

On the Design of Paid Sick Leave: A Structural Approach

Victoria Barone*
University of California at Los Angeles

October 7, 2022

JOB MARKET PAPER
[\[Click here for the latest version\]](#)

Abstract

This paper studies the optimal design of a paid sick leave system. To do so, I develop a model of sick pay provision where risk-averse workers face health risks and decide their sick pay coverage and utilization. These decisions are affected by workers' health and the generosity of paid sick leave. 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. Given these responses, a social planner provides sick leave coverage to maximize social welfare, which is a function of workers' welfare and the productivity losses associated with sick leave provision. I obtain estimates of the model's parameters exploiting individual-level data on sick pay utilization and labor supply from the Chilean Paid Sick Leave System. I leverage exogenous variation in the incentives for shirking behavior induced by the day of the week a sick leave claim is filed. Finally, I analyze the impact on welfare and behavior of implementing the optimal contract.

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, Adriana Lleras-Muney, Daniel Haanwinckel, 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 sick pay provision that illustrates how workers respond to a given paid sick leave program and decide their labor supply and sick leave utilization. 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. 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 sick pay utilization and labor supply from the Chilean Paid Sick

¹While 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 one's ability to work for a short period and has a foreseeable recovery.

²While 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).

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, facing a disutility cost of doing so and reducing aggregate welfare. Thus, the optimal level of benefits depends on the behavioral responses of workers' to the benefit design, the costs of sick work, and production losses—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. In fact, the optimal policy feature lower replacement rates for shorter absences and higher benefits for longer spells.

The empirical section of this paper leverages data from the Chilean Paid Sick Leave

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

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.⁴ 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.⁵ 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.⁶

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

⁴This 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.

⁵These 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).

⁶Panel (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).

occupation from the claims data combined with data from the Peruvian Handbook of Recovery Times to recover the observed component of the health state.⁷ 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 without paid leave.

Combining the model of sick pay provision with the estimated parameters, I derived the optimal paid sick leave system. This system features a low replacement rate for absences of up to 3 days, with an average replacement of about 40% of forgone wages. Relative to the current Chilean system, that offers no protection for sick leave claims of

⁷The 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.

up to 3 days, the optimal policy is more generous. This reflects welfare gains from providing protection even for small losses. Most of these gains are explained by workers who, when sick, were on the margin between asking for some days off and going to work, but find optimal to work under the Chilean system. Benefits are increasing for absences of longer duration but the optimal system does not feature full replacement. Additionally, the optimal system, while allowing for three brackets, features smooth increases on the replacement rate which help to contain moral hazard. That is, by providing more cost-sharing the optimal policy maximizes welfare and prevents workers from extending their absences in order to be on the full insurance region.

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 workers' compensation and disability insurance are informative for paid sick leave policies—these three programs condition benefits in a difficult to verify state. Nonetheless, the health shock insured by sick pay—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.

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. Exploit-

ing 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.⁸ 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 present the model of paid sick leave provision that I use to derive the optimal sick leave insurance contract. First, I model the choices of an expected utility maximizer worker who faces uncertainty about her health and her ability to work. Second, I describe how workers' choices and provision of sick pay affect production. Third, I propose a measure of aggregate welfare. In the rest of 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) their distribution of health shocks, (ii) preferences over leisure, and (iii) their degree of risk-aversion.

⁸Cronin, 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.

II.A Workers

I consider a static model of work-absence behavior. I start by presenting how workers decide their sick pay utilization. This decision is determined by workers' preferences over their valuation of time outside work ϕ and their propensity for moral hazard (κ). Then, I describe how workers' risk preferences affect their expected utility.

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 discrete and bounded between zero and M .⁹ 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.

Sick pay utilization. At the time of her utilization choice a worker knows her health realization θ . I assume that 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.¹⁰ 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

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

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 $s^*(\theta, \phi, \kappa) = \arg\max_s u(c(s), s; \theta, \phi, \kappa)$.

⁹The 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.

¹⁰I 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.

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 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”.¹¹

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.¹²

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

¹¹This 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).

¹²The 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.

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 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 .¹³ 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.¹⁴

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

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

¹³In this case, workers care about the effective valuation of time, i.e., $\frac{w(1-b)}{\phi}$.

¹⁴I assume a regular work schedule from Monday to Friday.

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 $\mathbb{1}\{\text{weekend}\}$ 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 $dow = \text{Friday}$, a sick leave claim of 4 days corresponds to values of s_c and s_l of 4 and 2, respectively. But if $dow = \text{Monday}$, 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 $s_c^*(\theta, dow) = \arg\max_{s_c} (u(c(s_c), s_l(s_c), s_c; \theta, dow))$.

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.

Expected utility. Ex ante, a worker aims to maximize her expected utility, taken over the distribution of health shocks $G(\theta, dow)$. I assume that the worker is risk averse 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 $u^*(\theta, dow)$. To see why this is the case, note that the utility maximization problem can be viewed as a two-stage problem (Einav et al., 2013). Once the health shock is realized, risk preferences become irrelevant and workers aim to maximize the contribution of the state-utility ($u^*(\theta, dow)$) to their expected utility ($E[v(u^*(\theta, dow))]$). Thus, expected utility is given by

$$U(\theta, dow, \phi, \kappa) = E[v(u^*(\theta, dow))] = \int v(u^*(\theta, dow)) dG(\theta, dow) ,$$

where $G(\theta, dow)$ is the joint distribution of health states and days of the week. 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 Production

Sick leave insurance provision affects workers’ incentives. Given a level of coverage, on any given day workers could decide to (i) show up to work sick or (ii) be absent from

work. Thus, changes to the generosity of sick pay could affect the pool of workers who are showing up to work or who are absent. Then, the design of an optimal policy requires incorporating this channel into aggregated welfare. To do so, I assume that there is a firm in the economy that hires workers and produces a consumption good. I describe how this firm operates next.

I assume that sickness is detrimental for worker's productivity. That is, when sick, a worker is less productive than in her healthy state. Nonetheless, the firm can not adjust workers' pay based on their health state. For each worker, I define "effective days worked" as the sum of days worked when healthy ($d_{healthy}$) and days worked when sick (d_{sick}) adjusted by the parameter ν that reflects the lost productivity from sickness. That is, effective days worked (d_e) are: $d_e = d_{healthy} + \nu d_{sick}$. Using these definitions, the profits generated by worker i are:

$$\pi(s_c^*) = g(d_{sick}(s_c^*) + \nu d_{healthy}(s_c^*)) - w(d_{healthy}(s_c^*) + d_{sick}(s_c^*))$$

I assume that the function $g(\cdot)$ features diminishing returns on effective days worked, that capital is fixed, and normalized the price of the good to one. While this is a very stylized version of the model of firm behavior, it allows to capture the production losses (or gains) of changes in the sick pay policy.

II.C Social Planner

In this section I discuss the definition of social welfare and how the social planner chooses the optimal sick pay contract. For this exposition, it is helpful to consider an economy populated by I workers and let $U_i(\theta_i, d_{ow}_i)$ represent the expected utility of worker i . This paper's main question is what is the policy $(B_n(s), \tau_n)$ 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.

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

The social planner chooses $B_n(s)$ and τ_n to maximize the sum of individual welfare:

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

where I assume that the social planner assigns equal weights to each worker. S represents the cost of the current Chilean system, i.e., $S = \sum_i s_i^*(B^{chi}(s))$. This constraint allows comparison across policies that have the same implementation cost. The set of contracts that I consider are the ones characterized by the following transfer function:

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

This function features a deductible of \underline{s} days and it can reproduce the current Chilean system. Additionally I assume that transfers are non-decreasing, i.e., $B(s+1) \geq B(s)$ and the system is at most as generous as the full coverage case: $B(s) \leq s$.

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.

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.¹⁵ 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

¹⁵These are called FONASA, for the Spanish name—*Fondo Nacional de Salud*—and IS-APRES—*Instituciones de Salud Previsional*—respectively.

covers healthcare in private healthcare institutions with high copayments.¹⁶ 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 providers 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 enrolled 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, 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.

¹⁶Plans 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.

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.¹⁷ 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.¹⁸ 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 (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.¹⁹ This non-payable period works like a deductible that resets for every new sick leave span.²⁰ 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

¹⁷The 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.

¹⁸These restrictions are independent of job tenure, which reflects that wage replacements are financed with employees' contributions to the healthcare system and not by employers.

¹⁹Cid (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.

²⁰These resettable deductibles are similar to those used in automobile or homeowners insurance: Separate deductibles apply to each loss.

11-day or longer duration are fully covered.²¹ 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 health-care 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 Appendix 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 pattern 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 every individual enrolled in the public insurance system, including sex, age, ZIP code, health insurance plan, annual earnings, and health indicators for chronic conditions.²² These data represent the universe of individuals enrolled in the government-run health-care 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

²¹If 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.

²²These 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.

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.²³ A second criterion is the recovery time associated with the primary diagnosis. For example, I exclude 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.

²³For example, these claims must be certified by a psychiatrist and require a comprehensive medical assessment at the time of filing.

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.²⁴ 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

²⁴Blue-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.

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 and wages. In the theoretical model, θ_i indicates the number of days a worker is sick and dow_i is the day a sick leave claim is filed. To construct the empirical joint distribution of $G(\theta, dow)$, I exploit the detailed data on sick leave claims. First, the day of the week a sick leave claim is filed is observed. I assume that workers can file sick leave claims only during week days. This assumption is based on two empirical facts, first only 6% of sick leave claims are filed during the weekend. Second, about 83% of workers in Chile work Monday to Friday. So, day of the week in the model reflect the day a worker files a sick leave claim rather than the day she falls sick. For this reason, when I estimate the model I use the probability of filing a sick leave claim on a given day as observed in the data. To construct the duration of a health shock I combine observed diagnoses (4 digits of ICD 10th) with detail data from the Peruvian Handbook of Recovery Times.

I allow for correlation between the health shocks and workers' wages to capture two empirical facts: (i) as shown on Tables 1 and 2, the duration of days on leave varies with income; (ii) the diagnoses prevalence changes with age and occupation of the workers. Thus, when estimating the model, I sample workers' wages and a vector of health shocks (θ, dow) allowing for any correlation between these variables.

Preference parameters. Worker i derives the following utility from her sick leave utiliza-

tion choice s_c given the health state realization θ_i :

$$u(s_{c,i}; \theta_i, \text{dow}_i) = c(B(s_{c,i})) + \phi_i \cdot \left[(s_{li}(s_{c,i}, \text{dow}_i) - \theta_i) + q \mathbb{1}\{\text{weekend}\} \right. \\ \left. - \kappa_{0i}(s_{c,i} - \theta_i)^2 \mathbb{1}\{s_{c,i} > \theta_i\} - \kappa_1 \mathbb{1}\{s_{c,i} = \theta_i + 1\} - \kappa_2 \mathbb{1}\{s_{c,i} = \theta_i + 2\} - \kappa_3 \mathbb{1}\{s_{c,i} = \theta_i + 3\} \right]$$

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_ϕ^2 such that

$$\log(\phi_i) \sim N(\mu_\phi, \sigma_\phi^2) .$$

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_κ^2 :

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

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 :

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

where γ is the constant relative risk-aversion coefficient, and the empirical analogue of $G(\theta, dow)$ is estimated non-parametrically from the claims data. 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.²⁵ 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 an additive error, that has mean zero and is uncorrelated with the “true” sickness level. That is, I 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_{me} , 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 to prescribe rest that are multiples of seven days. I take a non-parametric approach to implement this adjustment and simply assume that with some probability p_7 a sick leave claim of duration x is rounded up (dow) to 7 days.

