

Unemployment, Sick Leave and Health*

Matthias Schön[†]

October 13, 2015

JOB MARKET PAPER

Abstract

This paper studies the relationships among sick leave, income and unemployment. It investigates these relationships under the generous German sick leave regulation of 100% wage replacement, whereby workers do not bear any direct costs from sick leave. Using information from the German Socioeconomic Panel (SOEP), I identify three stylized facts of sick leave in Germany. First, the number of sick leave days shows a strong pro-cyclical pattern. Second, the average use of sick leave is hump-shaped over income quintiles. Third, the number of days of sick leave is a strong predictor of future unemployment. Using micro-evidence, I develop a structural model that rationalizes these facts. I argue that in the absence of direct costs of sick leave, the fear of future unemployment is the main driving force restraining sick leave. I then use the model, calibrated to the German labor market, to conduct counterfactual policy analysis. My results suggest that policies governing unemployment benefits significantly change the number of days of sick leave.

JEL Classification: D31, E24, J22, J32, H51, I14, I18.

Keywords: Labor Economics, Health Economics, Unemployment, Sick Leave, Inequality

*Part of this project has been realized while visiting the economics department of the University of Pennsylvania. I am indebted to Dirk Krueger for his invitation. I thank seminar participants at the iHEA Congress 2014, the University of Cologne, the University of Frankfurt, SMYE 2015, the University of Linz, the ESPE 2015, the SOLE/EALE World Meeting 2015, the Verein für Socialpolitik Annual Conference, the Workshop of Quantitative Dynamics at Konstanz, Hanming Fang, Max Groneck, David Jaeger, and Alexander Ludwig for their helpful comments.

[†]Center for Macroeconomic Research, University of Cologne; Albertus-Magnus-Platz, 50923 Köln, Germany; E-mail: m.schoen@wiso.uni-koeln.de

1 Introduction

Absence from work due to sickness poses a major threat to the economic situation of households. On the one hand, workers who take sick leave face direct opportunity costs arising from a reduction of working time for which they would otherwise be paid. In most countries, these costs are partially insured by paid sick leave schemes. The extent of this insurance coverage varies greatly across industrialized countries, cf. Scheil-Adlung and Sandner (2010).¹ On the other hand, workers have to take into account the indirect costs of sick leave that stem from reductions in future expected earnings. The layoff and promotion decisions of employers depend on workers' past days of sick leave, cf. Markussen (2012), Chadi and Goerke (2015) and further evidence below. To fully understand the role of sickness absence, it is therefore important to distinguish between the two types of costs.

In this paper, I identify the costs of sick leave stemming from lower employment prospects, and I analyze their economic implications. To do so, I focus on Germany, which features a very generous sick leave system that almost completely rules out direct opportunity costs of sick leave. In cases of work absence due to sickness, every (full-time, part-time or temporary) employee is eligible for six weeks of 100% wage replacement.²

The first objective of this paper is to identify and document patterns of sick leave utilization over business cycles and income groups in Germany. The second objective is to rationalize these empirical findings within a theoretical framework and to highlight the main mechanism: the decision to trade off utility-enhancing health against expected future earnings due to increased risk of job loss. The third objective is to analyze the distributional effects of indirect costs of sick leave and to evaluate counterfactual policies within a structural model calibrated to the German labor market.

For the first objective, I employ data from the German *Sozio-oekonomische Panel* (SOEP),

¹An extreme case among developed countries is the US, where there is no statutory paid sick leave, a situation US President Barack Obama recently urged to change in his State of the Union in 2015. Paid sick leave may be provided by employers on a contractual basis. According to the 2015 State of the Union, the number of workers without any sick leave scheme amounts to 43 million.

²This generosity is not without a price. The expenditures of paid sick leave, which are borne by employers, amounted to almost 40 billion € in 2013, or more than 1.5% of GDP, according to the German Federal Ministry of Labour and Social Affairs, cf. Bundeministerium für Arbeit und Soziales (2014). This number does not include the contribution to the social insurance system, which amounted to 6.9 billion €. For more information on the German regulations on paid sick leave, see Appendix A.

a nationally representative longitudinal dataset. With respect to aggregated data, I find three remarkable patterns of sick leave utilization in Germany. First, claims of sick leave exhibit a strong pro-cyclical pattern; i.e., workers are on average less absent from work in times of high unemployment. The correlation coefficient of average days of sick leave and the German unemployment rate is minus 0.6383. Second, the average number of days of sick leave in Germany displays a marked hump-shaped pattern across income quintiles. Workers in the medium income quintile have on average almost 10% more days of sick leave than workers in the bottom income quintile. This figure is noteworthy because average health monotonically increases with income.³ Assuming sick leave is driven only by health would therefore lead to a decrease in the number of days of sick leave between bottom and medium income quintiles, and this is not observed. Third, the variance of sick leave differs greatly between income quintiles. On the one hand, employees in the bottom quintile have the highest probability of not missing any day in a year. On the other hand, they also have the highest probability of missing more than two weeks. Top income employees miss a small number of days but do so at a higher frequency.

Regarding the second objective, the mechanism that rationalizes these three stylized facts of sick leave is that workers who become sick face the decision either to stay at home and recover or to go to work sick. Staying at home restores utility-enhancing health but simultaneously increases the risk of job loss. Exploiting the panel structure of the SOEP, I show that sick leave is one key predictor of future unemployment. Taking three additional days of sick leave increases the risk of becoming unemployed by a factor of 1.1. Going to work sick preserves expected future earnings, but a perpetual neglect of recuperation diminishes long-run health prospects.⁴ In times of high unemployment rates, workers face both higher overall firing rates and lower reemployment probabilities. Resulting higher marginal costs of unemployment shift the trade-off toward presenteeism and drive the cyclical pattern. Workers facing financial constraints, i.e., low skilled workers, are less able to smooth consumption over periods of unemployment and

³A finding well documented in the health economics literature, cf. Smith (1999).

⁴Time requirements for recovery in the health economics literature go back to the very beginning, c.f. the seminal paper of Grossman (1972). Clinical, experimental, and empirical evidence in support of this idea can be found in the bio-medical science, public health, psycho-biology, bio-sociology, and empirical health literature, which shows a negative effect of neglect of recuperation time, cf. Kivimäki et al. (2005), Bergström et al. (2009).

have an overall higher risk to become unemployed. Therefore, they are particularly compelled to go to work when sick. Consequently, optimal sick leave utilization differs across income groups. The rich constantly use fewer days absent from work to conserve their health. Poorer workers reduce their number of days of missing work to keep their job; however, this practice comes with the cost of a perpetual worsening of their health. In the long run, lower health increases the number of sick leave days for the poor compared to the rich. This outcome of the mechanism is affirmed by the finding of a widening sick leave gap between bottom and top income quintiles over the life cycle. Beginning at almost the same level of sick leave at age 20, workers in the bottom income quintile have 40% more days of sick leave at the end of their working life than those in the top income quintile.

To quantify the distributional effects of sick leave, the third objective, I develop a heterogeneous agent model with an endogenous health state. A shock process mimics acute sickness. Additionally, I implement central characteristics of the German labor market and worker protection system in my model.⁵ The government imposes a flat income tax on agents. The collected revenues are used to finance (i) expenditures due to sick leave payments, (ii) unemployment benefits, including means-tested welfare, and (iii) a retirement system.

To implement my quantitative analysis empirically, I first estimate and calibrate the model using SOEP data to match key statistics on sick leave, health status and unemployment. The estimated model is able to successfully explain the targeted features of the data in the estimation (e.g., the distribution of days of sick leave utilized by low income workers). It is also capable of reproducing other (non-targeted) dimensions, such as the hump-shaped pattern of average days of sick leave across income quintiles, the income gradient in health, and the cyclicity of claims of sick leave. I then use the parameterized version of the model as a laboratory to evaluate the consequences of different policy options. For this purpose, I contrast the benchmark economy, for instance, against an economy with no unemployment benefits. An immediate implication is that the average worker reduces her recuperation time by 1.2 days a year, and a worker in the bottom income quintile reduces it by more than 1.7 days a year.

⁵German workers have universal health insurance that covers medical expenditures. For this purpose, I omit medical expenditures in my model.

Related Literature There is a sizable body of literature that documents a positive relationship between workers' economic situation and their average claims of sick leave, cf. Leigh (1985), Pfeifer (2013). Arai and Thoursie (2005) and Askildsen et al. (2005) show that this procyclical variation in sickness absence is caused by established workers reducing sick time rather than the absence behavior of marginal workers entering or leaving the working population in various states of the business cycle. Another growing strand of literature studies the effect of sick leave on individuals' future labor market outcomes. Hesselius (2007) and Markussen (2012) both show that an increase in the sick leave rate lowers the probability of being employed in the future. Andersen (2010) also finds that a high number of days of sick leave not only affects employment status but also decreases post-sick leave earnings. I contribute to both strains of the literature by showing that these findings also hold for Germany. I add to the literature by merging the findings of sick leave, and I expand the discussion with a cross-sectional dimension.⁶

The structural model I present in this paper is part of a broad and growing body of literature that incorporates endogenous health into dynamic life-cycle models. Important related contributions include Grossman (1972), Ehrlich and Chuma (1990), Hall and Jones (2007), Ales et al. (2012), Halliday et al. (2012), Ozkan (2014), Cole et al. (2014). A small body of literature allows for interactions between endogenous health, employment and productivity. In a recent paper, Laun (2013) analyzes optimal insurance against unemployment and disability in a private information economy with endogenous health and search efforts. Little research has been conducted on such dynamic models distinguishing between long run health and the onset of acute illnesses. Gilleskie (1998) predicts changes in sickness-related absenteeism that arise with improvements in access to health care through more complete health insurance and sick leave coverage in the US. The paper, however, focuses only on the direct costs of work absence and does not take into account the risk of unemployment. It also falls short of providing a link to the endogenous health literature. To the best of my knowledge, this is the first paper to incorporate endogenous health and acute sickness into a heterogeneous agent life-cycle model.

⁶There is also considerable literature exploiting reforms of sick pay provision in Scandinavian countries (cf. Henrekson and Persson (2004), Johansson and Palme (2005), Dale-Olsen (2014)) and, more recently, Germany (cf. Ziebarth and Karlsson (2010), Ziebarth (2013), Ziebarth and Karlsson (2014), Puhani and Sonderhof (2010)). They found that an increase in generosity in the sick leave system induces a higher number of days missed at work, indicating a moral hazard.

The rest of the paper is organized as follows. In Section 2, I discuss the main data source, the methodology and the empirical findings. Then, I introduce a full structural model in Section 3. In Section 4, I discuss the estimation of the model and the model's fit to the data. Section 5 presents counter-factual policy experiments using the model. Finally, I conclude in Section 6.

2 Empirical Facts on Sick Leave Days

The purpose of this section is to carve out stylized facts of sick leave in Germany and to motivate the key modeling assumptions of the structural model in section 3. After discussing the data source and the methodology in section 2.1, I present in section 2.2 findings on aggregated data that show that taking sick leave is an endogenous choice of workers. Then, in section 2.3, I show results based on a panel analysis that underline the importance of this choice for the employment prospects of workers.

2.1 Data and Methodology

2.1.1 Description of the Survey

My empirical analysis is based on the German *Sozio-oekonomische Panel* (SOEP), a nationally representative longitudinal dataset. Starting in 1984 and conducted annually, it comprises 30 waves of household data. It oversamples foreigners, immigrants, and East Germans to allow for more precise estimates for population subgroups that may be of particular policy interest.⁷ The SOEP provides detailed information about demographic (e.g., sex, age), socioeconomic (e.g., educational level, marital status) and economic characteristics. The respondents report their current monthly income and their household income in the current and the previous year.⁸ The employment history contains the current employment status (e.g., full time, part time), the point in time of the layoff in the previous year, the length off the unemployment spell and information about the time worked for the same firm. Information about health since 1990 is requested. In addition to self-reported health, the SOEP contains information

⁷I include all sub-samples of SOEP with the appropriate cross-sectional weights.

⁸All monetary variables are deflated with the consumer price index contained in the SOEP using 2005 as the base year.

about the number of doctor visits and hospital stays.⁹ Further detailed information about the characteristics of the SOEP is provided in Wagner et al. (2007).

The key variable for the purpose of this paper is the number of working days missed due to sickness. The SOEP asks respondents to state whether they missed any day due to sickness in the previous year and, if so, how many days they missed. Puhani and Sonderhof (2010) show that, though self-reported, the SOEP adequately depicts the true number of days of absent from work. The SOEP also contains information about the number of spells that last longer than six weeks. However, the survey does not ask about occasions on which respondents go sick to work.

The only information that is used and not contained in the SOEP is the unemployment rate of Germany. I use official data from the federal employment agency, cf. Bundesagentur für Arbeit (2014).

2.1.2 Determination of the Sample

For the following empirical results, not all observations of the SOEP are used. Because I am interested in the sick leave utilization of workers, I focus only on the working age population. I drop all observations of respondents younger than 18 years and older than 65 years (the official German retirement age). I restrict the sample to respondents who report working in the current or the previous year and those who report being unemployed. I exclude people doing military service, people working in a sheltered workshop, and unemployed people not looking for work. The number of sick leave days is reported in absolute numbers and not in relative fractions of work time. The probability and the intensity of annual sick leave are biased when respondents work only a fraction of the year. To control for this bias, I exclude respondents who report fewer than five days or 35 hours of work a week. I also drop all respondents with a monthly income of less than 500 €.¹⁰

As for the time period, I use waves 1994 to 2013, corresponding to information about sick

⁹It also includes an SF-12 indicator of physical health. This measure combines several self-reported indicators; see Nuebling et al. (2007) for further information. Unfortunately, this measure is available only every other year since 2002 and is of limited use for the panel analysis.

¹⁰Altering the limits of days or hours of work a week or the minimum income does not change the qualitative results of the following empirical analysis.

leave from 1993 to 2012. Waves 1984, 1990 and 1993 do not contain information about sick leave. Waves 1991 and 1992 capture the unique economic situation of German reunification in 1990 and the liberalization of a state-owned socialist economy. I drop both periods because the income distribution and employment situation changed dramatically.¹¹ Waves 1985-1989 could potentially be used in the analysis of pro-cyclicality. I drop them for various reasons. First, these waves do not contain information about health and cannot be used in cross-sectional or panel regressions. Second, the unemployment rate varied in these years only between 7.9% and 8.1%. Hence, not much variation can be used to determine the cyclicity of sick leave. Third, I want to use an uninterrupted sample period for the time series analysis.

