

Unemployment, Sick Leave and Health*

Matthias Schön[†]

Draft - December 15, 2014

Abstract

This paper studies the relationship between sick leave, income and unemployment. In particular, it investigates this relationship under the generous German sick leave regulation of 100% wage replacement, i.e., in an environment where workers do not bear any direct costs from missing work due to sickness. Using information from the German Socioeconomic Panel (GSOEP) I identify three stylized facts of sick leave in Germany. First, sick days show a strong pro-cyclical pattern. Second, average use of sick days is hump-shaped over income quintiles. Third, the number of sick days is a strong predictor of becoming unemployed. Using this micro-evidence I develop a structural model that rationalizes these facts. I argue that in 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 to do counterfactual policy analysis.

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

Keywords: 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, the University of Cologne, the University of Frankfurt, Hanming Fang, Max Groneck, David Jaeger, and Alexander Ludwig for 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

Loss in labor income due to sick leave is one major economic risk for the working age population associated with health. These costs of sick leave could be either direct or indirect. Direct costs arise from a reduction of time worked which would be paid otherwise. The extent to which this cost is insured varies greatly among industrialized countries, cf. Scheil-Adlung and Sandner (2010). An extreme case is the US with no statutory paid sick leave. On the opposite side, Germany has one of the most generous sick payment systems. Every (full-time, part-time or temporary) employee is eligible for six weeks of 100% of wage replacement in case of sickness absence.¹ This generosity comes with a price. The expenditures of paid sick leave which are borne by employers amount to 35 billion € in 2010 or 1.4% of GDP according to the German Federal Ministry of Labour and Social Affairs (2011). Indirect costs stem from reductions in future expected earnings. Layoff or promotion decisions of employers depend on the past sick days of workers.

The first objective of this paper is to identify how indirect costs effect utilization of sick leave. The strategy is to empirically asses how sick leave is influenced by the economic situation of individuals under the German system of full wage replacement (ruling out direct costs). To this end, I employ data from the German Socioeconomic Panel (GSOEP) and document remarkable patterns of aggregated sick leave claims in Germany. First, average claims of sick leave show a strong pro-cyclical pattern, i.e., workers are on average less absent in times of high unemployment. Second, average sick days in Germany display a marked hump-shaped pattern over income quintiles. Workers in the medium income quintile have on average more number of sick days than worker in the bottom income quintile. This is noteworthy as average health is monotonically increasing in income. Assuming sick days are only driven by health would lead to the opposite pattern. Third, the variance of sick days differs greatly between income quintiles. Employees in the bottom quintile have the highest probability to not miss any day a year. They also have the highest probability to miss more than two weeks. Top income employees miss small number of sick days but on a high frequency.

In this paper, I rationalize these patterns by one further micro-level observation. Exploiting the panel structure of the GSOEP I estimate a logit model with fixed effects. The results show that sick days are one key predictor of future unemployment. Taking five additional sick days increase the risk of becoming unemployed by 10%. I argue that sick days are an endogenous choice. Workers who become sick face a trade-off to either stay at home and recover or going sick to work (presenteeism). Staying at home restores utility-enhancing health but at the same time increase the risk to be fired. Going to work sick preserves expected future earnings but a perpetual neglect of recuperation diminish long-run health prospects.

¹For more information on Germany regulation on sick leave see Appendix 7.1

In times of a high unemployment rate, workers face both higher overall firing rates as well as lower reemployment probabilities. Resulting higher marginal costs of unemployment shift the trade-off towards 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. Therefore they are particular compelled to go to work when sick. Over time this leads to differences in health between bottom and top income quintile. A fact that is well documented in the literature.² I further argue that the risk of a severe sickness depends positively on the overall health condition. It is more likely that a cold becomes the flu within an already poor immune system. When experience a severe sickness the agent has to take a large number of sick days. In the end, optimal sick day utilization differs between income groups. Rich take constantly small number of sick days to conserve health. Low skilled worker reduce sick days to keep their job. Resulting low health increases the however small probability to be hit by a severe shock and resulting high number of sick days.

The second objective of this paper is to quantify the distributional effects of the indirect effect. For this purpose, I develop a heterogeneous agent model with endogenous health and incomplete credit and insurance markets. The model also incorporates employment risk, i.e., the number of sick days increases the risk to be laid off. The model can account for the stylized facts and reflects the major trade-offs between consumption and health.

Additionally, I implement central characteristics of the German health care and worker protection system into my model. Individuals are offered universal health insurance that covers medical expenditures of agents (up to a certain amount) and are entitled to continued wage payments for up to six weeks per sickness episode. The government imposes the progressive German income tax schedule on agents. The collected revenues are used to finance (i) a PAYGO retirement system, (ii) expenditures due to health insurance and sick leave payments (iii) other government expenditures. The residual budget surplus or deficit is distributed in a lump-sum fashion to agents.

To empirically implement my quantitative analysis I first estimate and calibrate the model using GSOEP data to match key statistics on sick leave, health status and unemployment. In my calibration strategy I set some of the parameter values outside of the model (e.g., interest rate, preferences, policy parameter etc.). The income process and the layoff probability are estimated directly from the data. For remaining parameters (e.g., health transition probabilities, etc.) I use my model to choose their values. The model is stylized enough to allow me to identify its key parameters by the available data. The estimated model is able to successfully explain the targeted features of the data in the estimation (e.g., cyclicity of claims of sick leave, different utilization of sick days

²For example, in excellent health in 1984 have 74 percent more wealth than respondents in fair or poor health do (Smith 1999).

low and high skilled etc.) as well as other (non-targeted) salient dimensions.

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 including a wage dependent paid sick leave (Germany) with no mandatory paid sick leave (US), a wage independent amount of paid sick leave (UK) and a system with an unpaid grace period.

Related Literature The empirical part of this paper is related to two strands of the literature. First, this paper confirms for Germany the well documented negative relationship between a country's economic situation and its average sick rates, starting with Leigh (1985). This literature shows that there are two potential mechanisms regarding the negative relationship between unemployment rate and average sick days. One is an incentive effect, i.e., unemployment affect the propensity to take sick days. If an employee's higher sick rate increases the risk of job loss, a higher unemployment rate reduces the propensity to take sick days. A second alternative mechanism is related to the absence behavior of the marginal workers entering or leaving the working population in various states of the business cycle. When employers can choose whom to layoff, the most absence-prone workers are more likely to be laid off in an economic downturn. All agree that this pattern is mainly driven by the reduction of sick time of workers due to fear off job loss in recessions. Arai and Thoursie (2005) and Askildsen et al. (2005) show that the incentive effect is the dominant force.

Second, I provide additional evidence for this incentive effect on a micro basis, i.e., I estimate the impact of sick days on future layoff risk. Other work in this area is done by Hesselius (2007) and Markussen (2012). They find that sick days are a strong predictor off job loss. Andersen (2010) finds that many sick days not just affect employment risk but also decrease post sick leave earnings.

The structural model I present in this paper is part of a broad and growing macro health literature that incorporates endogenous health into dynamic models. Important related contributions include Grossman (1972), Ehrlich and Chuma (1990), Hall and Jones (2007), Ales et al. (2012), Halliday et al. (2012), Ozkan (2011), Cole et al. (2014).

Only a small literature distinguish in such dynamic models between long run health and the onset of acute illnesses. Gilleskie (1998) predicts the change in physician services use and illness-related absenteeism that arise with improvements in access to health care through more complete health insurance and sick leave coverage in the US.³ The paper however, only focus on the direct costs of work absence and does not take the risk of unemployment into account. It also falls short to provide a link to the endogenous health

³Health-enhancing leisure in the health economics literature goes back to e.g., Grossman (1972), Ruhm (2000), while clinical, experimental, and empirical evidence in support of this idea can be found in the bio-medical science, public health, psycho-biology, and bio-sociology, and empirical health literature.

literature.

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. Then I perform counter-factual policy experiments using the model in Section 5. Finally, I conclude in Section 6.

2 Empirical Facts

The purpose of this section is to motivate the key modeling assumption 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 prove that sick days are an endogenous choice of workers. Then, in section 2.3 I present results based on a panel analysis that underline the importance of the sick day choice for future income of workers.