²⁵While 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.

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 and earnings. I exploit the granularity of the Chilean data and take a sample of health states (duration of a health shock and days of the week) and wages. The duration of a health shock is not observed, to construct it, 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.

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.²⁶

Structurally estimated parameters. $q, \mu_\phi, \sigma_\phi^2, \mu_{\kappa_0}, \sigma_{\kappa_0}^2, \kappa_1, \kappa_2, \kappa_3, p_m^e, p_7$ I estimate the re-

²⁶I 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.

maintaining preference parameters, and the parameters that govern measurement error and rounding 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 $\phi q 1\{\text{weekend}\}$ 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 latter 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 filed that day.²⁷ That is:

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

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.²⁸

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

²⁷This computation excludes sick leave claims filed in a week of a national holiday.

²⁸Claims 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 Figure A4 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.

Figure 4 shows this moment graphically and Table 5 presents detailed computations.

Compliance costs $(\mu_\kappa, \sigma_\kappa^2, \eta_r, \eta_l)$: 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 :

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

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

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

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 func-

tion 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_ϕ^2): 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:

$$LC_i = \frac{\text{leisure}}{\text{consumption}_i} = \frac{1}{N_i} \sum_m \frac{w_{i,m} \times \text{Days on leave}_{i,m}}{w_{i,m} \times \text{Days worked}_{i,m} + \text{Sick pay}_{i,m}},$$

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 .

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.

VI The optimal paid sick leave contract

VII Conclusions

This paper studies the incentive problem created by paid sick leave insurance. To do so, it first develop a theoretical model of sick pay utilization. This model shows that workers choose sick pay utilization trading off the cost of a day away from work—lost consump-

tion—with the net gain in terms of leisure time—i.e., time outside work increases utility but it is costly. I quantify this trade-off empirically exploiting administrative data from the Chilean Paid Sick Leave System. This system insures all workers in Chile and provides income replacement for absences due to non-work related sickness. To estimate the model, I exploit variation in the temptation from shirking across days of the week. That is, I compare the choices of workers with comparable characteristics that fall sick on different days of the week. The model replicates salient features of reality internally and externally.

I find that moving the average worker from a non-paid to a sick leave contract with full-coverage induces sick pay utilization in 1.73 days. Policy evaluations suggest that shorter non-payable periods prevent working sick behavior and can reduce the spread of diseases in the work place. Nonetheless, these systems are more costly, they offer greater coverage. Additionally, discontinuities in the replacement rate generate behavioral responses: workers tend to bunch and extend sick pay utilization, resulting in inefficient sick leave taking.

This paper addresses a relevant but poorly understood question in the provision of social insurance: the optimal design of paid sick leave. When designing such programs, policymakers face several choices: What level of benefits should be paid? Should benefits rise or fall over the utilization span? How should the programs be financed? This paper's main contribution is to answer these questions and provide evidence on how the adoption (or modification) of a paid sick leave system could help to preserve workers' health. This paper speaks directly to the rules regarding generosity of sick pay and cost-sharing mechanisms that would provide enough coverage and limit shrinking behavior.

The empirical application of this paper exploits the Chilean context, but the insights of my model and results are informative for the United States context. Currently, many US workers do not have access to paid leave time, and there is substantial inequality in coverage across jobs. For instance, 97% of private-sector employees in the finance and insurance industry have access to paid sick leave while 41% of employees in the accommodation and food services industry have access to paid sick leave (Maclean, Pichler and Ziebarth, 2020). While there is no federal mandate of paid leave, there has been traction on paid sick leave policies at the state and local levels. Over the past decade, thirty-six US jurisdictions have implemented sick pay mandates enacting laws that enable eligible workers to accrue paid time off from their employer (A Better Balance, 2021). Additionally, opinion polls show large bipartisan support for mandating paid sick leave.

This paper has the potential to inform the policy discussion in two ways. First, it quantifies the main trade-off faced by policymakers when designing a paid sick leave system.

This trade-off applies to the US context: more generous sick pay contracts could reduce the number of workers showing up sick and help preserve workers' health at the cost of increased absenteeism, which entails product losses and increases program costs. Second, this paper shows that traditional cost-sharing rules—such as waiting periods—could contain moral-hazard behaviors, but contagious diseases undermine their effectiveness. Sick workers could spread diseases in the workplace, deteriorating public health. These externalities would justify more generous sick pay and highlight the relevance of enforcement mechanisms such as penalties for unjustified absenteeism.

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.

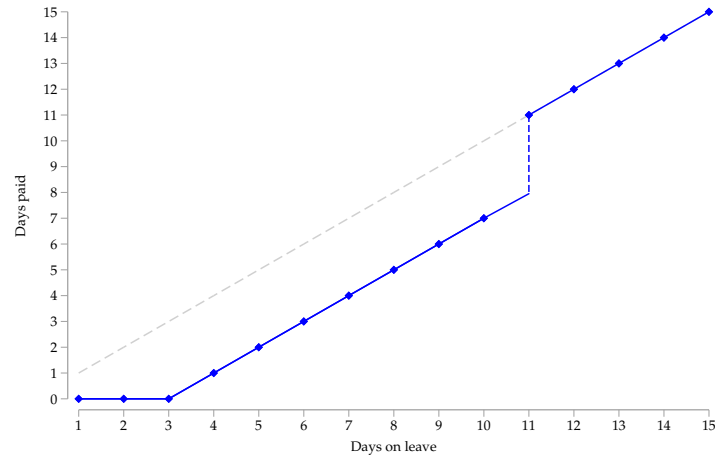
- Handel, Ben, Igal Hendel, and Michael D Whinston.** 2015. "Equilibria in health exchanges: Adverse selection versus reclassification risk." *Econometrica*, 83(4): 1261–1313.
- Hendren, Nathaniel.** 2017. "Knowledge of Future Job Loss and Implications for Unemployment Insurance." *American Economic 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.
- 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.

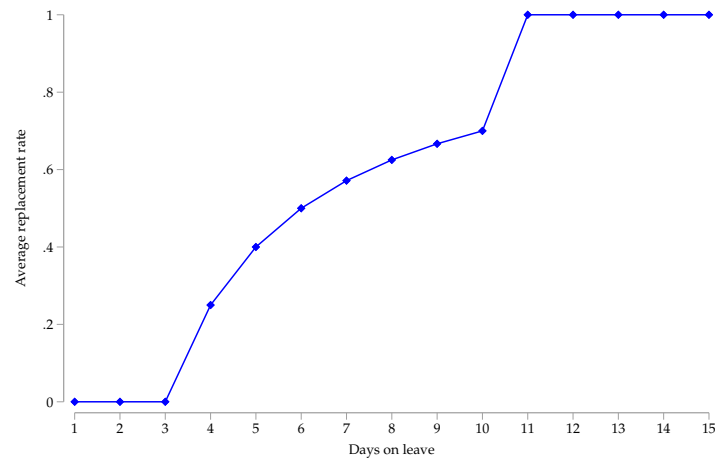
VIII Figures

Figure 1: Chilean Paid Sick Leave System

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

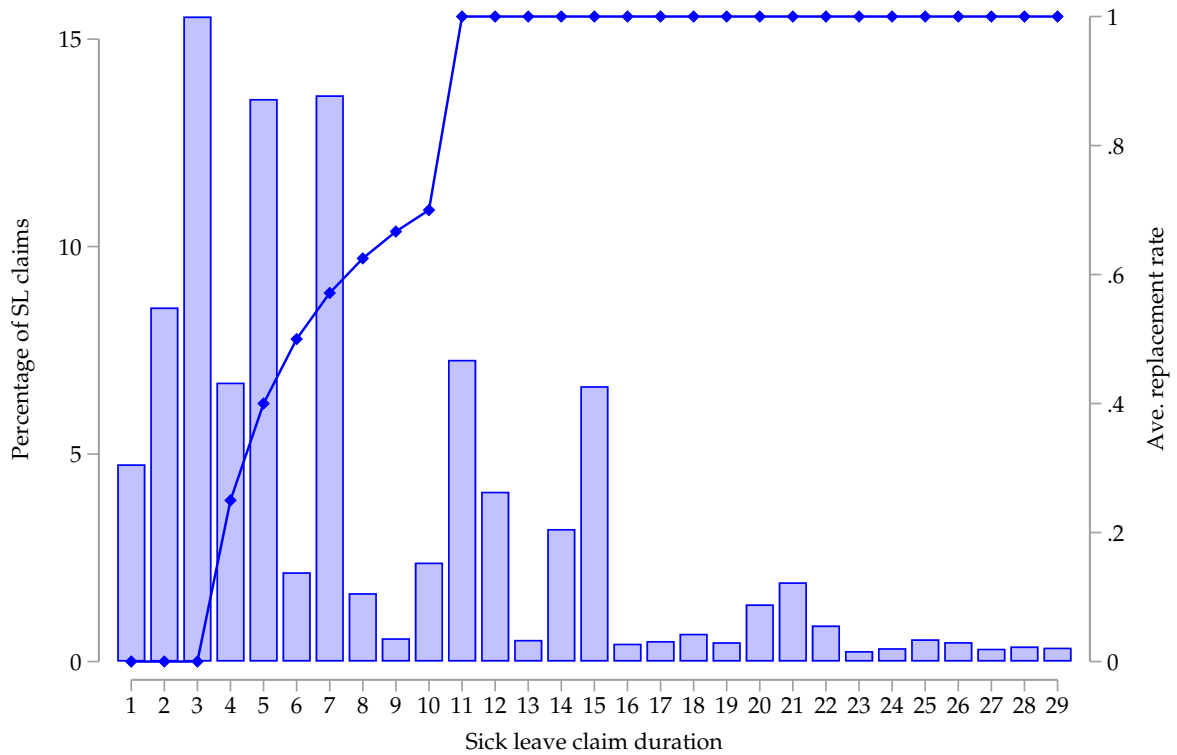


(b) Average replacement rate



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.

