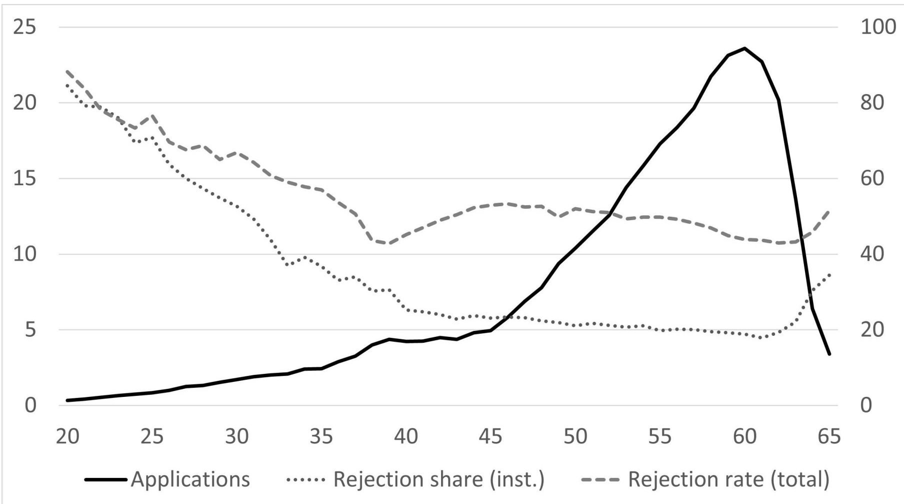
Appendix 1: Detailed description of data processing

In this appendix, we describe the parametrization of the model. After discussing the recent data on disability applications and inflows, we focus on the income process and its interaction with health shocks during the working phase of households. Finally, we discuss the data which is used to calibrate some of the model parameters internally.

1A: Applications and rejections for disability pensions in 2019

The most recent available data on disability applications is from the year 2019. We focus on the most recent applications since older data (i.e. around the year 2010) is only available in much less detail. In 2019 about 348.200 applications for disability pensions were submitted to the pension insurance, of which 170.100 were rejected. Figure 6 shows on the left vertical axis the absolute application numbers (in 1000) per age group and on the right axis the percentage rates of rejections. Less than 3 percent of all applications were submitted by households younger than 30. This group has a high rejection rate of more than 70 percent. In addition, most of the rejections were due to institutional reasons (which are not captured in our model). The latter refers to applicants with a total contribution record of less than five years or no contributions during the last three years or applications were sent to another insurance agency. After age 30 applications for disability pensions increase steadily until about age 45. About 14 percent of all applications were submitted by cohorts in the age group 31-44. The rejection rates fall in this group with rising age and institutional rejections become less and less important. Overall about 24 percent of rejections were due to institutional reasons.

Figure 6: DP applications and rejections in 2019



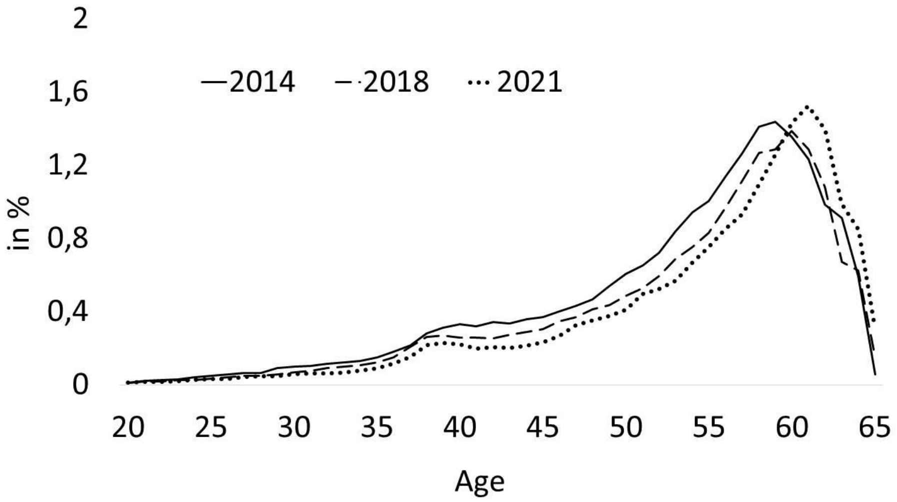
Starting with age 45 applications increase steadily until they peak roughly at age 60, before they fall again sharply to zero by age 65. Of course, after age 60 there are alternative routes into retirement (such as severely handicapped retirement), which is not captured by the model. Especially when they were formerly unemployed, households may also retire with reduced old-age pensions before, and can only retire with full old-age pensions in and after the normal retirement age. Note that the rejection rate is roughly constant after age 45 and that institutional rejections are very low.

