On the Design of Paid Sick Leave: A Structural Approach

Victoria Barone*

November 2024

[Click here for the latest version]

Abstract

This paper considers the design of an optimal paid sick leave contract and estimates the welfare gains of its implementation. I propose a model of paid sick leave provision. In this model, risk-averse workers face a health shock and decide how many days to be on leave. Given workers' behavioral responses a risk-neutral social planner chooses the optimal contract to maximize social welfare; which is a function of workers' utility and production losses and externalities induced by sick pay provision. Exploiting unique administrative data on paid sick leave utilization, I estimate workers' preferences over sick leave utilization. I use the estimated model to derive the optimal sick pay contract and estimate the welfare gains from its implementation. I find that relative to the current system, the optimal system would provide more insurance for short-term sickness and less insurance, i.e., lower replacement rates, for longer sickness spells. I estimate that workers are willing to give up 1.53% of their earnings to be insured under the optimal policy.

*Department of Economics, University of Notre Dame. Email: mbarone2@nd.edu. Website: baronevictoria.com. I thank Maurizio Mazzocco, Adriana Lleras-Muney, Michela Giorcelli, Martin Hackmann, Daniel Haanwinckel, Laura Wherry, Martha Bailey, Alexander Bloedel, Natalie Bau, Dora Costa, Felipe Goncalves, Juliana Londoño-Velez, Yotam Shem-Tov, and Bernardo Silveira for their comments and helpful conversations. I thank Fernanda Rojas-Ampuero, Kirill Ponomarev, Nicole Gorton, Jacob Kohlhepp, Ariadna Jou, Pablo Celhay, Nicolás Figueroa, Carolina Arteaga, Shooshan Danagoulian, Rajeev Dehejia, Christopher Flinn, Sherry Glied, Dmytro Hryshko, Andrew Jordan, Tim Layton, Catherine Maclean, Daniel Waldinger, Nicolas Ziebarth, and participants in the Albert Family Fund Proseminar in Applied Microeconomics at UCLA, Micro Lunch at NYU Econ, Wagner Research Colloquium, WEAI Graduate Student Workshop, 2022 ASHEcon Annual Conference, Young Economist Symposium (YES), EGSC at Washington University in St. Louis, seminar participants at Notre Dame, Wisconsin School of Business, Wisconsin Economics Department, and the University of Michigan for insightful comments and discussions. This 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 valuable protection against 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] If adequately designed, paid sick leave can be greatly beneficial. It allows workers to meet their personal health needs and smooth consumption. Nonetheless, the availability of sick pay could induce workers to request more sick days than should be assigned based on their health. This response could partially offset the welfare gains of sick pay. This raises the question: What is the optimal paid sick leave system?

The main contribution of this paper is to answer this question. To do so, I proceed in three steps. First, I use detailed data on paid sick leave claims to document how workers' characteristics and the institutional details of the sick pay system—e.g., the presence of deductibles—determine sick pay utilization. Second, I propose and estimate a model of sick pay provision. This exercise gives me key inputs to derive the optimal paid sick leave system: the value of risk protection, the costs of insurance provision in terms of moral hazard, the production cost of time off, and the underlying distribution of health shocks.[2] Finally, I use these estimates to determine the replacement rates that characterize the optimal sick paid system. I find that the optimal system features a low replacement rate for short claims, i.e., up to three days long claims are partially insured, with most of the cost on the worker side. Longer sickness spells are covered at a higher rate and the replacement rate is increasing with sick leave duration. I estimate that workers are willing to give up 1.53% of their earnings to be insured under the optimal policy.

I study this question in Chile, which is an ideal setting for this research for several reasons. First, it has a comprehensive paid sick leave system that covers all workers and features only one plan designed by the central government. Thus, workers do not choose their sick leave coverage.[3] The Chilean system provides no coverage for sickness lasting three days or fewer. This nonpayable period works like a deductible that resets with every new sick leave spell and is a common mechanism among sick leave programs.[4] Starting on the fourth day, there is full cover-

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

[2]In this paper, moral hazard refers to the responsiveness of workers' demand for sick leave to changes in the generosity of sick leave benefits.

[3]This alleviates adverse selection concerns. If workers could choose their sick pay coverage, we could expect individuals with preferences for more absences to self-select into plans with more generous provisions. The presence of adverse selection would result in an upward bias in the estimates of the moral hazard responses.

[4]These resettable deductibles are similar to those used in automobile or homeowners insurance: Separate de-

age of each missed day, i.e., the replacement rate is one. If the sick leave spans 11 days or more, the nonpayable period is reimbursed; this implies that the average replacement rate varies with the duration of a claim and jumps discretely at 11 days.[5]

The second advantage of this setting is that Chile has greatly detailed administrative data. I observe the universe of workers insured by the government eligible to file a sick leave claim between 2015 and 2019 and their utilization of sick leave benefits. This group accounts for about 70% of the Chilean workforce. This database includes rich demographic information at the worker and sick-leave claim levels. In particular, I observe the exact beginning and end dates and the primary diagnosis related to a sick leave claim at the International Classification of Diseases 10th revision (ICD-10) four-digit level. I combine the claims data with medical assessments from the Peruvian Handbook of Recovery Times (EsSalud, 2014). This handbook specifies the average recovery times for 2,763 unique disease codes at the ICD-10 four-digit level. These recommendations are adjusted based on workers' gender, age, and occupation.

Exploiting these data, I document three facts that provide qualitative motivation for the model and serve as quantitative targets in the estimation. First, workers' sick leave claim utilization varies with age and occupation. For example, on average, workers aged between 55 and 64 use an extra 2.72 days per year relative to their younger counterparts. Similarly, compared to white-collar workers, workers in blue-collar occupations use, on average, 1.15 more days. These patterns could reflect differences in the underlying distribution of health shocks and differences on their preferences. In the estimation of the model, I allow the distribution of health shocks to vary with age and occupation.

Second, I provide evidence that workers respond to the financial incentives induced by the benefit scheme. Specifically, I test whether workers bunch around the discontinuity in the replacement rate. To do so, I construct an underlying distribution of recovery times, leveraging the handbook's recommendation of how many rest days a worker needs to recover from a disease. I compare this distribution with the observed distribution of requested days and estimate that 11-day-long sick leave claims are 4.55 percentage points more likely than what the underlying distribution of health predicts. I rely on this empirical fact to assess the model performance and find that the proposed model can reproduce the excess mass observed at 11 days.

Third, I show that workers respond to nonmonetary shifts in the temptation to extend their leaves. To do so, I exploit the data on exact dates when sick leave claims are filed. I argue that the temptation to extend a sick leave claim varies with the day of the week when a worker falls

ductibles apply to each loss. Many European paid sick leave systems have a similar deductible. See Marie and Castello (2022) on the case of Spain and Pollak (2017) on the French experience. Table 1 summarizes sick pay systems for a sample of OECD countries with available data.

[5] Panel (a) of Figure 2 presents days paid as a function of days on leave for claims of different duration. Panel (b) shows the average replacement rate.

sick. For example, the incentives to file a two-day-long sick leave claim on a Thursday differ from those to file a two-day claim on a Tuesday. I consider the following exercise: I fix the duration of a sick leave claim and inspect the share of claims filed on each day of the week. I find an excess mass on combinations of days of the week and durations that allow the worker to extend her leave through the weekend. I refer to such combinations of durations and start days as "weekend-streak combinations". I document that workers are, on average, 12.33% more likely to file a weekend-streak claim than to file a sick leave claim of the same duration on any other day of the week. To capture this empirical regularity, the model allows workers' behavior to vary with the day of the week of a sick leave claim.

I use this evidence to develop and estimate a model of sick pay provision. The model has two agents: the workers and a social planner. Workers are risk-averse expected utility maximizers and choose their sick leave utilization. When choosing the demand for sick days, the worker trades off the utility of time off with the consumption loss from taking up sick leave. The latter depends on sick leave generosity. The former is a function of two key parameters: (i) workers' relative valuation of time outside work to recover from health shocks or engage in leisure and (ii) workers' propensity to overstate their sickness. Utility varies with the day of the week when a sick leave claim is filed to capture the empirical fact that the temptation to extend leaves varies at this level. In this setting, sick pay coverage lowers the cost of a day away from work and (weakly) increases sick pay utilization. This is the *traditional* moral hazard effect of insurance provision.

Sickness negatively impact workers' productivity. This reduction is justified by two forces. First, a *pure* productivity effect: a worker is less productive when she is sick relative to her healthy state. Second, individuals may be contagious at work, creating a negative externality. In the model, I capture this by allowing workers' productivity to vary with their health status. If the wage rate does not depend on the health-related productivity, this reduced productivity gives rise to a production externality.[6] Some workers find optimal to show up to work sick to avoid the consumption losses associated with an absence. Additionally, some workers find optimal to skip work even if their productivity is "high enough", e.g. some workers return to work too late. This production externality affects the main trade-off faced by the planner.

The optimal contract balances the benefits of risk protection with the cost associated with moral hazard and production losses. A more generous sick pay scheme would increase workers' well-being by offering more risk protection. Workers would respond to this policy change increasing sick pay utilization—moral hazard response. This response increase the pool of workers

[6]I assume that wages are not conditioned on workers' health-related productivity. This assumption is consistent with an information asymmetry between the firm and the worker—the worker knows her health state but the firm cannot observe it—and with contracting frictions—even if the firm could observe workers' health, it might not be able to offer a contract that induces worker to report their health state. The focus of this paper is on the consequences of this production externalities and not to model the reason why wages are independent of health-related productivity.

taking sick leave—more absences—and reduces production externalities—less individuals working sick. Thus, moral hazard is not necessarily welfare decreasing. The optimal level of benefits depends on risk preferences, workers' behavioral responses, production losses, and the distribution of risks. The principal empirical focus of this paper is to quantify these elements.

To determine the optimal system, I proceed in two steps. The first step concerns workers' sick pay utilization choices. In this step, I recover a vector of preference parameters—time valuation and compliance costs—from workers' observed leave-claiming behavior. The second step I find the replacement rates that maximize total welfare. The validity of this approach relies on the fact that the worker′s problem can be viewed as a two-stage problem. Once the health shock is realized, workers optimally choose their sick pay utilization. Risk preferences do not affect the utilization decision since the uncertainty has been resolved when this decision is made. Thus, the focus of the first step. Nonetheless, workers' expected utility depends on their risk preferences: more risk-averse workers would prefer a contract with more coverage. This is accounted for in the second step. The main strength of this approach is that I only rely on workers' observed decisions to estimate the model and do not need to assume that the current sick pay plan is optimal.

I estimate the model of workers' behavior by the simulated method of moments (SMM). The main empirical challenge is to disentangle the underlying distribution of health from the distribution of workers' preferences. To overcome this challenge, I build the underlying distribution of health exploiting the Peruvian Handbook of Recovery Times recommendations and the observed diagnoses. This approach has two advantages: (i) it provides an objective measure of recovery times constructed outside the structure of the Chilean system, and (ii) it does not impose parametric assumptions on this distribution. The empirical distribution of health states incorporates observed heterogeneity across workers: I allow for variation in age and occupation. That is, the same diagnosis has age- and occupation-specific associated recovery times. On average, older workers and workers employed in blue-collar occupations are assigned longer recovery times. In estimating the model, I also allow an arbitrary correlation between health states and workers' income to capture that wealthier workers tend to have better health and could require shorter absences.

To recover the distribution of workers' preference parameters, I employ the day of the week when a sick leave claim is filed as a quasi-exogenous shifter of the temptation to extend sick leave claims as the main source of variation. First, the excess of weekend-streak sick leave claims informs how workers' utility from a sick leave claim of the same duration varies with the day of the week on which the claim is filed. Second, I consider workers with similar characteristics and the same *assigned* recovery time—i.e., I hold workers' health, age, and occupation fixed—and compare their demand for sick pay across days of the week. I start by computing the share of claims filed for a duration that matches the assigned recovery time and compare this figure with the share of claims filed for an extra day. This difference is informative on how costly it is for individuals to ask for an extra day of leave. I restrict this comparison to claims filed for a combination of durations and days of the week representing a weekend streak. For example, I compare the share of two-day-long claims filed on a Thursday with that of three-day-long claims filed on a Wednesday for workers assigned two days of recovery. The larger the difference is, the more costly it is for workers to ask for an extra day of leave. Comparing these shares keeps the incentives for extending sick leave claims fixed since every combination implies that workers would be on leave through the weekend. These comparisons inform the distribution of compliance costs. Lastly, to learn the distribution of workers' time valuation, I construct the ratio of leisure to (a measure of) consumption from the claims data.

The estimation of the model incorporates observed heterogeneity across workers in the valuation of time outside work and the propensity for overstating sickness. Heterogeneity across the valuation of time outside work reflects variation in opportunity costs from missing work to recover from a disease—e.g., due to workers' role in the firm—or variation in tastes for leisure relative to consumption. Heterogeneity in the propensity for moral hazard behavior reflects variation in workers' preferences over behaving as expected or revealing their "true" health status. Additionally, the model of workers' behavior allows for heterogeneity in how workers perceive their sickness. That is, workers who suffer the same health shock can be affected by it differently. While the parameter that governs this perception is not separately identified, the derivation of the optimal policy does not require its identification. This derivation rather relies on workers' responses to the incentives generated by the provision of sick pay.

The estimated model provides a good fit for the targeted and nontargeted moments. I exploit the discontinuity at 11 days—a nontargeted moment—to assess the model performance. The model predicts that if a worker realizes a health state just under 11 days, she will take advantage of the proximity to the full-coverage region and fake her type to gain full coverage. The proposed model can reproduce the excess mass at 11 days quite well. I estimate that, in the data, the 11-day duration accumulates an additional 4.50% mass than its neighbors. Using the model-simulated sample, I estimate an additional 4.03%.

I use the estimated model to derive the optimal sick pay policy, i.e., to determine the replacement rates that maximize aggregate welfare. The optimal policy differs from the current system in three key ways. First, it offers partial replacement, with an average replacement rate of 0.36, for claims of up to three days. This shift increases the utility of sick workers who would not take sick leave under the current system but do under the optimal policy. At the same time, partial coverage constrains moral hazard since most of the cost of those absences is faced by workers.

Second, the optimal policy eliminates the discontinuity at 11 days and exhibits a higher average replacement rate between 4 and 10 days. Doing so curbs the cost of the behavioral responses to the program incentives and provides more risk protection. Implementing the optimal scheme would shift the distribution of sick leave duration: workers would be more likely to file sick leave claims of between 8 and 10 days and less likely to file claims for 11 days relative to the corresponding probabilities under the current Chilean system.

Third, the optimal policy does not offer full replacement for sick leave claims longer than 11 days. The average replacement rate is increasing, as in the current system, but it is less generous for longer claims. Taken together, these changes in the replacement rate reflect that the workers value a contract that offers more protection for shorter claims to smooth consumption across different health states. I estimate that workers are willing to give up 1.53% of their earnings to be insured under the optimal policy.

This paper contributes to several areas of the economics literature. First, it contributes to a large body of literature on public insurance programs. This literature 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 et al., 2010; Handel et al., 2015; Ho and Lee, 2020; Marone and Sabety, 2022) , and work-related injuries (Powell and Seabury, 2018; Cabral and Dillender, 2020). This paper contributes to this literature by being the first study to propose a theoretical framework for designing the provision of paid sick leave and quantifying the welfare gains from its implementation.

The closest programs to paid sick leave are disability insurance and workers' compensation. These three programs condition benefits on a difficult-to-verify state: the true impairment of working ability due to health deterioration. Thus, the trade-offs considered in the design of disability insurance and workers' compensation are relevant to the design of paid sick leave. Nonetheless, disability insurance and workers' compensation programs target specific groups of workers—elderly people and workers especially vulnerable to accidents—and provide protection for different set of health shocks—more permanent and more severe shocks. The nature of the health shock insured by paid sick leave provision—short spells of non–work-related illness—implies that virtually every worker could benefit from the risk protection under the program and demand paid sick leave.

This paper is close to Maclean et al. (2020), who evaluate the labor market effects of sick pay mandates in the United States and extend the Baily–Chetty framework of optimal social insurance to assess the welfare consequences of mandating sick pay. Their framework allows researchers to study the effects of policies that vary the share of employees eligible for the benefit. This paper differs in two critical dimensions. To begin with, I propose a structural approach to conducting welfare analysis. This approach does not rely on the assumption that policy changes are marginal.[7] Relaxing this assumption is important since it allows the optimal policy to differ freely from the actually implemented policy, i.e., it allows for non-marginal policy changes.

Second, this paper relates to the empirical literature on sick pay insurance. Exploiting arguably exogenous variations, the literature has documented a positive response of sick pay utilization to increases in benefit levels (Johansson and Palme, 2005; Ziebarth, 2013; De Paola et al., 2014; Ziebarth and Karlsson, 2014; Pollak, 2017;Böckerman et al., 2018; Cronin et al., 2022; Marie and Castello, 2022). This paper goes beyond workers' responses to policy changes and proposes a framework and an empirical strategy to quantify the welfare effects of these policy changes.

Additionally, this paper is the first to use administrative data on sick leave claims at the individual level.[8] These data allow me to study daily leave-taking behavior and estimate an individual demand for sick pay. In addition, 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 the incidence of absences from their length. Observing the length of absences is a crucial input for quantifying moral hazard responses, as workers could extend their absences to obtain more sick pay.

This paper also contributes to the literature on variation in leave claiming behavior across days of the week. Card and McCall (1996) and Campolieti and Hyatt (2006) provide evidence of a "Monday effect"—which refers to a spike in back injury and sprain claims on Mondays—among workers' compensation claimants. Thoursie (2004) shows that Swedish men were likelier to call in sick the day after popular skiing competitions were broadcast at night during the Winter Olympics in Calgary. Implementing a similar test, Cronin et al. (2022) document that teachers in Kentucky are not more likely to use sick leave while Keeneland is in session, on Mondays following Super Bowls, or on days when the University of Kentucky men's basketball team plays in the NCAA tournament. I provide new evidence regarding workers' behavior across weekdays by exploiting the exact dates of sick leave spans.

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 that I study and the data. Section IV presents the empirical implementation of the model. Section V presents the model estimates and main results. Section VI discusses the optimal policy. Section VII concludes.

[7]For the validity of the sufficient statistics approach, the analyzed policy changes should be infinitesimal or at least close enough to infinitesimal for first-order approximations to be precise Kleven, 2021.

[8]Cronin et al. (2022) constructs a similar dataset for the Scott County School District (SCSD) in Kentucky, which allows 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 data for Spain, though their data are at the spell rather than the individual level. The data in this paper capture the sick leave choices of 70% of the universe of Chilean workers.

II Theoretical Framework

In this section, I present a model of paid sick leave provision and discuss the derivation of the optimal sick leave insurance contract under the assumptions of this model. First, I model the choices of an expected utility–maximizing worker and outline a definition of moral hazard that applies to this setting. Second, I describe how workers' choices and provision of sick pay affect production. Third, I discuss the social planner's problem and the optimal system.

II.A Workers

A worker faces uncertainty about her health and her ability to work, she draws a stochastic health shock (θ) from a distribution $G(\theta)$, where θ represents the number of days that a worker is sick.[9] I assume that θ is discrete and bounded between zero and M and that higher values of θ are associated with longer sickness spells.[10] The sickness distribution $G(\theta)$ accumulates positive mass in the no-sickness realization; i.e., the value of zero for θ corresponds to the healthy state.

Sick pay utilization. Upon the realization of the health shock (θ), the worker decides her sick pay utilization to maximize her utility. I assume the worker derives utility over consumption (c) and time outside of work (s), given her budget constraint. The budget constraint is $c = y + w(B(s) - s)$, where y is disposable income, w is the daily wage rate, $B(s)$ represents the sick pay transfer function. Thus, $w(B(s) - s)$ is the net payment after discounting the time off. $B(s)$ is a piece-wise linear function, with marginal replacement rates (b_j) constant for sick leave claims in a duration bracket $[\underline{s}, \bar{s}]$. The worker's utility takes the following form:

$$u(s; \phi, f, w, B, \theta) = y + w(B(s) - s) + \phi\left(s - \theta + f(s - \theta)\right) \ . \tag{1}$$

The last term represents utility from time outside work. The preference parameter ϕ reflects the opportunity cost of time away from work relative to the time allocated to consumption. The term $(s - \theta)$ captures the utility cost of working while sick ($s < \theta$), and the gains from taking time off when not sick ($s > \theta$).[11] The compliance cost function $f(s - \theta)$ captures the utility cost of filling a sick leave claim. It is increasing in ($s - \theta$) and its normalized to zero if the difference between s and θ is non-positive. These costs are motivated on the risks and efforts associated with extending

[9] In this section, I omit i subscripts to simplify notation and present the baseline version of the model. I later describe how workers might vary across their distribution of health shocks and preferences.

[10] The sickness level is bounded to capture the fact that paid sick leave insurance aims to provide risk protection from impairments to working ability when full recovery is foreseeable.

[11] In the empirical application this expression depends on the day of the week a sick leave claim starts. See section IV for more details.

absences above the time needed for recovery.[12]

Workers choose sick pay utilization by trading off the cost of a day away from work $w(B'(s)-1)$ with its utility value, which depends on the duration of the claim. An additional day on leave beyond the worker's sickness level (i) lowers utility by increasing the compliance cost term in $\phi f'(s-\theta)$, (ii) increases utility in ϕ. If the marginal utility of a (extra) day off is above its cost, workers take time off. Let $s^*(\phi,f,w,B,\theta) = \operatorname{argmax} u(s;\phi,f,w,B,\theta)$ denote the optimal sick pay utilization.

Moral Hazard. Insurance provision lowers the marginal cost of sick leave—by lowering the cost of a day away from work—weakly increasing sick pay utilization. That is, $s^*(\cdot)$ is nondecreasing in the sick leave benefits function $B(s)$. In this setting, moral hazard refers to the responsiveness of the sick leave demand to varying the generosity of sick pay. Consider two alternative sick pay contracts B^0 and B^1. The contracts may specify different replacement rates given a particular piece-wise linear function or may differ in the shape of the function itself. For example, both contracts could feature a three-day-long deductible but differ in the marginal replacement rate for claims longer than three days. Alternatively, contract B^0 could have a three-days-long deductible, and contract B^1 could have a constant replacement rate. Moral hazard is the change in the demand for sick leave (Δs) when the worker is shifted from contract B^0 to contract B^1:

$$\Delta s = s^*(\phi,f,w,B^1,\theta) - s^*(\phi,f,w,B^0,\theta) \,. \tag{2}$$

This definition follows the conventional use of the term moral hazard in the healthcare literature.[13] In this literature, the term captures the notion that insurance coverage, by lowering the marginal cost of care to the individual, may increase healthcare use. Put another way, moral hazard refers to the responsiveness of consumer demand for healthcare to the price consumers pay for it. In the context of paid sick leave contracts, workers (consumers) demand time off (healthcare) by considering the share of wages that is forgone with an absence (price). Thus, moral hazard refers to the responsiveness of workers to the generosity of the contract. The literature on paid sick leave refers to this responsiveness as moral hazard as well (see Johansson and Palme, 2005 and Ziebarth and Karlsson, 2010).[14]

[12] For example, the compliance costs function could reflect the effort that the worker exerts to find a physician who would sign off on a longer leave. These mechanisms are captured in a reduced form way; i.e., I do not model the specific action that workers take to extend their absences or the risks associated to leaves above the recovery time.

[13] See Pauly (1968); Cutler and Zeckhauser (2000); Einav et al. (2013); Einav and Finkelstein (2018).

[14] This definition of moral hazard refers to "ex post moral hazard"; i.e., how workers respond to the generosity of sick pay. It abstracts from "ex ante moral hazard"; i.e., actions that workers can take to prevent deterioration of their health. Understanding how these actions are shaped by the generosity of the sick leave system is outside the scope of this paper.

Optimal utilization under linear contracts and quadratic penalties. To facilitate intuition, I impose the following functional form assumptions: (i) a linear benefit scheme $B(s) = bs$, where $b \in [0,1]$, (ii) a quadratic compliance cost function. Following this parametrization, the utility function is:[15]

$$u(s; \phi,\kappa,w,b,\theta) = y + w(b\, s - s) + \phi(s - \theta) - \phi\frac{1}{2\kappa}(s - \theta)^2 \times \mathbb{1}\{(s - \theta) > 0\}.$$

Thus, the optimal choice of sick leave duration from the worker's perspective, conditional on $s > \theta$, is:

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

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. Panels (b) to (d) present the partial coverage cases for different values of replacement rates and preference parameters. Panel (b) illustrates the definition of moral hazard proposed in equation 2. Consider worker i with preference parameters $\kappa = 3$ and $\phi = 4$, demand for time off increases with the replacement rate. Panel (c) shows that lower compliance costs (higher values of κ) are associated with longer claims: as κ increases, so does the demand for time off. Similarly, panel (d) shoes that a higher valuation of time outside work (lower ratio $\frac{w}{\phi}$) is associated with longer claims.

Scope of the model of workers' behavior. The proposed model aims to illustrate how illness affects the absence behavior of employed individuals and makes some simplifying assumptions. The model abstracts from the behavior of "when" to file a sick leave claim. Incorporating a filing-day choice in the model would require assumptions on the lag between the day a worker falls sick and when she files a claim as well as data on these events. I assume that workers claim on the day when they fall sick. Nonetheless, the model allows for strategic behavior in the duration margin of a sick leave claim, including not filing behavior. The model also abstracts from the interaction between workers and physicians. In reality, physicians write sick leave claims with (partial) information about workers' health. The main implication of not modeling this interaction is that estimates of the compliance cost function would not disentangle workers' and physicians' risks and costs.[16]

Expected utility. Ex ante, the worker aims to maximize her expected utility, taken over the dis-

[15]The quadratic compliance cost function assumes $\kappa > 0$.
[16]I present a more detailed discussion in section IV.A

tribution of health shocks $G(\theta)$. I assume that the worker is risk averse with a von Neumann–Morgenstern (vNM) utility function $v(u^*;\gamma)$, where $u^*(\theta)$ corresponds to the realized utility, i.e., $u(s^*;\phi,f,w,B,\theta)$ and γ captures risk aversion. Thus, expected utility is given by

$$U = \mathrm{E}_\theta\left[v(u^*(\theta))\right] = \int v(u^*(\theta))\, dG(\theta)\ . \tag{3}$$

This utility maximization problem can be viewed as a two-stage problem (Einav et al., 2013). Once the health shock is realized, the uncertainty is resolved, and workers aim to maximize the contribution of the *state* utility $u^*(\theta)$ to their expected utility $\mathrm{E}_\theta\left[v(u^*(\theta))\right]$ by optimally choosing the duration of a claim s. Put another way, given the health shock, risk preferences become irrelevant. That is, risk aversion does not affect workers' decision over sick pay utilization, and *all else equal*, variation in the utilization of paid leave across workers reflects variation in their preference parameters (ϕ and f).