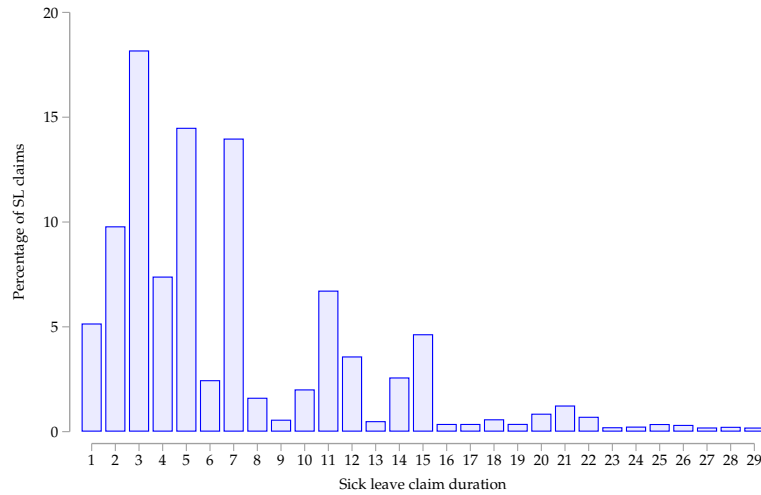
Figure 2: Duration of sick leave claims. Private-sector male workers



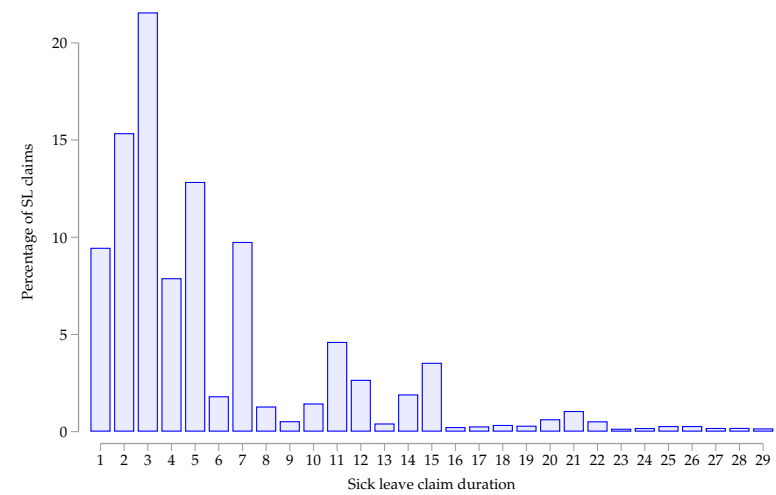
Notes: This figure shows the distribution of the duration of sick leave claims from male workers on the left-hand-side vertical axis and the average replacement rate on the right-hand-side vertical axis. The figure only includes sick leave claims of up to 29 days long, these represent 79% of all claims.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

Figure 3: Histogram of days on leave by workers characteristics

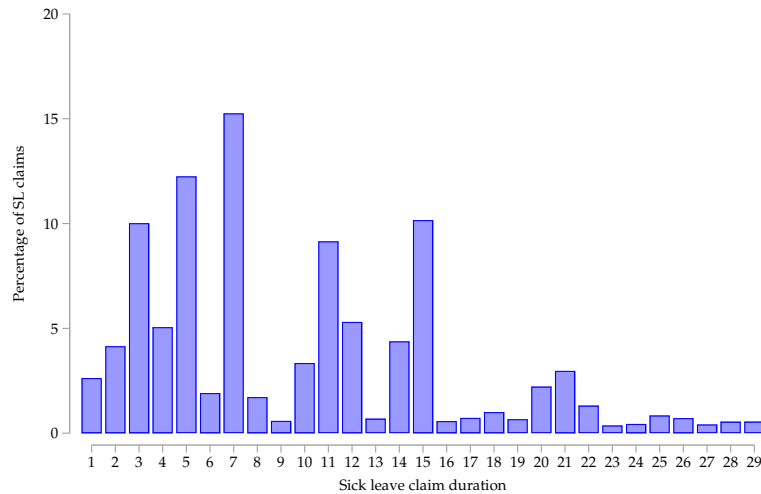
(a) 25-34 years old. Blue-collar



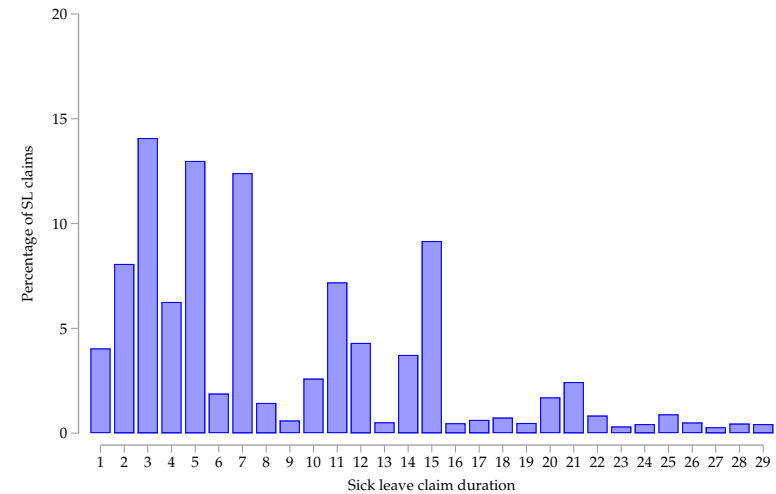
(b) 25-34 years old. White collar



(c) 55-64 years old. Blue-collar

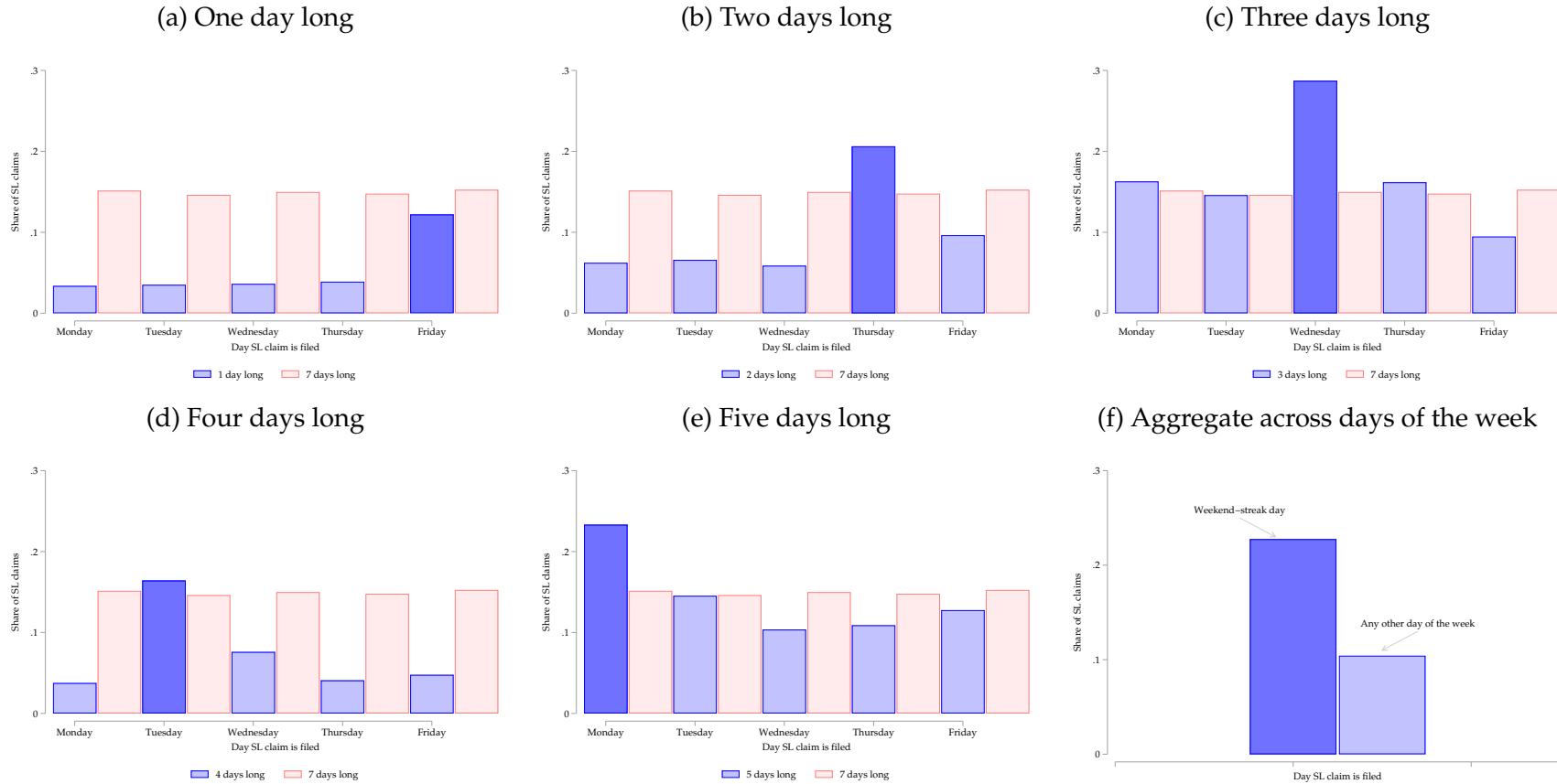


(d) 55-64 years old. White collar



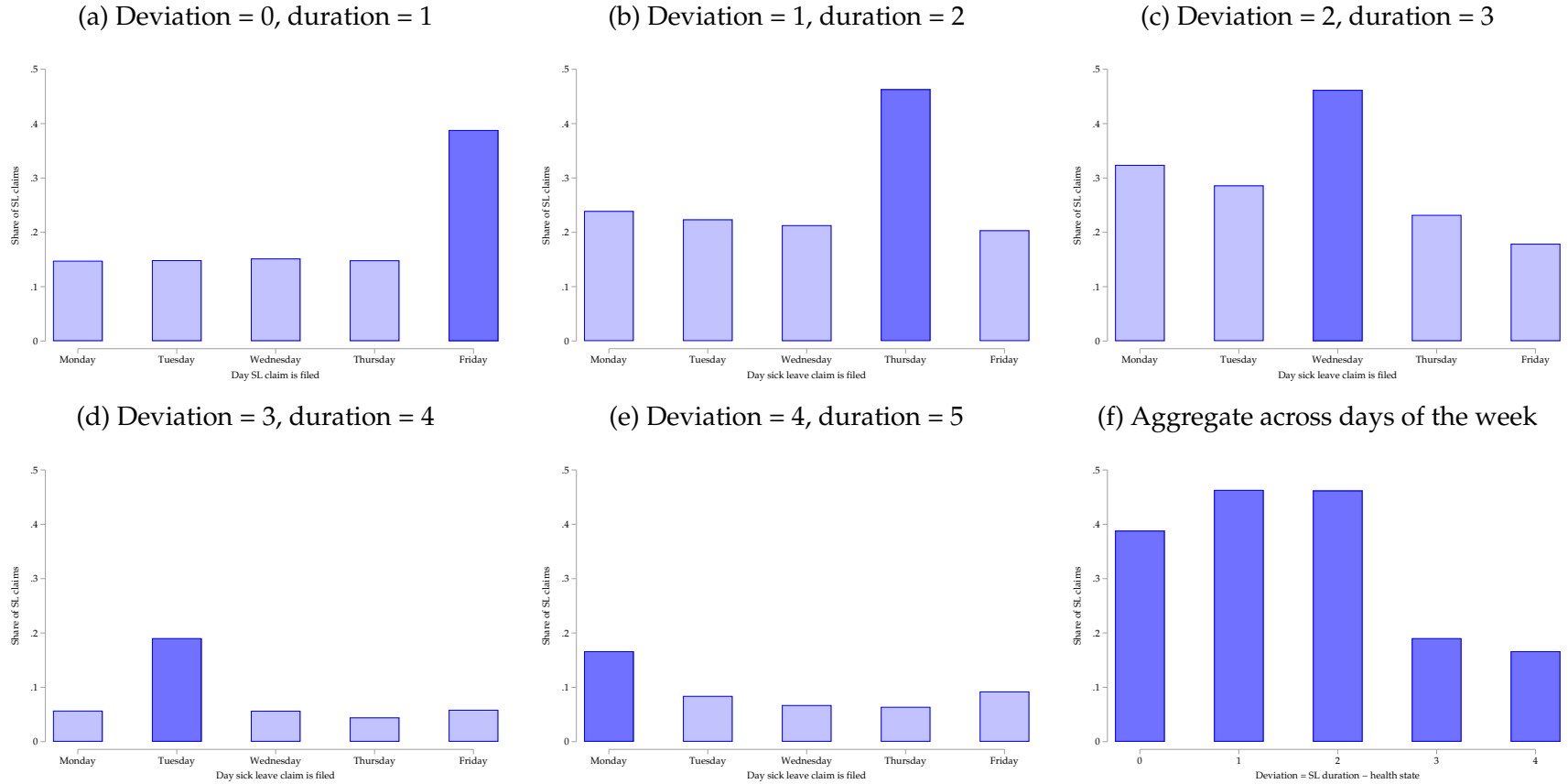
Notes: This figure shows the distribution of days on leave by workers' age and occupation for the youngest and oldest group of workers. Sample includes male private-sector employees. Blue-collar worker refers to workers who engage in hard manual labor, typically agriculture, manufacturing, construction, 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



