Sick Pay and Labor Supply

Victoria Barone*
University of California at Los Angeles

June 23, 2022

JOB MARKET PAPER
[Click here for the latest version]

Abstract

This paper studies the optimal design of a paid sick leave system. I develop a model of work-absence behavior where risk-averse workers who face health risks decide their sick pay coverage and utilization. The social planner provides insurance against these risks—i.e., it provides paid sick leave—maximizing social welfare, which is a function of workers' utility and population health. The optimal paid sick leave scheme trades off the value gained from risk protection against the costs associated with behavioral responses: moral hazard and externalities due to the spread of contagious diseases. I obtain estimates of the model's parameters using individual-level data on sick pay utilization, and labor supply from the Chilean Paid Sick Leave System. I leverage exogenous variation in the temptation for shirking behavior induced by the day of the week a sick leave claim is filed to identify the key parameters governing workers' behavior. Finally, I analyze the impact on welfare and behavior of implementing the optimal contract and compare them with the welfare gains of alternative reforms to the paid sick leave system.

Keywords: optimal social insurance; moral hazard; sick pay; welfare effects

JEL codes: I12, I13, I18, J22, J28, J38

*Email: mvbarone@ucla.edu; website: baronevictoria.com. I thank Maurizio Mazzocco, Daniel Haanwinckel, Adriana Lleras-Muney, Michela Giorcelli, and Laura Wherry for their continual support and guidance throughout this research project. I especially thank Fernanda Rojas-Ampuero for her help in developing the project in its early stages. The project was supported by the Lewis L. Clarke Fellowship Fund and the Stan Ornstein Summer Mentored Research Fellowship Program. I benefited from facilities and resources provided by the California Center for Population Research at UCLA (CCPR), which receives core support (P2C-HD041022, NICHD); and the Wagner School of Public Service.

I Introduction

Social insurance programs offer protection for a broad range of risks that could be detrimental to individuals' well-being, such as health deterioration that limits one's ability to work. In particular, paid sick leave provides income replacement for workers who suffer from short-term impairments caused by non-work-related sickness (e.g., common flu).[1]

Paid sick leave programs help to preserve workers' health and prevent the spread of diseases in the workplace (Pichler, Wen and Ziebarth, 2021). Additionally, support for paid sick leave has grown substantially in the United States—one of the few countries in the OECD that do not have federal mandates to ensure universal employee access to paid sick leave.[2] Conditional on deciding to insure this health risk, several policy choices must be made. For example, what level of benefits should be paid? Should benefits rise or fall over the utilization span? A large body of literature has answered these questions in the context of unemployment risks (Hopenhayn and Nicolini, 1997; Chetty, 2008; Hendren, 2017) disability risks (Gruber, 2000; Low and Pistaferri, 2015), healthcare risks (Cutler and Zeckhauser, 2000; Einav, Finkelstein and Cullen, 2010; Handel, Hendel and Whinston, 2015; Marone and Sabety, 2022), and work-related injuries (Powell and Seabury, 2018; Cabral and Dillender, 2020). Nonetheless, the design of paid sick leave remains understudied. The main contribution of this paper is to address these questions.

This paper develops a theoretical and empirical framework to study the optimal design of a paid sick leave program. This optimal design requires understanding how coverage generosity and other program features impact individual behavior and welfare. To do so, I propose a model of work-absence behavior that illustrates how workers respond to a given paid sick leave program and decide their labor supply, sick leave utilization, and coverage. Taking into account these responses, the social planner designs the optimal contract to maximize social welfare, which is a function of workers' utility and population health to capture externalities due to the spread of contagious diseases. In the empirical section of this paper, I quantify the benefits of implementing the optimal sick pay scheme. To do so, I obtain estimates of the model's parameters exploiting individual-level data on

[1]While closely related to worker's compensation programs—which provide income replacement and medical benefits in case of work-related sickness—and disability insurance programs—which provide income replacement in case of permanent impairments to work—paid sick leave programs offer protection against the risk of contracting a disease which reduces ones ability to work for a short period and has a foreseeable recovery.

[2]While there is no federal mandate for paid leave, there has been more traction on paid sick leave policies at the state and local levels. Over the past decade, thirty-six U.S. jurisdictions have implemented sick pay mandates enacting laws that enable eligible workers to accrue paid time off from their employer (A Better Balance, 2021).

sick pay utilization and labor supply from the Chilean Paid Sick Leave System. I show that the model replicates salient features of reality internally (targeted moments) and externally (non-targeted moments). Finally, I analyze the impact on welfare and behavior of implementing the optimal contract and compare them with the welfare gains of alternative reforms to the paid sick leave system

In the theoretical part of the paper, I study the incentive problem created by paid sick leave insurance. To do so, I build a model where a risk-averse expected utility-maximizing worker optimally chooses sick leave coverage and sick pay utilization. This model illustrates how institutional features of a paid sick leave scheme influence workers' decision-making. When choosing their demand for sick days, workers trade off the utility cost of working while sick with the consumption loss from missing work when taking up sick leave, which is affected by the generosity of sick pay. This model incorporates two unique features. First, in contrast to the previous literature, my model focuses on how sickness affects the absence behavior of employed individuals rather than how it affects labor force participation. Second, the model allows workers' behavior to vary with the day of the week a sick leave claim is filed. Previous literature has shown that the temptation for shirking varies with the day of the week.[3] I incorporate this mechanism in the model of workers' behavior explicitly and exploit how days of the week shift the temptation for shirking for identification of the model.

The challenge in designing sick pay is that this benefit is conditioned on a difficult-to-verify event—the true inability to work due to sickness. This unobservability raises moral hazard concerns: shirking behavior and potential health externalities. Thus, when choosing the level of sick pay, the social planner balances the benefits of risk protection with the cost associated with moral hazard. For example, a more generous sick pay scheme would increase workers' well-being but provide incentives to overstate sickness, resulting in lost production and increased program costs. On the other hand, if a contract is not generous enough, some workers would work while sick, spreading diseases and reducing aggregate welfare. Thus, the optimal level of benefits depends on the behavioral responses of workers' to the compensation benefit design, the costs of sick work, production losses, and health externalities—which I quantify in the empirical application of the paper. Low replacement rates that increase as a spell becomes longer allow the planner to minimize moral hazard and offer higher protection for bigger losses—as longer spells involve more severe diseases—but could be detrimental to public health when contagious rates are high. Externalities lessen the effectiveness of non-payable periods, justifying a

[3] For example, (Card and McCall, 1996) documents a spike in back injury and sprain claims for workers' compensation on Mondays.

more generous contract.[4]

The empirical section of this paper leverages data from the Chilean Paid Sick Leave System. The Chilean system constitutes a great setting to study the optimal design of sick-pay benefits. First, Chile has a mandatory paid sick leave system where insurers cannot design sick pay plans and must follow the rules set by the central government regarding eligibility criteria and benefits. Thus, workers do not choose their sick leave plans alleviating adverse selection concerns. If workers could choose their sick pay coverage, we could expect individuals with preferences for more absences to self-select into jobs or plans with more generous provisions. The presence of adverse selection would result in an upward bias in the estimates of the moral hazard effect. Second, I exploit unique administrative data. I observe the universe of workers eligible to file a sick leave claim between 2015 and 2019 and their utilization of sick leave benefits. This database includes rich demographic information at the worker level: sex, age, worker's occupation and industry, monthly wages, a set of health indicators for chronic conditions, and the primary diagnosis related to a sick leave claim at the fourth digit of the International Classification of Diseases (ICD) 10th revision. I exploit key features of these data and institutional components of the Chilean system to estimate the parameters of the work-absence behavior model.

In the Chilean paid sick leave system, benefits vary according to the duration of the sick leave episode.[5] During the first three days of the sick leave spell, the replacement rate is zero—i.e., employees do not receive any amount of benefits. This non-payable period works like a deductible that resets every new sick leave span and is a common mechanism among sick leave programs to deter unjustified absenteeism.[6] Starting on the fourth day, there is full coverage of each missed day—i.e., the replacement rate is one. If the sick leave spans 11 days or more, the non-payable period is reimbursed. That is, claims with an 11-day or longer duration are fully covered.[7]

[4]Low replacement rates that increase as a spell becomes longer are the most common way sick leave benefits are designed. Interestingly, this is almost diametrically opposed to how optimal unemployment insurance should be theoretically designed to maximize search effort Hopenhayn and Nicolini (1997). Studies on the impact of non-payable or waiting periods on sickness absences by Petterson-Lidbom and Thoursie (2013) and Pollak (2017) have shown that they may increase absenteeism.

[5]This system aims to provide risk protection from impairments to work that are temporary and where full recovery is foreseeable. A separate program provides disability insurance to workers in the case of permanent impairments to work.

[6]These resettable deductibles are similar to those used in automobile or homeowners insurance: Separate deductibles apply to each loss. Many European paid sick leave systems have a similar waiting period (see Marie and Castello (2022) for the case of Spain and Pollak (2017) for the Italian experience.

[7]Panel (a) of Figure 1 presents days paid as a function of days on leave for claims of different duration. Reimbursement of the non-payable period after 11 days implies that the average replacement rate jumps discretely at 11 days, and it is non-constant (see panel (b) of Figure 1).

I leverage the model of workers' behavior and the granularity of the data to recover the distribution of health states and preference parameters. The empirical distribution of health states incorporates observed and unobserved heterogeneity across workers. I assume that observed heterogeneity across workers is a function of the disease a worker contracts, her age, and her occupation. I use the primary diagnosis, worker's age, and occupation from the claims data combined with data from the Peruvian Handbook of Recovery Times to recover the observed component of the health state.[8] I leverage the structure of the model and the discontinuity of the average replacement rate to recover the distribution of parameters governing the unobserved heterogeneity component.

The identification and estimation of preference parameters leverage exogenous variation in the temptation for shirking behavior coming from the day of the week a sick leave claim is filed. First, I document excess mass in the share of sick leave claims filed on a day of the week for a duration that implies a streak of days off work that includes the weekend—which I term a weekend-streak combination. Workers are, on average, 12% more likely to file a sick leave claim on a weekend-streak day than on other days of the week. I use this empirical fact to pin down the utility value of a sick leave claim filed on a weekend-streak day relative to the rest of the week (Figure 4 illustrates this result). Second, I provide evidence that workers use their leave for the purpose of leisure. While I cannot perfectly observe illness or recreation, I compare the choices of workers with similar diagnoses and observable characteristics who fall sick on different days of the week. I document that workers who fall sick earlier on the week tend to ask for longer sick leave claims, even after controlling for workers' observable characteristics. In terms of workers' preferences, this behavior implies that compliance costs—i.e., the cost of reporting the *true* health shock—increases with the duration of a sickness spell non-monotonically.

The model matches the main patterns of the data quite well. For example, the model predicts that if a worker realizes a health state of just under 11 days on leave, she will take advantage of the proximity to the full insurance region and fake her type to gain full coverage, which is consistent with the patterns I document from the data. Preliminary estimates imply that the current paid sick leave increases workers' welfare by 32% —relative to the situation without paid leave. These welfare gains are concentrated among workers who suffer more severe health shocks. Nonetheless, the presence of moral hazard—i.e., workers who overstate their sickness level in response to discontinuities in the benefit scheme—results in a reduction in days worked of 19% relative to the situation

[8] The Peruvian Handbook of Recovery Times states the average number of days a worker would require to recover from a disease and proposes adjustments to those days based on workers' age and occupation. It covers 2,763 unique disease codes at the fourth digit level of the ICD 10th revision. Appendix Appendix C. provides further details on the handbook data.

without paid leave.

Beyond addressing an important policy question, this paper contributes to several areas of the economics literature. This paper contributes to a large body of literature on public insurance programs that has modeled the trade-offs between protection against risk and moral hazard present in unemployment risks (Hopenhayn and Nicolini, 1997; Chetty, 2008; Hendren, 2017) disability and retirement risks (Gruber, 2000; Low and Pistaferri, 2015), healthcare risks (Cutler and Zeckhauser, 2000; Einav, Finkelstein and Cullen, 2010; Handel, Hendel and Whinston, 2015; Marone and Sabety, 2022), and work-related injuries (Powell and Seabury, 2018; Cabral and Dillender, 2020). Building on this literature, this paper is the first to propose a theoretical framework for designing the provision of paid sick leave and quantify the welfare gains of its implementation.

The design of worker's compensation and disability insurance are informative for paid sick leave policies—these three programs condition benefits in a difficult to verify state. Nonetheless, two characteristics of paid sick leave call for a specific design. First, the health shock—short non-work-related illness—implies that virtually every worker would benefit from paid sick leave access. Worker compensation and disability insurance programs are (mainly) designed to focus on a specific group of workers—prompt-to-accidents and elderly workers, respectively. Second, externalities that arise from contagious diseases make paid sick leave different from other social insurance programs in that suboptimal provision of sick pay generates a negative externality in the form of "contagious presenteeism" or "going to work sick" (Pichler and Ziebarth, 2017). This paper focuses on deriving the optimal paid sick leave contract and shows the importance of accounting for externalities in the design of public insurance.

This paper is closest to Maclean, Pichler and Ziebarth (2020) who evaluate the labor market effects of sick pay mandates in the United States and extends the Baily-Chetty framework of optimal social insurance to assess the welfare consequences of mandating sick pay. However, I focus on a different dimension of optimal insurance. This paper derives conditions on the optimal level of sick pay and quantifies the welfare consequences of implementing the optimal system and alternative policy reforms. To address these aims, I propose a structural approach. This approach is better suited than the sufficient statistics approach to evaluate the welfare consequences of policy experiments, especially when studying policy changes that are not small in nature—such as the increase in benefits for sick leave from a zero replacement rate to a full-replacement rate.

This paper also contributes to the empirical literature on sickness insurance. Exploiting arguably exogenous variations, the literature has shown that workers adapt their absence behavior to increases in benefit levels (Johansson and Palme, 2005; Ziebarth, 2013;

De Paola, Scoppa and Pupo, 2014; Ziebarth and Karlsson, 2014; Pollak, 2017; Marie and Castello, 2022, Cronin, Harris and Ziebarth, 2022). I make important contributions to this literature. This paper is the first to use administrative data on sick leave claims at the individual level.[9] These data cover all workers in Chile enrolled in the government-run healthcare insurance—this group represents 70% of the universe of Chilean workers. These data allow me to study daily leave-taking behavior and estimate the individual demand for sick pay. Additionally, these data are less prone to measurement error. Many papers have used survey questions that ask respondents how many days of work they have missed due to illness in a reference period. The use of survey data raises the usual measurement error issues with self-reported recall data and prevents researchers from distinguishing between extensive (incidence of absences) and intensive (length of absences) margin effects.

This paper proceeds as follows. Section II presents the theoretical framework and discusses the optimal design of a paid sick leave system. Section III describes the empirical setting I study and the data. Section IV presents the empirical implementation of the model. Section V presents the model estimates and main results. Section VII evaluates welfare and distributional outcomes. Section VIII concludes.

II Theoretical Framework

In this section I characterize the optimal sick leave insurance contract between risk-averse workers and a risk-neutral principal. I consider a model where a risk-averse expected utility-maximizing worker first makes an optimal sick leave insurance coverage choice, using her available information on her expected absences needs. Given this choice, the worker observes her realized health and makes an optimal sick pay utilization decision. In this section, I omit i subscripts to simplify notation and present the baseline version of the model. I later describe how individuals might vary across (i) distribution of health shocks, (ii) preferences over leisure, and (iii) their degree of risk-aversion.

II.A Workers

Workers are subject to a stochastic health shock θ, drawn from a distribution $G(\theta)$ with p.d.f. $g(\theta)$. The sickness level, θ, represents the number of days a worker is sick; θ is dis-

[9] Cronin, Harris and Ziebarth (2022) constructs a similar dataset for the Scott County School District (SCSD) in Kentucky, which allows for a detailed study of teachers' use of paid sick leave. While the data structure is similar to the one used in this paper, I observe sick leave utilization regardless of workers' occupations. Marie and Castello (2022) also exploit administrative for Spain, though their data is at the spell rather than individual level.

crete and bounded between zero and M.[10] I assume that higher values of θ are associated with longer sickness spans. The sickness distribution $G(\theta)$ accumulates positive mass in the no-sickness realization; i.e., a value of zero for θ corresponds to the healthy state. I begin with the worker's utilization choice and then I discuss her coverage choice.

Sick pay utilization. At the time of his utilization choice a worker is characterized by her health realization θ and her preference parameters, which I introduce next. Workers derive utility over consumption (c) and time outside of work (s) to recover from a health shock. Workers make these choices given their budget constraint: $c = wM - ws + wB(s)$, where w is the daily wage rate, M is the number of workable days in a month, and $B(s)$ represents transfers with respect to sick pay.[11] I assume that $B(s)$ is piecewise linear, so that the marginal replacement rate b_s is constant within a sick leave duration bracket $[\underline{s},\bar{s}]$. If workers have access to unpaid sick leave, $B(s) = 0$ for any s. A linear contract with full insurance—$b_s = 1$ for any s—corresponds to the benefit's function $B(s) = s$.

Workers' utility is separable in consumption and leisure and can be written as $u(c,s;\theta,\phi) = c + \phi f(s-\theta)$. The function $f(s-\theta)$ is concave in $(s-\theta)$ and it captures the utility gains (or losses) from time outside of work relative to the realized health state. The preference parameter ϕ reflects the opportunity cost of time outside work relative to time devoted to consumption. One convenient parameterization of utility as a function of days on leave is

$$u(s;\theta,\phi,\kappa) = w\left[M - s + B(s)\right] + \phi\left[(s-\theta) - \frac{1}{2\kappa}(s-\theta)^2\right] .$$

Workers choose sick pay utilization by trading off the cost of a day away from work $w(1 - B'(s))$ with the net gain in terms of leisure time $\phi(1 - 1/\kappa(s-\theta))$. The optimal sick pay utilization is $s^*(\theta,\phi,\kappa) = \mathrm{argmax}_s\, u(c(s),s;\theta,\phi,\kappa)$.

In this representation, the term $\frac{1}{2\kappa}(s-\theta)^2$ captures a compliance cost—that is, how costly it is for the worker to ask for more (fewer) days on leave than her health state. This compliance cost is increasing in the term $\frac{1}{\kappa}$. The parameter κ reflects workers' preferences over behaving as expected or revealing their "true" health status in a reduced-form manner. For example, it reflects the worker's willingness to engage in doctor shopping—the

[10]The sickness level is bounded to capture the fact that paid sick leave insurance aims to provide risk protection from impairments to work when full recovery is foreseeable. In the empirical application, I focus on sick leave claims with a duration of up to 30 days.

[11]I interpret M as coming from the utility maximization problem of a worker who does not face health risks and chooses her labor supply. For example, for individuals working full time, M is the number of workdays in a month.

effort the worker could exert to find a physician who would sign off on a longer leave—or the cost of being caught filing an unjustified sick leave claim. The proposed functional form for $f(s-\theta)$ assumes that penalties or costs of asking for more or fewer days on leave than the health state are symmetric. That is, missing work for an extra day has the same penalty as coming to work sick for a day. While this representation is convenient to illustrate the main mechanisms of the model and facilitate intuition, in the empirical application I relax the symmetry assumption and discuss the implementation of a more flexible specification for $f(s-\theta)$.

Insurance provision, by lowering the marginal cost of sick leave, increases sick pay utilization. That is, $s^*(\cdot)$ is nondecreasing in sick leave benefits $B(s)$. I refer to this responsiveness of sick pay demand to insurance coverage as moral hazard, following the health literature. This definition of moral hazard refers to "expost moral hazard"; i.e., it focuses on the responsiveness of workers' demand for sick pay to varying levels of insurance generosity and abstracts from actions workers can take to prevent deterioration of their health—which could be referred to as "exante moral hazard".[12]

The model aims to understand how illness affects the absence behavior of employed individuals rather than how illness affects labor force participation. Consequently, the model does not explicitly consider the disutility of effort when the agent is employed. Additionally, I do not explicitly model the direct cost of applying for sick leave; for example, the cost of a doctor's appointment or the need to complete paperwork. If these costs do not vary with workers' health state, they shift utility levels but do not affect marginal choices regarding sick pay utilization. These modeling choices have no consequences for the qualitative properties of the optimal contract derived below.[13]

Optimal utilization under linear contracts. To facilitate intuition, consider the case of a linear benefit scheme $B(s) = bs$ where $b \in [0,1]$. Under this assumption, the optimal choice of sick pay duration is

$$s^*(\theta,\phi,\kappa) = \max\left[0, \theta + \kappa\left(1 - \frac{w}{\phi}(1-b)\right)\right] .$$

Abstracting from the potential truncation at 0, in the case of full coverage ($b = 1$) the worker optimally chooses $s^* = \theta + \kappa$. This case is presented in panel (a) of Figure A1, for

[12]This definition follows conventional use of the term "moral hazard" in the health insurance literature. A fuller discussion of this (ab)use of terminology in the health insurance literature can be found in Section I.B. of Einav et al. (2013).

[13]The presence of an application cost could, nonetheless, affect the program's take up. In Appendix AA, I consider a fixed cost c of applying for sick pay.

strictly positive values of κ sick pay utilization is above the worker's health state. Lower compliance costs (higher κ) are associated with greater deviations from the worker's health status. The not-paid sick leave ($b = 0$) is presented in panel (b) of Figure A1. If the worker's valuation of time outside work ϕ is greater than the wage rate, the worker would optimally choose to ask for longer sick leave claims. Panels (c) and (d) of Figure A1 consider the case of partial coverage—i.e., a strictly positive replacement rate less than one $b \in (0,1)$ —for different values of κ and ϕ.[14] In general, moral hazard is increasing in ϕ and κ. That is, *all else equal*, a greater valuation of time outside work (a higher ϕ) is associated with a longer sick leave claim. Similarly, lower compliance costs (higher κ), are associated with longer sick leave claims.