For the benchmark results, I exclude civil servants and self-employed workers from the sample. Self-employed workers do not receive paid sick leave; it is provided by the employer. Civil servants have paid sick leave but are not eligible for layoffs. Hence, they are not affected or are less affected by the indirect effect of fear of job loss.¹²

The days of sick leave have a highly skewed distribution with many observations at the 0 boundary and few observations at the highest value of 365. Of the observations, 95% percent report fewer than 42 days a year, and only slightly more than one percent report more than 120 days. Hence, many results, e.g., the average number of sick leave days, are prone to be driven by only a few observations. To control for outliers, I exclude in the benchmark results all observations that have one or more spells of sick leave that last longer than six weeks. Of the remaining sample, I cut off the highest two percentages, i.e., workers with more than 30 days of sick leave a year.

After sample selection, the sample used for the benchmark results consists of 100,526 observations.¹³ The sample includes 20 waves, and each wave has at least 3,910 observations.¹⁴

¹¹The following waves are also affected, but the effect is strongly mitigated over time. The classification in income quintiles is particularly disturbed because the income scale was lower in East Germany.

¹²The results of both groups are an additional argument for the proposed mechanism. The cyclical pattern is either not existent for the self-employed or much weaker for the civil servants, a result also found by Pfeifer (2013). Additionally, the income gradient in sick leave does not exist for both groups. Unfortunately, both groups vary from the rest of the sample in various respects (e.g., income, age, education). Therefore, they cannot be used as an adequate control group.

¹³Further details on the sample selection can be found in Appendix B.

¹⁴The size of the waves increases over time. There were refreshments of the SOEP in 1998, 2000, 2002 and 2006.

2.1.3 Empirical Approach

In Section 2.2, I run cross-sectional regressions of sick leave on various regressors, where I pool the observations of all waves in one sample. The regression equations for OLS and Logit estimation are

$$S_i = \alpha + \beta \log(W_i) + \beta H_i + \beta \mathcal{U}_i + X_i\theta + \varepsilon_i \quad (1)$$

$$\text{Logit} [S_i^{ext} = 1] = \mathcal{P}\{\alpha + \beta \log(W_i) + \beta H_i + \beta \mathcal{U}_i + X_i\theta + \varepsilon_i\} \quad (2)$$

where S_i is a countable variable used to denote days of sick leave, whereas S_i^{ext} is a dummy variable used to denote either missing any day in a year at work ($S_i^{ext} = 1$) or not ($S_i^{ext} = 0$). W_i is the monthly income of the respondent in the previous year.¹⁵ H_i is self-reported health, \mathcal{U}_i the German unemployment rate, $X_{i,t}$ is a set of control variables (e.g., sex, age, years of education, marital status, number of children living in the household, year dummies), and ε_i is the random error term.

In Section 2.3, I employ a logistic panel regression. I estimate the effect of the number of sick leave days in the previous period on the current probability of unemployment [$I_{i,t}^U = 1$]. I restrict the sample in this section to people who were employed at least six months in the previous year. The panel structure of the SOEP additionally allows me to use a fixed-effects model. The fixed effect will incorporate all unobserved characteristics of the agent. The regression equations are

$$\text{Logit} [I_{i,t}^U = 1] = \mathcal{P}\{\alpha + \beta S_{i,t-1} + X_{i,t-1}\theta + \epsilon_{i,t}\} \quad (3)$$

$$\text{Logit} [I_{i,t}^U = 1] = \mathcal{P}\{\alpha + \beta S_{i,t-1} + C_{i,t-1}\theta + a_i + \epsilon_{i,t}\} \quad (4)$$

where $C_{i,t}$ is a set of control variables that vary over time. It contains health, age and the log income in the previous year. Days of sick leave in the previous period are denoted by $S_{i,t-1}$. The a_i s represent the individual specific and time-invariant fixed-effect component, and $\epsilon_{i,t}$ is the random error term.

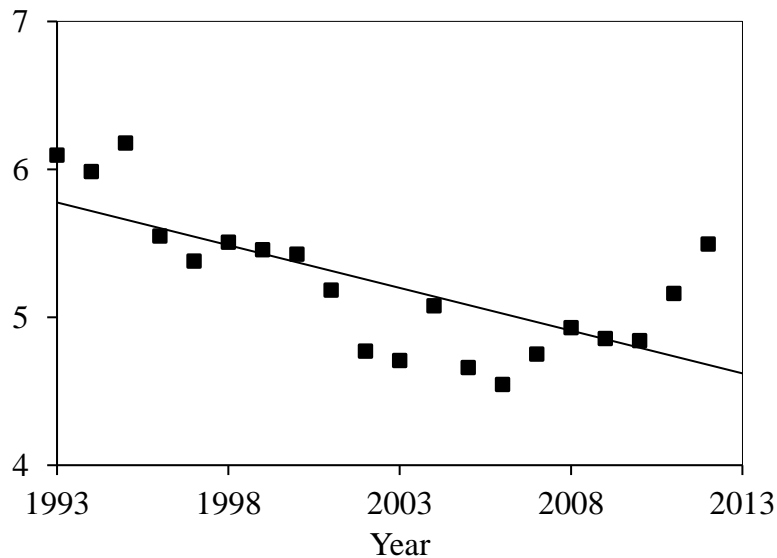
¹⁵I use other measures of income that are also included in the SOEP such as gross income or household net income. The qualitative results do not change.

2.2 Stylized Facts on Aggregated Data

2.2.1 Time Series

Figure 1 shows for the benchmark sample the average annual number of sick leave days per worker in the observed time period and a fitted linear trend. The average number of sick leave days varies between 4.5 and 6.2 days. An obvious first finding about the days of sick leave in Germany is the long-term decline. In the last 19 years, the average claims of sick leave have dropped by almost one day, or 20% relative to 1993.

Figure 1: Average Days of Sick Leave per Worker 1993-2012

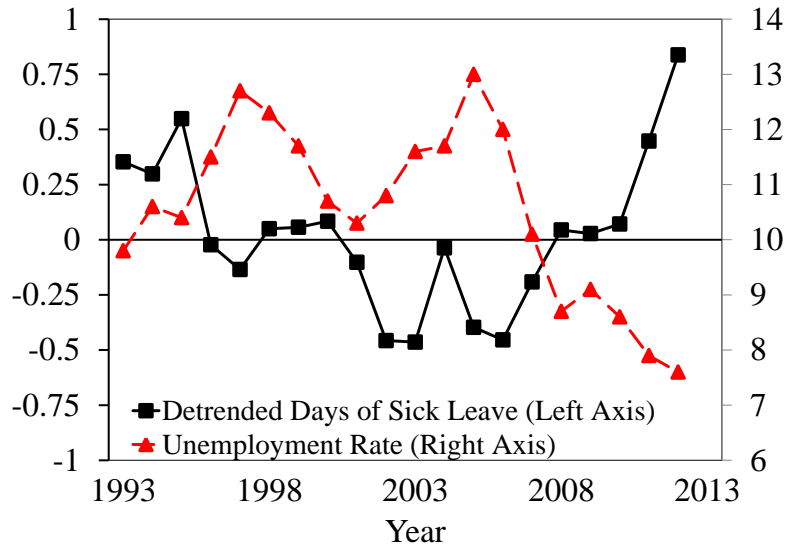


Notes: Dots: Average annual claims of sick leave per worker in benchmark sample. Solid line: Fitted linear trend, slope: -0.0578 .

A second characteristic is the strong pro-cyclical pattern of the average claims of sick leave in Germany once the time series is de-trended. Figure 2 depicts the absolute deviation of average days of sick leave from the linear trend (dashed blue line) and the unemployment rate for Germany (solid black line). The average number of days of sick leave is high when the German unemployment rate is low and vice versa. The correlation between the de-trended time series of days of sick leave and the German unemployment rate for the benchmark sample is minus 0.6383.¹⁶

¹⁶In Appendix C, I provide additional robustness checks using other measure of central tendencies, e.g., the number of days of sick leave for the median worker.

Figure 2: De-trended Average Days of Sick Leave and Unemployment Rate



Notes: Dashed line (left axis): Absolute deviation of average annual claims of days of sick leave per worker from linear trend in benchmark sample. Solid line (right axis): German unemployment rate.

To control for a composition effect, i.e., the absence behavior of the marginal worker, I construct a sub-sample consisting of workers who report never being unemployed for at least five consecutive years. This sub-sample shows, on the one hand, a lower number of average days of sick leave compared to the benchmark sample. On the other hand, the cyclical pattern of the always-employed sample is still distinctively negative, with a correlation of minus 0.6576.¹⁷ The remaining correlation supports the idea of an incentive effect. In times of low reemployment, workers reduce their days of sick leave to avoid unemployment. The incentive effect implicitly assumes that absence from work is not mechanically tied to the incidence of sickness. Workers are free to decide whether to go to work sick or stay at home and recover. This is a key assumption of the structural model in section 3.

2.2.2 Cross Section

This incentive effect, i.e., the economic trade-off between recuperation time on the one hand and an increased layoff probability on the other hand, can also be observed in cross-sectional

¹⁷Other composition effects would occur if specific occupation groups or sectors that exhibit many days of sick leave, e.g., the construction sector, are hit stronger by business cycles than others. In Appendix C, I show that the general pattern of pro-cyclical behavior holds for all occupation and sectors.

analyses. For each wave and age, I classify all respondents into one of five income quintiles based on their monthly earnings in the previous year.¹⁸ Figure 3 plots the average claims of sick leave (solid line) for each income quintile. It additionally shows the average self-reported health (dashed line) for each quintile. Workers in the top income quintile miss the fewest days at work due to sickness. Workers in the medium income quintile claim on average 10% more days of sick leave than workers in the bottom income quintile.¹⁹ Conversely, the health profile across income quintiles is monotonically increasing. The poorest workers have the lowest average health, whereas the top income quintile shows the highest average health.²⁰

Health and days of sick leave are naturally related; i.e., workers in generally bad health conditions are more likely to become sick and stay at home. Consequently, the observed differences in health could explain the small use of sick leave in the top income quintile compared to the rest of the workforce. Rich people are, on average, in better health condition; they are therefore less sick and require fewer days at home to recover. However, the same rationale is puzzling on the other side of the income distribution. The workers who are most unhealthy use fewer days at home to recover than medium income workers, who enjoy, on average, better health. This finding shows that sick leave is not mechanically tied to health, and the absence behavior of workers must have a second economic determinant.

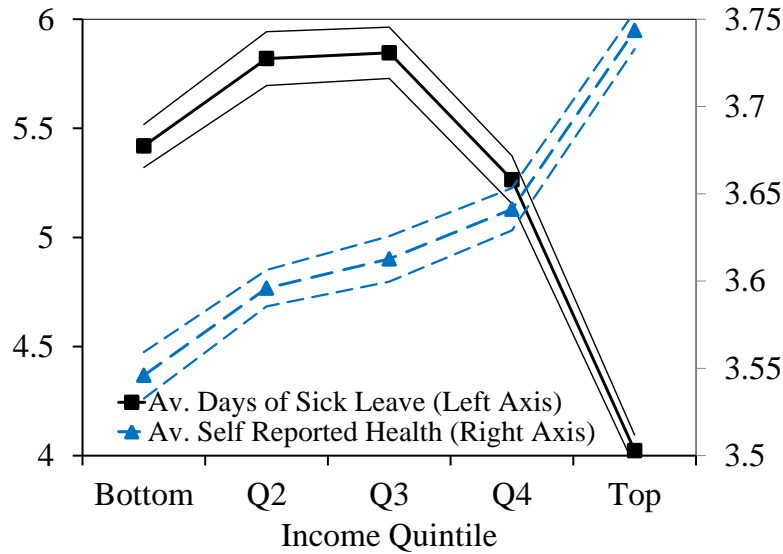
This graphical inspection is confirmed by estimating Equation (1) using days of sick leave as the dependent variable. Both income coefficients, for log income and for log income squared, are highly significant and suggest a hump-shaped relationship of income and days of sick leave. Health has the assumed protective effect against days of sick leave. Other coefficients confirm former results; see Table 1. There is a long-run negative trend in claims of sick leave of minus 0.0488 days per year. More importantly, related to the cyclicity of days of sick leave, the coefficient of the unemployment rate is negative and highly significant. This indicates that

¹⁸Controlling for age in the quintile classification is important; otherwise, older people are more likely to be in the top income quintile. Because older respondents are, on average, less healthy, the relationship between income and days of sick leave would be biased.

¹⁹These patterns also hold for each age bin separately and for both sexes; see Figure 14 and Figure 15 in Appendix C. The patterns also exist for the median number of days of sick leave and other cut-off levels for days of sick leave.

²⁰A simple probit regression of health (good or bad) on income controlling for age and sex confirms this pattern and yields a highly significant positive effect for log income. This income gradient in health is well established in the literature; see Smith (1999)

Figure 3: Average Days of Sick Leave and Health across Income Quintiles



Notes: Solid lines (left axis): Average days of sick leave of workers and 95% bootstrap confidence interval. Income quintiles are based on the monthly gross income of the respondent. Dashed lines (right axis): Average self-reported health and 95% bootstrap confidence interval. Health is reported on an ordinal five-point scale, where 0 denotes "bad" health and 4 denotes "very good" health.

during periods of high unemployment, the average days of sick leave are reduced.

Further insights into the characteristics of claims of sick leave are provided by distinguishing between the extensive margin, i.e., missing any day in a year or not, and the intensive margin, i.e., conditional on missing at least one day a year, the number of days the respondent is absent from work.

The left panel of Figure 4 shows the extensive margin of sick leave. Workers in the bottom income quintile exhibit the lowest probability of missing any day in a year. The higher the income quintile the higher is the probability to miss at least one day. Only at the very top of the income distribution does the extensive margin decrease. This pattern is confirmed by estimating Equation (2) using the extensive margin as a dependent variable. The results are presented as marginal effects at the means in the second column of Table 1. The estimate for log income is highly significant and confirms the positive relationship between income and the probability of missing any day. Further results show an unsurprising protective effect of self-reported health against missing any day, and that cyclicity is also present in the extensive margin.

Table 1: OLS and Logit Regressions of Days of Sick Leave on Income

Days of Sick Leave	(1)	(2)	(3)
Log Income	10.0384***	0.0496***	-1.7516***
Log Income Squared	-0.6633***	–	–
Health	-1.4626***	-0.0805***	-1.2924***
Wave	-0.0488***	0.0035***	-0.1383***
Unemployment Rate	-0.1599***	-0.0063**	-0.1878***
Other Controls	Yes	Yes	Yes
Observations	51,179	51,179	28,216

Notes: *** Significant at $\alpha=0.01$, ** Significant at $\alpha=0.05$, * Significant at $\alpha=0.1$.