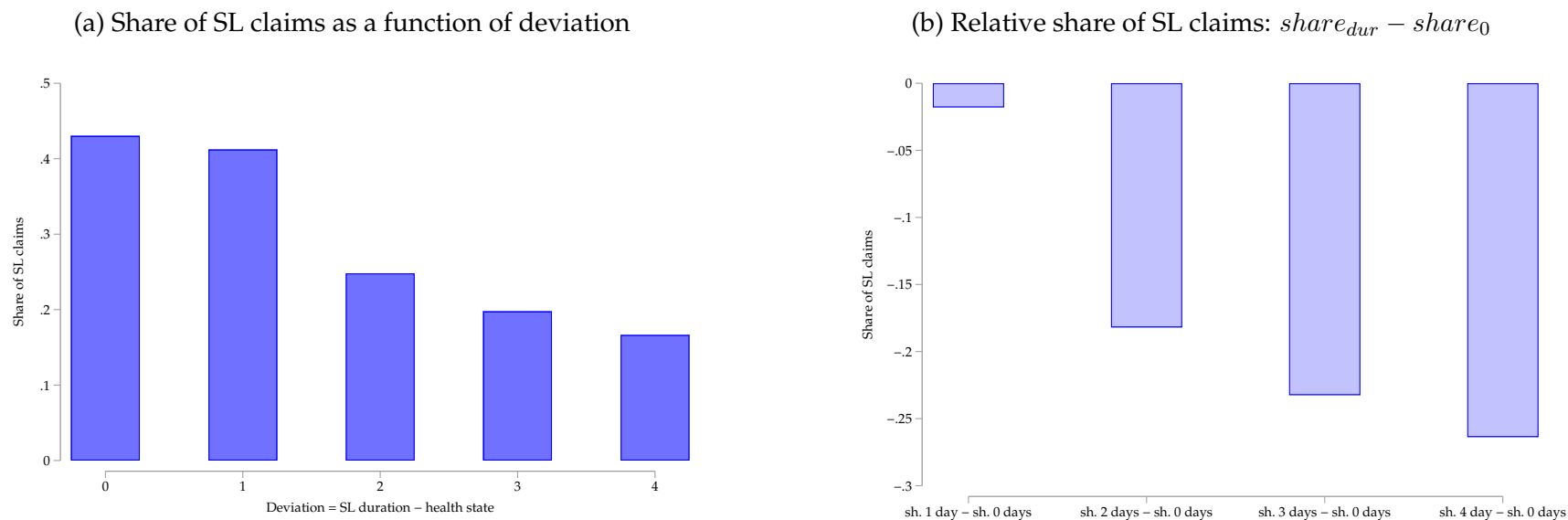
Notes: Panels (a) to (e) show the share of sick leave claims with duration s and the share of seven-days-long sick leave claims filed on each day of the week. Panel (f) aggregates across duration and days of the week: the first bar—labeled “weekend streak”—averages the share of one-to-five-days-long 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.



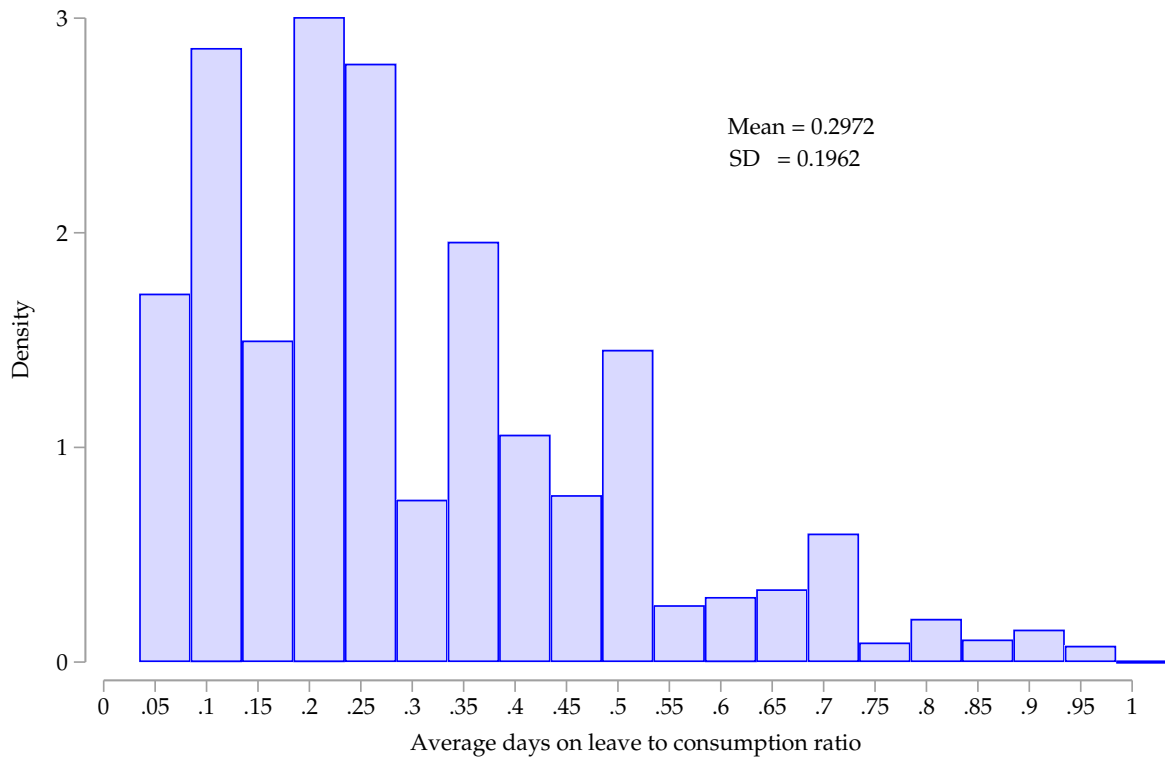
Notes: Panels (a) to (e) show the share of sick leave claims with duration s for workers whose main diagnose would implied a health state of 1 day on leave. Panel (f) aggregates the share of sick leave claims across days of the week, including only the weekend-streak combinations, e.g., from panel (a)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



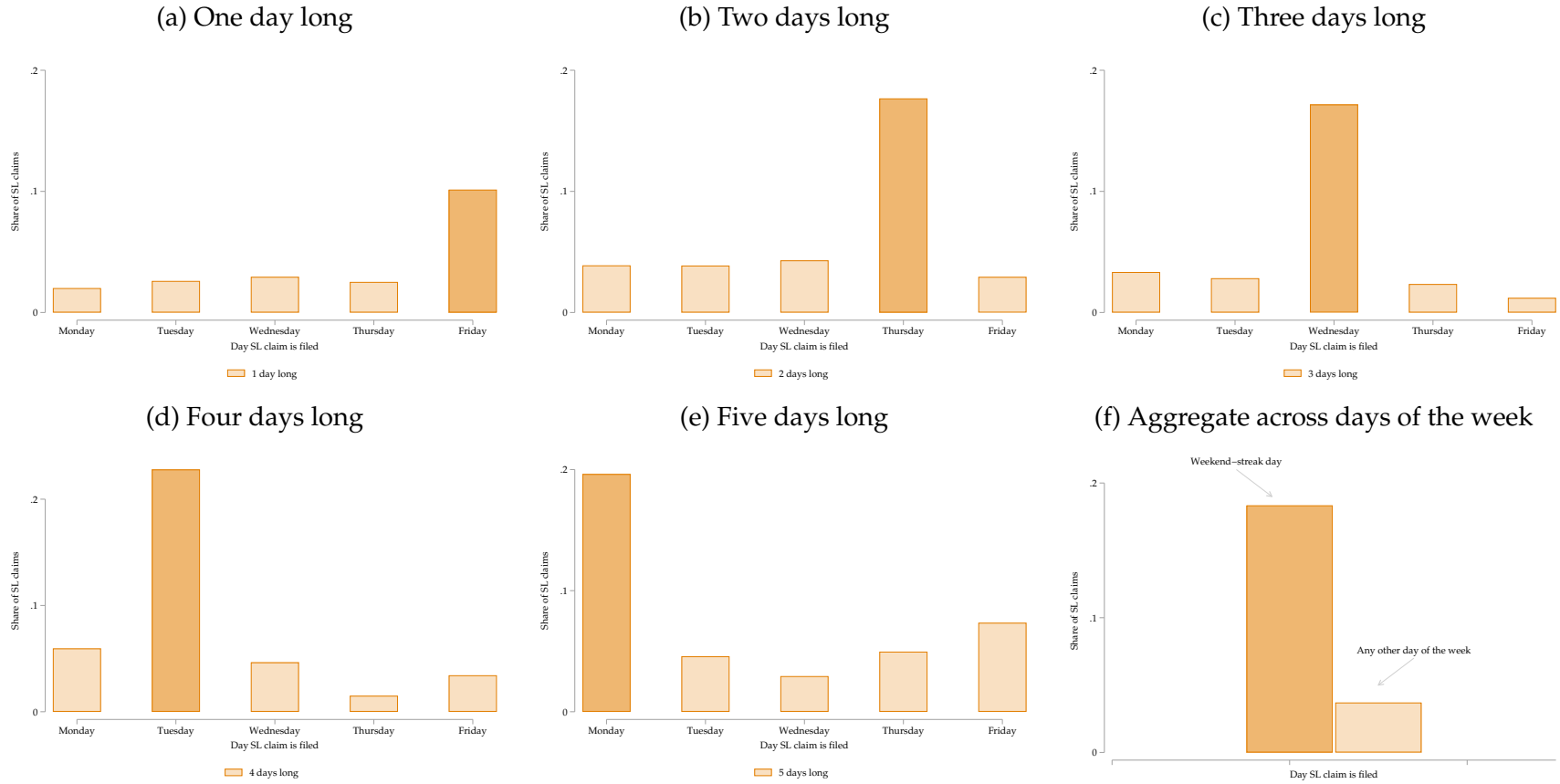
Notes: Panel (a) shows the average share of sick leave claims with deviations between 0 and 4 days. The average is computed over sick leave claims with primary diagnosis requiring 1, 2 or 3 days of rest filed on a weekend streak days. Each column is the weighted average of the probability for each health 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.