Day of the week. An additional determinant of sick pay utilization is the day of the week a claim is filed. Some combinations of days of the week and sick leave duration give workers more utility because of the presence of a weekend. For example, consider a worker who is sick on a Thursday. In this case, two days on leave would represent a streak of four days to recover from disease: The worker would be on leave on Thursday and Friday and then also have the weekend off.[15]

I extend the model of workers' behavior in two dimensions to capture how the day of the week affects workers' coverage choices. First, I assume that workers are subject to a stochastic health shock of duration θ that occurs on the day of the week dow. Thus, the worker's health state is characterized by (θ, dow), drawn from the distribution $G(\theta, dow)$. Second, I modify the utility function to account for the day of the week a sick leave claim is filed. I assume that the day of the week does not impact utility directly, but it affects the utility a worker derives from sick leaves of duration s. Thus, the worker's utility given her health state (θ, dow) is represented by

$$u(c,s_l,s_c;\theta,dow,\phi,\kappa) = c + \phi\left[(s_l - \theta) - \frac{1}{2\kappa}(s_c - \theta)^2\right] + \phi q \; \mathbb{1}\{\text{weekend}\}\ ,$$

where dow is a state variable that indicates the day of the week a worker falls sick, s_l represents the number of business days a worker is on leave, and s_c represents the *total* number of days a worker is on leave regardless of the day of the week. The term $\phi q \; \mathbb{1}\{\text{weekend}\}$ captures the extra utility a worker derives when the sick leave claim has a duration that allows the worker to not return to work until after the weekend. The indicator variable

[14] In this case, workers care about the effective valuation of time, i.e., $\frac{w(1-b)}{\phi}$.

[15] I assume a regular work schedule from Monday to Friday.

$\mathbb{1}\{\text{weekend}\}$ is defined as follows:

$$
\begin{aligned}
\mathbb{1}\{\text{weekend}\} &= 1 \text{ if } dow = \text{Monday and } \; s_c = 5 \\
&= 1 \text{ if } dow = \text{Tuesday and } \; s_c = 4 \\
&= 1 \text{ if } dow = \text{Wednesday and } \; s_c = 3 \\
&= 1 \text{ if } dow = \text{Thursday and } \; s_c = 2 \\
&= 1 \text{ if } dow = \text{Friday and } \; s_c = 1 \\
&= 0 \text{ otherwise}
\end{aligned}
$$

In the model without day of the week, the distinction between s_l and s_c is irrelevant; i.e., $s_l = s_c = s$. With days of the week, the mapping between s_l and s_c depends on the state variable dow. For example, if *dow=Friday*, a sick leave claim of 4 days corresponds to values of s_c and s_l of 4 and 2, respectively. But if *dow=Monday*, a sick leave claim of 4 days is associated with $s_c = 4$ and $s_l = 4$. Thus, the number of business days on leave can be written as a function of *total* days on leave and day of the week $s_l = s_l(s_c; dow)$ (see Appendix Table A1). Workers' payments are computed using the total duration of a sick leave claim and do not take into account the day of the week the claim is filed. Thus, the budget constraint is $c = w[M - s_c + B(s_c)]$ and consumption is a function of s_c, $c = c(s_c)$.

Workers choose sick pay utilization by trading off the cost of a day away from work $w(1 - B'(s_c))$ with its net gain in terms of leisure time. This net gain is a function of day of the week, *dow*, and the duration of the claim. An additional day on leave beyond the worker's sickness level (i) lowers utility increasing the compliance cost term in $\phi 1/\kappa(s_c - \theta)$, (ii) increases utility in ϕ if s_l increases by a unit—i.e., if $\frac{\partial s_l(s_c)}{\partial s_c} = 1$—and (iii) increases utility in q if the sick leave claim ends on a Friday. Thus, the net gain in terms of leisure time is given by the term $\phi \left[\frac{\partial s_l(s_c)}{\partial s_c} - 1/\kappa(s_c - \theta) + q\mathbb{1}\{\text{weekend}\} \right]$. The optimal sick pay utilization is $s_c^*(\theta,dow) = \text{argmax}_{s_c}(u(c(s_c), s_l(s_c), s_c; \theta, dow)$.

The definition of the indicator variable $\mathbb{1}\{\text{weekend}\}$ only considers the extra utility for sick leave claims with a duration of up to 5 days. A more general definition will be to have $\mathbb{1}\{\text{weekend}\}$ equal to one for each sick leave claim that ends on a Friday and assume different values of q for the first and second weekend. I argue that the extra utility from the first weekend is more salient in a worker's decision when filing a sick leave claim. The data presented in Figures 4 and A4 support this assumption.

The model abstracts from the behavior of "when" to file a sick leave claim—that is, in the model, a worker cannot choose the day she files a claim. Nonetheless, the model allows for strategic behavior on the duration margin of a sick leave claim, which is the main focus of this paper. While both margins play a role in the worker's choices, incorporating a filling-day choice in the model would require detailed data on when a worker falls sick and when she files a claim. Absent such data, I assume that workers claim on the day they fall sick.

Coverage choice. A sick pay insurance plan n is characterized by a benefit function $B_n(s)$ that indicates the coverage provided by plan n and the tax the government levies on workers (τ_n).[16] I assume that this tax does not depend on sick leave utilization. Thus, it only affects the choice over plans—since it affects disposable income—but not the worker's utilization choice. I assume that the worker is an expected utility maximizer with a von Neumann Morgenstern (vNM) utility function of the constant relative risk aversion (CRRA) type: $v(y) = y^{1-\gamma}/(1-\gamma)$, where y corresponds to the realized utility $u^*(\theta, dow)$. Before the health state is realized, expected utility is given by

$$U(\theta,dow,\phi,\kappa) = \mathrm{E}\left[v(u^*(\theta,dow,\phi,\kappa))\right] = \int v(u^*(\theta,dow,\phi,\kappa))\; dG(\theta,dow)\ ,$$

where $G(\theta,dow)$ is the joint distribution of health states and days of the week, $u^*(\theta,dow,\phi,\kappa) = u(s_c^*)$, and $c = w(M - s_c^* + B_n(s_c^*)) - \tau_n$. And the optimal coverage choice is $n^* = \arg\max U(\theta,dow,\phi,\kappa)$.

Under these assumptions, risk aversion affects the choice of plans but do not directly affect paid sick leave utilization. That is, more risk-averse workers would choose plans with a higher level of coverage. Nonetheless, *all else equal*, variation in the utilization of paid leave across workers in the same plan reflects variation in their preference parameters (ϕ, κ, and q).

The model also accommodates the traditional source of adverse selection in health insurance: Sicker individuals would choose plans that provide more health insurance coverage. This source of selection is captured by the fact that workers with greater $G(\cdot)$—in a first-order stochastic dominance sense—would purchase more coverage.

II.B Social Planner

In this section I discuss the definition of social welfare and how the social planner chooses the optimal sick pay contract.

Assume that the economy is populated by I workers and let $U_i(\theta_i, dow_i)$ represent the expected utility of worker i when she chooses the sick leave plan $(B_n(s), \tau_n)$. This paper's

[16] The tax τ_n is equivalent to the premium associated with an insurance plan p_n.

main question is what is the plan $(B_n(s), \tau_n)$ that maximizes social welfare. Thus, I assume that the planner offers only one contract and ask what is the level of replacement rates b_s, days-brackets, and tax τ that maximize social welfare. Additionally, I assume that the social planner only observes the duration of a sick leave. The replacement rates therefore depend only on them.

Regarding the design dimension of how many brackets the optimal system features, I restrict the space of contracts to those with up to three brackets. Thus, to derive the optimal contract, I proceed in two steps. In the first step, I consider each possible system, and for each one, I determine the optimal replacement rate.[17] In the second step, I compare their performances. To compare different systems among themselves I impose that all systems must cost at most as much as the current system.

Absent of moral hazard, the optimal sick leave contract would be one featuring full coverage for any duration, i.e., $B_s = b \ \forall \ s$. It would be socially optimal for all workers to be fully insured against health risks since leave's duration would equal their health state. In the presence of moral hazard, the optimal contract would feature some incomplete coverage to deter unjustified leave taking. Absent contagious diseases, the optimal contract trades off the value of risk protection and the social cost of moral hazard. In the presence of contagious diseases, incomplete take up could be detrimental to public health.

In what follows, I explore the more interesting (and more realistic) cases in which workers' behavior exhibit moral hazard, diseases could be contagious, and there is worker' heterogeneity.

The no-externalities case. Assume diseases are not contagious, i.e., a health shock only affects worker i utility. The social planner chooses $B_n(s)$ and τ_n to maximize the sum of individual welfare:

$$\max_{B_n(s),\tau_n} \quad W(\theta) = \sum_i U_i(\theta_i, dow_i) \ \text{ s.t. } \ \sum_i s_i^* B_n(s) \leq S \ ,$$

where I assume that the social planner assigns equal weights to each worker.

III Empirical Setting

In this section, I discuss the Chilean health care and sickness insurance systems, focusing on the institutional features relevant to my analysis. I then present the data and patterns in the data that motivate my modeling choices.

[17] Appendix B.II presents the contract space considered in the optimization procedures in detail.

III.A The Chilean Health Insurance System

In Chile, healthcare insurance providers fulfill two roles: they (i) offer healthcare insurance contracts and (ii) administer the paid sick leave system. The healthcare insurance system is composed of a government-run healthcare insurance provider and a handful of private insurance providers.[18] The government-run healthcare insurance provider offers four plans, whose eligibility is based on monthly salary and household composition. The lowest-tier plan provides coverage for individuals with no income at no cost in public system hospitals. As income increases, workers would qualify for a higher tier plan. This plan provides healthcare coverage in public system hospitals with low copayments and covers healthcare in private healthcare institutions with high copayments.[19] Private insurance companies provide tiered plans with financial vertically differentiated coverage levels—similar to the Gold, Silver, and Bronze plans offered by Affordable Care Act exchanges in the US. The plans offered by private insurance companies allow beneficiaries to obtain healthcare from private healthcare institutions, which provide a higher quality of care than public institutions.

Workers are mandated to purchase health insurance and allocate at least 7% of their salary to a healthcare plan. Workers can choose their health insurance provider. If they select one of the private provides they are free to contribute a higher proportion of their salary to qualify for the healthcare plan of their choice. If they select the government-run healthcare insurance, they would be enrolled in one of the four plans based on their monthly salary and household composition. For example, a single worker who earns USD $693 a month—the median salary in 2017—and chooses the government-run insurance system would be enroll in the highest-tier plan; and could not choose lower tier plans with lower copayments.

In 2017, 73% of workers enrolled in plans offered by the government-run healthcare insurance system; the remaining 27% enrolled in plans offered by one of the private providers (see Panel A of Table A2). Workers enrolled in government-run plans have observable characteristics that would predict that they are more costly to insure; they are older, more likely to be women, and have lower salaries. In contrast, those enrolled in government-run plans are older and their age distribution has a heavier right tail. For example, 31% of workers insured by the private provider are between 25 and 34 years,

[18] These are called FONASA, for the Spanish name—*Fondo Nacional de Salud*—and ISAPRES—*Instituciones de Salud Previsional*—respectively.

[19] Plans are indexed by letters, where A is the lowest-tier plan and D is the highest-tier plan. The highest-tier plan has a 20% copayment in public system hospitals and vouchers to use health care providers who participate in the plan's network at a discounted price.

and only 25% of workers insured by the government belong to this age group.

The second role of healthcare insurance providers is to *administer* the paid sick leave system. That is, insurers are in charge of receiving and screening sick leave claims and disbursing sick leave benefits. Insurers cannot design sick pay plans, and rather have to follow rules set by the central government in terms of eligibility criteria and benefits. Thus, workers do not choose their sick leave plan. This provides a unique setting to study workers' choices over the utilization of sick pay benefits, since it alleviates adverse selection concerns. If workers could choose their sick pay coverage, we could expect sicker individuals to choose greater insurance coverage. Whereas this mechanism could be at play in the choice of healthcare insurance provider, conditional on this decision, sicker and healthier individuals face the same sick pay coverage.

While the eligibility rules and structure of benefits are independent of insurer choice, there are differences in how each provider applies these rules in practice. Panel B of Table A2 shows that the rejection rate by private insurers is almost three times as high as that of the government-run insurer. I consider these differences in leniency to be suggestive evidence that private insurers might have different motives—such as minimizing sick leave payments—when screening sick leave claims. My empirical analysis focuses on workers enrolled in the government-run health insurance system—they represent about 73% of all Chilean workers. The main reason for this choice is that this paper focuses on the provision of paid sick leave as a social insurance system, which is closer to the behavior of the government-run healthcare provider.

III.B The Chilean Paid Sick Leave System

The Chilean paid sick leave system gives employees the right to call in sick and receive sick pay due to short-term, non-work-related sickness—e.g., the common flu or back pain.[20] The eligibility criteria to claim a paid sick leave requires that workers have (i) been enrolled in the social security system for 6 months before the sick leave starting date and (ii) made contributions to the health insurance system for 3 months before the leave starting date.[21] Upon experiencing an impairment to work, workers are required to get a physician's certificate of their sickness, which states the primary diagnosis and the number of days the physician considers the worker will need to recover from the disease. This certificate is reviewed by an insurance office that decides whether the sick leave claim is

[20] The sickness insurance system aims to provide risk protection from impairments to work that are temporary in nature and where full recovery is foreseeable. A separate program provides disability insurance to workers in the case of permanent impairments to work.

[21] These restrictions are independent of job tenure, which reflects that wage replacements are financed with employees' contributions to the healthcare system and not by employers.

(i) approved with no changes, (ii) approved with a reduction in its length, or (iii) denied.

Sick pay is computed as a function of sick leave claim duration subject to a maximum salary. For workers with salary below the maximum, benefits are computed as follows: The benefit scheme exhibits a non-payable period of 3 days, i.e., the replacement rate for the first three days of a sick leave span is zero.[22] This non-payable period works like a deductible that resets for every new sick leave span.[23] Starting on the fourth day, there is full coverage of each missed day—i.e., the replacement rate is one. If the sick leave spans 11 days or more, the non-payable period is reimbursed. That is, claims with an 11-day or longer duration are fully covered.[24] Panel (a) of Figure 1 presents days paid as a function of days on leave for claims of different duration. Reimbursement of the non-payable period after 11 days implies that the average replacement rate jumps discretely at 11 days and it is non-constant (see panel (b) of Figure 1).

Public sector employees—who could be enrolled in either the government-run healthcare insurance system or in a plan provided by a private insurer—are subject to a sick leave regime that provides income replacement at a rate equal to 100% of their wage for the entire duration of the sick leave span. This difference in benefits could imply that workers with preferences for more absences could self-select into jobs in the public sector due to the more generous sick leave provision. In Apendix Table A.10., I compare the observable characteristics of public and private sector workers and provide evidence that public sector workers have higher incidence rates while mean duration of each started spell is smaller. This patter is similar to the one documented by Marie and Castello (2022) in their sample of public and private sector workers in Spain. Given these differences in observable characteristics, the main analysis of this paper focuses on sick leave claims filed by private sector employees.

III.C Data and Descriptive Evidence

I exploit unique administrative data on sick leave claims matched to enrollment data for workers insured by the government-run healthcare system. These restricted-access data were provided directly by the government-run healthcare insurance office and covered the period 2015-2019. The enrollment data include rich demographic information about

[22] Cid (2006) documents that the origin of the 3-day non-payable period dates to a regulation implemented in 1952 that aimed to prevent abusive behavior and it has not been revised since.

[23] These resettable deductibles are similar to those used in automobile or homeowners insurance: Separate deductibles apply to each loss.

[24] If a worker files two (or more) consecutive claims, they are treated as one claim for the computation of benefits. To be consistent, I treat these claims as one claim in the analysis. Appendix table A3 presents counts and summary statistics of sick leave claims and sick leave spells.

every individual enrolled in the public insurance system, including sex, age, ZIP code, health insurance plan, annual earnings, and health indicators for chronic conditions.[25] These data represent the universe of individuals enrolled in the government-run healthcare insurance regardless of whether they have filed a sick leave claim. Additionally, I observe detailed information about the sick leave claim: start and end dates, prescribed days on leave, the primary diagnosis—coded following the 10th revision of the International Classification of Diseases (ICD-10)—physician identifiers, and the amount received for paid sick leave. I merge the sick leave claim data with the enrollment data to construct a claim-level dataset with detailed information on workers' demographic characteristics and leave-taking behavior.

I estimate the model of workers' behavior exploiting the linked claim-level dataset. My primary measures of leave-taking behavior are the duration of a sick leave claim, the day of the week the claim is filed, the disease code, and sick pay corresponding to the claim. I construct these measures at the sick-leave-spell level, i.e., I consider consecutive claims as one claim. Thus, the unit of analysis of the model is the same as the one used to compute sick leave benefits. Appendix table A3 presents counts and summary statistics of sick leave claims and sick leave spells.

To arrive at the analytic sample, I impose two additional restrictions. First, the estimation sample includes claims from private-sector male workers aged 25 to 64, a group with high labor market participation rates. Although women's sick leave-taking behavior is of high interest for the design of sick leave programs, women have much lower participation rates than men. For example, in Chile, women have more than 20 percentage points lower labor force participation rates—52.6% for women at the beginning of the sample period and 73.2% for men. Thus, a model of sick leave-taking behavior that explains women's choices would also require incorporating their decision to participate in the labor market. Nonetheless, abstracting from the participation decision simplifies the model exposition and estimation.

Second, the estimation sample includes claims from a subset of diseases. This paper focuses on non-mental-health sick leave claims; I exclude such claims because their filing process is more cumbersome than the one for non-mental-health claims.[26] A second criterion is the recovery time associated with the primary diagnosis. For example, I exclude

[25]These conditions are cerebral vascular accident, Alzheimer, juvenile arthritis, rheumatoid arthritis, bronchial asthma, lung cancer, diabetes, chronic obstructive pulmonary disease, chronic kidney disease, arterial hypertension, acute myocardial infarction, leukemia, lymphoma, lupus, multiple myeloma, and HIV.

[26]For example, these claims must be certified by a psychiatrist and require a comprehensive medical assessment at the time of filing.

diagnoses under the neoplasms category. Appendix C. describes in greater detail the criteria for selecting diseases to be included in the analysis.

Summary statistics. Table 1 presents summary statistics for all the workers in the sample and for those who have used sick pay during 2017. I split the last group based on the type of disease—included in the analysis or not—and duration of the sick leave claim. Almost 20% of Chilean workers filed a non-mental-health-related sick leave claim in 2017. The average worker in the sample is 44 years old, and the average worker who has used sick pay is about the same age (column 1 vs. column 2 comparison). Nonetheless, the average claimant has a higher salary than the average worker, and this difference is statistically and economically significant.

To better understand the differences between workers who have filed sick leave claims and those who have not, Table 2 presents characteristics of workers who have used sick leave benefits based on the duration of the claims. I group workers that had filed (i) at least one claim with a duration of up to 3 days, (ii) at least one claim with a duration between 4 and 10 days, and (iii) at least one claim with a duration of 11 days or longer. Shorter sick leave claims are associated with younger workers with higher average wages, who are also less likely to have chronic conditions. This pattern is compatible with the 3-day waiting period reducing the likelihood that lower-earning workers file a sick leave claim. Additionally, the association between workers' age and chronic conditions prevalence is consistent with older workers experiencing more severe conditions than their younger counterparts.

Figure 2 shows the distribution of the duration of sick leave claims of up to 29 days. Three main patterns characterize the distribution of days on leave. First, about 26.54% of sick leave claims have a duration of up to 3 days. Sick leave claims lasting between 4 and 10 days explain 41.06% of claims. Second, there is an excess of mass or bunching at 11 days. This coincides with the most significant jump in the average replacement rate, starting at 11 days long claims, workers are fully reimbursed for the time off work. This jump incentivizes workers to extend their leaves to enter the "full" insurance region. At the same time, it provides more generous coverage for more severe health shocks. Third, there is rounding at multiples of 5 and 7 days. This rounding is consistent with physicians being more likely to write recovery times that match with a work-week—five days—or a calendar week. Table XX presents formal tests that confirm the presence of bunching and rounding.

Figure 3 shows the histogram of sick leave claims duration by workers' characteristics. I group workers into eight groups or bins defined based on age and occupation type:

blue-collar and white-collar occupations.[27] Conditional on workers' occupation, older workers require a higher proportion of long sick leave claims. Their distribution of sick leave claims is shifted toward the right relative to the distribution of younger workers (comparison across rows of Figure 3). This pattern is consistent with workers requiring more time to recover from the same conditions as their age and workers suffering more severe underlying conditions. Comparisons across occupations for workers in the same age group indicate that claims from blue-collar workers are longer on average, with a smaller share of claims or up to 3 days. This comparison suggests that differences in the underlying distribution of health could be correlated with occupation type. Motivated on this results, I allow the underlying distribution of health to vary with workers' age and occupation in the estimation of the model.

IV Empirical Model

IV.A Parameterization

Following the model of sick pay utilization described in Section II, a worker is characterized by her wage rate (w_i), her health status (θ_i), and the day of the week she falls sick (dow_i). Given these parameters, worker i chooses to file a sick leave claim of duration s_i to maximize her utility; this choice depends on her preference parameters and the paid sick leave scheme $B(s)$. In this section, I discuss the empirical version of each of these elements and specify what form of heterogeneity I allow across individuals.