II.B Production

In this section, I propose a stylized version of a model of a firm to capture the effect of sick leave insurance and sickness on production. For the rest of the section, it is helpful to consider an economy populated by I workers indexed by i. Thus, $s^{i*}(\phi^i,f^i,w^i,B,\theta^i,dow^i)$ represents the optimal sick pay utilization choice of worker i when insured under contract B.

Changes in the generosity of a sick leave policy, i.e., switching worker i from policy B^0 to B^1 induces a labor supply response which in turn affects production. For example, a more generous paid sick leave policy could induce workers to take time off, decreasing production. In contrast, a less generous paid sick leave policy could increase the share of workers that work while sick. Let $d^i_{healthy}$ denote the number of healthy days worked, and d^i_{sick} the number of days that i works sick. These are a function of the optimal sick pay utilization s^{i*}. Table A1 summarizes these responses when worker i is shifted to a more generous contract. As long as workers respond to changes in the generosity of the contract, i.e., $\Delta s \geq 0$, the number of days work will decrease with the generosity paid sick leave. Nonetheless, this reduction would combine healthy and sick days depending on the initial choice of worker i. These responses affect production in two ways: (i) sickness might impair the ability to perform work; (ii) if a worker is contagious and infects her coworkers, further production losses could arise by direct absences or reducing other workers' productivity.

First, consider the case of non-contagious diseases. Let μ^i represent the average productivity of worker i in the healthy state and ν^i her productivity if she is sick.[17] The firm pays worker i

[17] In this derivation, I assume that all diseases affect productivity in the same way. This assumption simplifies the exposition of the model and allows to show the main result and intuition behind how sickness affects production.

a daily wage rate w^i independent of her realized health-related productivity. The posted wage does not depend on workers' health state since this is not observable by the firm.[18] Using these definitions, the expected profits generated by worker i are:

$$\pi^i(s^{i*}) = (1-p^i)\mu^i d^i(s^{i*}) + p^i\nu^i d^i(s^{i*}) - w^i d^i(s^{i*}) \ ,$$
$$\pi^i(s^{i*}) = \underbrace{\mu^i d^i(s^{i*})}_{\text{Product absent sickess}} - \underbrace{p^i\xi^i d^i(s^{i*})}_{\text{Costs of sick work}} - \underbrace{w^i d^i(s^{i*})}_{\text{Labor costs}} \ . \tag{4}$$

where p^i is the probability that worker i is sick, i.e., the realization of θ^i is different than zero, $p^i = P(\theta^i > 0)$; and ξ^1 represents the productivity loss of sickness, i.e., $\mu^i = \nu^i - \xi^i$. In this case, aggregated expected profits are given by: $\tilde{\Pi} = \sum_i \tilde{\pi}^i(s^{i*})$.

Second, consider the case with contagious diseases. Assume that if worker i is sick, she can transmit her sickness to worker j. Let p^j represent the probability that worker j is sick from sources other than contact with worker i, q^j the probability that worker j is sick because worker i infected them, and $(1-p^j-q^j)$ is the probability that worker j is healthy. Expected profits from worker j are given by:

$$\pi^j(s^{j*}) = (1-p^j-q^j)\mu^j d^j(s^{j*}) + (p^j+q^j)\tilde{\nu}^j d^j(s^{j*}) - w^j d^j(s^{j*}) \ ,$$
$$\pi^j(s^{j*}) = \underbrace{\mu^j d^j(s^{j*})}_{\text{Product absent sickess}} - \underbrace{p^j\xi^j d^j(s^{j*})}_{\text{Costs of sick work}} - \underbrace{q^j\xi^j d^j(s^{j*})}_{\text{Costs of contagious diseases}} - \underbrace{w^j d^j(s^{j*})}_{\text{Labor costs}} \ . \tag{5}$$

Thus, in an economy with I workers, expected aggregated profits are given by:

$$\Pi = \sum_i \left(\mu^i d^i(s^{i*}) - p^i\xi^i d^i(s^{i*}) - q^i\xi^i d^i(s^{i*}) - w^i d^i(s^{i*})\right) \ . \tag{6}$$

II.C Optimal Paid Sick Leave Contract

In this section, I derive the optimal contract. I assume that the planner offers one contract and only observes the duration of sick leave claims and this is the only contractible variable.[19] To facilitate intuition, I start with a stylized version of the model to illustrate the main trade-offs the social planner faces when designing the optimal policy. Then, I relax these assumptions and present the full-fledged model of insurance provision.

[18]This assumption is consistent with an information asymmetry between the firm and the worker—the worker knows her health state but the firm cannot observe it—and with contracting frictions—even if the firm could observe workers' health, it might not be able to offer a contract where wages are adjusted with health status.

[19]Understanding whether it would be optimal to offer more than one contract is beyond the scope of this paper. From a theoretical standpoint, the key condition determining whether the optimal menu features vertical choice is whether consumers with higher willingness to pay have a higher efficient level of coverage (see Marone and Sabety, 2022). In practice, almost all paid sick leave systems feature one contract, see Table 1.

II.C.1 The Textbook Case

Consider two simplifying assumptions: (i) the planner offers a linear contract, i.e., $B(s) = bs$, and (ii) workers are risk-neutral. The first assumption allows me to focus on one policy parameter: the replacement rate level b. The second assumption allows me to ignore the value of risk protection and focus on the role of moral hazard.

Worker i chooses her sick pay utilization by comparing the cost of an absence ($(1-b)w^i$) with the net gain from an extra day off (u_s). The former depends on the wage rate w^i and the replacement rate b. The marginal utility of an extra day off depends on the health shock and workers' preferences as derived in the previous section: $u_s^i = \phi^i \left(1 - f'^i(s^i - \theta^i)\right)$. Thus, when $u_s^i \leq w^i(1-b)$, the individual chooses to go to work, else she takes a day off.

No production cost of sickness. Assume that sickness does not affect workers' productivity. That is, worker i is (i) equally productive when healthy and when sick, and (ii) sickness is not contagious. The firm offers w_i to the worker, and in this case, this equals her marginal productivity ν^i, and $\mu^i = \nu^i$. Given this wage rate, worker i efficiently self-selects into working or not working depending on the marginal utility of a day off u_s^i.

Panel (a) of Figure 1 shows graphically the effect of sick pay provision on workers' choices and welfare. This figure puts together the worker and the firm problems. The vertical axis shows the marginal utility of a day off (u_s^i) and worker's marginal productivity (ν^i). The horizontal axis corresponds to the daily wage rate w^i. Thus, at the $45°$ line, wages equal marginal productivity $w^i = \nu^i$ for different productivity levels. Additionally, at the $45°$ line, the marginal utility of an extra day off equals its cost—the forgone wage. Thus, absent of sick pay, workers efficiently sort into working ($u_s^i \leq w^i$) and not working ($u_s^i > w^i$).

The provision of sick pay distorts workers' incentives by lowering the cost of absences and induces some individuals to take a day off. The blue shaded area represents *inefficient absenteeism*.[20] It comprises the pool of individuals with marginal utility for time off below their marginal product (ν^i) that takes a day off induced by sick pay provision:

$$\text{Inefficient absenteeism:} \qquad \underbrace{u_s^i \geq w^i(1-b)}_{\substack{\text{Worker's trade-off} \\ \text{do not work}}} \qquad \text{and} \qquad \underbrace{u_s^i \leq \nu^i}_{\substack{\text{Optimal employment} \\ \text{work}}} \tag{7}$$

If the replacement rate increases from b to b', more workers are induced to call in sick. This is the traditional moral hazard response: an increase in the replacement rate increases inefficient absenteeism, i.e., workers' behavioral responses reduce welfare.

[20]This terminology is close to the one proposed by Pichler and Ziebarth (2017).

Production externalities. Consider the case where sickness affects workers' productivity. Note that workers' decisions remain the same: given the wage rate w^i, worker i chooses to go to work if $u_s^i \leq w^i(1-b)$. Nonetheless, when sickness is detrimental to productivity, wages no longer reflect the marginal product when sick.

The firm pays worker i a daily wage w^i independent of her health-related productivity. From a welfare standpoint, optimal employment trades off the productivity from working sick ν^i—which incorporates any potential adverse effect of worker i on her coworkers—against the value of leisure u_s^i. That is, it would be efficient that workers who value time off more than their productivity do not work when sick.

Panel (b) of Figure 1 provides a graphical illustration. I define four regions based on the relation between wages, productivity, and the marginal value of a day off. The wage rate w^i is presented in the horizontal axis, and the value of a day off u_s^i in the vertical axis. As in Panel (a), at the $45°$ line, the marginal utility of an extra day off equals its (private) cost, i.e., the wage rate. In contrast with the previous case, in this figure, I consider one level of productivity ν^i given by the horizontal line labeled health-related productivity. This emphasizes that wages (horizontal axis) do not vary with the health-related productivity level ν^i.[21]

First, consider the top left area of Panel (b). I refer to this pool of workers as involved in *efficient absenteeism*. This is the pool of individuals who do not work and for whom this is efficient, given their productivity:

$$\text{Efficient absenteeism:} \quad \underbrace{u_s^i \geq w^i(1-b)}_{\substack{\text{Worker's trade-off} \\ \text{do not work}}} \quad \text{and} \quad \underbrace{u_s^i \geq \nu^i}_{\substack{\text{Optimal employment} \\ \text{do not work}}} \quad . \tag{8}$$

The bottom right area shows the *efficient presenteeism* case: a pool of individuals who do work ($u_s^i \leq w^i(1-b)$) and for whom this is the efficient response ($u_s^i \leq \nu^i$).

The two darker areas show *inefficient absenteeism* and *inefficient presenteeism*. These are the situations where the workers' choices do not coincide with what the planner would find optimal in terms of employment. Inefficient absenteeism was first described in Panel (a). It refers to the pool of workers who find it optimal to be absent when it would be efficient that they work. Inefficient presenteeism, on the other hand, refers to the situation where workers' value of time off is below the cost of the absence, so they work. However, that valuation is below their marginal

[21] Appendix Figure A2 presents each relevant trade-off separately in the absence of insurance provision.

productivity when sick:

$$\text{Inefficient presenteeism:} \quad \underbrace{u_s^i \leq w^i(1-b)}_{\substack{\text{Worker's trade-off}\\ \text{work}}} \quad \text{and} \quad \underbrace{u_s^i \geq \nu^i}_{\substack{\text{Optimal employment}\\ \text{do not work}}} \ . \tag{9}$$

What is the effect of an increase in the replacement rate? Consider the case where b increases to b'; this shifts the slope of the effective wage function $w^i(1-b')$. The welfare effects of moral hazard responses has two components. On the one hand, a higher replacement rate induces absences from workers with a relatively low value of time off; it increases inefficient absenteeism. This is the same response as in the no externalities case. On the other hand, a higher replacement rate reduces inefficient presenteeism: workers with relatively low productivity take time off. This exercise illustrates the main differences between the design of sick pay insurance and other insurance programs at the intersection of health and productivity (e.g., disability insurance or workers compensation). The presence of production externalities changes the welfare effect of moral hazard. A higher replacement rate induces workers who would take up "too little" time off to taking time off efficiently. Thus, workers' behavioral responses can be welfare improving and they do not necessarily make insurance provision more expensive.

II.C.2 The Full-fledged Model

In this section, I relax some simplifying assumptions imposed in the textbook case. First, I consider that the benefit function is piece-wise linear ($B(s)$). Second, I assume that workers are risk averse. Risk-averse individuals gain utility from insurance, because it lowers the uncertainty they face. Thus, the optimal design needs to incorporate the value of insurance provision. Aggregated welfare can be written as follows:

$$W(B(s)) = \sum_i^I \omega^i U^i(B(s); \theta^i, dow^i) \ , \tag{10}$$

where ω^i represents the Pareto weight assigned to worker i and U^i represents her expected utility. I assume that total profits equal the sum of individual workers' profits, which are defined in equation (4), thus:

$$\Pi(B(s)) = \sum_i^I \pi(s^{i*}) \ . \tag{11}$$

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

$$\max_{B(s)} \quad W(B(s)) + \Pi(B(s)) \quad \text{s.t.} \quad \sum_{i}^{I} s^{i*} B(s) \leq S \; , \tag{12}$$

where S represents the allocated funds to cover the cost of the sick pay system. This constraint allows comparisons across policies that have the same cost.

III Background and Data

In this section, I discuss the Chilean healthcare insurance and paid sick leave 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 serve two functions: (i) to offer healthcare insurance contracts and (ii) to administer the paid sick leave system. The healthcare insurance system is composed of a government-run healthcare insurance provider and a handful of private insurers.[22] Workers are mandated to purchase health insurance, allocating at least 7% of their salary to a healthcare plan offered by an insurer of their choice. The mandatory contribution would allow workers to be enrolled in a plan offered by the government-run healthcare insurance with no additional contributions. To select one of the private providers, workers might need to contribute a higher proportion of their salary to qualify for the healthcare plan of their choice.[23]

The government-run healthcare insurance features four plans and enrollees are assigned to plans based on household income and composition.[24] The lowest-tier plan provides coverage for individuals with no income at no cost in public system hospitals. As income increases, individuals are assigned to higher-tier plans. These plans provide healthcare coverage in public system hospitals with low (or zero) copays and access to private healthcare institutions with higher copays.[25] Private insurance companies, on the other hand, provide tiered plans with financially vertically

[22]These are called FONASA and ISAPRES, respectively, for their names in Spanish (*Fondo Nacional de Salud* and *Instituciones de Salud Previsional*).

[23]Plans offered by private insurers are highly regulated. These insurers can set prices based on observable characteristics—including age and (until April 2020) sex—and risk factors.

[24]For example, a single worker who earns USD$693 a month—the median monthly salary in 2017—and chooses the government-run insurance system will be enrolled in the highest-tier plan and cannot choose any other plan from this insurer.

[25]Plans are indexed by letters, where A is the lowest-tier plan and D is the highest-tier plan. In 2017, the highest-tier plan had a 20% copay in public system hospitals and vouchers to use healthcare providers who participate in the plan's network at a discounted price.

differentiated coverage levels—similar to the Gold, Silver, and Bronze plans offered by Affordable Care Act exchanges in the US. These plans allow beneficiaries to obtain healthcare from private healthcare institutions, which provide a higher quality of care than public institutions.

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 the 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.

The second function of healthcare insurance providers is to *administer* the paid sick leave system. Insurers receive and screen sick leave claims and disburse sick leave benefits. Insurers cannot design sick pay plans and must follow the rules set by the central government regarding eligibility criteria and benefits. Nonetheless, there are differences in how each provider applies these rules in practice. For example, panel B of Table A2 shows that the rejection rate by private insurers is almost three times that of the government-run insurer. These differences in leniency are 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 sick pay utilization of 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.[26] Workers can use sick leave to meet their own health needs but not to care for family members. The eligibility criteria for claiming paid sick leave requires that workers (i) have been enrolled in the social security system for six months and (ii) have made contributions to the health insurance system for three months. The paid sick leave system is financed through mandatory contributions to the healthcare system. Between 2015 and 2019, the paid sick leave portion of the system was financed with 2.6% of the contributions (see table A2).

Upon falling sick, workers must obtain a physician's certification of their sickness stating the primary diagnosis and the number of days the physician considers needed to recover from the disease. This certificate is necessary to justify the absence from work and must be requested

[26]The aim of the sickness insurance system is to provide risk protection from impairments to working ability that are temporary and from which full recovery is foreseeable. A separate program provides disability insurance to workers in case of permanent impairments.

regardless of the duration of the sick leave claim; it is reviewed by an insurance office, which decides whether the sick leave claim is (i) approved with no changes, (ii) approved with a different length, or (iii) denied.

Sick leave payments are a function of sick leave duration.[27] The benefit scheme exhibits a nonpayable period of 3 days; i.e., the replacement rate for the first three days of a sick leave claim is zero. This nonpayable period works like a deductible that resets for every new sick leave span.[28] Starting on the fourth day, there is full coverage of each additional missed day—i.e., the replacement rate is one. If the sick leave lasts 11 days or more, the nonpayable period is reimbursed. That is, claims with an 11-day or longer duration are fully covered. Panel (a) of Figure 2 presents days paid as a function of days on leave for claims of different duration. Reimbursement of the nonpayable period after 11 days implies that the average replacement rate jumps discretely at 11 days, and it is nonconstant (see Panel (b) of Figure 2).

The design of the Chilean system is similar to many European paid sick leave systems making it ideal to study the research question of this paper. Table 1 presents the main features of paid sick leave systems for 22 countries. Twelve of these are "bracket systems" characterized by (i) a first bracket with a low or zero replacement rate and (ii) two or three brackets with a higher replacement rates. Specifically, in 9 of these 12 countries the replacement rate for the first days is zero. An additional advantage of the Chilean setting is that its system is comprehensive and features only one plan designed by the central government. This alleviates adverse selection concerns that could arise from plan choice.[29] Lastly, Chile has greatly detailed administrative data which I describe in the next section.

III.C Data

I exploit unique administrative data on workers insured by the government-run healthcare system provided directly by the government-run healthcare insurance office and cover the period 2015–2019. In these data, I observe the universe of individuals enrolled in government-run healthcare insurance regardless of whether they have filed a sick leave claim. The data contain individuals' demographic and economic characteristics: sex, age, annual earnings, and health indicators for

[27]The benefit scheme described here applies to private-sector workers with earnings bellow a maximum. This maximum is updated every year. In my sample, less than 1% of workers earn above this threshold; see Figure A3. Additionally, public sector employees benefits are computed following different rules. I exclude these workers from the analysis.

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

[29]If workers could choose their sick pay coverage, we could expect sicker individuals to choose plans offering more generous insurance coverage. While 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.

chronic conditions.[30] Additionally, I observe the individual's health insurance plan, which allows me to exclude those enrolled in the lowest-tier plan from the analysis, as they are not active in the labor market. For each individual who has filed a sick leave claim, I observe detailed information about it: exact start and end dates, prescribed days on leave, primary diagnosis (coded following the ICD-10), physician identifiers, and amount received for paid sick leave. I also observe the occupation in which the worker is employed at the moment of filling the sick leave claim.

The Peruvian Handbook of Recovery Times (EsSalud, 2014) contains medical guidelines about average recovery times for 2,763 unique conditions. These recommendations are coded at the ICD-10 four-digit level and are adjusted based on workers' sex, age, and occupation. Table A3 provides an example of the average recovery times for three common diagnoses—lumbago with sciatica, common cold, and infectious gastroenteritis—and the correction factors proposed by the handbook. The main advantage of exploiting this external source of data is that it provides an objective measure of recovery times constructed outside the structure of the Chilean system. That is, it is not affected by the brackets used in the paid sick leave benefit function.

Based on these two sources, I construct a claim-level dataset with detailed information on workers' demographic characteristics, their leave-taking behavior, and the average recovery time. My primary measure of leave-taking behavior is the exact duration of a sick leave claim filed on a given day of the week. I consider consecutive claims as one claim, since this is how sick leave benefits are computed.[31] Exploiting the granularity of the data, I assign a benchmark recovery time to each sick leave claim based on the disease code and workers' characteristics.

The estimation of the model relies on the sick leave taking behavior from male workers aged 25 to 64 years in 2017. I focus on this group because of its 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 labor force participation rates that are more than 20 percentage points lower—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 choices would also require incorporating their decision to participate in the labor market. Nonetheless, abstracting from the participation decision simplifies the model estimation.

The estimation sample includes claims associated with a subset of diseases. I exclude mental-health diagnoses because their filing process is more cumbersome than the one for non–mental health claims.[32] Among non–mental health diagnoses, I exclude diagnoses for which it is hard

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

[31] Appendix Table A5 presents counts and summary statistics of sick leave claims and sick leave spells.

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

to assign a recovery time or for which recovery is not foreseeable. For example, I exclude claims with codes corresponding to neoplasms. Table A4 lists the conditions included in the analysis and the share of claims recorded under each diagnosis. The final sample includes 90.19% of all non–mental health sick leave claims.

My final dataset includes 1,916,138 workers of whom 13.8% have filed at least one sick leave claim in 2017, this amounts to 329,312 sick leave claims. Exploiting these data, I document the main patterns regarding sick leave claim utilization, workers' behavioral responses, and estimate the parameters of the model of workers behavior.

III.D Descriptive Evidence

Determinants of Sick Leave Utilization. Table 2 presents summary statistics for all the workers in the sample and for those who used sick pay benefits. I group workers who filed (i) at least one claim with a duration of up to three days, (ii) at least one claim with a duration of between 4 and 10 days, and (iii) at least one claim with a duration of 11 days or longer. In 2017, approximately 13.8% of workers in the estimation sample filed at least one sick leave claim. These workers are on average 42 years old, and about 1.7 years younger than the average worker. The age distribution of workers who filed a sick leave claim skews towards younger workers, especially for up to 3 days long sick leave claims. For example, the share of workers aged between 25 and 34 who filed a sick leave claim for up to three days is 38% while for claims longer than 11 days, this figure is 25% (Columns (3) vs (5) of Table 2). Higher earners are more likely to take time off, and the difference in average monthly wages between those who take time off and those who do not is greater for shorter sick leave claims. This pattern is compatible with the 3-day waiting period, representing a binding constraint for workers and thus reducing the likelihood that lower-earning workers file a sick leave claim.

Distribution of Sick Leave Duration and Conditions. One of the main advantages of the administrative data is that I can leverage exact begin and end dates of paid sick leave to construct a reliable measure of sick leave duration. Figure 3 presents the distribution of days on leave. Two main patterns characterize this distribution. First, approximately 26.54% of sick leave claims have a duration of up to 3 days. This provides evidence that workers are completing the filling process to justify their absences even when not paid for them. Sick leave claims lasting between 4 and 10 days account for 41.1% of claims. Thus, 32.4% of claims have a duration between 11 and 29 days.[33] Second, some durations accumulate more mass than others. For example, three days on

[33]Claims filed for 30 days are used either by workers with illness requiring a longer recovery or those transitioning to disability insurance. I do not have access to data that would allow me to differentiate between these outcomes. Thus, my analysis focus on claims of up to 29 days. This restriction approximates the universe of workers that suffer conditions with foreseeable recovery.

leave is the most common duration representing 15.5% of claims, follow by five and seven days (accounting for 13.6% and 13.6%, respectively). This pattern could be explained by the underlying distribution of recovery times or behavioral responses to the incentives provided by the sick leave benefit scheme; disentangling these is one of the papers' aims.

Figure A4 shows the histogram of sick leave claim duration by worker characteristics. I group workers into eight groups or bins defined based on age and occupation type: blue-collar and white-collar occupations.[34] Conditional on worker 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 for younger workers (comparing across rows in Figure A4). This pattern is consistent with workers requiring more time to recover from the same conditions as they age and with older workers suffering more severe underlying conditions. Comparisons across occupations for workers in the same age group indicate that claims made by blue-collar workers are longer on average, with a smaller share of claims of up to 3 days. This comparison suggests that differences in the underlying distribution of health could be correlated with occupation type. Motivated by these results, I allow the underlying distribution of health to vary with workers' age and occupation in the estimation of the model.

Evidence of behavioral responses – Excess mass at 11 days. Figure 4 compares the underlying distribution of recovery times with the observed distribution of days on leave. It shows that three rest days are the most recommended recovery time, consistent with 3-days-long sick leave claims being the most frequent duration. In contrast, there is a broader gap when comparing the share of claims with five and seven days as suggested recovery time and the observed claims of such duration. This pattern is consistent with physicians being more likely to write recovery times that correspond to a workweek—five days—or a calendar week and multiples of these. Panel (b) of Figure 4 shows the ratio of the difference between the share of sick leave claims for a given duration and the share implied by the underlying distribution to the latter. This figure illustrates that the greater gaps are at 5, 7, and 11 days.

Finally, the excess of mass or bunching at 11 days coincides with the most significant jump in the average replacement rate: starting at 11-day-long claims, workers are fully reimbursed for the time off. This jump incentivizes workers to extend their leaves to enter the "full" insurance region. Panel (b) of Figure 4 shows missing mass in durations just below the eleven-day jump: eight, nine, and ten days. I estimate that 11-day-long sick leave claims are 4.55 percentage points more likely than what the underlying distribution of health predicts. I interpret these patterns as suggestive evidence that workers respond to the discontinuity in the replacement rate at 11

[34] A blue-collar worker refers to an individual who performs manual labor. For example, operators, assemblers, and laborers are considered blue-collar workers. A white-collar worker refers to an individual who performs professional, desk, managerial or administrative work. For example, sales representatives are considered white-collar workers.

days.[35]

Evidence of behavioral responses – Days of the week. Incentives to take time off vary with the day of the week when a worker falls sick. For example, the incentives to file a two-day-long sick leave claim on a Thursday differ from the incentives to file a two-day-long claim on a Tuesday. The first combination implies four continuous days on leave while the second combination implies two days. I refer to the first case as a "weekend-streak combination".

Figure 5 shows the share of sick leave claims filed on each day of the week. 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 day . That is:

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

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 5). Similarly, two-days long claims are more likely to be filed on a Thursday than any other day of the week.[36] This pattern is present for one- to five-day-long claims. Crucially, when inspecting 7-day-long claims, the share is constant across days of the week.[37] I document that workers are 12% more likely to file a weekend-streak claim than a claim of comparable duration on any other day of the week. To capture this empirical regularity, the model allows workers' behavior to vary with the day of the week of a sick leave claim.

