JOHANNES KEPLER
UNIVERSITY LINZ

Physicians, Sick Leave Certificates, and Patients' Subsequent Employment Outcomes

by

Alexander AHAMMER

Working Paper No. 1605
January 2018

Johannes Kepler University of Linz
Department of Economics
Altenberger Strasse 69
A-4040 Linz - Auhof, Austria
www.econ.jku.at

alexander.ahammer@jku.at

Physicians, Sick Leave Certificates, and Patients' Subsequent Employment Outcomes[*]

Alexander Ahammer

Department of Economics, Johannes Kepler University Linz
Christian Doppler Laboratory Aging, Health, and the Labor Market

Accepted for publication at *Health Economics*

January 17, 2018

Abstract

I analyze how general practitioners (GPs) indirectly affect their patients' employment outcomes by deciding the length of sick leaves. I use an instrumental variables framework where spell durations are identified through supply-side certification measures. I find that a day of sick leave certified only because the worker's GP has a high propensity to certify sick leaves decreases the employment probability persistently by 0.45–0.69 percentage points (pps), but increases the risk of becoming unemployed by 0.28–0.44 pps. These effects are mostly driven by workers with low job tenure. Several robustness checks show that endogenous matching between patients and GPs does not impair identification. My results bear important implications for doctors: Whenever medically justifiable, certifying shorter sick leaves to protect the employment status of the patient may be beneficial.

JEL Classification: I10, J21, J60
Keywords: Sick leave duration, employment, general practitioners.

1 Introduction

Today, sick leaves are implemented in most industrialized countries as an institution to allow workers to recover from medical conditions and maintain at least part of their regular income while being absent from work. Instead of protecting employment, however, a higher sick leave take-up rate has been found to induce unemployment and wage reductions (Andersen, 2010; Hansen, 2000; Markussen, 2012; Scoppa and Vuri, 2014). In this paper I shed light on the role of general practitioners (GPs) in explaining these adverse effects.

Physicians differ substantially in their treatment behavior, both across and within geographic regions, even when health status of the patient is held fixed (e.g., Grytten and Sørensen, 2003). This is because every physician has a unique set of preferences and beliefs with regard to the necessity and efficiency of different treatments, with both medical and legal leeway for them to adjust their behavior accordingly. This has immediate implications for sick leaves as well (Aakvik *et al.*, 2010): While some GPs may generally

[*]A previous version of this paper was circulated under the title *"How Physicians Affect Patients' Employment Outcomes Through Deciding on Sick Leave Durations."* I thank the editor Pilar García-Gómez and two anonymous referees; René Böheim, Peter Egger, Martin Halla, Michael Lechner, Gerald Pruckner, Nicole Schneeweis, Rudolf Winter-Ebmer, and Ivan Zilic; the seminar participants in Innsbruck, Linz, at the 2016 Labor Seminar in St. Anton, and the WUWAETRIX-IV in Vienna, and the conference participants at the 2016 EEA/ESEM in Geneva, the 2016 EuHEA in Hamburg, the 2016 ESPE in Berlin, and the 2016 NOeG-SEA in Bratislava for numerous fruitful discussions and valuable comments. Furthermore, I am indebted to Tom Schober for providing part of the data. The usual disclaimer applies.

prefer to certify four days of sick leave for a common flu, others may certify only three. The question whether workers' employment outcomes differ by consulting GPs in one or the other group is precisely the focus of this paper.

In order to tackle this question empirically, I use the prescription behavior of GPs in Upper Austria as an instrumental variable (IV) for the duration of sick leave spells they certify. This yields a local average treatment effect (LATE) capturing the effect of a *marginal* day of sick leave—namely, a day of sick leave certified only because the worker's GP prefers to certify longer sick leaves—on subsequent employment. Returning to the previous example, this essentially means that I compare two identical workers who are equally sick but consult different doctors. While one doctor grants three days of sick leave, the other grants four days; in this case the LATE captures exactly the difference in employment outcomes between the two workers. An important distinction between my framework and the rich literature on *absenteeism* (i.e., absence from work despite being healthy) is that in my framework, the decision whether to stay home for an additional day is taken *not* by the worker, but by the doctor who certifies the sick leave.

Apart from highlighting the GP's role, I contribute to the literature in several other ways. This is the first study to use the duration of individual sick leave spells as the main explanatory variable instead of aggregate sickness absence measures. The former is indeed the more obvious and immediate decision variable for GPs in day-to-day medical care and entails more practicable policy recommendations. I also suggest a novel robustness check that utilizes random patient–GP matches; this check shows that my analysis is not impaired by endogenous sorting of patients to GPs. Finally, most of the existing evidence linking sick leaves to labor market outcomes stems from Scandinavian countries. Although Austria has a similar social security system and economic structure in general, it is important to consider other countries as well so as to gain a more comprehensive picture.

Using social security data and health records from Upper Austria, I find that a marginal day of sick leave decreases employment probabilities persistently during the first 18 months after a sick leave spell by 0.45–0.69 percentage points (pps). The risk of unemployment increases by at least 0.28 pps, but this effect approaches zero comparably quicker over time. Stratifying the population into different subsamples, I find that these effects are largely driven by workers with low job tenure; the effects do not appear to differ by the workers' gender or citizenship.

My findings hold important implications for policy makers, and, more importantly, for doctors. In line with the existing literature, I show that each additional day of sick leave is detrimental to the patients' employment outcomes. Since a large part of this negative effect can be explained by the high propensity of doctors to certify longer sick leaves, in case of doubt, doctors should certify shorter sick leaves whenever possible to protect the employment status of their patients.

1.1 Review of the Literature

The association between sick leave take-up and labor market outcomes has gained increasing attention from both labor and health economists in recent years, originating mainly in Scandinavia. Using Norwegian administrative data, Markussen (2012), for instance, finds that a one pp increase in a worker's sick leave take-up rate is associated with a 0.5 pp reduction in employment probability and a 1.2% reduction in earnings two years later. As in this paper, the identification of the sick leave take-up rate is based on propensity-to-prescribe measures estimated using a competing risks survival model.

Studies focusing on wage outcomes include Hansen (2000) for Sweden and Andersen (2010) for Denmark. Swedish workers are covered by a national health insurance that reimburses their earnings loss due to sick leave. Hansen (2000) investigates the effects of a 1991 reform that led to a substantial reduction in replacement rate. He finds negative wage effects due to an increase in sick leave take-up rate, but only for women. Andersen (2010) exploits a similar policy reform in Denmark that changed the sick leave reimbursement scheme, placing an additional financial burden on municipalities that should provide incentives to speed up the case work for workers on sick leave. Andersen finds that a one-month increase in aggregated sick leaves reduces wages up to two years later by 4.4%–5.5%, which is a rather small, but statistically significant effect.