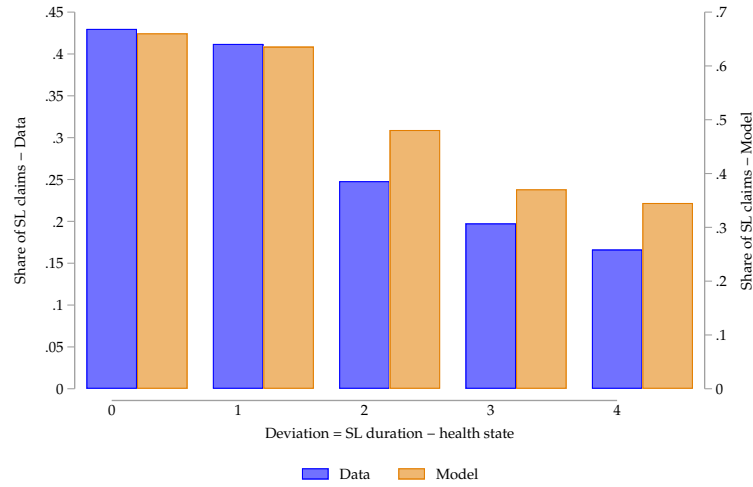
Figure 8: Model's prediction: weekend streak utility moments



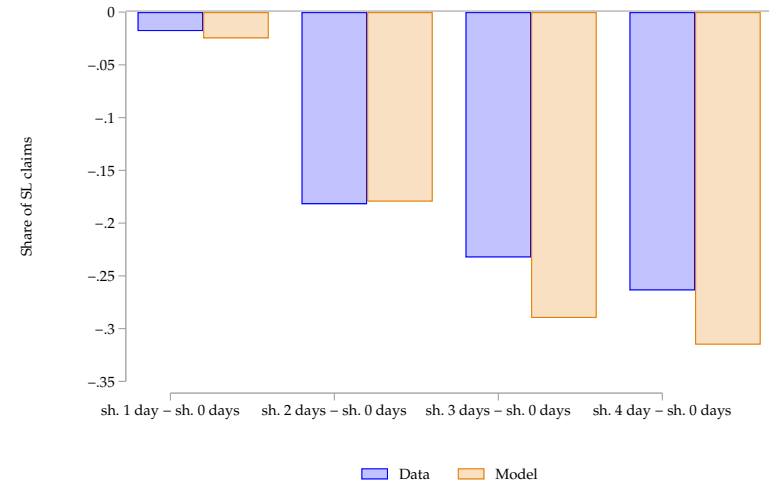
Notes: This figure presents the estimated-analogue of Figure 4. Panels (a) to (e) show the share of sick leave claims with duration s from a model-simulated sample. Panel (f) aggregates across duration and days of the week: the first bar—labeled “weekend streak”—averages the share of one-to-five-days-long 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

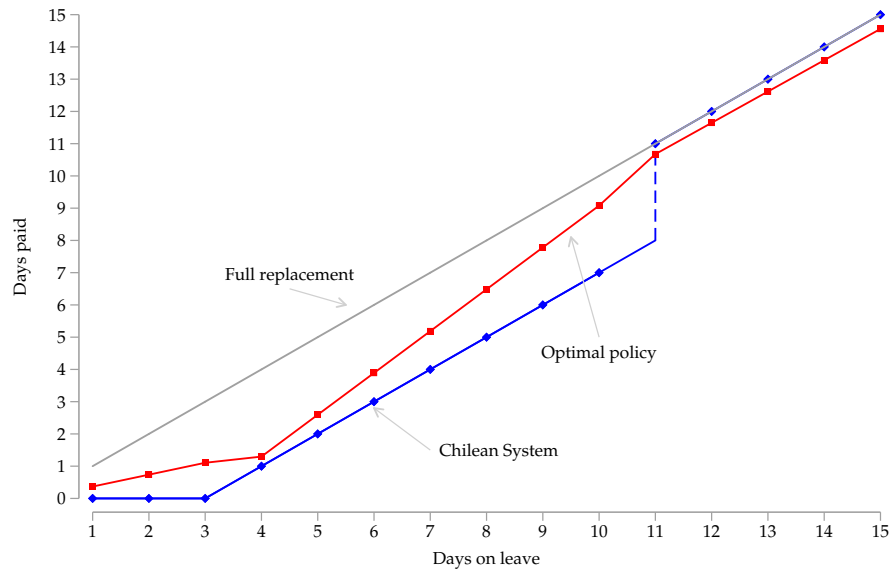


(b) Relative share of SL claims: $share_{dur} - share_0$



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: The Optimal Sick Pay System



Notes: This figure compares the Chilean system with the optimal system. This figure is referenced in Section VI.

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)
<i>Age</i>				
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
<i>Income (monthly USD)</i>				
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
<i>Region (%)</i>				
Central	34.97	40.76	41.22	41.92
Mining intensive	8.96	8.49	8.32	7.62
<i>Health - chronic conditions (%)</i>				
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)
<i>Age</i>				
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
<i>Income (monthly USD)</i>				
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
<i>Region (%)</i>				
Central	48.35	59.10	48.05	43.78
Mining intensive	6.01	4.09	5.30	7.89
<i>Health - chronic conditions (%)</i>				
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

ICD group	Main diagnoses	25 - 34 years old		35 - 44 years old		45 - 54 years old		55 - 64 years old	
		Blue c.	White c.	Blue c.	White c.	Blue c.	White c.	Blue c.	White c.
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(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 \hat{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		Difference
	Weekend streak	Non-weekend streak	
	(1)	(2)	
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: Parameters estimates

Parameter	Description	Value
<i>Preference parameters</i>		
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 positive deviations	2.0314
η_l	Curvature of cost function, curvature for negative deviations	3.9576
<i>Measurement error</i>		
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
<i>Rounding</i>		
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.

Table 8: Counterfactual Analysis

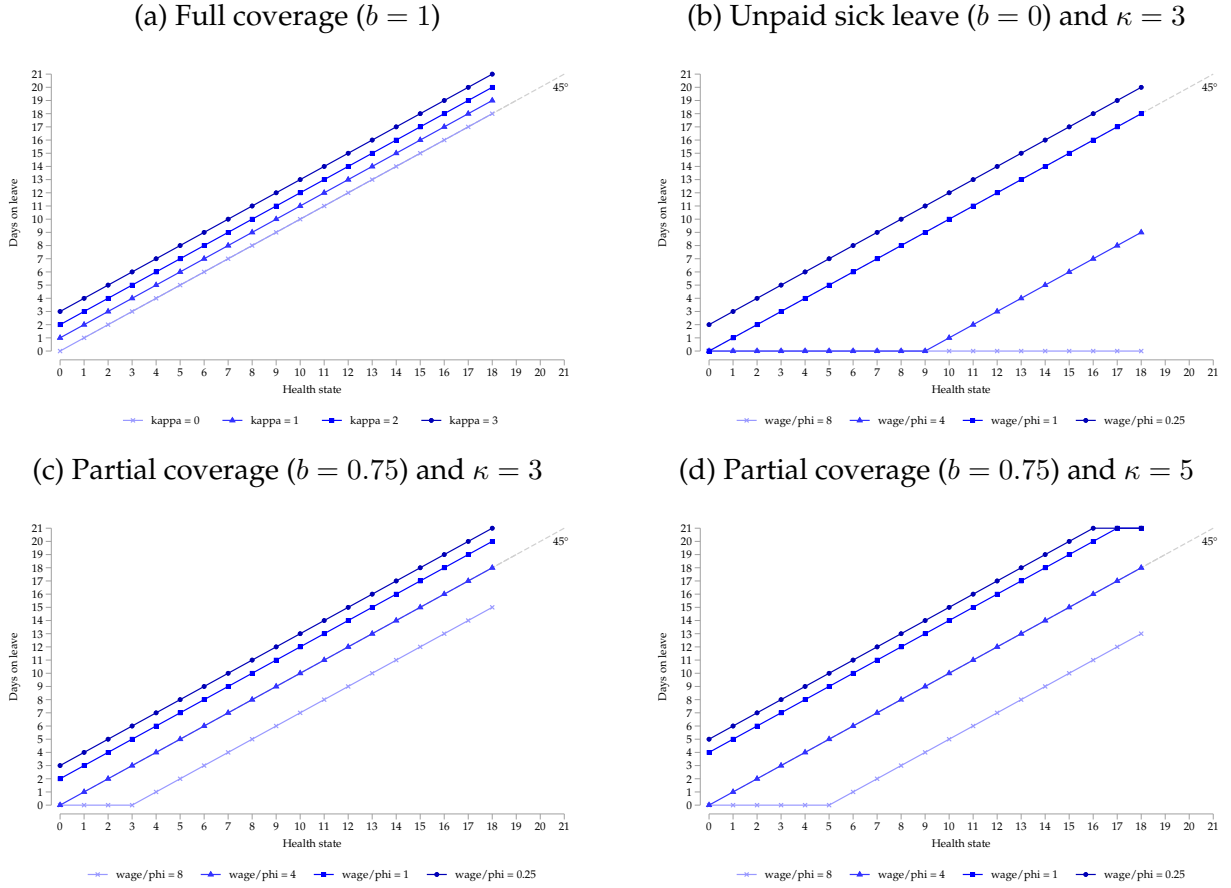
	Alternative systems				
	Chilean system	Eliminate jump at 11 days	Non-payment to 2 days	Constant rep. rate (b=0.8)	Spanish system
	(1)	(2)	(3)	(4)	(5)
<i>Relative to no insurance</i>					
Unjustified leave	1.85	1.35	1.46	1.07	0.71
Working sick (share)	0.18	0.20	0.09	0.13	0.46
Workers' welfare gain	34.24	27.64	21.06	33.64	15.79
<i>Relative to current system</i>					
Cost index	100	76.464	88.234	88.859	42.186
<i>Change in cost due to</i>					
Mechanical effect		19.02	8.06	5.65	51.19
Behavioral effect		4.51	3.71	5.49	6.63

Notes: This table presents the performance of the current Chilean and alternative paid sick leave systems. I consider the following alternative systems: (i) the current Chilean system without the reimbursement of the non-payable period at 11 days, i.e., without the discontinuity at 11 days, (ii) the current system without discontinuity at 11 days and a shorter non-payable period, (iii) a constant replacement with an average rate of 0.8 of forgone wages, and (iv) the Spanish system for private sector employees. First, I quantify the moral hazard responses induced by sick pay, relative to a situation with no insurance, i.e., relative to the non-paid sick leave case. Second, I compute the aggregate welfare gain of insured workers relative to the no insurance case. Third, I compute the cost of implementing each system relative to the current one. This table is referenced in Section VI.

Appendix A. Additional Figures and Tables

A.I Additional Figures

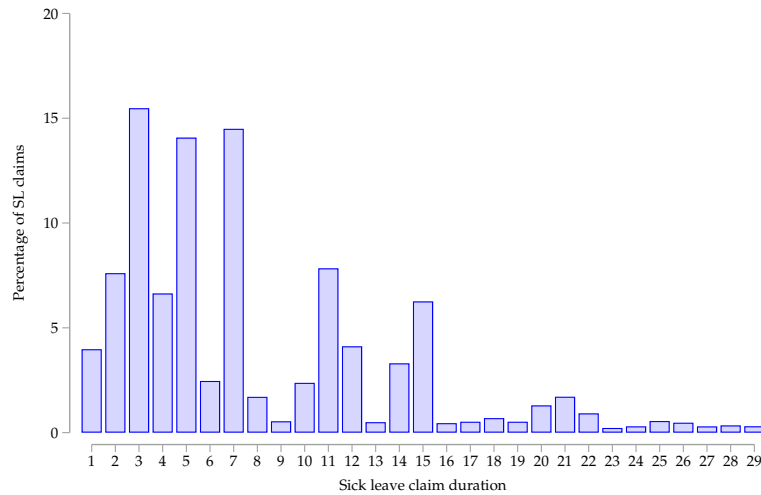
Figure A1: Sick Pay Utilization with Linear Contract



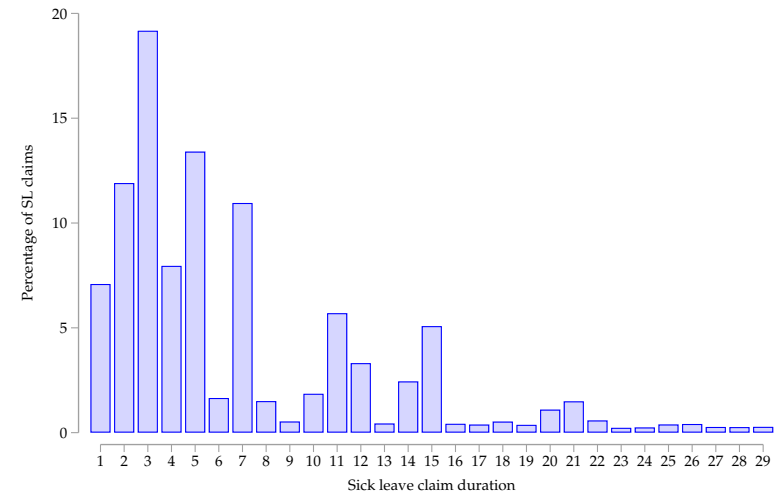
Notes: This figure shows the optimal demand of days on leave $s^*(\theta)$ as a function of worker's health status (θ) under the assumption of linear contracts with different levels of coverage. This figure is referenced in Section II.