Distribution of health states. In the theoretical model, θ_i indicates the number of days a worker is sick and is distributed $G(\theta)$. In the empirical setting, each sick leave claim not only specifies a duration but also a disease. That is, a health state is characterized by a probability of contracting disease d and the duration of the sickness spell θ_i^d. Let p_i^d represent the probability that worker i contracts disease d. For simplicity, I assume that workers can contract only one disease at a time and $\sum_{d=1}^{D} p_i^d = 1$. I assume that this probability varies with individual characteristics and that workers with similar observable characteristics—age and occupation—face the same risk of getting disease d. Thus, for each group of workers or bin b the risk a worker contracts disease d is characterized by the vector $P_b^d = \{p_b^1, p_b^2, \ldots, p_b^D\}$.

[27]Blue-collar worker refers to an individual who performs manual labor. For example, operators, assemblers, and laborers are considered blue-collar workers. White-collar worker refers to an individual who performs professional, desk, managerial or administrative work. For example, sales representatives are considered white-collar workers. Table A9 details the occupations classified as blue-collar and white-collar.

I define workers' groups or bins as an interaction between age group categories—i.e., individuals aged 25 to 34 years old, individuals aged between 35 and 44 years old, and so on—and occupation categories—white- and blue-collar workers. I provide more details on the rational for these categories in the estimation section.

The duration of a sickness spell can be written as

$$\theta_i^d = \overline{\theta_b^d} + \varepsilon_i^d \ ,$$

where $\overline{\theta_b^d}$ is a disease-specific component that captures the average number of days a worker in group b needs to recover from disease d. That is, I assume this average varies with individuals' age and occupation. Variation across age captures the fact that younger (older) workers might need less (more) time to recover from the same disease. Variation across occupation reflects the physical-work requirements of different occupations. Thus, for each bin b the duration of a sickness spell is characterized by the vector of average recovery times $\Theta_b = \{\overline{\theta_b^1}, \overline{\theta_b^2}, \ldots, \overline{\theta_b^D}\}$.

The individual-specific component ε_i^d captures the idea that, given her age and occupation, individual i could get a particular bad (mild) realization of disease d. I assume that ε_i is a mean-zero measurement error with variance given by σ_ε to be estimated.

Additionally, I assume that health states and days of the week are distributed independently: $G(\theta{,}dow) = G(\theta)F(dow)$, where $F(dow)$ is the distribution of days of the week and $G(\theta)$ is the distribution of the duration of sickness states *unconditional* on the disease the worker contracts:

$$g(\theta) = P(\theta_i = x) = \sum_{d=1}^{D} p_i^d \ P(\theta_i^d = x | disease_i = d),$$

where p_i^d is the probability that worker i contracts disease d as previously defined, and $P(\theta_i^d = x | disease = d)$ is the probability that the sickness spell has a duration x conditional on disease d. Workers in the same bin face the same sickness duration distribution. Thus, for each bin b the vector of unconditional probabilities is given by $P_b = \{p_{b,0}, p_{b,1}, \ldots, p_{b,M}\}$ where M is the maximum number of days a worker could sick in a month.

Days of the week. I assume that individuals face a risk of falling sick on day dow that is uniform across business days. That is, the state variable that represents days of the week dow is a random variable of the discrete type that can take five values $dow = \{Monday, \ldots, Friday\}$, and the probability of any of these values is the same. This implies that $P(dow = day) = 0.20$ for each day of the week.

Wage process. I model the wage process semi-parametrically. I assume the worker's wage rate w_i follows a lognormal distribution with mean $\mu_{w,b}$ and variance $\sigma^2_{w,b}$ which vary by worker's group. That is, I assume that the wage rate varies across workers and wages follow the same distribution for workers in the same bin b. The wage process is given by

$$\log(w_i) \sim N(\mu_{w,b},\sigma^2_{w,b}) \; .$$

Preference parameters. Worker i derives the following utility from her sick leave utilization choice s_c given the health state realization θ_i:

$$\begin{aligned} u(c,s_l,s_c;\theta,dow) = w_i\,(M - s_c + B(s_c)) + \phi_i\Big[(s_l - \theta) \\ - \frac{1}{2\kappa_i}(s_c-\theta)^{\eta_r}\mathbb{1}\{s_c > \theta\} - \frac{1}{2\kappa_i}(\theta - s_c)^{\eta_l}\mathbb{1}\{s_c < \theta\} + q\;\mathbb{1}\{\text{weekend}\}\Big] \; , \end{aligned}$$

where ϕ_i and κ_i govern the valuation of time outside work and compliance cost and, ultimately, the degree to which optimal utilization varies across workers. This function represents a more flexible parameterization of the costs of deviating from the true type, that is it allows for asymmetric penalties for positive and negative deviations with curvature parameters η_r and η_l to be estimated.

I assume that $\log(\phi_i)$ is drawn from a normal distribution with mean μ_ϕ and variance σ^2_ϕ such that

$$\log(\phi_i) \sim N(\mu_\phi,\sigma^2_\phi) \; .$$

That is, I allow for heterogeneity in the valuation of time outside work, but I assume that each worker draws a realization from the same distribution. Heterogeneity across this dimension reflects different tastes for leisure relative to consumption or, put another way, different opportunity costs of missing work to recover from a disease.

Variation in compliance cost parameter κ_i reflects variation in the worker's preferences over behaving as expected or revealing their "true" health status. Additionally, job characteristics can justify variation in κ. For example, if a worker's job can easily be performed by a coworker, workers might face high compliance costs (low κ) and ask for time outside work that closely follows their health status. I capture both of these mechanisms in a reduced-form manner. I assume that $\log(\kappa_i)$ follows a normal distribution with mean and variance μ_κ and σ^2_κ:

$$\log(\kappa_i) \sim N(\mu_\kappa, \sigma^2_\kappa) \ .$$

The term $\phi_i q$ captures the extra utility a worker derives when the sick leave claim has a duration that allows the worker to not return to work until after the weekend. I assume that all of the variation in this term is governed by the parameter ϕ_i; thus q does not vary across workers and is constant across sick leave duration. The first assumption is justified by the fact that the valuation of time outside work is already captured by the parameter ϕ_i. The second assumption implies that a sick leave claim of 5 days on Monday provides the same extra utility as a sick leave claim of 2 days on Thursday. However, the utility level associated with each claim will differ because of the value of consumption and time are functions of duration. Thus, while having a constant q could seem restrictive, it keeps the model simple—and variation coming from the duration of a claim is already accounted for.

Worker i derives the following expected utility from her utilization choice s_c:

$$U_i = \sum_x \sum_{day} \frac{u^*(s_c(\theta_i, dow_i))^{(1-\gamma)}}{1-\gamma} P(\theta_i = x) P(dow_i = day) \ ,$$

where γ is the constant relative risk-aversion coefficient, and the empirical analogue of $G(\theta, dow)$ uses independence between the health state distribution and days of the week distribution. In the main version of the model, I assume that individuals do not vary in their degree of risk aversion. In extensions of the model, I allow for preference heterogeneity.

Measurement error and rounding. To better explain the main patterns of the data I include two additional mechanisms in the estimation of the model. These mechanisms aim to capture the behavior of physicians who prescribe sick leave claims in a reduced-form way.[28] First, I allow the duration of sick leave claims assigned to a worker to differ from the one optimally chosen by the worker. This discrepancy allows the model to accommodate (i) informational frictions between a worker and a physician and (ii) observed sick leaves with a combination of duration and day of the week that is not predicted by the model. I assume that the duration of sick leave claims is measured with and additive error, that has mean zero and is uncorrelated with the "true" sickness level. That is, I

[28] While explicitly modeling physician behavior is relevant for the design of paid sick leave, the lack of available data on physicians characteristics limits the availability to address the question empirically.

assume that given the optimal sick leave duration s^*, the physician prescribes $\tilde{s}$:

$$\tilde{s} = s^* + \delta \ ,$$

where δ is a mean-zero random variable with support $[-3,3]$. With probability p_0, δ takes the value zero, with probability $\frac{p_1}{2}$ it takes the value one, with probability $\frac{p_1}{2}$ it takes the value negative one. Probabilities p_2 and p_3 are defined analogously.

Second, I adjust sick leave duration to take into account the rounding or heaping observed in the data. I interpret this pattern as coming from physicians being more likely prescribe rest that are multiples of five or seven days. I take a non-parametric approach to implement this adjustment and simply assume that with some probability p_{r_5} and p_{r_7} a sick leave claim of duration x is rounded up (dow) to 5 (or 7) days. I assume that these procedures are independent, and to estimate the model I randomly select which adjustment to implement first.

IV.B Estimation and Identification

The main empirical challenge is to disentangle the underlying distribution of health states from workers' preferences (time-valuation and compliance costs). To address this challenge, I exploit (i) detailed data on health states, to construct and underlying distribution of health and (ii) variation in the temptation for shirking behavior coming from the day of the week a sick leave claim is filed, to identify workers' preference parameters. In this section, I discuss the identification and estimation of the empirical model.

Distribution of health states. P_b^d,Θ_b,P_b, σ_ε. To estimate the elements of the vector P_b^d I group diseases in 22 categories, with each representing at least 2% of the sick leave claims; this implies that $D = 22$. Let j index sick leave claims, the estimated probability that individual i in bin b contracts disease d is given by

$$\widehat{p_b}^d = \frac{\sum_j \mathbb{1}\{\text{ main diagnosis}_j = d, i \in b\}}{\sum_j \sum_d \mathbb{1}\{\text{ main diagnosis}_j = d, i \in b\}} \ ,$$

where the denominator counts the number of sick leave claims filed by individuals in bin b for any disease. I estimate the vector P_b^d using sick leave claims data. The advantage of this approach is that by exploiting the details of the sick leave claim data I (i) estimate age-occupation-disease-specific probabilities (ii) avoid imposing parametric assumptions on the distribution of health states. Table 3 presents the estimation of these probabilities and Figure A3 shows them graphically.

A limitation of this approach is that the claims data reflect not only the probabilities of contracting a disease (the "true" p^ds) but also the behavior of insured workers and their physicians. Thus, if workers fake how sick they are or their symptoms and this is reflected in physicians' diagnoses, $\widehat{p}_b^{\,d}$ will be a biased estimator of p_b^d. Note that for $\widehat{p}_b^{\,d}$ to be biased, it should be the case that physicians diagnoses' differ from the true underlying condition—regardless of the number of days assigned for a given diagnoses—since the estimator of p_b^d only uses the diagnostic information and not the duration of a sick leave claim. The direction of the bias depends on the nature of moral hazard behaviors. If workers tend to overstate their sickness in a way that leads physicians to diagnose a more severe condition, I would overestimate the probability of these severe conditions and underestimate the probability of milder conditions. I would also overestimate the probability of the occurrence of conditions that are easier to fake, regardless of their severity.

To estimate the elements of the vector Θ_b I exploit the Peruvian Handbook of Recovery Times. The handbook states the average number of days a worker would require to recover from disease d and proposes adjustments to those days based on workers' age and occupation. Table A7 outlines examples of these calculations. The advantage of exploiting this external data source to estimate the average recovery time is that these averages are not subject to the concerns discussed previously regarding use of the claims data. Table A8 presents the average number of days for each group of workers. The average recovery time increases with age. Across occupations, blue collar workers require longer recovery periods than white collar workers.

I identify the variance in the individual-specific component of duration σ_ε by exploiting the structure of the model and estimate it jointly with the preference parameters; thus, I discuss their identification and estimation jointly with the preference parameters.

I estimate the unconditional distribution of the duration of a health shock as follows:

$$\widehat{p}_{x,b} = \frac{\sum_i \mathbb{1}\{\ \theta_i = x, i \in b\}}{\sum_x \sum_i \mathbb{1}\{\ \theta_i = x, i \in b\}}\ ,$$

where the numerator is the frequency that individual i falls sick for x days and the denominator is the summation over all possible duration of a sick leave spell. Estimation of these probabilities relies on estimation of the conditional probabilities p_b^d and the term σ_ε. With estimations of the vector P_b^d and σ_ε, I use Monte Carlo simulations to construct the vector P for each group of workers b.

Wage process parameters. $\mu_{w,b}$, $\sigma^2_{w,b}$ The parameters of the wage process are estimated outside the structure of the model. For each bin, I estimate the mean and standard deviation of log wages. Estimates of the parameters $\mu_{w,b}$ and $\sigma^2_{w,b}$ are presented in Table 4. Wages of workers on white collar occupations are on average 19% greater than those in blue collar occupations. Wages increase with workers' age between the first and third bins and show a small decrease for the last bin.

Relative risk aversion coefficient γ. The theoretical model of worker's behavior assumes that risk aversion only affects workers' choices over sick pay plans, not their utilization choice. Thus, identification of γ requires variation in plan choices across workers. Nonetheless, the Chilean Paid Sick Leave System does not offer choice over sick pay plans. Absent this variation, I calibrate γ using results from the literature. A concern in calibrating γ is that risk aversion can vary significantly across the scale of shocks (Chetty and Szeidl, 2007). In the model, γ captures risk aversion relative to wage losses from short-term, non-work-related health shocks. The losses associated with such shocks on average are thought to be smaller than losses faced in the case of unemployment shocks. I assume that $\gamma = 2$ and present results with two alternative specifications that allow for preference heterogeneity.[29]

Structurally estimated parameters. $q, \mu_\phi, \sigma^2_\phi, \mu_\kappa, \sigma^2_\kappa, \eta_r, \eta_l, p_1, p_2, p_3, p_{r_5}, p_{r_7}, \sigma^2_\varepsilon$ I estimate the remaining preference parameters, the parameters that govern measurement error and rounding, and the variance of health shocks structurally using the simulated method of moments (SMM). For this estimation, I select informative moments from the sick leave claims data. Even though the parameters are jointly estimated, I provide a heuristic discussion of the most relevant moment for each parameter.

Weekend-streak utility (q). The term $\phi q 1\{weekend\}$ captures the extra utility a worker derives when the interaction of sick leave claim duration and day of the week implies a streak of days off work that includes the weekend, which I term a weekend-streak combination. To identify q I exploit variation across days of the week on which a sick leave claim of duration s is filed. That is, I rely on the fact that the temptation for shirking behavior varies between days of the week. For example, a 2-days-long sick leave claims is more attractive on a Thursday than a 1-day-long claim, since the later entails one rest day while the former entails four days of rest. Figure 4 illustrates this variation. For each day of the week, I compute the share of sick leave claims, indexed by j, of duration s that are

[29] I follow Herbst and Hendren (2021) and assume that the coefficient of relative risk aversion, γ, is drawn from a uniform distribution. I consider two alternative specifications (i) a uniform distribution between 1 and 3 and (ii) a uniform distribution between 0 and 4. Both specifications retain the mean risk aversion coefficient at 2, but introduce different degrees of heterogeneity.

filed that day.[30] That is:

$$\text{share}_s^{day} = \frac{\sum_j \mathbb{1}\{dow_j = day, s_j = s\}}{\sum_j \mathbb{1}\{dow_j = day\}} .$$

where j indexes sick leave claims, s represents the duration of a claim, and day represents the day of the week. Consider 1-day-long sick leave claims; the share of claims filed on Friday is about three times higher than the share of claims filed on any other day of the week (see Panel (a) of Figure 4). This pattern is present for claims that are between 1 and 5 days long. Crucially, when inspecting 7-day-long claims, the share is constant across days of the week.[31]

I identify q using the difference between the share of 1-to-5-day-long sick leave claims filed on a weekend-streak day and the share of 1-to-5-day-long claims filed any other day of the week. That is, I pooled all the weekend-streak combinations to compute the average share of claims on those days and compare it with the average share of claims during the rest of the week. If a larger difference is observed, the model requires a higher q to rationalize the data. This comparison relies on the idea that, the share of sick leave claims of duration s on a non-weekend-streak day is a good counterfactual to estimate the effect of filling a sick leave claim of duration s on a weekend-streak day. The last panel of Figure 4 shows this moment graphically and Table 5 presents detailed computations.

Compliance costs ($\mu_\kappa, \sigma^2_\kappa, \eta_r, \eta_l$): The terms $\frac{1}{2\kappa}(s_c - \theta)^n_r$ and $\frac{1}{2\kappa}(s_c - \theta)^n_l$ captures how costly is for a worker to be on leave for a longer or shorter duration than her health state, respectively. To identify $1/\kappa$ I exploit variation across days of the week and sick leave claims duration conditional on workers' health. For each day of the week and health state, I compute the share of sick leave claims, indexed by j, of duration s filed by workers with health state θ:

$$\text{share}_{s,\theta}^{day} = \frac{\sum_j \mathbb{1}\{dow_j = day, s_j = s, \theta_j = x\}}{\sum_j \mathbb{1}\{dow_j = day, \theta_j = x\}} ,$$

where the denominator counts the number of sick leave claims filed on day of the week day with primary diagnoses that would require x days on leave and the numerator counts how many of these claims have duration s. For example, the share of workers with a 1-

[30]This computation excludes sick leave claims filed in a week of a national holiday.

[31]Claims of duration longer than 6 days exhibit a similar pattern. I use 7 days as a reference point since the share of these claims in the data is greater than the share of 6-days-long claims. Appendix FigureA4 presents the distribution of share of sick leave claims by day of the week for claims with duration between 8 and 15 days, pooled in 2-days groups.

day-long health shock on a Friday who ask for a one-day-long leave, i.e., is given by

$$\text{share}_{1,1}^{Friday} = \frac{\sum_j 1\{dow_j = Friday, s_j = 1, \theta_j = 1\}}{\sum_j 1\{dow_j = Friday, \theta_j = 1\}} \ .$$

Figure 5 illustrates this computation for sick leave claims with a health shock that requires 1-day-long recovery. The first panel shows that, conditional on the health state, the share of one-day-long sick leave claims on Fridays is more than twice as high as the share any other day of the week. Panel (b) fixes the health state and day of the week and looks at 2-days-long sick leave claims, or a one-day deviation from the "true" health state. Similarly, by changing the duration, I can trace the share of 2-days, 3-days and 4-days-long deviations. Panel (f) summarizes the probabilities of not-asking for extra days on leave, asking for one extra day on leave, two extra days on leave and up to four extra days on leave conditional on filing a sick leave claim on a weekend-streak day.

Identification of the distribution of compliance costs parameters relies on the comparison of the average share of claims with no deviations, 1-day deviations, 2-day deviations, and up to 4-day deviations. These shares are presented in Panel (a) of Figure 6. First, to compute the average share of claims with a given deviation, I average across claims from health states requiring up to 3 days on leave (see Figures A5 and A6). Second, I compute the difference between the share of sick leave claims with deviation dev relative to the no-deviation share. The bigger this difference, the more costly it is to deviate from the *true* health state. Thus, these differences are informative of the distribution of compliance costs. The sharper the decrease in these shares, the lower the curvature in the cost function that rationalizes the data. Panel (b) of Figure 6 shows this computation. The pattern in the data suggest that a one day long deviation is not too costly relative to truth-telling, two days deviations are more costly as reflected by the lower share of sick leave claims in the third column of this graph.

Value of leisure (μ_ϕ, σ^2_ϕ): The parameter ϕ captures the taste for leisure relative to the taste for consumption. It can therefore be identified by the average ratio of leisure to consumption. I leverage data on wages, duration of sick leave claims, and sick pay to compute this ratio. I compute consumption as the net earnings in a month using data on wages and sick pay, that is, which is the consumption measure implied by the model. To compute leisure, I use the number of days a worker is on leave. For worker i this ratio is computed as follows:

$$LC_i = \frac{\text{leisure}}{\text{consumption}}_i = \frac{1}{N_i} \sum_m \frac{w_{i,m} \ \times \ Days\ on\ leave_{i,m}}{w_{i,m} \ \times Days\ worked_{i,m} + \ Sick\ pay_{i,m}} \ ,$$

where m indexes month of the year and N_i is the number of months in the year that worker i has used at least one sick leave claim. The numerator estimates worker i valuation of leisure in month m and the denominator estimates her consumption in month m. Thus, the ratio LC_i is the average relative valuation of leisure for individual i. Figure 7 shows the distribution of LC_i, the mean and standard deviation of this distribution inform the distribution of ϕ, which I assume log-normal with mean μ_ϕ and standard deviation σ_ϕ.

Measurement error (p_1,p_2,p_3): I use the difference between the share of 3-days-long sick leave claims filed on non-weekend-streak days and on Wednesday, conditional on health shocks with a 1-day recovery, to pin down the distribution of the measurement error term δ (see panel (c) of Figure 6). Given the share of claims filed on Wednesday, a smaller difference implies that more sick leave claims have been moved away from the most-profitable duration. That is, the smaller the difference, the more likely the observed duration is not the optimally one in terms of workers' utility.

Rounding (p_{r_5},p_{r_7}): The distribution of sick leave claims shows excess mass around sick leave claims with duration that are multiples of 5 and 7. The empirical model allows sick leave claims to be rounded up (or down) to such duration. To inform these probabilities I construct a simple measure of heaping on 5 and 7 days as proposed by Roberts and Brewer (2001) :

$$h_z = f(z) - \frac{f(z-1) + f(z+1)}{2} \ ,$$

where z corresponds to 5 and 7, and $f(\cdot)$ indicates the frequency of sick leave claims with duration z. Thus, h_z gives the difference between a duration frequency and the average of the frequencies of the two immediately neighboring duration. It indicates how much a duration 'sticks out' from the pattern suggested by its neighbors.