IV Model Estimation and Identification Discussion

The mechanics of the model are as follows: each worker i draws a realization of a health shock (θ^i,dow^i) and knows their wage rate (w^i) and labor income (y^i), which are determined outside the structure of the model. Upon the realization of the health shock, the worker decides her paid sick leave utilization. First, I present the parametric assumptions imposed on the model of workers' behavior. Second, I explain the construction of the health shocks and the estimation procedure. Third, I discuss identification of these parameters.

[35] Missing mass at eight, nine, and ten days could also be explained by the rounding at seven days. In the model, I allow for a rounding process that rounds up (down) sick leave claims with durations in three days neighbor.

[36] In the data, less than 6% of sick leave claims are filed on weekends. Thus, in the rest of the paper I focus on claims filed between Monday and Friday.

[37] Claims of durations longer than six days exhibit a similar pattern. I use seven days as a reference point since the share of these claims in the data is greater than the share of 6-day-long claims. Appendix Figure A6 presents the distribution of the share of sick leave claims by day of the week for claims with a duration of between 8 and 15 days, pooled in 2-day groups.

IV.A Parameterization of Workers' Preferences

Parameterization of the utility function. I represent the theoretical model fully in terms of parameters to estimate. I assume the following utility function:

$$u(s^i;\phi^i,\boldsymbol{\kappa}^i,w^i,B,\theta^i) = y^i + w^i(B(s^i)-s^i) + \phi^i(s_l^i(s^i;dow^i)-\theta^i) + \phi^i q\ \mathbb{1}\{\text{weekend}\} \\ - \phi^i\left[\kappa_0^i(s^i-\theta^i)^2\mathbb{1}\{s^i-\theta^i>0\} + \sum_{j=1}^{3}\kappa_j\mathbb{1}\{s^i-\theta^i=j\}\right], \tag{13}$$

where i indexes workers. I assume that the transfer function $B(s)$ is captured by a piece-wise linear function with day brackets corresponding to those currently implemented in the Chilean system. The compliance cost function takes a flexible functional form: it allows for quadratic penalties and specific costs of asking for one, two, or three days above the health state θ.

This formulation of the utility function modifies Equation (1) letting utility to vary by the day of the week to better capture the patterns observed in the data. The (dis)utility from working sick is a function of week-days, i.e., $\phi(s_l(s;dow)-\theta)$. In this expression, s_l indicates business days, a function of total days on leave and the day of the week a sick leave claim starts.[38] The term $\phi^i q\ \mathbb{1}\{\text{weekend}\}$ captures the extra utility that a worker derives when the sick leave claim allows her to not return to work until after the weekend. The indicator variable $\mathbb{1}\{\text{weekend}\}$ takes the value of one for sick leave claims ending on a Friday and with a duration of up to 5 days.[39]

The preference parameter ϕ^i governs the valuation of time outside work. I assume that $\ln(\phi^i)$ is drawn from a normal distribution with mean μ_ϕ and variance σ_ϕ^2 such that

$$\ln(\phi^i) \sim N(\mu_\phi,\sigma_\phi^2)\ .$$

Heterogeneity in the valuation of time outside work reflects variation in the opportunity costs of missing work to recover from a disease or in tastes for leisure relative to consumption. The term ϕq captures the extra utility that a worker derives when the sick leave claim has a weekend-streak duration. I assume that all of the variation in this term is governed by the parameter ϕ; thus, q does not vary across workers or sick leave duration. Under this assumption, variation in the value of a weekend streak combination across workers reflects their valuation of time off.

Heterogeneity in the compliance cost parameter κ_0 reflects variation in workers' preferences

[38] For example, two days long sick leave claim starting on a Monday represents two business days away from work, while a sick leave claim starting on Friday implies one day away from work. See Appendix Table A8.

[39] A more general definition would be to have $\mathbb{1}\{\text{weekend}\}$ equal one for each sick leave claim that ends on a Friday and assume different values of q for the first and second weekends. 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 5 and A6 support this assumption.

over behaving as expected or revealing their "true" health status. Additionally, job characteristics can justify variation in κ_0. For example, to capture variation in how easily a coworker could perform one's job and how this affects utilization. Alternatively, high risks of extending a sick leave claim above ones' health would be captured by a high compliance cost. I model both of these mechanisms in a reduced-form manner. I assume that $\ln(\kappa_0^i)$ follows a normal distribution with mean and variance μ_{κ_0} and $\sigma^2_{\kappa_0}$:

$$\ln(\kappa_0^i) \sim N(\mu_{\kappa_0},\sigma^2_{\kappa_0}) \ .$$

I interpret κ_1, κ_2, and κ_3 as shifters of the compliance cost for deviating for one, two, and three days, respectively. For example, the value of the compliance cost function for an extra day off is: $f_1^i = f^i(s^i = \theta^i + 1; \theta^i) = \kappa_0^i + \kappa_1$. Thus, heterogeneity in κ_0 implies that the cost of deviating for one day varies across individuals—a similar argument applies for two- and three-day-long deviations. Nonetheless, this specification assumes that workers with high (low) compliance costs face a high (low) cost of deviating for one, two, three, four, or any number of days.

In sum, workers' behavior is characterized by a vector of preference parameters: $\{q, \mu_\phi, \sigma^2_\phi, \mu_{\kappa_0}, \sigma^2_{\kappa_0}, \kappa_1, \kappa_2, \kappa_3\}$.

Rounding and measurement error. I include two additional mechanisms when estimating the model to capture the behavior of physicians.[40] 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 the model does not predict. I assume that the duration of sick leave claims is measured with an additive error that has a 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}, it takes the values one or negative one; i.e., it shifts the duration of a sick leave claim by one day. With probability $p^2_{me} = p_{me} \times p_{me}$, the duration of the sick leave claim is shifted two days. That is, δ takes the values two or negative two. And a similar argument works for p^3_{me}.

Second, I adjust the sick leave duration to consider the rounding or heaping observed in the data. I interpret this pattern as coming from physicians being more likely to prescribe rest for a

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

number of days that is a multiple of seven. I assume that with some probability p_7, a sick leave claim of duration m is rounded up (down) to seven days.

IV.B Health Shocks

Worker i health shock is characterized by a recovery time (θ^i) and a day of the week (dow^i) when the worker falls sick. I use the Peruvian Handbook of Recovery Times to assign the average number of days that a worker would need to recover from the condition reported in the sick leave claim data, i.e., θ^i. These recovery times vary by occupation and age allowing for heterogeneity in θ. I assume that workers fall sick on the day they start a recovery spell. Thus, I construct dow^i from the observed day when a sick leave claim is filed from the data. [41]

Heterogeneity in health shocks "intensity". The model of workers' behavior allows for heterogeneity in how workers suffer a health shock. To see this, let the parameter α reflect how sickness affects a worker: workers with a higher α benefit more from time outside work. Consider two workers such that worker a is more affected by the symptoms of any disease than worker b: $\alpha^a > \alpha^b$. For example, workers a and b fall sick with the common cold on Monday, i.e., they suffer the same health shock, but the realization that a gets is worse and she would need more time to recover. The proposed model implies that worker a would file long sick leave claims: $s^{*a} > s^{*b}$. Nonetheless, the perception parameter (α) is not identified. To see this, note that the utility of worker i explicitly accounting for her perception of a shock is:

$$u(s^i;\phi^i,\boldsymbol{\kappa}^i,w^i,b^i,\theta^i,\alpha^i) = w^i(M-s^i) + w^iB(s^i) + \tilde{\phi}^i\ \ \alpha^i(s_l^i(s^i;dow^i)-\theta^i) + \tilde{\phi}^i q\ \mathbb{1}\{\text{weekend}\}$$
$$-\tilde{\phi}^i\big[\kappa_0^i(s^i-\theta^i)^2\mathbb{1}\{s^i-\theta^i>0\} + \sum_{j=1}^{3}\kappa_j\mathbb{1}\{s^i-\theta^i=j\}\big]\ .$$

The parameter α^i is not separately identified from ϕ^i, i.e., $\tilde{\phi}^i\alpha^i$ is observational equivalent to ϕ^i. Nonetheless, it is not necessary to separately identify α^i to derive the optimal sick pay policy. What matters for the optimal design of the policy are workers' responses to the incentives generated by the provision of sick pay. These responses are a function of the parameters $\phi, q, \kappa_0, \kappa_1, \kappa_2,$ and κ_3 which are identified.

[41] Relaxing this assumption would require an additional source of data that distinguishes between the day that a worker falls sick and the day that she starts an absence spell or files a sick leave claim. Absent such data, I use the starting day of a recovery span as the day when the worker falls sick. I assume that workers file sick leave claims from Monday to Friday and that their work schedule is precisely Monday to Friday. In the data, less than 6% of sick leave claims are filed on weekends. Additionally, 83% of Chilean workers have a regular work schedule (Aguayo Ormeño, 2019).

IV.C Estimation Procedure

I estimate a vector of ten parameters: $\Lambda = \{q, \mu_\phi, \sigma^2_\phi, \mu_{\kappa_0}, \sigma^2_{\kappa_0}, \kappa_1, \kappa_2, \kappa_3, p_{me}, p_7\}$. For this estimation, I select informative moments from the sick leave claims data and use the SMM to estimate the vector of parameters that minimize the criterion function. Let $G(\Lambda)$ represent the vector of simulated moments and G^E their empirical counterpart. I aim to find the vector of parameters Λ that minimizes the squared distance between the simulated moments and the moments computed from the data:

$$\min_{\Lambda} \sum_{t=1}^{10} \left(\frac{G_t(\Lambda) - G^E}{G^E} \right)^2 .$$

To compute the simulated moments, I draw a representative sample of the data. This sample consists of a vector of wages, recovery times, and days of the week and is stratified at the workers' group level.[42] The main strength of this approach is that it does not impose parametric assumptions on the distributions of wages and health shocks. That is, this strategy allows for arbitrary correlation between the health shocks and workers' wages to capture two empirical facts: (i) the duration of days on leave varies with income; and (ii) diagnosis prevalence changes with the age and occupation of workers.

In the estimation, I exploit workers' responses to the incentives created by sick insurance provision. That is, the estimation procedure relies only on workers' observed decisions and does not impose optimality of the current policy. That is, in the estimation of the model, I do not assume that the current policy is the optimal one, I only need to assume that workers are utility maximizers. This result relies on the fact that the worker's problem can be viewed as a two-stage problem. Once the health shock is realized, workers optimally choose their sick pay utilization. Neither risk preferences nor production effects affect workers' utility. Risk preferences do not affect the utilization decision since the uncertainty is resolved once the health shock is realized. The production effects are not internalized by workers: wages and the health distribution are independent of their labor supply decision, i.e., each worker is infinitesimal.

IV.D Moments and Identification

Even though the parameters are jointly estimated, below I provide a heuristic discussion of the most relevant moment for each parameter.

[42] Table A.11 verifies balance in terms of workers' characteristics and sick leave utilization between the sample drawn for estimation of the model and the sample used to document workers' behaviors and compute data moments.

Weekend-streak utility (q). The term $\phi q 1\{weekend\}$ captures the extra utility that a worker derives when the interaction of the 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 to extend a sick leave claim varies between days of the week. For example, a 2-day-long sick leave claim is more attractive on a Thursday than on a Tuesday. Figure 5 illustrates this variation.

The identification of q relies on 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. I pool all the weekend-streak combinations to compute the average share of claims on those days and compare it with the average share of claims made during the rest of the week. The model requires a higher q to rationalize the data if a larger difference is observed. 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 filing a sick leave claim of duration s on a weekend-streak day. The last panel of Figure 5 shows this moment graphically, and Table A9 presents detailed computations.

Compliance cost function ($\mu_{\kappa_0}, \sigma^2_{\kappa_0}, \kappa_1, \kappa_2, \kappa_3$): I exploit variation across days of the week and sick leave claims duration conditional on workers' health to inform the distribution of compliance costs. I consider the pool of workers with similar characteristics and the same *assigned* recovery time—i.e., I hold fixed workers' health, age, and occupation—and compare their demand for sick pay across days of the week. For each day of the week and assigned recovery time, I compute the share of sick leave claims, indexed by j, of duration s filed by workers with health θ:

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

where the denominator counts the number of sick leave claims filed on day of the week day with primary diagnoses that would require x days of 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 is given by

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

Figure 6 illustrates this computation for sick leave claims with a health shock that requires a 1-day-long recovery. I start by computing the share of claims filed for a duration that matches the assigned recovery time on a weekend-streak day and compare this share with the share of claims filed for an extra day on a weekend-streak day. That is, I compare Panel (a) vs. Panel (b) of Figure 6. This difference is informative on how costly it is for individuals to ask for an extra day of leave. I restrict this comparison to claims filed for a combination of duration and day of the week representing a weekend streak. This conditioning keeps the incentives for extending a sick leave claim fixed. That is, every combination implies that workers would be on leave through the weekend. These are the darker columns in Figure 6. Panel (c) of Figure 6 shows how costly it would be to ask for two extra days of leave. Panel (f) summarizes the probabilities of not asking for extra days of leave, asking for one extra day of leave, asking for two extra days of leave, and asking for up to four extra days of leave conditional on filing a sick leave claim on a weekend-streak day.

I perform these comparisons for sick leave claims with diagnoses assigned one, two, and three days of rest (see Figures A7 and A8). To inform the distribution of compliance costs, I compute the average share of claims with a given deviation. These shares are presented in Panel (a) of Figure 7. The pattern in the data suggest that a one-day-long deviation is not too costly relative to truth-telling while two-day deviations are more costly, as reflected by the lower share of sick leave claims in the third column of this graph.

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

$$LC^i = \frac{\text{leisure}^i}{\text{consumption}^i} = \frac{1}{N^i}\sum_m \frac{w^i_m \times \ Days\ on\ leave^i_m}{w^i_m \ \times Days\ worked^i_m + \ Sick\ pay^i_m},$$

where m indexes the month of the year. N^i is the number of months in the year in which worker i used at least one sick leave claim. The numerator estimates worker i's 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 A9 shows the distribution of LC^i; the mean and standard deviation of this distribution inform the distribution of ϕ, which I assume to be log-normal with mean μ_ϕ and standard deviation σ_ϕ.

Rounding and measurement error. I use the difference between the share of 5-day-long sick leave claims filed on a Monday relative to the share of claims filed on a Tuesday, conditional on health shocks with a 1-day recovery, to pin down the success probability of the measurement error term δ (see Panel (e) of Figure 6). Given the share of claims filed on a Monday, a smaller difference implies that more sick leave claims have been moved away from the most-profitable duration. That is, the smaller the difference, the more likely it is that the observed duration is not the optimal one in terms of workers' utility. To inform the probability of a sick leave claim being rounded to duration that is a multiple of seven, I use the share of seven-day-long sick leave claims.

V Results: Workers' Behavior

V.A Parameter Estimates

Table 3 presents the values of the estimated parameters. 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 (q). I estimate that, *all else equal* the utility of a worker increases in 0.79 for filing a sick leave claim weekend-streak combination.

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. Panel (b) of Figure 7 compares 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 $\mu_{\kappa_0} = 0.82$ and $\sigma_{\kappa_0} = 1.77$.

I use the distribution of the ratio of leisure to consumption to identify the distribution of values of time outside work, i.e., the distribution of the parameter ϕ. I estimate that, on average, workers value time off work, either to recover from disease or to engage in leisure, about 45% 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.

V.B Workers' responses

Exploiting these estimates, I document how sick leave taking behavior varies with changes in the replacement rate. First, I document how workers' behavior changes if the jump at 11 days is reduced. I consider three alternative systems that keep the marginal replacement rates in each bracket fixed—the slope of each payment function is one for claims above 4 days—and reduce the size of the jump at 11 days. Panel (a) of Figure 8 presents the alternative payment schemes, each scheme reduces the jump at 11 days in one day, the less generous alternative features no discontinuity. Panel (b) shows the share of sick leave claims filed under each alternative system.

Two main patterns arise. First, the mass at 11 days decreases monotonically as the jump at 11 days is reduced. Second, the share of sick leave claims filed for eight, nine, and ten days increases as the discontinuity decreases. This result suggests that workers who would extend their time off to 11 days to enter the "full insurance" region, find this behavior less attractive as the jump at 11 days decreases. I estimate that, on average, the share of 11-days-long sick leave claims declines in 0.1538 percent for each day that is not reimbursed. That is, when move from the current system to the one that pays 10 days out of 11 days, the share of claims filed for 11 days decreases in 0.1417 percent. Similarly, when move from the current system to the one that pays 9 out of 11 days, the share of claims decreases in 0.1565 percent, on average, per day.

V.C Model Fit

Matched moments. With the estimated parameters, the model matches the most relevant moments, presented in Table 4. There are, on average, 12.33% more sick leave claims on weekend-streak days. The share generated by the model is very close: on average, I estimate 14.64% more sick leave claims on weekend-streak days relative to non–weekend-streak days.

Panel (b) of Figure 7 compares the share of sick leave claims with non, one, and up to three-days long deviation implied by the data and by a model-simulated sample. I overestimate the share of claims with non deviations—i.e., claims with duration equal to the assigned diagnosis. I slightly underestimate the share of claims for a day above the assigned diagnosis. Nonetheless, the model replicates the decay in the share of sick leave claims with positive deviations quite well. For example, a two-day deviation is more costly than a one-day deviation, as reflected by the lower share of sick leave claims in this category.

The distribution of the ratio of leisure to consumption is assumed log-normal. Under this assumption, the mean generated by the model is slightly higher than the observed in the data, while the variance, on the other hand, is very close.

Specification Tests. I test how well the model matches data moments not used in the estimation. Figure 9 compares the share of claims filed for a duration of 8 to 13 days 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. In particular, using the measure of heaping proposed by Roberts and Brewer (2001), I estimate that, in the data, the 11-day duration accumulates an additional 4.50% mass than its neighbors. Using the model-simulated sample, I estimate an additional 4.03% mass relative to its neighbors.[43] The derivation of the optimal policy requires an estimate of the moral hazard costs associated with this discontinuity, thus the importance of a precise estimate of workers' responses to this feature of the paid sick leave contract.

Additionally, I construct the demand for days on leave as a function of the duration of the health shock. For each duration, I compute how many days; on average, workers request to be on leave. Figure A10 compares the average days on leave from the data and a model-simulated sample. This figure tests the model's ability to replicate workers' sick leave utilization choices and provides evidence that the model can replicate workers' responses to different health shocks. It is important that the model performs well on this dimension since the derivation of the optimal policy relies on estimates of workers' responses to changes in the paid sick leave policy.

I also propose an out-of-sample exercise exploiting data not used in the estimation of the model. Using data on sick leave claims filed in 2019, I compute the vector of moments used to estimate preference parameters and the share of claims with a duration in the neighborhood of 11 days. I compare these moments with their model-simulated counterparts to test the model performance. To obtain the latter, I simulate the model based on a representative sample drawn from the 2019 data and the estimated vector of preference parameters. Table 5 presents the results of this exercise. The results suggest that the model performs reasonably well out of the sample: preference parameters and the share of sick leave claims of selected durations are comparable in magnitude. Additionally, the model reproduces (i) the decay of the share of sick leave claims with positive deviations and (ii) the excess mass at 11 days.

Robustness Checks: Moments' computation. The estimation of the model relies the computation of moments using data on all sick leave claims filed during the 2017. I ask whether the main moments are affected by restricting the sample to specific times of the year, e.g., Winter.

Weekend-streak utility (q). In figures A11 to A14, I compute this moment restricting the sample to claims filed during each quarter of the year. These figures show the same qualitative pattern than Figure 5 providing evidence that claims from a particular time of the year, e.g., winter do not drive the patterns in the data.

Compliance cost function ($\mu_{\kappa_0}, \sigma^2_{\kappa_0}, \kappa_1, \kappa_2, \kappa_3$): Figure A15 proposes a similar exercise. I show the distribution of share of sick leave claims with none and positive deviations for each quarter of the year. This figure tells a similar story: the share of sick leave claims with no or one day deviations are almost the same, and the share of sick leave claims corresponding to longer deviations

[43] Roberts and Brewer (2001) proposes the following measure: $h_z = f(z) - \frac{f(z-1)+f(z+1)}{2}$, where z corresponds to 11 days, and $f(\cdot)$ indicates the frequency of sick leave claims with duration z. Thus, h_z gives the difference between a duration frequency and the average of the frequencies of the two immediately neighboring duration. It indicates how much a duration sticks out from the pattern suggested by its neighbors.

decreases monotonically.

VI The Optimal Sick Pay Contract: Derivation and Counterfactuals

I use the estimated model of workers' behavior to determine the sick paid leave system that maximizes aggregate welfare. In this section, first I present the set of assumptions that I impose when solving the social planner maximization problem. Second, I discuss what the optimal system is and compare it with the current system. Finally, I present counterfactual analyses.

VI.A Solving the Social Planner's Problem

I consider solutions of the social planners problem in the set of piece-wise linear contracts with three-brackets. That is, when solving the welfare maximization problem I aim to find the marginal replacement rate b within a sick leave duration bracket $[\underline{s}, \bar{s}]$ that maximizes welfare given the budget constraint. I restrict attention to contracts where $\underline{s}$ equals three days. That is, I constraint the solution to those contracts that reproduce the bracket systems summarized in Table 1.[44] Motivated by the features of such contracts, I also assume that transfers are nondecreasing, i.e., $B(s+1) \geq B(s)$. Additionally, I constrain the system to be at most as generous as the full-coverage case: $B(s) \leq s$.

The derivation of the optimal contract relies on the estimates of workers' preference parameters discussed in the previous section and requires an estimate of risk preferences and the parameters of the production function. In the baseline estimates, I consider the case where the pareto weights are the same for all workers.

Risk Aversion. Identification of γ would require, for instance, 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. I assume that $\gamma = 2$ and present results with two alternative specifications that allow for preference heterogeneity.

Production costs. To quantify the cost of the production losses associated to the changes in the sick pay policy, I need an estimate of the toll of sickness on workers' productivity, i.e., an estimate of ν. Unfortunately, my data do not allow a direct estimate. Thus, I rely on the estimate proposed by Maestas et al. (2021). This estimate is based on the American Working Conditions Survey (AWCS), which asks a nationally representative sample of U.S. adults to estimate their reduced

[44]Using the proposed framework, one could allow $\underline{s}$ to be a choice variable. Nonetheless, adding the brackets' limit as an additional choice variable increases the dimensionallity of the problem quickly. For example, if we consider the set of contracts with two brackets, and allow bracket's limits to take any value between 2 and 30 days there are 406 contracts. Then, for each contract one should find the vector of replacement rates that maximize welfare.

work productivity when working sick. In the main estimates of the optimal policy, I calibrate $\nu = 0.77$ and consider sensitivity checks regarding this assumption.

VI.B The Optimal Paid Sick Pay Contract

Figure 10 presents the total payment function implied by the optimal sick pay contract. The total payment function could be interpreted as the monetary payment a worker with wage $w = 1$ would receive for a sick leave claim with the indicated duration. The optimal policy differs from the current system in three key ways. First, it offers partial replacement, with an average replacement rate of 0.36, for claims of up to three days. This shift increases the utility of workers who would not take sick leave under the current system but do under the optimal policy. At the same time, partial coverage constraints moral hazard since most of the cost of those absences is faced by workers.

Second, the optimal policy eliminates the discontinuity at 11 days and exhibits a higher average replacement rate between 4 and 10 days. This feature curbs the cost of the behavioral responses to the program incentives and provides more risk protection. Implementing the optimal scheme would shift the distribution of sick leave duration relative to the distribution of claims under the current Chilean system: workers would be more likely to file sick leave claims between 8 and 10 days and less likely to file claims for 11 days.

Third, the optimal policy does not offer full replacement for sick leave claims longer than 11 days. The average replacement rate is increasing, as in the current system, but is less generous for longer claims. Taken together, these changes in the replacement rate reflect that workers value a contract that offers more protection for shorter claims to smooth consumption across different health states. I estimate that workers are willing to give up 1.53% of their earnings to be insured under the optimal policy.

Changes in compliance cost function. In this section, I examine how the optimal sick pay policy changes when workers' are more (less) reluctant to extend sick leave claims. First, *all else equal*, I reduce the cost of filing a sick leave claim longer than the health state (θ). For example, the cost of filing a sick leave claim for an extra day (f_1) is given by:

$$f_1 = f(s = \theta + 1; \theta) = \kappa_0^i + \kappa_1 \ .$$

I use the estimates of $(\mu_{\kappa_0}, \sigma_{\kappa_0}, \kappa_1, \kappa_2, \kappa_3)$ and construct a new distribution of compliance costs by shifting the mean of κ_j such that $E(\kappa_j) = \mu_{\kappa_j} \times (1 + \varepsilon)$ for $\varepsilon = 0.10$. The second exercise considers the case where $\varepsilon = -0.10$.

Using these counterfactual compliance costs distributions, I first show workers' choices assuming that they are insured under the Chilean paid sick leave system. Panel (a) of Figure 11 shows that when compliance costs are higher, the average number of days on leave more closely reflects workers' health state. In contrast, when compliance costs are low, workers' ask for longer claims, given their health state. That is, in the scenario with low compliance costs, the average duration of a sick leave claim for a given health shock is longer.

Panel (b) of Figure 11 presents the optimal policy for higher compliance costs and the policy for lower compliance costs and compares it to that in the benchmark case. This exercise provides two main lessons. First, when workers use sick leave claims that closely reflect recovery times, their duration is shorter and the optimal contract is more generous. That is, the optimal contract offers more coverage for all sick leave claims using the same budget as the baseline policy. Financing a higher level of coverage is possible due to shorter sick leave claims—this is a mechanical effect—and smaller production losses. Second, when workers' are more prompt to extending sick leave claims, the optimal policy aims to contain these responses by lowering coverage for all durations. The new policy is bellow the baseline case. This reduction is more marked for longer sick leave claims. This result indicates that the savings from providing less coverage to longer claims outweighs its utility costs—a reduction in coverage lowers the utility value of sick pay provision.