To my knowledge, only one study has taken the length of individual sick leave spells explicitly into account: Hesselius (2007) splits sick leaves into short (1–7 days), medium (8–28 days), and long (more than 28 days) spells, and analyzes how the total number of leaves taken in each of these categories affects the risk of unemployment. Using proportional hazards models, he finds a positive relationship between sick leave take-up and unemployment, with effects being strongest the longer spells are. My paper is different from Hesselius (2007) in several ways. Most importantly, Hesselius considers an average treatment effect where variations in sick leave may be caused by both patients *and* GPs (and potentially also other factors). In my framework, only the GP-side variation is used to determine the effects on employment outcomes. Furthermore, Hesselius does not consider the potential endogeneity of sick leaves. Although controlling for a rich set of covariates in his estimates, unobserved heterogeneity may still induce bias (for example, when more motivated patients have fewer days of sick leave and better employment outcomes). Moreover, the main explanatory variable in Hesselius (2007) is the aggregate number of sick days, whereas I focus on the duration of individual spells.

Similar effects on unemployment have been reported in other studies as well. Using Italian data, Scoppa and Vuri (2014) find strong positive effects of absences on the risk of unemployment, controlling for a variety of confounding factors. Similarly, Amilon and Wallette (2009) show that sick leaves increase both temporary employment and unemployment probability of men in Sweden.

1.2 Institutional Background

Austria has a Bismarckian welfare system with almost universal health care access. Social pension, health, and work accident insurance are covered by a total of 22 social insurance institutions organized through the umbrella organization *"Main Association of Austrian Social Security Organisations."* Once employed, workers are automatically insured with one of these 22 institutions depending on their industry affiliation, place of residence, and whether they are employed in the private or public sector. In this paper, I focus on employees insured with the *Upper Austrian Sickness Fund* (UASF), which covers around one million members representing roughly 75% of the population in Upper Austria, one of the nine Austrian provinces. Except for workers in the railway and mining industries, all employed individuals in Upper Austria are insured with the UASF. Farmers, self-employed persons, and civil servants are insured with other institutions.

Sick leave insurance in Austria is designed to compensate workers for earnings loss due to both occupational and non-occupational diseases. Depending on their job tenure, employees receive full salary during the first six weeks (for workers with less than five years of tenure) to twelve weeks (for workers with more than 26 years of tenure). After this period of full reimbursement, workers receive half their salary for another six to twelve weeks (again, depending on job tenure), and then one quarter of the full salary for another four weeks (Federal Ministry of Social Affairs, 2014).

Sick employees are obliged to inform their employer as soon as they become incapacitated for work.

In most cases, sickness certificates are issued by GPs, who act as gatekeepers to the Austrian health care system. Hospitals or specialists certify sick leaves only in rare circumstances. The certificate itself contains mainly the starting date of the sick leave and its expected duration. The latter is binding only in one way, meaning that the actual absence must not exceed the recommended duration, but may fall short of it in case the employee decides to return to work earlier. In such cases, the firm has to notify the insurance fund immediately. If a worker is still sick at the end of a spell, she has to consult her GP again, who can then prolong the leave. Sickness certificates do not reveal a specific diagnosis, because the law does not grant employers the right to learn about their employees' diagnoses.

One particularity in the Austrian system is that workers are legally not obliged to produce certificates for absences of less than three days, unless the firm explicitly requires it. Firms are free to enforce such a rule if the newly hired employees are informed about the policy and their employment contracts include such a clause. I therefore expect measurement error in my estimations, because I do not observe very short sick leave spells for some firms in the data. As long as a firm's personnel planning is unrelated to whether it requires certificates for short sick leaves or not, the estimations should not be affected by this type of sample truncation. In any case, dropping sick leave spells that do not exceed two days does not affect the estimation results reported below at all.

In principal, employment contracts can be terminated at any point of time by either the employer or the employee giving a period of notice. Apart from this requirement, employers have no other formal procedures to follow when dismissing a worker, just as workers when terminating their contract. In case both parties reach a consensual agreement on the termination, the contract would end according to that agreement. When no agreement is reached, there is a cancellation period of usually six weeks or more (depending on the worker's seniority). Certain groups are protected against dismissal by law, most notably apprentices, pregnant mothers and parents on parental leave, members of the works council, disabled employees, and employees on military leave. Temporary work contracts cannot be terminated, because they end on expiry. Note importantly that workers are *not* protected from dismissal while on sick leave. In such cases, however, the employer has to pay wages until the expiry of leave.

2 Data

I combine data from the UASF, the *Austrian Social Security Database* (ASSD), and tax reports from the *Austrian Ministry of Finance*. The UASF database comprises individual-level information on health care service utilization in both the inpatient and outpatient sector for members of the sickness fund. I extract the sick leave durations, diagnoses, and certain health indicators from these data. Information on employment histories, wages, and certain demographics is taken from the ASSD, which is a longitudinal matched employer–employee dataset covering the universe of Austrian workers from the 1970s onward (Zweimüller *et al.*, 2009). Since wages are right-censored up to a tax cap, I augment the ASSD with income data from the *Austrian Ministry of Finance*. The data used for my empirical analysis cover all the sick leaves certified between 2005 and 2012 in Upper Austria by GPs who have a contract with the sickness fund and attend to at least 50 patients on average during this period.

I construct a panel data set where each observation is a single sick leave spell.[1] Each worker may have multiple non-intersecting sick leaves ordered by their ending date. Sick leave spells are recorded as the total time an employee stays away from work, not the actual time the GP certified (these may not

[1]More information on how spells are arranged in the data is provided in section A.1 of the web appendix.

necessarily coincide, for example, if the worker decides to return to work earlier).

Starting with 3,920,075 observations, I drop 445,807 apprenticeship spells, 230,481 spells whose subsequent spell is either retirement or a maternity leave spell (workers belonging to these three groups are protected against dismissal by law), and 19,904 spells of employees who are either younger than 18 years or older than 65 years. Another 37,983 observations whose sick leave duration is above the 99$^{\text{th}}$ percentile at 44 days are dropped as well.[2] Finally, I follow Correia (2015) and drop 97,898 singleton observations in order to ensure proper inference and improve computational efficiency in the fixed effect regressions outlined below. Finally, I am left with a total of $N^* = 3{,}125{,}759$ sick leave spells granted to $N = 423{,}352$ workers in $J = 43{,}297$ firms. Each worker has on average 7.19 distinct sick leave spells during the observation period of eight years.

As mentioned earlier, the employer has the discretion to require certificates for absences that do not exceed two days. Since approximately 20% of all sick leaves in the data fall into this category, it seems quite normal for firms to require certificates for very short leaves as well. In fact, out of 47,944 firms in the data, 29,756 (62%) have required a certificate for short leaves at least once during the observation period. These firms are on average older, bigger in terms of firm size, pay higher wages, and have a slightly lower intra-firm wage inequality (measured via the standard deviation in wages), compared to those that never required a certificate for short absences. Industry distributions for these two types of firms are highly similar, but firms that do not require certificates for short leaves are more often in the construction sector. Dropping sick leaves of less than three days does not affect the estimation results at all.