Besides the above-described institutional reasons, most rejections were declared because applicants did not pass the medical examination (about 56 percent). In the remaining cases, rejections were mostly due to either non-cooperation or withdrawal by applicants. This behavior might indicate that the applicants were afraid not to pass the medical test. The average processing time of an application was about 160 days, whereas the average acceptance took roughly three weeks longer. However, there is a huge spread in the processing time ranging from 56 days (when sent to another agency) up to 200 days (rejection because of non-cooperation).

1B: DP inflow during 2014 - 2021

While applications and rejection rates are not directly available to the public, detailed inflow data is available to the public even for the most recent years. To see whether and how the considered reforms have already affected retirement behavior in practice, Figure 7 provides the age-specific DP inflow for the years 2014, 2018, and 2021. As one can see the age-shape of the inflow has hardly changed but the peak has shifted to the right. While in the year 2014 most households retired with DP at age 59, this number increased to 60 and 61 in the years 2018 and 2021, respectively. As a consequence, the average DP retirement age has increased from age 51.8 in the year 2014 to 54 years in the year 2021.

Figure 7: DP inflow in 2014, 2018, 2021



Since our simulations consider situations in the long run, when the reforms are fully implemented, the increase in DP retirement ages shown in Figure 7 might reflect a transitional situation (not captured by our model) where the impact of the Reform 2007 on retirement ages is much stronger than the contrary effect of the Reform 2018.

1C: Health transitions for age and education groups

The transition matrices of the health state for individuals older than age 45 are obtained from the German sub-sample of the Survey of Health, Ageing and Retirement in Europe (SHARE). A detailed description of the data set can be found in Börsch-Supan et al. (2013). The SHARE is a longitudinal data set that includes a wide range of micro-data on socio-economic status, social and family networks, as well as health across European countries. The latest SHARE Release 7.0.0 includes about 140,000 individuals aged 50 or older in 28 countries. The German sub-sample currently includes seven waves between 2004 and 2018 with a total of 8,788 persons. However, the original sample had to be reduced considerably for our purposes. As shown in Table 15, we need to subtract the first 337 persons because of missing information on some health variables, age, or educational background. The remaining sample of 8,451 persons with 17,444 observations was used to construct the health

Table 15: Selection of sample from German SHARE data 2004-2018

	High school ($s = 0$)	College ($s = 1$)	$\sum$
Original sample			8,788
- missing identification variables			-337
Sample for health index weights			8,451
- only one observation			-3,221
Final sample for transition matrix	3,666	1,564	5,230
Transitions	5,946	2,768	8,714

index weights which are discussed below. To construct the transition matrices for the two education groups and age brackets we needed to exclude further all persons who only provided full information in one wave. This reduced our sample for the transition matrix further to 5,230 persons with 8,714 health transitions. Table 15 shows that our final sample roughly contains the double number of people and observations for high school graduates compared to college graduates. In both educational groups, the transitions are roughly the same in the two age brackets 45-64 and 65+.