V Results

V.A Parameter Estimates

The model predicts the main patterns of sick leave-taking behavior quite well. Table 6 summarizes the targeted moments from the data. I use the spikes in the share of claims filed on weekend-streak days relative to non-weekend-streak days to identify the parameter governing the utility that workers derive from sick leave claims that end on a Friday, i.e., q. There are, on average, 12.33% more sick leave claims on weekend-streak days. My estimates follow a similar pattern: on average, I estimate 14.64% more sick leave claims on a weekend-streak day relative to non-weekend-streak days. Figure 8 shows the simulated shares of sick leave claims by day of the week and duration. It reproduces Figure 4 using a model-simulated sample. Even if I only target the difference in the average shares, the model provides a good fit for the patterns observed by duration and days of the week.

Conditional on their health shocks, I exploit the share of sick leave claims observed on weekend-streak days to identify the parameters of the compliance cost function. Specifically, I exploit the differences relative to the no-deviation case to identify the curvature parameters of this function. Figure 9 compares targeted and non-targeted moments from the data and a model-simulated sample. The model matches the distribution of compliance costs—i.e., the cost of reporting the *true* health shock—reasonable well with $\mu_\kappa = 2.59$ and $\eta_r = 2.03$. I test the fit of the model in two ways. First, the model predicts that a one-day deviation is not too costly relative to truth-telling and that a two-day deviation is more costly, as reflected by the lower share of sick leave claims in this group. In particular, the model matches the difference between the shares quite well (see panel b).

Second, the model predicts that if a worker realizes a health state just under 11 days on leave, she will take advantage of the proximity to the full-coverage region and fake her type to gain full coverage, which is consistent with the patterns I document from the data. Figure A8 compares the distribution of days on leave from the data and a model simulated sample. The model captures the main pattern observed in the data: sick leave claims spike at 11 days, with lower mass at 8, 9, and 10 days. Using the measure of heaping proposed in the previous section, I estimate that, in the data, the 11-day duration accumulates an additional 4.04% mass than its neighbors. Using a model-simulated sample, I estimate an additional 4.35% mass at this point relative to its neighbors.

I use the distribution of the ratio of leisure to consumption to identify the parameter ϕ. I estimate that, on average, the ratio of wages to the value of leisure is 0.7120, implying that workers value time off work to recover from disease about 40% more than their wages. To put this estimate into context, consider that 26.54% of sick leave claims involve non-paid time off, and a total of 67.60% of claims involve partial paid for workers—, i.e., 67.60% of claims have a duration of up to 10 days for which the replacement rate is less than one.

Sick leave-taking behavior. What do the estimated preference parameters imply for workers' sick leave-taking behavior? I propose two exercises to answer this question. First, to facilitate intuition, I consider the case of a linear benefit scheme $B(s) = bs$ where $b \in [0,1]$. I compute the optimal sick leave duration, i.e., the duration that would maximize the worker's utility given her health state under alternative values of the replacement rate and preference parameters.[32] I present these results in Figure 10 and Table A10. Increasing the replacement rate from $b = 0$ to $b = 1$, that is, moving the average worker from a non-paid- to paid-sick leave contract with full-coverage, induces sick pay utilization in 1.74 days, conditional on no-weekend effect (i.e., assuming that the sick leave claim does not end on a Friday). To illustrate the role of compliance costs, I consider the exercise of moving a worker with high compliance costs parameter—i.e., a worker in the first quartile of the κ distribution—from a situation of no coverage to full coverage. The optimal sick leave utilization increases by one day. In contrast, when compliance costs are low, i.e., a worker in the top quartile of the κ distribution, sick leave utilization increases by 2.14 days. Relative value of time also plays a key role to explain sick leave-taking behavior. Moving a worker with high valuation of time outside work relative to consumption from the least to the most generous linear contract increases sick leave utilization in 0.70 days. This response is almost four times higher for a worker with low valuation of time: this worker is more responsive to the monetary incentives given by a higher replacement rate while workers with high value of time are less responsive to these incentives.

The second exercise uses the estimated parameters and assuming that the benefit scheme follows the Chilean contract—presented in Figure 1—constructs the demand for days on leave as a function of workers' health state. These results are presented in panel (b) of Figure 10. Two main patterns emerge from this figure (i) workers respond to the discontinuities in the replacement rate by extending sick leave claims, especially in for low-coverage duration, e.g. up to 5 days; (ii) the model captures the bunching at 11 days and predicts that workers with lower compliance costs are more likely to bunch than those with higher compliance costs. Put another way, workers with lower compliance

[32] Appendix B.I presents the derivation of the optimal utilization under the assumption of linear contracts and the compliance cost function used in the estimation. This expression is analogous to the one presented in the theoretical section but allows for days of the week (s_l vs. s_c), weekend-streak utility, and a flexible compliance costs functional form.

costs are more responsive to the monetary incentives than their counterparts.

VI Counterfactual Sick Leave Schemes

VII Conclusions

References

Balance, A Better. 2021. "Overview of Paid Sick Time Laws in the United States." *retrieve August 30, 2021.*

Cabral, Marika, and Marcus Dillender. 2020. "The Impact of Benefit Generosity on Workers' Compensation Claims: Evidence and Implications." National Bureau of Economic Research.

Card, David, and Brian P McCall. 1996. "Is Workers' Compensation Covering Uninsured Medical Costs? Evidence from the "Monday Effect"." *ILR Review*, 49(4): 690–706.

Chetty, Raj. 2008. "Moral Hazard Versus Liquidity and Optimal Unemployment Insurance." *Journal of Political Economy*, 116(2): 173–234.

Chetty, Raj, and Adam Szeidl. 2007. "Consumption Commitments and Risk Preferences." *The Quarterly Journal of Economics*, 122(2): 831–877.

Cid, Camilo P. 2006. "Evaluación del Subsidio por Incapacidad Laboral por Enfermedad Común en Chile: 1990-2000." *Superintendencia de Salud. Documento de Trabajo*.

Cronin, Christopher J, Matthew C Harris, and Nicolas R Ziebarth. 2022. "The Anatomy of US Sick Leave Schemes: Evidence from Public School Teachers." NBER Working Paper No. 29956.

Cutler, David M, and Richard J Zeckhauser. 2000. "The Anatomy of Health Insurance." In *Handbook of Health Economics*. Vol. 1, 563–643. Elsevier.

De Paola, Maria, Vincenzo Scoppa, and Valeria Pupo. 2014. "Absenteeism in the Italian public sector: The effects of changes in sick leave policy." *Journal of Labor Economics*, 32(2): 337–360.

Einav, Liran, Amy Finkelstein, and Mark R Cullen. 2010. "Estimating Welfare in Insurance Markets Using Variation in Prices." *The Quarterly Journal of Economics*, 125(3): 877–921.

Einav, Liran, Amy Finkelstein, Stephen P Ryan, Paul Schrimpf, and Mark R Cullen. 2013. "Selection on Moral Hazard in Health Insurance." *American Economic Review*, 103(1): 178–219.

Gruber, Jonathan. 2000. "Disability Insurance Benefits and Labor Supply." *Journal of Political Economy*, 108(6): 1162–1183.

Handel, Ben, Igal Hendel, and Michael D Whinston. 2015. "Equilibria in health exchanges: Adverse selection versus reclassification risk." *Econometrica*, 83(4): 1261–1313.

Hendren, Nathaniel. 2017. "Knowledge of Future Job Loss and Implications for Unemployment Insurance." *American Economic Review*, 107(7): 1778–1823.

Herbst, Daniel, and Nathaniel Hendren. 2021. "Opportunity Unraveled: Private Information and the Missing Markets for Financing Human Capital." National Bureau of Economic Research.

Hopenhayn, Hugo A, and Juan Pablo Nicolini. 1997. "Optimal Unemployment Insurance." *Journal of Political Economy*, 105(2): 412–438.

Johansson, Per, and Mårten Palme. 2005. "Moral Hazard and Sickness Insurance." *Journal of Public Economics*, 89(9-10): 1879–1890.

Low, Hamish, and Luigi Pistaferri. 2015. "Disability Insurance and the Dynamics of the Incentive Insurance Trade-Off." *American Economic Review*, 105(10): 2986–3029.

Maclean, Johanna Catherine, Stefan Pichler, and Nicolas R Ziebarth. 2020. "Mandated Sick Pay: Coverage, Utilization, and Welfare Effects." National Bureau of Economic Research.

Marie, Olivier, and Judit Vall Castello. 2022. "Sick Leave Cuts and (Unhealthy) Returns to Work." *accepted at Journal of Labor Economics*.

Marone, Victoria R, and Adrienne Sabety. 2022. "When Should There Be Vertical Choice in Health Insurance Markets?" *American Economic Review*, 112(1): 304–42.

Pichler, Stefan, and Nicolas R Ziebarth. 2017. "The pros and cons of sick pay schemes: Testing for contagious presenteeism and noncontagious absenteeism behavior." *Journal of Public Economics*, 156: 14–33.

Pichler, Stefan, Katherine Wen, and Nicolas R Ziebarth. 2021. "Positive Health Externalities of Mandating Paid Sick Leave." *Journal of Policy Analysis and Management*, 40(3): 715–743.

Pollak, Catherine. 2017. "The Impact of A Sick Pay Waiting Period on Sick Leave Patterns." *The European Journal of Health Economics*, 18(1): 13–31.

Powell, David, and Seth Seabury. 2018. "Medical Care Spending and Labor Market Outcomes: Evidence from Workers' Compensation Reforms." *American Economic Review*, 108(10): 2995–3027.

Roberts, John M, and Devon D Brewer. 2001. "Measures and tests of heaping in discrete quantitative distributions." *Journal of Applied Statistics*, 28(7): 887–896.

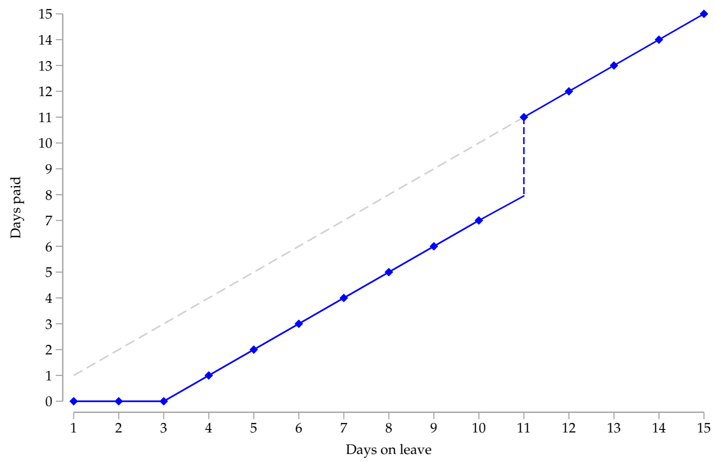
Ziebarth, Nicolas R. 2013. "Long-term Absenteeism and Moral Hazard - Evidence From A Natural Experiment." *Labour Economics*, 24: 277–292.

Ziebarth, Nicolas R, and Martin Karlsson. 2014. "The Effects of Expanding the Generosity of the Statutory Sickness Insurance System." *Journal of Applied Econometrics*, 29(2): 208–230.

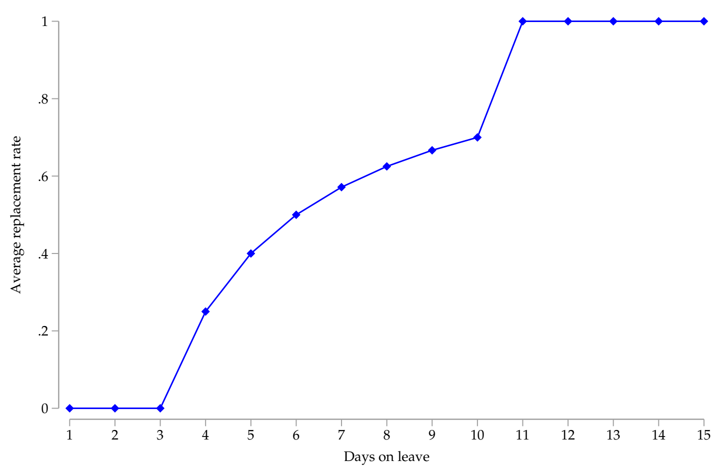
VIII Figures

Figure 1: Chilean Paid Sick Leave System

(a) Days paid as a function of days on leave

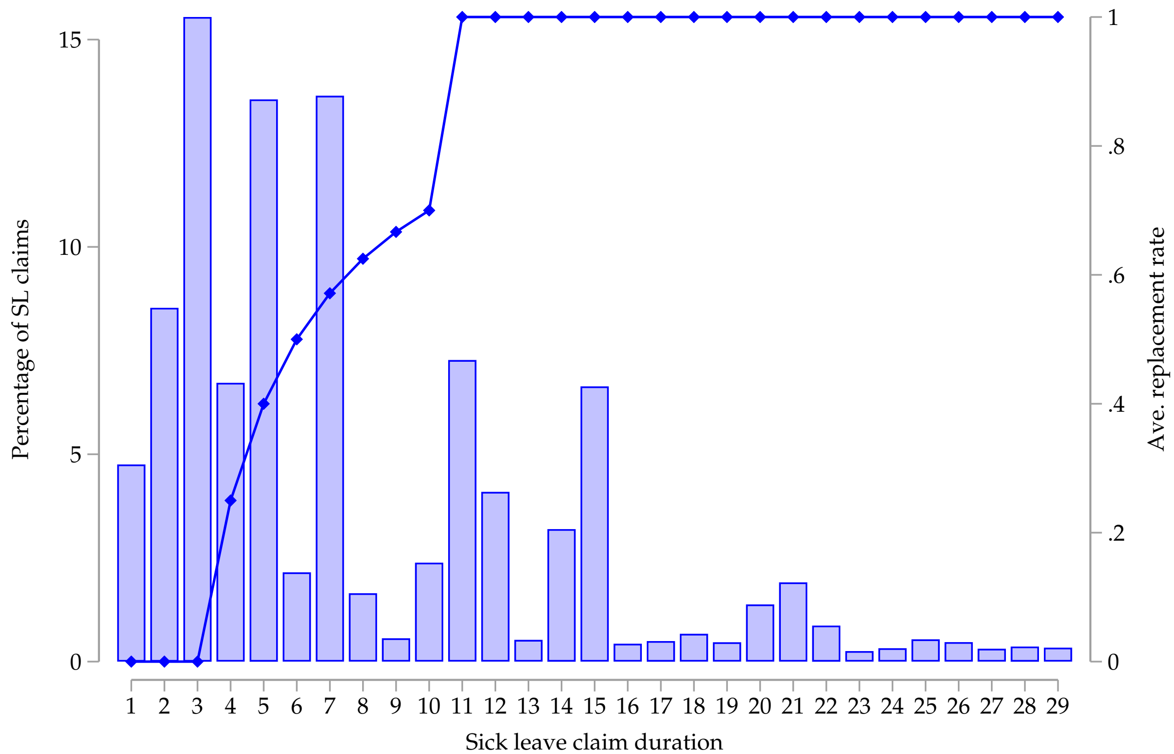


(b) Average replacement rate



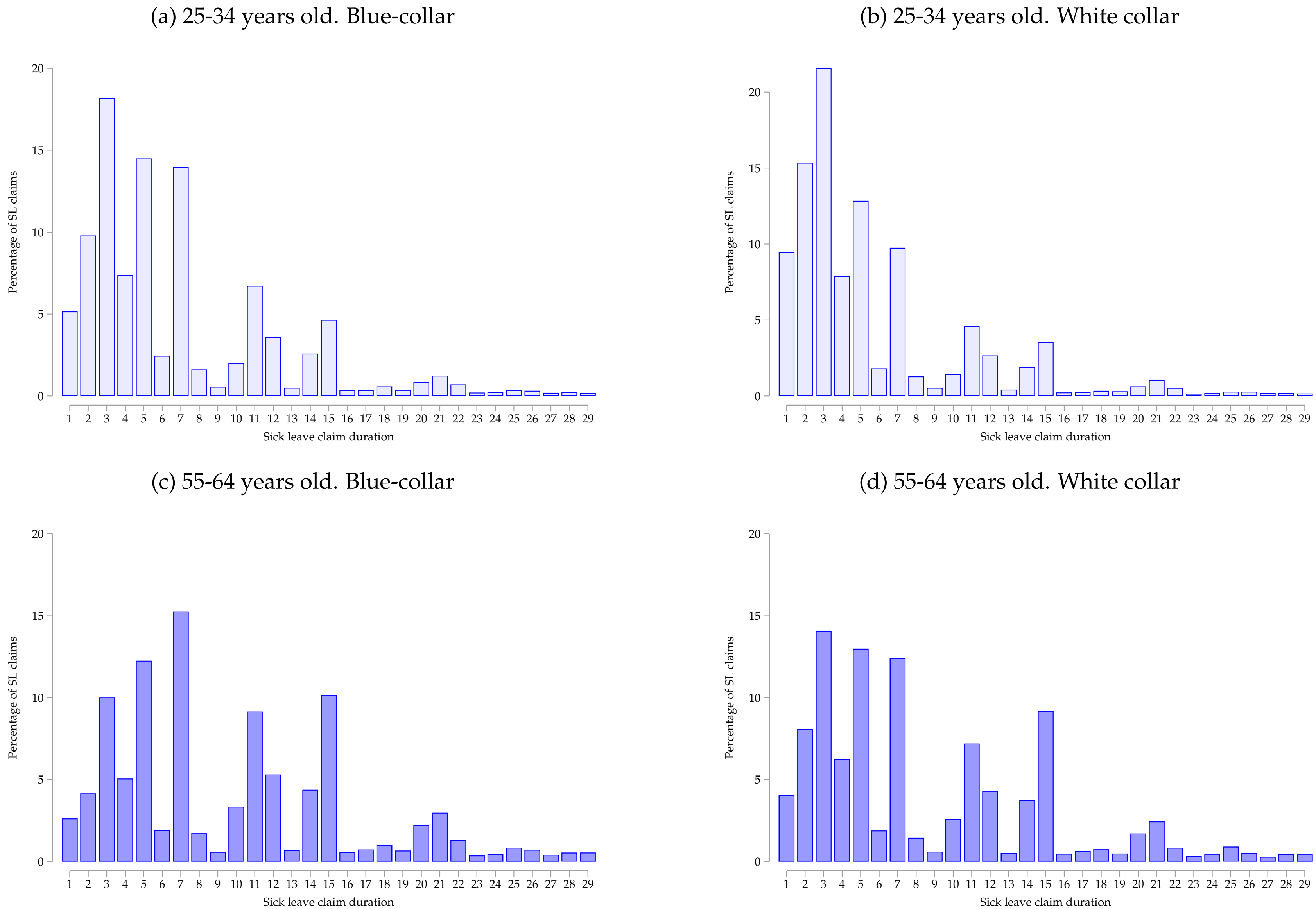
Notes: This figure shows the paid sick leave benefit scheme for private-sector employees. Panel (a) shows the number of days paid as a function of days on leave. Panel (b) shows the average replacement rate, i.e., the ratio between the number of days paid and the number of days on leave. This figure is referenced in Section III.B.

Figure 2: Duration of sick leave claims. Private-sector male workers



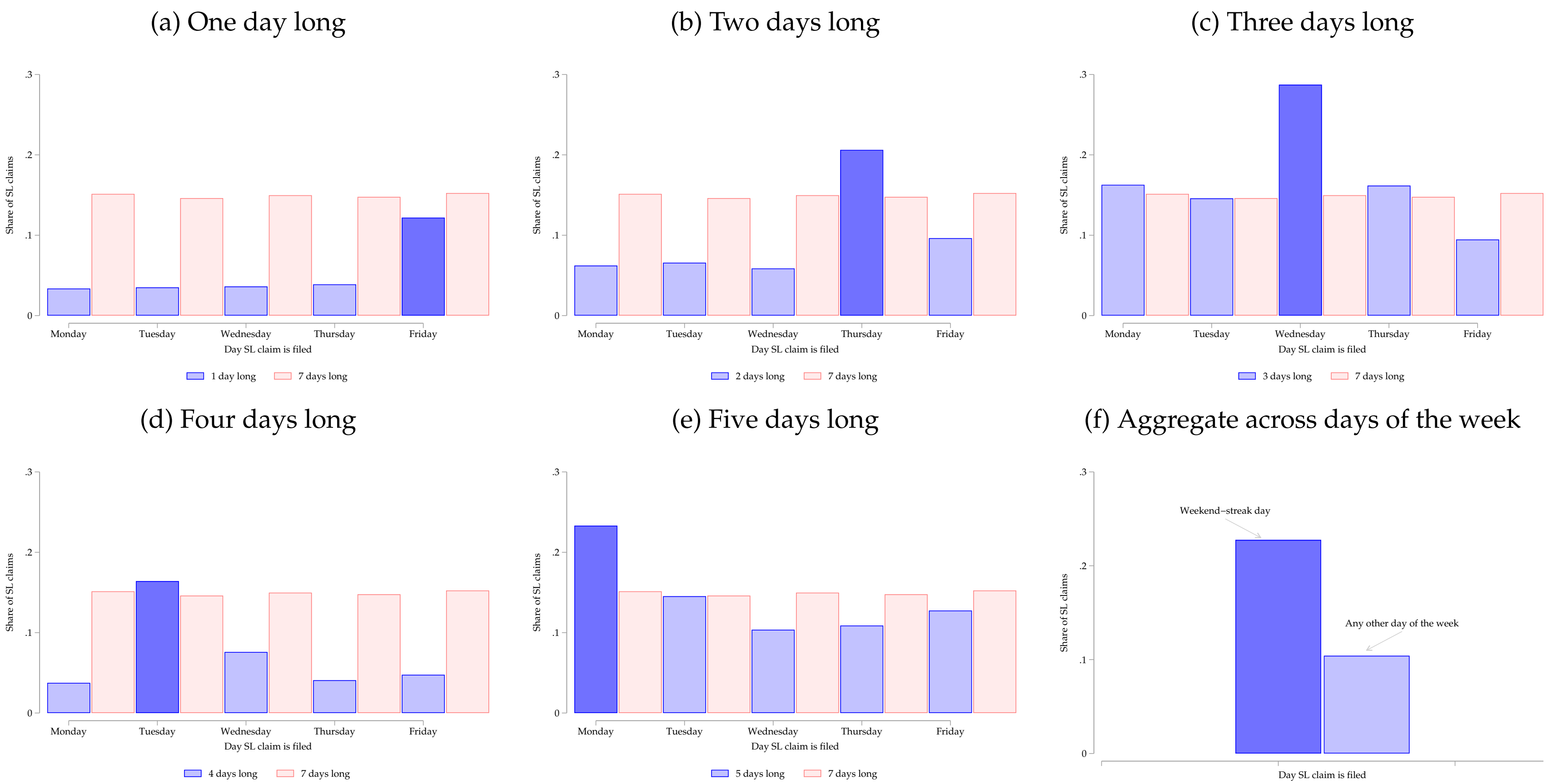
Notes: This figure shows the distribution of the duration of sick leave claims from male workers on the left-hand-side vertical axis and the average replacement rate on the right-hand-side vertical axis. The figure only includes sick leave claims of up to 29 days long, these represent 79% of all claims. This figure is referenced in Section III.C.