Detailed descriptive statistics are discussed in section A.2 of the web appendix. The average sick leave spell is around 6 days (the median is 5 days), while the average employment spell is 8 years. After a sick leave, the average *remaining* employment spell is 2.69 years. A small yet negative reduced-form relationship can be observed in the raw data: sick leaves certified by below-average propensity-to-certify doctors are followed by employment spells of around 0.059 years ($\approx$ 22 days) longer than those following sick leaves certified by more lenient doctors (this difference is statistically significant at the 5% level).

With a probability of 27%, the spell that follows sick leave is an unemployment spell rather than an employment spell at a new firm. As expected, sick leaves certified by physicians with an above-average propensity to certify are on average 0.597 days longer ($p < 0.05$). After two years, 51% of all workers still belong to the same firm they worked for initially, while 22% registered at the unemployment office and 27% transitioned to a different firm.

3 Methodology

My main outcome measure is the duration of the remaining employment spell e_{ik} following sick leave $k = 1, \ldots, K_i$ of individual $i = 1, \ldots, N$. Instead of modeling the duration of e_{ik} itself (e.g., by means of a hazard rate model), I use monthly employment transition probabilities over a period of two years after the sick leave spell as outcome variables to account for the time dimension of e_{ik}.[3] The first set of

[2] Note that none of these sample restrictions changes my results significantly. Regression results for samples extended with doctors who have less than 50 patients on average during the observation period, with employees who are younger than 18 or older than 65, or with sick leaves which last longer than 44 days are available upon request.

[3] Although e_{ik} is naturally a duration outcome, I refrain from using survival analysis in this paper, mainly because I rely heavily on the local average treatment effect interpretation obtained via *two-stage least squares* estimation of the treatment effects (Imbens and Angrist, 1994). Apart from that, I am unaware of estimators that deal with endogeneity in a survival analysis framework when the endogenous variable is continuous. A notable exception is Li *et al.* (2015), who essentially propose a control function approach where the second stage is specified as an additive hazards model. This is not practicable, however, because (1)

outcomes $\mathbf{y}_{ikm} = (y_{ik1}, y_{ik2}, \ldots, y_{ik24})$ is defined as a series of binary variables indicating whether worker i is still employed $m = 1, \ldots, 24$ months after sick leave k, or zero else. Although I do not observe whether workers are laid-off or terminate their contract themselves, I examine whether the subsequent spell is an unemployment spell or another employment spell at a different firm (i.e., a firm-to-firm transition). This allows me to use unemployment transition probabilities as a second set of outcome variables.

For every month m after sick leave k, the linear probability *two-stage least squares* (2SLS) model I estimate reads,

$$\begin{aligned} y_{ikm} &= \rho_m \hat{n}_{ik} + \mathbf{x}_{ik}' \mathbf{\Theta}_m + \omega_i + \varepsilon_{ikm}, \qquad m = 1, \ldots, 24 \\ n_{ik} &= \delta \Lambda_{d(ik)} + \mathbf{x}_{ik}' \mathbf{\Gamma} + \omega_i + \xi_{ik}, \end{aligned} \tag{1}$$

where y_{ikm} is the outcome variable of interest, n_{ik} is the length of sick leave spell k, $\Lambda_{d(ik)}$ is a binary IV indicating whether GP d who certifies observation i's sick leave k has an above-average certification propensity (see Section 3.1 for more details), $\mathbf{x}_{ik}$ is a vector of exogenous control variables, ω_i is a vector of worker fixed effects, and ε_{ikm} and ξ_{ik} are stochastic mean-zero error terms.

The model amounts to 24 separate second-stage regressions, where the coefficients $(\rho_m, \mathbf{\Theta}_m)$ are indexed by m, indicating that they are allowed to differ every month after the sick leave.[4] Inference throughout the paper is based on heteroskedasticity-robust and worker-level clustered standard errors.[5] This clustering is necessary to account for autocorrelation amongst the observations, because each worker may take multiple sick leaves during the sample period.

The vector of control variables $\mathbf{x}_{ik}$ comprises age squared, initial wage, tenure, experience, log firm size, and binary variables indicating whether the worker is a part-time and a blue collar worker, all measured at the beginning of the sick leave spell. As a proxy for health status, I use total drug expenses incurred two years prior to the leave in logarithmic form, along with total days spent in hospital in the previous two years. Finally, I use industry-specific unemployment rates as well as full sets of region and year dummies to capture macroeconomic fluctuations. Another reason why controlling for unemployment rates is important is that they act as worker discipline devices, thereby having a direct impact on absences as well (Scoppa and Vuri, 2014).

3.1 Estimating the Instrumental Variable

To obtain the certification propensity measure from the data, I decompose aggregated certified days of sick leave into time-varying observable patient characteristics and time-invariant patient and GP fixed effects. Consider the following two-way additive fixed effects model proposed by Abowd *et al.* (1999, AKM hereafter):[6]

$$\tilde{n}_{it} = \mathbf{x}_{it} \mathbf{\Pi}' + \theta_i + \psi_{d(it)} + u_{it}, \tag{2}$$

it requires assumptions on the underlying hazard function that are unrealistic given my empirical framework, and (2) incorporating a large set of fixed effects makes its computation infeasible.

[4]Equation (1) can also be understood as a system of seemingly unrelated regression (SUR) equations, with each regression being related to the others through their errors ε_{ikm}. Estimating each equation individually is consistent yet inefficient. Since I obtain very precise estimates anyways, the potential gain from using SUR-type estimation methods (which are computationally burdensome) is negligible.

[5]Bootstrapped standard errors which account for the variance of the (estimated) IV are similar to the analytical ones reported here and are available upon request.

[6]The idea of using the AKM model to estimate an IV from the data is based on Ahammer *et al.* (2017), who analyze the effect of labor income on mortality in Austria. In a similar vein, Markussen (2012) uses fixed effects obtained from competing risks survival models as IVs for sick leave take-up (see Section 1.1 for further details).

where subscripts $i = 1, \ldots, N$ again denote patients, $d = 1, \ldots, D$ denote GPs with $d(it)$ being the dominant GP of patient i in year $t = 1, \ldots, T_i$,[7] and $\tilde{n}_{it}$ are number of days of sick leave certified by doctor d for patient i in year t. Time-invariant effects are split into a patient-specific fixed effect θ_i and a GP fixed effect ψ_d. While θ_i is a time-invariant health stock unique to patient i, I interpret the GP fixed effect ψ_d as an inherent propensity to certify sick leaves.[8] Observable time-varying health characteristics, including a cubic in age, a binary variable equal to unity if i was pregnant in year t, number of days spent in hospitals where the referral was not initiated by a GP in $t-1$, and with a vector of region binary variables are captured within the vector $\mathbf{x}_{it}$.

