 1960, that is, men who reach their 60th birthday between 1997 and 2020, so that we observe complete labor market histories for workers up to age 60. The combined data of the different annual samples includes about 120,000 workers for which we observe the full employment histories. Appendix A.1 provides details on the number of workers per year of birth.[8]

Earnings in the data are reported as pension points. These pension points express monthly earnings relative to macroeconomic annual average earnings, an index updated since 1957 based on gross earnings data from the national accounts. The fact that earnings are expressed relative to an annual average allows us to easily compare the dynamics of earnings over time by automatically removing the time trend of average earnings.[9] Some earnings observations are capped at the social security contribution limit, and we focus on uncapped earnings observations for most of our analysis.[10]

The data provide us with monthly gross earnings (in terms of pension points, hereafter: income), monthly employment status, and days of employment in the respective month. We aggregate the monthly data to an annual frequency by summing the monthly income and days of employment. The ratio of annual income to annual employment days gives the average annual wage.

Generally, we consider workers with labor income from all types of employment statuses. However, we often require workers to have at least some income in a year coming from employment. We identify employment based on the monthly employment information. In cases when workers are not employed, they might still collect pension points, for example, in case of unemployment as the employment agency pays contributions to the social security system on behalf of the worker.

[7]We identify (sworn-in) civil servants ("Beamte") using the fact that they do not contribute towards the public pension system, i.e., we exclude all workers where all observations after age 40 are missing. This is based on the implicit assumption that civil servants do not return to the public pension system at a later stage. Returning to the public pension system meant for the civil servant a significant loss in pension claims, so the incentives to avoid this were very high. Only in recent years, there have been changes in this system, but they only affect cohorts not in our data.

[8]Notably, not all of these workers have complete employment histories – neither in the sense of always positive earnings nor in the sense of not having missing data years, see below for details.

[9]Note that the construction of the data typically does not provide a representative cross section of all employed workers at each calendar date. This is due, for example, to changing demographics. This means that the average number of points at different calendar dates may differ significantly from one another. See Appendix A.2 for a discussion.

[10]We show the time series of the Social Security cap in the Appendix A.3. The cap has varied over time between 1.5 and 2.1 times average earnings (= points). The share of capped observations varies with age from less than one percent at age 25 to about 15 percent at age 45. We provide more details in the Appendix A.3.

Table 1 Descriptive statistics of the Social Security and SOEP samples

Variable	Mean	Median	P10	P25	P75	P90
Social Security Data						
Age	45	44	32	37	52	58
Income	1.11	1.15	0.16	0.80	1.53	1.85
SOEP Data						
Age	47	47	33	38	56	61
Income	1.01	1.04	0.00	0.62	1.44	1.86
Wealth	2.95	1.00	0.00	0.01	3.77	7.57
Illiquid Share (Narrow)	63.79	76.05	0.00	23.58	96.02	100.00

Notes: Income and wealth are reported in terms of pension points, that is, relative to the average annual earnings for the respective year. The illiquid share is reported in percent of total wealth. The sample is restricted to workers between age 30 and 65.

2.2 Wealth and portfolio data

We combine the social security data and the SOEP data by imputing SOEP wealth data in the social security data. We select the SOEP sample that is most comparable to the one we use for the social security data: we focus on male employees who do not live in East Germany. We do not use data from refugee samples and exclude individuals identified as self-employed or civil servants. All wealth information is at the individual level.[11] Table 1 reports descriptive statistics of the social security and the SOEP samples. We convert a worker's earnings in the SOEP data from euro amounts to pension points using the average annual earnings for the respective year.

We find that age and income align closely between the social security data and the SOEP data. Individuals in the SOEP data are on average 2 years older and their income is 10 percent lower as share of average income (0.1 pension points). Average income in the SOEP is very close to 1 and therefore to average macroeconomic income.

For the financial assets, we focus on the level and liquidity of wealth. We follow the construction of these variables as commonly used in the literature (Bhutta et al., 2020; Kuhn, Ríos-Rull, et al., 2016; Kuhn, Schularick, and Steins, 2020). We construct individual wealth as the net worth of the individual by taking the difference between the value of all assets and the value of all debt.[12] We consider two versions of a liquid-illiquid

[11] More specifically, we exclude individuals who were mainly self-employed or mainly civil servants during their lifetime. We declare an individual to be self-employed in a year if the annual income from self-employment is greater than the income from any other potential first job. We report an individual as a government employee based on an identifier measured in the SOEP.

[12] The SOEP collects wealth data on an individual's assets as well as liabilities. We calculate net real estate wealth (primary residence plus other real estate net of mortgages), business wealth, tangible assets,

split of assets. In the narrow definition, we consider only net real estate and insurance assets as illiquid. This is our baseline. In a robustness check (wide definition), we include savings in home loan and savings contracts in the illiquid assets. We compute portfolio returns based on average historical returns for Germany for the years 1970 to 2020, see Table A.2 in the Appendix.

We also transform wealth into pension points to simplify interpretation and to harmonize the unit of observation between income and wealth and across the SOEP waves we use. Table 1 reports descriptive statistics of wealth and the illiquid share. On average, individuals in the SOEP have wealth equal to three times average income. The median is with one annual income in wealth much lower and shows that the distribution of wealth is highly skewed. We also find that the bottom 25 percent of households have hardly any wealth. By contrast, the 90^{th} percentile of the wealth distribution is at more than 7.5 times average annual income. We also find that most of the wealth is illiquid (64 percent). More than half of workers have more than 3 out of 4 euros in illiquid wealth and almost 1 out of four workers holds all of his wealth in illiquid assets.

The overlap in terms of observed variables and their close alignment between the social security and the SOEP data allows us to combine the two datasets and to accurately impute wealth to the pension data. We describe the predictive mean matching algorithm below. The result is a data set that contains both highly detailed earnings histories and detailed wealth and portfolio information.

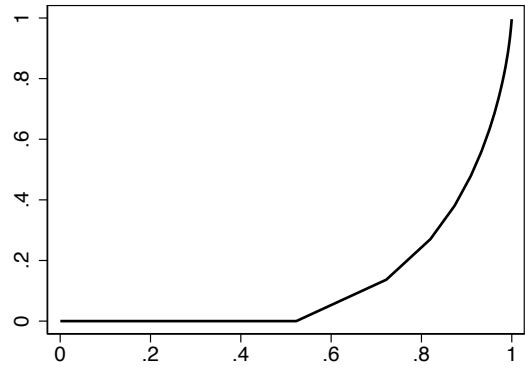
The social security data also allow us to observe two important life events for earnings dynamics: unemployment and (long-term) sickness. Unemployment spells are directly coded in the social security data. Sickness is observed as soon as a worker is sick for more than six weeks. This reflects that during the first six weeks of sickness a worker is entitled to the full salary paid by the employer. After six weeks, the worker is entitled to a replacement income covered by the public health insurance system. Only these long-term sickness spells are observed in the social security data. As we observe the entire employment history of workers in the social security data, we can compute the number of unemployment and sickness spells during a worker's working life. We use this information for as a first approach to document the heterogeneity in earnings dynamics. Figure 1 shows the Lorenz curves of accumulated unemployment and sickness spells of 50-year-old workers in the social security data.

The results are striking. Looking at the concentration at unemployment spells in Figure 1a, we find that more than half of the workers had no single unemployment spell

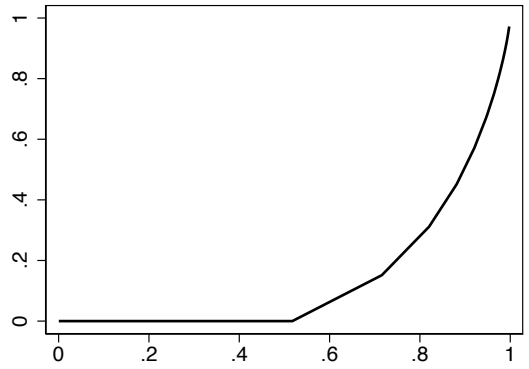
net liquid financial wealth (financial assets net of financial liabilities, see below), and net building loan and insurance wealth. Financial assets include savings accounts, bonds, stocks and investments. These may differ in terms of liquidity. However, since most of the financial assets of German households are liquid and there is no way of identifying, for example, shares, we classify the item financial assets as liquid in the following. Financial liabilities are loans other than mortgages and building loans, i.e., consumer credit.

Figure 1 Earnings and wage growth rate distribution

(a) Unemployment



(b) Sickness



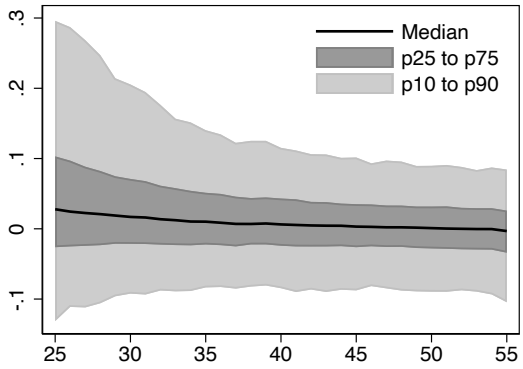
Notes: Lorenz curve of accumulated unemployment and sickness spells of 50-year-old workers in the social security data. Left panel shows the Lorenz curve of transitions into unemployment over their working life (unemployment spells) for workers at age 50. Right panel shows the Lorenz curve of all transitions into long-term sickness over their working life (sickness spells) for workers at age 50.

at age 50. At the same time, the 20 percent of workers with most unemployment spells account for 80 percent of all unemployment spells. Even more strikingly, when we look at the concentration of sickness in Figure 1b, we find almost the identical concentration of sickness spells. Again, more than half of the workers had no single long-term sickness spell at age 50 whereas 20 percent of workers again account for four out of 5 sickness spells in a cross-section of 50-year-old workers. Overlaying the two Lorenz curves shows that unemployment spells are slightly more concentrated than sickness spells but the difference is very small.