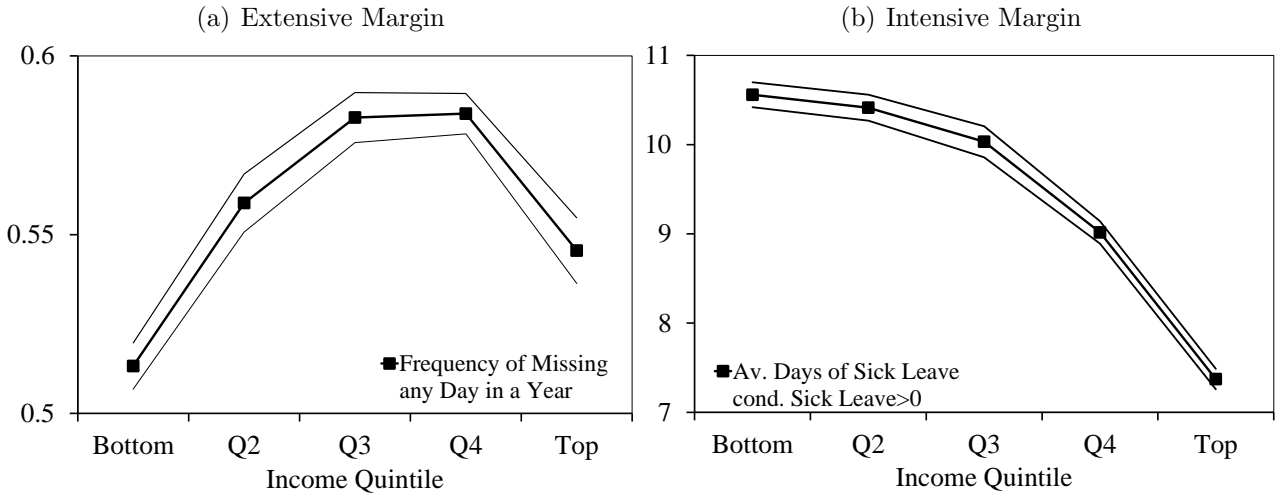
All regressions are based on benchmark sample. Other controls include sex, age, age², years of education, marital status, number of children and dummies for industrial sector. Robust standard errors are clustered on the personal level. Entries in column (2) show marginal effects at means.

The right panel of Figure 4 displays the average days of sick leave conditional on missing at least one day (intensive margin). In contrast to the extensive margin, the intensive margin is monotonically decreasing across income quintiles. The decline in conditional averages originates from different distributions of claims of sick leave. The upper income quintiles have a higher probability of experiencing few days of sick leave (1 up to 14 days). Workers in the lower income quintile have a higher probability of claiming many days of sick leave (more than 14 days) once they miss one day.²¹ The third column of Table 1 presents results of estimating Equation (1) using the intensive margin as the dependent variable. The results confirm that a high income has a protective effect against many days of sick leave. As the income increases, the number of days of sick leave a year conditional on being sick decreases. Health has again a protective effect against days of sick leave. The coefficient of the unemployment rate also confirms the cyclical pattern for the intensive margin.

Summarizing, there is a remarkable difference in the utilization of sick leave between income groups. The top income quintile has the lowest average days of sick leave but a high probability of missing a day. The medium income quintile has the highest average days of sick leave. The bottom income quintile has the worst health but only a moderate number of days of sick leave. Furthermore, the lowest income group has the highest probability of both not missing any day and missing more than 14 days compared to other income groups.

²¹The density functions of days of sick leave for the bottom and top income quintiles are shown in Figure 13 in Appendix C.

Figure 4: Composition of Sick Leave across Income Quintiles



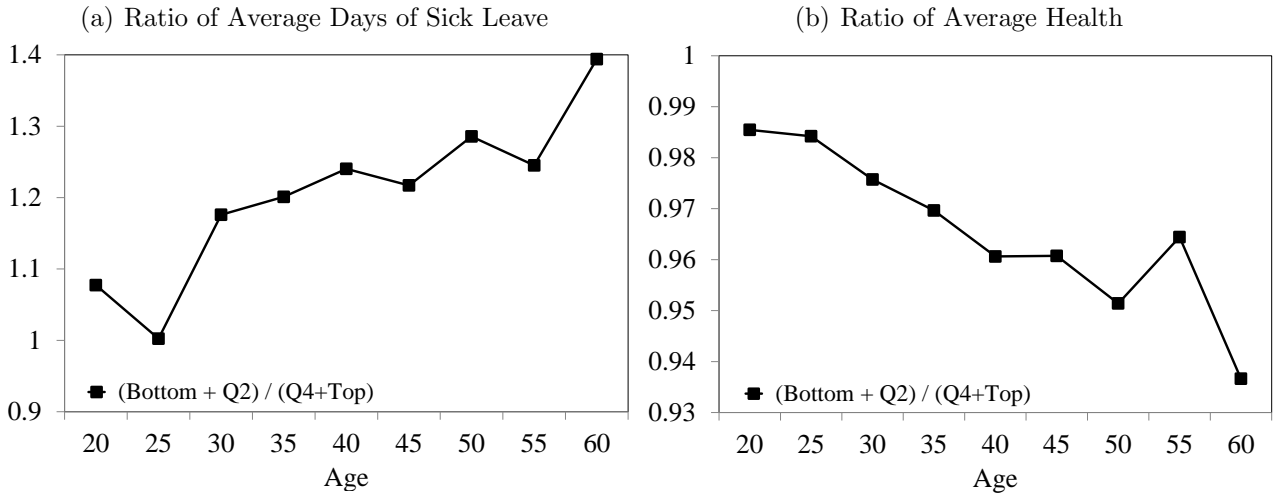
Notes: Left panel: Frequency of absence at least one day a year from work (extensive margin) of workers and 95% bootstrap confidence interval separated for income quintiles. Right panel: Average days of sick leave conditional on being at least one day absent from work (intensive margin) and 95% bootstrap confidence interval separated for income quintiles.

2.2.3 Life Cycle

The left panel of Figure 5 presents the ratio of the average days of sick leave of the bottom two to the top two income quintiles. It shows that the average number of days of sick leave is almost the same across income quintiles for workers in their twenties. Over the life cycle, the gap between the bottom and top income groups widens, and shortly before retirement, low income workers have almost 40% more days of sick leave than their high-income peers. The right panel of Figure 5 presents the same ratio but for average health. Here, I similarly find that the gap between rich and poor workers widens over the life cycle. At the beginning of working life, the average health of both groups is almost the same. At retirement, the average health of the bottom two income groups is considerably lower than average health of the top two income groups.

These empirical findings are further evidence for the proposed mechanism that poor people reduce their recuperation time to retain their jobs and hazard the consequence of a reduction in overall health over time.

Figure 5: Ratio of Bottom vs. Top of Sick Leave and Health over Life Cycle



Notes: Left Panel: Ratio of average days of sick leave of the bottom two income quintiles (Bottom & Q2) to the top two income quintiles (Q2 & Top) over 5-year age bins. Right Panel: Ratio of average self-reported health of the bottom two income quintiles to the top two income quintiles over 5-year age bins.

2.3 Micro Evidence Using Panel Data

The panel structure of SOEP allows me to carve out further features of sick leave in Germany. A first finding is that days of sick leave are persistent. People who claim sick leave in the last period also have higher claims of sick leave in the current period. Including lagged days of sick leave in estimating Equation (1) returns a positive and significant estimate. All other results remain qualitatively unchanged.

More important is the relationship of days of sick leave and the employment prospect of workers. Table 2 shows descriptive statistics of workers who report continuous employment in the entire previous year. In columns (2) and (3), this sample is separated by employment status in the current period. Workers who switched from employment to unemployment are not surprisingly poorer, less educated and less healthy than respondents who stayed employed. They are also slightly younger. Regarding the days of sick leave, respondents who are unemployed in the current period missed in the previous year on average one day more from work than workers who stayed employed.

Table 3 shows the odds ratios of estimating Equation (3). In the first column are results of a regression that includes only controls contained in the structural model. These coefficients will

Table 2: Summary Statistics of Workers Employed in t-1

Employed in t-1	Whole Sample	Employed in t	Unemployed in t
Days of Sick Leave in t-1	4.83	4.81	5.85
Days of Sick Leave in t-2	4.46	4.44	5.31
Age	41.1	41.1	40.2
Income in t-1	2,934€	2,952€	1,977€
Years of Education	12.3	12.4	11.5
Health in t-1	2.65	2.66	2.46
Male	64%	64%	52%
Observations	102,125	99,025	3,100

Notes: Descriptive statistics for the sample used in the panel logit model. Only workers who report employment the entire previous year. Unemployed people in t are classified as people who report at least one month of unemployment in t.

later be used in the calibration of the model. The second column contains results of regressing Equation (3), additionally including other control variables, e.g., health, age, and education. In both columns, income in the previous period has an odds ratio smaller than 1; i.e., the highest probability of becoming unemployed is among workers in the bottom income quintile. To be in good health, on the other hand, has a protective effect against the risk of unemployment. Naturally, the risk of unemployment is large in times of a high overall unemployment rate. The key results here are the coefficients for the days of sick leave. In both columns, days of sick leave show odds ratios that are greater than 1 and highly significant. Three additional days of sick leave lead to an increase in the probability of being unemployed in the next period by a factor of 1.1; e.g., a 10% layoff probability would then be 11%.

To account for unobserved heterogeneity in workers, I estimate Equation (4) including a fixed-effect component (column 3). The effect of sick leave on the probability of becoming unemployed is qualitatively not affected and is still highly significant. The results for the effect of income and the unemployment rate on the risk of unemployment remain unchanged. Health still has the same qualitative sign but becomes insignificant in the model with fixed effects.

In the relationship between sick leave and employment exists the problem of reverse causality. Workers who know that they will lose their job could be tempted to take sick leave because they do not fear retaliation. Note that the period length is one year, and therefore, this effect should not be present in the preceding period. Estimating Equation (4) using days of sick leave

in t-2 instead of t-1 still yields an odds ratio that is significantly greater than 1 (column 4).²²

Table 3: Panel Results for Sick Leave and Unemployment

Unemployed in t	(1)	(2)	(3)	(4)
Days of Sick Leave in t-1	1.0328***	1.0287***	1.0270***	–
Days of Sick Leave in t-2	–	–	–	1.0149**
Income Quintile in t-1	0.6132***	0.6132***	0.9129**	0.9519
Unemployment Rate in t-1	1.1626***	1.0798***	1.2376***	1.2542***
Health in t-1	–	0.8873***	0.9248	0.9242
Other Controls	No	Yes	Yes	Yes
Fixed Effects	No	No	Yes	Yes
Observations	60,052	60,052	7,578	4,868

Notes: *** Significant at $\alpha=0.01$, ** Significant at $\alpha=0.05$, * Significant at $\alpha=0.1$.

Unemployed in t is defined as reporting unemployment for at least one month in the current year. Other controls include age, sex, years of education, number of children, marital status, and year dummies. All columns report odds ratios at the population average. Robust standard errors are clustered.

3 Structural Model

In this section, I describe a dynamic stochastic model of work absence decisions over the life cycle. It captures both the standard consumption-saving decisions and sequential decision-making behavior of employed individuals regarding their health. I will later use this model to evaluate the consequences of economic inequalities for the utilization of sick leave and health as well as the consequences emanating from different policies.

3.1 Household's Problem

The economy is populated by overlapping generations of a continuum of agents that live up to a maximum age of J^T .²³

²²This specification, however, has the disadvantage of reducing the sample size and potentially biasing the results. In looking at the effect of days of sick leave in t-2 on the employment status in t, I have to exclude all respondents who became unemployed in period t-1. Many days of sick leave in t-2, however, increases the likelihood of unemployment in t-1 and therefore downward biases the results seen in column (4).

²³Unlike most of the health economics literature, I abstract for simplification from survival rates. Making the survival rate health dependent would further strengthen the mechanism because rich people would have an additional incentive to invest in health.

Health, Acute Sickness, and Sick Leave One important feature of this paper is the distinction between the general health status, H_t , the event of acute sickness, S_t , and the time an agent missed work, l_t .

Health status, H_t , reflects the overall constitution of agents. It is a persistent state that adjusts only gradually. Agents start their economic life in a certain health state, H_0 . At the end of each period, agents may drop into the next lower health state with probability Π^W , ascend into the next higher health state with probability Π^B , or stay in the same health state with probability $1 - \Pi^B - \Pi^W$.

In contrast, acute sickness, S_t , has a transitory notion and mimics the contraction of an illness or an injury, e.g., the flu or back pain. At the beginning of each period, individuals face the risk of either staying well, S_t^0 , or contracting one of m types of acute illnesses, S_t^m , which vary in severity with $S_t^m < S_t^{m+1}$.

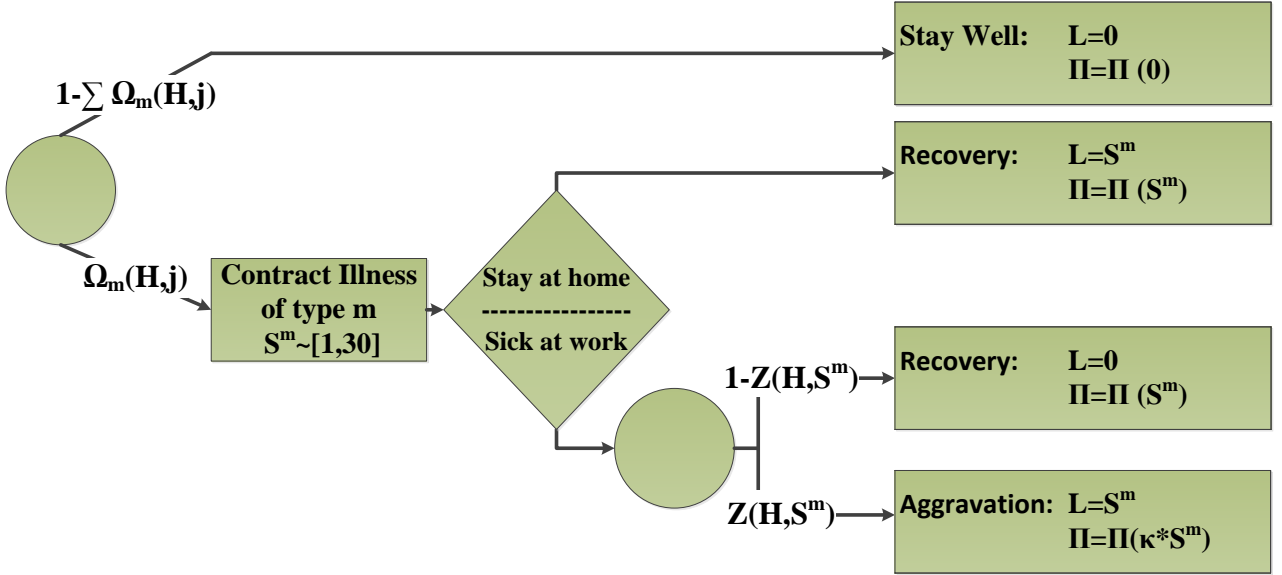
Health and acute sickness are interdependent. On the one hand, the probability of contracting an illness of type m , $\Omega_m(H_t, j)$, depends on the overall health status and the age, j , of an individual. As the general health status decreases, the probability both to become ill (extensive margin) and to contract a more severe illness (intensive margin) increases. The severity of the contracted sickness, on the other hand, affects the health transition probabilities, with $\Pi^B(S_t^m) < \Pi^B(S_t^{m+1})$ and $\Pi^W(S_t^m) > \Pi^W(S_t^{m+1})$.