Consistent estimation of the AKM model requires that all time-varying observables, the patient fixed effect, the GP fixed effect, and the error term u_{it} contribute additively to prescribed days of sick leave. This implies that the mobility between patients and GPs is exogenous, conditional on these factors. In particular, it implies that the motives for transition of patients to new GPs are orthogonal to the error term.[9]

In order to estimate equation (2), I build a panel for 2005–2012 comprising 1,294,460 patients and 857 GPs, with a total of 8,743,451 observations. This sample is larger than that used to estimate (1), because it contains individuals with no employment (for instance, pensioners, students, or unemployed people) and children as well. Additionally, patients having zero days of sick leave in a given year are included as well, as long as they are insured.

Using the estimated GP fixed effects $\hat{\psi}_d$, define the instrument for GP d as a binary variable equal to unity if $\hat{\psi}_d$ is above its sample mean, that is,

$$\Lambda_d \equiv \mathbf{1}\{\hat{\psi}_d > \bar{\hat{\psi}}_d\}, \tag{3}$$

where $\mathbf{1}\{\bullet\}$ denotes an indicator function and $\bar{\hat{\psi}}_d = D^{-1} \sum_{d=1}^{D} \hat{\psi}_d$ is the sample mean of the estimated GP fixed effects. Note that different specifications of the instrument, for instance, defining Λ_d to be equal to one if $\hat{\psi}_d$ is above its sample median or the 90$^{\text{th}}$ percentile of the GP fixed effect distribution, or simply using $\hat{\psi}_d$ as a continuous instrument, yield similar results.

3.2 Identification

In order to interpret the $\hat{\rho}_m$'s from the model in (1) as weighted averages of unit causal responses, two main assumptions are necessary (Angrist and Imbens, 1995). First, a first-stage is required. First-stage regression results are discussed in section A.3 of the web appendix: the null hypothesis that $\delta = 0$ can easily be rejected at $p < 0.01$. Second, I have to impose an *exclusion restriction* on the certification propensity IV. Here, the biggest threat to identification is endogenous matching between patients and GPs. If patients select GPs based on their propensity to certify sick leaves, and this mobility decision is correlated with unobserved characteristics affecting employment or wages as well, the exclusion restriction would be violated. I address this issue through various robustness checks in Section 4.2.

[7]The "dominant" GP is defined as the GP who billed the highest amount of fees to the health insurance for patient i in year t.

[8]Note that there are indeed two layers of certification behavior, namely one which is time-invariant (determined by, e.g., the GP's personality or medical education), and one which might change over time (e.g., certain preferences for treatments). Since I use only the time-invariant layer as the IV, the time-variant part would only interfere with identification if it is (1) causally related to the time-invariant part (which is of course plausible), but also (2) causally related to the employment status of the patient (which is rather unlikely).

[9]This assumption is weaker compared to the exclusion restriction necessary for the IV framework in equation (1) to be valid. For (2) to be identified, mobility between patients and GPs can also be conditioned on the GP fixed effect ψ_d, while this is obviously not the case for (1), where ψ_d (or, rather, its binary-coded counterpart Λ_d) is excluded from the second-stage regression. Thus, if the identifying assumptions discussed in section 3.2 hold, also the less general exogenous mobility assumption must hold.

Additionally, identification requires that the doctors' time-invariant propensities to certify sick leaves be independent of their patients' employment outcomes. Since these propensities could be seen as an inherent trait, something doctors are born with or develop during their studies, this assumption seems reasonable. On a related issue, it is crucial to properly control for the patients' health status to avoid omitted variable bias. Note also that principal-agent problems—in the sense that patients and GPs may negotiate about the length of sick leaves (see e.g., Nilsen *et al.*, 2011)—do not pose problems for my empirical analysis. The IV in my framework is orthogonal to any observed or unobserved patient characteristics, thus the LATE does not capture patient-side bargained days of sick leave.[10]

4 Results

In order to identify a causal effect of supply variation in sick leaves on employment, I estimate the model outlined in section 3. My main results are presented in Figure 1, where 2SLS estimates of the local average treatment effects $\hat{\rho}_m$ are plotted along with their 95% confidence intervals against time.[11] The left graph depicts the effect of sick leave duration on employment probabilities, while the right graph considers the estimated effects on unemployment probabilities. For comparison, *ordinary least squares* (OLS) point estimates are plotted as dashed lines.

I find that a marginal day of sick leave decreases employment probabilities persistently during the first 18 months, with dips at months 3 and 16. From month 18 onward, the effects converge quickly toward zero and become statistically insignificant. Conversely, the LATE on unemployment probabilities peaks in month 3 and then slowly converges to zero. After month 6, the effect remains non-significant at the 5% confidence level until the end of the observation period.

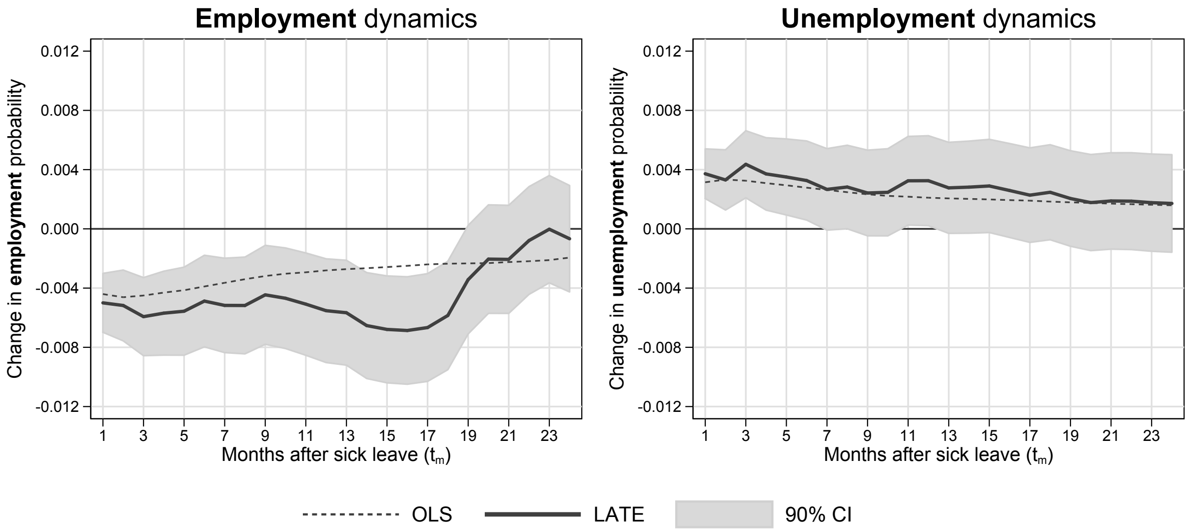
In terms of magnitudes, the LATE on employment probabilities varies between -0.0045 (week 9, p = 0.03) and -0.0069 (week 16, $p < 0.01$), whereas it ranges between 0.0028 (week 12, $p = 0.10$) and 0.0044 (week 3, $p < 0.01$) for unemployment probabilities. Thus, each marginal day of sick leave leads, *ceteris paribus*, to a decrease in employment probabilities between 0.45 and 0.69 pps, and to an increase in unemployment probabilities between 0.28 and 0.44 pps. Full regression results for month 3 after the sick leave are provided in the web appendix—including standard errors, covariate coefficients, and first-stage statistics.