VII Conclusions

This paper addresses a relevant but poorly understood question in the provision of social insurance: What is the optimal paid sick leave system? I answer this question by combining a unique dataset on sick pay utilization and a model of insurance provision. I start by providing descriptive evidence of the main determinants of workers' behavior. I show three main empirical facts (i) workers' sick leave claim utilization varies with age and occupation; (ii) workers respond to the discontinuity in the replacement rate bunching at 11 days; (iii) workers respond to nonmonetary shifts in the temptation to extend their time off through the weekend.

Based on these facts, I develop a model of sick pay provision. The model gives three main insights. First, workers demand sick pay by trading off the utility cost of working while sick with the consumption loss from missing work when taking sick leave. The provision of sick pay lowers the cost of absences, increasing sick pay utilization. This trade-off governs the moral hazard cost of insurance provision. Second, sick leave insurance could generate production externalities arising from extended absences and workers showing up sick (when their productivity is lower). Third, the model provides intuition on the trade-off faced by the social planner (the insurer). The optimal policy balances the benefits of risk protection with the cost associated with moral hazard and production losses.

I use the estimated model of workers' behavior to determine the sick paid leave system that maximizes aggregate welfare. I limit attention to those payment schemes of the piece-wise linear family. The optimal policy differs from the current system in three key ways. First, it offers partial replacement, with an average replacement rate of 0.36, for claims of up to three days. Second, the optimal policy eliminates the discontinuity at 11 days and exhibits a higher average replacement rate between 4 and 10 days. Doing so curbs the cost of the behavioral responses to the program incentives and provides more risk protection. Third, the optimal policy does not offer full replacement for sick leave claims longer than 11 days. The average replacement rate is increasing, as in the current system, but it is less generous for longer claims. I estimate that workers are willing to give up 1.53% of their earnings to be insured under the optimal policy.

The empirical application of this paper exploits the Chilean context but the insights are informative in other contexts and more generally for the discussion on sick pay policy design. Many paid sick leave systems use the replacement rate as the relevant policy parameter. This paper provides a framework to study and quantify the main trade offs that arise when considering changing this rate.

References

Aguayo Ormeño, Irina. 2019. "Estadísticas de Jornada Laboral de Trabajadores del Sector Privado." Comisión Trabajo y Previsión Social de la Cámara de Diputados.

Böckerman, Petri, Ohto Kanninen, and Ilpo Suoniemi. 2018. "A kink that makes you sick: the effect of sick pay on absence." *Journal of Applied Econometrics*, 33(4): 568–579.

Cabral, Marika, and Marcus Dillender. 2020. "The Impact of Benefit Generosity on Workers' Compensation Claims: Evidence and Implications." NBER Working Paper No. 26976.

Campolieti, Michele, and Douglas E Hyatt. 2006. "Further evidence on the "Monday Effect" in workers' compensation." *ILR Review*, 59(3): 438–450.

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.

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.

Einav, Liran, and Amy Finkelstein. 2018. "Moral Hazard in Health Insurance: What We Know and How We Know It." *Journal of the European Economic Association*, 16(4): 957–982.

EsSalud. 2014. "Normas y Procedimientos Para la Emision, Registro Y Control De Las Certificaciones Médicas Por Incapacidad y Maternidad." Seguro Social de Salud.

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.

Ho, Kate, and Robin S Lee. 2020. "Health insurance menu design for large employers." NBER Working Paper No. 27868.

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.

Kleven, Henrik J. 2021. "Sufficient Statistics Revisited." *Annual Review of Economics*, 13: 515–538.

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." NBER Working Paper No. 26832.

Maestas, Nicole A, Kathleen J Mullen, and Stephanie Rennane. 2021. "Absenteeism and Presenteeism Among American Workers." *Journal of Disability Policy Studies*, 32(1): 13–23.

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.

Pauly, Mark V. 1968. "The Economics of Moral Hazard: Comment." *The American Economic Review*, 58(3): 531–537.

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.

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

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

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

Thoursie, Peter Skogman. 2004. "Reporting Sick: Are Sporting Events Contagious?" *Journal of Applied Econometrics*, 19(6): 809–823.

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. 2010. "A natural Experiment on Sick Pay Cuts, Sickness Absence, and Labor Costs." *Journal of Public Economics*, 94(11-12): 1108–1122.

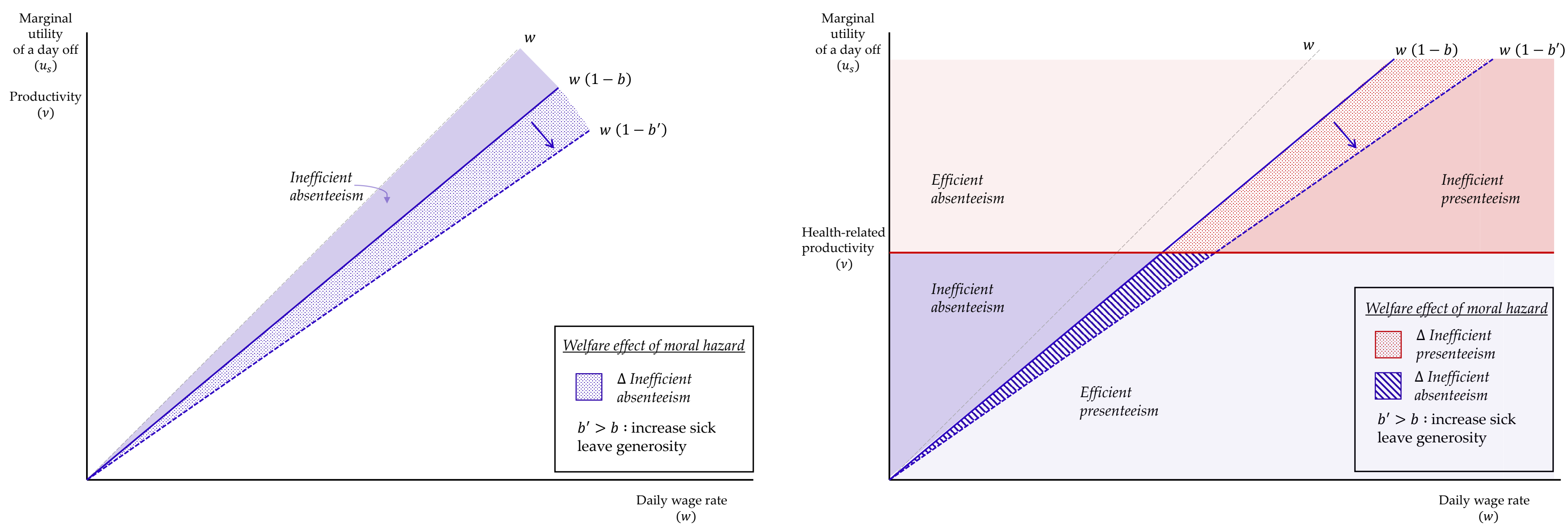
Ziebarth, Nicolas R, and Martin Karlsson. 2014. "The Effects of Expanding the Generosity of the Statutory Sickness Insurance System." *Journal of Applied Econometrics*, 29(2): 208–230.

VIII Figures

Figure 1: The Welfare Effects of Moral Hazard

(a) No externalities case

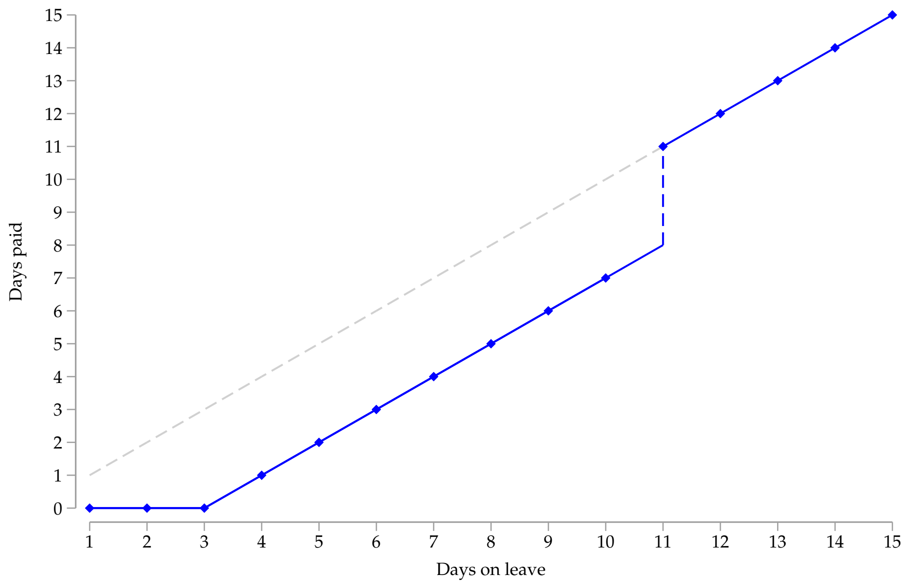
(b) Externalities case



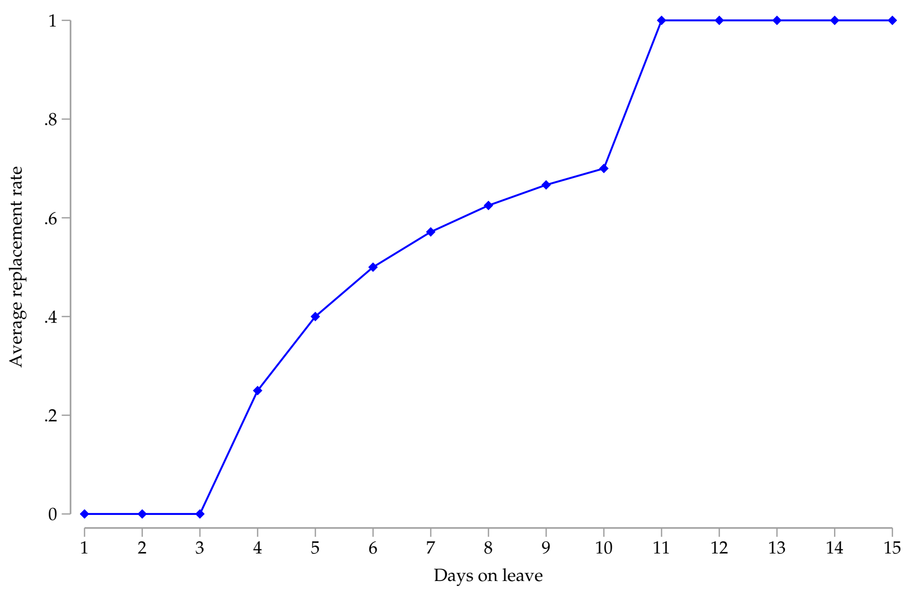
Notes: Panel (a) shows the effect of an increase in the replacement rate on absences in the no production externalities case. The provision of sick benefits gives rise to inefficient absenteeism: a pool of individuals with marginal utility for time off below their marginal product (ν^i) takes a day off. An increase in the replacement rate accentuates this response increasing inefficient absenteeism. Panel (b) presents the case where there are production externalities ($\nu < 1$). The relation between wages, productivity, and the marginal value of a day off defines four regions. The top left area corresponds to the pool of individuals who do not work ($u_s > w(1-b)$) and for whom this is efficient given their productivity ($u_s > \nu$). I refer to this pool of workers as involved in *efficient absenteeism*. The bottom right area shows the opposite situation: a pool of individuals who do work ($u_s < w(1-b)$) and for whom this is the efficient response ($u_s < \nu$). I refer to this pool of workers as involved in *efficient presenteeism*. The other two (darker) areas show inefficient absences ($u_s > w(1-b)$ and $u_s < \nu$) and inefficient work ($u_s < w(1-b)$ and $u_s > v$). This figure is referenced in Section II.C.

Figure 2: Chilean Paid Sick Leave System: Benefits Computation

(a) Days Paid as a Function of Days on Leave

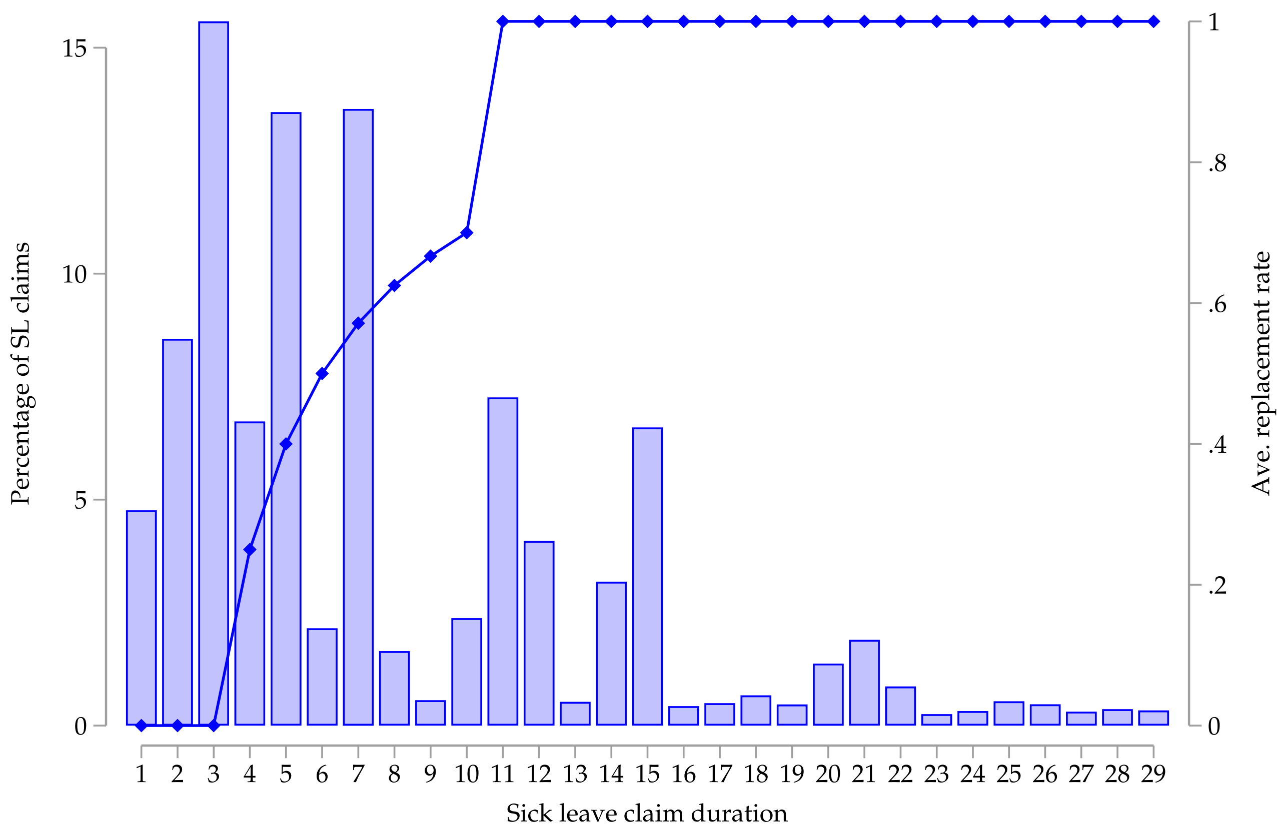


(b) Average Replacement Rate



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. The replacement rate for the first three days of a sick leave spell is zero. Starting on the fourth day, there is full coverage of each missed day—i.e., the replacement rate is one. If the sick leave lasts 11 days or more, the nonpayable period is reimbursed. 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.

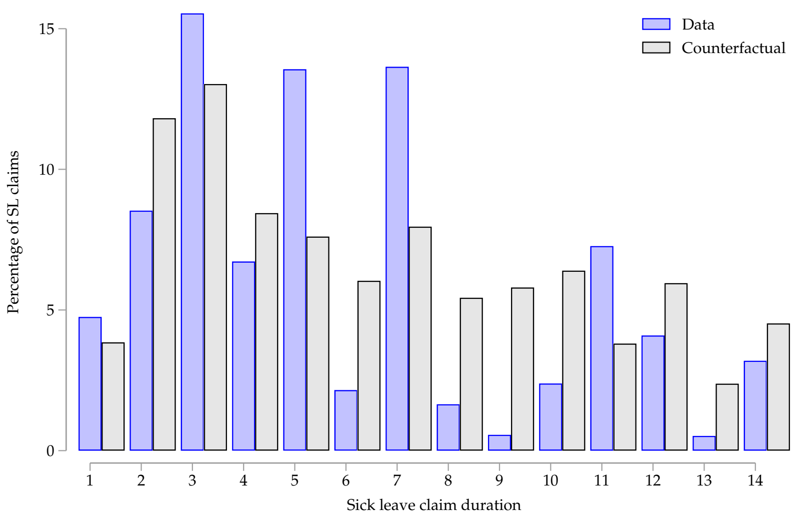
Figure 3: Duration of Sick Leave Claims



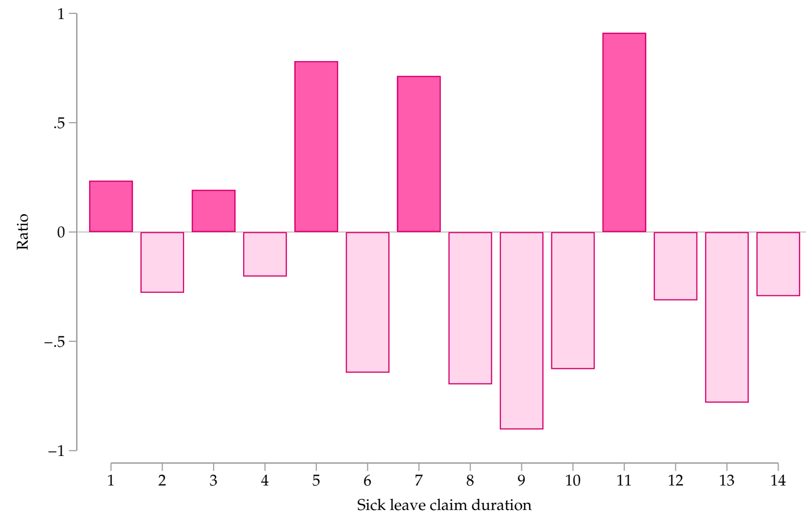
Notes: This figure shows the distribution of the duration of sick leave claims made by male workers on the left-hand-side vertical axis (bars) and the average replacement rate on the right-hand-side vertical axis (solid line). The figure includes only sick leave claims of up to 29 days; these represent 89% of all claims. This figure is referenced in Section III.D.

Figure 4: Sick leave duration: data and counterfactual distribution

(a) Distributions: data vs. counterfactual

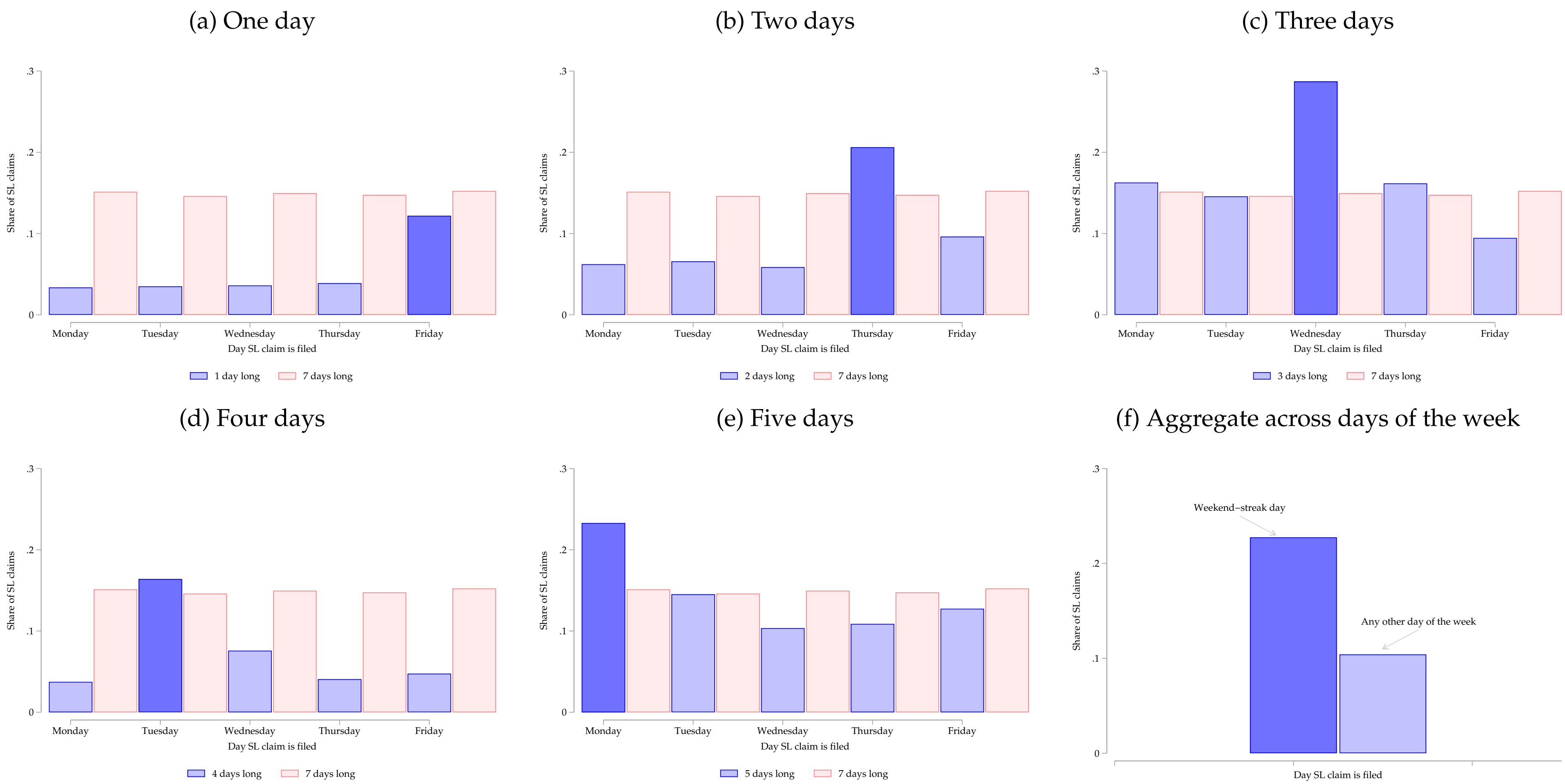


(b) Proportional differences



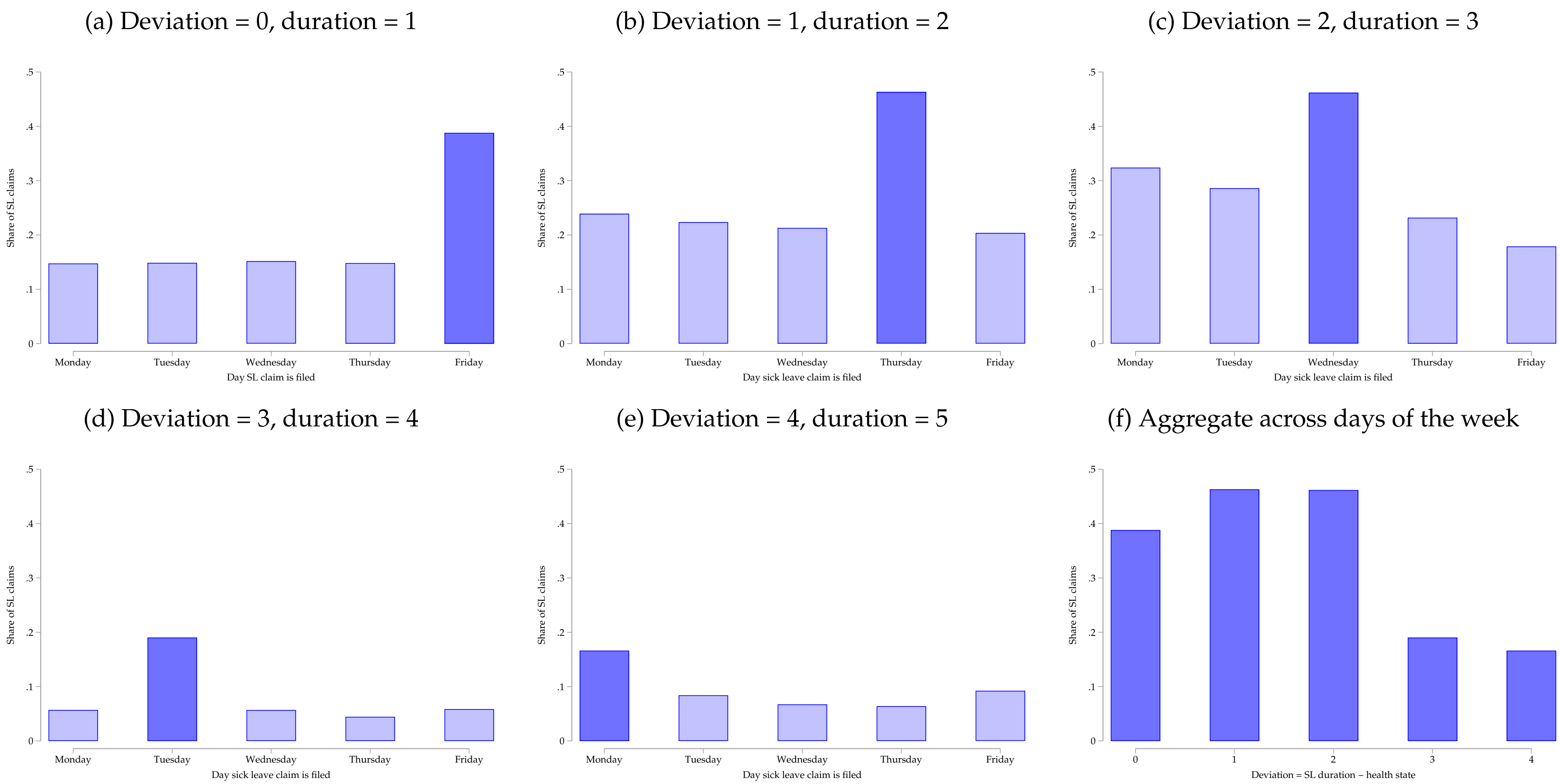
Notes: This figure shows the distribution of days on leave coming from the data, as shown in Figure 3, and the counterfactual distribution of days on leave. The latter is constructed assigning to each sick leave the recovery time suggested by the Peruvian Handbook of Recovery Times, adjusted by worker age and occupation. This figure is referenced in Section III.D.