Figure 3: Histogram of days on leave by workers characteristics



Notes: This figure shows the distribution of days on leave by workers' age and occupation for the youngest and oldest group of workers. Sample includes male private-sector employees. Blue-collar worker refers to workers who engage in hard manual labor, typically agriculture, manufacturing, construction, mining, or maintenance. White-collar worker refers to workers whose daily work activities do not involve manual labor—e.g., teachers or administrative staff. Additional groups are presented in Appendix Figure A2. This figure is referenced in Section III.C.

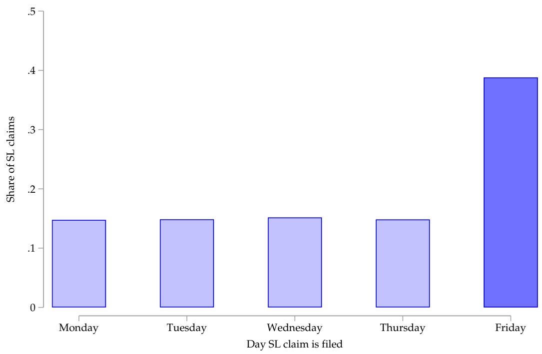
Figure 4: Identification of weekend-streak utility parameter (q): moments from raw data



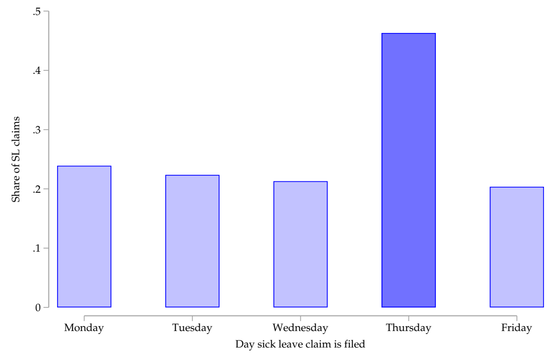
Notes: Panels (a) to (e) show the share of sick leave claims with duration s and the share of seven-days-long sick leave claims filed on each day of the week. Panel (f) aggregates across duration and days of the week: the first bar—labeled "weekend streak"—averages the share of one-to-five-days-long sick leave claims that end of a Friday and are filed any day of the week. For example, one-day-long on Friday, two-days-long on a Thursday, and so on. The second bar—labeled "non-weekend streak"—averages the share one-to-five-days-long sick leave claims filed any other day of the week. For example, two-days-long claims file on Friday. Table 5 reports the estimated shares and moments. This figure is referenced in Sections II.A and IV.B.

Figure 5: Identification of compliance costs parameter $(1/\kappa)$:
Sick leave claims by duration and day of the week. Health shock (θ) equals 1-day-long.

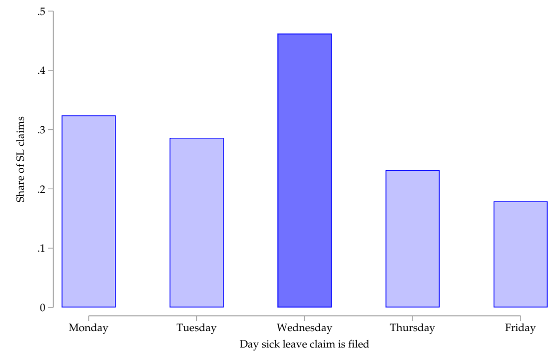
(a) Deviation = 0, duration = 1



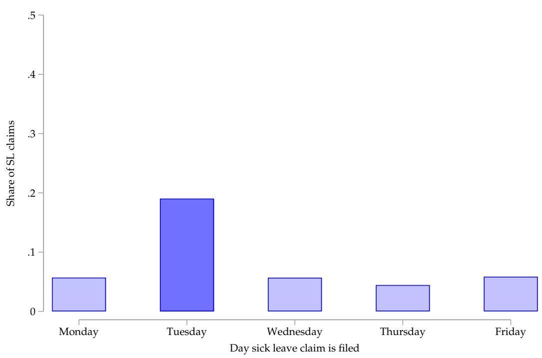
(b) Deviation = 1, duration = 2



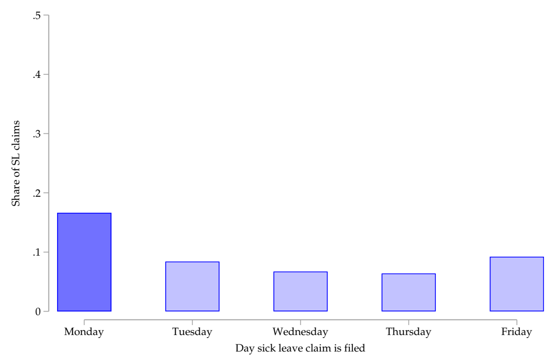
(c) Deviation = 2, duration = 3



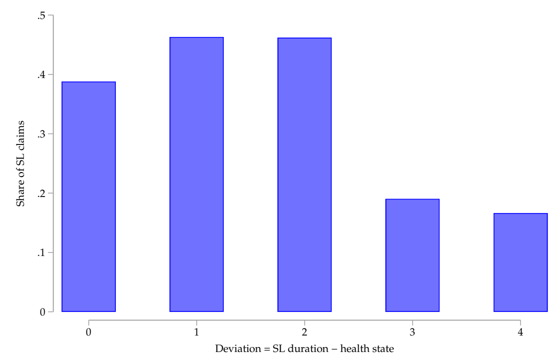
(d) Deviation = 3, duration = 4



(e) Deviation = 4, duration = 5



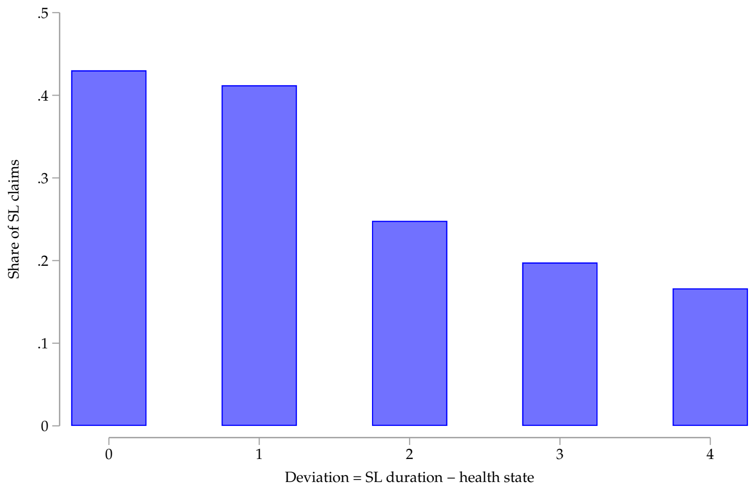
(f) Aggregate across days of the week



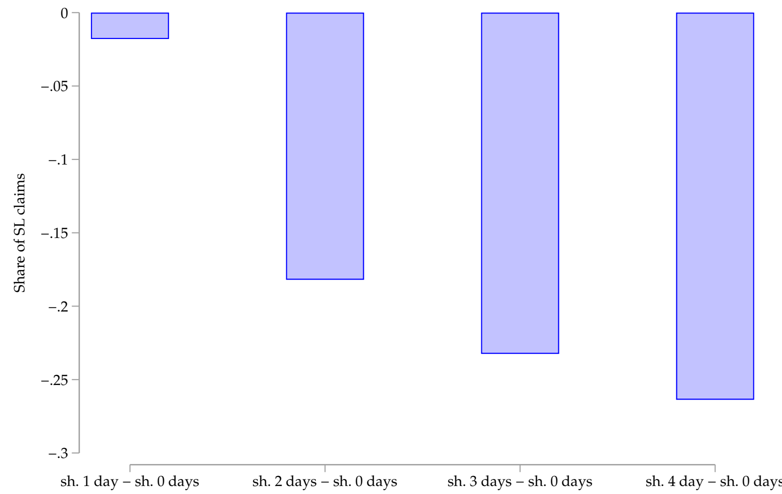
Notes: Panels (a) to (e) show the share of sick leave claims with duration s for workers whose main diagnose would implied a health state of 1 day on leave. Panel (f) aggregates the share of sick leave claims across days of the week, including only the weekend-streak combinations, e.g., from panel (a) I only consider the share for Friday. This figure is referenced in Section IV.B.

Figure 6: Identification of compliance costs parameter $(1/\kappa)$: moments from raw data

(a) Share of SL claims as a function of deviation

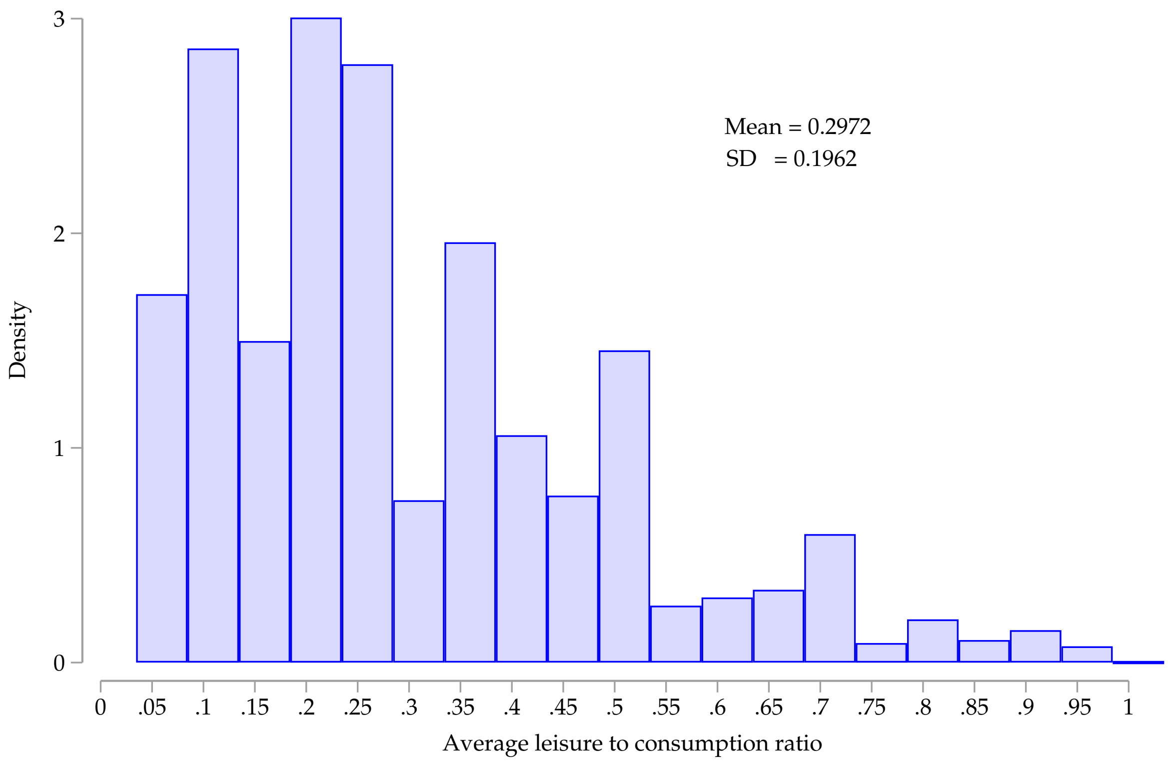


(b) Relative share of SL claims: $share_{dur} - share_0$



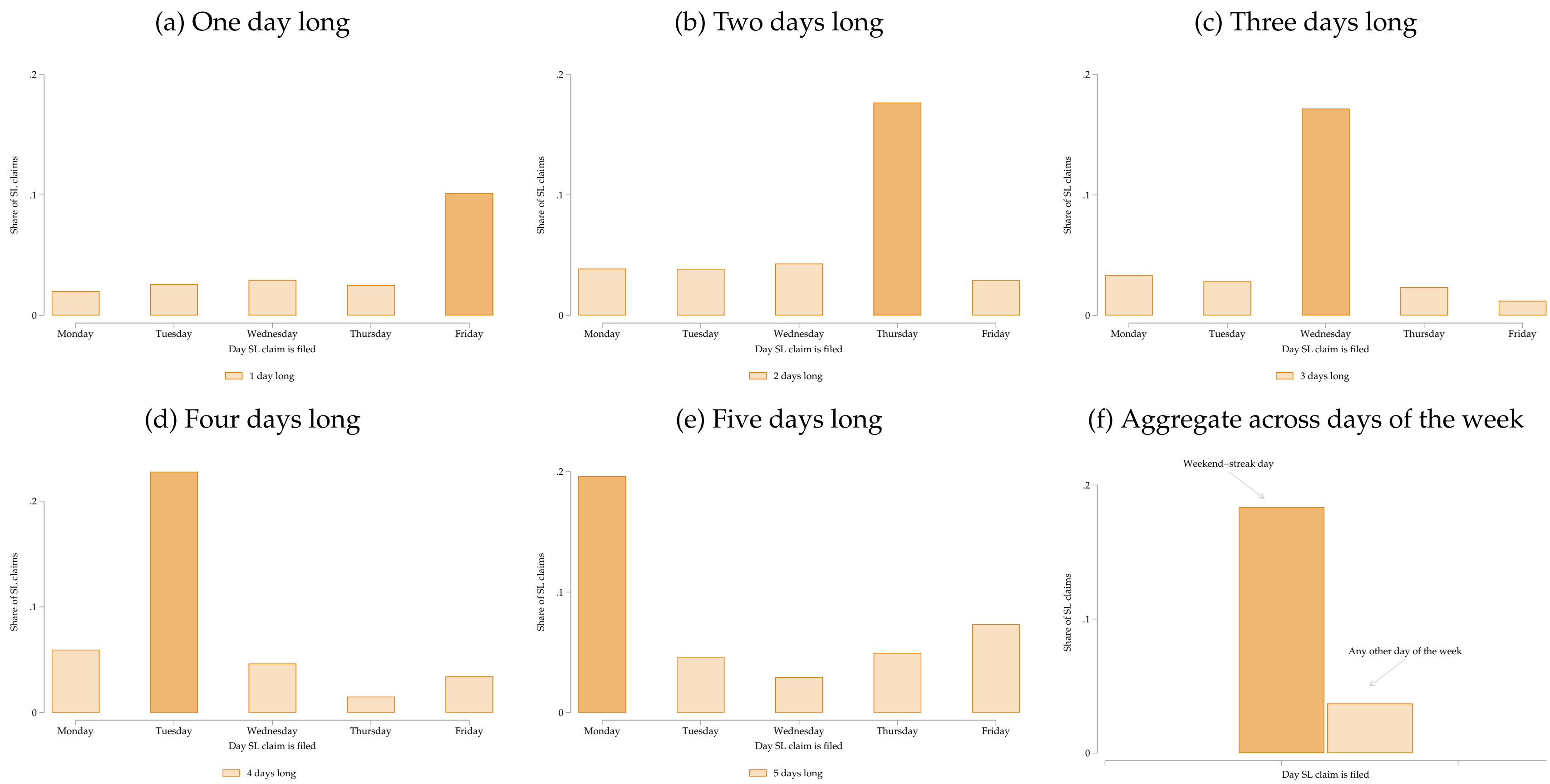
Notes: Panel (a) shows the average share of sick leave claims with deviations between 0 and 4 days. The average is computed over sick leave claims with primary diagnosis requiring 1, 2 or 3 days of rest filed on a weekend streak days. Each column is the weighted average of the probability for each health state. Panel (b) reports the difference between the share of SL claims with d deviation and SL claims with no-deviation. This figure is referenced in Section IV.B.

Figure 7: Identification of value of leisure parameter (ϕ):
Distribution of leisure to consumption ratio from raw data



Notes: This figure shows the distribution of the leisure to consumption ratio LC_i. This figure is referenced in Section IV.B.

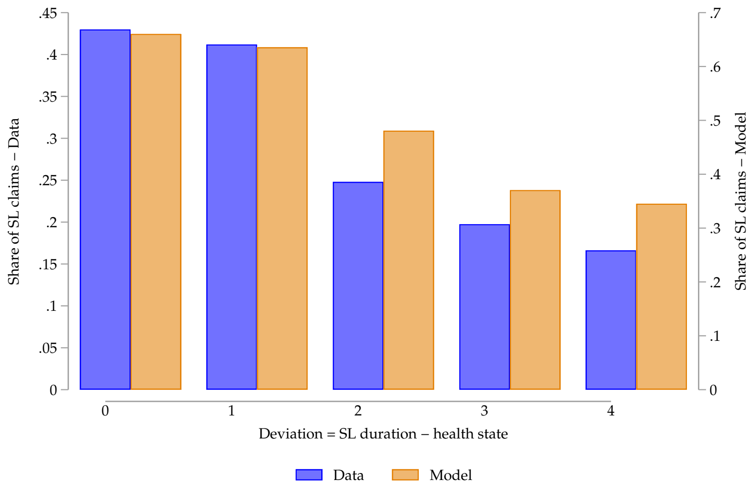
Figure 8: Model's prediction: weekend streak utility moments



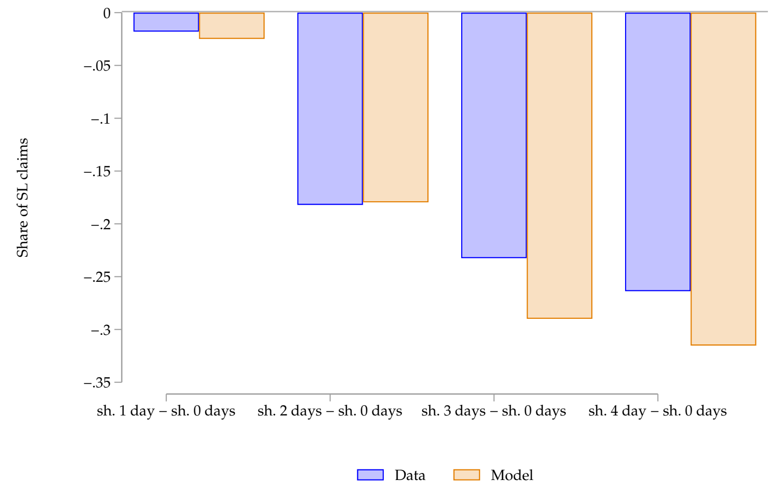
Notes: This figure presents the estimated-analogue of Figure 4. Panels (a) to (e) show the share of sick leave claims with duration s from a model-simulated sample. Panel (f) aggregates across duration and days of the week: the first bar—labeled "weekend streak"—averages the share of one-to-five-days-long sick leave claims that end of a Friday and are filed any day of the week. This figure is referenced in Section V.A.

Figure 9: Sick leave taking behavior: estimated days on leave relative to worker's health state

(a) Share of sick leave claims as a function of deviation



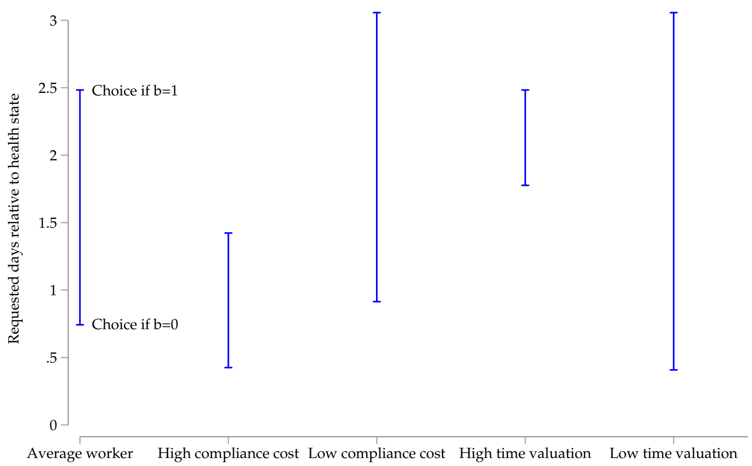
(b) Relative share of SL claims: $share_{dur} - share_0$



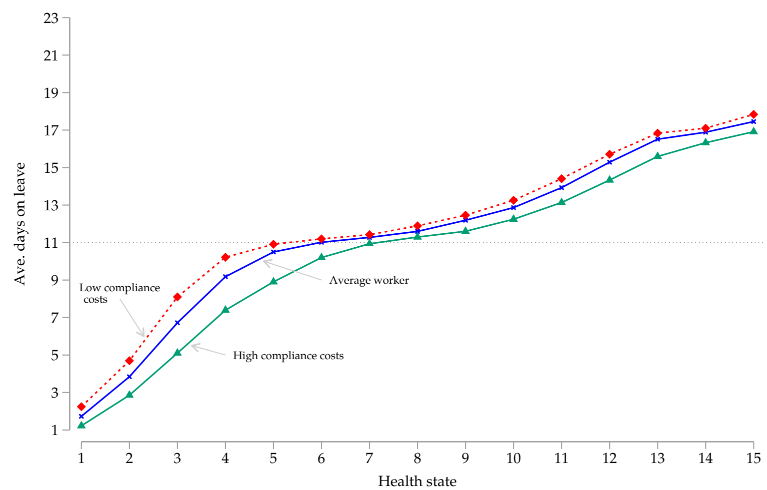
Notes: Panel (a) shows the comparison between the average share of sick leave claims with deviations between 0 and 4 days from the data and computed from a model-simulated sample. These are not-targeted moments. Panel (b) reports the difference between the share of SL claims with d deviation and SL claims with no-deviation from the data and from a model-simulated sample. The first three sets of bars correspond to targeted moments, the last set is a non-targeted moment. This figure is referenced in Section V.A.