Economically, the size of these estimates appears rather small at first glance. For example, a LATE of 0.59 pps in month three translates into a reduction in average employment probability from 86% to 85.41%. However, a comparison with other coefficients in the model (see Table A.4) shows that an annual wage increase of 3% would be necessary to compensate for the reduction in employment probability induced by the LATE. This is a rather high wage increase, given that it compensates only for one additional *GP-induced* day of sick leave. Similarly, the LATE corresponds to roughly 10% of the difference in employment probabilities between blue and white collar workers. These are sizable effects, given the specific nature of

[10] From the same argument it follows that my estimated effects do not just reflect different characteristics of the patient population (for example along demographic or socioeconomic lines), because these are controlled for and not captured by the LATE in any case.

[11] Because only between 40 and 60 percent of observations have predicted values between 0 and 1, I compare summary statistics of observations whose predictions lie inside the unit interval with observations whose predictions lie outside the interval in Table B.1 (web appendix). Although most averages are statistically significantly different from each other, their absolute differences are very small, which suggests no systematic differences between observation that are inside and outside the unit interval. In any case, linear probability models can approximate the local average treatment effect reasonably well under the exclusion restriction outlined in section 3.2 irrespective of the data generation process underlying the outcome variable (Angrist and Pischke, 2008).

Figure 1 — Effect of a marginal day of sick leave on employment and unemployment.



Notes: These figures plot the estimated local average treatment effects, $\hat{\rho}_m$, $m = 1, \ldots, 24$, of a marginal day of sick leave on employment probabilities (left graph) and unemployment probabilities (right graph) for the full sample (N = 3,125,759). Each coefficient for $t_1, \ldots, t_{24}$ is estimated from a separate regression based on the model in (1). LATEs are estimated by 2SLS, and OLS effects (where the length of sick leaves n_{ik} is treated as exogenous) are plotted as dashed lines for comparison.

the local treatment effect.

A complier analysis (see section A.5 of the web appendix for more details) shows that these effects are driven almost exclusively by workers with diseases of the respiratory system (especially acute colds and flus) and sick leave durations between three and nine days. Furthermore, compliers are more likely to be migrants, less likely to have at least an A-level degree, less likely to be part-time workers, and more likely to be blue collar workers. Furthermore, compliers are on average 35.7 years old, earn 26,715 euros, and have around 15.3 years of experience and roughly 5.3 years of tenure. Compliers appear to be largely located near the means of independent variables in the model; this is highly beneficial in terms of external validity.

These results compare well with those Markussen (2012) finds for Norway, but the employment effects seem to fade out quicker in my case. Aggregating days of sick leave on a yearly basis and running similar 2SLS models as those in model (1), I find that a one pp increase in number of days of sick leave decreases employment probability half-a-year later by 0.44 pps ($p < 0.05$).[12] One year later, however, the effect becomes statistically insignificant. Markussen estimates a pp decrease of 0.5 two years later.

4.1 Heterogeneous Effects

Next, I compare the above estimated effects between different subsamples of the population. The employment dynamics are provided in Figure B.1 and unemployment dynamics are provided in Figure B.2, both in the web appendix. Again, the solid lines show the evolution of the estimated LATEs up to 24 months after the sick leave spell, while the dashed lines provide the baseline estimates from Figure 1 for comparison.

[12] In order to obtain these estimates, I run 2SLS fixed effects models similar to the ones used for the main estimations (e.g., in Figure 1), but—instead of looking at individual sick leaves—the treatment variable is the sick leave rate in year t. The outcomes are employment probabilities measured at different points after the end of t. Continuous covariates are averaged over all entries in year t, and for categorical covariates the first entry in year t was taken for every worker i. Fixed effects are included on the worker-level and standard errors are robust to heteroskedasticity and clustered on the worker-level.

In Figure B.3, I test whether the differences between the subsamples are statistically significant.

First, I split the sample by gender. For both genders, estimates are close to the baseline. The coefficients are somewhat greater for men than for women, but their difference is not statistically significant on any conventional level. Second, I stratify the sample by tenure levels. One might suspect that workers with lower job tenure get punished harder for longer absence, because they had less time to reveal their inherent productivity or convince the employer of their trustfulness. On the other hand, firms may have a preference for younger workers, which could have the opposite effect. In fact, I find that the LATE is insignificant for workers with more than five years of tenure, and is positive and significant after month 18. One explanation could be that high-tenure workers do not get punished for longer sick leaves, but eventually a positive health effect could increase their employment probabilities. For workers with less than five years of tenure, estimates are similar to the baseline effects. In terms of unemployment probabilities, neither group differs significantly from the baseline.

Third, the estimated effects seem to be slightly stronger for migrants compared to Austrian citizens, but the differences in estimated coefficients are largely not statistically significant. For Austrians, employment dynamics behave very similar to the baseline, yet negative employment effects seem to be more persistent in their case (the differences between the two subgroups become significant also towards the end of the observation period). For migrants, both the employment and unemployment coefficients are not significantly different from the baseline.

Finally, one important question remains to be answered: Why do sick leaves seem to entail adverse employment effects in spite of being designed as an institution to protect workers? There are two plausible explanations for this: Workers are penalized by their employers for being absent from work, or sick leave itself causes the workers' health to deteriorate and lead to lower employability later. The first explanation implies that employers discriminate against workers with longer sick leaves because they interpret them as either signals of low work effort or motivation, or signs of decrease in permanent productivity, depending on whether the worker showed obvious indications of health issues before or after the sick leave.[13] Apart from the employer-side penalization, there may also be a pure health channel implying that the sickness absence *itself* drives this negative effect by preventing the worker from engaging in regular activity.