The decision to take sick leave, l_t , depends essentially on the presence of an acute sickness shock. When staying well, $S_t = S_t^0$, there is no reason to be absent from work, $l_t = S_t^0 = 0$. Only when becoming ill do individuals have to decide whether to take sick leave. By being absent from work, I assume for simplicity of the model that individuals are required to take a fixed amount of days of sick leave according to the severity of their illness, $l_t = S_t^m > 0$. By going to work sick, workers face two possible consequences. First, with the health state dependent probability $1 - Z(H_t, S_t^m)$, workers are fortunate and recover without taking recuperation time, i.e., $l_t = 0$. The health transition probabilities are the same as if they had stayed at home. Second, with probability $Z(H_t, S_t^m)$, the sickness aggravates; e.g., a cold becomes the flu. In this case, workers are forced to stay at home, $l_t = S_t^m$. In addition, an aggravation of sickness leads to lower health prospects; i.e., the sickness factor in health transition is multiplied by a

factor of κ .

Note that the decision of going to work sick or not is made simultaneously with the consumption-saving decision at the beginning of the period. Workers cannot change their consumption if their sickness aggravates. The model also ignores preventive treatment, so individuals cannot improve their overall health by taking sick leave without being sick.

Figure 6: Events and Decisions - Sickness, Sick Leave and Sickness Aggravation



Notes: Circles denote events, rectangles denote realizations, diamonds denote decisions.

Preferences Individuals value consumption, c_t , and general health, H_t , over their life cycle according to a standard time-separable utility function

$$\mathbb{E} \left\{ \sum_{j=0}^J \beta^j u(c_t, H_t) \right\}$$

where β is the raw time discount factor, and expectations are taken over a stochastic employment, health and sickness history. Days of sick leave do not enter the utility function directly but rather yield utility from an increased probability of being in a higher health state in the future.²⁴

²⁴Including both sick leave and overall health in the utility function would not alter the qualitative predictions of the model; the main trade-off between consumption and health/sick leave remains. Augmenting the utility function by an additional argument, however, significantly complicates the identification of preferences

Labor Income and Employment Status There are two sources of heterogeneity that affect an agent’s labor productivity, $\Gamma_{j,k}$, in this model. First, the labor productivity differs according to the age of an agent. Second, each household belongs to a particular group, k , that shares the same productivity. Differences in groups stand in for differences in education or ability, characteristics that are fixed at entry into the labor market and affect a group’s relative income.²⁵ It is important to note that neither the general health status nor acute sickness directly affect labor productivity.²⁶ Agents get paid only for the time they work; i.e., working time is reduced by sick leave l_t .²⁷ The wage rate is w . Labor income $y_{j,k,t}$ is then given by

$$y_{j,k,t} = \Gamma_{j,k}w(1 - l_t)$$

In addition to labor productivity and days of sick leave, the labor income of agents depends crucially on their employment status \mathcal{I}_t . At the end of each period, agents may be dismissed with probability $1 - \Phi^e$. Central to this paper is that workers are able to influence their layoff probability by reducing absence from work, and they take this into their optimization reasoning.²⁸ The probability of staying employed, Φ^e , depends, in addition to days of sick leave, on the skill type of workers and the current unemployment rate, Π_t .

$$\Phi^e = \Phi^e(l, k, \mathcal{U})$$

The probability of finding a new job when unemployed, Φ^u , is again determined by workers’

parameters.

²⁵By focusing on innate differences across workers instead of idiosyncratic skill shocks as the main source of heterogeneity I follow Keane and Wolpin (1997), Storesletten et al. (2004) and Huggett et al. (2011). They estimate that innate heterogeneity is a major contributor to the observed variations of lifetime earnings, contributing 50-90%.

²⁶Health-dependent labor productivity would reinforce the inequality between ability groups because low-productivity workers are less healthy. For simplicity of the model, I refrain from this additional channel.

²⁷The maximum amount of working days per year in this economy is set to 250.

²⁸For simplification reasons, I refrain from modeling the demand side of the labor market. There are various theoretical explanations for the described linkage from an employer’s perspective. First, health is important for the productivity of a worker. However, employers might be unable to directly observe health and instead use days of sick leave as a signal for health. Table 1 shows the high correlation of sick leave and health. Second, related to the literature on shirking, employers can only imperfectly monitor workers in terms of work absence. As the numbers of days of sick leave increases, so does the probability of being discovered when shirking. Third, due to the structure of the German sick leave system, days of sick leave are costly for employers. Days of sick leave are persistent over time. A simple test for auto-correlation yields a coefficient of 0.34. Therefore, the incentive for employers is high to get rid of workers with many days of sick leave to decrease these costs. In all these cases, a higher number of sick leave days leads to a higher layoff probability.

skill type and the current unemployment rate. Unemployed workers do not take sick leave.

$$\Phi^u = \Phi^u(k, \mathcal{U})$$

Unemployment Rate The evolution of the unemployment rate, \mathcal{U}_t , in my model is exogenous (i.e., I do not model general equilibrium effects). I assume that the unemployment rate in the model follows an AR(1) process approximated by a 5-state Markov process.

3.2 Government Policies

The government imposes a flat income tax, τ . The collected revenues are used for three main purposes:

(i) to finance unemployment benefits b^U . The unemployment insurance (Arbeitslosengeld I) replaces a constant fraction of the previous net income and therefore depends on the age and skill of an individual. When this unemployment insurance falls below some consumption floor c^W , workers might become eligible for government provided welfare (Arbeitslosengeld II). This welfare is means tested; i.e., before workers receive welfare, they have to run down their assets to a certain amount, a^W . Unemployment benefits are then given by

$$b_{j,k,i}^U = \begin{cases} c^W & \text{if } \rho^U(1-\tau)\Gamma_{j,k}w < c^W \quad \cap \quad a_i \leq a^W \\ \rho^U(1-\tau)\Gamma_{j,k}w & \text{else} \end{cases}$$

where ρ^U is the unemployment insurance net replacement rate, and a_i are the assets of worker i .

(ii) to finance paid sick leave b^S . For the time absent from work, l_t , workers get reimbursed by the government with payments depending on their regular labor income.²⁹ Sick leave benefits are given by

$$b_{j,k}^S = \rho^S(1-\tau)\Gamma_{j,k}w$$

where ρ^S is the sick leave net replacement rate.

(iii) to finance retirement benefits, b^R . Workers receive these skill-dependent retirement

²⁹This reimbursement is actually provided not by the government but by the employer. Because I do not model the firm site, I make this shortcut.

benefits after their fixed retirement age J^R .³⁰ Retirements benefits are given by

$$b_{j,k}^R = \rho^R(1 - \tau)\Gamma_{j^R,k}w$$

where ρ^R is the retirement system net replacement rate.

3.3 Individual's Dynamic Program

I model the decisions to miss work during an episode of acute illness as the choices of workers solving a discrete choice stochastic dynamic programming problem. At each discrete period, the forward-looking individual chooses whether to miss work based on expected utility maximization. Individuals can accumulate assets, a , at a constant interest rate R . They are not allowed to borrow. They allocate their total resources between consumption, c , and asset holdings a_{t+1} for the next period.

At the beginning of period t , individuals are indexed by their age j , their skill group k , their asset holdings a , their health status H , their realization of acute sickness S , and their employment status \mathcal{I} . To simplify the analysis, I assume that the factor prices are exogenous. Each individual starts her life in a specific health state H_0 and employment state \mathcal{I}_0 and is endowed with initial assets a_0 . Thus, her maximization problem reads as

$$\begin{aligned} V(j, k, a_t, H_t, S_t, \mathcal{I}_t) = & \max_{c_t, l_t, a_{t+1}} u(c_t, H_t) \\ & + \beta \sum_{H_{t+1}} \Pi(j, k, H_t, S_t, l_t) \sum_{S_{t+1}} \Omega(j, H_{t+1}) \sum_{\mathcal{I}_{t+1}} \Phi(k, l_t, \mathcal{U}_t) \\ & V(j + 1, k, a_{t+1}, H_{t+1}, S_{t+1}, \mathcal{I}_{t+1}) \quad (5) \end{aligned}$$

³⁰To reduce the state space in the quantitative approach, these retirement benefits do not depend on the history of idiosyncratic employment shocks. This is a deviation from the actual German system. Lower pension benefits from increased periods of unemployment would strengthen the underlying mechanism of the model because, again, more days of sick leave reduce future expected earnings.

subject to the constraints

$$a_{t+1} + c_t = \begin{cases} (1 - \tau) (\Gamma_{j,k} w (1 - l_t) + l_t b_{j,k}^S) + Ra_t & \text{if } j < J^R \cap \mathcal{I}_t = 1 \\ b_{j,k}^U + Ra_t & \text{if } j < J^R \cap \mathcal{I}_t = 0 \\ b_{j,k}^R + Ra_t & \text{if } j \geq J^R \end{cases}$$

$$a_{t+1} \geq 0$$

4 Quantitative Analyses

In this section, I begin by discussing the specification of the model parameters. Then, in Section 4.2, I present simulation results and their counterparts in the data to evaluate the model's performance. These results contain cross-sectional distribution and lifetime profiles of days of sick leave and health and the average days of sick leave for a time series of unemployment rates.

4.1 Parameter Estimation and Calibration

To fully determine the structural model, I need to choose parameters that govern the employment status, incidence of sickness, health transitions, preferences, and policy settings. The determination of the model parameters proceeds in three steps. First, I fix a subset of parameters exogenously. Second, parts of the model parameters can be estimated from SOEP data directly. These include parameters governing labor productivity Γ ; probabilities to keep employment Φ^e , to find a new job, Φ^u , and to contract acute sicknesses Ω ; health transition probabilities Π ; and the Markov process of the overall unemployment rate. Third (and given the parameters obtained in steps 1 and 2), the remaining parameters (i.e., governing sickness aggravation Z and κ) are determined through a method of moments estimation. I now describe these three steps in detail.

4.1.1 A Priori Chosen Parameters

In the first step, I choose parameters from the literature or set them according to the data. All values of the a priori chosen parameters are shown in Table 11 in Appendix D.

General Settings The model period is one year. Workers start their economic life at age 20, retire at age 65 and live until age 80. Because I do not model childhood of a household explicitly, I denote its 20th year of life by $j=0$, its retirement age by $J^R=45$ and the terminal age of life by $J^T=60$. After retiring, the problem of agents is reduced to a consumption-saving decision.

I use self-reported health status as an empirical counterpart from the SOEP for the model's health state. Thus, H_t takes one of five values: 0-"bad", 1-"not so good", 2-"satisfactory", 3-"good", or 4-"very good".

I assume that the interest rate, R , is determined exogenously by world factors in an open-economy equilibrium, and following Siegel (2002), I set $R=1.0402$. The wage rate is normalized to $w=1$.

Preferences I choose for the instantaneous utility function the specification of Finkelstein et al. (2013). They estimate using household panel data a health state-dependent utility function with the functional form:

$$u(c_t, H_t) = (1 + \psi(4 - H_t)) \frac{c_t^{1-\sigma}}{1 - \sigma}$$

where σ determines the inter-temporal elasticity of substitution, and ψ captures state dependence because it allows the state of health to affect the marginal utility of consumption.

Finkelstein et al. (2013) estimate that a one standard deviation increase in the number of chronic diseases is associated with an 11% decline in the marginal utility of consumption relative to this marginal utility when the individual has no chronic disease. This implies that $\psi=0.112$.³¹ I choose $\sigma=2$ to obtain an inter-temporal elasticity of substitution of 0.5, which is a value widely used in the literature (e.g., Fernandez-Villaverde and Krueger (2007)). Consistent

³¹I deviate from Finkelstein et al. (2013) by substituting the number of chronic diseases (0-5) by the ordinal five health-state variables used in this model.

with values commonly used in the quantitative macroeconomics literature, I choose a time discount factor of $\beta=0.96$ per annum.

Policy Parameters For the benchmark calibration, I choose the current institutional setting for Germany. It shuts down two direct effects of income on health. First, Germany has universal health care coverage, so individuals do not have to pay for standard medical expenditures, e.g., doctor visits. Second, I set the benchmark paid sick leave coverage to 100% of the current wage, $\rho^S=1$, ruling out any reduction in the current income of workers due to sick leave; this step is important to isolate the indirect effects of sick leave via the risk of unemployment. I set the unemployment insurance net replacement rate to $\rho^U=0.60$. This is the current German setting for the first 12-24 months in Germany. The means-tested welfare provision, c^W , I set to 0.25 ($\triangleq 700 \text{ €}(2005)$), which represents the German basic tax-free allowance. The amount of assets of an individual that are protected without losing eligibility of welfare, a^W , I set to 0.44 ($\triangleq 1220 \text{ €}(2005)$).³² Retirement benefits are set to $\rho^R=0.50$ of the former net labor income of a worker. I set the tax rate τ to 33%.

4.1.2 Parameters Estimated Directly from the Data

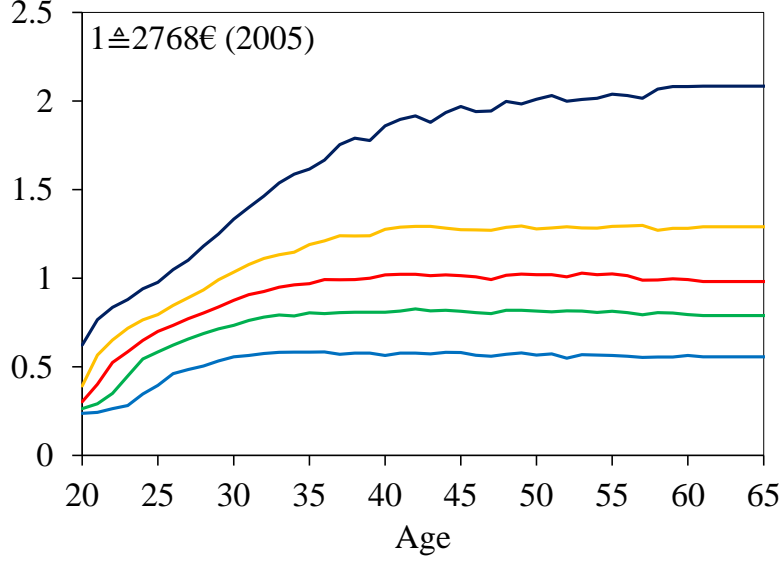
In a second step, I estimate a part of the model parameters directly from the SOEP data. Detailed results for the estimated parameters are shown in Table 12 - 13 in Appendix D.