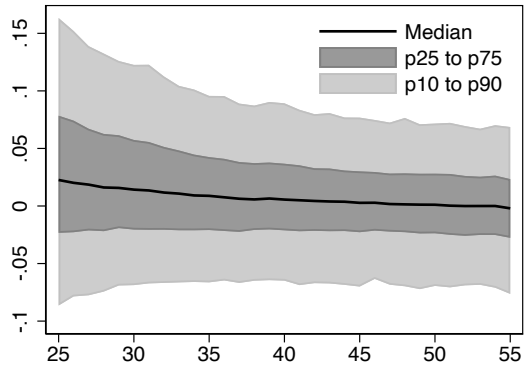
To impute asset information into the social security data we use a predictive mean matching algorithm. Specifically, we impute for each worker a wealth level, the two measures of the share of illiquid assets in a worker's portfolio, worker's wealth-to-income ration, and the worker's average portfolio return. As predictive regressors, we use year, age, state, unemployment, and sickness indicators, as well as the interactions between age and year, age and unemployment, age and sickness, and state and year. We also use cumulative unemployment and sickness spells. Finally, as main predictors, we use a large set of statistics on workers' earnings, in particular measures of current and past earnings levels, means and volatilities, as well as interactions of means and volatilities. We regress the variables of interest on these regressors in the SOEP and then predict the values in the pension data. Then, instead of using the predicted value, the predictive mean matching algorithm finds the closest neighbors of this predicted value and averages them. This captures nonlinearities in the data and results in a more similar distribution in the original and imputed data. For each worker, we compute ten different implicates by drawing from the error term distribution of the predictive regression. Further details and quality checks for the imputation are reported in the Appendix A.5.

Figure 2 Earnings and wage growth rate distribution

(a) Earnings growth by age



(b) Wage growth by age



Notes: Percentiles of the annual earnings (the left panel) and wage (the right panel) growth rate (log-differences) distribution by age (in terms of pension points, i.e., relative to aggregate earnings growth). See Section 2 for details on the sample and computation. For wages, we focus on fulltime employment.

Before using the imputed information, we first zoom in on the labor market part described in Section 2.1 and study workers' labor market risks in detail.

3 Heterogeneity and persistence in earnings risk

Based on these data, we first document the earnings and wage growth distribution over the life cycle. Second, we show that there is a large, significant, and persistent heterogeneity in earnings risk across workers. Third, we find that risk and earnings levels are negatively correlated. In this sense, there is no risk-return trade-off for human capital between workers (this does not exclude that each worker faces such a trade-off). Ultimately, we show that risk heterogeneity also has sizable effects on workers' portfolios.

3.1 Wage and earnings growth dispersion over a worker's life

Figure 2 shows percentiles of the distributions of earnings and wage growth (as log differences) by the age of the worker. Figure 2a shows earnings growth rates and Figure 2b shows wage growth rates.[13] Importantly, our evidence comes from observing entire working lives, rather than merging short panels of workers as in most of the literature.

Looking at the median growth rate, we find a declining, slightly convex profile over the working life for both wages and earnings. The peak is at entry and then the median growth

[13]To compute wage and earnings growth between age j and age $j+1$, we compute growth rates as log first differences (log growth rates). We require that workers work at least 90 days in both years. Thus, to compute growth rates between two years, we require that workers are at least partially attached to the labor market in both years. We also require that income is not censored in both years. For the annual data, we compute annual income by multiplying annual pension points with the average annual income. If this implied income is above 99% of the annual social security contribution limit, we set the censoring indicator to one.

rates decline to a slightly negative value at age 55. The declining convex growth rate profiles provide the commonly observed concave life-cycle profiles of wages and earnings levels (for Germany, see Bayer and Kuhn, 2019).

More importantly, both figures highlight the overall large dispersion of growth rates across workers, with the dispersion of earnings growth rates being larger than that of wage growth rates. There is also a clear age pattern. The dispersion of wage and earnings growth rates is highest after labor market entry and declines rapidly within 5 years. The dispersion decreases with age until around the age of 40 and then remains fairly constant until shortly before retirement.

For earnings, we observe that the 90th and 10th percentiles differ by about 40 percentage points at labor market entry. For wages, this dispersion is still large (over 20 percentage points), but much smaller than that for earnings. The difference highlights the importance of employment risk in the years after labor market entry (Jung and Kuhn, 2018). The differences in the dispersion of wage and earnings growth decrease with age and almost converge in the second half of the working life. This implies that earnings risk is dominated by wage risk later in life, when employment risk is less of a concern.[14]

As workers age, the distributions of wage and earnings growth rates not only become more similar and compressed, but also more symmetric. While the distributions are strongly right-skewed at labor market entry, this skewness is largely lost in the second half of the working life. The changing growth rate distribution suggests an underlying dynamic in the earnings process over the life cycle (Karahan and Ozkan, 2013). The income growth dispersions between age 25 and 55 are dwarfed by that outside these bounds because younger and older workers have a looser labor market attachment (not shown).

3.2 Non-Gaussian income growth distribution and risk types

Although the growth rate distributions become more symmetric around age 40, they are still significantly over-dispersed.[15] In fact, there is a surprisingly large mass of workers with earnings growth rates close to zero, that is, close to the average earnings growth in the economy.

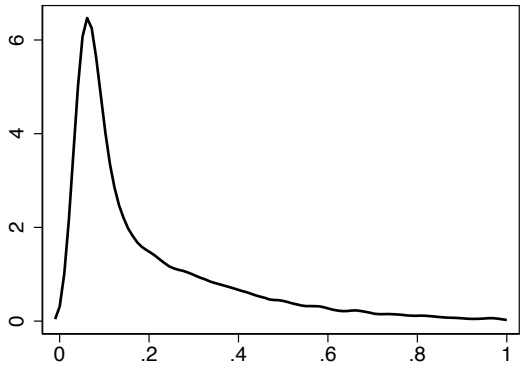
A standard way to model such an earnings growth distribution is to assume mixing. Workers, or some workers, sometimes draw from a high-variance distribution and sometimes from a low-variance distribution (Guvenen et al., 2021). Aside from being a modeling device, we can also assign an economic interpretation to such mixing. Some

[14] The inter-quartile ranges for wage and earnings growth rates are still large, at just over 10 percentage points, but of a similar magnitude at labor market entry. In general, the decline in earnings variability is greater than that for wages, and the decline is most pronounced between ages 25 and 30.

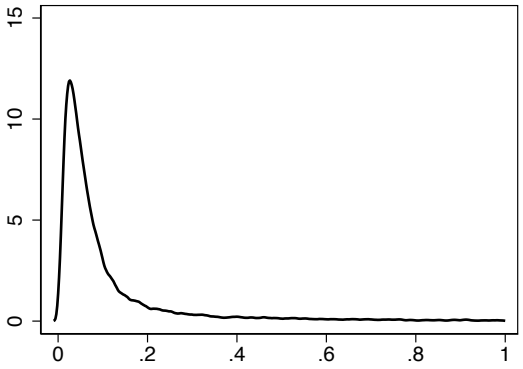
[15] The strong departure from a normal distribution has been emphasized by Guvenen et al. (2021) for the distribution of earnings growth in the United States. Busch et al. (2022) and Drechsel-Grau et al. (2022) provide recent evidence for Germany.

Figure 3 Distribution of realized earnings growth rate volatility

(a) Lifetime earnings growth volatility



(b) 5-year volatility at age 40



Notes: The left panels show the distribution of the realized earnings growth volatility during the working life, computed as the standard deviation of annual earnings growth rate between age 25 and 55. The right panel shows the distribution of the realized short-run earnings growth volatility at age 40, computed as the 5-year rolling standard deviation of the annual earnings growth rate. Kernel density estimates are shown. See Section 2 for details on the sample and computation.

workers go through a more turbulent labor market than others. An interesting question in applying this interpretation is whether the risk types are transitory, persistent, or permanent. For simplicity, in the rest of the paper, we focus on the earnings growth distribution and approximate it by a mixture with only two risk types.[16]

Our data offer a unique opportunity to take this economic interpretation seriously and to ask whether there are indeed distinctive risk types of workers; whether some workers face persistently more labor market turbulence. To understand the persistence and cross-sectional heterogeneity of earnings risk, we, first, look at the distribution of (within-person) earnings growth standard deviations over the entire working life[17] and, second, look at the (rolling) 5-year standard deviations of (within-person) earnings growth rates.[18] Figure 3a and Figure 3b provide estimates for the distributions of earnings-risk estimates across workers.

Looking first at the lifetime volatility in Figure 3a, we find a large dispersion of earnings growth volatilities. On average, the lifetime standard deviation of individual earnings growth rates is 0.20. The median is 0.12, reflecting a large heterogeneity in volatilities. At one end of the distribution, one in ten workers has a standard deviation of less than 0.05. At the same time, the ten percent of workers with the most volatile earnings have a standard deviation of annual earnings growth greater than 0.46. The 90-50 ratio of the

[16]This captures most of the features of the earnings growth distribution and simplifies the discussion. Allowing for more worker types will mechanically improve the fit of the data, especially with respect to the tails of the distribution.

[17]We require at least 20 non-missing growth rate observations within a worker's working life.

[18]Specifically, we compute for each worker at each age j the standard deviation of growth rates using growth rates at ages $j, j-1, \ldots, j-4$. We require that all growth rate observations for the current 5-year window are non-missing.

distribution is 3.8, indicating that at least 10 percent of workers have a working life that is more nearly four times as volatile as that of the typical (median) worker.

These fluctuations in earnings growth over the lifetime could reflect periods of predictable high or low earnings growth. To filter these out, we also look at a standard deviation over shorter horizons, calculating for each worker the standard deviation of earnings over a 5-year window at age 40. Figure 3b shows the distribution of these measures. Strikingly, there is now even a mass of workers with (almost) zero earnings volatility, suggesting that a substantial fraction of workers experience wage growth rates at age 40 that have almost no volatility. The distribution of volatilities at age 40 is even more skewed than the lifetime distribution. The average volatility is 0.10, the median is 0.06, and the 10 most volatile percent face a standard deviation of wage growth of 0.24. This means that the 90-50 ratio is about 4.

This higher dispersion of volatilities at a given age relative to lifetime suggests some mean reversion in risk. At the same time, the large dispersion of lifetime volatility is hardly consistent with purely transitory episodes of high earnings volatility.

3.3 Income process with time-varying risk types

To study the persistence of risk heterogeneity, we use a model of an income process where income shocks are drawn from normal mixtures. However, we assume that the mixture is not necessarily i.i.d., but that the probability of drawing from a distribution is driven by a hidden state space/Markov switching model.