Figure 5: Days of the week and sick leave claim duration



Notes: Panels (a) to (e) show the share of sick leave claims with duration s and the share of seven-day-long sick leave claims filed on each day of the week. Panel (f) aggregates across durations and days of the week: The first bar—labeled "weekend streak"—averages the share of one- to five-day-long sick leave claims that end on a Friday and are filed on any day of the week (for example, one-day-long claims filed on a Friday, two-day-long claims filed on a Thursday, and so on). The second bar—labeled "non–weekend streak"—averages the share one- to five-day-long sick leave claims filed on any other day of the week (for example, two-day-long claims filed on a Friday). Table A9 reports the estimated shares and moments. This figure is referenced in Sections II.A and IV.D.

Figure 6: Identification of compliance cost function: Sick leave claims by duration and day of the week (one-day recovery time)

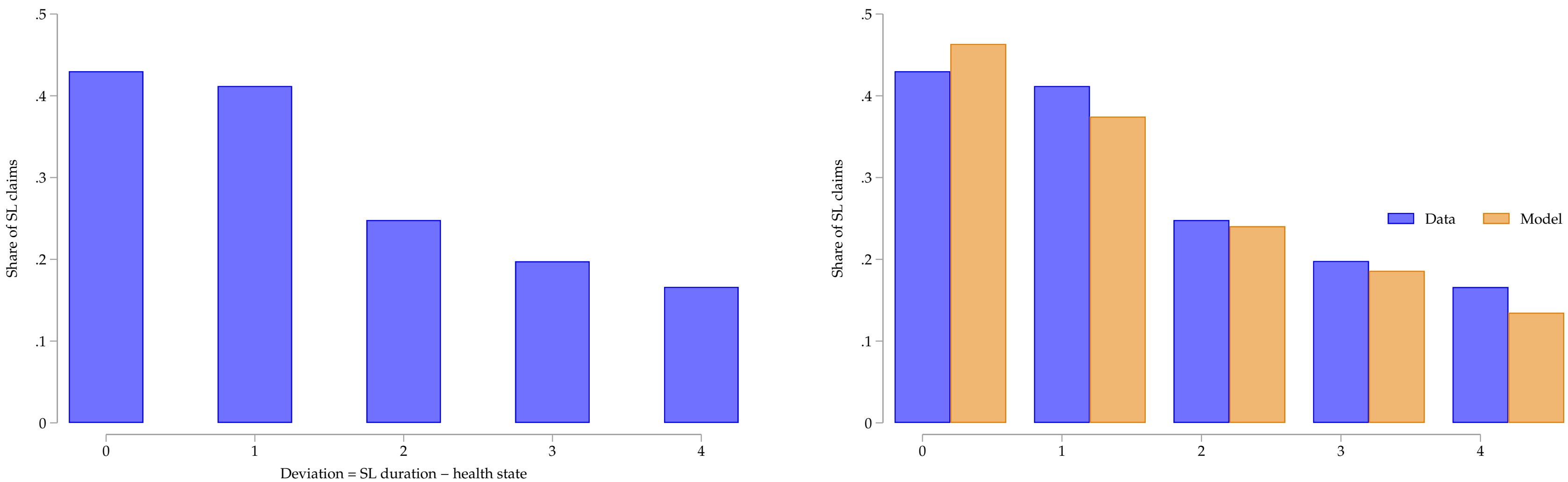


Notes: Panels (a) to (e) show the share of sick leave claims with duration s for workers whose main diagnosis would imply a health state of 1 day on leave. Panel (f) aggregates the share of sick leave claims across days of the week, including only weekend-streak combinations; e.g., from Panel (a), I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

Figure 7: Compliance cost function

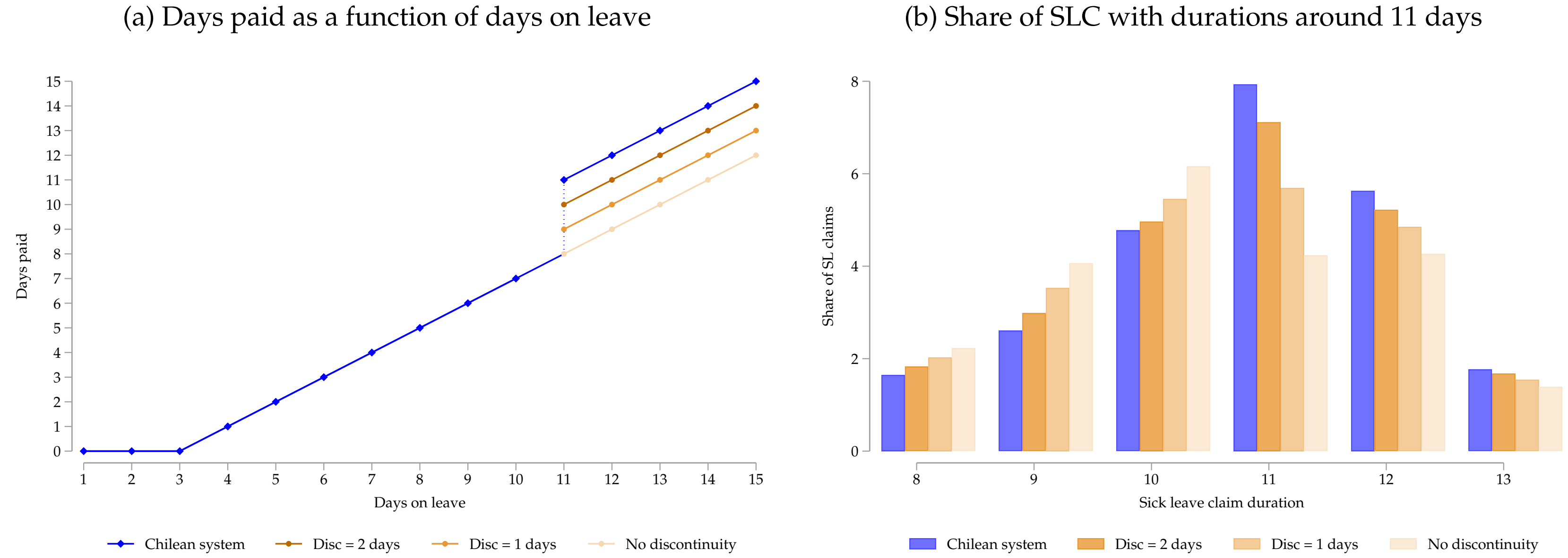
(a) Share of sick leave claims as a function of deviation

(b) Model fit: Data vs. model moments



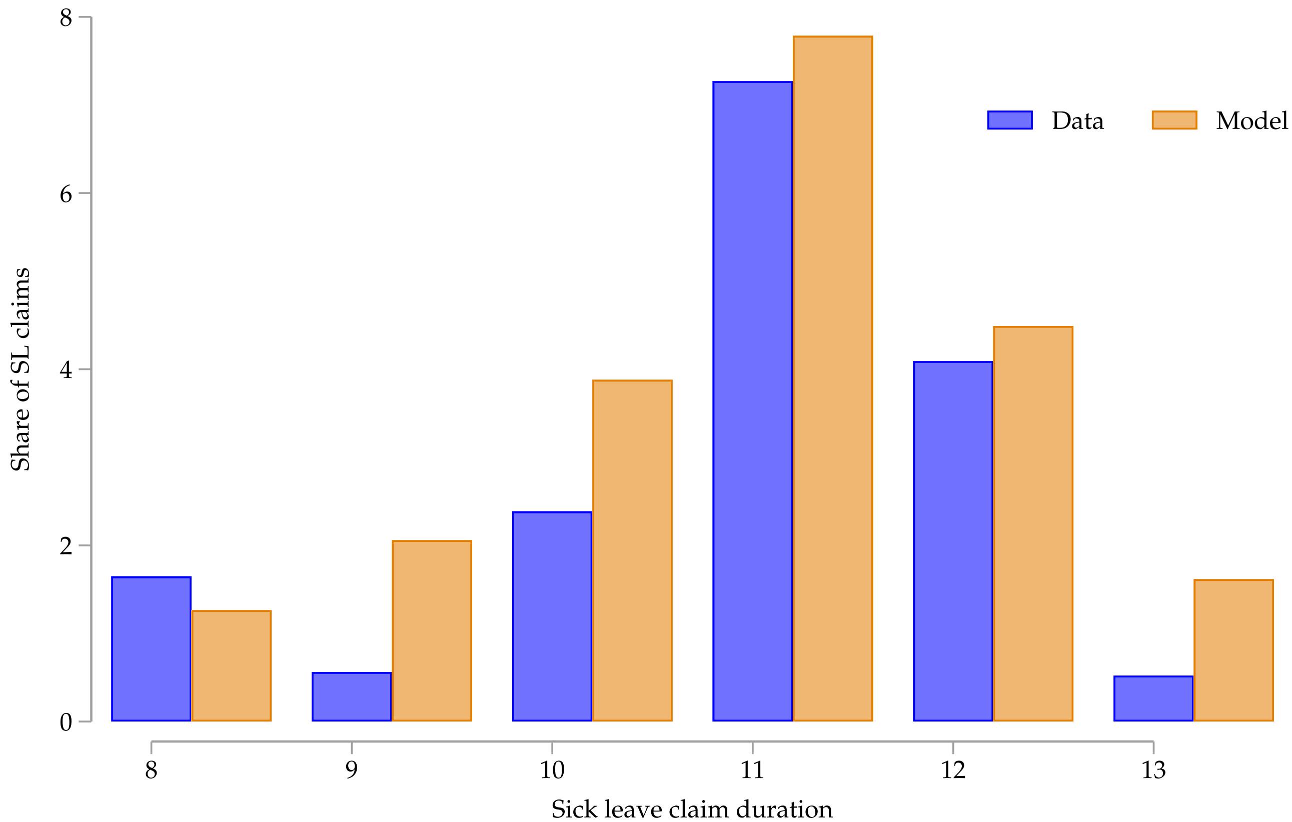
Notes: Panel (a) shows the average share of sick leave claims with deviations between 0 and 4 days. The average is computed over sick leave claims with primary diagnoses requiring 1, 2 or 3 days of rest filed on weekend-streak days. Each column is the weighted average of the probability for each health state. Panel (b) replicates this figure and adds the moments computed from the simulated data. This figure is referenced in Section IV.D.

Figure 8: Workers' behavior: Changes in the discontinuity at 11 days



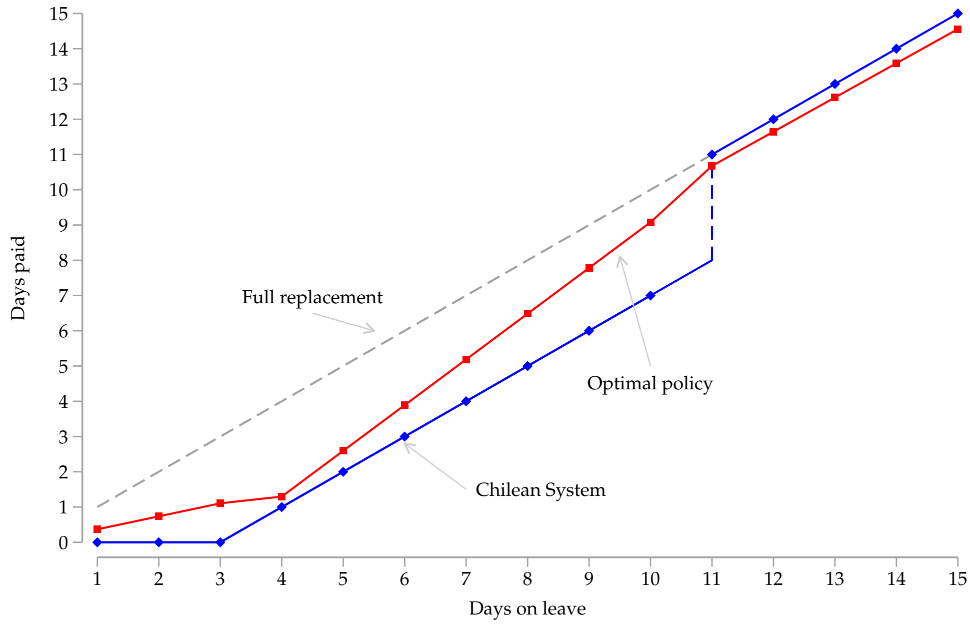
Notes: This figure shows workers responses to changes in the paid sick leave benefit scheme, relative to the current Chilean system. I consider three alternative systems that keep the marginal replacement rates in each bracket fixed—the slope of each function is one for claims above 4 days—and reduce the size of the jump at 11 days. Panel (a) presents the alternative payment schemes and Panel (b) shows the share of SLC filed under each alternative payment scheme. This figure is referenced in Section V.A.

Figure 9: Model's fit: 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 neighborhood of 11 days from the data and a model-simulated sample. Using the measure of heaping proposed by Roberts and Brewer (2001), I estimate that, in the data, the 11-day duration accumulates an additional 4.50% mass than its neighbors. Using the model-simulated sample, I estimate an additional 4.03% mass relative to its neighbors. This measure approximates how much a duration 'sticks out' from the pattern suggested by its neighbors. Roberts and Brewer (2001) proposes: $h_z = f(z) - \frac{f(z-1)+f(z+1)}{2}$, where z corresponds to 11 days, and $f(\cdot)$ indicates the frequency of sick leave claims with duration z. Thus, h_z gives the difference between a duration frequency and the average of the frequencies of the two immediately neighboring duration. This figure is referenced in Section V.C.

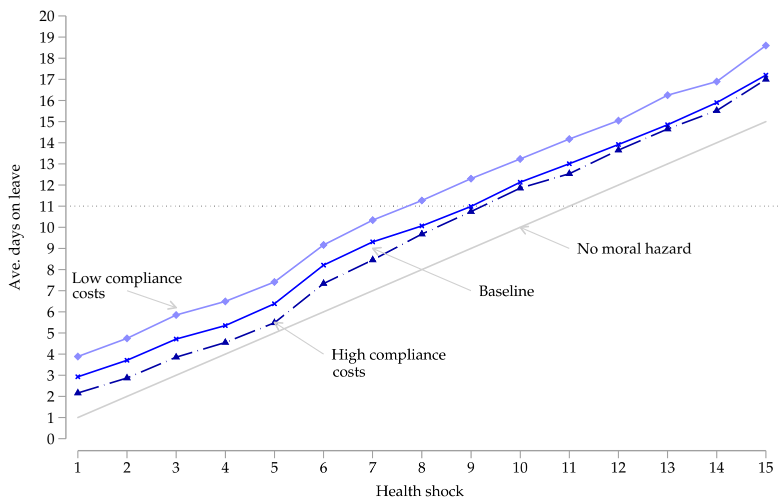
Figure 10: The Optimal Sick Pay System



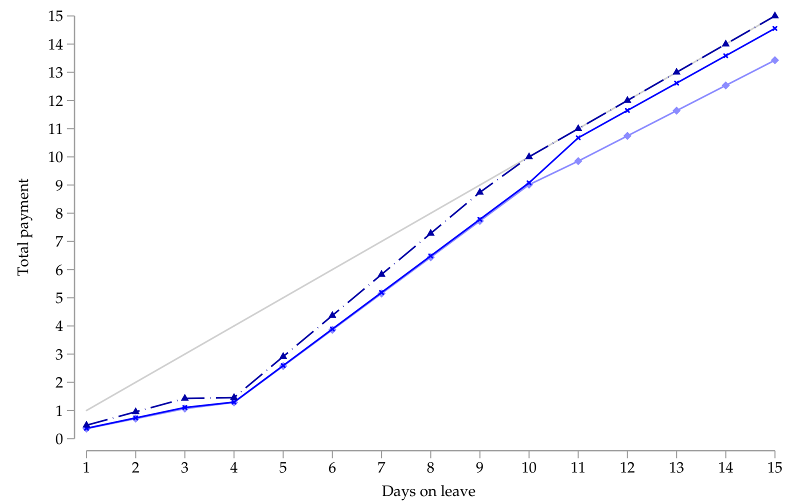
Notes: This figure compares the Chilean system with the optimal paid sick leave system. It presents the total payment function under each contract. This function could be interpreted as the monetary payment a worker with wage $w = 1$ would receive for a sick leave claim with the duration indicated in the horizontal axis. That is, this presents sick pay as a function of the duration of a claim for a unitary wage. This figure is referenced in Section V.A.

Figure 11: Counterfactual exercise: Changes in compliance cost

(a) Workers' responses: Ave. number of days on leave



(b) Optimal paid sick leave contract



Notes: This figure presents the optimal policy under alternative distributions of compliance costs. Panel (a) summarizes workers behavior. It shows the average duration of sick leave claims for the estimated distribution of compliance costs and for two alternative distributions. To construct the high compliance costs distribution, I increase the average cost of an extra day off by 10%. Similarly, the low compliance costs distribution is constructed by decreasing the average cost of an extra day off by 10%. Panel (b) shows the optimal policy under each of these distributions. This figure is referenced in Section V.A.

IX Tables

Table 1: Paid Sick Leave Systems

Country	Design	Benefits' computation
Bulgaria	bracket system	- Days 1 to 3: rep. rate = 0.7
		- Day 4 onward: rep. rate = 0.8
Chile	bracket system	- Days 1 to 3: rep. rate = 0
		- Days 4 to 10: rep. rate = 1
		- Days 11 onward: rep. rate = 1
		and the waiting period is reimbursed
Estonia	bracket system	- Days 1 to 3: rep. rate = 0
		- Day 4 and onward: rep. rate = 0.7
Finland	bracket system	- Days 1 to 9: rep. rate = 0
		- Day 10 and onward: rep. rate = 0.7
France	bracket system	- Days 1 to 3: rep. rate = 0
		- Day 4 and onward: rep. rate = 0.5
Greece	bracket system	- Days 1 to 3: rep. rate = 0.5
		- Days 4 to 30 are paid with a rep. rate of 1
Ireland	bracket system	-Days 1 to 3: 6 working days
		- Rates vary by earnings
Hungary	bracket system	- Days 1 to 15: rep rate = 0.7
		- Days 16 onward: rep. rate = 0.5
Italy	bracket system	- Days 1 to 3: rep rate = 0
		- Days 4 to 20: rep. rate = 0.5
		- Days 21 onward: rep. rate = 0.66
Portugal	bracket system	- Days 1 to 3: rep rate = 0
		- Days 4 onward: rep rate between 0.65 and 0.75
Spain	bracket system	- Days 1 to 5: rep rate = 0
		- Days 6 to 20: rep. rate = 0.60
		- Days 21 onward: rep. rate = 0.75
United Kingdom	bracket system (hybrid)	- Waiting period = 3 days , rep. rate = 0
		- 99.35 pounds per week
Denmark	linear system	Replacement rate = 0.9
Netherlands	linear system	Replacement rate = 0.7
Norway	linear system	Replacement rate = 1
Poland	linear system	Replacement rate = 0.8
Switzerland	linear system	Replacement rate = 1
Germany	linear system*	- Days 1 to 42 : rep. rate = 1
		- Week 7 onward: rep. rate = 0.7
Australia	credit account	10 sick days per year
Austria	credit account	Six weeks full paid sick leave
Belgium	credit account	30 sick days per year
United States	credit account	Average private-sector: $<$ 10 days per year

Notes: This table summarizes sick paid systems for 22 countries. Bracket system refers vary the replacement rate based on the duration of a leave. Linear system feature a constant replacement rate. Credit account refers to the case where paid leave is earned over time and unused leave accumulates, producing an employee-specific "leave balance." The UK system is hybrid because the second bracket proposes a lump sum transfer. The German system a hybrid because it features a change in the replacement rate that kicks in after a long period. This table is referenced in Secntion II.C and in Section III.B.

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

		Workers who had used SL benefits			
	All workers	Any	1 to 3 days	4 to 10 days	11 to 29 days
	(1)	(2)	(3)	(4)	(5)
Age					
Mean	43.94	42.24	40.35	42.43	44.46
Share of workers aged					
25 - 34 years old	0.26	0.32	0.38	0.32	0.25
35 - 44 years old	0.24	0.25	0.26	0.25	0.24
45 - 54 years old	0.27	0.24	0.21	0.24	0.26
55 - 64 years old	0.22	0.19	0.14	0.20	0.24
Income (monthly USD)					
Mean	772.00	918.02	1014.27	912.01	887.71
Standard deviation	367.27	390.03	415.64	381.88	372.60
25th percentile	484.45	601.77	685.83	603.06	581.80
Median	682.15	845.74	947.32	841.81	815.54
75th percentile	997.97	1161.29	1271.04	1149.22	1120.91
90th percentile	1328.04	1496.41	1619.20	1477.68	1443.80
Region (shares)					
Central	0.35	0.42	0.51	0.41	0.38
Mining intensive regions	0.09	0.08	0.05	0.07	0.10
Share of workers (%)	100	13.78	4.69	7.50	5.70
Observations	1,916,138	263,951	89,880	143,615	109,213

Notes: This table presents summary statistics for all male workers aged 25 to 64 years old. Column (1) presents characteristics for all workers in the sample, regardless of sick claim utilization. Column (2) presents characteristics for workers who have filed at least one claim with duration of up to 30 days for conditions included in the analysis. Columns (3) to (5) present characteristics of workers by duration of the sick leave claims filed. Workers can be included in more than one category based on the claims they have filed. This table is referenced in Section III.D.

Table 3: Parameter Estimates

Parameter	Description	Value	Std. error
Preferences parameters			
q	Weekend-streak utility	0.7894	0.2110
$\tilde{\mu}_{\phi}$	Value of time off relative to consumption, mean	56.8930	26.8567
$\tilde{\sigma}_{\phi}$	Value of time off relative to consumption, std. dev.	39.8246	13.1373
$\tilde{\mu}_{\kappa_0}$	Compliance costs, mean	0.8290	0.2320
$\tilde{\sigma}_{\kappa_0}$	Compliance costs, standard deviation	1.7714	0.2616
κ_1	Cost of one day deviation	0.3594	0.0960
κ_2	Cost of two days deviation	0.3217	0.0998
κ_3	Cost of three days deviation	0.0041	0.0021
Measurement error			
p_{me}	Prob. physician assigns one day more (less) than asked	0.2992	0.0430
Rounding			
p_7	Prob. SL duration is round to the closest multiple of 7	0.5746	0.1751

Notes: This table presents the estimated parameters for the model of workers' behavior. I assume that the value of time off relative to consumption is distributed log-normal, i.e., $\ln(\phi) \sim N(\mu_{\phi},\sigma^2_{\phi})$. Thus, $\phi \sim$ Lognormal$(\tilde{\mu}_{\phi},\tilde{\sigma}^2_{\phi})$. I report $(\tilde{\mu}_{\phi},\tilde{\sigma}_{\phi})$. Similarly, I assume that the compliance cost parameter κ_0 follows a log-normal distribution and I report $(\tilde{\mu}_{\kappa_0},\tilde{\sigma}_{\kappa_0})$ which are the moments that characterize: $\kappa_0 \sim$ Lognormal$(\tilde{\mu}_{\kappa_0},\tilde{\sigma}_{\kappa_0})$. The standard errors are based on 200 bootstrap simulations. This table is referenced in Section V.

Table 4: Moments used in the estimation.

Moments	Data (1)	Model (2)
Preferences parameters		
Weekend streak utility		
Weekend streak days relative to non-streak days	0.1233	0.1546
Compliance costs		
Sh. of SLC with 0 day deviation	0.4300	0.4635
Sh. of SLC with 1 day deviation	0.4120	0.3745
Sh. of SLC with 2 days deviation	0.2480	0.2401
Sh. of SLC with 3 days deviation	0.1975	0.1858
Sh. of SLC with 4 days deviation	0.1663	0.1371
Value of time outside work		
Mean time outside work to consumption ratio	0.2972	0.3342
SD time outside work to consumption ratio	0.1962	0.1965
Measurement error		
Sh. Monday SLC - sh Tuesday SLC conditional to 5-days-long and a day of recovery	0.0822	0.0802
Rounding		
Share of 7-days-long claims	0.1364	0.1050

Notes: This table presents the moments used to estimate the model's parameter. Column 2 reports the data moments. Column 3 reports simulated moments.This table is referenced in Section V.C.