2.1 Data and Methodology

2.1.1 Description of the Survey

My empirical analysis is based on the German Socio-Economic Panel (GSOEP), a nationally representative longitudinal data set. Starting in 1984 and conducted annually and it comprises 30 waves of data. It oversamples foreigners, immigrants, and East Germans to allow for more precise estimates for population subgroups that may be of particular policy interest.⁴ GSOEP provides detailed information about demographic (sex, age, ...), socioeconomic (educational level, marital status, ...) and economic characteristics. The respondents report their monthly income in the current and the previous year.⁵ The employment history contains the current employment status (full time, part time...), point in time of layoff in previous year, length off unemployment spell and information about the time worked for the same firm. Information about health is asked since 1990. The GSOEP contains information about self-reported health, number of doctor visits and hospital stays.⁶ Further detailed information about the characteristics of the GSOEP is provided in Wagner et al. (2007).

Key variable for the purpose of this paper is the number of working days missed due to sickness. The GSOEP asks the respondents to state if they missed any day due to sickness in the previous year and if so how many. Puhani and Sonderhof (2010) show

⁴I include all sub-samples of GSOEP with the appropriate cross-sectional weights.

⁵Both income variables are deflated with the consumer price index contained in the GSOEP using 2005 as base year.

⁶It also includes a 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 only available every second year since 2002 and is only of limit use for the panel analysis.

that, though self-reported, the GSOEP adequately depicts the true number of days of absence from work. The GSOEP also contains information about the number of spells that last longer than six weeks. However, the survey does not record information about occasion respondents going sick to work.

The only information that I used and is not contained in the GSOEP is the unemployment rate of Germany. I use official data from the federal employment agency, cf. Bundesagentur fuer Arbeit (2014). To investigate the cyclical behavior I additionally construct a dummy variable indicating periods with a high rate of unemployment, "Recession".⁷

2.1.2 Determination of the Sample

For the following empirical results not all of the observations of GSOEP are used. As I am interested on the sick leave use of workers I focus only on the working age population. I drop all observations of respondents younger than 18 and older than 65 (official German retirement age). I restrict the sample to respondents that either report to work in the current or in the previous year or report to be unemployed. This excludes people that are doing their military or social service as well as people that report not to be employed and not looking for work. I also exclude part time worker and respondents that report a monthly income of less than 400 €. ⁸ The probability and the intensity of annual sick leave are biased when respondents only work a fraction of the year.

As time period I use the waves 1994 to 2011, corresponding to information about sick days from 1993 to 2010. Waves 1984, 1990 and 1993 do not contain information about sick days. Wave 1991 and 1992 captures the unique economic situation of German reunification in 1990 and the liberalization of a state-owned socialist economy. I dropped both periods as income distribution and employment situation changed dramatically.⁹ 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 section or panel regressions. Second, the unemployment rate only varied in these year between 7.9% and 8.1%. Hence, not much variation can be used for figuring out cyclicality of sick leave. And 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 from the sample. Self-employed workers do not receive paid sick leave as it is provided by the employer. Civil servants do get paid sick leave but are not eligible for layoffs. Hence, they are not

⁷For the results in Table 7 set an unemployment rate of 11% as cut off level. This gives me equally many periods marked as recessions as not. The following results are robust for other cut off levels.

⁸All monetary variables are deflated with the base year 2005.

⁹Following waves are also affected but the effect is strongly mitigated over time. Especially the classification in income quintiles is disturbed as the income scale was lower in East-Germany. **Still to do:** Do the analysis for West Germany only.

or less affected by the indirect effect of fear of job loss.¹⁰

Sick days have a highly skewed distribution with many observations on the 0-boundary and few observations at the highest value of 365. 95% percent of the observations report less than 40 sick days a year and only one percent report more than 120 days. Hence, many results, e.g., the average number of sick 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 days that last longer than six weeks. Of the remaining sample I cut off the highest one percent, i.e., workers with more than 40 sick days a year.

After sample selection, the sample used for benchmark results consists of 138,569 observations.¹¹ It includes 19 waves and each wave has at least 6,180 observations.¹²

2.1.3 Empirical Approach

In Section 2.2 I run cross section regressions of sick days on various regressors. The regression equation for OLS and Logit are

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

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

where S_i is a countable variable denoted for sick days whereas S_i^{ext} is a dummy variable for either miss any day ($S_i^{ext} = 1$) a year or not ($S_i^{ext} = 0$). W_i is the monthly income of the respondent in the previous year.¹³ H_i is self-reported health, \bar{u}_i the unemployment rate and $X_{i,t}$ is a set of control variables, e.g., sex, age, years of education, year dummies and ε_i is the random error term.

In Section 2.3 I employ a logistic panel regression. I estimate the effect that (accumulated) sick days in the previous period have on the current probability of unemployment. I restrict the sample in this section to people that were employed at least six month in the previous year. The panel structure of GSOEP additionally allows me to use a fixed effects model. The fixed effect will incorporate all unobserved characteristics of the agent.

¹⁰Results 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). Also the income gradient in sick days 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.

¹¹Further details of the sample selection can be found in Appendix 7.2.

¹²The size of the waves increases over time. There were refreshments of the GSOEP in 1998, 2000, 2002 and 2006.

¹³I used other measures of income that are also included in the GSOEP. The qualitative results do not change.

The regression equations are

$$\text{Logit}[Unemployment_{i,t} = U] = \Phi\{\alpha + \beta S_{i,t-1} + C_{i,t-1}\theta + \epsilon_{i,t}\} \quad (3)$$

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

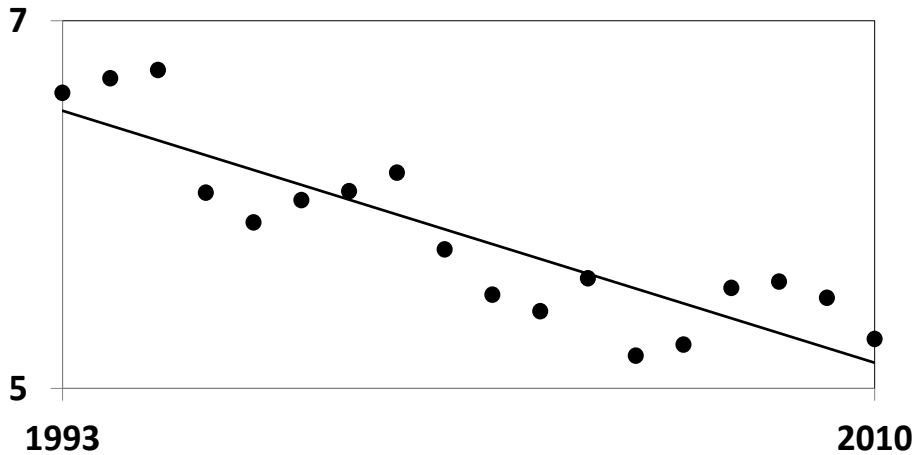
where $C_{i,t}$ is a set of control variable that do vary over time. It contains lagged health, age and lagged log income. For sick days, $S_{i,t-1}$, I use two different definitions. First, sick days reported only in the previous year. Second, accumulated sick days, i.e., the average sick days of the respondent over the last three years. The a_i s represent the individual specific and time-invariant fixed effect component and $\epsilon_{i,t}$ is the random error term.

2.2 Facts on Aggregated Data

2.2.1 Time series

Figure 1 shows for the benchmark sample the average annual number of sick days of workers in the observed time period and a fitted linear trend. Average sick days varies between 5.2 and 6.8 days. An obvious first finding about the number of sick days in Germany is the long term decline. In the last 19 years average claims of sick days have declined by more than 1.5 days or 20% relative to 1993.¹⁴

Figure 1: Average Number of Sick Days per Worker 1993 - 2010



Notes: Dots: Average annual claims of sick days for benchmark sample. Solid line: Fitted linear trend.