Following Jürges et al. (2015) or Poterba et al. (2017) health is measured by constructing a health index after performing a principal component analysis using data from six waves of SHARE.[33] Using all observations at once a continuous health index is computed from the first principal component of twenty-one health indicators listed in Table 16. The weights (or loadings) of the first principal component reported in Table 16 are chosen to maximize the variance of the projected health data. As one can see, the weights are fairly stable across all waves. The highest weights are given daily life health measures (i.e. difficulties in walking, lifting, climbing, etc.), while much less weight is given to questions about whether the respondent ever experienced specific health problems. This is very much in line with the results of Poterba et al. (2017), who argue that this probably reflects a high correlation between many self-reported measures.

Using the weights from Table 16 the individual data is converted into a "raw health index" in each wave for each person with more than one observation. In the next step, the raw health indices of each wave are put together and normalized on the $[0, 1]$-interval, where a higher value indicates a worsening of health. The interval is split into three areas $[0 - k_0, k_0 - k_1, k_1 - 1]$ and all normalized indices are shifted back to their original wave. Now we can observe transitions and allocate them back into the cells of the respective (h, h^+) matrix. Finally, the resulting transitions matrix has to be recalculated on a yearly basis. Given the annual transition probabilities, we can compute the distribution and dynamics of the health status over the life cycle and compare it with the data. The calibrated interval borders $k_0 = 0.43$ and $k_1 = 0.80$ then minimize the difference between the model generated and the observed dynamics of the health status. Tables 4 and 5 show the final transition matrices derived with this optimizing procedure.

[33] The third wave contained only retrospective data on early lives.

Table 16: Health index weights (principal component loadings)

Health measure	Loads
Difficulty walking 100m	.300
Difficulty lift/carry	.294
Difficulty push/pull	.294
Difficulties with an ADL	.297
Difficulty climbing stairs	.276
Difficulty stoop/kneel/crouch	.288
Difficulty getting up from chair	.279
Difficulty reach/extend arms up	.257
Difficulty sitting two hours	.224
Difficulty picking up a coin	.179
Ever experience heart problems	.123
Ever experience stroke	.130
Ever experience high blood pressure	.119
Ever experience lung disease	.103
Ever experience diabetes	.119
Ever experience cancer	.074
Nursing home stay during last year	-.089
Hospital stay during last year	-.149
BMI at beginning of period	.114
Self-reported health	.299
Psychological problems (Euro-D-scale)	.233
N	17,444

1D: Productivity profiles and productivity risk

The parameters to calibrate the earnings process and earnings risk over the life cycle are taken from Kindermann and Püschel (2021), who use administrative data from the German public pension insurance system. The scientific use file of the Versichertenkontenstichprobe 2017, contains information from the insurance accounts of 69,520 individuals, with information on age, gender, education, and (most importantly) a monthly history of accumulated pension claims which could be used to compute individual earnings. The estimations are based on observations between the years 2000 and 2016 of males aged between 25 and 60 with available information on educational background. Labor earnings comprise income from regular work, marginal employment, and short-term unemployment (up to one year), all other sources of pension claims (like times of care for children or sickness) are counted as zero earnings. Individuals with a full year of zero earnings were excluded from the sample. This selection procedure left a total of 15,242 individuals and 189,184 annual earnings observations. The share of individuals with a college degree in this group was 23.73 percent, all remaining persons were assumed to have a high school degree. Due to the contribution ceiling, the raw data was top-coded and due to part-time jobs or so-called mini-jobs a substantial fraction of the sample showed an earnings level below 25 percent of average income. Consequently, a simple log-normal AR(1) process is not rich enough to describe the earnings dynamics of households in Germany. While there exists a "′normal"′ earnings process that follows some regular AR(1) dynamics, individuals can also experience very low earnings episodes. To develop a statistical model that can fit the data on low earners by age and education, Kindermann and Püschel (2021) split the data set into two parts where the first group contains all labor incomes below an earnings threshold of 23 percent of average labor earnings. These low-earnings individuals can be thought of as having some months of temporary unemployment or non-employment throughout a year or as being marginally employed. All individuals with labor earnings above the threshold have normal labor earnings. To specify a statistical model with low earnings episodes, Kindermann and Püschel (2021) assume that individuals either follow a stable or an unstable career path, where both groups have the same probability (i.e. $\omega_m = 0.5$) independent of the educational background. Those who are on a stable career path are always in the normal earner group, while those on the unstable career path may experience earnings shocks, which put them temporarily in the low earner group.