Concretely, we estimate the following model of workers' earnings. As usual, we decompose worker i's log earnings into a life-cycle component and a residual $e_{i,t}$ that consists of a persistent component $\eta_{i,t}$ and a transitory component $\tau_{i,t}$. Often, the literature assumes a model of homogeneous parameters and risk across workers. That is,

$$\begin{aligned} \eta_{i,t} &= \rho\,\eta_{i,t-1} + \mu_{\varepsilon,0} + \sigma_{\varepsilon,0}\varepsilon_{i,t} && \text{with } \varepsilon_{i,t} \sim \mathcal{N}(0,1) \\ \tau_{i,t} &= \mu_{\xi,0} + \sigma_{\xi,0}\xi_{i,t} && \text{with } \xi_{i,t} \sim \mathcal{N}(0,1). \end{aligned}$$

Instead, we allow for the income shocks to be drawn from a mixture of two normals. If in state $s_{it} = 0$, income shocks are drawn from a low-risk distribution and in state $s_{it} = 1$ they come from a high-risk distribution. The transitions between the two types are stochastic. The earnings process for worker i is then

$$\eta_{i,t} = \rho\,\eta_{i,t-1} + \mu_{\varepsilon,s_{i,t}} + \sigma_{\varepsilon,s_{i,t}}\varepsilon_{i,t} \qquad \text{with } \varepsilon_{i,t} \sim \mathcal{N}(0,1) \tag{1}$$

$$\tau_{i,t} = \mu_{\xi,s_{i,t}} + \sigma_{\xi,s_{i,t}}\xi_{i,t} \qquad \text{with } \xi_{i,t} \sim \mathcal{N}(0,1) \tag{2}$$

with $s_{i,t} \in \{0,1\}$ and Markov transition matrix Π. Importantly, it nests both the normally distruted income shocks case when parameters are identical across s and also the case of

the usual mixture of two normals, when the probability distribution of s_{it} is independent from s_{it-1} (s being i.i.d.).

This specification of the income process falls in the class of regime-switching models studied in Hamilton (1988, 1989), Kim (1994), and Kim and Halbert (1999). What we add, however, is the cross-sectional dimension of workers. Given that the income processes of workers are assumed to be independent (after taking out time fixed effects in the first stage) and parameters are constant across workers, we can construct the log-likelihood of the data by summing over workers. To the best of our knowledge, this is the first time such a model has been estimated on micro data to estimate a complex earnings process. A sensible estimation is only possible with the long earnings histories we have at hand.

We estimate this model using an panel-augmented version of the Kim filter – which, itself, is an extension of the Kalman filter. With the Kim filter, we compute the likelihood function which we then maximize using maximimum likelihood estimation.[19] For this, we pose the model in state-space form.

$$e_{i,t} = \eta_{i,t} + \mu_{\xi,s_{i,t}} + \xi_{i,t} \qquad \text{with } \xi_{i,t} \sim \mathcal{N}(0, \sigma^2_{\xi,s_{i,t}}) \tag{3}$$

$$\eta_{i,t} = \mu_{\epsilon,s_{i,t}} + \rho\eta_{i,t-1} + \varepsilon_{i,t} \qquad \text{with } \varepsilon_{i,t} \sim \mathcal{N}(0, \sigma^2_{,\varepsilon,s_{i,t}}) \tag{4}$$

with (3) being the observation and (4) the state space model itself. In addition, the regimes $s_{i,t} \in \{0, 1\}$, the transition matrix Π defined as

$$\Pi = \begin{bmatrix} \pi_{00} & 1 - \pi_{00} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix}$$

with $\pi_{00}, \pi_{11} \in [0, 1]$. We have estimated the model with state dependent transitory or persistent income level shifts $(\mu_\varepsilon, \mu_\xi)$ separately and have choosen the one with the higher likelihood. The data prefers the version with a transitory intercept.[20] For all models, we enforce that the unconditional mean of the process is zero (recall, we standardized the data beforehand). For that, we only estimate $\mu_{\xi,0}$ and assume $\mu_{\xi,1} = -\pi_0\mu_{\xi,0}/\pi_1$, where π_0 and π_1 are the steady-state probabilities of the two types.

For the estimation, we use a sample of male workers with complete and always positive earnings between age 30 and 50. Additionally, we require that workers do not have earnings below half of the poverty line in more than two years between age 30 and 50. This is to ensure that we do not include workers with very low earnings which most likely are not attached to the labor market.

Table 2 shows the estimation results for several model versions. Column (1) shows the model with just one type. Column (2) shows the case with two permanent types. Column (3) the one with i.i.d. switching between the two types. Column (4) finally shows the full

[19] The results are robust to using Bayesian estimation methods.

[20] Results with intercepts on the persistent component are available upon request.

Table 2 Estimation results for the earnings process of workers

Parameter	(1) 1 type	(2)	(3) 2 types	(4)
		Permanent	i.i.d.	Persistent
π_{00}	–	–	0.892	0.934
	(–)	(–)	(0.891, 0.892)	(0.933, 0.934)
π_{11}	–	–	–	0.578
	(–)	(–)	(–)	(0.574, 0.582)
ρ	0.921	0.946	0.977	0.975
	(0.921, 0.922)	(0.945, 0.946)	(0.977, 0.978)	(0.975, 0.975)
$\sigma_{\varepsilon,0}$	0.101	0.069	0.041	0.041
	(0.100, 0.101)	(0.069, 0.069)	(0.040, 0.041)	(0.041, 0.041)
$\sigma_{\varepsilon,1}$	–	0.180	0.205	0.163
	(–)	(0.178, 0.181)	(0.204, 0.206)	(0.161, 0.164)
$\sigma_{\xi,0}$	0.057	0.045	0.009	0.011
	(0.057, 0.057)	(0.045, 0.046)	(0.009, 0.009)	(0.010, 0.011)
$\sigma_{\xi,1}$	–	0.128	0.225	0.212
	(–)	(0.127, 0.129)	(0.223, 0.226)	(0.211, 0.214)
$\mu_{\xi,0}$	-0.001	0.031	0.012	0.018
	(-0.003, 0.000)	(0.030, 0.032)	(0.012, 0.012)	(0.018, 0.019)
Log-likelihood	652,941	835,749	1,175,260	1,193,411

Notes: Estimates from maximum likelihood estimation of the simple and regime-switching model of earnings dynamics, likelihood computed using the Kalman filter and Kim filter, respectively. The numbers in parentheses show the 90-% confidence intervals.

Data Source: VSKT. West German male workers born between 1937 and 1960, excluding civil servants and self-employed workers. We use their earnings histories between age 30 and 50 and require complete and always positive earnings between age 30 and 50, including unemployment benefits. We exclude workers that earn less than half the poverty line in more than two years between age 30 and 50.

model. For all of these cases, the model with two types and transitions has the highest likelihood. The second-best model is the one with i.i.d. switching between the two types (the standard Gaussian mixtures model).[21]

We find that high-risk workers ($s = 1$) have not only more volatile income but also lower incomes. The estimated transition probabilities imply that each period, about 6.6 percent of low-risk workers become high-risk and 42.1 percent of high-risk workers become low-risk. On average, a worker remains high-risk for 2.4 years and low-risk for 15.2 years. This implies that approximately 13.5 percent of workers are high-risk at any given time.

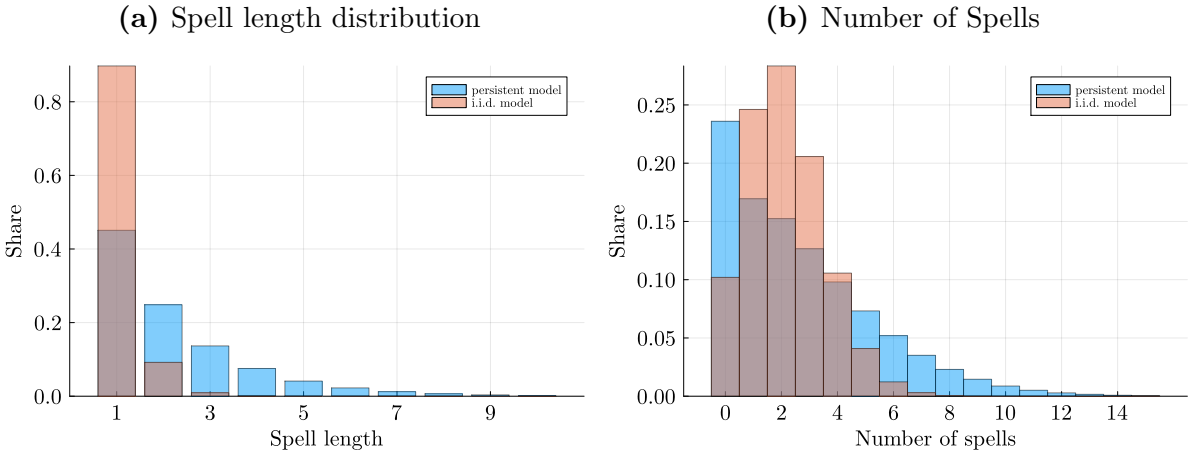
There are economically significant differences in income risk across risk states, with

[21] A formal test would take the form of a likelihood ratio test. Denote as ℓ_R the restricted and ℓ_U the unrestricted log-likelihood model. The degrees of freedom of the test are the difference in the number of parameters between the two models, $\text{df} = \dim(\theta_U) - \dim(\theta_R)$. The test statistic is then

$$\lambda_{LR} = 2\left(\ell_U - \ell_R\right) \sim \chi^2_{\text{df}}.$$

The differences in likelihood are significant at all usual levels of significance.

Figure 4 Comparison of the i.i.d. and persistent model in terms of high-risk spells



Notes: Spell length and number of spell distribution over a 20 year horizon, initializing the distribution of workers over high- and low-risk states according to the ergodic distribution of the estimated Markov chain, see Table 2.

similar point estimates. Persistent income is approximately four times more volatile for a high-risk worker than for a low-risk one. For the transitory component, the difference is even more pronounced: low-risk workers face close to no transitory volatility, while high-risk ones face a standard deviation of 0.21.