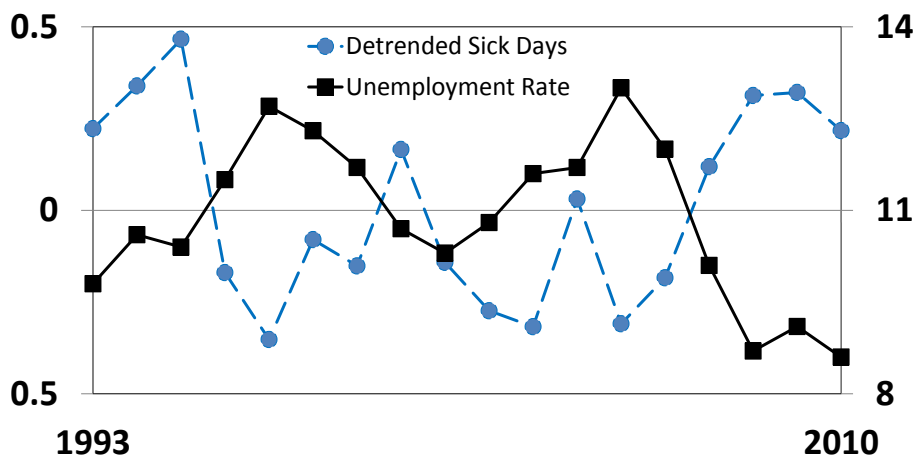
A second empirical characteristic is the strong pro-cyclical pattern of average claims of sick leave in Germany once the time series is de-trended.¹⁵ Figure 2 depicts the absolute deviation of average number of sick days from the linear trend (dashed line) and the

¹⁴A potential explanation might be pressure from abroad on the German labor market. Another explanation might be that the decline is due to technological progress in medical treatment.

¹⁵Pro-cyclicality of sick days is also documented for other countries, see Leigh (1985), Askildsen et al. (2005).

unemployment rate for Germany (solid line). Average number of sick days is high when the German unemployment rate is low. The correlation between the de-trended time series of sick days and the German unemployment rate is for the benchmark sample -0.7494 .¹⁶

Figure 2: Deviation of Sick Days from Linear Trend and Unemployment Rate



Notes: Dashed line (left axis) is absolute deviation of average number of sick days from the linear trend for benchmark sample; solid line (right axis) is the German unemployment rate.

To control for the composition effect, i.e. the absence behavior of the marginal worker, I construct a sub-sample consisting of workers that never report to be unemployed and have been observed for at least five consecutive years. This sub-sample shows on the one hand a lower number of average sick days compared to the benchmark sample. On the other hand the cyclical pattern of always-employed sample is still distinctive negative with a correlation of -0.5402 .¹⁷ The remaining correlation supports the incentive effect, i.e., in times of low reemployment workers reduce their number of sick days 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

The economic trade-off between taking sick days and recover on the one side and increased layoff probability on the other side is also shown in another pattern of sick days. A cross sectional analyses of sick days exhibits remarkable differences of average sick leave use

¹⁶In Appendix 7.3 I provide additional robustness checks using other measure of central tendencies, e.g., the number of sick days for the median worker.

¹⁷Other composition effects would occur if specific occupation groups or sectors that exhibit high sick days, e.g. the construction sector, are hit stronger by business cycles than others. In Appendix 7.3 I show that the general pattern of pro-cyclical behavior holds for all occupation and sectors.

between income groups.¹⁸ Figure 3 plots the average number of sick days (solid line) for each income quintile. It additionally shows the average self-reported health (dashed line) for each quintile. The used sample is restricted to workers between age 40 to 50. Controlling for age is important as age is highly correlated with sick days, health and income. Pooling all observations would lead to a bias in the results as poor people are more likely to be young and therefore healthy and using less sick days.¹⁹ Workers in the top income quintile have the lowest number of average sick days. Workers in the medium income quintile claim on average more sick days than workers in the bottom income quintile.²⁰ On the contrary, the health profile over income quintiles is monotonically increasing. The poorest workers have the lowest average health where the top income quintile shows the highest average health.²¹

Health and sick days are naturally related, i.e., worker in bad health are more likely to be sick and stay at home. Differences in health could potentially explain the small use of sick leave in the top income quintile compared to the rest of the workforce. Rich people are on average less sick and do not need to stay at home to recover. However, the same rationale is puzzling on the other side of the income distribution. The ones that are most unhealthy use less sick days to recover than medium income workers that have on average a better health. This is further evidence that sick days are not mechanically tied to health and absence behavior of workers has a second determinant.

This graphical inspection is confirmed by estimating Equation (1) using number of sick days as 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 sick days. Health has the assumed protective effect against sick days. Other coefficients in Table 1 confirm former results. There is a long run negative trend in sick days of -0.0936 days per year. More importantly relating to the cyclicity of sick days, the coefficient of the unemployment rate is significantly negative. This means during periods of high unemployment average number of sick days are reduced.

Further insights to the characteristics of sick day use are provided by distinguishing between extensive margin, i.e., missing any day a year or not, and intensive margin, i.e., conditional on missing at least one day a year how many days the respondent is not at work.

The left panel of Figure 4 shows the extensive margin of sick days for workers (40–50).

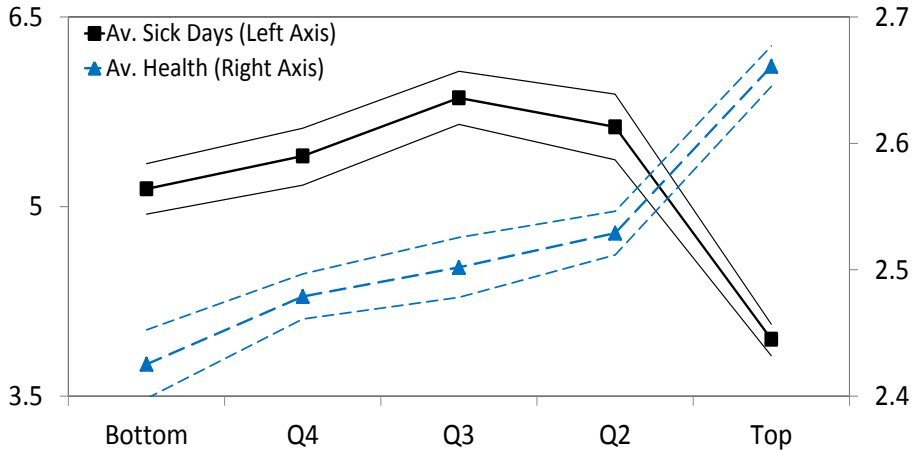
¹⁸Respondents are classified into income quintiles based on their monthly income in the previous year. If they report to be unemployed in the last year they are not considered for this section as they also have no sick days.

¹⁹All results hold also for other age bins, see Figure 7 in Appendix 7.3. See Appendix 7.3 for a version of these figures where gender effects are controlled for. The qualitative results are not changing.

²⁰This pattern also exists for the median number of sick days in each income quintile and other sick days cut off levels.

²¹A simple probit regression of income on health(good or bad) 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 Sick Days and Health over Income Quintiles



Notes: Dashed lines (right axis) show average self-reported health and 95% bootstrap confidence interval. Health is reported on ordinal five point scale where 0 denotes "bad" health and 4 denotes "very good" health. Solid lines (left axis) present average sick days of workers(40 – 50) and 95% bootstrap confidence interval separated for income quintiles.