Γ – Labor Productivity Using SOEP data on labor income, I compute age-dependent productivity for five different income quintiles. For each age (20-65) and wave separately, I split the sample into five income quintiles. Within each age-income cell, I take the median income as the labor productivity $\Gamma_{j,k}$ of workers in that quintile. Figure 7 shows the resulting income profiles over the life cycle. It exhibits the well-known increase in income at younger ages and the flattening out at older ages. The distance between the bottom and the top income quintiles widens over age. I normalize the life-cycle profiles using the income third quintile at age 40 as a basis ($1 \triangleq 2768 \text{ €}(2005)$).³³

³²The German unemployment insurance system underwent a major reform in 2005, which might limit the historical comparison between the model predictions and the data.

³³These results are broadly in line with other findings on the German income structure, cf. Hujer et al. (2001).

Figure 7: Life Cycle Profile of Labor Income by Income Quintiles



Notes: Labor income over the life cycle for five quintiles computed for the benchmark sample. Quintiles are defined for each age and wave separately.

Φ – Employment Transition For each combination of days of sick leave, skill group, and unemployment rate, the model predicts a certain probability of retaining the job Φ^e , which is given by

$$\widehat{\phi}^e = 6.775832 - 0.0331853 * l + \sum_k \beta_k^e * k^{Dummy} - 0.1506627 * \mathcal{U}$$

$$\Phi^e(l, k, \mathcal{U}) = \frac{e^{\widehat{\phi}^e}}{1 + e^{\widehat{\phi}^e}}$$

The used parameters are obtained by a panel logit regression using SOEP data. The resulting odds ratios of this estimation are shown in Section 2.3.

The probability of finding new employment when unemployed in the last period, Φ^u , is computed the same way except using days of sick leave.

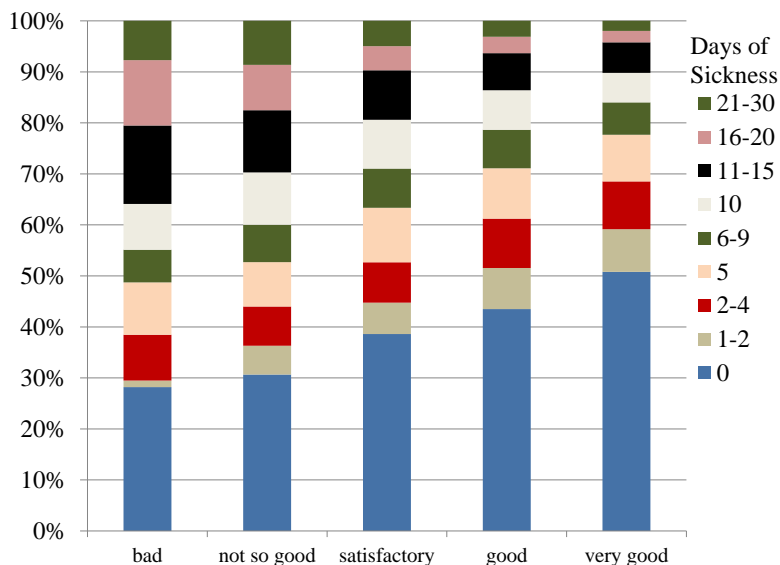
$$\widehat{\phi}^u = 4.347449 + \sum_k \beta_k^u * k^{Dummy} - 0.199108 * \mathcal{U}$$

$$\Phi^u(k, \mathcal{U}) = \frac{e^{\widehat{\phi}^u}}{1 + e^{\widehat{\phi}^u}}$$

Ω – Incidence of Sickness One shortcoming of the dataset is that it contains only days missed at work, l , and not the incidence of sickness, S . To compute the probabilities of sickness, I make the identifying assumption that workers with a high productivity always stay at home to recover, and therefore, observed sick leave is equal to the occurrence of a sickness, $S = l$.³⁴

The incident of sickness S depends on the current health state H of an agent. Individuals with lower health have both a higher probability of contracting a sickness and a higher probability that the sickness is more severe. In addition to health, the number of days of sick leave also seems to depend on the age of workers.³⁵ To take this trend into account, I estimate incident probabilities conditional on the health status and the age group of workers. To ensure sufficient observations for each sickness state, I restrict the number of sickness shock realizations to $m = 9$ states, $S \in [0, 1-2, 3-4, 5, 6-9, 10, 11-15, 16-20, 21-30]$. Figure 8 shows the frequency of days of sick leave conditional on health for the top two income quintiles for the specific age group 40-50.³⁶

Figure 8: Frequency of Days of Sick Leave Conditional on Health and Age



Notes: Frequency of days of sick leave conditional on self-reported health state. The sample consists of the fourth income quintiles for the age group 40-50.

³⁴In the benchmark calibration, I use only the second to the top income quintiles. The empirical analysis and the later simulation results suggest that the top income quintile is an outlier. In a robustness check, I include all observations in the estimation of the probabilities of sickness. The qualitative patterns remain the same.

³⁵This age dependency of sick leave can be observed in Figure 14 in Appendix D. During early working life, workers use on average more days of sick leave than later in their working life.

³⁶The results for all age-groups are shown in Table 13 in Appendix D.

II – Health Transition To estimate the effect of acute sickness on the general health status, it is important to differentiate between sickness and sick leave. Taking sick leave should be beneficial to health, whereas an acute sickness itself should harm overall health. As in the estimation of Ω , it is therefore necessary to restrict the sample to those for whom observed sick days are equal to sickness, i.e., the high-productivity workers.

To estimate health transition probabilities, I define a new variable

$$\Delta_{i,t,t-2} = \begin{cases} 1 & H_{i,t} > H_{i,t-2} \\ 0 & H_{i,t} = H_{i,t-2} \\ -1 & H_{i,t} < H_{i,t-2} \end{cases}$$

which marks a change, either better or worse, in the health status of an individual between period $t-2$ and t .³⁷ To assess the transition probabilities of health, I employ an ordered logistic panel regression. The regression equation is

$$OLogit[\Delta_{i,t,t-2}] = \mathcal{F}\{\alpha + \beta_1 S_{i,t-1} + H_{i,t-2}^{Dum} \beta_2 + \beta_3 k_i + \beta_4 j_{i,t-1} + \epsilon_{i,t}\}$$

where days of sick leave are denoted by $S_{i,t-1}$, $H_{i,t-2}^{dum}$ are dummies for each health state, k_i is the income quintile (skill type), and $j_{i,t-1}$ the age of the respondent. $\epsilon_{i,t}$ is a random error term.

Table 4 shows the results for the specification used for calibration (column 1) and a regression in which numerous additional controls are included (column 2). In both specifications, sickness has a negative effect on the probability of an improvement in health and consequently a positive effect on the probability of a deterioration in health. Age has a detrimental effect on health, reflecting a worsening health status over the life cycle. Being in a high income quintile has, in contrast to age, a protective effect against health deterioration. Including years of education makes the income effect insignificant, reflecting the strong correlation between income and education.

³⁷I choose a lag of two periods to compute health transitions. Doing so is necessary because the question about health refers to the point in time at which the interview is conducted (distributed over the year), whereas the question about days of sick leave refers to the time span of the whole previous year. This might lead to a situation in which the acute sickness occurs later in the year than the respondent answers the health question. In a robustness check, I look at a one lag health transition and find qualitatively similar results.

Table 4: Ordered Logit Results for Changes of Health

$\Delta_{t,t-2}$	(1)	(2)
Days of Sick Leave (t-1)	-0.0459***	-0.0414***
Age (t-1)	-0.0292***	-0.0296***
Income Quintile (t-1)	Yes	Yes
Health State Dummies(t-2)	Yes	Yes
Other Controls	No	Yes
Observations	40,460	34,935

Notes: *** Significant at $\alpha=0.01$, ** Significant at $\alpha=0.05$, * Significant at $\alpha=0.1$.

Other controls include sex, years of education, number of children, marital status, recession, and year dummies. All columns report the regression coefficient.

\mathcal{U} – Unemployment Rate I estimate the unemployment rate with an AR(1) process using the time span from 1994 to 2012 (with $\rho=0.9152$). This continuous AR(1) process is then approximated by a finite state Markov process applying Tauchen’s method, cf. Tauchen (1986). I restrict the number of unemployment rate states to $n=5$. Doing so leads to an unemployment rate state grid $\mathcal{U} \in [7.5\%, 9\%, 10.5\%, 12\%, 13.5\%]$ and a transition matrix, Ξ , given by

$$\Xi = \begin{pmatrix} 0.7939 & 0.2056 & 0.0005 & 0 & 0 \\ 0.0733 & 0.7752 & 0.1512 & 0.0002 & 0 \\ 0.0001 & 0.1072 & 0.7853 & 0.1072 & 0.0001 \\ 0 & 0.0002 & 0.1512 & 0.7753 & 0.0733 \\ 0 & 0 & 0.0005 & 0.2055 & 0.7939 \end{pmatrix}$$

4.1.3 Parameters Calibrated Within the Model

In the final step, I use my model to find parameters that govern the probability and severity of sickness aggravation. All estimated parameters and the fit of the targeted moments to the data can be found in Table 16 in Appendix D.

Sickness Aggravation Important for the decision to stay at home to recover or go to work sick is the aggravation probability, Z , and the aggravation factor κ .

For the aggravation probability, Z , I make two assumptions. On the one hand, it depends negatively on the overall health constitution, i.e., that people in a bad health state are much

more likely to face an aggravation than people in very good health. On the other hand, the aggravation probability depends positively on the severity of the sickness; i.e., being hit by a bad illness makes it very unlikely not to recover without taking sick leave. The functional form for Z is given by

$$\widehat{Z} = (H - 1)S^\zeta$$

$$Z(H, S) = 2 \frac{e^{-\widehat{Z}}}{1 + e^{-\widehat{Z}}}$$

where $0 \leq Z \leq 1$ and $\zeta \leq 0$. Parameters ζ and κ are calibrated to minimize the distance between the distribution of days of sick leave of the bottom income quintile found in the data and predicted by the model.

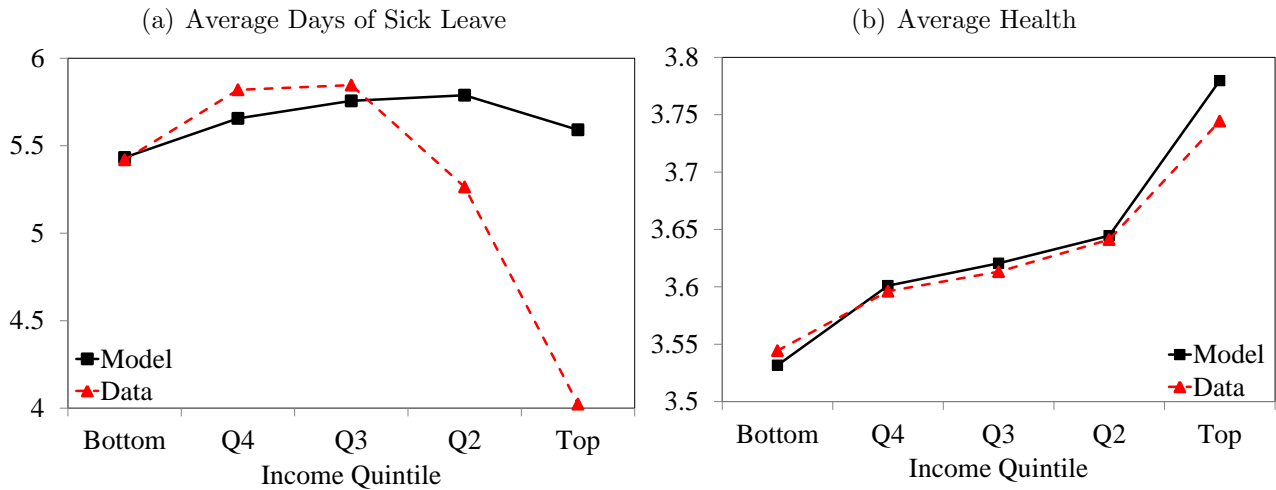
4.2 Model Fit and Benchmark Results

In this section, I examine the fit of the model to the non-targeted data moments. Moments of the model are computed by aggregating over the state distribution of the population. The initial states for health, H_0 , and employment, \mathcal{I}_0 , are drawn from distributions conditional on the skill type that match the data; see Table 14 - 15 in Appendix D. All individuals start their lives with no initial assets, $a_0=0$. For cross-sectional and life-cycle patterns, I fix the general unemployment rate constant at $\mathcal{U}=10.8\%$.

Cross Section The left panel of Figure 9 plots the simulated average days of sick leave across income quintiles (the solid black line) and the data counterpart (the dashed red line). The model is able to endogenously generate the key hump-shaped profile of average days of sick leave, particularly the increase in days of sick leave between the bottom and the medium income quintiles. This increase is driven by the indirect cost effect, i.e., that low income people reduce their recuperation time to reduce their layoff probability. Note that the model is also able to reproduce the decrease in the average days of sick leave for high-income individuals compared to medium-income workers, although the difference is not as large as that observed

in the data.³⁸ The right panel of Figure 9 shows the simulated average health for each income quintile and the data counterpart. The cross-sectional pattern is very close to the data moments and reflects the income gradient in health. Only in the very top income quintile does the model predict a higher average health than is found in the data.