Interestingly both the risk differences across states as well as the average share of workers in the high risk state are not very different in the model with persistent risk from the standard mixture model (i.i.d.). However, in terms of the distribution of risk across workers there are large differences across the two models of the earnings process. In the i.i.d. model, the expected duration of the high risk state is, with 1.1 years, 55% shorter compared to the persistent risk model. Looking at tail-risk, the difference becomes even starker. In the persistent risk state model, conditional on entering the high-risk state, 11.8 percent of workers remain in the high-risk state for five years or more. In the i.i.d. model, this is true for only 0.01 percent of workers. Figure 4 compares the two models in terms of the distribution of the lengths of high-risk spells (conditional on being in the high-risk state) and the number of high-risk spells a worker faces in 20 years between the ages of 30 and 50. The model with persistent risk produces a much more unequal distribution of high-risk phases. With persistent risk types, roughly a quarter of workers are never in the high-risk state in their working life. At the same time, 22 percent of all workers experience five or more years of high risk within these 20 years. For the standard mixture model (i.i.d.), only ten percent of workers never face high risk and only six percent of workers would experience five or more years of high risk.

3.4 Earnings risk, health, unemployment, and wealth

One advantage of the state-space model is that we can use the Kim smoother to obtain (among other things) smoothed probabilities of being in the high-risk state for each worker at each age. We use these probabilities as a continuous measure of risk in the following. We calculate two measures of risk for a worker i at time t. First, the probability that the worker is in risk state $p_{it} = P(s_{it} = 1)$, second the average probability to have been in in the high-risk state until time t, $\bar{p}_{it} = \frac{1}{t}\sum_{s=0}^{t} p_{it}$. To minimize the impact of the initialization of the smoother at age 30, we use the smoothed probabilities from age 35 on in our regressions.

First, we ask how the risk type relates to earnings without enforcing the specific model we estimate. Further, we show that high risk workers are more often in unemployment and also are more often on long-term sickness leave. Table 3 shows the results of this exercise, a regressions of earnings, cumulative earnings, unemployment months, and sickness months on our two measures of income risk: current and average risk. The regression coefficients on the current risk status can be interpreted as hypothetically assigning a probability of one to the high-risk state in period t, the regression coefficient

In line with the estimation results from Table 2, Table 3 shows that all measures of risk are negatively correlated with both earnings and cumulative earnings. The effects are statistically significant and sizable. A hypothetical worker who would for his entire life certainly be in the high-income-risk state (having an average probability of one at age 50) would have earned 11.6 pension points (annual average incomes) lower income cumulatively between age 30 and 50. Risk is positively correlated with being unemployed or in long-term sickness leave. This holds true for the current employment/health status, as well as for the cumulated number of months in unemployment/sickness, and also for the number of spells. Again looking at the cumulative effects of hypothetically being permanently with probability one in the high-risk state, we find that this permanent-high-risk worker has 62 extra months (7.6 extra spells) of unemployment and 20 extra months (4.5 extra spells) of sickness.

Table 3 Earnings risk, earnings levels, unemployment, and sickness

	(1) Earnings	(2) Unemployment	(3) Sickness	(4) Cum. earnings
Risk measure	-0.35***	2.53***	1.03***	-11.65***
	(-204.29)	(148.53)	(94.90)	(-82.72)
	(5) Cum. unemployment	(6) Cum. sickness	(7) Cum. spells unemployment	(8) Cum. spells sickness
Risk measure	61.98***	19.59***	7.57***	4.46***
	(74.64)	(46.26)	(64.95)	(48.01)

*** significant at the 1% level.

Regressions: Results of regressions of earnings in year t, months of unemployment in year t and months of sickness in year t on the current risk (all between age 35 to 50) – columns (1) to (3) – as well as the cumulated earnings up to year t, cumulated months in unemployment and cumulated months of sickness up to year t, and the cumulated number of spells of unemployment and sickness up to year t on the lifetime average risk (all at age 50) – columns (4) to (8). Additional controls are year, age, and state dummies. Robustly estimated T-statistics in parentheses.

Definitions of variables: Earnings are expressed in "pension points", i.e., relative to the average labor earnings in Germany in that year. Unemployment and sickness are expressed in months. Cumulated earnings are the sum of earnings up to year t (starting at age 30). Cumulated months in unemployment and sickness are the sum of months in unemployment and sickness up to year t (starting at age 30). Cumulated spells of unemployment and sickness are the sum of spells up to year t (starting at age 30). The current risk is imputed by the Kim-smoother as the probability of being in the high risk state. We drop the first five years of the sample to avoid an initial imputation bias and hence use the probabilities between age 35 and age 50. The lifetime average risk is the average of the high risk probabilities from age 35 to 50.

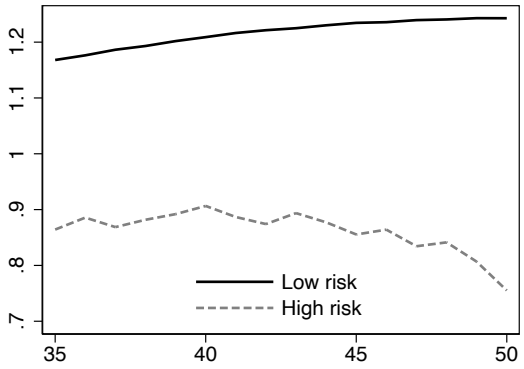
Data source and sample: VSKT data, birth cohorts 1937 to 1960, ages 35 to 50. For the regressions on current risk, we have 794,624 observations. For the regressions on lifetime average risk, we have 49,664 observations.

Figure 5 repeats the estimation exercise in a slightly more flexible way and reports the results graphically. We regress cumulative earnings, unemployment, and sickness on an age dummy and an interaction of average risk with the age dummy. Then we plot the life-cycle pattern for two hypothetical workers: one who never was in the high-risk state and one worker who always was in the high-risk state. The low-risk worker has higher earnings and accumulates virtually no unemployment and sickness spells, while the high-risk one sees a significant and slightly convex pattern of accumulated unemployment and sickness spells and months. Of course, one needs to take into account that these two worker types would be hypothetical in that the probability of being 20 years in the high-risk state is basically zero even in our model of persistent income risk. At the same time, we have seen that never being in the high-risk state is something a quarter of workers experiences (based on the persistent risk model estimate).

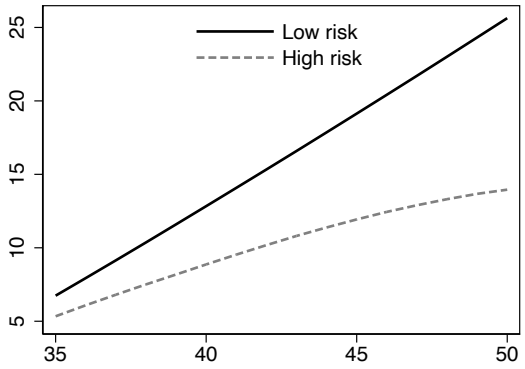
Next, we ask whether workers respond systematically to the fluctuations in risk they face. Economic theory suggests that workers facing higher income risk would save more and make their portfolios more liquid. At the same time, workers would be expected to

Figure 5 Lifetime income by risk group

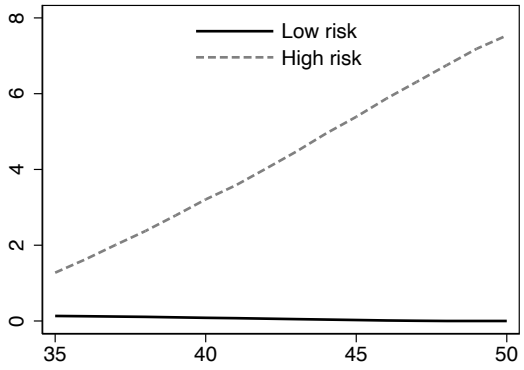
(a) Earnings by age



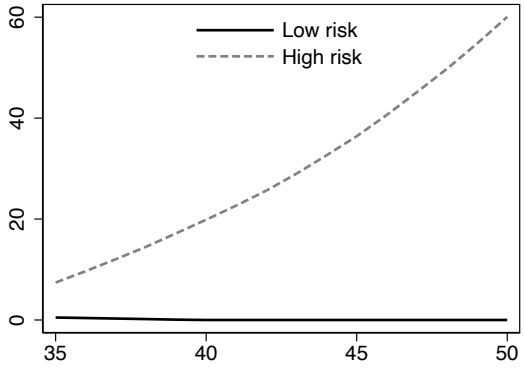
(b) Cumulative earnings by age



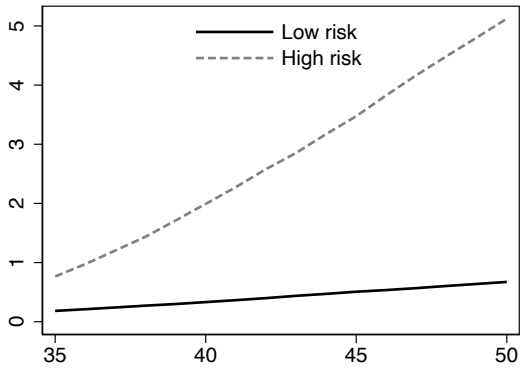
(c) Cum. unemployment spells by age



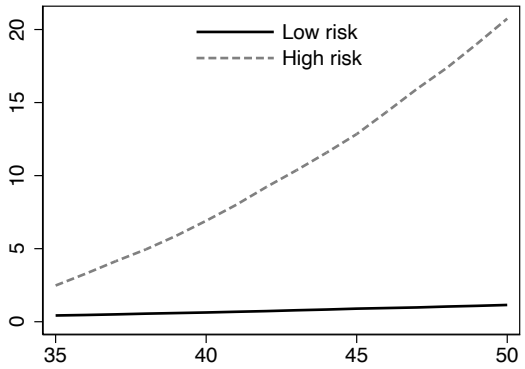
(d) Cum. unemployment duration by age



(e) Cum. sickness spells by age



(f) Cum. sickness duration by age



Notes: Estimated life-cycle profiles from a regression of the outcome variable on an age dummy and a risk-age dummy interaction. High risk refers to a hypothetical worker who is always in the high risk state, low risk to a worker always in the low risk state. The top left panel uses current risk at age t. All other panels use average risk up to age t as risk measure. Risk measures in the data from running the Kim smoother on the VSKT data. The top row looks at earnings measured in pension points. The center row at unemployment (left: cumulated number of spells, right cumulated duration in months). The bottom row considers long-term sickness leave (left: cumulated number of spells, right cumulated duration in months).