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

Sick Days	(1)	(2)	(3)
Log income	17.8502***	0.0736***	-1.9705***
Log income squared	-1.1450***	–	–
Health	-1.8104***	-0.0900***	-1.6077***
Wave	-0.0876***	0.0013	-0.1758***
Unemployment rate	-0.1317***	-0.0054**	-0.1435***
Observations	65,653	65,653	36,108

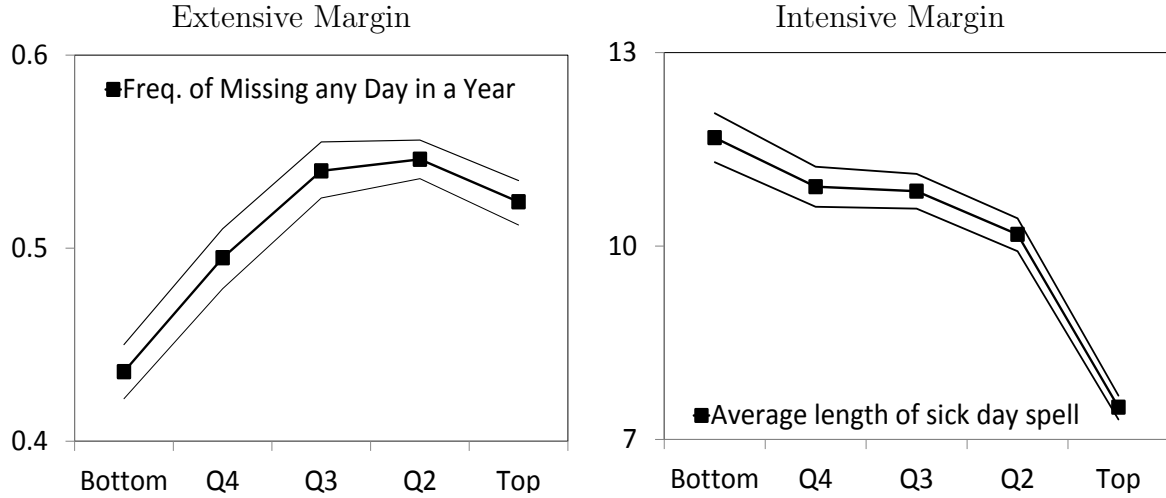
Notes: *** Significant at $\alpha = 0.01$, ** Significant at $\alpha = 0.05$, * Significant at $\alpha = 0.1$.

All regressions used the benchmark sample. Additionally to shown controls the regressions include sex, age, age² years of education and health². Robust standard errors are clustered on the personal level. Column (2) shows marginal effects at means.

Workers in the bottom income quintile exhibit the lowest probability to miss any day in a year. The higher the income group the higher the higher the probability to miss at least one day. Only at the very top the extensive margin seems to decrease. This pattern is also confirmed by estimating Equation (2) using the extensive margin as dependent variable. The results are presented as odd ratios in the second column of Table 1. The estimate for log income is highly significant and confirms the positive relationship between income and a high probability to miss any day. Other results show an unsurprising protective effect of self-reported income against missing any day.

The right panel of Figure 4 shows the intensive margin of sick days for workers (40 – 50). In contrast to the extensive margin the intensive margin is monotonically decreasing over income quintiles. The decline in conditional averages originates from

Figure 4: Composition of Sick Days over Income Quintiles



Notes: Left panel: Frequency of absence at least one day a year (extensive margin) at work of workers(40 – 50) and 95% bootstrap confidence interval separated for income quintiles. Right panel: Average number of sick days conditional on being at least one day sick (intensive margin) and 95% bootstrap confidence interval of workers(40 – 50).

different distribution length of sick day spells. The top income quintile have a higher probability of experience short sick spells (up to 14 days). On the other hand workers in the bottom income quintile have a high probability to have longer sick spells (more than 14 days). The third column of Table 1 displays results of estimating Equation (1) using the intensive margin as dependent variable. The results confirm that a high income has a protective effect against sick days. The higher the income the fewer are the number of sick days a year conditional on being sick. Coefficients on health and recession have again the expected sign.

Summarizing, there is a remarkable difference in utilization of sick leave between income groups. The top income group has a low average but the highest probability to miss a day and the highest probability of a short spell when sick. The medium income quintile has the highest average of sick days. The bottom income quintile has the worst health but the lowest number of sick days on average. It also shows a huge discrepancy in terms of lengths of sick spells. It has the highest probability of missing not one days at work but the lowest probability of missing only few days and then again the highest probability of missing more than 14 days.

2.3 Micro Evidence Using Panel Data

The panel structure of GSOEP allows to carve out further facts about sick days. First of all sick days are persistence. People that report sick days report to have also higher sick days in the next year. Including lagged sick days in estimating Equation (1) yields a positive and significant estimate. All other results remain qualitatively unchanged.

More important is the relation of sick days and unemployment. Table 2 shows descriptive statistics of workers that report employment in the previous year. This group is separated in two sub-samples. One in which respondents also work in the current period and one in which worker report to be unemployment in the current period. Workers who are laid off miss in the previous year on average 1.7 days more at work. I additionally compute the average of sick days of a worker over the last three years. The average sick days are more than 0.9 days higher for those that lost their job. Laid off workers are also poorer, less healthy and more likely to be male.

Table 2: Summary statistics of employed workers

	employed in prev. period	stayed employed	become unemployed
Sick days in prev. year	5.62	5.54	7.25
Av. sick days in prev. 3 years	5.14	5.11	6.02
Age	40.66	40.71	39.56
Income in prev. year	2,643€	2,684€	1,704€
Health in prev. year	2.64	2.64	2.55
Male	65%	65%	61%
Observations	103,852	99,435	4,417

Notes: Descriptive statistics for sample used in panel Logit model. Only workers that were employed in the previous year.

Table 3 shows the results of estimating Equation (3). The first column present the regression coefficients that later will be used in the structural model. The second column presents the results for the same regression but also including additional controls variables. e.g. sex and education. In both columns lagged income quintile has a negative sign, i.e., the highest probability of becoming unemployed is among workers in the bottom income quintile. Good health on the other side has a protective effect against the risk of unemployment. Healthier worker are less likely to become laid off. The key results are the coefficients of number of sick days on the risk to become unemployed. In both columns sick days show positive and highly significant results. The higher the number of sick days the more likely it is to become unemployed in the next year. ²²

In a next step I show that these results also hold estimating Equation (4) including the fixed effect component. For a better understanding I report in the last two columns the results as odd ratios. The effect of sick days on unemployment is qualitatively not effected and still highly significant. The results for the effect of income and the recession

²²In the relationship between sick days and layoffs might also exist the problem of reverse causality. Workers that know that they will lose their job could take sick days without fear of retaliation. This effect should be attenuated in the average sick days regression. I also run a regression for lagged sick days, i.e. number of sick days in the second to last year. They also show positive significant results. The problem with this specification is that the sample size is heavily reduced. I exclude in this specification all respondents that were unemployed in the previous year. Higher number of sick days the second to last year however increased the likelihood of unemployment in the last year and therefore the exclusion.

dummy on risk of unemployment remain unchanged. Health has still the same qualitative sign but becomes insignificant in the model without fixed effects. In the last column I replace sick in the previous year for accumulated sick days over the last two years. The effect seems robust.²³

Table 3: Panel Results for Unemployment I

Unemployment	Coefficients		Odd Ratios	
	(1)	(2)	(3)	(4)
Sick Day prev. Year	0.0256***	0.0260***	1.0192***	–
Sick Day prev. 3 Years	–	–	–	1.0162***
Lagged Income Quintile	-0.6807***	-0.7134***	0.7709***	0.8344***
Lagged Health	-0.1029***	-0.1138***	0.9381	0.9243
Age	0.0131***	0.0140***	1.1645	0.8602
Recession	0.3197***	0.2643***	1.7424***	1.7363***
Controls	No	Yes	Yes	Yes
Fixed Effects	No	No	Yes	Yes

Notes: *** Significant at $\alpha = 0.01$, ** Significant at $\alpha = 0.05$, * Significant at $\alpha = 0.1$. Controls include sex, years of education, and year dummies. Column (3) and (4) report odd ratios at population average. Robust standard error clustered.

3 Full Model

In this section I describe a dynamic stochastic model of work absence decisions. It captures both standard consumption-saving decision and sequential decision-making behavior of employed individuals with acute illnesses. In order to do so I distinguish between acute sickness and the overall health status. Both are interconnected but are two separate things. I will later use this model to evaluate the consequences of economic inequalities for utilization of sick leave and health as well as the consequences emanating from different paid sick leave systems.