Figure 9: Health and Days of Sick Leave across Income Quintiles



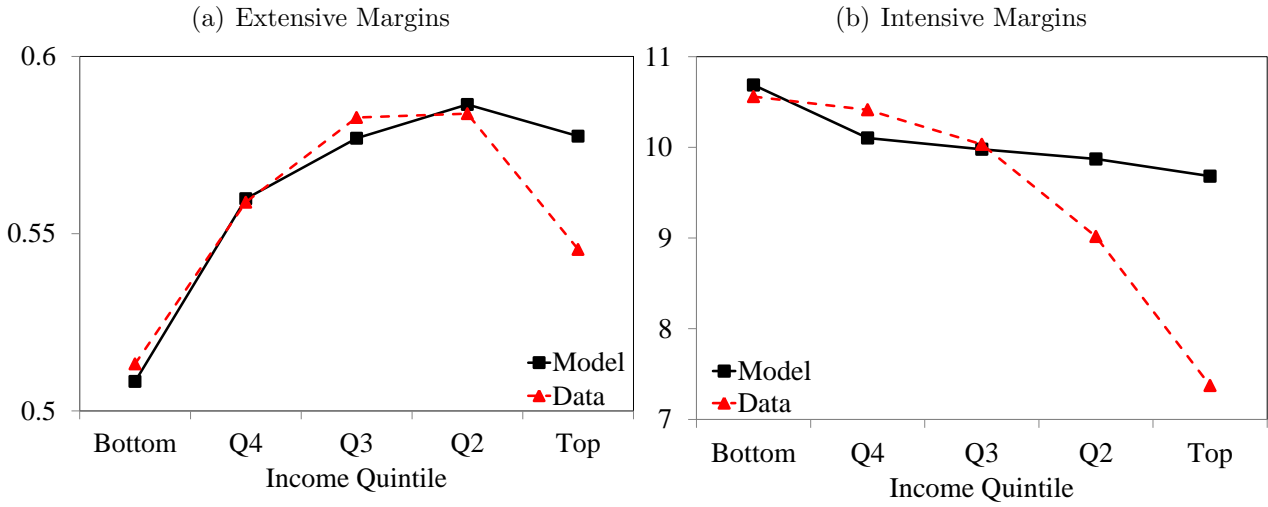
Notes: The dashed red lines show moments taken from the SOEP data. The solid black lines show averages from simulating the structural model using the parameters explained in Section 4.1.

Figure 10 shows the decomposition of the average days of sick leave into the extensive (left panel) and intensive (right panel) margins. The fit of the model prediction to the data for the probability of missing any day in a year is good. The model can account for the sharp increase in the probability of missing any day across the income quintiles. Again, the top income quintile has the worst fit. Looking at the intensive margins (the number of days missed conditional on being absent for at least one day), the model can generate the monotonic decrease over income quintiles. However, particularly for the top two income quintiles, the simulated level of the intensive margins exhibits a poor fit to the data moments.

Life Cycle In addition to cross-sectional moments, the model can replicate the widening gap in average health across the income quintiles over the life cycle. The left panel of Figure 11

³⁸The not-so-good fit of the model for the top income quintile hints that characteristics of very high income jobs, e.g., high responsibility, might also have an effect on the utilization of sick leave. Additionally, a higher work flexibility (e.g., a home office) allows for fewer days of sick leave compared to workers on an assembly line.

Figure 10: Extensive and Intensive Margins of Days of Sick Leave across Income Quintiles



Notes: The dashed red lines show moments taken from the SOEP data. The solid black lines show averages from simulating the structural model using the parameters explained in Section 4.1.

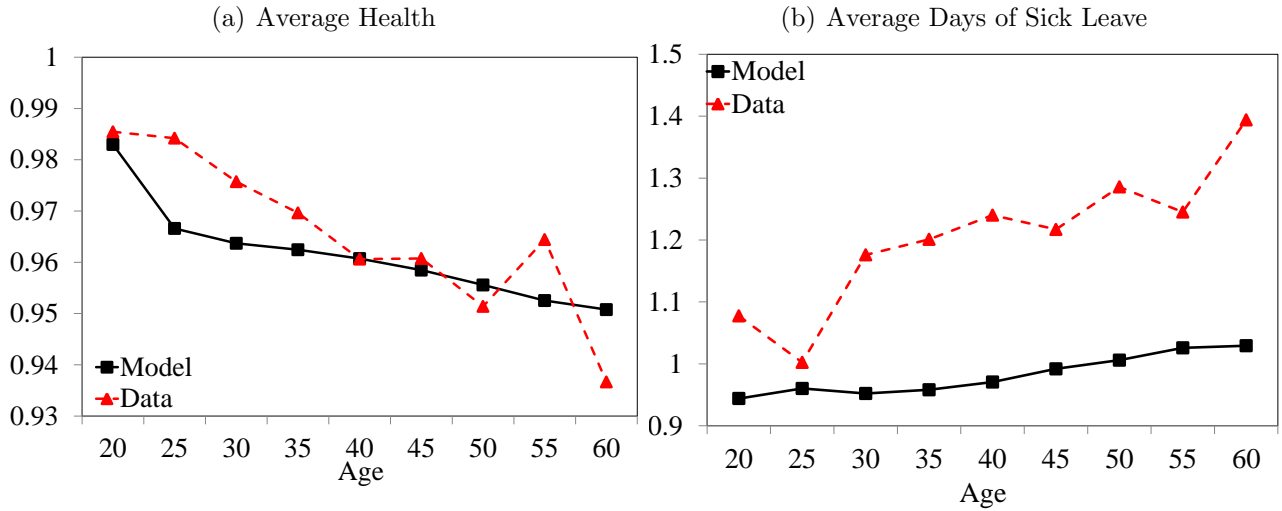
shows the ratio of the two bottom to the two top income quintiles for average health. At the beginning of working life at age 20, average health is almost identical between the bottom and the top.³⁹ Over the life course, the simulated health ratio decreases (the health gap widens). At the end of working life at age 60, the model predicts a clear income gradient in health, as observed in the data.

The right panel of Figure 11 shows the ratio for the average days of sick leave of the bottom to top income groups. The simulated model is able to reproduce the increase in the ratio, i.e., the low-income workers used more days of sick leave compared to their high-income peers as they advanced in age. However, the match of the level of this ratio is poor, which, as mentioned, is due to the fact that the very top income quintile has a low utilization that cannot be explained only by their overall good health.

Business Cycle The model is able to represent the cyclical behavior of the average claims of sick leave in Germany. Figure 12 shows the simulated average days of sick leave for all workers. The unemployment is set according to the German unemployment rate in the time span from 1994 to 2012. When the unemployment rate is high, the average days of sick leave are reduced

³⁹The prediction of the model at age 20 is matched to the data by construction; I set the initial values of the model according to the data.

Figure 11: Inequalities in Days of Sick Leave and Health over Life Cycle



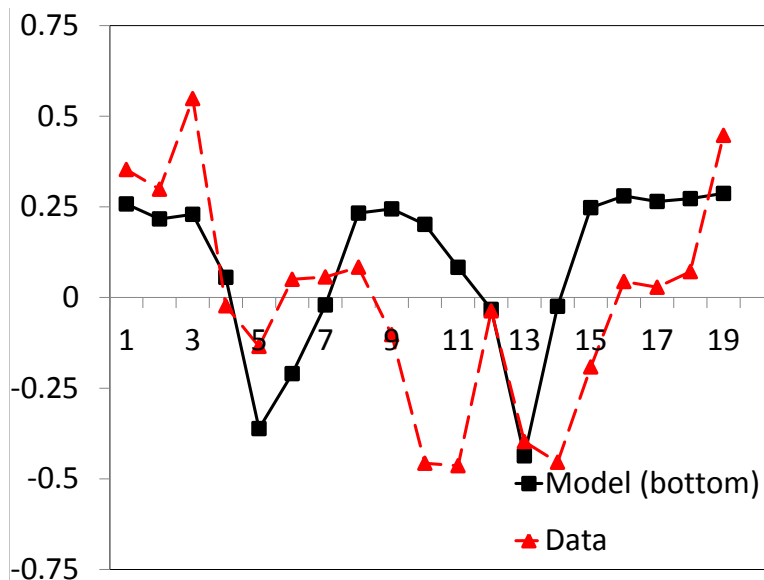
Notes: The dashed red lines show moments taken from the SOEP data. The solid black lines show averages from simulating the structural model using the parameters explained in Section 4.1.

and vice versa. The magnitude is, however, less distinctive than that observed in the data. The model is particularly able to reproduce the drop in days of sick leave in times of high unemployment. The strong increases in the days of sick leave in times of low unemployment are not well matched.

4.3 Effects of Indirect Costs of Sick Leave

To illustrate the quantitative dimension of the indirect costs of sick leave, I simulate a counterfactual model in which the layoff probability, Φ , is independent of workers' past days of sick leave. Table 5 shows the central sick leave moments for the population average and bottom income quintile workers. The average number of days of sick leave for all workers is increased by 4% compared to the benchmark economy. This difference in work absence is primarily driven by low skilled workers. The bottom income group would increase their average number of days of sick leave by more than 11%. A similar pattern is shown for the extensive margin of sick leave, where the bottom income is 10% higher than the benchmark case. The drop in the intensive margin follows from the fact that more people stay at home and reduce the conditional average of sick leave.

Figure 12: Cyclicity of Average Days of Sick Leave for German Business Cycle



Notes: The dashed red lines show moments taken from the SOEP data. The solid black lines show averages from simulating the structural model using the parameters explained in Section 4.1.

Table 5: Quantitative Effect of Indirect Costs of Sick Leave

	Benchmark Case	Sick Leave Independent Layoff	Difference
Average days of sick leave	5.64	5.84	+4%
Average days of sick leave (Bottom)	5.43	6.02	+11%
Extensive Margin	0.56	0.59	+3%
Extensive Margin (Bottom)	0.50	0.60	+10%
Intensive Margin	10.06	9.90	-2%
Intensive Margin (Bottom)	10.69	10.06	-6%

Notes: Results are shown for the benchmark economy and a model without a sick-leave-dependent layoff probability.

5 Policy Evaluation

The determination of policy-invariant structural parameters allows for the introduction and evaluation of different counterfactual policies that affect the financial constraints of a worker's decision-making problem. These policies include variations in the paid sick leave coverage and the benefit structure of unemployed workers.

5.1 Paid Sick Leave Coverage

Having identified the importance of sick leave for households, the model allows for the analysis of the effect of introducing direct costs of sick leave, i.e., a reduction in the replacement rate of paid sick leave. Table 6 shows the average days of sick leave of the whole work force and of workers in the bottom income quintile for different sick leave replacement rates. Introducing the additional costs of being absent from work naturally leads to a decrease in the average days of sick leave. A reduction from 100% (benchmark economy) to only 80% of the current income reduces the days an average worker is absent from work by more than one day or 19% relative to the benchmark economy. These numbers are broadly in line with those of Ziebarth and Karlsson (2010), who estimated the reduction of sick leave for such a policy change as 12% using a natural experiment. Reducing the replacement rate further to zero (as in the US) would lower the average days of sick leave to 3.29 days or only 58% of the days of sick leave of the benchmark economy.

Although both direct and indirect costs of sick leave reduce the number of days missed at work, it is important to note that these reductions are borne by different groups of the income distribution. Although indirect costs primarily affect the bottom income quintile (see results above), Table 6 shows that direct costs of sick leave also have a strong effect on the higher income groups. The reduction of paid sick leave to 80% reduces the days of sick leave of the average worker by 20%. Bottom income quintile workers reduce their days of sick leave by only 17%. The underlying reason for this finding is that the sick leave replacement rate is a fraction of income, and therefore, high-income agents lose more income in absolute terms than their low-income peers.

Table 6: Variation of Income Replacement Rate of Paid Sick Leave

Policy Change	All Workers		Bottom Income Quintile	
	Days of Sick Leave	Difference to Benchmark	Days of Sick Leave	Difference to Benchmark
$b^S=100\%$ (benchmark)	5.64	–	5.43	–
$b^S=99\%$	5.61	-0.5%	5.38	-0.9%
$b^S=90\%$	5.27	-6.5%	4.73	-12.9%
$b^S=80\%$	4.52	-19.8%	4.51	-16.9%
$b^S=50\%$	4.07	-27.8%	3.89	-28.3%
$b^S=0\%$	3.29	-41.7%	3.08	-43.2%

Notes: Simulations of structural model using the parameters explained in Section 4.1, varying the income replacement rate of sick leave.

5.2 Unemployment Benefits

Corresponding to the replacement rate b^S for direct costs of sick leave, unemployment benefits are a main determinant of the indirect costs. Altering the financial situation for workers who are laid off changes the incentives to go to work sick; e.g., full unemployment insurance would completely eliminate the indirect costs of sick leave. Table 7 shows the average days of sick leave of all workers and of workers in the bottom income quintile for various changes in the unemployment benefit structure, i.e., the unemployment insurance net replacement rate b^U , the level of means tested welfare c^W , and the amount of protected assets a household is allowed to save before losing welfare eligibility a^W .

A 10% increase in the unemployment insurance net replacement rate leads to a 2% increase in the days of sick leave for all workers. This increase is strongest among low-income workers, whose days of sick leave would increase by more than 5%. Similarly, a 10% reduction in the unemployment benefit level would reduce sick leave by 1%. Important to note here is that due to the unemployment benefit structure in Germany, when most low-income workers become unemployed, they will fall directly into means-tested welfare and are therefore not affected by this reduction.

Reducing means-tested welfare has, on average, a weaker effect on the population average but a much stronger effect on low-income workers. Reducing the level of welfare by 10% reduces the days of sick leave by almost 3%, whereas increasing the level of welfare by 10% increases

the days missed at work by almost 3%. Taken to the extreme, setting both the unemployment benefits and the welfare to zero reduces by more than 1.2 days the days missed at work for all workers and by more than 1.7 days for poor workers in absolute terms; in relative terms, these reductions are 22% for all workers and 32% for poor workers.

In addition to the level of welfare benefits, it is also important to look at the amount of assets that unemployed people are allowed to save without reducing their welfare. These protected assets enable workers to smooth their consumption over an unemployment period and therefore ease the negative effect of unemployment. Setting a^W to zero has the same effect as setting c^W to 450€. Both reduce the days of sick leave by 5%.

Table 7: Reduction of Income Replacement Rate of Paid Sick Leave

Policy Change	All Workers		Bottom Income Quintile	
	Days of Sick Leave	Difference to Benchmark	Days of Sick Leave	Difference to Benchmark
$b^U=70\%$	5.75	+2.0%	5.71	+5.1%
$b^U=60\%$ (benchmark)	5.64	–	5.43	–
$b^U=50\%$	5.58	-1.1%	5.38	-0.9%
$c^W=770\text{€}$	5.68	+0.7%	5.58	+2.8%
$c^W=700\text{€}$ (benchmark)	5.64	–	5.43	–
$c^W=630\text{€}$	5.61	-0.7%	5.29	-2.6%
$c^W=0\text{€}$	5.56	-1.4%	5.10	-9.4%
$b^U=0\% \& c^W=0\text{€}$	4.41	-21.8%	3.70	-31.9%
$a^W=5000\text{€}$	5.68	+0.7%	5.61	+3.3%
$a^W=1200\text{€}$ (benchmark)	5.64	–	5.43	–
$a^W=0\text{€}$	5.58	-1.1%	5.16	-5.0%

Notes: Simulations of structural model using the parameters explained in Section 4.1, varying the income replacement rate of unemployment benefits and changing the means-tested welfare consumption floor and the asset exemption.