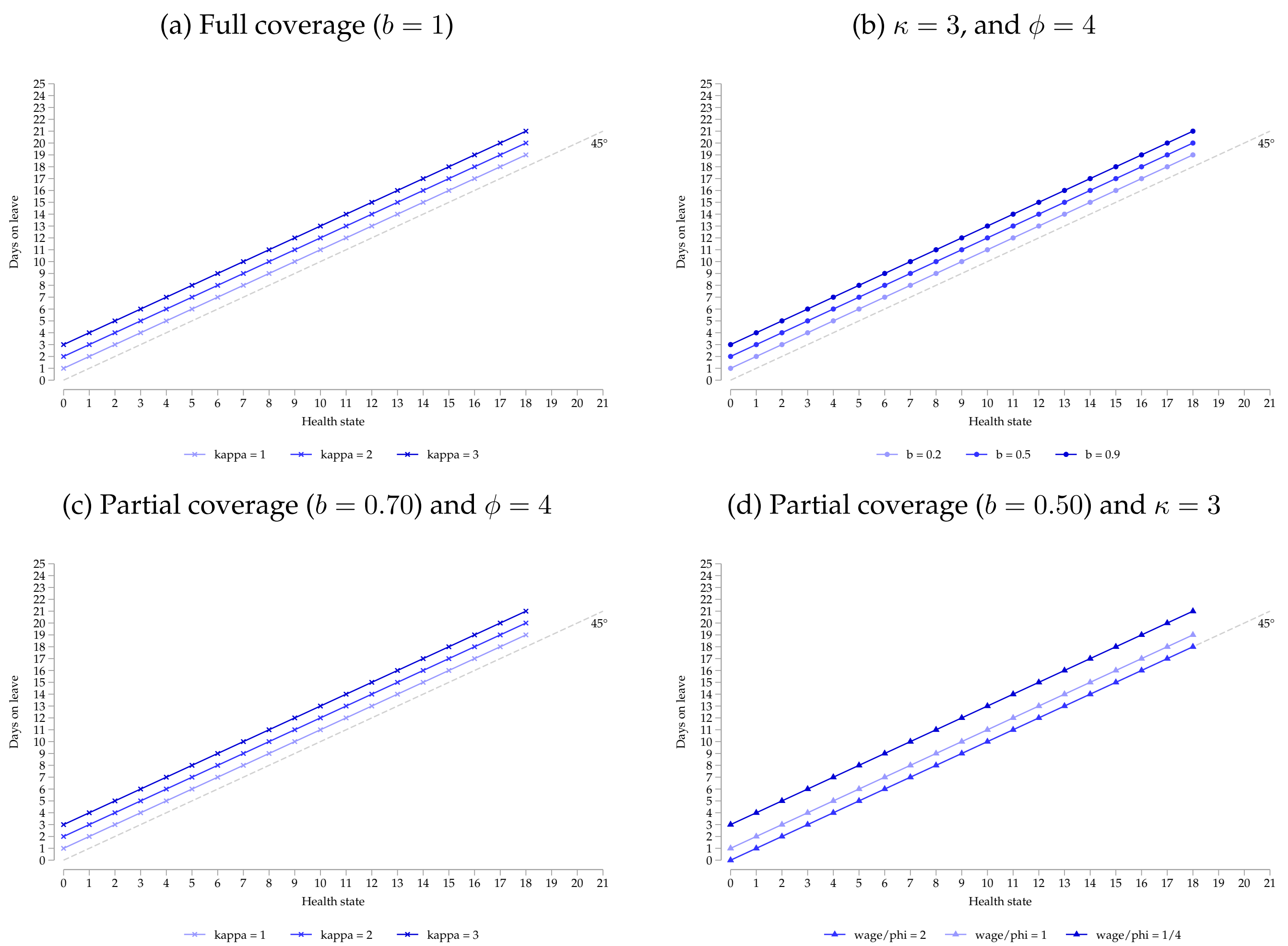
Table 5: Out-of-sample: selected moments from 2019 data and simulated counterparts

Moments	Data (1)	Model (2)
Preference parameters		
Weekend streak utility		
Weekend streak days relative to non-streak days	0.1158	0.1428
Compliance costs		
Sh. of SLC with 0 day deviation	0.4385	0.4611
Sh. of SLC with 1 day deviation	0.4223	0.3819
Sh. of SLC with 2 days deviation	0.2293	0.2330
Sh. of SLC with 3 days deviation	0.1843	0.1683
Sh. of SLC with 4 days deviation	0.1650	0.1232
Value of leisure		
Mean leisure to consumption ratio	0.2993	0.3297
SD leisure to consumption ratio	0.1985	0.1980
Share of SLC - selected durations		
7 days	0.1305	0.1053
8 days	0.0162	0.0162
9 days	0.0057	0.0262
10 days	0.0211	0.0461
11 days	0.0834	0.0939
12 days	0.0404	0.0535
13 days	0.0051	0.0172

Notes: This table presents results from an out-of-sample test of the model performance. Column (1) is constructed using data on sick leave claims filed in 2019. The moments included correspond to those used for the estimation of preference parameters and the shares of claims with duration in the neighborhood of 11 days. To test the performance of the model, I compare these moments with their model-simulated counterparts. These are presented in column (2). To construct column (2) I simulate the model based on a representative sample drawn from the 2019 data and the estimated vector of preference parameters. This table is referenced in Section V.C.

Appendix A. Additional Figures and Tables

Figure A1: Sick Pay Utilization with Linear Contract and Quadratic Penalties



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

(b) $\kappa = 3$, and $\phi = 4$

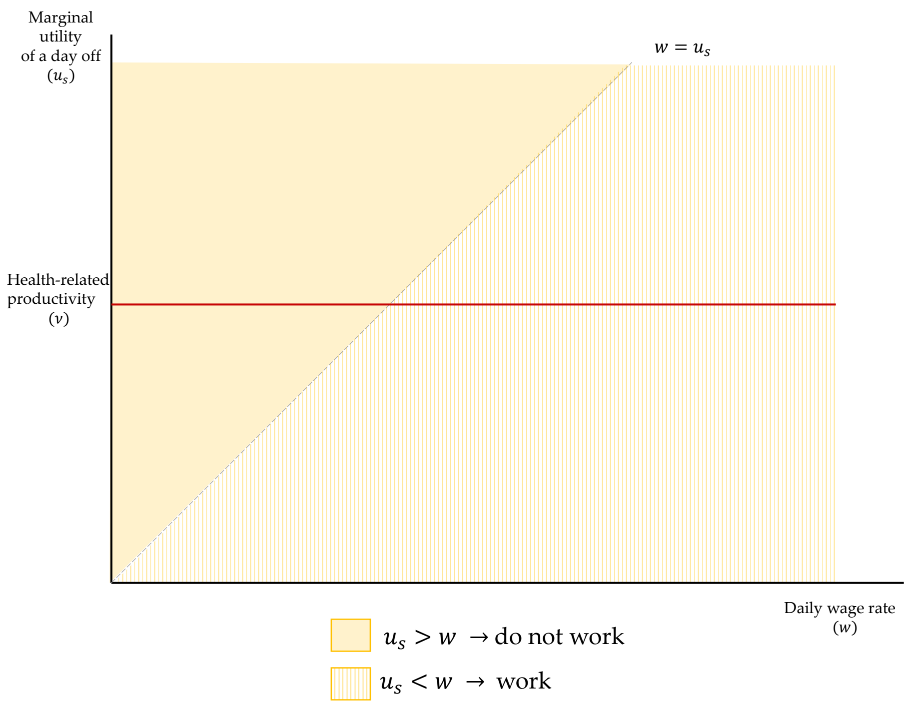
(c) Partial coverage ($b = 0.70$) and $\phi = 4$

(d) Partial coverage ($b = 0.50$) and $\kappa = 3$

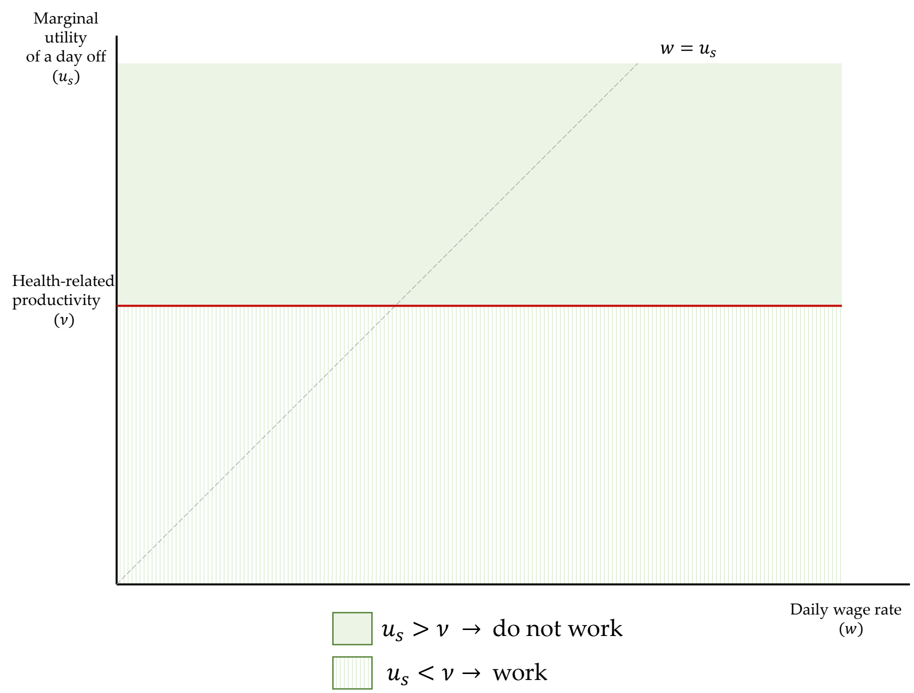
Notes: This figure shows the optimal demand of days on leave $s^*(\theta)$ as a function of worker's health status (θ) under the assumptions of linear contracts and quadratic penalties with different levels of coverage. It illustrates the definition of moral hazard given by equation (2). Panels (a) and (b) show the full coverage and no coverage cases. Panels (b) and (c) show choices when partial coverage is offered for alternative values of κ and $wage/\phi$. This figure is referenced in Section II.

Figure A2: Worker vs Social Planner trade-offs: externalities and no insurance provision

(a) Worker trade-off

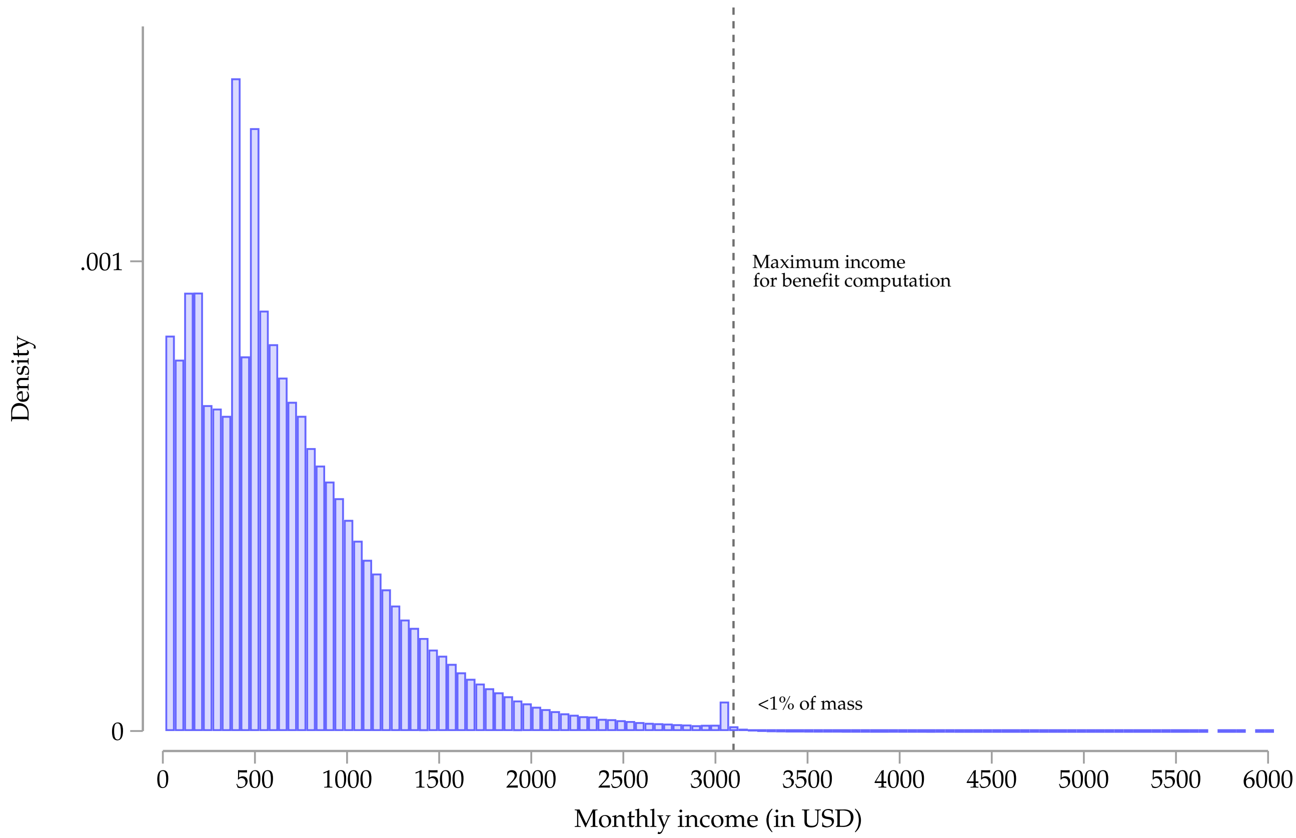


(b) Social Planner trade-off: optimal employment



Notes: Panel (a) summarizes workers' choices. Absent of sick pay, worker i takes a day off if $u_s^i > w^i$, this corresponds to the solid fill area. Note that some workers optimally choose to work even if their value of a day off is above their productivity (ν). This is the pattern fill area located above the horizontal line. This is a consequence of the fact that wages do not longer reflect productivity. Panel (b) shows the optimal employment decision. This trade-off compares the productivity of working with the value of a day off. Absent of sick pay, it would be efficient that worker i takes a day off if her valuation is above her marginal product when sick, i.e., when $u_s^i > \nu$. This corresponds to the solid fill area located above the horizontal line. Note that some workers would find optimal to not work regardless: those with $u_s^i > w^i$. This figure is referenced in Section II.C.

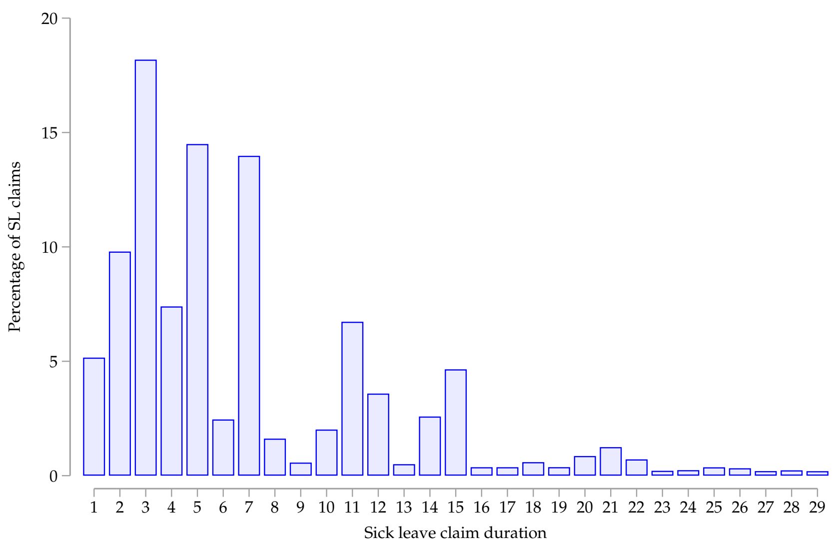
Figure A3: Distribution of monthly income among eligible workers



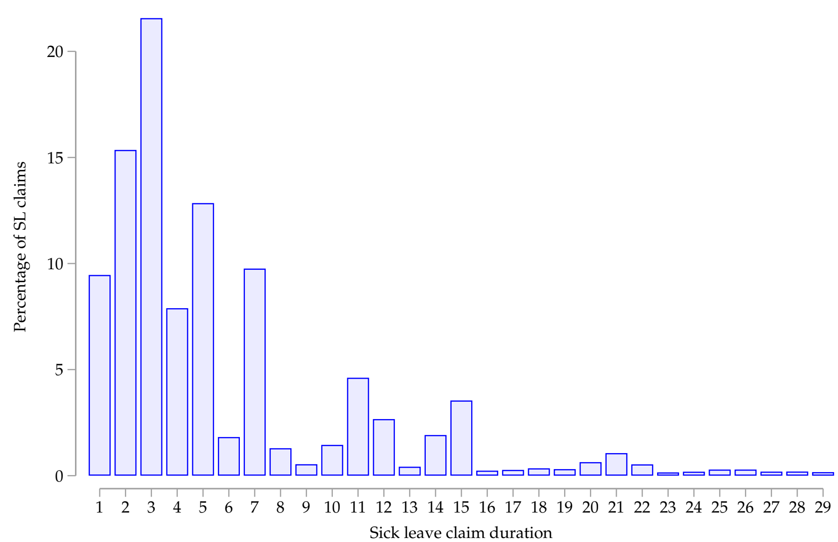
Notes: This figure shows the distribution of monthly income (in USD) for workers eligible to file a sick leave claim in 2017. The vertical line indicates the income level associated with the maximum benefit threshold. This figure is referenced in Section III.B.

Figure A4: Histogram of days on leave by worker characteristics

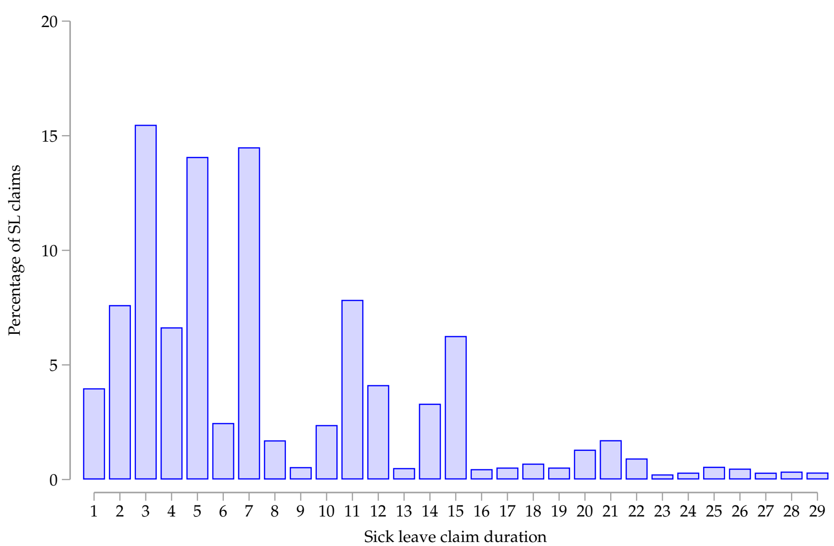
(a) 25–34 years old, blue-collar



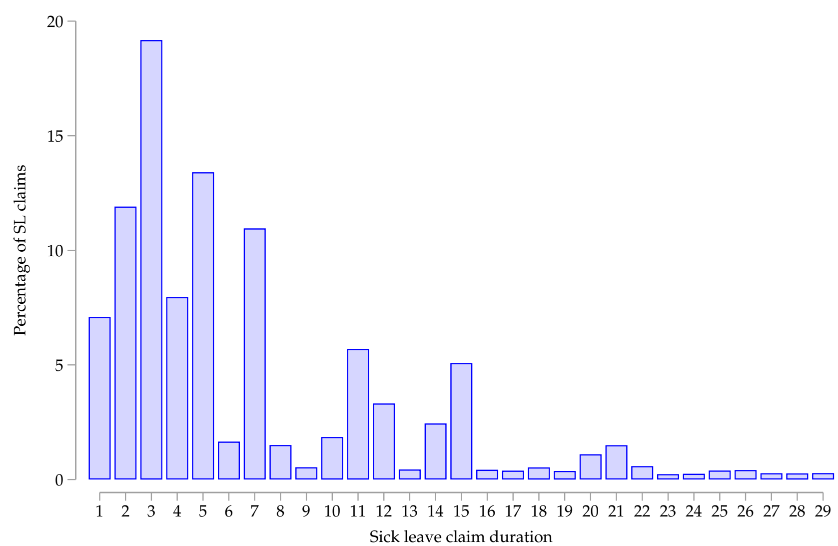
(b) 25–34 years old, white-collar



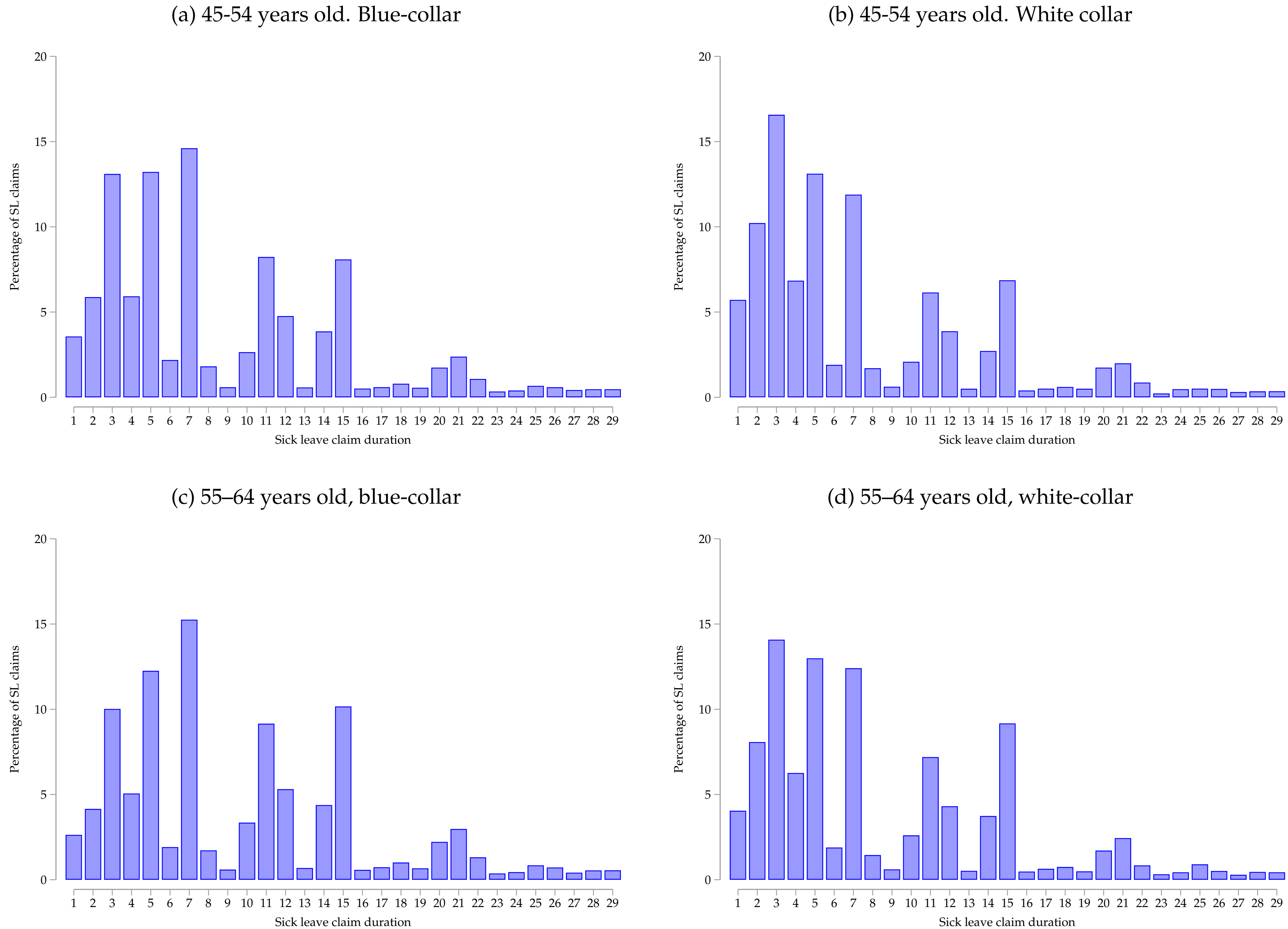
(c) 35-44 years old. Blue-collar



(d) 35-44 years old. White collar



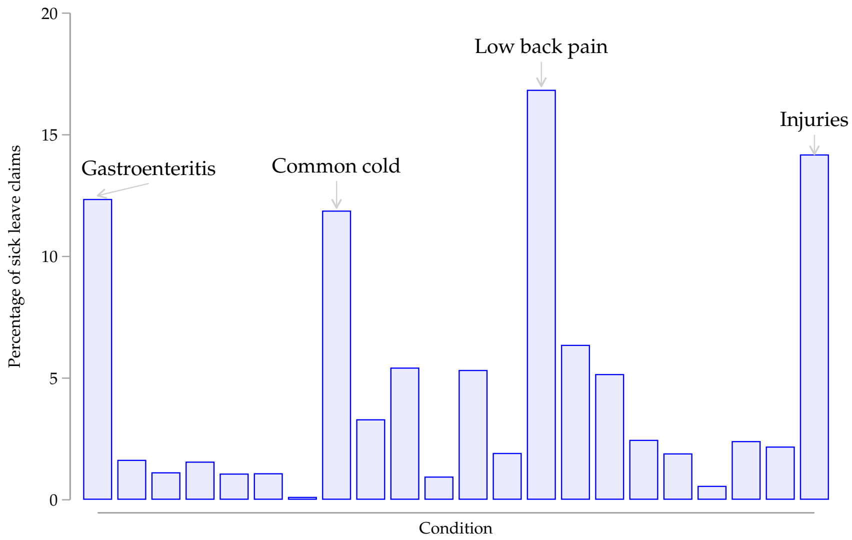
(continue)



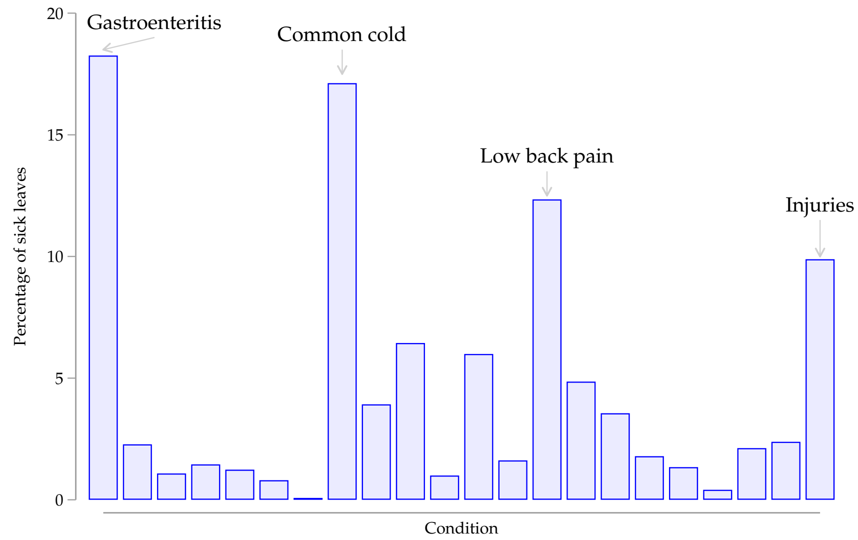
Notes: This figure shows the distribution of days on leave by workers' age and occupation. Blue-collar worker refers to workers who engage in manual labor, typically agriculture, manufacturing, construction, mining, or maintenance. White-collar worker refers to workers whose daily work activities do not involve manual labor—e.g., teachers or administrative staff. This figure is referenced in Section III.C and in Section IV.A.