3.1 Household's Problem

Agents start their economic life at age 20, retire at age 65. Since I do not model childhood and the retirement explicitly, I denote its 20th year as $j = 0$ and the terminal age of life $J = 45$. Following Gourinchas and Parker (2002) I postulate a retirement value function that summarizes the household problem at retirement.²⁴

²³In later sections it will be necessary to see whether the effect of sick days on unemployment is cyclical or not. To investigate this I estimate Equation (3) with fixed effects using an interaction term of *Recession* and the two measures of sick days. Both regressions show that the interaction term is positive but not significant. It seems that the influence on sick days on the probability of being laid off is the same in a recession as in a boom. Results are provided in the Appendix

²⁴Unlike most of the health economics literature I abstract from survival rates.

Health and Acute Illness The model distinguishes between the state of health, h_t , and a flow variable measuring acute sickness, Λ_t . An acute sickness affects utility and drives the evolution of overall health status whereas the overall health status determines the distribution of acute sickness realizations.

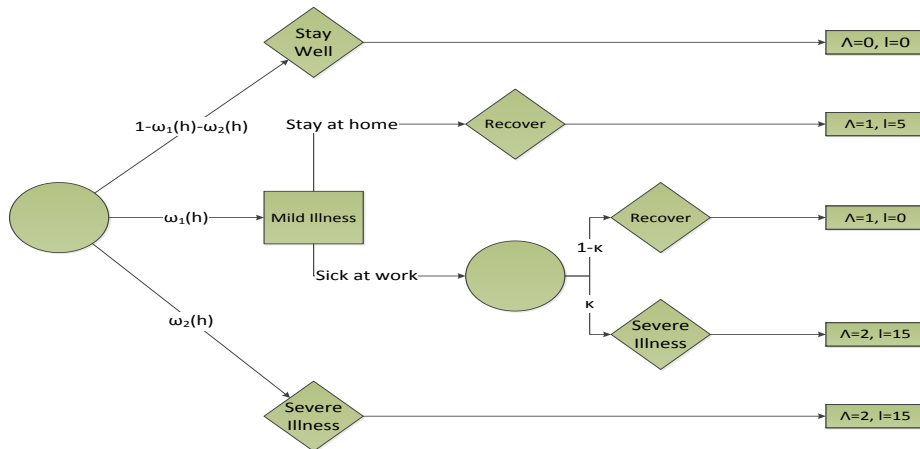
At each period an individual faces the risk to either stay well ($\Lambda = 0$) or contracting one of two types of acute illnesses Λ ($\Lambda = 1, 2$) that differ in their severity. The probability of contracting an illness of type k , $\omega_k(h_t)$, depends on the overall health status h of an individual. Upon becoming ill with the less severe type ($\Lambda = 1$), individuals can decide whether or not to be absent from work. By staying home from work, the individual takes $l_t = \tilde{l}$ sick days. By going sick to work, the individual does not take sick days $l_t = 0$. In this case, however, the mild type ($\Lambda = 1$) might become the severe type ($\Lambda = 2$) with a probability κ . When ill with the severe type ($\Lambda = 2$) individuals have no choice but to stay at home and take sick days $l_t = \bar{l}$ where $\bar{l} > \tilde{l}$. This is regardless whether the individual contracted the severe type first-hand or it evolved out of the mild type.

To examine the dynamic effect of the decision of sick to work trade-off through future health risks, the model includes a health production function. This allows future overall health to depend on current health, sickness and related treatment and a random element, ε_t^H .

$$h_{t+1} = H(h_t, \Lambda_t, \varepsilon_t^H) \quad (5)$$

Note that individuals face only a decision regarding their health when contracting the mild illness. This decision -about going to work sick or not- is made simultaneously to the saving consumption decision at the beginning of the period. The analysis ignores preventive treatment, so employed individuals who are free from acute illness ($\Lambda = 0$) have no illness-related absences as they cannot improve health.

Figure 5: Events and Decisions - Acute Illness



Notes: Setting for acute sickness.

Preferences Agents value current consumption c_t and dislike acute sickness $\Lambda_t \in (1, 2)$. I will assume that their preferences are additively separable over time and they discount the future at time discount factor β . Expectations are taken over stochastic employment and health history

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

I assume that the period utility function is of the familiar CRRA form given by

$$u(c_t, h_t) = \frac{[\psi c_t^\gamma + (1 - \psi) \Lambda_t^\gamma]^{\frac{1-\sigma}{\gamma}}}{1 - \sigma}.$$

The parameter ψ measures the relative importance of the consumption in the utility function. Noting that there are situations where being well and consumption are complements (e.g., marginal utility of a ski trip is lower for a sick person) and other situations where they are substitutes (e.g., marginal utility of hiring a maid is higher for a sick person), the functional form allows for flexibility and assume the elasticity for substitution between consumption and being well to be determined by $\frac{1}{1-\gamma}$. The parameter σ determines the inter-temporal elasticity of substitution.

Income, Employment Status and Unemployment Rate Agents differ with respect to their age and their predetermined type. These sources of heterogeneity affect an agent's labor productivity which is given by

$$\Gamma_k \lambda_j.$$

First, the labor productivity differs according to the age of an agent: λ_j denotes average age-specific productivity of cohort j . Second, each household belongs to a particular group k that shares the same average 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 groups relative income. I introduce these differences in order to generate part of the cross-sectional income and thus wealth dispersion, cf. Krueger and Ludwig (2007). Notice that health does not directly affect labor productivity.

The agent gets only paid the wage rate w , for the time she works $1 - l_t$. For the sick time l_t the household could get reimbursed by the government with wage dependent replacement rate b^S .²⁵

Central to this paper is the relationship between sick days and the probability to become unemployed. Theoretically there could be different explanations. First, health

²⁵This reimbursement is actually not done by the government but by the employer. As I do not model the firm side I make this shortcut. However, it would be interesting to see how this different setting would alter the result in a general equilibrium.

is important for the productivity of a worker. Employers cannot directly observe health and use sick days as a signal for health. This assumes that sick days are driven by health as shown in table 1. Second, related to the shirking literature, employer can only imperfectly monitor workers in terms of sick days. The higher the number of sick days the higher is the probability to be discovered. Third, due to the structure of the German system, sick days are costly for the employer. Sick days are also persistent over time. Therefore employer might get rid of worker with high sick days to save these costs. In all cases higher sick days lead to a higher layoff probability. For the following structural model it is not important why sick days are a good predictor of future unemployment. It is important that worker know about this fact and take it into account when optimizing their number of sick days.

Most importantly, the income of a household depends crucially on the employment status I . While working the household earns labor income. At the end of each period the agent may be dismissed. The probability to keep the job, ϕ , depends on age, skill type, health of the worker as well as the current unemployment rate. Additionally, workers can reduce the layoff probability in the next period by reducing the number of sick days.²⁶

$$\phi = \phi(k, j, h, \bar{u}, l)$$

While unemployed the agent gets unemployment benefits, $b_{j,k}^U$, which depend on age and type of the agent. The probability to find a new job when unemployed, $\bar{\phi}$, again is determined by age, skill health, and general economic conditions. They ,however, do not depend on the number of sick days.

$$\bar{\phi} = \bar{\phi}(k, j, h, \bar{u})$$

The evolution of the unemployment rate in my model is exogenous (i.e., I do not model general equilibrium effects or the firm side) and is the main driving force behind the model. For simplicity I model only two states of the general economic conditions, i.e., "boom" and "recession".

Budget constraint and borrowing limit 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 for next period:

$$a_{t+1} + c_t = (1 - \tau_t)(1 - l_t)\Gamma_k\lambda_j w + l_t b^S I_t + b^U(1 - I_t) + R a_t$$

²⁶It would be nice to do that for the accumulated sick days over some periods but this would require an additional endogenous state and is computational burdensome.

3.2 Government Policies

The government imposes a flat income tax, τ . The collected revenues are used for three main purposes: (i) to finance the unemployment insurance b^U , (ii) to finance the paid sick leave b^S and (iii) finally, to finance the government expenditure, G , that does not yield any direct utility to consumers (because of either corruption or waste). The residual budget surplus or deficit, T , is distributed in a lump-sum fashion to all households regardless of age. I assume that the budget of the government is balanced at all times.

$$\sum_i \tau_t y_{t,i} I_{t,i} = G + \sum_i [b^U (1 - I_{t,i}) + b^S I_{t,i} l_{t,i}] \quad (6)$$

3.3 Individual's Dynamic Program

I model the decisions to miss work during an episode of acute illness as the sequential choices of workers solving a discrete choice stochastic dynamic programming problem. At each discrete period of an illness the forward-looking individual chooses whether or not to miss work based on expected utility maximization.