The normal earner group

The normal earner sample is then split between high school and college graduates and the earnings dynamics of both groups are described by a standard AR(1) process in logs. The estimated statistical model

$$\log(y_{isjt}) = \hat{\kappa}_{t,s} + \hat{\theta}_{j,s} + \hat{\eta}_{isjt} \quad \text{with} \quad \hat{\eta}_{isjt} = \hat{\rho}_s \hat{\eta}_{is,j-1,t-1} + \hat{\epsilon}_{isjt} \quad \text{and} \quad \hat{\epsilon}_{isjt} \sim N(0, \hat{\sigma}^2_{\epsilon,s})$$

explains individual labor earnings y_{isjt} of an individual i with education s, age j and year t with a year fixed effect $\hat{\kappa}_{t,s}$ that controls for earnings changes along the business cycle, an age fixed effect $\hat{\theta}_{j,s}$ that informs us about the age-earnings relationship and a noise term $\hat{\epsilon}_{isjt}$ that is assumed to follow a normal distribution with mean zero. The method of moment estimation controls for the top-coding and the truncation at the lower earnings threshold, see Kindermann and Püschel (2021) for more details. The point estimates of the age-fixed effects show a steep increase for both education groups up to age 45 and a stagnation or slight decline afterwards for both education groups. The college wage premium implied by the profiles is about 60 percent, which seems fairly realistic, see OECD (2016). The profiles of the age-fixed effects are used to calibrate the age-productivity profiles for the two education groups in the model.

Table 17 reports the parameter estimates for the residual earnings process. Both earnings groups exhibit a high persistence in labor earnings with an unconditional variance of roughly 15-20 percent. Following Kindermann and Püschel (2021), we directly apply the estimated autocorrelation terms in our model but calibrate the innovation variance to replicate the unconditional variances with our model.

Table 17: Parameters of residual log-earnings process

	High school ($s=0$)	College ($s=1$)
Autocorrelation $\hat{\rho}_s$	0.9869	0.9900
Innovation variance $\hat{\sigma}^2_{\epsilon,s}$	0.0046	0.0039
Unconditional variance $\frac{\hat{\sigma}^2_{\epsilon,s}}{1-\hat{\rho}_s}$	0.1780	0.1982

The low earner group

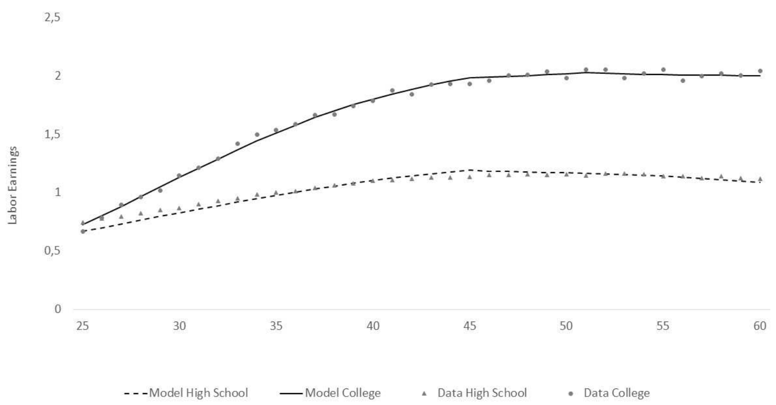
The fraction of individuals in an age cohort that are in the low earnings group declines over the life cycle for both educational groups. However, college-educated mostly experience low labor earnings early in their career and after age 35 their share converges to almost zero. High school workers experience low earnings episodes much less at younger ages but this sharer declines also much less later in life. The data also shows that individual labor earnings in the low earnings group are by and large independent of age and education and amount on average to roughly 10 percent of average labor earnings (i.e. $\exp(\eta_0) = 0.1$). Kindermann and Püschel (2021) therefore distinguish within each educational level two groups that exhibit different degrees of career stability. Some workers follow a stable career path, while others frequently transition into and out of employment. Individuals therefore draw at the beginning of working life a discrete shock whether they face a stable or unstable career path during their working life. They assume the same probabilities $\omega_m = 0.5$ to belong to one of the two groups. The transition into and out of low earnings for those who face unstable careers is then modeled as a first-order discrete Markov process for both educational groups. Households with unstable careers $(m = 1)$ face an education specific transition matrix