Figure A5: Histogram of Health Conditions by Workers Characteristics

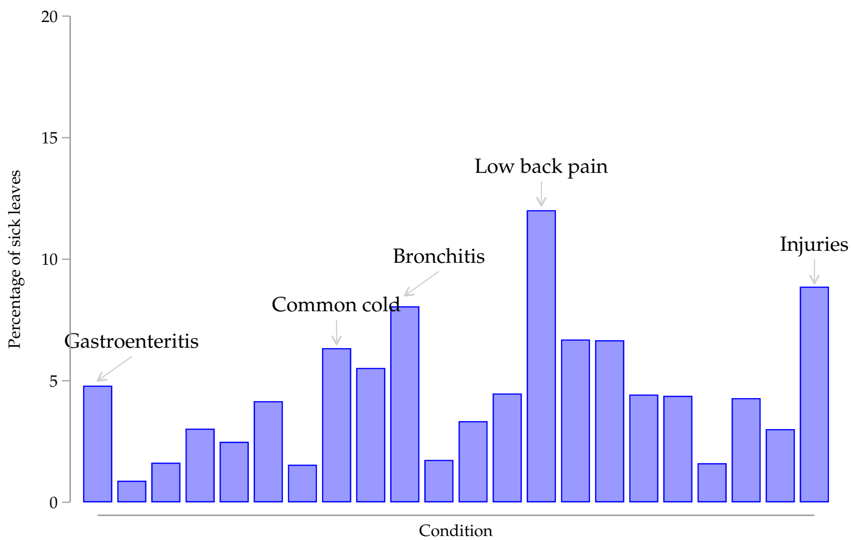
(a) 25-34 years old. Blue-collar



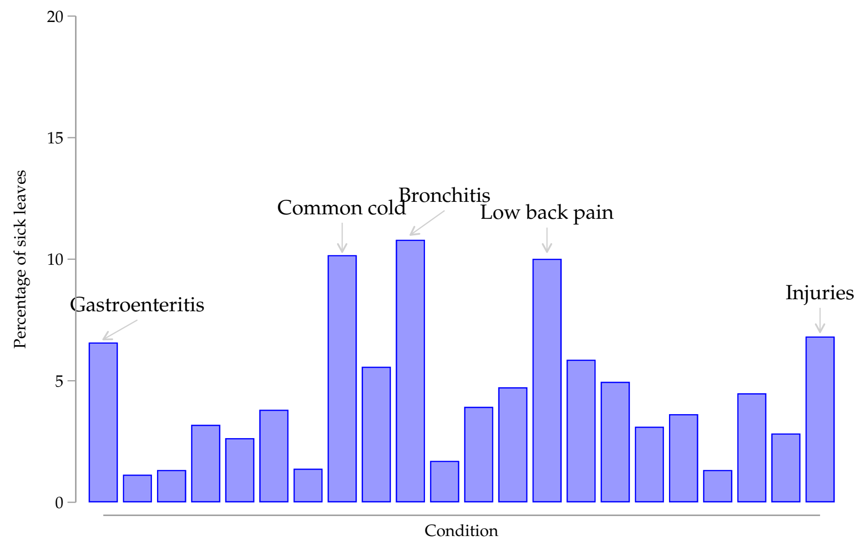
(b) 25-34 years old. White collar



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

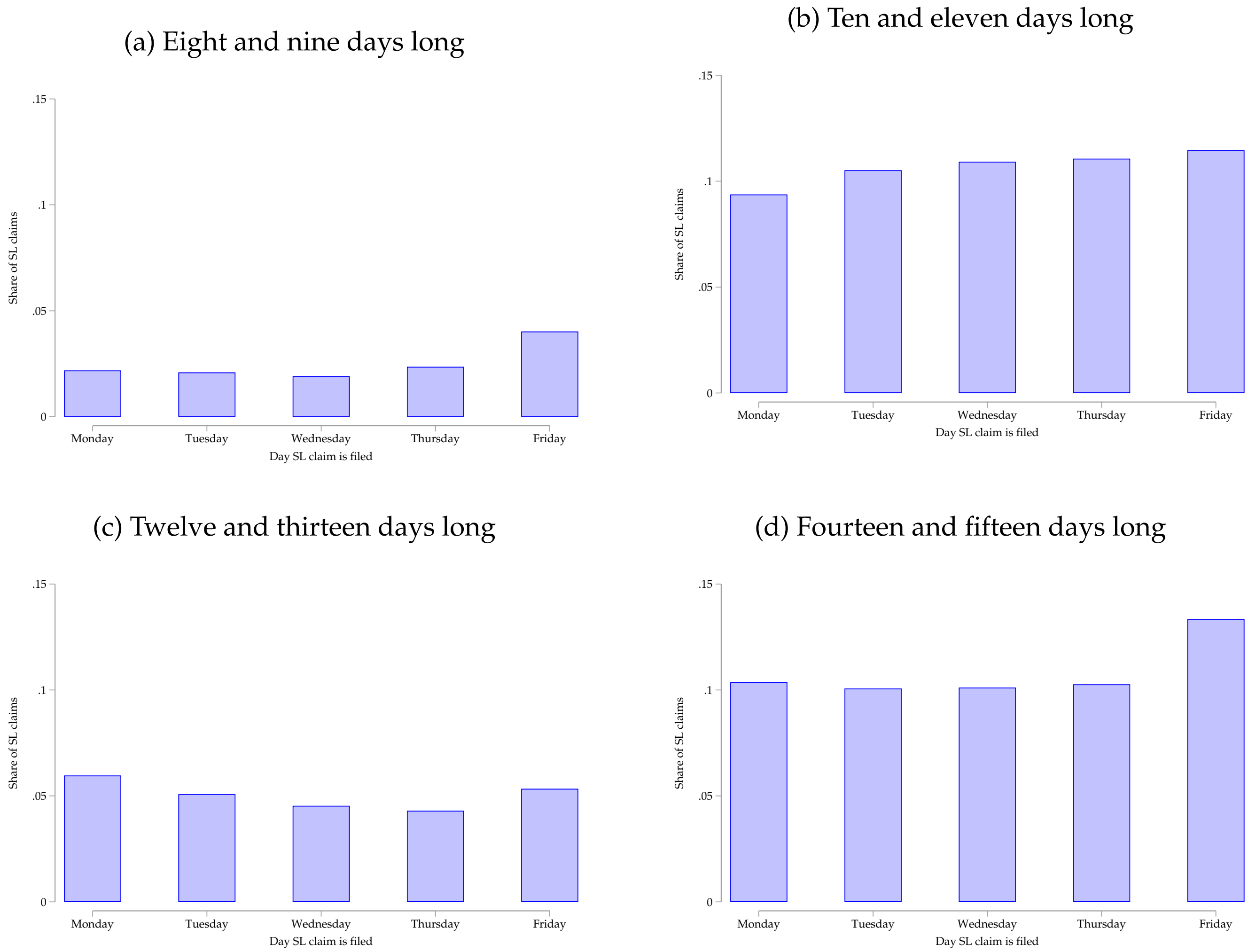


(d) 55-64 years old. White collar



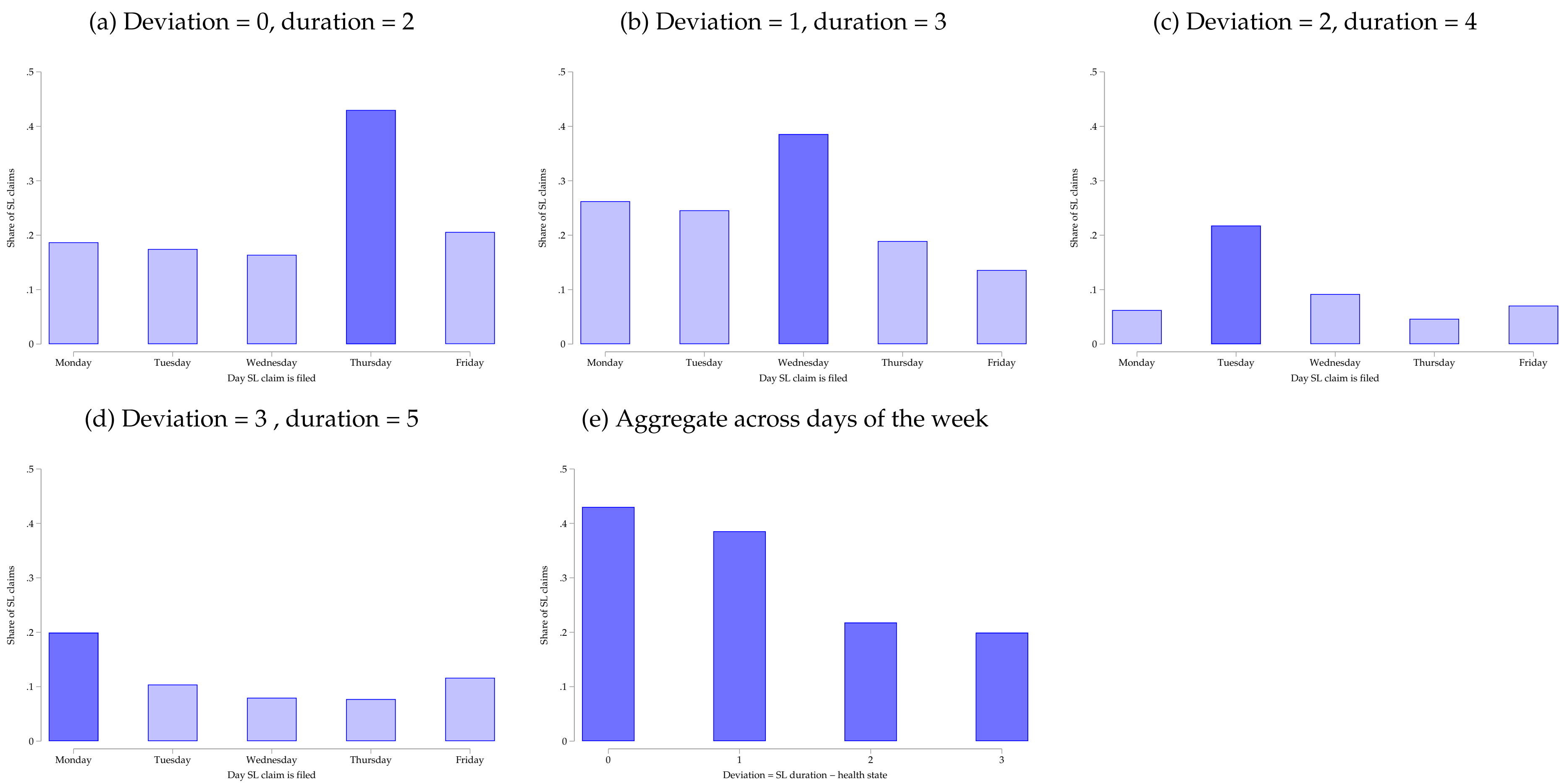
Notes: This figure shows the probability that a worker contracts disease d by by workers' age and occupation for the youngest and oldest group of workers. Diseases are ordered as presented in Table A7. Sample includes male private-sector employees. Blue-collar worker refers to workers who engage in hard 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.D.

Figure A6: Distribution of sick leave claims by duration and day of the week.



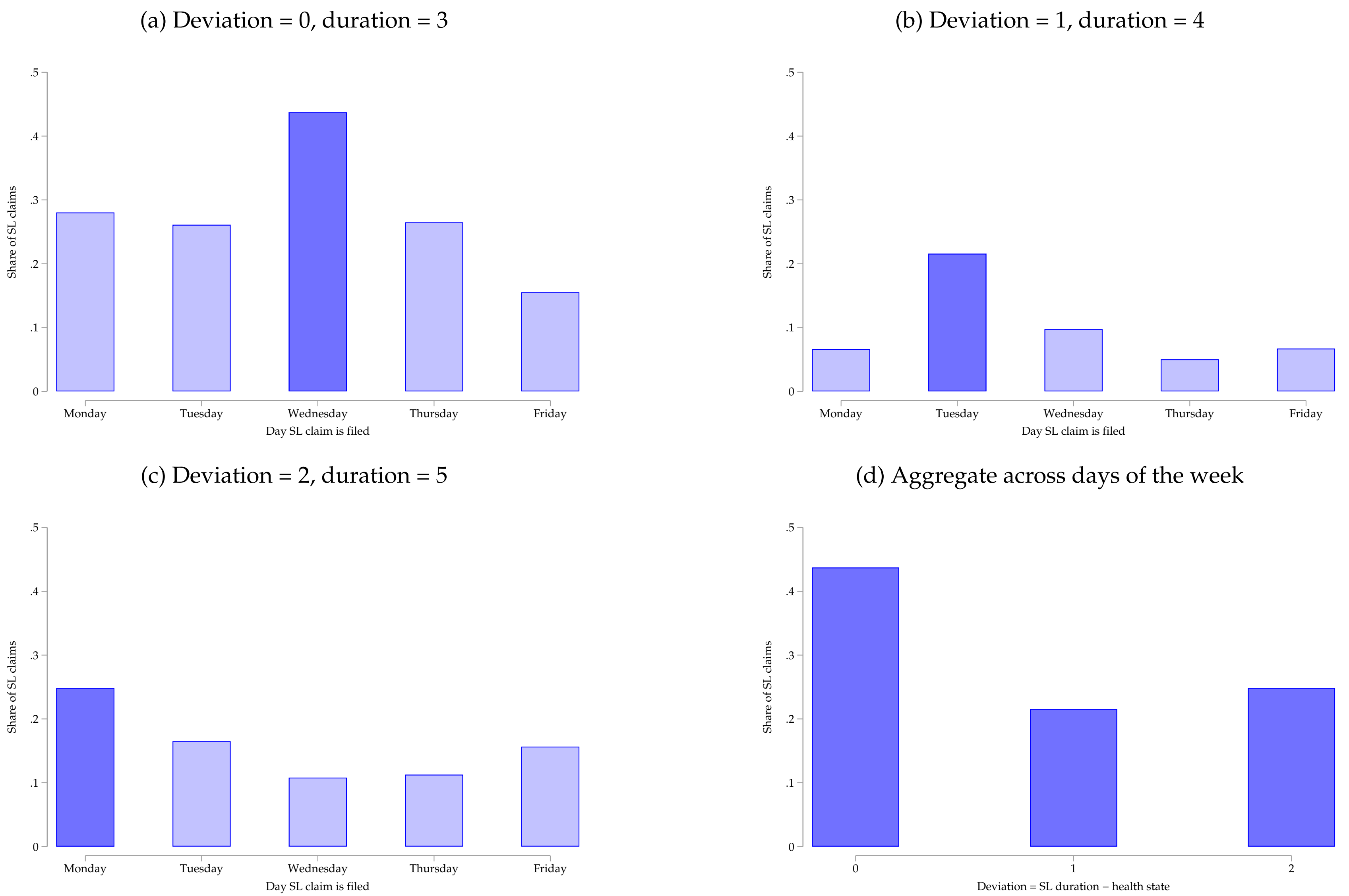
Notes: This figure shows the share of sick leave claims of duration s filed on each day of the week. Each panel aggregates sick leave claims with consecutive duration as stated in the title. This figure is referenced in Section IV.D.

Figure A7: Identification of compliance costs parameter: Sick leave claims by duration and day of the week.
Health shock (θ) equals 2-days-long.



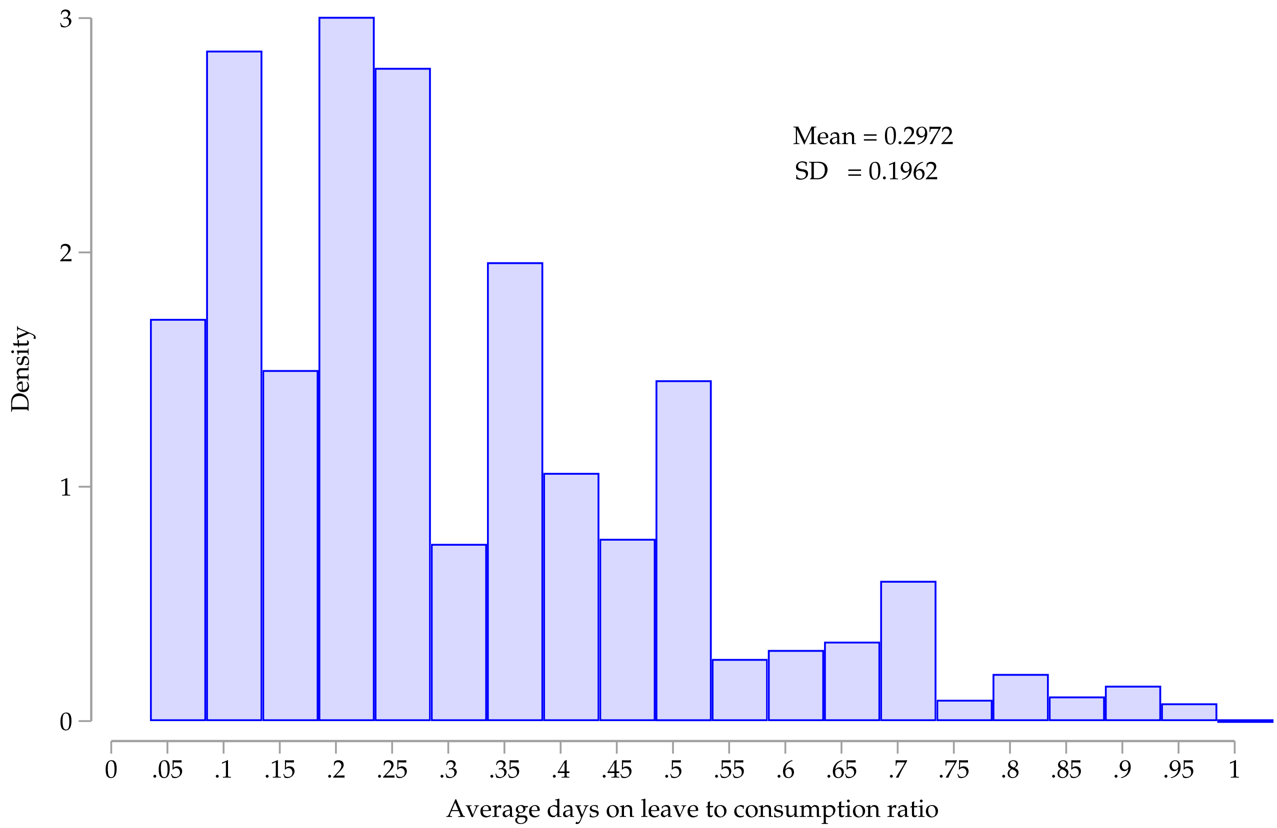
Notes: Panels (a) to (d) show the share of sick leave claims with duration s for workers whose main diagnose would implied a health state of 2 days on leave. Panel (e) aggregates the share of sick leave claims across days of the week, including only the weekend-streak combinations, e.g., from panel (a)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

Figure A8: Identification of compliance costs parameter: Sick leave claims by duration and day of the week. Health shock (θ) equals 3-days-long.



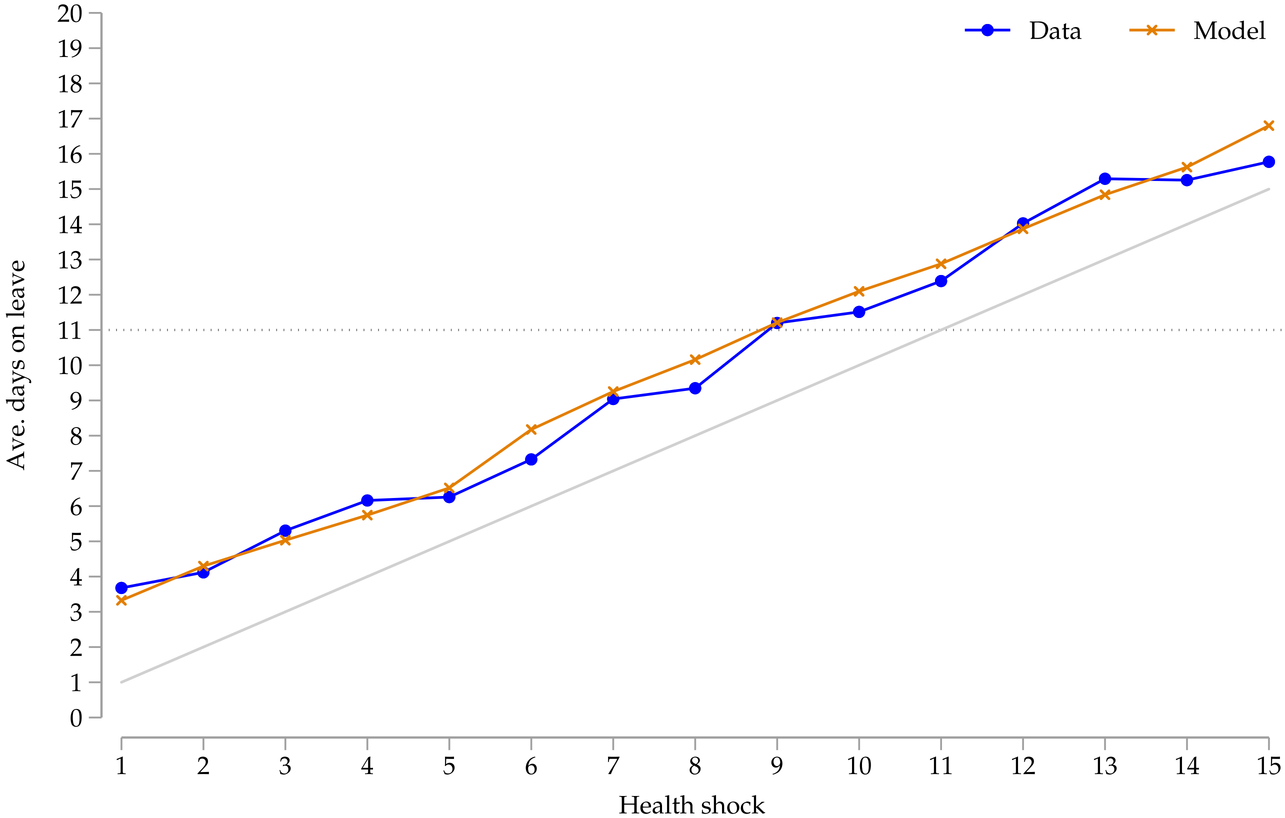
Notes: Panels (a) to (c) show the share of sick leave claims with duration s for workers whose main diagnose would implied a health state of 3 days on leave. Panel (d) aggregates the share of sick leave claims across days of the week, including only the weekend-streak combinations, e.g., from panel (a)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

Figure A9: Distribution of leisure to consumption ratio from raw data



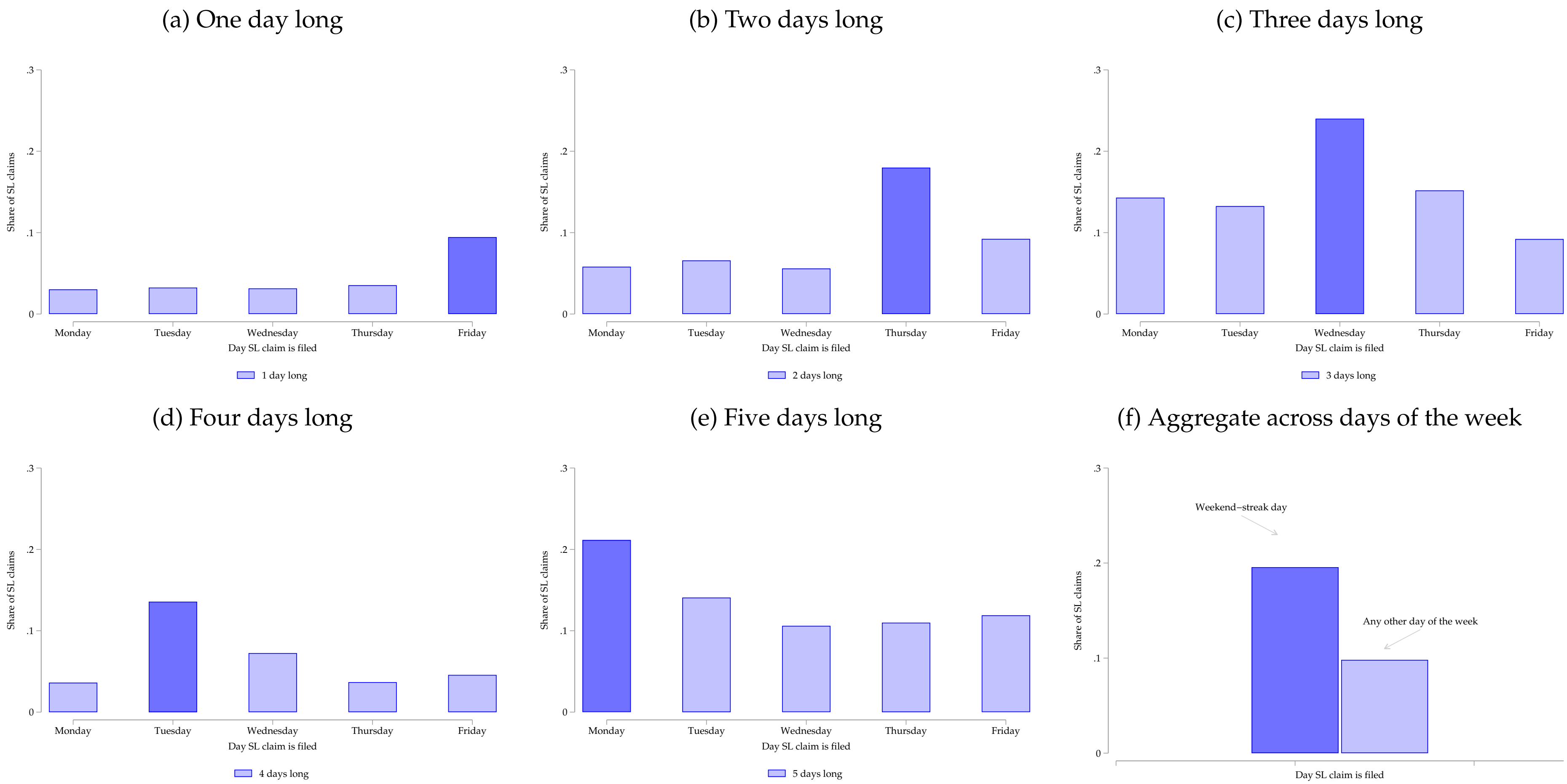
Notes: This figure shows the distribution of the leisure to consumption ratio LC_i. This figure is referenced in Section IV.D.