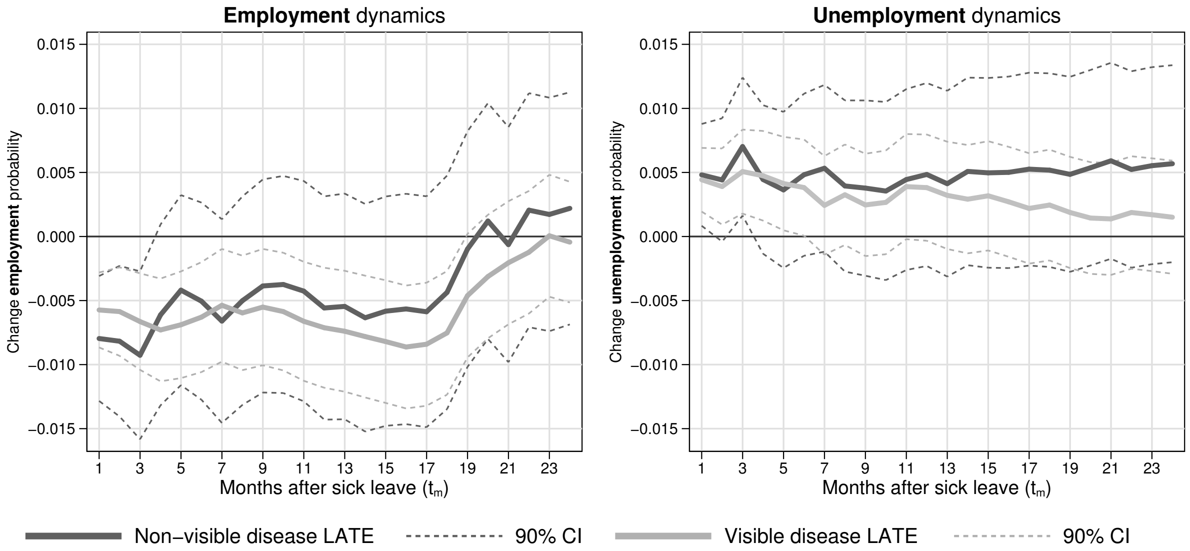
In order to gain an insight into the prevalence of these mechanisms, I divide sick leaves according to diagnoses that may be associated with shirking behavior and those that indicate "true" illness.[14] Employers do not learn about diagnoses directly, unless the illnesses are clearly visible to them; headaches, for example, often do not entail obvious visual signs of illness, while broken bones do. If an illness is visible, the employer may not associate the sick leave with shirking behavior. If negative employment effects are found for sick leaves that follow visible illnesses, it would indicate that employers statistically discriminate against sick workers because they expect their productivity to decline. Negative employment effects for non-visible diseases, on the other hand, would indicate that employees are penalized for shirking.

In Figure 2, I present the differential employment effects for the most common non-visible diseases; *acute nasopharyngitis* and *acute upper respiratory infections of multiple and unspecified sites* (the common cold, ICD-10 codes J04 and J06), *low back pain* (M54.5), and *headache* (R51), and compare them with all other (visible) diseases. Roughly 35% of sick leaves in my sample are non-visible according to this definition. Non-visible diseases lead to strong reductions in employment probability during the first four

[13] Note that employers probably do not observe whether the marginal increase in sick leave duration was caused by the patient or the doctor. They may hold the patient responsible, although it was in fact the doctor's decision.

[14] Thanks to one of the referees for suggesting this analysis.

FIGURE 2 — Difference in employment effects between sick leaves with diagnoses potentially associated with shirking behavior.



Notes: These figures plot the estimated local average treatment effects, $\hat{\rho}_m$, $m = 1, \ldots, 24$, of a marginal day of sick leave on both employment (left graph) and unemployment (right graph) probabilities for a sample of sick leaves diagnosed with one of the following ICD-10 codes: J04, J06, M54.5, and R51 (termed as *"non-visible,"* N = 1,096,598); for the rest of the sample (*"visible,"* N = 2,029,161). Each coefficient for $t_1, \ldots, t_{24}$ is estimated from a separate regression based on the model in (1).

months after sick leave, but become insignificant from month 4 onward. Visible diseases involve more persistent long-term consequences, as their effect on employment is significant until month 18 after sick leave. In terms of unemployment risk, those diagnosed with non-visible diseases generally seem to fare worse. Hence, it seems that there is a short period immediately after sick leave where the shirking effect dominates. However, in the long run the statistical discrimination channel is relatively stronger. Note that sick leave may indeed entail negative health effects itself, which could translate into a reduction in employment probability as well.