$$\Pi^s_{low} = \begin{bmatrix} 1 - \pi^s_{low,0} & \pi^s_{low,0} \\ 1 - \pi^s_{low,1} & \pi^s_{low,1} \end{bmatrix}$$

where the probability $\pi^s_{low,0}$ indicates the likelihood of a normal earner transitioning into the low earnings state in the next period, while $\pi^s_{low,1}$ is the probability to remain in the low earnings state. Furthermore, they assume that, at age 25, a fraction ω^s_{low} of individuals of education s starts as a low earnings individual. The point estimates which are reported in the calibration section provide the best fit to the data in a least squares sense.

Age-earnings profile

Figure 8: Education-specific age-productivity profiles



The original data is taken from Kindermann and Püschel (2021), while the straight line is simulated with the parameters from the model.

1E: Scientific use file of fully insured lives (Vollendete Versicherungsleben) SUFVVL2016

To calibrate parameters and the initial equilibrium of the model we use the administrative data set of fully insured lives in 2016 which is available as a scientific use file (SUF) from the research center of the German pension insurance. The representative sample contains 25 percent of all pensioners who received in this year the first time a pension benefit in Germany. The data set contains biographical information on sex, age, address, citizenship, family status, number of children, and education, but also all information necessary to compute the benefit level (sum of earning points, etc.). In the year 2016, 832,664 people entered retirement, so the full sample contains 208,166 insured persons. As Table 18 documents, we have excluded all persons where we assume that their employment history differs significantly from a regular work history (women, foreigners, no residence, no contribution record, long absence). In addition, we only wanted to consider retirement with old age or with disability

Table 18: Selection from German VVL 2016 sample

	Initial full sample (persons)	208,166
-	women	-110,679
-	pensions from unemployment, severely handicapped, etc.	-7,827
-	retirement before 2015	-1,745
-	pensions without contribution record	-2,533
-	no German citizenship/residence	-3,981
-	no information on education	-25,931
-	long absence from pension insurance	-403
=	final database	55,067

pensions who had information on their educational background. This procedure left us with roughly more than 55,000 persons of which were 44,483 high school graduates and the remaining 10,584 had a college degree.

Appendix 2: Additional material from the simulation model

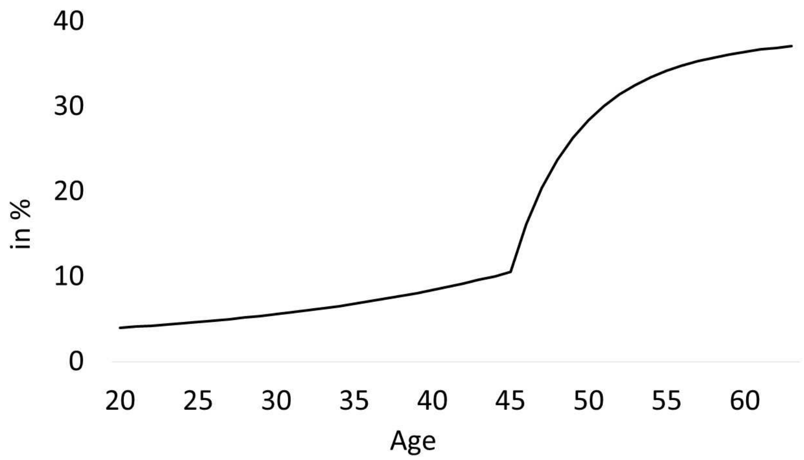
In this appendix, we provide additional material concerning the model calibration and sensitivity checks.