Table 4 Earnings risk, wealth, and portfolios

	(1) Net wealth	(2) Illiquid share	(3)	(4) Portfolio return
		Narrow	Wide	
Risk measure	-2.76***	-19.34***	-8.46***	-1.38***
	(-49.36)	(-28.52)	(-13.04)	(-38.30)

*** significant at the 1% level.

Regressions: Results of regressions of net wealth in year t, portfolio return in year t and several measures of the illiquid share in year t on the lifetime average risk (up to year t). Additional controls are year, age, and state dummies. T-statistics in parentheses, standard errors are clustered at the worker level (using ten wealth implicates per worker).

Definitions of variables: Net wealth is expressed in "pension points", i.e., relative to the average labor earnings in Germany in that year. Illiquid shares are expressed in percent. Portfolio returns are expressed in percentage points. Risk is imputed by the Kim-smoother as the probability of being in the high risk state. We use the average of the high-risk probability from age 35 to up to time t. We restrict the sample to workers that are between 45 and 55 years. If a worker is older than 50, we use the risk measure of age 50. Narrow illiquid asset definition: real estate wealth and life-insurance contracts net of mortgaged debt. Wide definition includes savings in home loan and savings contracts.

Data source and sample: VSKT data, with imputed wealth based on the SOEP data, see Section 2.2. Birth cohorts 1937 to 1960, ages 45 to 55. Wealth data is only available in years 2002, 2007, 2012, and 2017.In total 618,120 observations (61,812 workers with 10 wealth implicates each).

respond to the negative realization of the income shock associated with the transition to the high-risk state by drawing down their (liquid) wealth. Thus, in theory, the exact wealth and portfolio response is not unambiguous. To explore the relationship in the data, we use imputed wealth in the VSKT (based on estimates from the SOEP). Specifically, we regress net worth, portfolio liquidity, and average portfolio returns for each individual worker on his risk type and a set of fixed effects for year, age, and state.

Table 4 shows the results of this regression. We focus on roughly 50-year-old workers (ages 45-55) in the years for which the SOEP contains wealth information. As a measure of risk, we use the average high risk probability up to the year of the survey. This means that we can interpret the regression coefficients again as the effect a hypothetical worker faces who was in the high-risk state all his life. Such worker has 2.76 annual average labor income less of wealth (at age 50), using the 2025 average German labor earnings this amounts to approximately €140,000.

Looking at the share of illiquid assets in their portfolio, we see that the high-risk worker has an eight to 19 percentage point lower share of illiquid assets in his portfolio, depending on whether we count building savings-loan contract as liquid or not. Consistent with this, we also find economically and statistically significant effects on portfolio returns. Because more liquid and safer assets have lower returns, and because high-risk workers hold a larger share of them, the high-risk worker has a 138 basis point lower return of his assets.

4 Life-cycle choices of heterogeneous risk types

The fact that high-risk households accumulate less wealth is surprising at first glance given that they would have a stronger precautionary savings motive. However, the fact that high-risk episodes often come in conjunction with low-earnings can possibly explain this surprising fact. To check whether the differences in risk are in line with forward-looking behavior of households, we build a life-cycle model of agents who are heterogeneous with respect to their labor earnings risk. Risk heterogeneity in this model matters for portfolio choices, and thereby affects wealth accumulation over the life cycle. We calibrate the model, inform it by our empirical results, and use it to obtain quantitative results for the consequences of risk heterogeneity.

Our model is an incomplete markets life-cycle model with portfolio choice. Agents in the model are risk averse and maximize their expected life-time utility, applying a discount factor β. The life cycle consists of J periods, indexed by $j = 1, ..., J$. We distinguish two phases of the life cycle, a working phase $j = 1, ..., J_{ret} - 1$, where agents work and receive stochastic labor income, and a retirement phase $j = J_{ret}, ..., J$, where they receive retirement benefits. Agents in the model make investment decisions to self-insure against idiosyncratic income fluctuations and to smooth consumption over the life cycle.

In every period of their life cycle agents can invest in two assets. These assets differ in terms of their return, risk, and liquidity. The first asset is a safe and liquid asset, denoted by a, with a risk-free return π, which is perfectly liquid in the sense that it can be traded without any frictions. The other asset is an illiquid risky asset, denoted by k. The valued stock of the illiquid asset is exposed to shocks δ and evolves therefore stochastically, thereby driving its risky return which fluctuates around its mean return r. The illiquid asset is subject to a trading friction. Agents can invest in this asset only by committing to a saving plan d, which can be changed at a utility cost κ. Borrowing is not allowed in either asset. The return on the liquid asset is lower than the mean return on the illiquid asset, $\pi < r$, thus inducing a trade-off between return and liquidity.

The higher average rate of return makes the illiquid asset an attractive vehicle for accumulating wealth over the life-cycle. However, holdings of the liquid asset are crucial for smoothing consumption in the presence of earnings risk. The extent and structure of earnings risk will therefore affect how agents trade off the benefits from each of the assets available. The resulting portfolio composition will therefore depend on the risk types of agents.

Agents face heterogeneous earnings processes distinguished by their risk type and whose calibration is explained in detail below. In the following formal description of the life-cycle decision problem in its recursive form $y(y_{p,j}, y_{\tau,j}, j)$ denotes earnings, allowed to depend on a persistent state $y_{p,j}$, a transitory state $y_{\tau,j}$, and on age j. This formulation

accommodates persistent risk types by considering the risk types as part of the definition of relevant persistent states.

4.1 Recursive formulation

We now formulate the agent's decision problem recursively. The state of an agent is described by age j, liquid asset position a_j, illiquid asset position k_j, current savings plan d_j, and income states $y_{p,j}$ and $y_{\tau,j}$. To ease exposition, we drop any indexation, say i, referring to an individual, since all state variables of the individual decision problem are individual by nature and only index by age.

The decision problem is framed as featuring a discrete choice between the two branches of non-adjustment and adjustment of the saving plan. On each of these two branches the decision problem then specializes in determining the appropriate financial decisions, and thereby the corresponding level of consumption.

Conditional on the discrete choice of not adjusting the saving plan, the recursive problem is the following:

$$\begin{aligned} V_{N,j}(a_j, k_j, d_j, y_{p,j}, y_{\tau,j}) &= \max_{(c_j, a_{j+1})} u(c_j) + \tilde{\beta}_j \mathbb{E}_{p,\tau} V_{j+1}(a_{j+1}, k_{j+1}, d_{j+1}, y_{p,j+1}, y_{\tau,j+1}) \quad (5) \\ s.t. \quad & c_j + a_{j+1} = (1+\pi)a_j + rk_j + y(y_{p,j}, y_{\tau,j}, j) - d_{j+1} \\ & k_{j+1} = (k_j + d_{j+1})\delta_{j+1}, \quad d_{j+1} = d_j \\ & a_{j+1} \geq 0 \end{aligned}$$

Non-adjustment of the saving plan, leaving the state of the saving plan unchanged, $d_{j+1} = d_j$. This choice is feasible only if the implied illiquid asset position is not negative, $k_{j+1} \geq 0$. Non-adjustment the saving plan will still lead to accumulation of illiquid wealth according to the previously chosen plan d_j. In this formulation $u(c_j)$ is utility from current-period consumption and $\tilde{\beta}_j$ is the effective discount factor, i.e. the product of the discount factor β and the probability of surviving from age j to age $j+1$.

Conditional on the discrete choice of adjusting the saving plan, the recursive problem is as follows:

$$\begin{aligned} V_{A,j}(a_j, k_j, d_j, y_{p,j}, y_{\tau,j}) &= \max_{(c_j, a_{j+1}, d_{j+1})} u(c_j) + \tilde{\beta}_j \mathbb{E}_{p,\tau} V_{j+1}(a_{j+1}, k_{j+1}, d_{j+1}, y_{p,j+1}, y_{\tau,j+1}) \quad (6) \\ s.t. \quad & c_j + a_{j+1} + d_{j+1} = (1+\pi)a_j + rk_j + y(y_{p,j}, y_{\tau,j}, j) \\ & k_{j+1} = (k_j + d_{j+1})\delta_{j+1} \\ & a_{j+1} \geq 0, \quad k_{j+1} \geq 0 \end{aligned}$$

The choice d_{j+1} of a flow for the saving plan has an immediate impact on the budget

constraint relevant for financing consumption from liquid resources. In case of future non-adjustment of the saving plan the same impact will remain present, as long as the amount committed to this per-period flow is unchanged.

The state-dependent value of the problem, which takes into account both branches, is obtained by considering the following maximization:

$$V_j(a_j, k_j, d_j, y_{p,j}, y_{\tau,j}) = \mathbb{E}_\zeta \max\{V_{N,j}(a_j, k_j, d_j, y_{p,j}, y_{\tau,j}) + \zeta_{N,j}, \\ V_{A,j}(a_j, k_j, d_j, y_{p,j}, y_{\tau,j}) - \kappa + \zeta_{A,j}\} \tag{7}$$

For the discrete choice of adjustment the utility cost parameter κ is taken into consideration. We allow for random components in the valuation of discrete choice options. Using the assumption that these random components are distributed according to an extreme-value (type I) distribution, the relevant expectation is obtained by the familiar log-sum formula for such a discrete choice problem.

4.2 Calibration

We calibrate the model to German data. We set the model period to correspond to a year and assume a constant relative risk aversion utility function with risk aversion parameter $\gamma = 2$. Workers work for 45 years and then spend (subject to uncertain survival) up to 20 years in retirement.

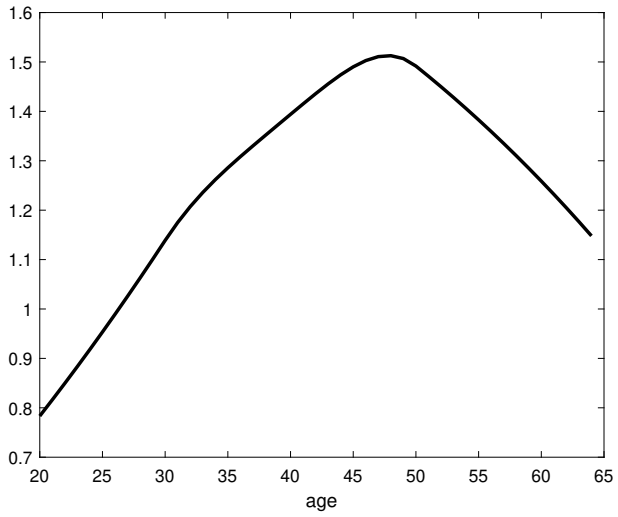
In the working phase agents face a stochastic process for gross log-income y, composed of a deterministic life-cycle component and a residual stochastic component. For the process of the stochastic component the calibration uses the estimates for the specification with Markov switching between risk types, as already described in detail in Section 3.3. We use the estimated parameter values of the "persistent" risk types model of Table 2. When numerically solving the model, we approximate the estimated earnings processes by a discrete Markov chain using the method proposed by Tauchen (1986).