Figure A2: Histogram of days on leave by workers characteristics

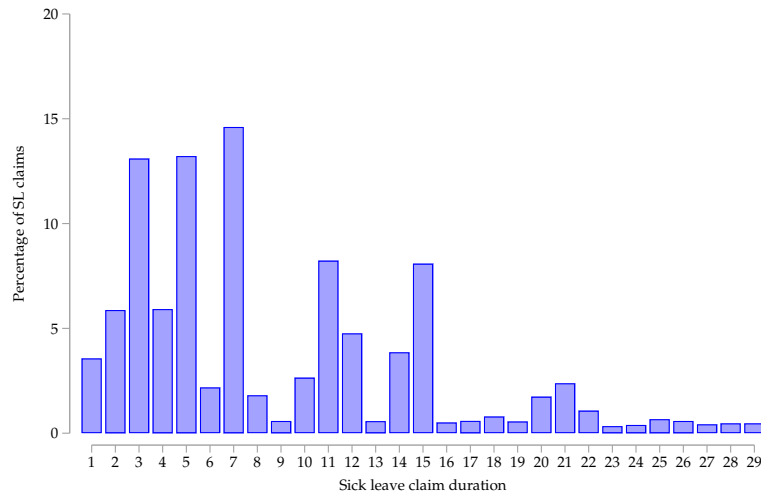
(a) 35-44 years old. Blue-collar



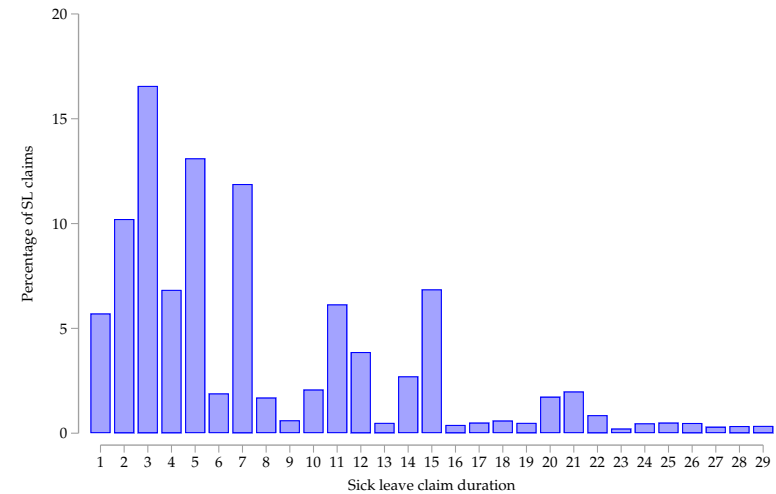
(b) 35-44 years old. White collar



(c) 45-54 years old. Blue-collar



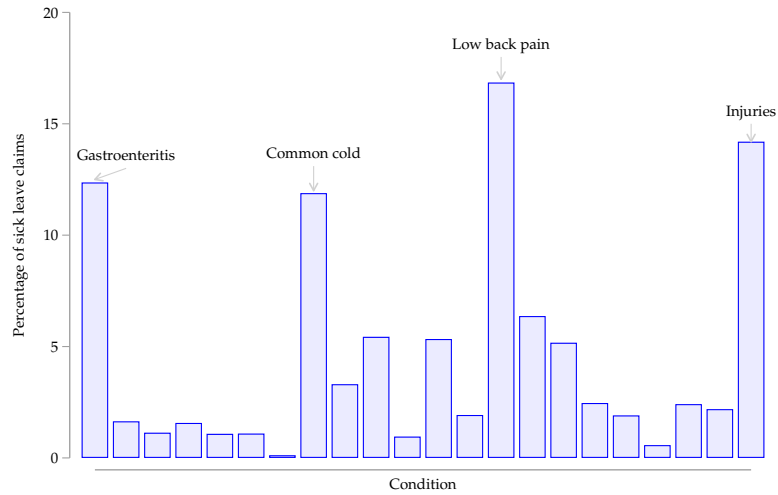
(d) 45-54 years old. White collar



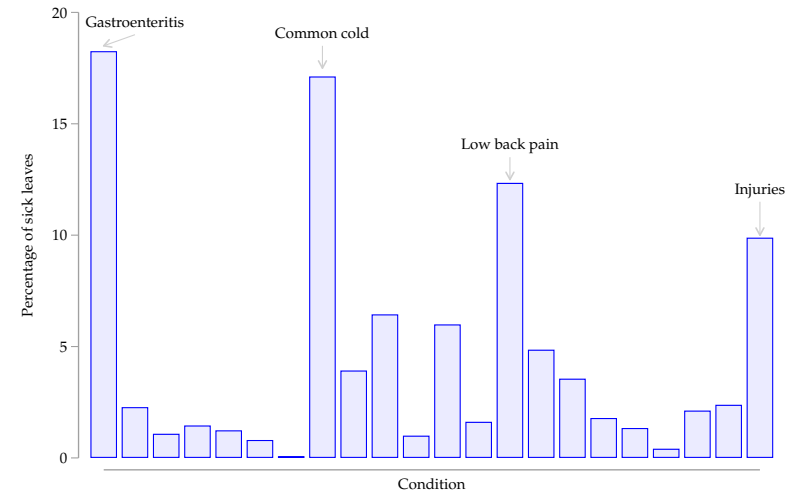
Notes: This figure shows the distribution of days on leave by workers' age and occupation for the youngest and oldest group of workers. Sample includes male private-sector employees. Blue-collar worker refers to workers who engage in hard manual labor, typically agriculture, manufacturing, construction, 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