6 Conclusion

In this paper, I have studied disparities in the utilization of sick leave across income quintiles, over the life course and during business cycles. Using data from the SOEP I have found first

that in times of high unemployment days of sick leave are on average low. Second, low income workers use surprisingly few days of sick leave once taken their general low overall health state into account. Furthermore I have documented that the effect of days of sick leave on future employment is empirically relevant and serve as the economic rationale behind these stylized facts of aggregated days of sick leave. Based on this finding I have developed and estimated a life cycle model including an endogenous health state. I have estimated this model intensively using micro panel data and showed that the is able to replicate data moments on aggregated days of sick leave in many dimensions, e.g., the hump-shaped pattern across income quintiles. I found that costs of sick leave stemming from reductions in future expected earnings mostly affect the lowest income quintile whereas costs stemming from reductions in current income (i.e., reduction paid sick leave coverage) also affect the higher income quintiles. Furthermore my results from counterfactual policy experiments suggest that changing the unemployment benefits (especially means tested welfare) would lead to sizeable changes in the number of sick leave in Germany.

References

- Ales, L., R. Hosseini, and L. E. Jones (2012). Is There "Too Much" Inequality in Health Spending Across Income Groups? NBER Working Papers 17937, National Bureau of Economic Research, Inc.
- Andersen, S. H. (2010). The cost of sickness: On the effect of the duration of sick leave on post-sick leave earnings. *Social Science & Medicine* 70(10), 1581–1589.
- Arai, M. and P. S. Thoursie (2005). Incentives and selection in cyclical absenteeism. *Labour Economics* 12(2), 269–280.
- Askildsen, J. E., E. Bratberg, and ivind Anti Nilsen (2005). Unemployment, labor force composition and sickness absence: a panel data study. *Health Economics* 14(11), 1087–1101.
- Bergström, G., L. Bodin, J. Hagberg, T. Lindh, G. Aronsson, and M. Josephson (2009). Does sickness presenteeism have an impact on future general health? *International Archives of Occupational and Environmental Health* 82(10), 1179–1190.
- Bundeministerium für Arbeit und Soziales (2014). Sozialbudget 2013. Technical report, Bundesministerium für Arbeit und Soziales.
- Bundesagentur für Arbeit (2014). Arbeitslosigkeit im zeitverlauf. Technical report, Bundesagentur für Arbeit.
- Chadi, A. and L. Goerke (2015, March). Missing at Work Sickness-related Absence and Subsequent Job Mobility. IAAEU Discussion Papers 201504, Institute of Labour Law and Industrial Relations in the European Union (IAAEU).
- Cole, H. L., S. Kim, and D. Krueger (2014). Analyzing the Effects of Insuring Health Risks: On the Trade-off between Short Run Insurance Benefits vs. Long Run Incentive Costs. PIER Working Paper Archive 14-023, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.
- Dale-Olsen, H. (2014). Sickness absence, sick leave pay, and pay schemes. *Labour* 28(1), 40–63.

- Ehrlich, I. and H. Chuma (1990). A Model of the Demand for Longevity and the Value of Life Extension. *Journal of Political Economy* 98(4), 761–82.
- Fernandez-Villaverde, J. and D. Krueger (2007). Consumption over the Life Cycle: Facts from Consumer Expenditure Survey Data. *The Review of Economics and Statistics* 89(3), 552–565.
- Finkelstein, A., E. F. P. Luttmer, and M. J. Notowidigdo (2013). What good is wealth without health? the effect of health on the marginal utility of consumption. *Journal of the European Economic Association* 11, 221–258.
- Gilleskie, D. B. (1998). A dynamic stochastic model of medical care use and work absence. *Econometrica* 66(1), 1–46.
- Grossman, M. (1972). On the concept of health capital and the demand for health. *Journal of Political Economy* 80(2), 223–55.
- Hall, R. E. and C. I. Jones (2007). The Value of Life and the Rise in Health Spending. *The Quarterly Journal of Economics* 122(1), 39–72.
- Halliday, T., H. He, and H. Zhang (2012). Health investment over the life-cycle. Working Papers 201210, University of Hawaii at Manoa, Department of Economics.
- Henrekson, M. and M. Persson (2004). The Effects on Sick Leave of Changes in the Sickness Insurance System. *Journal of Labor Economics* 22(1), 87–114.
- Hesselius, P. (2007). Does sickness absence increase the risk of unemployment? *The Journal of Socio-Economics* 36(2), 288–310.
- Huggett, M., G. Ventura, and A. Yaron (2011). Sources of Lifetime Inequality. *American Economic Review* 101(7), 2923–54.
- Hujer, R., B. Fitzenberger, R. Schnabel, and T. E. MaCurdy (2001). Testing for uniform wage trends in West-Germany: A cohort analysis using quantile regressions for censored data. *Empirical Economics* 26(1), 41–86.
- Johansson, P. and M. Palme (2005). Moral hazard and sickness insurance. *Journal of Public Economics* 89, 1879 – 1890.

- Keane, M. P. and K. I. Wolpin (1997). The Career Decisions of Young Men. *Journal of Political Economy* 105(3), 473–522.
- Kivimäki, M., J. Head, J. E. Ferrie, H. Hemingway, M. Shipley, J. Vahtera, and M. G. Marmot (2005). Working while ill as a risk factor for serious coronary events: The whitehall ii study. *American Journal of Public Health* 95(1), 98102.
- Laun, T. (2013). Optimal Social Insurance with Endogenous Health. SSE/EFI Working Paper Series in Economics and Finance 742, Stockholm School of Economics.
- Leigh, J. P. (1985). The effects of unemployment and the business cycle on absenteeism. *Journal of Economics and Business* 37(2), 159–170.
- Markussen, S. (2012). The individual cost of sick leave. *Journal of Population Economics* 25(4), 1287–1306.
- Nuebling, M., H. H. Andersen, A. Muehlbacher, J. Schupp, and G. G. Wagner (2007). Computation of Standard Values for Physical and Mental Health Scale Scores Using the SOEP Version of SF12v2. *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift fr Wirtschafts- und Sozialwissenschaften* 127(1), 171–182.
- Ozkan, S. (2014). Preventive vs. Curative Medicine: A Macroeconomic Analysis of Health Care over the Life Cycle. Technical report, University of Toronto.
- Pfeifer, C. (2013). Cyclical absenteeism among private sector, public sector and self-employed workers. *Health Economics* 22(3), 366–370.
- Puhani, P. A. and K. Sonderhof (2010). The effects of a sick pay reform on absence and on health-related outcomes. *Journal of Health Economics* 29(2), 285–302.
- Scheil-Adlung, X. and L. Sandner (2010). The case for paid sick leave. World Health Report Background Paper 9, World Health Organization.
- Siegel, J. (2002). *Stocks for the Long Run*. McGraw-Hill Education.
- Smith, J. P. (1999). Healthy bodies and thick wallets: The dual relation between health and economic status. *The Journal of Economic Perspectives* 13(2), pp. 145–166.

- Storesletten, K., C. I. Telmer, and A. Yaron (2004). Consumption and risk sharing over the life cycle. *Journal of Monetary Economics* 51(3), 609–633.
- Tauchen, G. (1986). Finite state markov-chain approximations to univariate and vector autoregressions. *Economics Letters* 20(2), 177–181.
- Wagner, G. G., J. R. Frick, and J. Schupp (2007). The german socio-economic panel study (soep) scope, evolution and enhancements. *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift fr Wirtschafts- und Sozialwissenschaften* 127(1), 139–169.
- Ziebarth, N. R. (2013). Long-term absenteeism and moral hazard - Evidence from a natural experiment. *Labour Economics* 24(C), 277–292.
- Ziebarth, N. R. and M. Karlsson (2010). A natural experiment on sick pay cuts, sickness absence, and labor costs. *Journal of Public Economics* 94(11-12), 1108–1122.
- Ziebarth, N. R. and M. Karlsson (2014). The effects of expanding the generosity of the statutory sickness insurance system. *Journal of Applied Econometrics* 29(2), 208 – 230.

Appendix A German Sick Leave Policy

Compulsory sick pay with 100% wage replacement was established 1930 for white collar employees and 1969 for blue collar workers. The current regulation of sick leave (*Entgeltfortzahlung im Krankheitsfall*) in Germany is determined in the *Entgeltfortzahlungsgesetz*. According to the law eligible for paid sick leave are those employees (also including part time and temporary workers) that fulfill following conditions:

- The employment has to be in place for four weeks.
- The worker has to be incapable of working.
- The incapability has to be a consequence of an illness.
- The illness is not a result of a gross negligence.

Sick pay has to be provided by the employer from the first day (no grace period) up to six weeks. After six weeks, sick pay is provided by the health insurance with a reduced wage replacement rate of 80%. The claim of full wage replacement, however, renews if the worker contracts a different illness, or more than 6 months has elapsed in which the worker was sick with the same illness. Workers receive the average earnings that they would have earned if they had not been sick. These earnings include the fixed salary potential commissions. Furthermore, if workers become sick while they are on vacation, holiday entitlement is not reduced. Workers absent from work have to inform the employer immediately about their incapability to work. On the fourth day of a sick spell, workers have to send a sick certificate issued by a health practitioner.

Monitoring the worker is highly restricted. The German Federal Labor Court decided that the observation of employees by their employer is illegal without concrete evidence supporting the suspicion of fraud. (Court decision 19. February 2015 - 8 AZR 1007/13). The employer can only request that the employee be reexamined by the practitioner of the Medical Service of the Health Funds (Medizinischer Dienst der Krankenversicherung).

Between October 1996 and December 1998, there was a temporary change in the law. The main changes were a reduction of wage replacement from 100% to 80%. However, this

reduction applied to only a fraction of the German workforce, as collective labor agreements between unions and firms mostly kept 100% wage replacement. Empirical research on this law discontinuity is conducted by Ziebarth and Karlsson (2010) and Ziebarth (2013).

Appendix B Sample Selection

Table 8 shows the descriptive statistics before and after sample selection. I use in these statistics weighted samples. Weights are provided by the SOEP to match the German micro-census. The final sample is younger due to a focus on the working age population. The higher percentage of men in the sample can be explained by their higher participation rate in the labor force. The average number of reported days of sick leave is lower as the upper tail of the sick day distribution is cut off.

Table 8: Descriptive Statistics for Sample Selection

	Whole Sample	Benchmark Sample
Male	46%	63%
Age	49	41
Years of Education	11.7	12.1
Health	2.32	2.60
Income	2,359€	2,814€
Unemployed	6.80%	5.40
Sick Leave	9.73	5.26
Observations	393,245	100,526

Notes: Descriptive statistics before and after sample selection. A benchmark sample used in the cross-sectional and panel analysis.

Appendix C Robustness Check of Empirical Part

C.1 Different Measures of Sick Leave

Workers' number of sick leave days has a skewed distribution. Table 9 provides results for the correlation of the unemployment rate and different measures of sick leave days. First, it shows the result for the median worker. The second column shows the correlation with the extensive margin, i.e., whether the respondent has missed a day or more. In the last three columns,

different cut-off levels for the maximum days of sick leave are used. All results are negative and in the same range as the benchmark result. The pro-cyclical pattern is extremely robust.

Table 9: Days of Sick Leave and Unemployment - Different Sick Leave Measures

Correlation	Median	Ext. Marg.	Max 120	Max 60	Max 40
Unemployment Rate	-0.5605	-0.4960	-0.7033	-0.6951	-0.6347

Notes: Times series correlation of different measures of days of sick leave and unemployment rate. First, the days of sick leave of the median respondent; second, the cyclical behavior of the extensive margin. The last three columns represent the correlation of the mean with different cut-off levels for the maximum number of sick days.

C.2 Composition Effects

A potential different explanation of the cyclicity of days of sick leave could arise if sectors (e.g., construction sector) with high usual high number of days of sick leave are more prone to business cycles than the rest of the economy. To control that the general effect is not driven by this reason I check for different sector whether their exclusion alter the general finding. Table 10 shows the exclusion of the construction sector does not alter the benchmark result. The correlation coefficient is only slightly reduced to -.7188.

I also check whether this cyclical behavior is different for different occupation type. SOEP provides the ISCO88 classification and using the white/blue collar distinction as in the European working conditions surveys. Table 10 shows that for both subgroups the pro-cyclicity of days of sick leave holds.

Table 10: Days of Sick Leave and Unemployment - Different Sectors and Occupations

Correlation	No Construction	Blue Collar	White Collar	Never Unemployed
Unemployment Rate	-0.7032	-0.6289	-0.6272	-0.6576

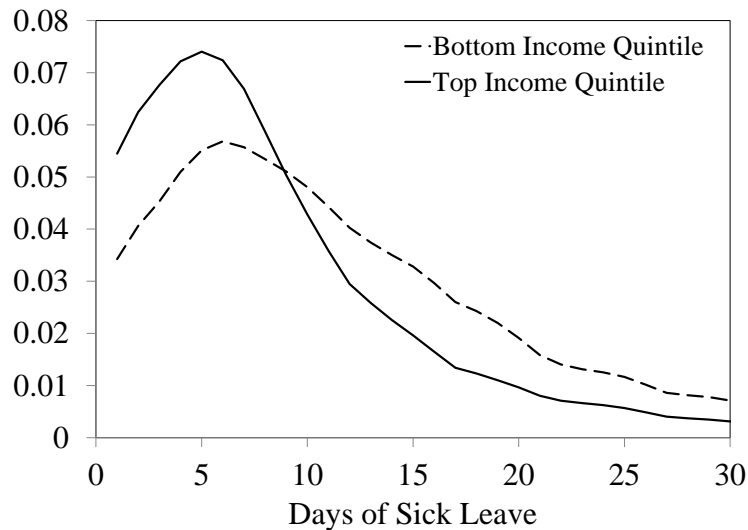
Notes: Time-series correlation of average days of sick leave and unemployment rate for selected subgroups.

C.3 Density Function of Days of Sick Leave

Figure 13 shows the density functions for the top and the bottom income quintiles conditional on missing any day in a year. The frequency of a low number of sick leave days is higher for the

top income quintile, whereas the frequency of a high number of days of sick leave is higher for the bottom income quintile. Combined with the differences in the extensive margin, this shows that workers in the bottom income quintile have either no or many days of sick leave, whereas workers in the top income quintile have few days of sick leave but at a higher frequency.

Figure 13: Density Function of Days of Sick Leave for Bottom and Top Income Quintiles



Notes: Density function of days of sick leave for bottom and top income quintiles. Graphs exclude the probability of having no days of sick leave.