2A: Fraction of non-contributors over the life cycle

We have calibrated the probability of a non-contributory year π_h^{nw} which depends on the health status so that we match the fraction of those who reach a skill-specific threshold of 35 contribution years. Since - mostly for technical reasons - a year without contributions is not a specific dimension of the state space, we cannot derive a Markov matrix that gives the transitions into and out of this situation for specific social groups. However, we can compute the age-specific fraction of those who do not pay contributions (and in our model do not receive labor income).

As Figure 9 shows, this fraction is quite low (at less than 10 percent) for ages younger than 45 years. Then it increases sharply and approaches almost 40 percent for older cohorts. It is difficult to compare these generated data with some observed data since non-contributors in our model cover many different social facts not directly covered by the model. According to the German Statistical Office (StaBu, 2023), the employment rate of 55 to 64-year-olds in Germany rose markedly during the last decade from 62% in 2012 to just under 72% in 2021. This means that between 30 and 40 % of the German workforce at that age did not contribute which fits quite well with this figure. Of course, there are many reasons why people do not contribute temporarily (i.e. illness, self-employment, children, the shadow economy, etc.), but for us only the consequences for early retirement eligibility matter.

Figure 9: Fractions of non-working households without contributions



2B: Sensitivity analysis: No income during DP application

The simulations above always assume that households who apply for DP can only work 20 percent of their time endowment during that year. This assumption is justified for two reasons. On the one side most people still work one year before they receive DP benefits and on the other side DP retirement requires severe work limitations that are defined by this time interval. To check how important this assumption is for our results, we repeat the three major simulations in Table 19 assuming that one cannot earn income during the application period.

Table 19: Sensitivity analysis: No income during DP application

	No reform equilibrium	Reform 2007	Reform 2018
Retirement ages (av.)	62.5	63.5	62.8
OAP	64.1	64.8	64.7
DP	53.7	55.4	55.0
Pension budget			
Total budget (in % of GDP)	12.3	11.2	11.8
DP (in % of pension budget)	4.9	4.9	8.3
DP (in % of OAP benefits)	64.2	68.9	76.1
DP applications (50-65)	1.0/7.3	0.8/7.4	1.7/9.1
Labor supply	–	1.5	0.9
Capital	–	3.2	1.1
Wage rate	–	0.5	0.1
Interest rate (in pp)	3.0	-0.1	0.0
Contribution rate (in pp)	18.5	-1.7	-0.7
Consumption tax rate (in pp)	16.2	-1.1	-0.5
Welfare	–	3.4	1.4

Changes are reported in % of initial equilibrium if not stated otherwise.

The left column of Table 19 shows that this assumption significantly reduces DP applications and DP retirement already in the initial equilibrium. Consequently, the fraction of DP benefits in the total pension budget is lower, the total budget is lower and the contribution rate is slightly lower. However, OAP retirement is hardly affected. The reform of 2007 has almost identical consequences at the micro and macro levels as reported in Table 13. Households now retire later and substitute for DP retirement, which reduces the pension budget and the contribution rate. The increase in labor supply and wages boosts income taxes so that the consumption tax rate can be lowered. Similarly, the Reform 2018 induces households to switch towards DP retirement, which in turn lowers the retirement age of both OAP and DP, increases the fraction of DP benefits within the pension budget, and increases the overall pension budget. As it seems, the substitution towards DP is a bit less pronounced than in Table 13 above. Consequently, Reform 2018 decreases welfare, but less than before.

Overall we think that this sensitivity check strengthens the central message of our paper.

Literature

Börsch-Supan A., M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck and S. Zuber (2013): Dataresource profile: the survey of health, ageing and retirement in Europe (SHARE), *International Journal of Epidemiology 42*, 1-10.

Poterba, J., S. Venti and D. Wise (2017): The asset cost of poor health, *Journal of the Economics of Ageing 9*, 172-184.

Statistisches Bundesamt (StaBu) (2023): Employment of older people in Germany and the EU markedly up in the last 10 years, Press release No. N003 of 19 January 2023, Wiesbaden.