Figure A10: Demand for days on leave as a function of health shock



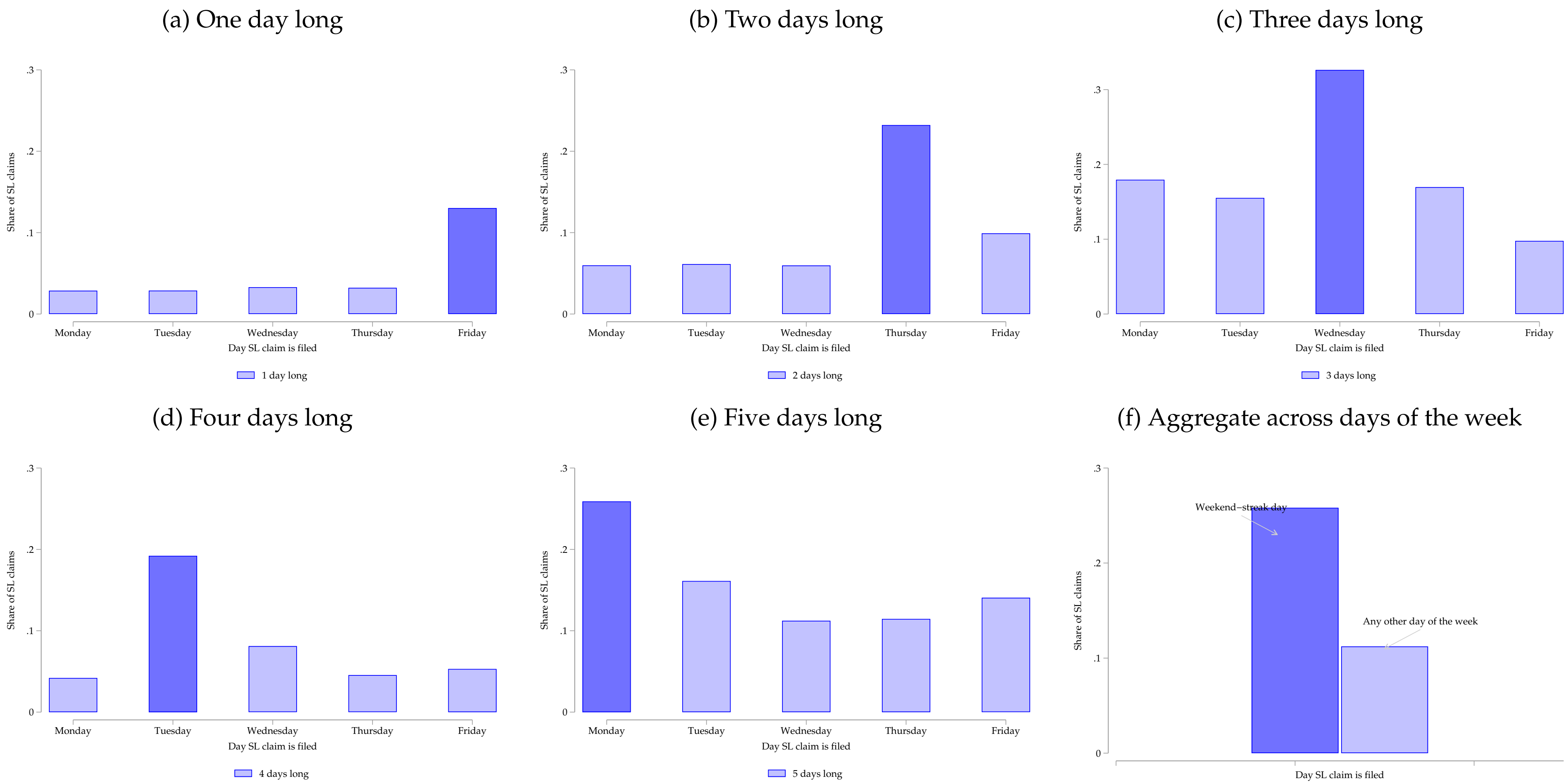
Notes: This figure shows the demand for days on leave as a function of the duration of the health shock from the data and a model-simulated sample. For each duration, I compute how many days; on average, workers request to be on leave. The 45 degrees line can be interpreted as the demand for days on leave when workers report their health. The horizontal line at 11 days indicates the position of the discontinuity in the sick pay scheme. This figure is referenced in Section V.C.

Figure A11: Days of the week and sick leave claim duration. Conditional to first quarter of the year (Summer quarter).



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. This figure is restricted to sick leave claims filed during the first quarter of the year (Summer quarter in Chile). This figure is referenced in Section **??**.

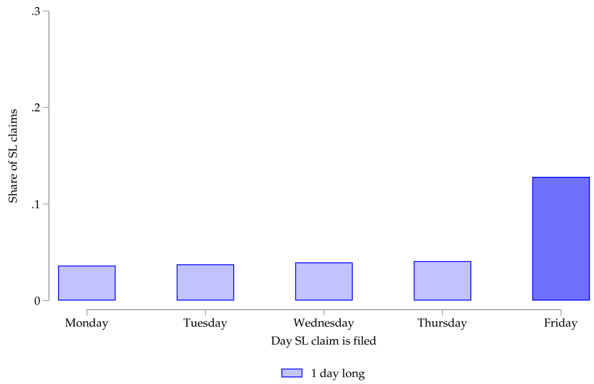
Figure A12: Days of the week and sick leave claim duration. Conditional to the second quarter of the year (Fall quarter).



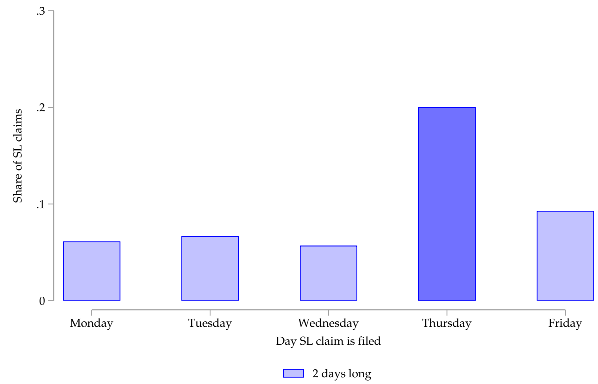
Notes: Panels (a) to (e) show the share of sick leave claims with duration s and the share of seven-days-long sick leave claims filed on each day of the week. Panel (f) aggregates across duration and days of the week: the first bar—labeled "weekend streak"—averages the share of one-to-five-days-long 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. This figure is restricted to sick leave claims filed during the second quarter of the year (Fall quarter in Chile). This figure is referenced in Section **??**.

Figure A13: Days of the week and sick leave claim duration. Conditional to the third quarter of the year (Winter quarter).

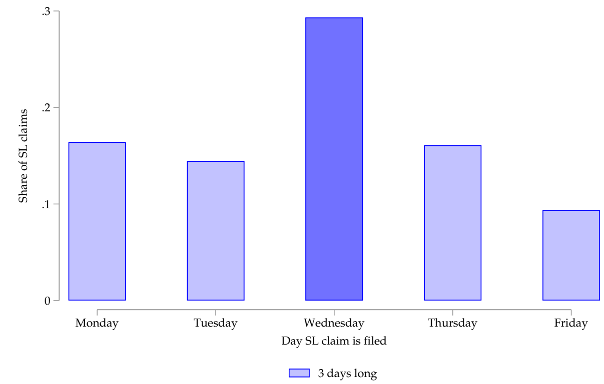
(a) One day long



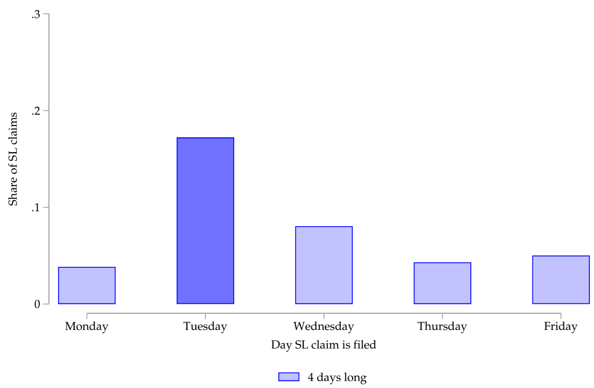
(b) Two days long



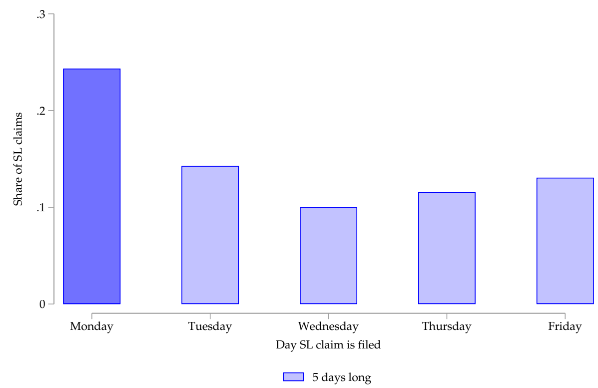
(c) Three days long



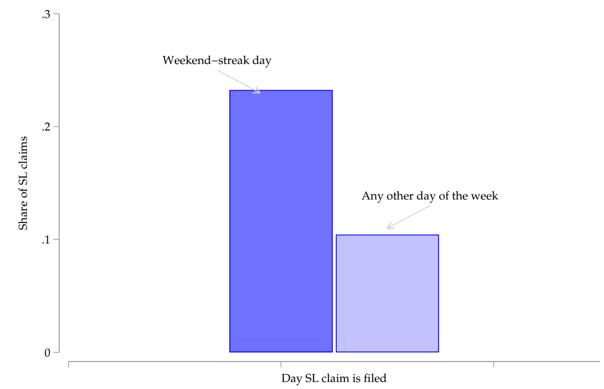
(d) Four days long



(e) Five days long

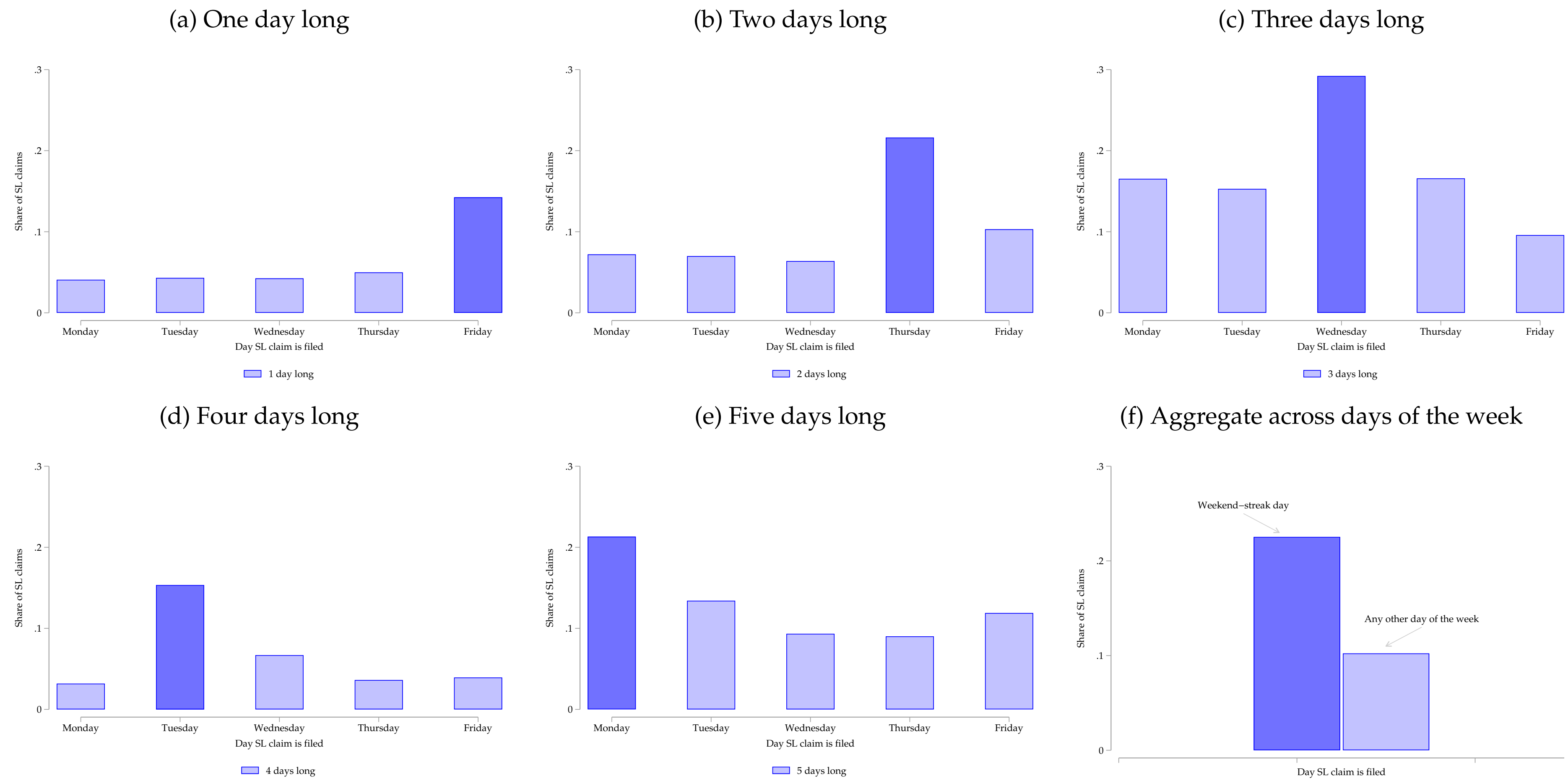


(f) Aggregate across days of the week



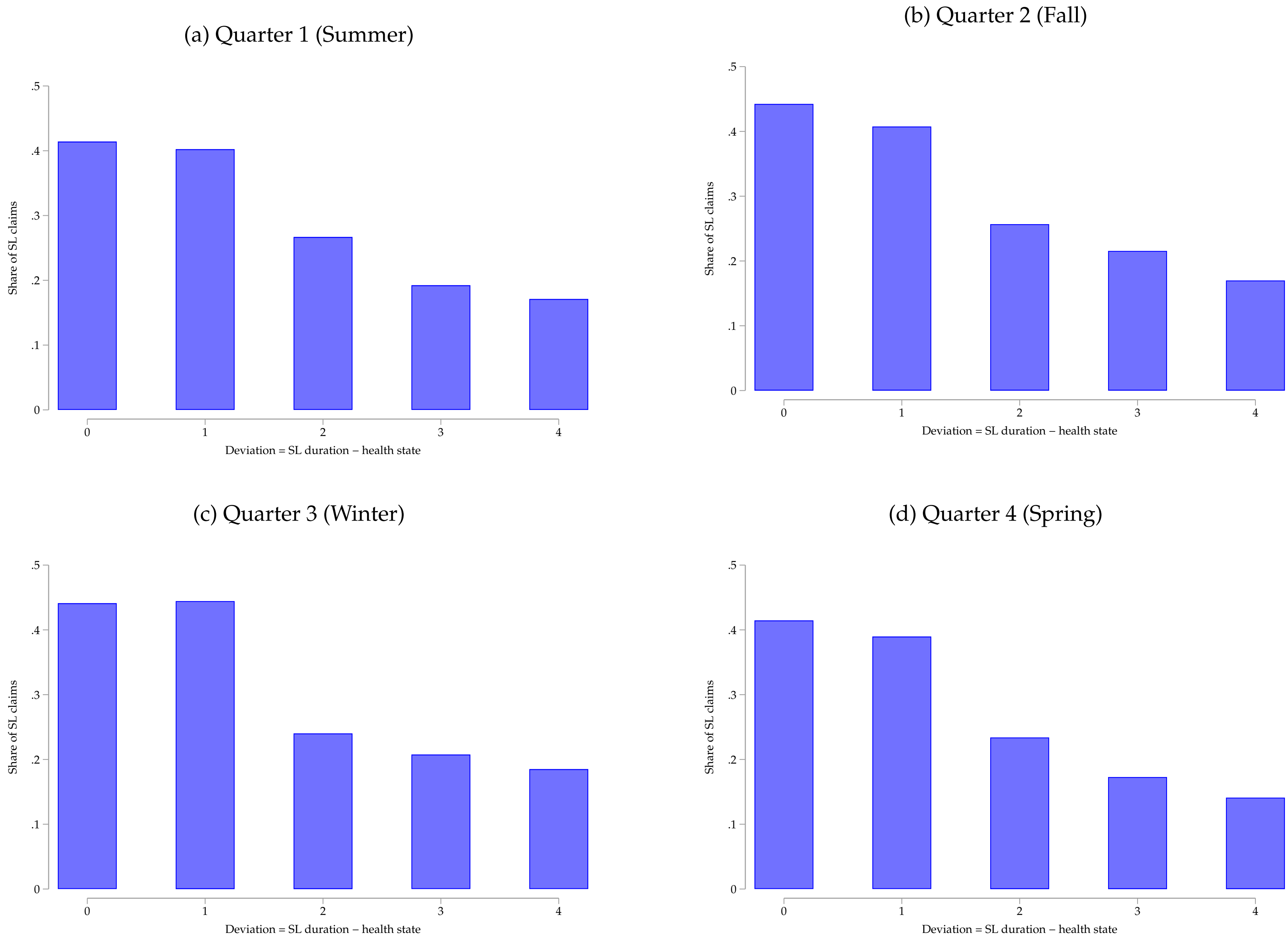
Notes: Panels (a) to (e) show the share of sick leave claims with duration s and the share of seven-days-long sick leave claims filed on each day of the week. Panel (f) aggregates across duration and days of the week: the first bar—labeled "weekend streak"—averages the share of one-to-five-days-long 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. This figure is restricted to sick leave claims filed during the third quarter of the year (Winter quarter in Chile). This figure is referenced in Section **??**.

Figure A14: Days of the week and sick leave claim duration. Conditional to the third quarter of the year (Spring quarter).



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. This figure is restricted to sick leave claims filed during the fourth quarter of the year (Spring quarter in Chile). This figure is referenced in Section **??**.

Figure A15: Compliance cost function by quarter.



Notes: Panels (a) to (d) show the average share of sick leave claims with deviations between 0 and 4 days for each quarter. Summer quarter goes from Jan to March. 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. This figure is referenced in Section **??**.

Table A1: Effects of Labor Labor Supply Responses on Days Worked

Time off (s)		Days work		Effect
B^0	B^1	B^0	B^1	
[0,θ)	[0,θ)	$d_{healthy} = M - \theta$	$d_{healthy} = M - \theta$	$d_{healthy}(B^1) - d_{healthy}(B^0) = 0$
		$d_{sick} \in (0,\theta]$	$d_{sick} \in (0,\theta]$	$d_{sick}(B^1) - d_{sick}(B^0) \leq 0$
[0,θ)	[θ,M]	$d_{healthy} = M - \theta$	$d_{healthy} \in (0, M - \theta]$	$d_{healthy}(B^1) - d_{healthy}(B^0) \leq 0$
		$d_{sick} \in (0,\theta]$	$d_{sick} = 0$	$d_{sick}(B^1) - d_{sick}(B^0) \leq 0$
θ	(θ,M]	$d_{healthy} = M - \theta$	$d_{healthy} \in (0, M - \theta]$	$d_{healthy}(B^1) - d_{healthy}(B^0) \leq 0$
		$d_{sick} = 0$	$d_{sick} = 0$	$d_{sick}(B^1) - d_{sick}(B^0) = 0$
(θ,M]	(θ,M]	$d_{healthy} \in (0, M - \theta]$	$d_{healthy} \in (0, M - \theta]$	$d_{healthy}(B^1) - d_{healthy}(B^0) \leq 0$
		$d_{sick} = 0$	$d_{sick} = 0$	$d_{sick}(B^1) - d_{sick}(B^0) = 0$

Notes: This table presents the direction of labor supply responses as a function of the demand for time off when worker i is shifted from contract B^0 to B^1. M represents the maximum number of workable days in a month. Thus, time off is bounded between no days off and M days off. The first column partitions the optimal time off choices in whether these are smaller, equal, or greater than the health shock θ. The second column presents the days work as a function of θ and M for each contract. The last column summarizes the impact on production. This table is referenced in Section II.B.

Table A2: Summary Statistics by Healthcare Insurance Provider

	Government-run insurance (1)	Private insurance (2)	Year(s) (3)
Panel A. Enrollees Characteristics			
Share of enrollees aged			
25 - 34	25.06	31.76	2015-2019*
35 - 44	21.51	28.84	2015-2019*
45 - 54	21.11	20.12	2015-2019*
55 - 64	14.64	11.47	2015-2019*
Share female enrollees	0.44	0.35	2015-2019*
Wages (in USD monthly)			
Average	761.27	1,824.94	2015-2019*
Enrollees w/ wage above median (%)	34.44	86.69	2015-2019*
Metropolitan region (%)	38.04	60.01	2015-2019*
Mining sector (%)	0.50	2.49	2015-2019*
N of enrollees	4,503,474	1,689,240	2015-2019*
Share (%)	72.72	27.28	2015-2019*
Panel B. Sick Leave Claims			
Ratio SL claims to enrollees (%)			
2015	77.66	86.53	2015
2019	98.42	90.66	2019
Approved SL claims (%)	91.94	74.54	2015-2019
Rejected SL claims (%)	5.31	14.76	2015-2019
Ratio days on leave to SL claim	13.09	10.24	2015-2019
Annual cost per enrollee (in USD)	240.69	463.61	2015-2019
Ratio of total annual cost to paid days on leave	24.91	58.90	2015-2019
Annual cost			
percentage of mandatory contribution	2.6	2.1	2015-2019
as percentage of GDP	0.51	0.37	2015-2019
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: Average Recovery Times - Examples From Peruvian Handbook

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

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

Table A4: Conditions Included in the Analysis by ICD-10 Group

ICD Group	Description	Included (=1 if yes)	Sick leave claims Number	Sick leave claims 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	<50	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	<50	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. There are two criteria for excluding conditions from the analysis. The first excluded group is conditions not listed in the Peruvian handbook. These conditions originate in the perinatal period (codes in groups P00-P96) and congenital malformations, deformations, and chromosomal abnormalities (codes in groups Q00-Q99). The second group of conditions corresponds to diagnosis for which a benchmark recovery time is hard to assess or full recovery might not be foreseeable. This group includes: poisonings and burns (codes in group T00-T98), "special purposes codes" (codes U00-U85), external causes of morbidity (codes V00-Y99), and factors influencing health status and contact with health services (codes Z00-Z99). Examples of the latter are leg amputations and organ 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. This table is referenced in Section III.C.

Table A5: Sick Leave Claims and Sick Leave Spells

	(1)	(2)	(3)
Number of sick leave claims	657,125	551,647	488,180
Number of sick leave spells	437,418	365,127	329,312
N of SL claims in a spell (% of claims)			
One claim	51.71	51.22	52.77
Two claims	17.19	17.30	17.06
Three claims	7.93	8.00	7.78
Four claims	5.06	5.12	4.94
Five claims	3.63	3.67	3.52
Six or more claims	14.48	14.70	13.92
Among sick leave spells with more than 1 claim* (% of claims)			
Diagnoses change within spell			
Yes — 4 digits disease code	30.16	30.13	30.39
Yes — 3 digits disease code	27.99	27.96	28.11
Physician change within spell	31.12	31.09	31.29
Sample			
Gender	Male	Male	Male
Age	18-70	25-64	25-64
Conditions	All	All	Included

Notes: This table presents counts and summary statistics of sick leave claims and sick leave spells for all male workers, for male workers aged 25 to 64, and for this group when diagnoses are conditioned to those included in the analysis (see Table A4). 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 present the composition of a spell in terms of the 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.C.

Table A6: Average recovery time 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	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 recovery time by workers' age and occupation type for 22 disease groups. Blue c. stands for blue collar and white c. stands for white collar. Table **??** indicates what occupations and industries are classified as blue and white collar. This table is referenced in Section III.C and Section IV.D.

Table A7: Probability of filing a SLC for each disease group by workers' characteristics

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

Notes: This table shows the probability of filling a sick leave claim for each disease (d) by workers' group (b). 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 A5. This table is referenced in Section III.D.

Table A8: Number of business days on leave (s_l)

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

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

Table A9: Identification of weekend-streak utility parameter (q): estimates from raw data

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

Notes: This table presents the distribution of sick leave claims by duration and day of the week. Weekend streak refers to the day of the week a sick leave claim should start to finish on Friday. For example, when duration is one day, weekend streak refers to Friday, when duration is two days, it refers to Thursday. The non-weekend streak category groups all the other days of the week. The share of sick leave claims of duration s filed on day dow is computed as the ratio between the number of claims with duration s filed on dow and the number of claims of filed on dow with duration between one and fifteen days. Figure 5 presents this table graphically. This table is referenced in Section IV.D.