Figure 10: Comparison between model's prediction and observed distributions: compliance costs and leisure to consumption ratio

(a) Linear contract



(b) Chilean contract



Notes: Panel (a) shows workers' choices, relative to their health state (i.e., $s^* - \theta$), as a function of their preference parameters under the assumption of a linear coverage contract. The lower bound of each segment corresponds to the choice under the least generous contract and the upper bound corresponds to the choice under the most generous contract. Panel (b) reports the demand for days on leave as a function of workers' health shock under the Chilean sick leave contract—shown in Figure 1. High compliance costs (low value of κ) correspond to the case where κ is calibrated to first quartile of its distribution. Low compliance costs (high value of κ) corresponds to the case where κ is calibrated to the third quartile of its distribution. Similarly, high (low) valuation of time corresponds to calibrating the ratio of wages to ϕ to the first (third) quartile. This figure is referenced in Section V.A.

IX Tables

Table 1: Summary statistics: all workers and workers who use sick leave insurance

	All workers	Workers who had used SL benefits		
		Any	Included conditions	
			All	Up to 30 days
	(1)	(2)	(3)	(4)
Age				
Mean	43.94	43.41	43.30	42.24
Share of workers aged (%)				
25 - 34 years old	26.35	28.90	29.14	32.11
35 - 44 years old	24.48	24.10	24.34	25.35
45 - 54 years old	26.70	24.71	24.66	23.73
55 - 64 years old	22.47	22.28	21.86	18.81
Income (monthly USD)				
Mean	772.00	904.70	909.42	918.02
Standard deviation	367.27	388.79	389.99	390.03
25th percentile	484.45	587.51	591.79	601.77
Median	682.15	829.84	835.53	845.74
75th percentile	997.97	1,146.82	1,152.48	1,161.29
90th percentile	1,328.04	1,483.17	1,489.20	1,496.41
Region (%)				
Central	34.97	40.76	41.22	41.92
Mining intensive	8.96	8.49	8.32	7.62
Health - chronic conditions (%)				
Hypertension	12.90	16.12	15.96	13.93
Diabetes	6.04	7.95	7.51	6.19
Share of workers (%)	100	18.50	17.12	13.78
Observations	1,916,138	354,469	328,053	263,951

Notes: This table presents summary statistics for all male workers in the sample (column 1) and for workers who have used sick leave benefits in the past year based on the conditions and duration of sick leave claims (columns 2 to 4). The sample includes private and public sector employees age 25 to 64 years old. Income statistics are based on the winsorized distribution where the lowest and highest 5% of the income values are excluded. Sick leave claims of up to 30 days account for 95% of all claims filed in a year. This table is referenced in Section III.C.

Table 2: Summary statistics: workers who use sick leave insurance by duration.

	All	Sick leave claims duration		
		1 to 3 days	4 to 10 days	11 to 29 days
	(1)	(2)	(3)	(4)
Age				
Mean	42.00	39.40	41.79	43.97
Share of workers aged (%)				
25 - 34 years old	33.07	41.74	33.74	26.38
35 - 44 years old	25.11	26.22	25.23	24.56
45 - 54 years old	23.50	19.82	23.15	26.10
55 - 64 years old	18.32	12.23	17.88	22.96
Income (monthly USD)				
Mean	870.30	953.50	852.48	849.98
Standard deviation	367.30	386.19	353.57	358.76
25th percentile	569.20	645.54	564.03	552.85
Median	797.88	892.11	782.44	776.38
75th percentile	1,103.05	1,201.53	1,075.22	1,072.83
90th percentile	1,418.81	1,523.37	1,380.31	1,390.51
Region (%)				
Central	48.35	59.10	48.05	43.78
Mining intensive	6.01	4.09	5.30	7.89
Health - chronic conditions (%)				
Hypertension	14.34	11.46	14.48	17.17
Diabetes	6.22	4.64	6.21	7.92
Share of workers (%)	100	32.00	49.76	38.08
Observations	177,531	56,803	88,345	67,599

Notes: This table presents summary statistics for all male workers who have used sick leave insurance in the past year. Column (1) presents characteristics of all workers who have filed at least one claim with duration of up to 30 days for conditions included in the analysis of this paper (see Table A6 for more details). Columns (2) to (4) present characteristics of workers by duration of the sick leave claims filed. Columns (2) to (4) are not exhaustive, that is, workers can be included in more than one category based on the claims they have filed. This table is referenced in Section III.C.

Table 3: Probability of contracting disease d by workers' characteristics

		25 - 34 years old		35 - 44 years old		45 - 54 years old		55 - 64 years old	
ICD group	Main diagnoses	Blue c. (1)	White c. (2)	Blue c. (3)	White c. (4)	Blue c. (5)	White c. (6)	Blue c. (7)	White c. (8)
A00-A99	Infectious gastroenteritis	12.36	18.26	8.91	13.06	6.51	9.36	4.80	6.57
G00-G99	Migraine and headaches	1.64	2.28	1.57	1.86	1.19	1.64	0.89	1.14
G00-G99	Carpal tunnel syndrome	1.13	1.08	1.31	1.30	1.48	1.31	1.63	1.33
H00-H59	Conjunctivitis	1.57	1.46	1.92	2.40	2.40	2.40	3.03	3.19
H60-H95	Vertigo	1.08	1.24	1.32	1.23	1.85	1.68	2.49	2.64
I00-I99	Hypertension	1.09	0.80	1.88	1.70	2.95	2.85	4.16	3.81
I00-I99	Myocardial infarction	0.12	0.08	0.31	0.34	0.77	0.71	1.55	1.38
J00-J06	Common cold	11.89	17.12	10.64	14.87	8.62	12.79	6.34	10.17
J09-J18	Influenza and pneumonia	3.31	3.92	3.78	4.66	4.68	4.88	5.53	5.58
J20-J22	Bronchitis	5.44	6.44	5.91	7.33	7.18	8.35	8.06	10.80
J23-J99	Other respiratory diseases	0.96	1.00	1.11	1.13	1.30	1.37	1.74	1.71
K00-K95	Noninfective gastroenteritis	5.34	5.99	4.19	4.99	3.95	4.57	3.34	3.93
K00-K95	Inguinal hernia	1.93	1.62	3.02	2.87	3.82	4.02	4.48	4.73
M50-M54	Chronic low back pain	16.85	12.35	16.73	12.78	14.17	11.43	12.02	10.02
M50-M54	Lumbago with sciatica	6.37	4.86	8.11	6.33	7.54	6.60	6.70	5.87
M60-M79	Tendinitis	5.17	3.56	5.99	4.37	6.55	4.98	6.67	4.95
M60-M79	Shoulder lesions	2.47	1.79	3.78	2.49	4.62	3.61	4.43	3.11
Other M	Arthritis	1.91	1.34	2.25	1.79	3.11	2.63	4.38	3.63
Other M	Knee injuries	0.57	0.41	0.70	0.66	1.16	0.99	1.61	1.33
N00-N99	Renal colic	2.42	2.12	3.06	3.23	3.63	3.47	4.29	4.48
R00-R99*	Abdominal and pelvic pain	2.19	2.38	1.99	2.37	2.38	2.72	3.01	2.83
S00-S99	Injuries (e.g., sprain ankle)	14.19	9.89	11.53	8.25	10.13	7.64	8.87	6.82

Notes: This table shows the probability $\widehat{p}_b^d$ that individual i, characterized by his age and occupation type, contracts disease d. Each of these probabilities is computed as the ratio of sick leave claims with diagnosis d and all claims from group b, thus columns add up to 100. Main diagnoses indicates the most common condition for a disease group. Blue c. and white c. stand for blue-collar and white-collar occupations respectively. These probabilities are plotted in Figure A3. This table is referenced in Section IV.B.

Table 4: Estimates of the wage process parameters by age and occupation

Age	Type of occupation	Ln wages		Wages		Observations
		Mean ($\mu_{w,b}$)	SD ($\sigma_{w,b}$)	Mean	SD	
		(1)	(2)	(3)	(4)	(5)
25 - 34 years old	Blue collar	6.4894	0.5763	763.43	398.61	35,748
25 - 34 years old	White collar	6.5818	0.5723	837.42	445.83	17,482
35 - 44 years old	Blue collar	6.6580	0.5830	908.45	487.20	29,789
35 - 44 years old	White collar	6.8197	0.5673	1,064.70	580.74	12,570
45 - 54 years old	Blue collar	6.6535	0.5557	895.87	476.15	32,242
45 - 54 years old	White collar	6.8169	0.5541	1,058.68	583.94	10,283
55 - 64 years old	Blue collar	6.5894	0.5300	831.51	431.96	27,791
55 - 64 years old	White collar	6.7577	0.5586	1,002.68	573.55	8,317

Notes: This table presents estimates of the wage process parameters: the mean wage ($\mu_{w,b}$) and the standard deviation ($\sigma_{w,b}$) for each group of workers. Columns (1) and (2) present moments for the logarithm of wages and columns (3) and (4) for the wages in levels. This table is referenced in Section IV.B.

Table 5: Identification of weekend-streak utility parameter (q): estimates from raw data

Duration	Day of the week		
	Weekend streak	Non-weekend streak	Difference
	(1)	(2)	(3)
1 day long	0.1219	0.0355	0.0864
2 days long	0.2062	0.0672	0.1391
3 days long	0.2872	0.1482	0.1390
4 days long	0.1640	0.0489	0.1151
5 days long	0.2330	0.1216	0.1114
Simple average	0.2025	0.0843	0.1182
Weighted average	0.2274	0.1041	0.1233

Notes: This table presents the distribution of sick leave claims by duration and day of the week. Weekend streak refers to the day of the week a sick leave claim should start to finish on Friday. For example, when duration is one day, weekend streak refers to Friday, when duration is two days, it refers to Thursday. The non-weekend streak category groups all the other days of the week. The share of sick leave claims of duration s filed on day dow is computed as the ratio between the number of claims with duration s filed on dow and the number of claims of filed on dow with duration between one and fifteen days. Figure 4 presents this table graphically. This table is referenced in Section IV.B.

Table 6: Moments used in the estimation

Moments	Data (1)	Model (2)
Weekend streak utility		
Weekend streak days relative to non-streak days	0.1233	0.1464
Compliance costs		
Share 1 day dev - share no deviation	-0.0180	-0.0249
Share 2 day dev - share no deviation	-0.1820	-0.1796
Share 3 day dev - share no deviation	-0.2325	-0.2899
Share 11 days-long sick leave claims	0.0727	0.1043
Value of leisure		
Mean leisure to consumption ratio	0.2972	0.3216
SD leisure to consumption ratio	0.1962	0.1607
Streak utility, conditional to 5-days-long and a day of recovery		
Monday - Tuesday	0.0822	0.1158
Monday - Wednesday	0.0990	0.1180
Monday - Thursday	0.1023	0.1218
Heaping points		
5 days	0.0912	0.0234
7 days	0.1175	0.0678

Notes: This table presents the moments used to estimate the model's parameter. Column 2 reports the data moments. Column 3 reports simulated moments based on 10,000 simulations. This table is referenced in Section IV.B and in Section V.

Table 7: Preference parameters estimates

Parameter	Description	Value
Preference parameters		
q	Weekend-streak utility	0.1464
$\mu_{wage/\phi}$	Value of leisure relative to consumption, mean	0.7120
$\sigma_{wage/\phi}$	Value of leisure relative to consumption, standard deviation	0.7090
μ_{κ}	Compliance costs, mean	2.5949
σ_{κ}	Compliance costs, standard deviation	1.7008
η_r	Curvature of cost function, curvature for	2.0314
η_l	Curvature of cost function, curvature for	3.9576
Measurement error		
p_1	Prob. physician assigns one day more (less) than asked	0.0909
p_2	Prob. physician assigns two days more (less) than asked	0.0722
p_3	Prob. physician assigns three days more (less) than asked	0.0508
Rounding		
p_{r_5}	Prob. sick leave duration is rounded to the closest multiple of 5	0.1999
p_{r_7}	Prob. sick leave duration is rounded to the closest multiple of 7	0.2075

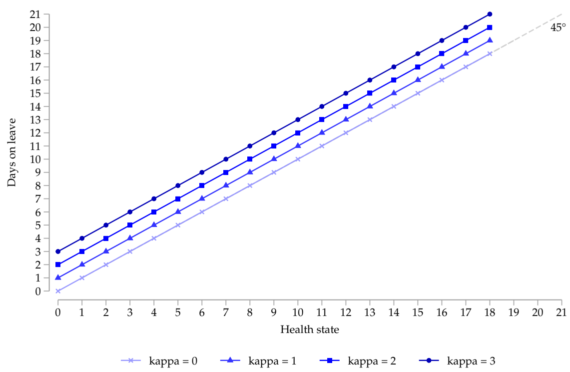
Notes: This table presents estimates of preference, measurement error and rounding parameters using the Simulated Method of Moments. Estimates are based on 10,000 simulations. This table is referenced in Section V.

Appendix A. Additional Figures and Tables

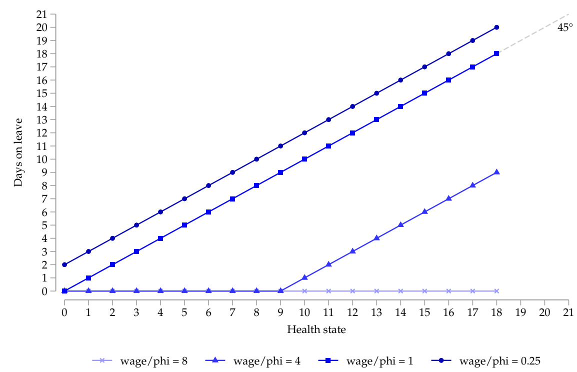
A.I Additional Figures

Figure A1: Sick Pay Utilization with Linear Contract

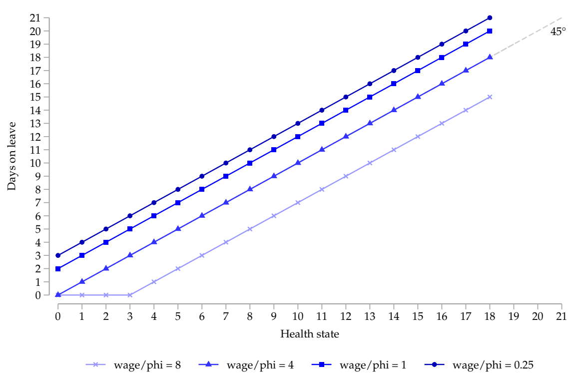
(a) Full coverage ($b = 1$)



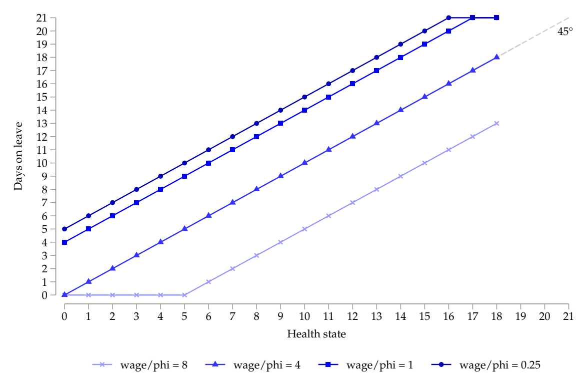
(b) Unpaid sick leave ($b = 0$) and $\kappa = 3$



(c) Partial coverage ($b = 0.75$) and $\kappa = 3$

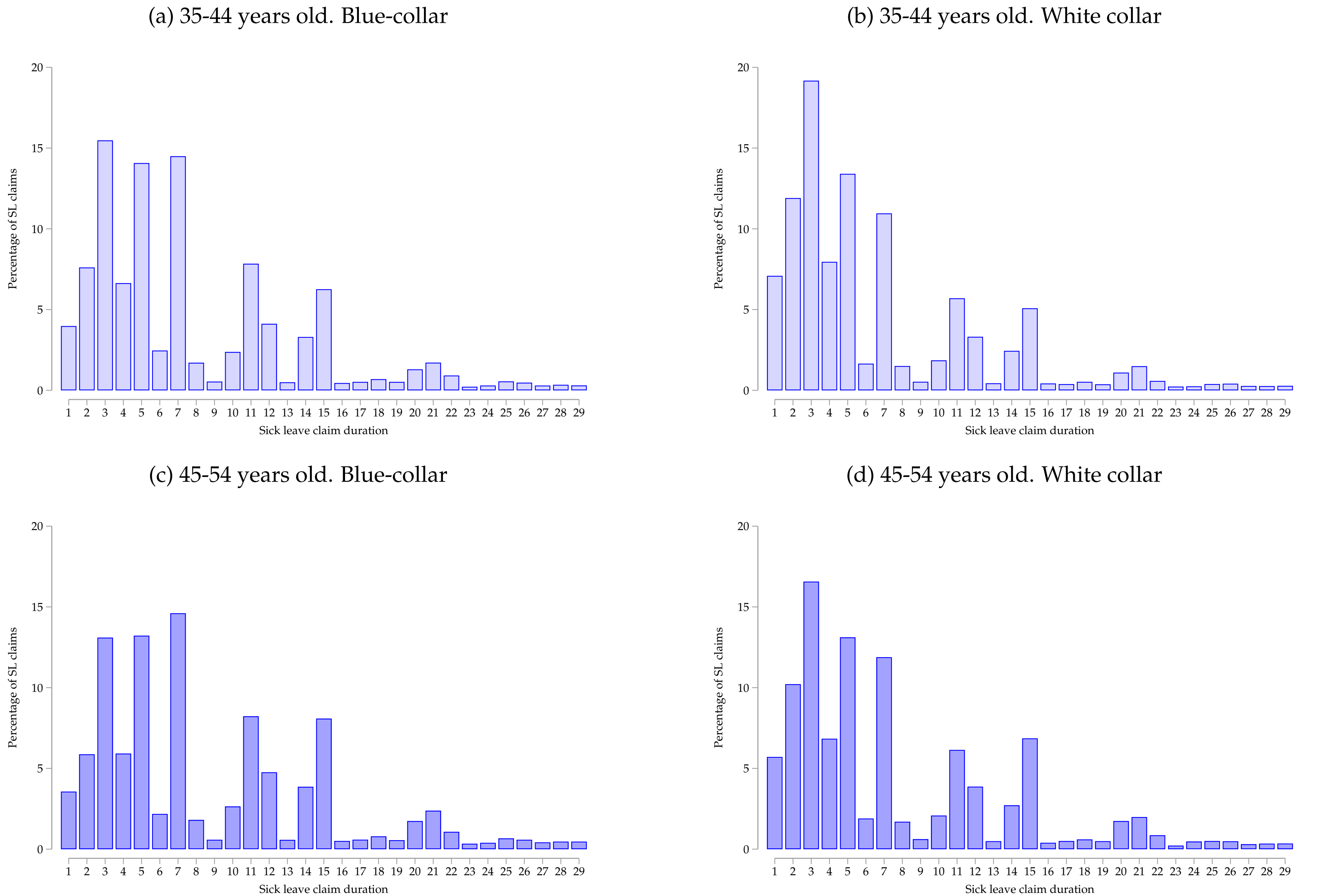


(d) Partial coverage ($b = 0.75$) and $\kappa = 5$



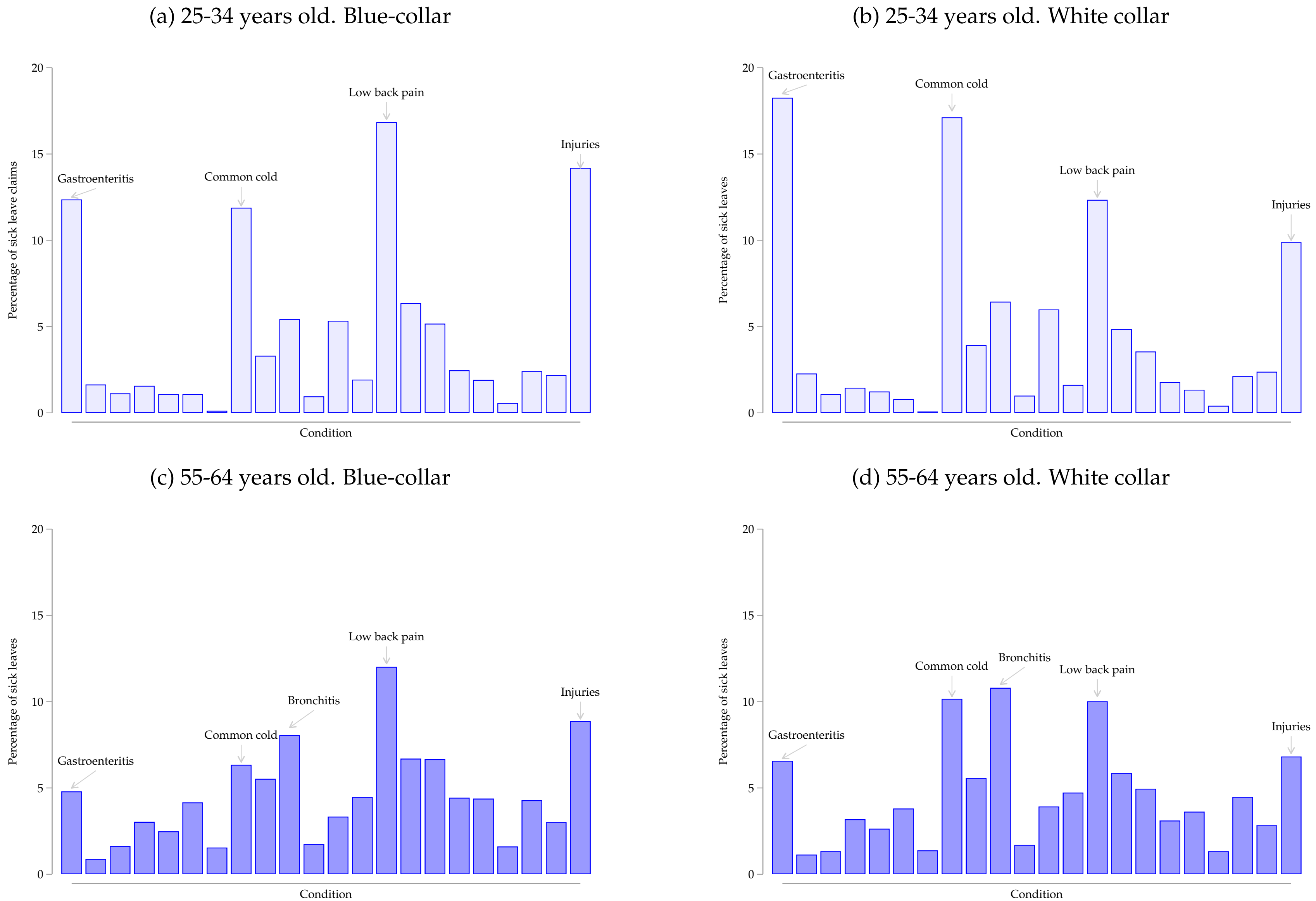
Notes: This figure shows the optimal demand of days on leave $s^*(\theta)$ as a function of worker's health status (θ) under the assumption of linear contracts with different levels of coverage. This figure is referenced in Section II.