The deterministic age earnings-age profile is based on a fourth-order polynomial of age, estimated on the data for ages 30 to 50, and extrapolated based on derivatives of the polynomial to cover all working ages. In the life-cycle decision problem we let individuals expect aggregate growth at a rate of 1%. Figure 6 shows the resulting life-cycle profile of earnings.

German retirement accounts provide only information on earnings before taxes. To convert gross earnings to net earnings, we apply a proportional tax rate of 45%.

The calculation of retirement benefits uses some parameters from Hintermaier and Koeniger (2018), namely the adjustment factor for pre-retirement earnings (the valorisation rate), set to 1%, and the number of earning years considered for the determination of retirement benefits, set to 35 years. The approximation of pension benefits is based on the last pre-retirement income state, considering the average working-age earnings according

Figure 6 Life-cycle earnings profile



to the distribution of paths of earnings which lead to the relevant terminal income state. We then apply a replacement rate of 58%.

The returns on both assets are calibrated using average historical returns between 1970 and 2020. We use the real results in the table in Appendix A.4. Therefore, we set the return on the liquid asset to -0.3% and the return on the illiquid asset to 5%.

We use the historical housing returns data for Germany from Jordà et al. (2019) to calibrate the volatility of the return on the illiquid asset. We find that the standard deviation of the housing return between 1970 and 2020 equals 3.4%.

We use the empirically observed patterns of wealth accumulation over the life cycle to pin down the value of the discount factor β and the life-cycle evolution of the portfolio composition to estimate the parameters κ and σ_ζ that govern the portfolio adjustment frictions.

The calibration targets three statistics for asset holdings and matches those statistics fairly well: The parameterzation of the model implies a median wealth-to-income ratio at age 50 of 0.71 (0.70 in the data) and for the 50th and the 75th percentiles of the illiquid wealth share at age 50 values of 0.71 and 0.96 (0.75 and 0.96 in the data). The resulting calibrated parameters together with the other calibrated parameters are shown in Table 5.

4.3 Model results

We simulate workers to obtain combinations of earnings, liquid assets, illiquid assets, and risk state for each age. A key advantage of the model is that we also observe the risk type for each period.

As a first step, we assess whether the simulated data yields the same estimated income process as observed in the actual data. Specifically, we apply our modified Kim filter

Table 5 Overview of model parameters

Parameter	Value	Description
γ	2	Relative risk aversion
π	-0.003	Liquid asset return
r	0.05	Illiquid asset return
σ_r	0.034	Volatility of the illiquid return
β	0.97	Discount factor
κ	2	Adjustment cost
σ_ζ	0.4	Randomness of adj. cost

(and smoother) to estimate the income process and impute the high-risk probabilities. The estimates closely align with the "true" values. The correlation between the imputed probability and the risk type is strong, at 0.76.

Next, we conduct an exercise that is not feasible with the data due to the lack of high-frequency wealth information. We examine how agents adjust their portfolios when they switch risk types. For this, we compute the first differences in risk type, earnings, and liquid and illiquid asset holdings. That is, the independent variable risk type change is 1 if a worker becomes high risk, 0 if the worker stays in the same type and -1 if the worker becomes low risk. The results are presented in Table 6.

Column (1) reinforces a key feature of the income process: when a worker changes risk type, their income decreases by 0.122 pension points.

Columns (2) to (4) examine the impact of a risk-type change on a worker's liquid asset holdings. On one hand, the decline in income prompts the worker to draw on their savings to smooth consumption. On the other hand, transitioning to a high-risk state tends to be persistent, which encourages the worker to maintain their savings for precautionary reasons. The effect of this adjustment is non-linear and depends on the worker's initial level of liquid wealth. Intuitively, a worker with little or no wealth cannot afford to reduce their assets when becoming high risk. Conversely, workers with higher initial liquid wealth are more likely to reduce their liquid assets in response to the risk change. The cutoff of 0.01 corresponds roughly to the median level of liquid wealth, while the cutoff of 0.02 is around the 57th percentile.

Column (5) shows that illiquid wealth remains unaffected by the risk-type change, which is expected given the presence of adjustment costs. This result holds even when restricting the sample to workers with higher wealth (not shown in the table).

Columns (6) and (7) demonstrate the movement of the wealth-to-income (WTI) ratio, which increases following a risk-type change. This indicates that workers reduce their savings by less than their income decreases. The effect is weaker for workers with better insurance coverage.

Table 6 Risk type change

	(1) Earnings	(2)	(3) Liquid wealth	(4)
		–	$liquid_{t-1} > 0.01$	$liquid_{t-1} > 0.02$
Risk type change	-0.122***	0.004***	-0.001	-0.003***
	(-75.39)	(9.37)	(-1.27)	(-3.25)
	(5) Illiquid wealth	(6)	(7) WTI	(8) Illiquid share
		–	$liquid_{t-1} > 0.1$	
Risk type change	-0.0003	6.311***	2.611***	-0.821***
	(-0.69)	(51.94)	(10.64)	(-13.46)

*** significant at the 1% level.

Regressions: Results of regressions of the first differences of earnings, liquid wealth, illiquid wealth, the wealth-to-income (WTI) ratio and the illiquid share between year t and year $t-1$ on the first difference of the risk type between year t and year $t-1$ average risk (up to year t). Additional controls are age dummies. In columns (3), (4) and (7), we restrict the sample to workers above a threshold of liquid wealth. Robustly estimated T-statistics in parentheses.

Definitions of variables: Earnings as well as wealth are expressed in "pension points", i.e., relative to the average labor earnings in that year. Illiquid shares and wealth-to-income ratios are expressed in percent. Portfolio returns are expressed in percentage points. The risk type is one if the worker is high-risk and zero otherwise.

Data source and sample: Simulated model data of 13,561 workers for age 30 to 50.

Finally, column (8) shows the change in the illiquid wealth share. The illiquid share decreases by nearly one percentage point upon the transition to a high-risk state.

Details and further experiments TBA.

4.4 Changes in income risk and risk-type transitions

Finally, we use the model to understand the macroeconomic importance of the persistent differences in risk types. For simplicity, we focus on the partial equilibrium household response to changes in the riskiness of incomes. As income risk changes trigger demand fluctuations, the aggregate effects of these then depend on monetary and fiscal policy (see Bayer et al., 2019, for the case of risk fluctuations in the standard single risk state earnings model). To understand the importance of the persistent income risk process, we compare first solve our model under the baseline and then consider three alternative scenarios. First, we assume that the standard deviation of income shocks increases for four years by 10 percent in both risk states. Second, we consider a scenario, where the standard deviation increases by 10% for four years because the transition into the high-risk state becomes more likely, and, third, we permanently replace the transition matrix of the risk state by the i.i.d. transition matrix that keeps the share of agents in the high-risk state

unchanged. **TBA ...**

5 Conclusion

TBA

References

Angerer, X. and Lam, P.-S. (2009). Income Risk and Portfolio Choice: An Empirical Study. *The Journal of Finance* **64** (2), 1037–1055.

Antoni, M., Ganzer, A., and Vom Berge, P. (2019). Stichprobe Der Integrierten Arbeitsmarktbiografien Regionalfile (siab-R) 1975 - 2017. FDZ-Datenreport. Documentation on Labour Market Data 201904 (de). Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg [Institute for Employment Research, Nuremberg, Germany].

Arellano, M., Blundell, R., and Bonhomme, S. (2017). Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework. *Econometrica* **85** (3), 693–734.

Bayer, C. and Kuhn, M. (2019). Which Ladder to Climb? Decomposing Life Cycle Wage Dynamics. Manuscript. IZA Discussion Paper.

Bayer, C., Luetticke, R., Pham-Dao, L., and Tjaden, V. (2019). Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk. *Econometrica* **87** (1), 255–290.

Bhutta, N., Bricker, J., Chang, A. C., Dettling, L. J., Goodman, S., Volz, A. H., Hsu, J. W., Moore, K. B., Reber, S., Windle, R., et al. (2020). Changes in US Family Finances from 2016 to 2019: Evidence from the Survey of Consumer Finances. Federal Reserve Bulletin 5. Board of Governors of the Federal Reserve System (US).

Blundell, R., Graber, M., and Mogstad, M. (2015). Labor Income Dynamics and the Insurance from Taxes, Transfers, and the Family. *Journal of Public Economics* **127**, 58–73.

Busch, C., Domeij, D., Guvenen, F., and Madera, R. (2022). Skewed Idiosyncratic Income Risk Over the Business Cycle: Sources and Insurance. *American Economic Journal: Macroeconomics* **14** (2), 207–242.

Chang, Y., Hong, J., Karabarbounis, M., and Wang, Y. (2020). Income Volatility and Portfolio Choices. Federal Reserve Bank of Richmond Working Papers 01. Federal Reserve Bank of Richmond.

Drechsel-Grau, M., Peichl, A., Schmid, K. D., Schmieder, J. F., Walz, H., and Wolter, S. (2022). Inequality and Income Dynamics in Germany. *Quantitative Economics* **13** (4), 1593–1635.

Fagereng, A., Guiso, L., and Pistaferri, L. (2018). Portfolio Choices, Firm Shocks, and Uninsurable Wage Risk. *The Review of Economic Studies* **85** (1), 437–474.

Gálvez, J. and Paz-Pardo, G. (2022). Richer Earnings Dynamics, Consumption and Portfolio Choice Over the Life Cycle. Banco de Espana Working Paper.

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., and Schupp, J. (2018). The German Socio-Economic Panel (SOEP). *Jahrbücher für Nationalökonomie und Statistik* **239** (2), 345–360.

Guiso, L., Jappelli, T., and Terlizzese, D. (1996). Income Risk, Borrowing Constraints, and Portfolio Choice. *The American Economic Review* **86** (1), 158–172.