Individuals, at the beginning of period t are indexed by their age j , their group k , their asset holdings a , their health state h , and their job employment status I . To simplify the analysis, I assume that the factor prices are exogenous. Each individual starts their life in the best health states $h_0 = 1$ and is endowed with initial assets a_0 . Thus their maximization problem reads as

$$\begin{aligned} W(j, k, a_t, h_t, I_t, \Lambda_t^{ante}) = & \max_{c_t, l_t, a_{t+1}} u(c_t, \Lambda_t^{post}) \\ & + \beta \sum_{h_{t+1}} \sum_{I_{t+1}} \Pi(h_{t+1} | h_t) \Phi(I_{t+1} | I_t) \sum_{\Lambda_{t+1}^{ante}} \omega(h_{t+1}) W(j+1, k, a_{t+1}, h_{t+1}, I_{t+1}, \Lambda_{t+1}^{ante}) \quad (7) \end{aligned}$$

subject to the constraints

$$\begin{aligned}
a_{t+1} + c_t &= (1 - \tau_t) y_t I_t + b^U (1 - I_t) + R a_t \\
y_t &= (1 - l_t) \Gamma_k \lambda_j w + l_t b^S \\
a_{t+1} &\geq 0 \\
I_{t+1} &= \begin{cases} 1 & \text{with probability } \Phi_t \\ 0 & \text{with probability } 1 - \Phi_t \end{cases} \\
\Phi_t &= \begin{cases} \phi(k, j, h_t, \bar{u}_t, l_t + \xi_t), & \text{if } I_t = 1 \\ \bar{\phi}(k, j, h_t, \bar{u}_t), & \text{if } I_t = 0 \end{cases} \\
\Lambda_t^{ante} &= \omega(h_t) \\
\Lambda_t^{post} &= \chi(\lambda^{ante}, l_t) \\
h_{t+1} &= \Pi(h_t, \Lambda_t^{post}, \varepsilon_t^H)
\end{aligned}$$

Definition 1 A stationary competitive equilibrium of this economy for given sick pay schemes b^{sick} , tax rate τ , wage w , and risk-free interest rate r is a set of decision rules, $\{c_t(z), l_t(z), a_{t+1}(z)\}$ and value functions W where $z = (t, j, k, a_t, h_t, I_t, \Lambda_t)$ such that:

1. Given initial conditions $W(t, \cdot)$ solves Eq. (7) and decision $c(t, \cdot), l(t, \cdot), a_{t+1}(t, \cdot)$ are the associated policy functions.
2. Government policies satisfy Eq.(6) in every period.

4 Parameter Estimation and Calibration

In this section I discuss the specification of the model parameters. I need to choose parameters governing the employment status, 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 the GSOEP data directly. These include the parameters governing the probability to get laid off $\phi(l, \bar{u})$; as well as the productivity difference Γ_k and λ_t . Third, (and given the parameters obtained in step 1 and 2) the remaining parameters (mainly those governing the health transition Π , acute illness shock process *omega*, χ and preferences) are then determined through a method of moments estimation of the model with GSOEP wage and health data. I now describe these three steps in greater detail.

4.1 A Priori Chosen Parameters

Table 7.5 show the parameter that are fixed exogenously with their values. The model period is one year. The life span of an individual is $J = 45$ periods. I assume that the interest rate, R is determined exogenously by world factors in an open-economy equilibrium and following Fernandez-Villaverde and Krueger (2011) I set $R = 4\%$. Then I select two preference parameters. Consistent with values commonly used in the quantitative macroeconomics literature I choose a risk aversion parameter of $\sigma = 2$ and a time discount factor of $\beta = 0.96$ per annum. 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)).

Policy Parameters For the benchmark calibration I choose the current institutional setting for Germany. It shut 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. The second I set the benchmark paid sick leave coverage to 100% of the current wage. So there is no direct reduction in income of sick leave for individuals. This is important to isolate the indirect effect of income on health via risk of unemployment. I set the unemployment benefits to 67% of the former wage of a worker. This is the current German setting for the first 12-24 month in Germany.²⁷ The German income tax I assume to be 35%, the government revenue that is not needed for the paid sick leave and unemployment benefits I assign to government consumption G .

4.2 Parameters Estimated Directly from the Data

In a second step I estimate part of the model parameters directly from the data, without having to rely on the equilibrium of the model. First, I obtain the probability of becoming unemployed Φ directly from the data. The following additional model elements are also directly determined from the data.

Job Keeping Probability The probability to lose the job $1 - \phi$ is directly taken from the estimation in section 2.3. The model computes for each combination of skill k , age j , health status h , economic state \bar{u} and number of sick days l the predicted probability

²⁷The unemployment setting has undergone a major reform in 2005. This might limit the historical comparison between model output and GSOEP data.

to retain the job. The used coefficients are taken from the first column of Table 3.

$$\hat{y} = \alpha_{0,1} + \alpha_{1,1} * k + \alpha_{2,1} * j + \alpha_{3,1} * h + \alpha_{4,1} * \bar{u} + \alpha_{5,1} * l$$

$$1 - \phi(k, j, h, \bar{u}, l) = 1 - \frac{e^{\hat{y}}}{1 + e^{\hat{y}}}$$

The probability to stay unemployed $\bar{\phi}$ is computed the same way except using sick days.²⁸

$$\hat{y} = \alpha_{0,2} + \alpha_{1,2} * k + \alpha_{2,2} * j + \alpha_{3,2} * h + \alpha_{4,2} * \bar{u}$$

$$1 - \bar{\phi}(k, j, h, \bar{u}, l) = 1 - \frac{e^{\hat{y}}}{1 + e^{\hat{y}}}$$

Labor Productivity Using the GSOEP data on income I compute the age-dependent productivity λ_j . I classify five different income skill types according to their annual income. Each individual therefore has a certain skill productivity Γ_k .

4.3 Parameters Calibrated Within the Model

In a final step I now use my model to find parameters governing the transition probabilities for health, the distribution of acute sickness shocks, and preference parameters governing disutility from acute sickness.

Health Transition Empirically the health stock, h_t , is defined using self-reported health status. Thus h_t takes one of five values: 1-”poor”, 2-”fair”, 3-”good”, 4-”very good”, 5-”excellent”. Following Khwaja (2010), the stochastic health production technology is specified to have a multi-nomial logit form with the index function for transition from health stock level h_t to health stock level $q = 1, \dots, 5$, at time t , $t = 1, \dots, T$, given by

$$\hat{y} = \beta_0 + \beta_1 h_t + \beta_2 \lambda^{post}$$

$$\Pi(h_t, \Lambda^{post}) = 1 - \frac{e^{\hat{y}}}{1 + e^{\hat{y}}}$$

I estimate the parameters of the health production function to match health average health and differences between age groups.

Acute Sickness Shock The health shock should simulate the acute illness case where the agent is forced to stay at home and recover. I make the ad-hoc assumption that an

²⁸The corresponding regression is contained in Appendix Table 3.

illness is severe if the worker takes more than three weeks (15 days) of sick leave. The probability to face an acute health shock is conditioned on health.

Health Preference The preference parameter for consumption and acute sickness are important so that high income individuals choose to invest in their health via sick days. Borrowing constraint households cannot follow this optimal strategy.

4.4 Model Fit and Benchmark Results

Using the parameter, I compute the model using standard numerical methods.

Figure 6 shows the sick day distribution of bottom and top income quintiles for different health states. One can see that the model can replicate the difference in sick days in each income quintile in both important features. First poor people take less often sick days and second when missing any day there average is higher. This is remarkable as it reflects an economic decision. The parameters that govern the onset of acute sickness are independent of the skill type and only depend on the health state.