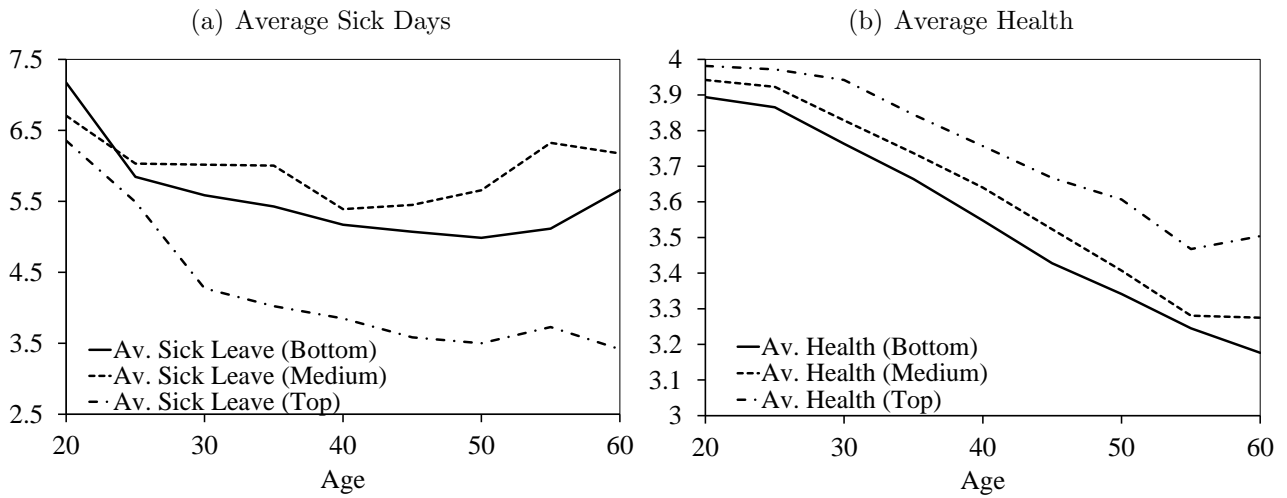
C.4 Age Profiles of Days of Sick Leave and Health

The left panel of Figure 14 confirms that the observed hump-shaped income profile also holds within all but one age group (20-25). It is noteworthy that there is no significant increase in days of sick leave with age, which is driven in part by the exclusion of a very high number of sick leave days (>30). The right panel shows that the income gradient in health increases with age.

C.5 Controlling for Gender in Sick Day Profiles over Income Quintiles

Figure 15 shows the average annual claims of days of sick leave and average self-reported health separated by gender. Both groups show a hump-shaped profile in average days of sick leave

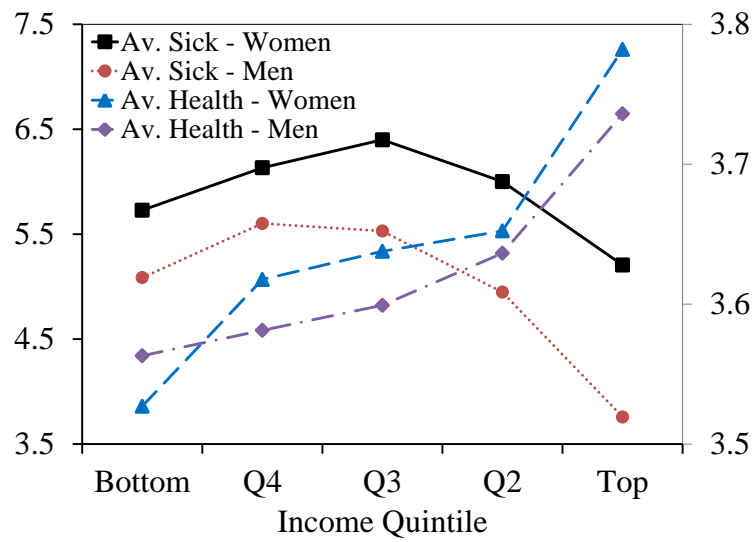
Figure 14: Days of sick leave and Health over Life Cycle by Income Quintiles



Notes: Left Panel: Average self-reported health of bottom (solid), medium (dashed) and top (dotted) income quintiles over the life cycle. Right Panel: Average number of sick leave days of the bottom (solid), medium (dashed) and top (dotted) income quintiles over the life cycle. Age bins: 18-22, 23-27, 28-32, 33-37, 38-42, 43-47, 48-52, 53-57, 58-62.

across income quintiles. On average women miss more days at work due to sickness than men. Health increase monotonically over income for both groups with better average health for woman than for men.

Figure 15: Average Days of Sick Leave and Health over Income Quintiles - Gender



Notes: Dashed and dotted-dashed lines (right axis): Average self-reported health separated by gender. Health is reported on an ordinal five-point scale, where 1 denotes "bad" health, and 5 denotes "very good" health. Solid and dotted lines (left axis): Average days of sick leave of workers separated by income quintile and gender.

Appendix D Parameters of Structural Model

Table 11: Fixed Parameters

Parameter	Description	Value
<i>Economy</i>		
J^T	Life Span	60
J^R	Retirement Age	45
R	Interest Rate	1.04
w	Wage Rate	1
<i>Preferences</i>		
β	Time Discount Factor	0.9659
σ	Inter-temporal Elasticity of Substitution	2
ψ_0	Health Weight (level)	0.011
ψ_1	Health Weight (marginal)	0.19
<i>Policy</i>		
ρ^U	Unemployment Benefit (ALGI)	60%
c^W	Consumption Floor Welfare (ALGII)	$0.25 \triangleq 700\text{€}$
a^W	Asset Limit for Welfare (ALGII)	$0.44 \triangleq 1225\text{€}$
ρ^S	Sick Leave Replacement Rate	100%
ρ^R	Retirement Benefit	50%
τ	Tax Rate	33%

Notes: Parameters used in the structural model and taken from the literature or chosen to match the German labor market.

Table 12: Labor Productivity of Income Quintile over Life Cycle - Γ

	K_1	K_2	K_3	K_4	K_5
J_{20}	0.224826037	0.328405423	0.456578007	0.587303446	0.732877732
J_{21}	0.250005566	0.409269009	0.560519145	0.685354277	0.850657669
J_{22}	0.311148441	0.509774966	0.622307323	0.724721362	0.884406497
J_{23}	0.350360594	0.552927485	0.656849717	0.754174695	0.924644522
J_{24}	0.425638928	0.598862981	0.700761760	0.809373763	0.979296359
J_{25}	0.453304887	0.633908213	0.749711757	0.848907970	1.043437128
J_{26}	0.486555339	0.666411652	0.784269193	0.909526063	1.105854970
J_{27}	0.505469995	0.689869703	0.819220306	0.957395861	1.196744849
J_{28}	0.526567728	0.729833055	0.848928302	1.009396469	1.256001953
J_{29}	0.534257324	0.749711757	0.892406923	1.043371438	1.333614431
J_{30}	0.552043113	0.775403025	0.923040554	1.094749559	1.398805958
J_{31}	0.588814584	0.811998956	0.964703602	1.141444947	1.461369255
J_{32}	0.581567748	0.813786468	0.967081217	1.143405056	1.507091996
J_{33}	0.590600670	0.827264526	0.997438473	1.182487624	1.598022908
J_{34}	0.610923651	0.844675221	1.0	1.195602273	1.638440613
J_{35}	0.605390496	0.850657669	1.019549932	1.236543287	1.696821211
J_{36}	0.587309058	0.841163678	1.035278366	1.258038975	1.730506925
J_{37}	0.595460368	0.850998814	1.034804554	1.262927700	1.803819125
J_{38}	0.588814584	0.851471889	1.046094062	1.291190800	1.845452180
J_{39}	0.598372424	0.854338956	1.050974415	1.293505670	1.844733642
J_{40}	0.596219571	0.856083687	1.053135456	1.324056682	1.900608466
J_{41}	0.596219571	0.850717011	1.053102151	1.303667321	1.891957676
J_{42}	0.598372424	0.854338956	1.051454666	1.293505670	1.899733341
J_{43}	0.587912961	0.852549696	1.051454666	1.299490326	1.910960195
J_{44}	0.597444858	0.842776110	1.030489463	1.275986504	1.913979709
J_{45}	0.601222195	0.844669241	1.028788249	1.275986504	1.913979709
J_{46}	0.595460368	0.845894251	1.033453225	1.261745563	1.913979709
J_{47}	0.588814584	0.842776110	1.027148308	1.265043937	1.910960195
J_{48}	0.590662450	0.850657669	1.035278366	1.275986504	1.921742678
J_{49}	0.598372424	0.862337174	1.043399683	1.262927700	1.934141825
J_{50}	0.591030413	0.851508322	1.035278366	1.257473805	1.974699741
J_{51}	0.592397659	0.842936838	1.035278366	1.275231717	1.986756456
J_{52}	0.575729560	0.854603003	1.034199823	1.284230276	1.958125531
J_{53}	0.574158805	0.840822441	1.031918443	1.264579325	1.971260512
J_{54}	0.571549022	0.837721413	1.018490066	1.275348100	1.955603380
J_{55}	0.565836095	0.838220617	1.033453225	1.278164842	1.987566444
J_{56}	0.560519145	0.834560214	1.020789203	1.263601985	2.009223412
J_{57}	0.570354465	0.818330276	0.993699316	1.259025884	1.919270208
J_{58}	0.566708552	0.827973127	0.995205761	1.232760613	1.976734648
J_{59}	0.562409792	0.835161909	0.991687686	1.213715589	1.943266795
J_{60}	0.573288049	0.824682942	0.992062779	1.222188079	2.009503467
J_{61}	0.567814006	0.857883160	1.031918443	1.254869129	2.060114657
J_{62}	0.575595834	0.849653188	1.050129374	1.262336586	2.084111243
J_{63}	0.609652318	0.891687833	1.095787160	1.333614431	2.155842844
J_{64}	0.611094040	0.807900162	0.993699316	1.217998479	2.135917358

Notes: Parameters directly estimated from the SOEP.

Table 13: Incident Probability of Sickness Conditional on Age & Health - Ω

J	S_m	H_1	H_2	H_3	H_4	H_5
20-30	0	0.2736842	0.2898424	0.3794949	0.4438789	0.5247102
	1-2	0.0210526	0.0078809	0.0160531	0.02507	0.0359976
	3-4	0.0105263	0.0402802	0.0410959	0.0558666	0.0457596
	5	0.0	0.0420315	0.0524401	0.0567574	0.0506406
	6-9	0.0210526	0.0262697	0.031036	0.0349962	0.0280659
	10	0.0842105	0.0954466	0.0997432	0.092899	0.0841977
	11-15	0.0	0.0131349	0.0205479	0.0187071	0.0183038
	16-20	0.0315789	0.0288967	0.0265411	0.0273607	0.0207444
	21-30	0.0105263	0.0297723	0.0278253	0.0223976	0.0195241
30-40	0	0.2736842	0.2898424	0.3794949	0.4438789	0.5247102
	1-2	0.0210526	0.0078809	0.0160531	0.02507	0.0359976
	3-4	0.0105263	0.0402802	0.0410959	0.0558666	0.0457596
	5	0.0	0.0420315	0.0524401	0.0567574	0.0506406
	6-9	0.0210526	0.0262697	0.031036	0.0349962	0.0280659
	10	0.0842105	0.0954466	0.0997432	0.092899	0.0841977
	11-15	0.0	0.0131349	0.0205479	0.0187071	0.0183038
	16-20	0.0315789	0.0288967	0.0265411	0.0273607	0.0207444
	21-30	0.0105263	0.0297723	0.0278253	0.0223976	0.0195241
40-50	0	0.2736842	0.2898424	0.3794949	0.4438789	0.5247102
	1-2	0.0210526	0.0078809	0.0160531	0.02507	0.0359976
	3-4	0.0105263	0.0402802	0.0410959	0.0558666	0.0457596
	5	0.0	0.0420315	0.0524401	0.0567574	0.0506406
	6-9	0.0210526	0.0262697	0.031036	0.0349962	0.0280659
	10	0.0842105	0.0954466	0.0997432	0.092899	0.0841977
	11-15	0.0	0.0131349	0.0205479	0.0187071	0.0183038
	16-20	0.0315789	0.0288967	0.0265411	0.0273607	0.0207444
	21-30	0.0105263	0.0297723	0.0278253	0.0223976	0.0195241
50-65	0	0.2736842	0.2898424	0.3794949	0.4438789	0.5247102
	1-2	0.0210526	0.0078809	0.0160531	0.02507	0.0359976
	3-4	0.0105263	0.0402802	0.0410959	0.0558666	0.0457596
	5	0.0	0.0420315	0.0524401	0.0567574	0.0506406
	6-9	0.0210526	0.0262697	0.031036	0.0349962	0.0280659
	10	0.0842105	0.0954466	0.0997432	0.092899	0.0841977
	11-15	0.0	0.0131349	0.0205479	0.0187071	0.0183038
	16-20	0.0315789	0.0288967	0.0265411	0.0273607	0.0207444
	21-30	0.0105263	0.0297723	0.0278253	0.0223976	0.0195241

Notes: Parameters directly estimated from the SOEP.

Table 14: Initial Distribution of Health States Conditional on Income Quintile - H_0

	H_1	H_2	H_3	H_4	H_5
K_1	0.53%	5.60%	20.68%	56.28%	17.23%
K_2	0.33%	4.54%	19.50%	57.13%	18.83%
K_3	0.37%	4.88%	18.88%	57.69%	18.55%
K_4	0.45%	4.36%	17.97%	58.66%	19.01%
K_5	0.25%	3.37%	17.21%	57.61%	21.81%

Notes: Parameters estimated directly from the SOEP.

Table 15: Initial Distribution of Unemployment Conditional on Income Quintile - \mathcal{I}_0

	K_1	K_2	K_3	K_4	K_5
$\mathcal{I}_0 = 0$	8%	5%	2.5%	1.5%	0.4%
$\mathcal{I}_0 = 1$	92%	95%	97.5%	98.5%	99.6%

Notes: Parameters estimated directly from the SOEP.

Table 16: Calibrated Parameters - Sickness Aggravation

Param.	Statistics	Data	Model
$\zeta=0.1135$	Probability of $l \in \{0\}$ for K_1	0.4874	0.4885
	Probability of $l \in \{1-2\}$ for K_1	0.0449	0.0315
	Probability of $l \in \{3-4\}$ for K_1	0.0632	0.0613
	Probability of $l \in \{5\}$ for K_1	0.0751	0.0845
	Probability of $l \in \{6-9\}$ for K_1	0.0633	0.0721
	Probability of $l \in \{10\}$ for K_1	0.0760	0.0824
$\kappa=4.021$	Probability of $l \in \{11-15\}$ for K_1	0.0920	0.0883
	Probability of $l \in \{16-20\}$ for K_1	0.0462	0.0436
	Probability of $l \in \{21-30\}$ for K_1	0.0517	0.0475

Notes: Calibrated parameter values and match of model to data.