Guvenen, F., Karahan, F., Ozkan, S., and Song, J. (2021). What Do Data on Millions of U.s. Workers Reveal about Lifecycle Earnings Dynamics? *Econometrica* **89** (5), 2303–2339.

Hall, R. E. (1982). The Importance of Lifetime Jobs in the U.S. Economy. *The American Economic Review* **72** (4), 716–724.

Hamilton, J. D. (1988). Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates. *Journal of Economic Dynamics and Control* **12** (2), 385–423.

— (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* **57** (2), 357–384.

Heaton, J. and Lucas, D. (2000). Portfolio Choice and Asset Prices: The Importance of Entrepreneurial Risk. *The Journal of Finance* **55** (3), 1163–1198.

— (2001). Portfolio Choice in the Presence of Background Risk. *The Economic Journal* **110** (460), 1–26.

Himmelreicher, R. K. and Stegmann, M. (2008). New possibilities for socio-economic research through longitudinal data from the Research Data Centre of the German Federal Pension Insurance (FDZ-RV). *Journal of Contextual Economics–Schmollers Jahrbuch* (4), 647–660.

Hintermaier, T. and Koeniger, W. (2018). Household debt and crises of confidence. *Quantitative Economics* **9** (3).

Jordà, Ò., Knoll, K., Kuvshinov, D., Schularick, M., and Taylor, A. M. (2019). The Rate of Return on Everything, 1870–2015. 3, 1225–1298.

Jung, P. and Kuhn, M. (2018). Earnings Losses and Labor Mobility Over the Life Cycle. *Journal of the European Economic Association* **17** (3), 678–724.

Karahan, F. and Ozkan, S. (2013). On the Persistence of Income Shocks Over the Life Cycle:

Evidence, Theory, and Implications. *Review of Economic Dynamics* **16** (3), 452–476.

Karahan, F., Ozkan, S., and Song, J. (2019). Anatomy of Lifetime Earnings Inequality: Heterogeneity in Job Ladder Risk Vs. Human Capital. FRB of New York Staff Report 908.

Kim, C.-J. (1994). Dynamic Linear Models with Markov-Switching. *Journal of Econometrics* **60** (1), 1–22.

Kim, C.-J. and Halbert, D. C. R. (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. The MIT Press.

Kuhn, M. and Ploj, G. (2020). Job Stability, Earnings Dynamics, and Life Cycle Savings. IZA Discussion Paper 13887.

Kuhn, M., Ríos-Rull, J.-V., et al. (2016). 2013 Update on the US Earnings, Income, and Wealth Distributional Facts: A View from Macroeconomics. Federal Reserve Bank of Minneapolis Quarterly Review 1. Federal Reserve Bank of Minneapolis.

Kuhn, M., Schularick, M., and Steins, U. I. (2020). Income and Wealth Inequality in America, 1949–2016. *Journal of Political Economy* **128** (9), 3469–3519.

Larkin, K. P. (2023). Job Risk, Separation Shocks and Household Asset Allocation. Manuscript.

Manuel, A., Blundell, R., and Stéphane, B. (2017). Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework. *Econometrica* **85** (3), 693–734.

Morchio, I. (2020). Work Histories and Lifetime Unemployment. *International Economic Review* **61** (1), 321–350.

Mukoyama, T. and Şahin, A. (2006). Costs of Business Cycles for Unskilled Workers. *Journal of Monetary Economics* **53** (8), 2179–2193.

Palia, D., Qi, Y., and Wu, Y. (2014). Heterogeneous Background Risks and Portfolio Choice: Evidence from Micro-Level Data. *Journal of Money, Credit and Banking* **46** (8), 1687–1720.

Rentenversicherung (FDZ-RV), F. der (2020). *Versichertenkontenstichprobe 2020 (Version 1.0)*. (Dataset).

Schmillen, A. and Möller, J. (2012). Distribution and Determinants of Lifetime Unemployment. *Labour Economics* **19** (1), 33–47.

Storesletten, K., Telmer, C. I., and Yaron, A. (2004). Cyclical Dynamics in Idiosyncratic Labor Market Risk. *Journal of Political Economy* **112** (3), 695–717.

Tauchen, G. (1986). Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions. *Economics Letters* **20** (2), 177–181.

A Details on the data

A.1 Sample size

Figure A.1 Number of workers per birth year (cohort)

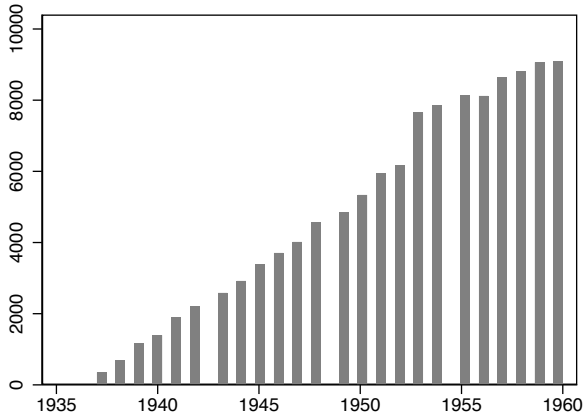
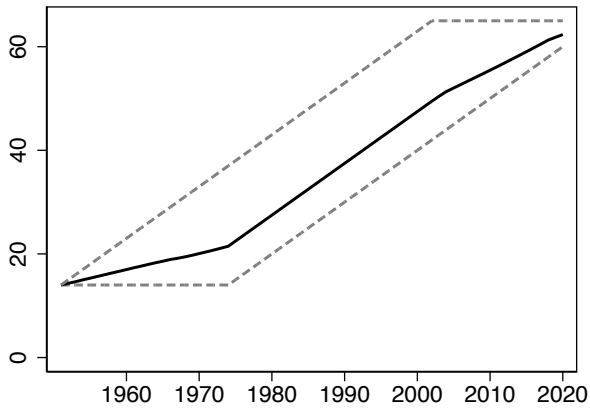


Figure A.2 shows the average age of all workers by calendar year in the data. It also shows the support of the age distribution by indicating the observed minimum and maximum age.

Figure A.2 Mean, minimum, and maximum age by year



Notes: The solid line shows the average age of all workers by calendar year in the data. The dashed lines show the minimum and maximum age in the data. Age is indicated on the vertical axis in years. Calendar time in years is on the horizontal axis.

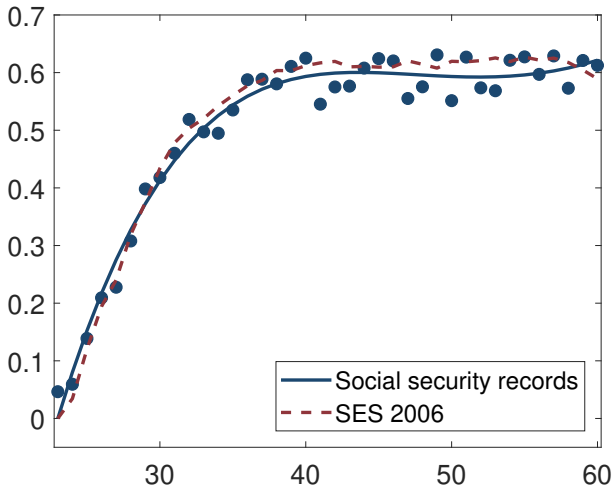
Table A.1 Number of workers per birth year (cohort)

	Freq.	Percent	Cum. percent
1937	353	0.298	0.298
1938	698	0.589	0.887
1939	1161	0.979	1.866
1940	1396	1.178	3.043
1941	1884	1.589	4.632
1942	2209	1.863	6.496
1943	2585	2.180	8.676
1944	2921	2.464	11.14
1945	3392	2.861	14.00
1946	3708	3.128	17.13
1947	4017	3.388	20.52
1948	4556	3.843	24.36
1949	4849	4.090	28.45
1950	5338	4.503	32.95
1951	5937	5.008	37.96
1952	6172	5.206	43.17
1953	7651	6.454	49.62
1954	7840	6.613	56.23
1955	8136	6.863	63.10
1956	8120	6.849	69.95
1957	8651	7.297	77.24
1958	8820	7.440	84.68
1959	9063	7.645	92.33
1960	9097	7.673	100
Total	118554	100	

A.2 Representativeness

To analyze the representativeness of the data, we compare the data from the social security records to independent data from the 2006 Structure of Earnings Survey ("Verdienststrukturerhebung"). The SES data is a linked employer-employee cross-sectional data that covers all German workers in dependent employment. Earnings data in the SES is not capped at the social security contribution limit. To make the data comparable, we therefore use annual gross earnings and drop all observations above the social security contribution limit for 2006. We transform all pension points in the worker's account for a given calendar year as described in Appendix A.3 and divide by the number of days employed. Figure A.3 compares the log annual earnings profiles from the two data sets. We see that the two data sets are very similar in slope and shape.

Figure A.3 Age-earnings profile for 2006



Notes: The red dashed line shows the age-earnings profile from the 2006 Structure of Earnings Survey ("Verdienststrukturerhebung"). The blue points show the raw data for the age-earnings profile from the social security records and the blue solid line shows the smoothed profile (fourth-order polynomial). All profiles have been normalized to zero at age 23. For the social security data, we use earnings from the smoothed profile.