Figure A2: Histogram of days on leave by workers characteristics



Notes: This figure shows the distribution of days on leave by workers' age and occupation for the youngest and oldest group of workers. Sample includes male private-sector employees. Blue-collar worker refers to workers who engage in hard manual labor, typically agriculture, manufacturing, construction, mining, or maintenance. White-collar worker refers to workers whose daily work activities do not involve manual labor—e.g., teachers or administrative staff. Additional groups are presented in Appendix Figure 3. This figure is referenced in Section III.C.

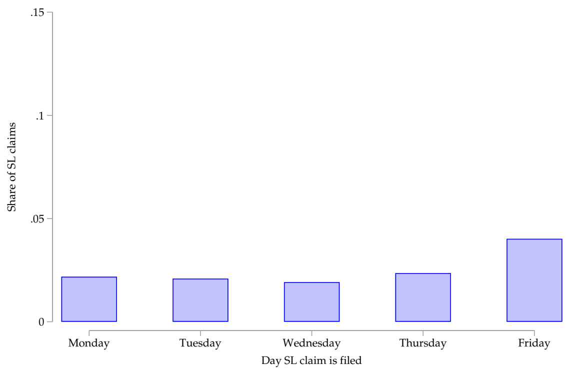
Figure A3: Histogram of days on leave by workers characteristics



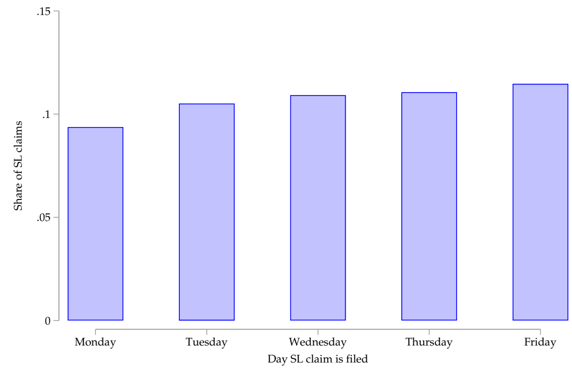
Notes: This figure shows the probability that a worker contracts disease d by by workers' age and occupation for the youngest and oldest group of workers. Diseases are ordered as presented in Table 3. Sample includes male private-sector employees. Blue-collar worker refers to workers who engage in hard manual labor, typically agriculture, manufacturing, construction, mining, or maintenance. White-collar worker refers to workers whose daily work activities do not involve manual labor—e.g., teachers or administrative staff. This figure is referenced in Section IV.B.

Figure A4: Distribution of sick leave claims by duration and day of the week.

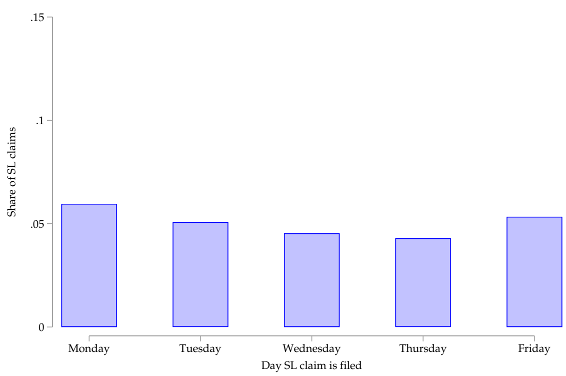
(a) Eight and nine days long



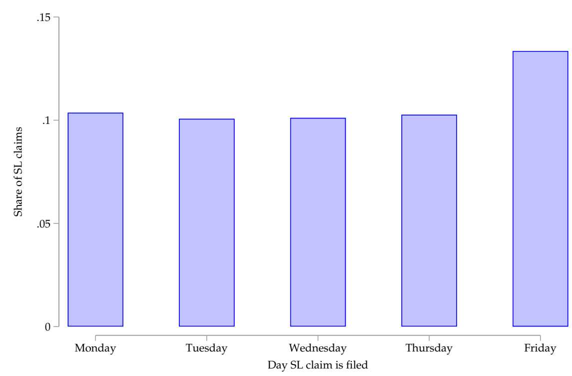
(b) Ten and eleven days long



(c) Twelve and thirteen days long

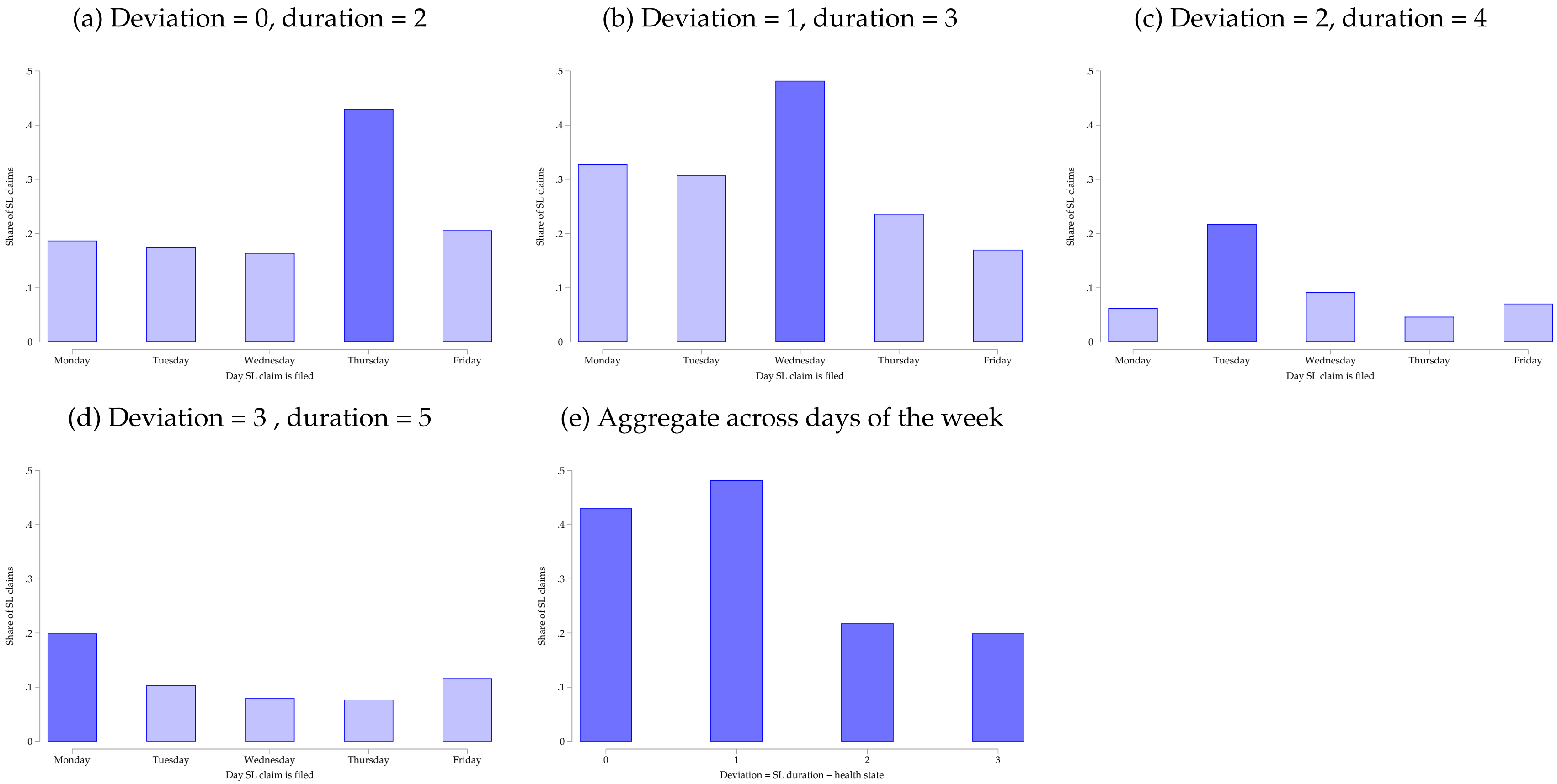


(d) Fourteen and fifteen days long



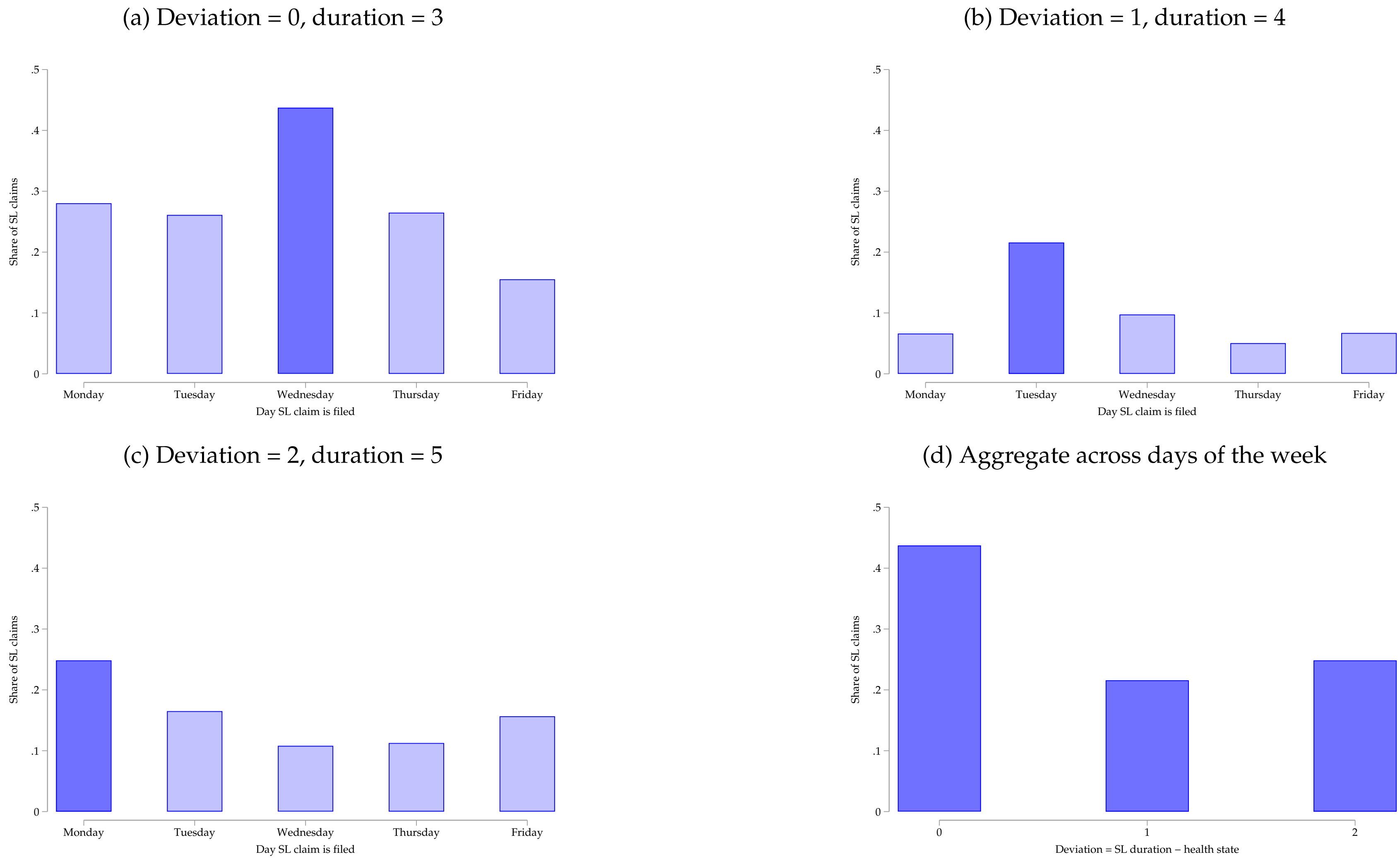
Notes: This figure shows the share of sick leave claims of duration s filed on each day of the week. Each panel aggregates sick leave claims with consecutive duration as stated in the title. This figure is referenced in Section IV.B.

Figure A5: Identification of compliance costs parameter $(1/\kappa)$:
Sick leave claims by duration and day of the week. Health shock (θ) equals 2-days-long.



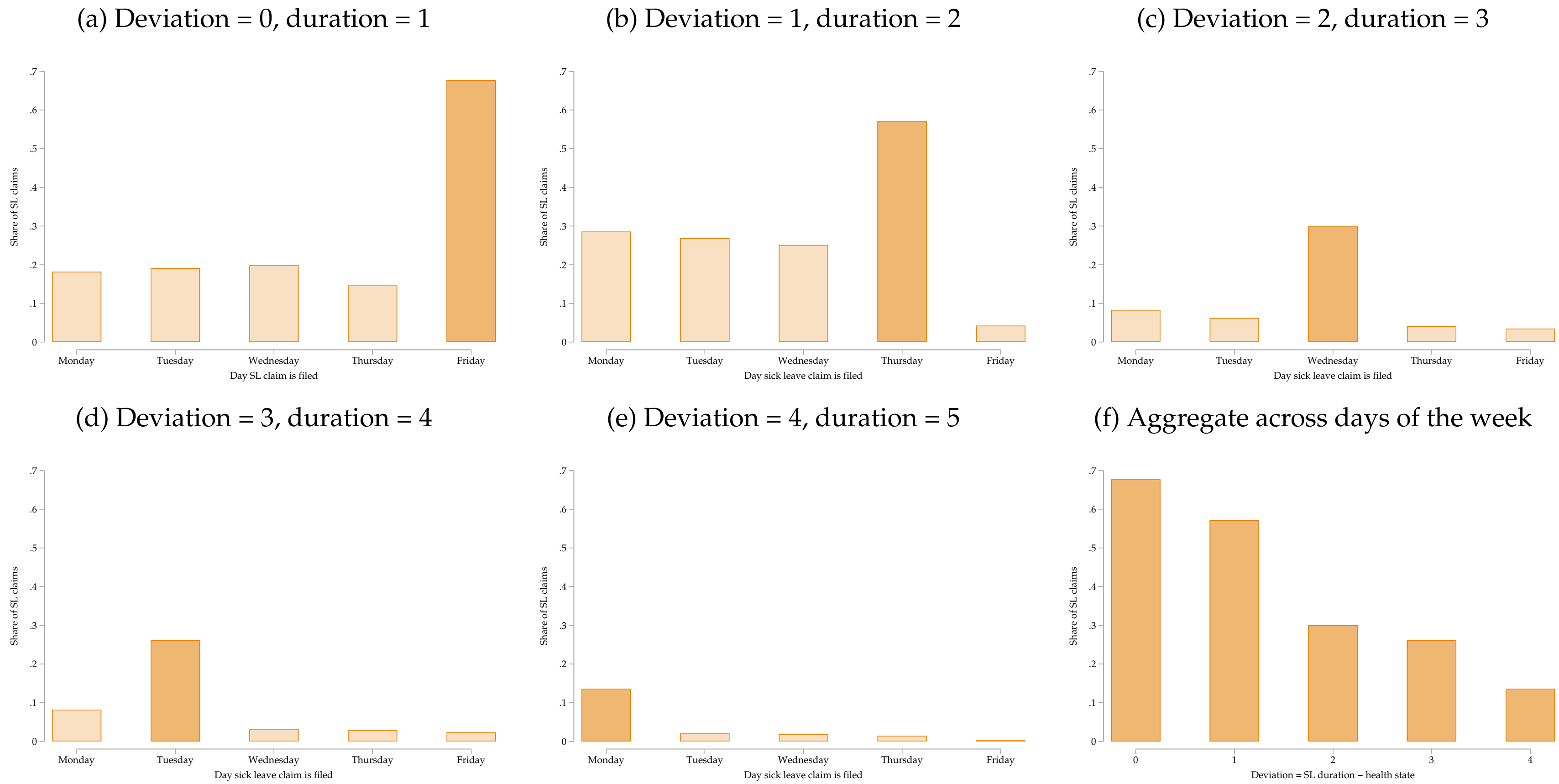
Notes: Panels (a) to (d) show the share of sick leave claims with duration s for workers whose main diagnose would implied a health state of 2 days on leave. Panel (e) aggregates the share of sick leave claims across days of the week, including only the weekend-streak combinations, e.g., from panel (a) I only consider the share for Thursday. This figure is referenced in Section IV.B.

Figure A6: Identification of compliance costs parameter $(1/\kappa)$:
Sick leave claims by duration and day of the week. Health shock (θ) equals 3-days-long.



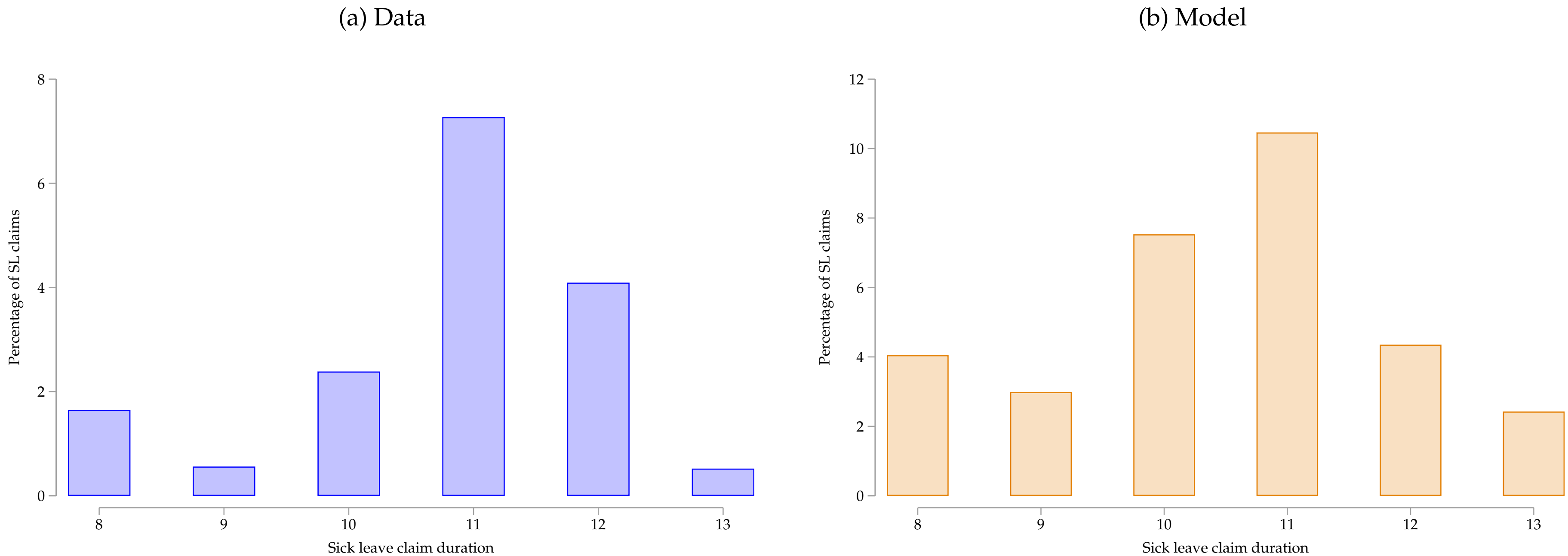
Notes: Panels (a) to (c) show the share of sick leave claims with duration s for workers whose main diagnose would implied a health state of 3 days on leave. Panel (d) aggregates the share of sick leave claims across days of the week, including only the weekend-streak combinations, e.g., from panel (a) I only consider the share for Wednesday. This figure is referenced in Section IV.B.