5 Policy Evaluation

The determination of policy-invariant structural parameters allows for the introduction and evaluation of different policies that affect the financial constraints of a consumer's decision-making problem. The policy instrument in the paper involves sick leave coverage. [To be added]

US - No mandatory SP

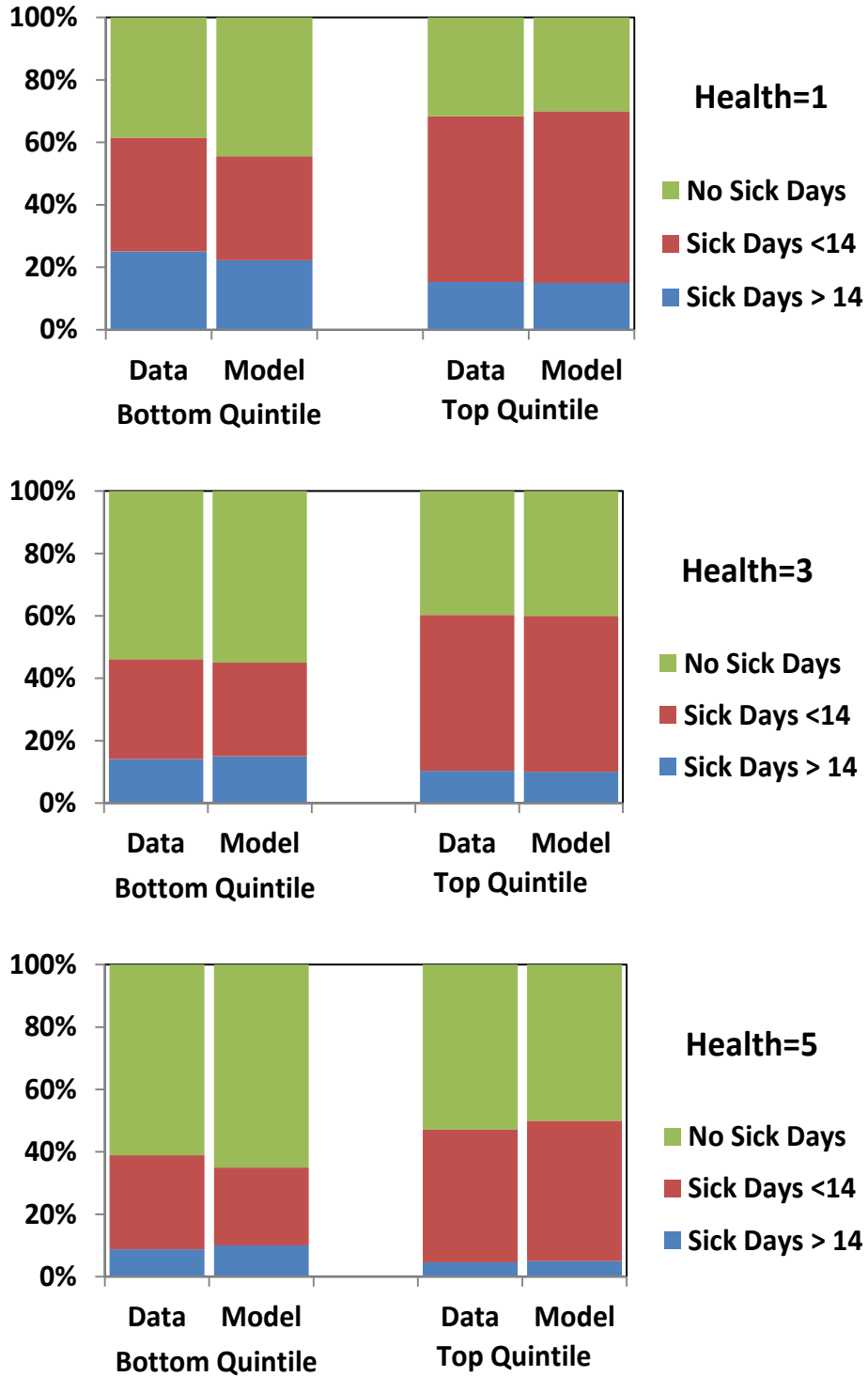
UK - Wage independent SP

France - Grace period for SP

6 Conclusion

In this paper, I studied the relationship between sick leave and unemployment and its effect on the evolution of health. Using data from the GSOEP I document new empirical facts on sick leave by income quintiles. First, the average number of sick leave exhibits a hump shaped pattern over income quintiles. Second, bottom and top income worker differ significantly in the pattern of sick leave. Poor individuals try to not miss any day or take a high number of sick days. Rich individuals take constantly a moderate number of sick days.

Figure 6: Number of Sick days for bottom and top income quintile



Notes: Left: Data. Right: Model Outcome.

I develop and estimate a life cycle model of health that can account for these facts. The main feature of my model is entangle the employment risk with health-related decisions. Moreover, I incorporate important features of the German labor market system into my model, such as universal health care and sick leave coverage.

I estimate my model using both micro and macro data. Then I use my model to analyze the macroeconomic effects of a counterfactual policy analysis.

References

- Ales, L., R. Hosseini, and L. E. Jones (2012, March). 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 Iivind Anti Nilsen (2005). Unemployment, labor force composition and sickness absence: a panel data study. *Health Economics* 14(11), 1087–1101.
- Bundesagentur fuer Arbeit (2014). Arbeitslosigkeit im zeitverlauf. Technical report, Bundesagentur fuer Arbeit.
- Chetty, R. (2008). Moral hazard versus liquidity and optimal unemployment insurance. *Journal of Political Economy* 116(2), pp. 173–234.
- Cole, H. L., S. Kim, and D. Krueger (2014, April). 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.
- Ehrlich, I. and H. Chuma (1990, August). A Model of the Demand for Longevity and the Value of Life Extension. *Journal of Political Economy* 98(4), 761–82.
- Fernndez-Villaverde, J. and D. Krueger (2007, August). Consumption over the Life Cycle: Facts from Consumer Expenditure Survey Data. *The Review of Economics and Statistics* 89(3), 552–565.
- Fernndez-Villaverde, J. and D. Krueger (2011, November). Consumption And Saving Over The Life Cycle: How Important Are Consumer Durables? *Macroeconomic Dynamics* 15(05), 725–770.
- Gilleskie, D. B. (1998, January). A dynamic stochastic model of medical care use and work absence. *Econometrica* 66(1), 1–46.
- Gourinchas, P.-O. and J. A. Parker (2002, January). Consumption Over the Life Cycle. *Econometrica* 70(1), 47–89.
- 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, 02). 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, June). Health investment over the life-cycle. Working Papers 201210, University of Hawaii at Manoa, Department of Economics.
- Hesselius, P. (2007). Does sickness absence increase the risk of unemployment? *The Journal of Socio-Economics* 36(2), 288–310.
- Khwaja, A. (2010, May). Estimating willingness to pay for medicare using a dynamic life-cycle model of demand for health insurance. *Journal of Econometrics* 156(1), 130–147.
- Krueger, D. and A. Ludwig (2007, January). On the consequences of demographic change for rates of returns to capital, and the distribution of wealth and welfare. *Journal of Monetary Economics* 54(1), 49–87.
- 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, October). 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. (2011). Income Differences and Health Care Expenditures over the Life Cycle. 2011 Meeting Papers 478, Society for Economic Dynamics.
- 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.
- Ruhm, C. J. (2000). Are recessions good for your health? *The Quarterly Journal of Economics* 115(2), 617–650.
- Scheil-Adlung, X. and L. Sandner (2010). The case for paid sick leave. World Health Report Background Paper 9, World Health Organization.
- 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.
- 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.

7 Appendix

7.1 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.

The sick pay has to be provided by the employer for the length of six weeks. If a worker becomes sick again with the same sickness then the sick days are summed up until the six weeks are reached. The claim of sick pay renews if the worker suffers from a different illness or it has been more than 6 months that the worker was sick with the same illness. The worker receives the wage that she would have earned if she hasn't got sick. There is no grace period. If a worker becomes sick while she is on vacation her holiday entitlement is not reduced. The worker has to tell her employer immediately about the incapability of work. On the fourth day of the sick spell the worker has to send a sick certificate issued by practitioner.