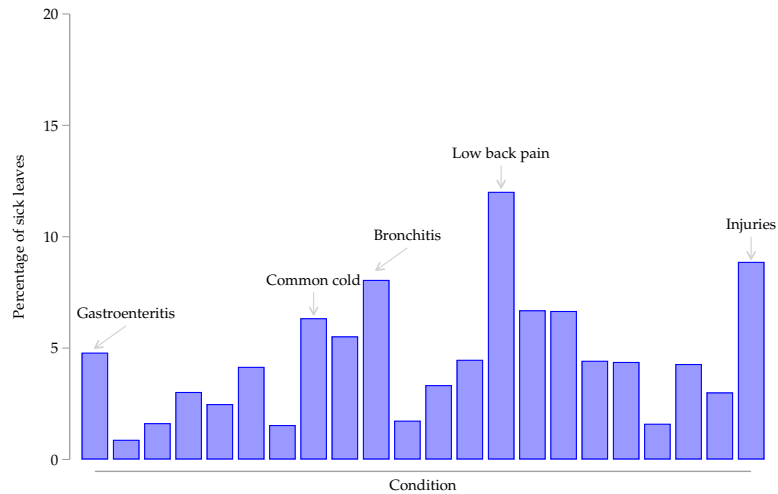
(a) 25-34 years old. Blue-collar



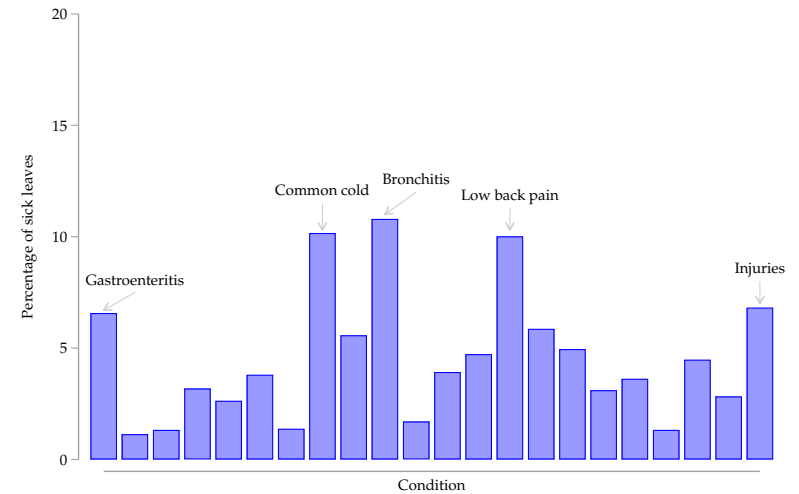
(b) 25-34 years old. White collar



(c) 55-64 years old. Blue-collar

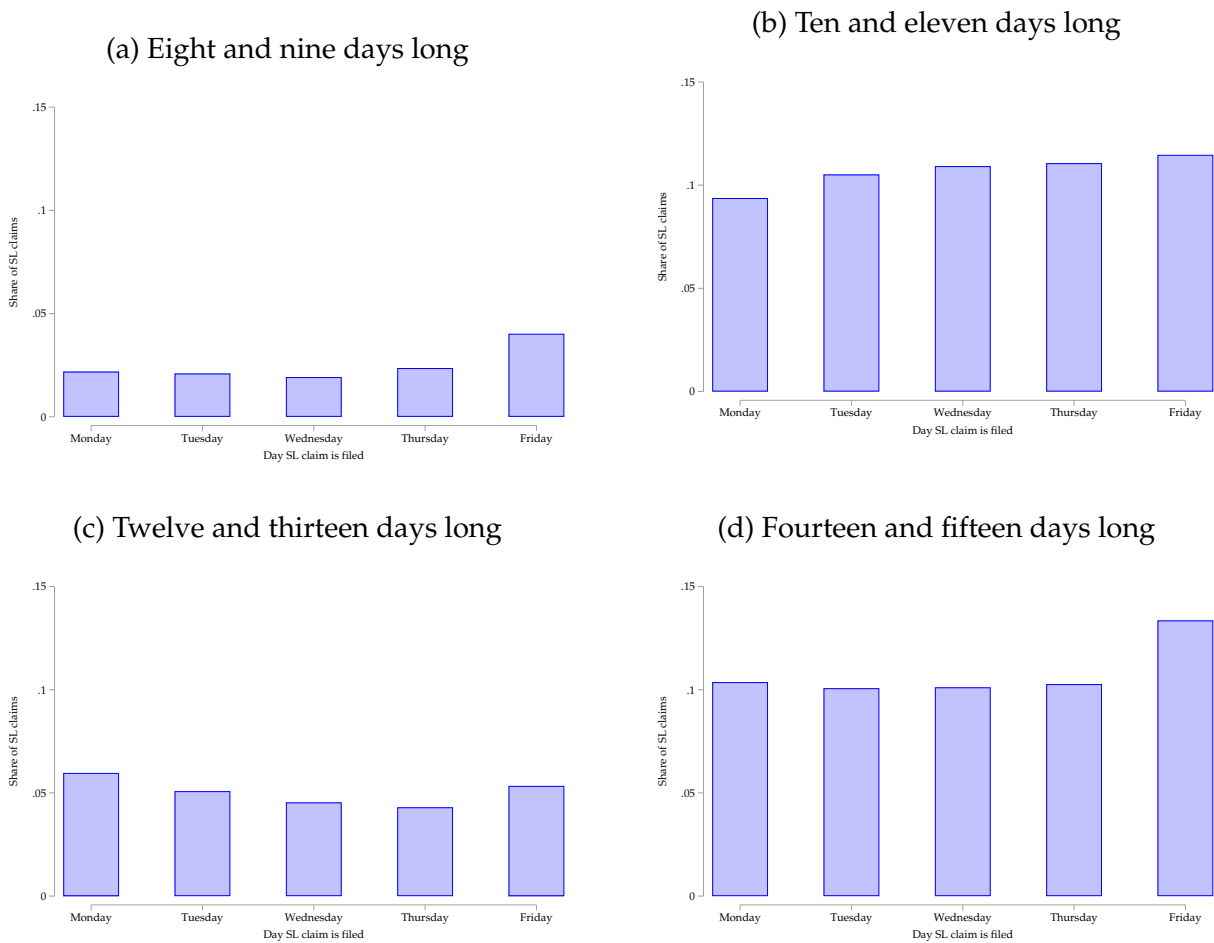


(d) 55-64 years old. White collar



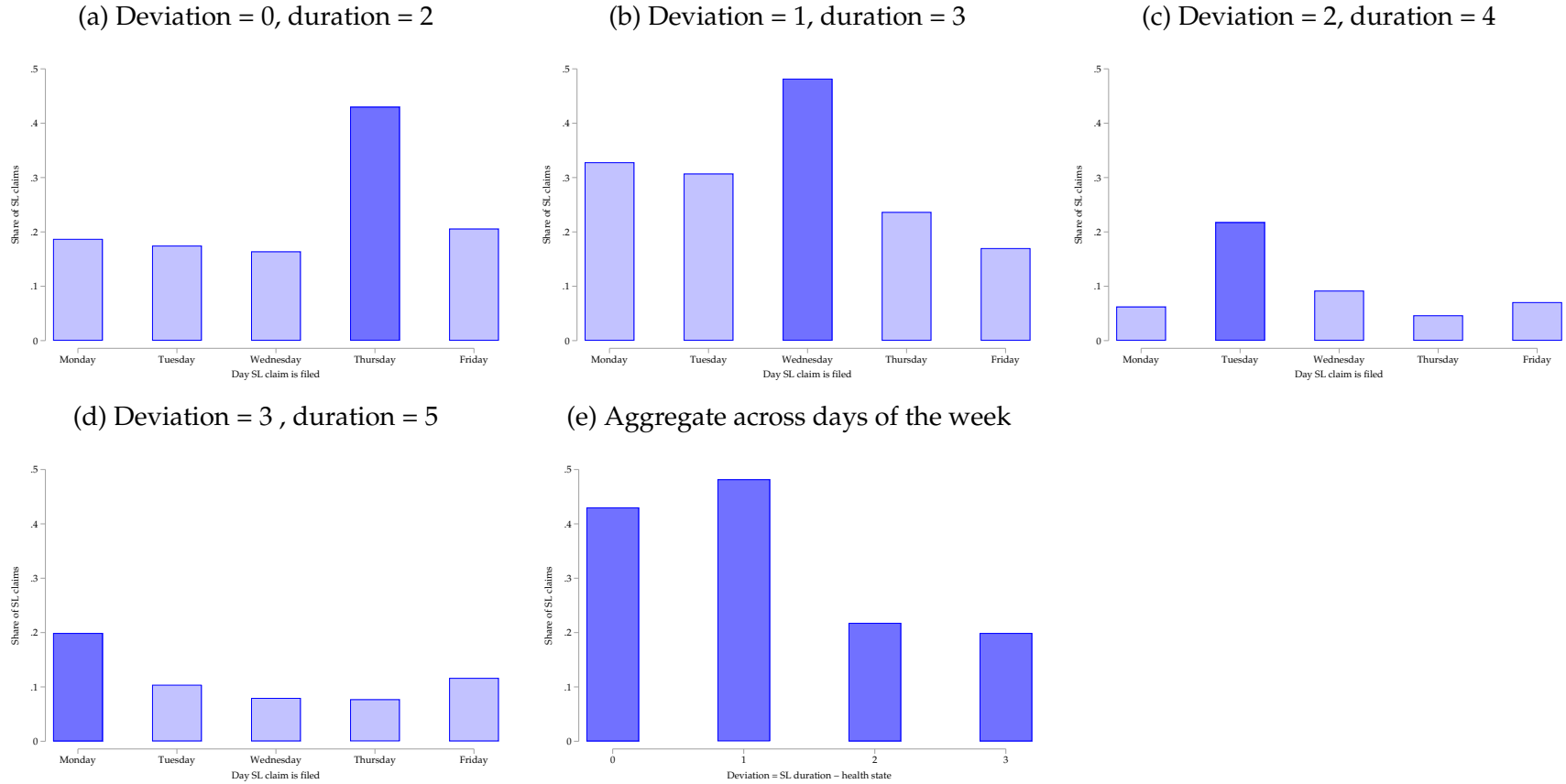
Notes: This figure shows the probability that a worker contracts disease d by workers' age and occupation for the youngest and oldest group of workers. Diseases are ordered as presented in Table 3. Sample includes male private-sector employees. Blue-collar worker refers to workers who engage in hard 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.



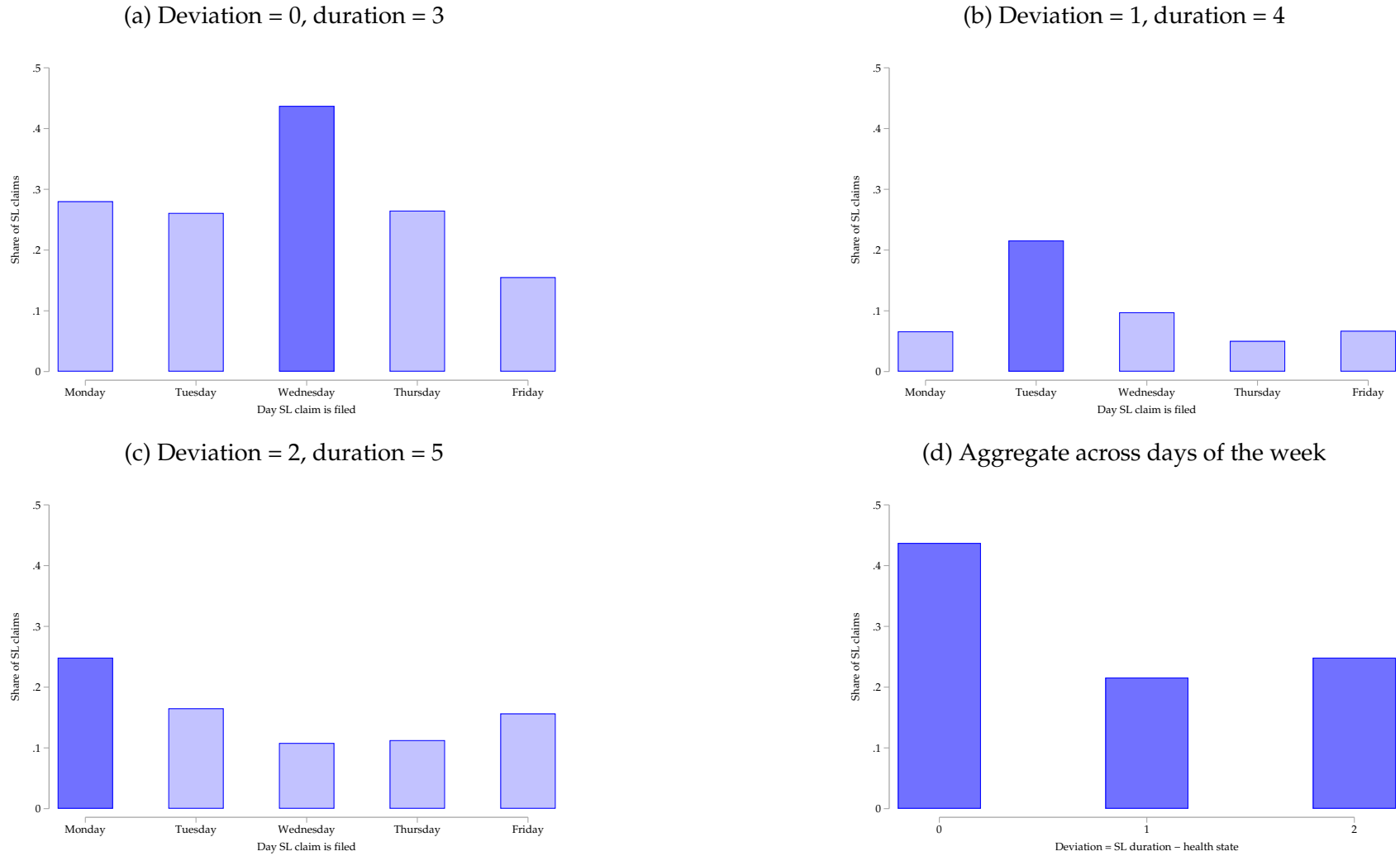
Notes: This figure shows the share of sick leave claims of duration s filed on each day of the week. Each panel aggregates sick leave claims with consecutive duration as stated in the title. This figure is referenced in Section IV.B.

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.