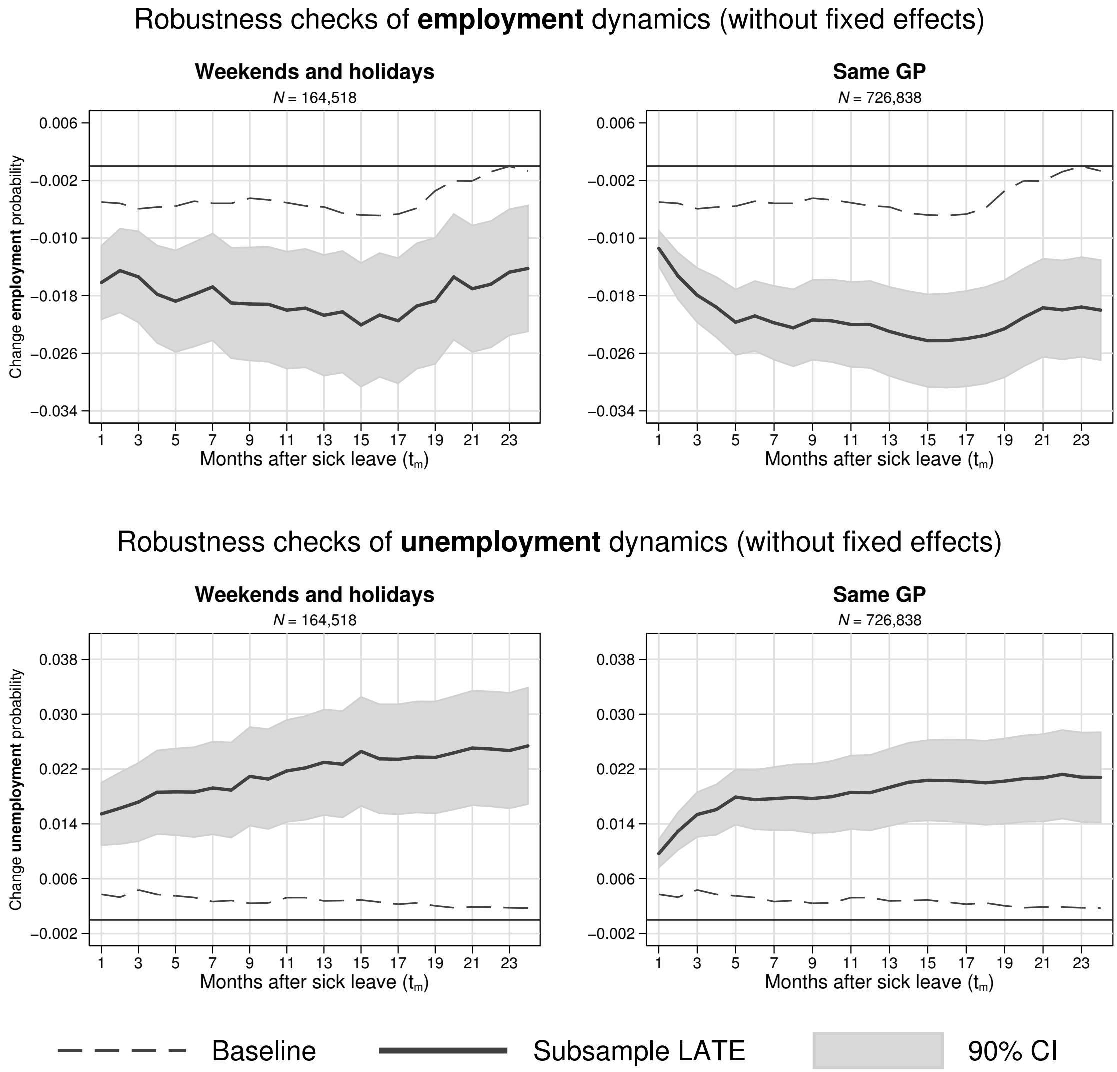
4.2 Robustness

The main threat to identification is endogenous matching between patients and doctors. In this section, I therefore analyze different subsamples of the population where either mobility is restricted or the motives of transitions can be assumed to be caused by factors other than prescription behavior of the new GP. Whenever results hold, it is likely that the effect of sorting is negligible. Another important requirement for identification is that the patients' health status is adequately controlled for. Thus, I provide two further robustness checks: A set of regression where I perform my main estimations on a specific subsample which can be considered as homogeneous with regard to health status, and another set of regressions where I control also for the type of diagnosis the respective sick leave is based on.

First, I restrict the sample to sick leaves starting either on weekends or public holidays, when doctors typically close their practice. In order to maintain the provision of basic health care on such days, each district in Upper Austria has a schedule of rotating GPs who provide out-of-hours services. Thus, the assignment of patients to GPs is more or less random on weekends and holidays, because it depends solely on the rotation schedule.[15] Although the purpose of such services is to offer assistance in medical *emer-*

[15]Note, however, that there are some problems associated with this assignment mechanism: First, the resulting sample might be selected, as patients will typically wait until their family doctor's practice is open again, unless they suffer from an (what

Figure 3 — Robustness checks for subsamples where fixed effects estimation is infeasible.



Notes: These figures plot the estimated local average treatment effects $\hat{\varsigma}_m$, $m = 1, \ldots, 24$, obtained from model (4), which does not incorporate fixed effects but controls for gender, education, and migratory status instead. The outcomes in the upper two graphs are the employment probabilities estimated from separate regressions for each month $t_1, \ldots, t_{24}$, while the outcomes in the lower two graphs are unemployment probabilities for each month $t_1, \ldots, t_{24}$. *Weekends and holidays:* the sample is restricted to sick leaves certified on weekends or public holidays, *same GP:* the sample is restricted to workers who never changed GPs during the observation period. The dashed lines show the baseline results from Figure 1.

gencies, patients may avail them irrespective of their actual condition. In fact, people hardly consult GPs on weekends or holidays for serious (potentially life-threatening) conditions. The first six most common diagnoses certified on weekends or holidays are identical to those for the full sample.

One drawback of this robustness check is that it disallows the use of fixed effects for estimations. Since

they perceive as) acute condition which requires immediate treatment. Second, ambulances are open on weekends and holidays as well, most importantly to treat serious life-threatening conditions such as strokes or heart attacks. However, in areas where hospitals are reachable in a few minutes patients may also prefer to go to the ambulance for less serious conditions, rather than consulting an emergency GP. Supposedly, workers living in rural areas will therefore be overrepresented in this subsample. Third, I do not observe the actual day of consultation. Although law prohibits sick leaves being certified retroactively, it is possible that consultations preceding spells which start on weekends or holidays in fact took place during the week. However, this applies only to employees who work on weekends but not during the week, which is indeed a rather unusual type of working contract. Hence, the bias induced by such observations should be rather small. Finally, as noted below, I cannot use worker-level fixed effects in this specification, because only few patients consult a doctor twice or more on weekends or holidays.

workers hardly consult GPs more than once on weekends or holidays during the observation period, the sample used for this robustness check does not provide enough within-worker variation in sick leaves for fixed effects estimations. Formally, the main regression model in (1) without fixed effects translates into,

$$\begin{aligned} y_{ikm} &= \varsigma_m \hat{n}_{ik} + \mathbf{x}'_{ik}\mathbf{\Omega}_m + \mathbf{z}'_i\mathbf{\Xi}_m + \upsilon_{ikm}, \qquad m = 1, \ldots, 24 \\ n_{ik} &= \iota\Lambda_{d(ik)} + \mathbf{x}'_{ik}\mathbf{\Phi} + \mathbf{z}'_i\mathbf{\Psi} + \zeta_{ik}, \end{aligned} \tag{4}$$

where a vector $\mathbf{z}_i$ of time-invariant control variables consisting of a female dummy, a migrant dummy, and education in categorical form is incorporated in place of the fixed effect.