Between October 1996 and December 1998 there was a temporary change in the law. The main changes were a reduction of wage replacement from the 100% to 80%. However this reduction only applied to a fraction of the German work force as collective labor agreements between unions and firms mostly kept the 100% wage replacement. Empirical research on this law discontinuity is done by Ziebarth and Karlsson (2010) and Ziebarth (2013).

7.2 Sample selection

Table 4 shows the descriptive statistics before and after sample selection. I use in these statistics weighted samples. Weights are provided by the GSOEP to match the German micro-census. The final sample is younger due to focusing on 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 sick days is slightly increased after sample selection. The higher unemployment rate is due to the exclusion of the part time and temporary worker.

Table 4: Descriptive Statistics for Sample Selection

	Whole Sample	Benchmark Sample
Male	48%	63%
Age	49	41
Years of Education	11.68	11.87
Health	2.32	2.49
Income	2,199€	2,576€
Unemployed	6.85%	12.39
Sick	9.61	10.35
Observations	345,760	137,910

Notes: Descriptive statistics before and after sample selection. Benchmark sample used in the cross section and panel analysis.

7.3 Robustness Check in Empirical Part

Different measures of sick days

Sick days of worker have a skewed distribution. Table 5 provides results for the correlation of the unemployment rate and different measures of sick days. First, it shows the results 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 or not. In the last three columns different cut of level for the maximum sick days are used. All of results are negative and in the same range as the benchmark result. The pro-cyclical pattern is extremely robust.

Table 5: Sick days and unemployment - Different sick day measures

Correlation	median	ext. marg.	max 120	max 60	max 30
Unemployment rate	-0.5357	-0.6093	-0.7789	-0.7970	-0.7586

Notes: Times series correlation of different measures of sick days and unemployment rate. First the sick days 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 maximum sick days.

Composition Effects

A potential different explanation of the cyclicity of sick days could arise if sectors (e.g., construction sector) with high usual high number of sick days 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 6 shows the exclusion of the construction sector does not alter the benchmark result. The correlation coefficient is only slightly reduced to -.65

I also check whether this cyclical behavior is different for different occupation type. GSOEP provides the ISCO88 classification and using the white/blue collar distinction as in the European working conditions surveys. Table 6 shows that for both subgroups the pro-cyclicality of sick days holds.

Table 6: Sick days and unemployment - Different sectors and occupations

Correlation	Without construction	Blue collar	White collar	Never unemployed
Unemployment rate	-0.7368	-0.6672	-0.6731	-0.5402

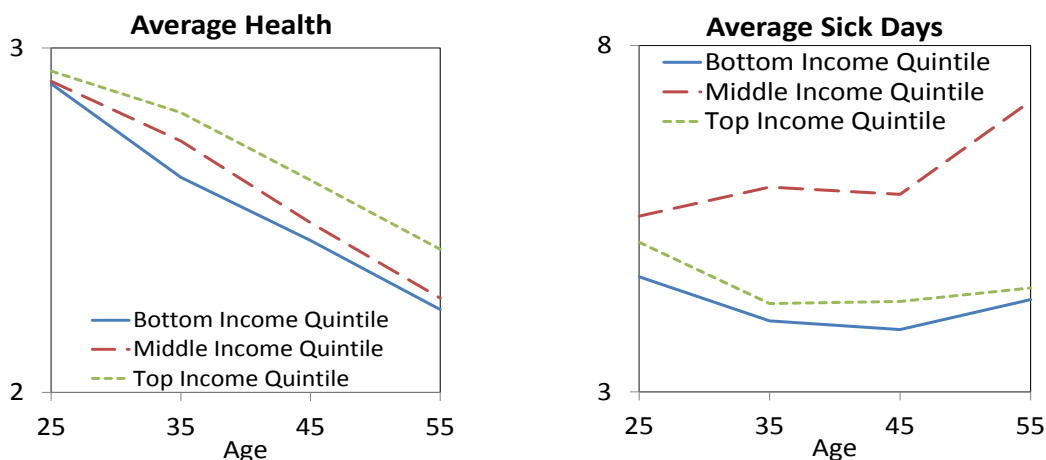
Notes: Times series correlation of average sick days and unemployment rate for different subgroups.

Age profiles in sick days and health

The right panel of Figure 7 confirms that the observed hump shaped income profile hold for all age-groups. The left panel also reveals that the income gradient in health is increasing over age. The difference in health between bottom and top income quintile almost five times as high for 50 – 60 year old workers than for 20 – 30 year old.

Needs to be reviewed. Control for cohort effects.

Figure 7: Sick Days and Health over Life Cycle by Income Quintiles



Notes: Left Panel: Average self-reported health of bottom (solid), medium (dashed) and top income (dotted) quintile over the life cycle. Right Panel: Average sick days of bottom (solid), medium (dashed) and top income (dotted) quintile over the life cycle. Age bins: 20 – 30, 30 – 40, 40 – 50, 50 – 60

Controlling for gender in sick day profiles over income quintiles

still to be added

Interaction Recession Sick Days

Table 7 show the relation between unemployment, unemployment rate and sick days.

Table 7: Panel Results for Unemployment II

Unemployment	(1)	(2)
Sick Day prev. Year	1.0206***	–
Sick Day prev. 3 Years	–	1.0579***
Recession	1.6159***	1.5475***
Interaction	1.0052	1.0055
Fixed Effects	Yes	Yes

Notes: *** Significant at $\alpha = 0.01$, ** Significant at $\alpha = 0.05$, * Significant at $\alpha = 0.1$. All regressions include age, income, health and year dummies. Reported are odd ratios at population average. Robust standard error clustered.

7.4 Computational Details

7.5 Estimation Results

Table 8: Fixed Parameters

Parameter	Description	Value
<i>Economy</i>		
T	Life span	10
R	Interest rate	0.04
w	Wage rate	1
<i>Preferences</i>		
β	Discount factor	0.9659
σ	Inter-temporal elasticity of substitution	2
θ	share of c in c - l combination	1
<i>Policy</i>		
b^U	Unemployment benefit	0.67
b^S	Sick leave replacement rate	1

Notes: Parameters taken from the literature.

Table 9: Directly Estimated Parameters

Parameters	Description	Value
$\alpha_{0,1}$	Constant ($I = 1$)	-3.2145
$\alpha_{1,1}$	Skill ($I = 1$)	-0.4172
$\alpha_{2,1}$	Age ($I = 1$)	0.0019
$\alpha_{3,1}$	Health State ($I = 1$)	-0.1296
$\alpha_{4,1}$	Unemployment rate ($I = 1$)	0.2554
$\alpha_{5,1}$	Sick Days ($I = 1$)	0.0171
$\alpha_{0,2}$	Constant ($I = 0$)	10.032
$\alpha_{1,2}$	Skill ($I = 0$)	0.2006
$\alpha_{2,2}$	Age ($I = 0$)	-0.0762
$\alpha_{3,2}$	Health State ($I = 0$)	0.2202
$\alpha_{4,2}$	Unemployment rate ($I = 0$)	-0.2379
Γ	Skill wage distribution	
λ	Age wage distribution	

Notes: Parameters estimated directly from GSOEP.

Table 10: Data Targets

Param.	Description	Data Targets
<i>Acute Illness and Health Transition</i>		
ω_1	Prob. to contract mild illness cond. health	Average Sick days less than 14 days
ω_2	Prob. to contract mild illness cond. health	Average Sick days more than 14 days
κ	Prob. untreated mild become severe illness	Difference in sick days between rich and poor
Π_1	Prob. to drop in health status after mild illness	Health differences over life cycle
Π_2	Prob. to drop in health status after mild illness	Health differences over life cycle
\tilde{l}	Required Time to Recover from mild illness	Average number of sick days
\bar{l}	Required Time to Recover from severe illness	Average number of sick days
<i>Preference for Acute Illness</i>		
ψ	Share of c in utility	Difference between bottom and top income quintile
γ	Elasticity b/w consumption and health	Difference between bottom and top income quintile

Notes: Calibrated parameters and their data targets.