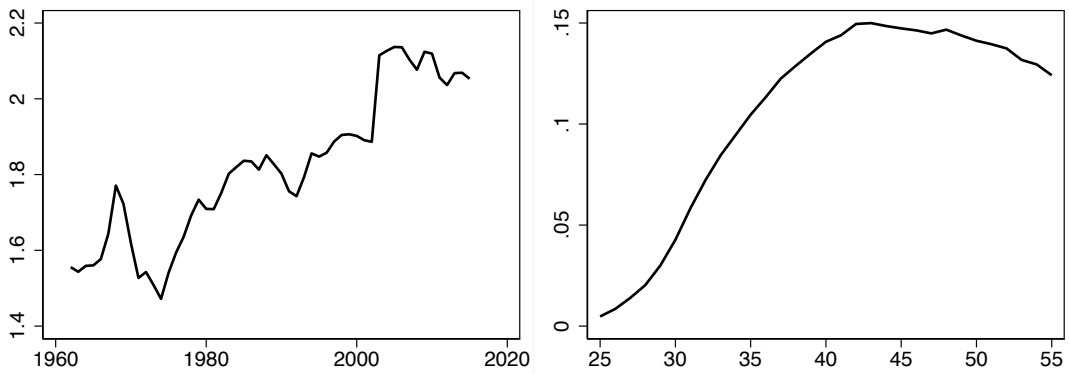
A.3 Transform to Euro earnings

The German old age pension scheme ("Deutsche Rentenversicherung (DRV)") uses the so-called "Durchschnittsentgelt" (average earnings) to express worker contributions. The individual accounts carry so-called "Rentenpunkte" that are the ratio of the workers earnings and average earnings. If a worker gets one point credited to her account, then she earned the "Durchschnittsentgelt" of the respective year. Importantly, the "Durchschnittsentgelt" does not correspond to the average income of all workers in the old-age pension scheme but is an earnings index that is constructed and updated since 1957. The growth rate comes from NIPA data on gross earnings per employee of all employees ("Bruttolöhne und Gehälter je Arbeitnehmer"). The initial 1957 value was based on full-time employees in 1957. Since earnings data are reported in pension points, we use the time series of "Durchschnittsentgelt" to get nominal values for earnings.

It is important to note, though, that Social security contribution in Germany are capped at an upper bound which changes over time. Figure A.4a shows the historical evolution of the social security contribution limit in pension points. Due to this cap, earnings for a certain fraction of workers are top-coded in the data. Figure A.4b shows that, in line with the the observed life cycle increase in average earnings, almost no workers are affected in early twenties, however, at age 50 around 15% of workers are affected by censoring.

Figure A.4 Social security contribution limit and the share of capped observations

(a) Social security contribution limit **(b)** Share of earnings observations



Notes: The left panel shows the social security contribution limit expressed in the number of pension points. The right panel shows the share of earnings observations that are affected by the social security contribution limit along the dimension of age.

A.4 Return data

Table A.2 Average historical returns, real

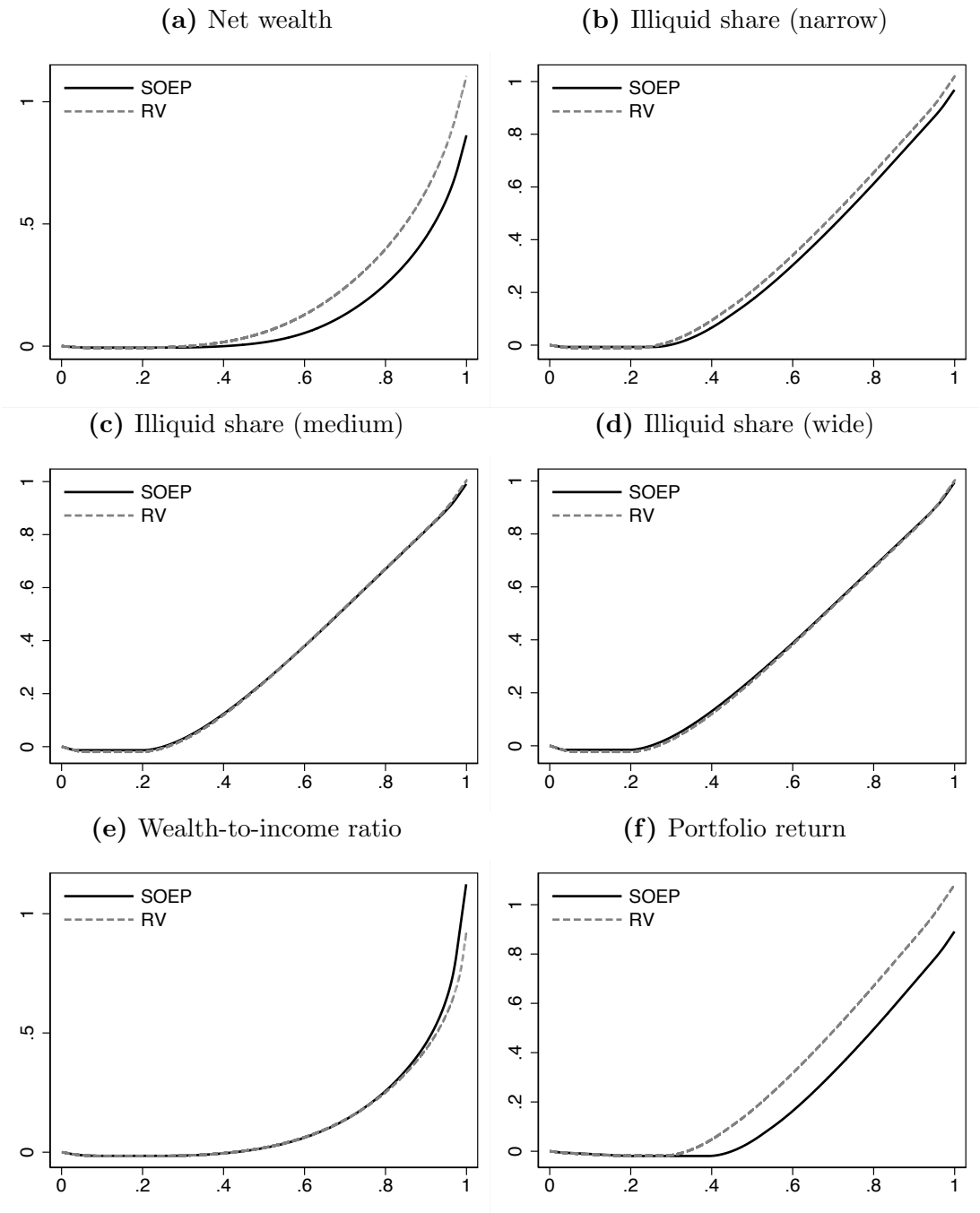
	(1) Equity	(2) Housing	(3) Bonds	(4) Bills	(5) Deposits
Constant	5.8030	4.7334	3.6345	1.3852	-0.2123
Observations	50	50	50	50	50

Notes: The table reports average historical returns based on German data between 1970 and 2020. The first four columns use data from Jordà et al. (2019), the last column uses data by the Bundesbank.

A.5 Imputation

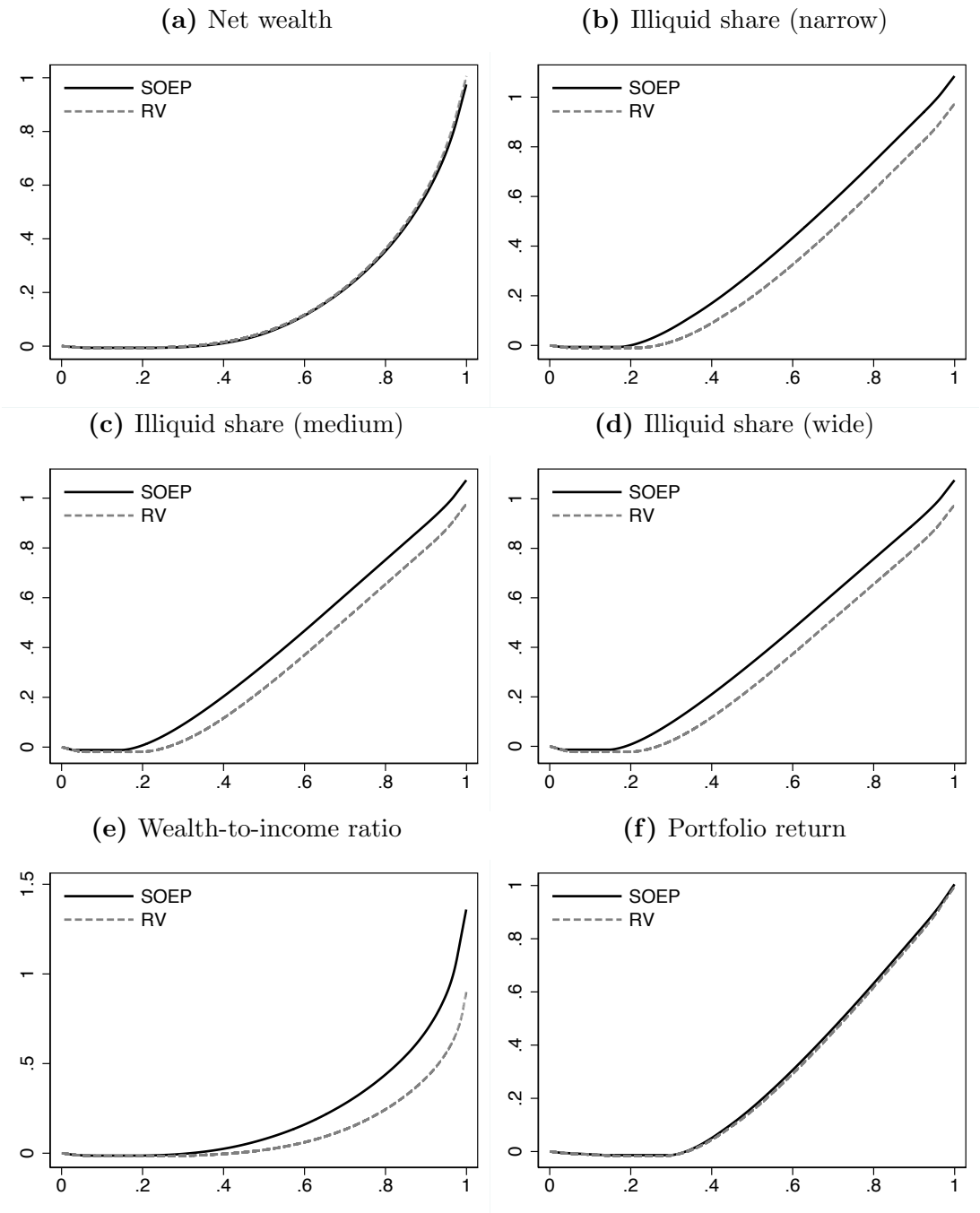
Figure A.5 and Figure A.6 show the result of the imputation by comparing the implied distributions, first for the full sample and second for a sample where the age range is restricted to be identical.

Figure A.5 Comparison of distributions



Notes: Lorenz curves comparing the distributions in the SOEP with the imputed distributions in the pension data (RV).

Figure A.6 Comparison of distributions, conditional on same age range



Notes: Lorenz curves comparing the distributions in the SOEP with the imputed distributions in the pension data (RV).