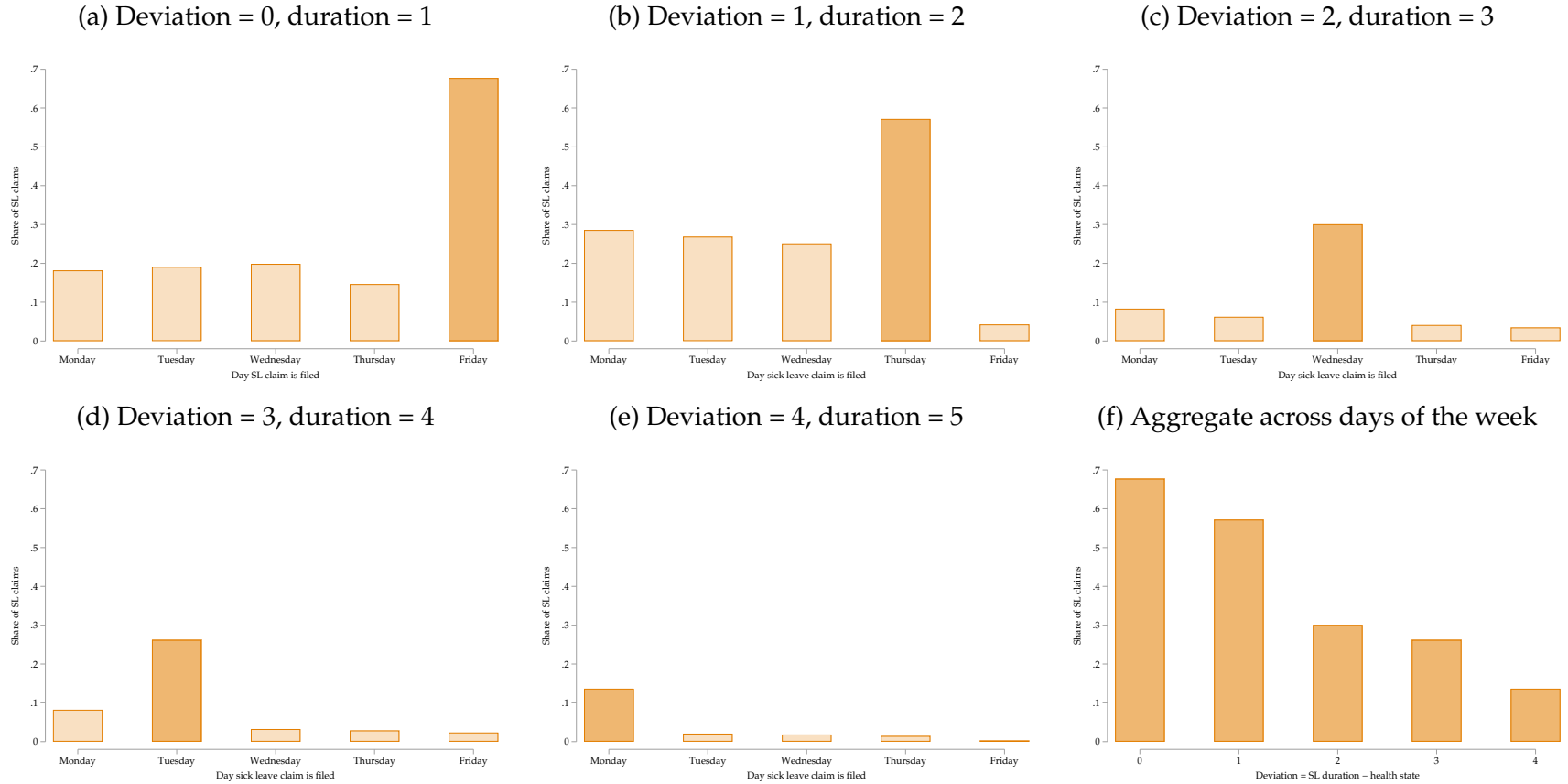
Notes: Panels (a) to (d) show the share of sick leave claims with duration s for workers whose main diagnose would implied a health state of 2 days on leave. Panel (e) aggregates the share of sick leave claims across days of the week, including only the weekend-streak combinations, e.g., from panel (a)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



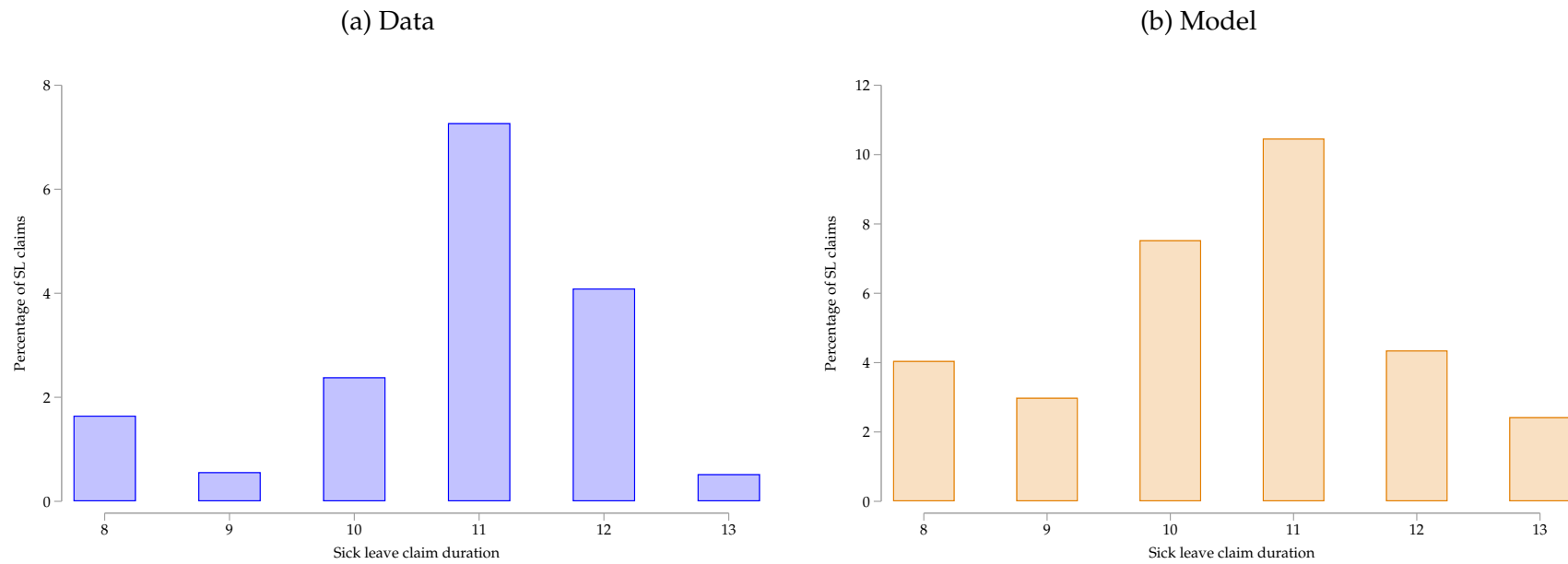
Notes: Panels (a) to (c) show the share of sick leave claims with duration s for workers whose main diagnose would implied a health state of 3 days on leave. Panel (d) aggregates the share of sick leave claims across days of the week, including only the weekend-streak combinations, e.g., from panel (a)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



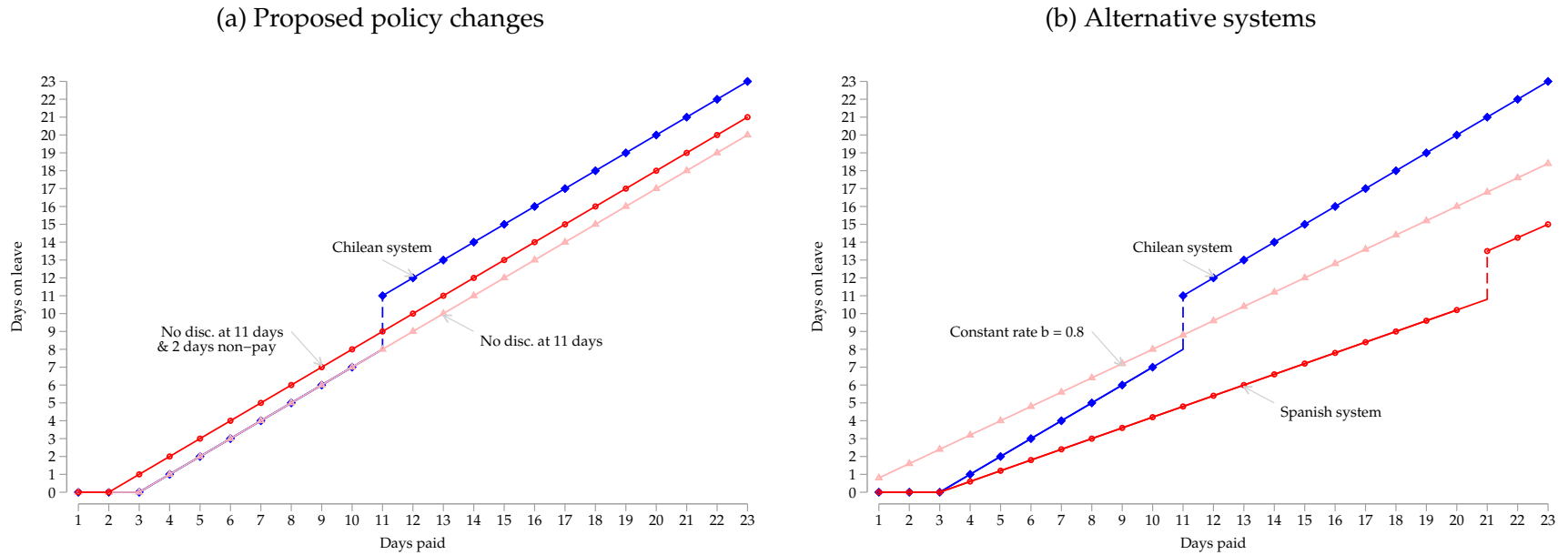
Notes: This figure shows the estimated-analogue of Figure A7. Panels (a) to (e) show the share of sick leave claims with duration s for workers whose main diagnose would implied a health state of 1 day on leave from a model-simulated sample. Panel (f) aggregates the share of sick leave claims across 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.

Figure A9: Sick leave schemes used for counterfactual analysis.



Notes: This figure presents the paid sick leave benefit schemes used in counterfactual analysis. Both panels include the Chilean system, replicating Figure 1. Panel (a) shows two alternative policies that modify the current system (i) eliminating the reimbursement of the non-payable period at 11 days and (ii) additionally reducing the non-payable period to 2 days. Panel (b) shows a simpler scheme, with a constant replacement rate equal to 0.8 of wages for any sick leave duration. Thus, the slope of this function is 0.8. This panel also shows the Spanish system for private sector employees, as studied in [Marie and Castello \(2022\)](#). This system replaces wages at a 60% rate starting in the fourth day and for sick leave claims of up to 20 days, and from the 21st day it replaces 75% of wages. This figure is referenced in Section VI.

A.II Additional Tables

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

Day of the week (dow)	Number of days on leave (s_c)							
	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)
<i>Panel A. Enrollees Characteristics</i>			
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*
<i>Panel B. Sick Leave Claims</i>			
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 within 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
<i>Panel A. Sick leave claims</i>	
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
<i>Panel B. Sick leave spells</i>	
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)
<i>Age</i>			
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
<i>Income (monthly USD)</i>			
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
<i>Region (%)</i>			
Central	46.78	47.32	48.35
Mining intensive	6.77	6.67	6.01
<i>Health - chronic conditions (%)</i>			
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 (=1 if yes)
		Number	Share (%)	
		(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
<i>Example 1</i>	
43 years old	1.05
Operator/manual worker (blue collar)	1.5
Lumbago with sciatica (M544)	14
Optimal time	22.05
<i>Example 2</i>	
25 years old	0.87
Teacher (white collar)	0.75
Common cold (J00)	3
Optimal time	2
<i>Example 3</i>	
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.	White c.	Blue c.	White c.	Blue c.	White c.	Blue c.	White c.
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(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	Average worker		High compliance costs		Low compliance costs		High valuation of time		Low valuation of time	
rate	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(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:

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

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 ²⁹

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:

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

²⁹This sub-section is referenced in footnote ??.

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 $\tilde{s} \in \{3,4,5,6,\dots,15\}$. In the three-brackets' case these restrictions imply: $\underline{s} \geq 3$, $\bar{s} \geq \underline{s} + 3$, $\underline{s} \in \{3,4,5,6,\dots,15\}$, $\bar{s} \in \{6,7,8,\dots,30\}$. 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 2,690.³⁰ Estimating the model with such level of granularity is unfeasible: it would require estimating a probability for each disease and group of observable characteristics ($2,690 \text{ diagnoses} \times 4 \text{ age group} \times 2 \text{ occupation groups} = 21,520$). For this reason, I group diagnoses in more aggregated categories.

I use 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.

³⁰I use codes F01-F99 to define mental health conditions, these are grouped under the chapter “Mental, Behavioral and Neurodevelopmental disorders”.