Figure A7: Compliance costs parameter $(1/\kappa)$:
Simulated sick leave claims by duration and day of the week. Health shock (θ) equals 1-day-long.



Notes: This figure shows the estimated-analogue of Figure A7. Panels (a) to (e) show the share of sick leave claims with duration s for workers whose main diagnose would implied a health state of 1 day on leave from a model-simulated sample. Panel (f) aggregates the share of sick leave claims across days of the week, including only the weekend-streak combinations, e.g., from panel (a) I only consider the share for Friday. This figure is referenced in Section V.A.

Figure A8: Distribution of sick leave claims with duration around 11 days.



Notes: This figure compares the distribution of sick leave claims with a duration in the 11-days-neighborhood from the data and from a model-simulated sample. This figure is referenced in Section V.A.

A.II Additional Tables

Table A1: Number of business days on leave (s_l)

	Number of days on leave (s_c)							
Day of the week (dow)	1	2	3	4	5	6	7	8
Monday	1	2	3	4	5	5	5	6
Tuesday	1	2	3	4	4	4	5	6
Wednesday	1	2	3	3	3	4	5	6
Thursday	1	2	2	2	3	4	5	6
Friday	1	1	1	2	3	4	5	6

Notes: This table shows the number of business days on leave (s_l) as a function of (total) days on leave (s_c) and day of the week (dow) a sick leave claim is filed. This table is referenced in Section II.A.

Table A2: Summary statistics by healthcare insurance provider

	Government-run insurance (1)	Private insurance (2)	Year(s) (3)
Panel A. Enrollees Characteristics			
Share of enrollees aged			
25 - 34	25.06	31.76	2015-2019*
35 - 44	21.51	28.84	2015-2019*
45 - 54	21.11	20.12	2015-2019*
55 - 64	14.64	11.47	2015-2019*
Share female enrollees	0.44	0.35	2015-2019*
Wages (in USD monthly)			
Average	761.27	1,824.94	2015-2019*
Enrollees w/ wage above median (%)	34.44	86.69	2015-2019*
Metropolitan region (%)	38.04	60.01	2015-2019*
Mining sector (%)	0.50	2.49	2015-2019*
N of enrollees	4,503,474	1,689,240	2015-2019*
Share (%)	72.72	27.28	2015-2019*
Panel B. Sick Leave Claims			
Ratio SL claims to enrollees (%)			
2015	77.66	86.53	2015
2019	98.42	90.66	2019
Approved SL claims (%)	91.94	74.54	2015-2019
Rejected SL claims (%)	5.31	14.76	2015-2019
Ratio days on leave to SL claim	13.09	10.24	2015-2019
Annual cost per enrollee (in USD)	240.69	463.61	2015-2019
Ratio of total annual cost to paid days on leave	24.91	58.90	2015-2019
Annual cost (% of GDP)	0.51	0.37	2015-2020
N of sick leave claims	3,910,482	1,473,540	2015-2019
Share (%)	72.63	27.37	2015-2019

Notes: Panel A presents summary statistics of individuals enrolled in plans offered by each healthcare insurance provider. Only individuals eligible to file a sick leave claim are included in the computations. Panel B shows characteristics of the sick leave claims handled by each insurer. Data come from the Annual Statistics of the Sick Leave System published by the Social Security Administration (SUSESO, 2020; 2019; 2018; 2017; 2016). The reported data are annual counts. Statistics in this table correspond to averages for 2015 - 2019, * indicates that data for 2018 are not available. The median monthly wage is computed from the 2017 CASEN survey, and using this figure I compute the share of workers with monthly salary above the median. GDP data comes from the World Bank national accounts data. SL stands for sick leave. This table is referenced in Section III.A.

Table A3: Sick leave claims and sick leave spells definitions

	(1)	(2)	(3)
Number of sick leave claims	1,483,103	657,125	551,647
Number of sick leave spells	1,030,613	437,418	365,127
N of SL claims in a spell (% of claims)			
One claim	55.43	51.71	51.22
Two claims	16.85	17.19	17.30
Three claims	7.30	7.93	8.00
Four claims	4.49	5.06	5.12
Five claims	3.19	3.63	3.67
Six or more claims	12.75	14.48	14.70
Among sick leave spells with more than 1 claim* (% of claims)			
Diagnoses change within spell			
Yes — 4 digits disease code	30.83	30.01	29.97
Yes — 3 digits disease code	28.67	27.86	27.82
Physician change within spell	31.21	30.87	30.83
Sample: Private sector workers			
Gender	All	Male	Male
Age	18-70	18-70	25-64

Notes: This table presents counts and summary statistics of sick leave claims and sick leave spells. A spell is a group of consecutive claims—these are considered one claim for the computation of sick leave benefits. The first row counts each sick leave claim as one observation and the second row considers the number of sick leave spells. The subsequent rows explore the composition of a spell in terms of number of claims and whether diagnoses and physicians changed withing a spell. * indicates that proportions are computed for spells composed by two to five sick leave claims. This table is referenced in Section III.B and in Section III.C.

Table A4: Sample construction

	2017
Panel A. Sick leave claims	
Single claims - from clean dataset	2,698,993
Observations without demographic information	22,757
Workers' age not in the interval [18,70]	50,369
Worker is not Chilean	61,740
Worker not enrolled in a public insurance plan	33,560
Observations without income information	8,920
Observations	2,521,647
Panel B. Sick leave spells	
Single spells	1,825,904
Condition on private sector workers	1,030,613
Condition on male workers	437,418
Condition on ages 25-64	365,127
Condition on diagnoses included in analysis	329,312

Notes: Panel A of this table shows the counts of sick leave claims drop due to each sample selection criterion. Panel B shows the counts of sick leave spells—consecutive claims with continuous start and end dates—for each sample selection criterion. A complete list of diagnoses included in the analysis is provided in Table A6. This table is referenced in Section III.C.

Table A5: Summary statistics: Private sector workers who have used SL benefits

	Any	Included conditions	
		All	Up to 30 days
	(1)	(2)	(3)
Age			
Mean	43.39	43.25	42.00
Share of workers aged (%)			
25 - 34 years old	29.21	29.51	33.07
35 - 44 years old	23.71	23.98	25.11
45 - 54 years old	24.70	24.63	23.50
55 - 64 years old	22.38	21.89	18.32
Income (monthly USD)			
Mean	857.97	862.72	870.30
Standard deviation	367.59	368.79	367.30
25th percentile	555.28	559.42	569.20
Median	782.85	788.21	797.88
75th percentile	1,089.33	1,095.32	1,103.05
90th percentile	1,408.30	1,414.46	1,418.81
Region (%)			
Central	46.78	47.32	48.35
Mining intensive	6.77	6.67	6.01
Health - chronic conditions (%)			
Hypertension	16.89	16.71	14.34
Diabetes	8.24	7.79	6.22
Share of workers (%)	100.00	92.59	72.16
Observations	246,017	227,797	177,531

Notes: This table presents summary statistics for all male workers who had used sick leave benefits in the past year based on the conditions and duration of sick leave claims. The sample includes private sector employees age 25 to 64 years old. Income statistics are based on the winsorized distribution where the lowest and highest 5% of the income values are excluded. Sick leave claims of up to 30 days account for 95% of all claims filed in a year. This table is referenced in Section III.C.

Table A6: Conditions included in the analysis by ICD group

ICD Group	Description	Sick leave claims		Included
		Number	Share (%)	(=1 if yes)
		(1)	(2)	(3)
A00-B99	Certain infectious and parasitic diseases	1	31,244	8.56
C00-D49	Neoplasms	0	6,515	1.78
D50-D89	Blood and blood-forming organs	0	478	0.13
E00-E89	Nutritional and metabolic diseases	0	3,842	1.05
G00-G99	Nervous system	1	8,758	2.40
H00-H59	Eye and adnexa	1	6,141	1.68
H60-H95	Ear and mastoid process	1	6,246	1.71
I00-I99	Circulatory system	1	15,139	4.15
J00-J99	Respiratory system	1	64,823	17.75
K00-K95	Digestive system	1	25,854	7.08
L00-L99	Skin and subcutaneous tissue	0	8,762	2.40
M00-M99	Musculoskeletal system	1	108,908	29.83
N00-N99	Genitourinary system	1	11,605	3.18
O00-O9A	Pregnancy and childbirth	0	29	0.01
P00-P96	Certain conditions of the perinatal	0	149	0.04
Q00-Q99	Congenital malformations	0	331	0.09
R00-R99	Abnormal clinical and laboratory findings	1	9,840	2.69
S00-S99	Injuries	1	44,922	12.30
T00-T88	Poisoning and external causes	0	4,385	1.20
U00-U85	Codes for special purposes	0	1	0.00
V00-Y99	External causes of morbidity	0	3,578	0.98
Z00-Z99	Contact with health services	0	3,577	0.98
Total included			329,312	90.19
Total			365,127	

Notes: This table reports the health conditions included in the analysis, the number of sick leave claims filed in 2017, and what share these represent of the universe of claims. The criteria for excluding the selected health conditions is discussed in Section Appendix C.. This table is referenced in Section .

Table A7: Average recovery times - Examples from Peruvian Handbook

Workers' characteristics and diagnoses	Correction factor and recovery time
Example 1	
43 years old	1.05
Operator/manual worker (blue collar)	1.5
Lumbago with sciatica (M544)	14
Optimal time	22.05
Example 2	
25 years old	0.87
Teacher (white collar)	0.75
Common cold (J00)	3
Optimal time	2
Example 3	
57 years old	1.3
Office manager (white collar)	0.75
Infectious gastroenteritis (A09)	2
Optimal time	2

Notes: This table presents examples on how to construct the average recovery time based on workers' characteristics and sick leave diagnoses. This table is referenced in Section IV.B.

Table A8: Average recovery time by workers characteristics $\overline{\theta_b^d}$.

ICD group	Main diagnoses	25 - 34 years old		35 - 44 years old		45 - 54 years old		55 - 64 years old	
		Blue c. (1)	White c. (2)	Blue c. (3)	White c. (4)	Blue c. (5)	White c. (6)	Blue c. (7)	White c. (8)
A00-A99	Infectious gastroenteritis	2	1	3	2	3	2	3	2
G00-G99	Migraine and headaches	3	2	4	2	4	2	5	3
G00-G99	Carpal tunnel syndrome	13	8	17	10	17	10	21	13
H00-H59	Conjunctivitis	5	3	7	4	7	4	9	5
H60-H95	Vertigo	4	2	5	3	5	3	6	4
I00-I99	Hypertension	4	3	6	3	6	3	7	4
I00-I99	Myocardial infarction	16	10	21	13	21	13	26	16
J00-J06	Common cold	3	2	4	2	4	2	5	3
J09-J18	Influenza and pneumonia	4	3	5	3	5	3	6	4
J20-J22	Bronchitis	5	3	7	4	7	4	8	5
J23-J99	Other respiratory diseases	8	5	9	6	9	6	11	7
K00-K95	Noninfective gastroenteritis	2	1	2	1	2	1	3	2
K00-K95	Inguinal hernia	6	4	9	5	9	5	11	7
M50-M54	Chronic low back pain	10	6	12	7	12	7	14	8
M50-M54	Lumbago with sciatica	10	6	12	7	12	7	14	8
M60-M79	Tendinitis	8	5	9	6	9	6	10	6
M60-M79	Shoulder lesions	8	5	9	6	9	6	10	6
Other M	Arthritis	9	5	10	6	11	6	12	7
Other M	Knee injuries	12	7	14	8	14	8	16	10
N00-N99	Renal colic	4	3	5	3	5	3	6	4
R00-R99*	Abdominal and pelvic pain	2	1	3	2	3	2	3	2
S00-S99	Injuries (e.g., sprain ankle)	14	8	16	9	16	9	18	11

Notes: This table shows the average time required for worker i characterized by his age and occupation type to recover from disease d, these correspond to the $\overline{\theta_b^d}$ in the specification of the model. This table is referenced in Section IV.B.

Table A9: Workers' occupation, industry and manual work classification

Occupation	Industry	Blue collar (=1 if yes) (1)
Executive, managers	Any	0
Professor, lecturer, teacher	Any	0
Other professional	Any	0
Sales representative	Any	0
Admin staff	Any	0
Factory worker	Any	1
Trabajador de casa particular	Any	1
Technician	Any	1
Unknown	Agriculture	1
Unknown	Natural Resources and mining	1
Unknown	Manufacturing	1
Unknown	Construction	1
Unknown	Utilities	1
Unknown	Retail trade	0
Unknown	Transportation, warehousing and telecommunications	1
Unknown	Service-Providing Industries	0
Unknown	Public administration	0
Unknown	Not specified	n.a.

Notes: This table reports workers' occupation, industry, and whether its combination implies the worker is considered a blue-collar (or manual) worker or not. If information is available on occupation and industry, I use worker's occupation to classified the worker as a blue-collar. If occupation is not available, I use workers' industry information. When neither occupation or industry is available, I drop observations for this worker. "n.a." stands for not applicable. This table is referenced in Section III.C.

Table A10: Estimated sick leave utilization. Linear contract.

Replacement rate	Average worker (1)	(2)	High compliance costs (3)	(4)	Low compliance costs (5)	(6)	High valuation of time (7)	(8)	Low valuation of time (9)	(10)
0	0.7427	1.1064	0.4256	0.6340	0.9143	1.3619	1.7764	2.1315	0.4083	0.8645
0.1	0.9201	1.2818	0.5273	0.7345	1.1326	1.5778	1.8474	2.2022	0.6820	1.1329
0.2	1.0964	1.4564	0.6283	0.8346	1.3497	1.7928	1.9183	2.2727	0.9522	1.3994
0.3	1.2719	1.6304	0.7289	0.9343	1.5656	2.0070	1.9891	2.3432	1.2199	1.6642
0.4	1.4465	1.8038	0.8290	1.0337	1.7807	2.2204	2.0599	2.4136	1.4858	1.9278
0.5	1.6206	1.9767	0.9287	1.1327	1.9949	2.4332	2.1306	2.4839	1.7502	2.1902
0.6	1.7940	2.1491	1.0281	1.2315	2.2084	2.6455	2.2012	2.5542	2.0134	2.4516
0.7	1.9669	2.3210	1.1271	1.3301	2.4212	2.8571	2.2717	2.6244	2.2755	2.7122
0.8	2.1393	2.4926	1.2260	1.4284	2.6335	3.0683	2.3422	2.6945	2.5366	2.9719
0.9	2.3113	2.6638	1.3245	1.5265	2.8452	3.2791	2.4126	2.7646	2.7969	3.2310
1	2.4829	2.8347	1.4229	1.6244	3.0564	3.4894	2.4829	2.8347	3.0564	3.4894
$\Delta_{b=1-b=0}$	1.7402	1.7283	0.9972	0.9904	2.1421	2.1275	0.7065	0.7031	2.6481	2.6250
Preference parameters										
$\frac{w}{\phi}$	0.7120	0.7120	0.7120	0.7120	0.7120	0.7120	0.2920	0.2920	0.8746	0.8746
κ	2.5949	2.5949	1.4613	1.4613	3.2151	3.2151	2.5949	2.5949	2.5949	2.5949
Weekend dummy	0	1	0	1	0	1	0	1	0	1

Notes: This table reports estimated sick leave utilization, expressed in days relative to the worker health state—i.e., it reports the optimal deviation $(s^* - \theta)$—for alternative values of preference parameters. High compliance costs (low value of κ) correspond to the case where κ is calibrated to first quartile of its distribution. Low compliance costs (high value of κ) corresponds to the case where κ is calibrated to the third quartile of its distribution. Similarly, high (low) valuation of time corresponds to calibrating the ratio of wages to ϕ to the first (third) quartile. All the estimations assume a linear contract: $B(s) = bs \forall s$. The row labeled Δ presents the difference between the optimal choice with $b = 1$ and the optimal choice with $b = 0$: $s^*(b = 1) - s^*(b = 0)$. This table is referenced in Section V.A.

Appendix B. Derivations and Proofs

B.I The linear contract case

Assume that $B(s_c) = bs_c \forall s_c$. Worker i derives utility from her sick leave utilization choice s_c given the health state realization θ_i following:

$$u(c,s_l,s_c;\theta,dow) = w_i\left(M - ws_c(1-b)\right) + \phi_i\Big[(s_l-\theta) \\ - \frac{1}{2\kappa_i}(s_c-\theta)^{\eta_r}\mathbb{1}\{s_c>\theta\} - \frac{1}{2\kappa_i}(\theta-s_c)^{\eta_l}\mathbb{1}\{s_c<\theta\} + q\ \mathbb{1}\{\text{weekend}\}\Big],$$

The corresponding optimal sick leave utilization is defined by the following expression:

$$s^* - \theta = \left[\frac{2\kappa}{\eta_l}\left(\frac{\partial s_l}{\partial s_c} - \frac{w}{\phi}(1-b) + q\ \mathbb{1}\{\text{weekend}\}\right)\right]^{\frac{1}{\eta_l-1}}$$

where $\frac{\partial s_l}{\partial s_c}$ takes the values zero or one depending on whether the extra day on leave represents a business day or not.

B.II The sick leave contract space [33]

I assume that contracts are piece-wise linear, i.e., the marginal replacement rate b is constant within a sick leave duration bracket $[\underline{s}, \bar{s}]$. When solving to welfare maximization problem, I consider contracts of the following form:

1. Constant replacement rates for any duration: $B(s) = b\ \forall\ s$

2. Two-brackets contracts of the form:

$$\begin{aligned} B(s) &= b_1 \text{ for } [1, \tilde{s}] \\ &= b_2 \text{ for } s > \tilde{s} \end{aligned}$$

3. Three-brackets contracts of the form:

$$\begin{aligned} B(s) &= b_1 \text{ for } [1, \underline{s}] \\ &= b_2 \text{ for } (\underline{s}, \bar{s}] \\ &= b_3 \text{ for } s > \bar{s} \end{aligned}$$

[33] This sub-section is referenced in footnote 17.

The determination of the optimal replacement rate in the case of a contract with one bracket is straightforward: b is chosen to maximize social welfare. Determination of the optimal replacement rate in contracts with more than one bracket require defining the brackets' limits. I restrict the set of contracts to bracket lengths of at least three days and at most 15 days. That is, in the two-brackets' case these restrictions imply: $\tilde{s} \geq 3$ and $\tilde{s} \in \{3,4,5,6,\ldots,15\}$. In the three-brackets' case these restrictions imply: $\underline{s} \geq 3$, $\bar{s} \geq \underline{s} + 3$, $\underline{s} \in \{3,4,5,6,\ldots,15\}$, $\bar{s} \in \{6,7,8,\ldots,30\}$. Thus, there are 169 (i.e., 13^2) combinations of $\{\underline{s}, \bar{s}\}$ that characterize the structure of a three-bracket contract. For each of these combinations, I ask what are the optimal replacement rates.

Appendix C. Distribution of health states

Diagnoses included in the analysis

The Peruvian Handbook of Recovery Times specifies an average recovery time for 2,763 unique disease codes at the fourth digit level of the 10th revision of the ICD. This paper focuses on non-mental health conditions; excluding these diagnoses reduces the number of unique diseases to 2,690.[34] Estimating the model with such level of granularity is unfeasible: it would require estimating a probability for each disease and group of observable characteristics (2,690 diagnoses $\times$ 4 age group $\times$ 2 occupation groups = 21,520). For this reason, I group diagnoses in more aggregated categories.

I these categories considering (i) the type of diseases they represent and (ii) how frequently these diagnoses are used in the claims data. Table A6 lists the conditions included in the analysis by their ICD 10th revision group and the share of sick leave claim data with these diagnoses. To compute these shares, I used the sample constructed for the quantitative analysis of this paper. Table A4 provides details on sample construction.

The first criteria for excluding a group of conditions are those not listed in the Peruvian handbook. These are conditions originating in the perinatal period (codes in groups P00-P96) and congenital malformations, deformations, and chromosomal abnormalities (codes in groups Q00-Q99). In fact, 0.15% of the sick leave claim data is reported under these diagnoses. I dropped such observations.

The second criteria for excluding a group of conditions is the nature of the diagnosis which makes very challenging to assign a benchmark recovery time. I exclude poisonings and burns (codes in group T00-T98)—these diagnoses accumulate 1.22% of the sick leave claim data. Additionally, I exclude diseases coded under "special purposes codes" (codes U00-U85), external causes of morbidity (codes V00-Y99), and factor influencing health status and contact with health services (codes Z00-Z99). All these together represent 3.16% of the sick leave claims. These conditions associated with longer recovery times or impairments where full recovery might not be foreseeable, for example, leg amputations and organs transplants. Finally, I exclude conditions with diagnoses C00-D49; these codes are used for neoplasms, which in most cases, are chronic conditions or diseases that would require a longer recovery time. In terms of claims data, these represent 1.78% of the claims. The final sample includes about 86% of all sick leave claims filed by private-sector male workers.

[34] I use codes F01-F99 to define mental health conditions, these are grouped under the chapter "Mental, Behavioral and Neurodevelopmental disorders".