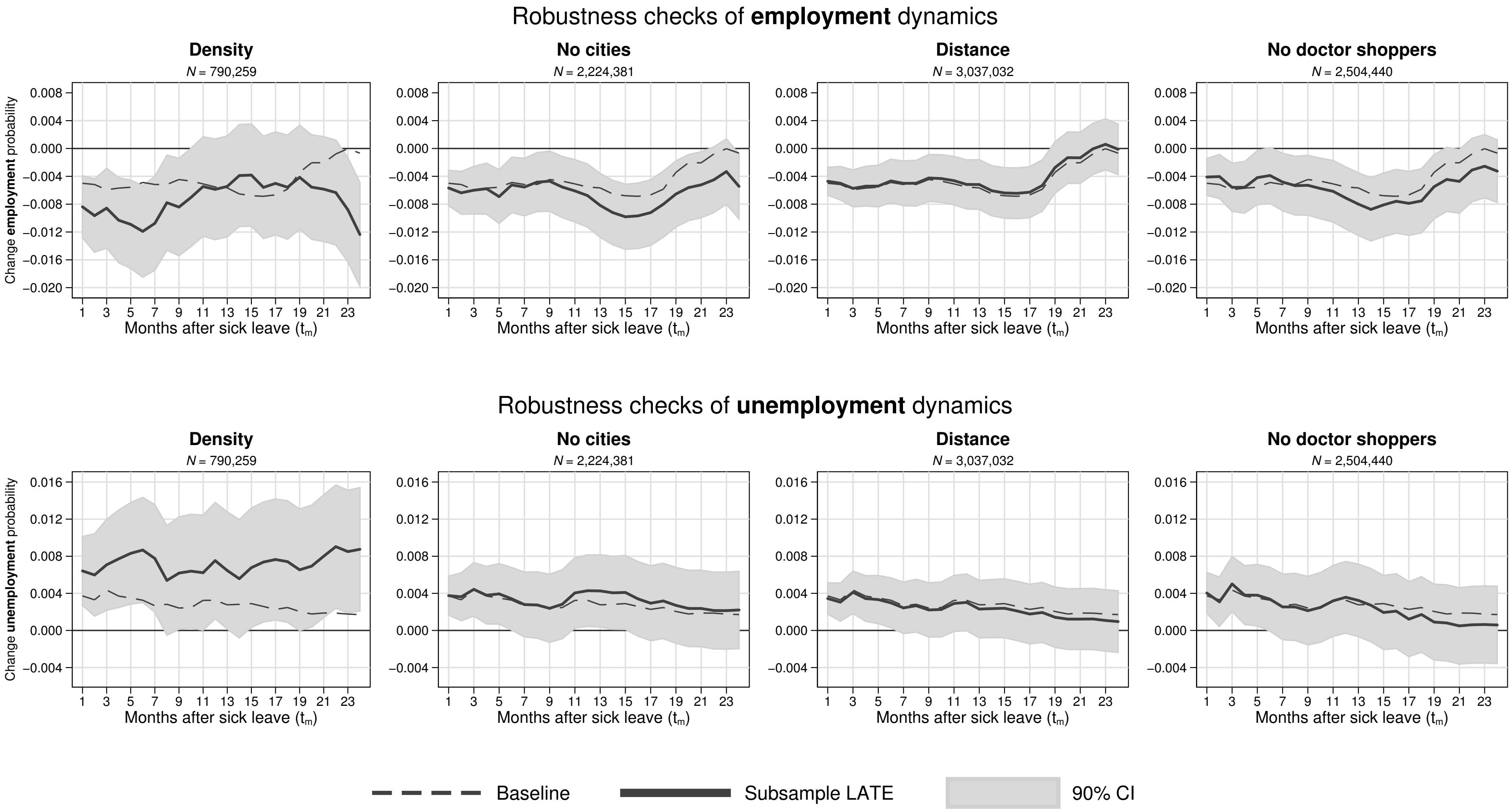
The distribution of the weekdays on which sick leave spells start and end is shown in Figure A.5 (web appendix). Most sick leaves start on Mondays and end on Fridays. A total of 159,856 spells (approximately 5.2% of the full sample) start either on a Saturday, a Sunday, or on a public holiday. The estimated employment and unemployment dynamics for this sample are illustrated in Figure 3. The main conclusions hold for this sample of randomly assigned patient–GP matches too. Coefficients have the expected sign and are roughly three times as high as those in the baseline model. Three months after a spell, the employment probability decreases by 1.86 pps ($p < 0.01$) through a marginal day of sick leave. The estimated coefficient for unemployment probability is even higher at 1.93 pps.

Next, I consider only workers who never changed their GP during the observation period. For these patients, endogenous matching is obviously a problem only if it took place before 2005. The sample is reduced to 707,624 observations, or roughly 23% of the original data. Again, because worker fixed effects coincide with the IV in this subsample, where patients stick to one GP over time, I estimate the model in (4) instead. The results are provided in Figure 3. Once again, each additional day of sick leave has a strong negative effect on employment and a positive effect on unemployment, with the coefficients being relatively large in magnitude. Both effects appear to persist well beyond the observational period of two years. This is in contrast to the evolution of baseline estimates approaching zero toward the end of the observed time horizon. Three months after the leave, the estimate of ς_3 suggests a 1.99 pps ($p < 0.01$) decrease in employment probability and a 1.72 pps ($p < 0.01$) increase in unemployment probability for each marginal day of sick leave.

An important restriction for matching is certainly the competition between doctors. In areas with high competition, patients can easily change doctors if they do not match their demands. In low-density areas, patients face only a small set of doctors they can choose from. As another robustness check, I restrict the sample to areas with a density of less than 0.63 doctors per 100,000 inhabitants at the community level (this roughly corresponds to the 25$^{\text{th}}$ percentile of the density distribution). The results are shown in Figure 3. In this subsample, the initial effect found during the first seven months is robust to both employment and unemployment probabilities.

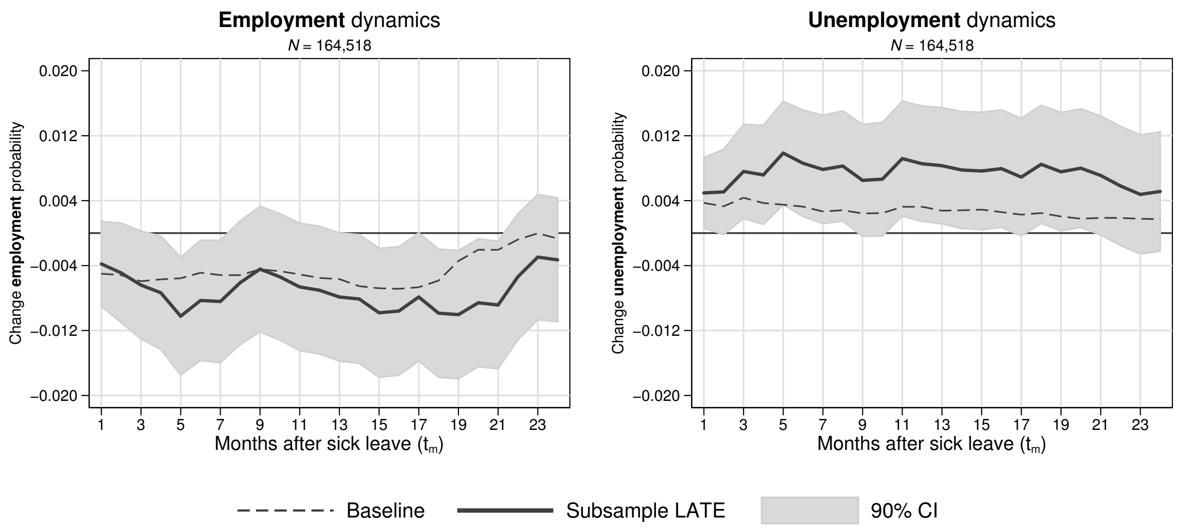
Next, I drop areas with more than 18,705 inhabitants (the population size of the smallest city in Austria in 2016) from the data. This approach is motivated by the notion that workers living in rural areas face a limited variety of different doctors and thus are restricted in mobility. Because roughly 28% of workers live in cities, the sample size remains relatively stable after dropping these. The results provided in Figure 4 are similar to the baseline specification. In a similar vein, I consider only the patient–GP pairs where the geographical distance between the two is low. I argue that if the distance is shorter than 10 kilometers, the patient likely selected his GP based on close proximity rather than the doctor's practice style. The evolution of effects over time is roughly the same as for the full sample (see Figure 4).

FIGURE 4 — Robustness checks for subsamples where fixed effects estimation is possible.



Notes: These figures plot the estimated local average treatment effects, $\hat{\rho}_m$, $m = 1, \ldots, 24$, of a marginal day of sick leave on both employment (upper four graphs) and unemployment (lower four graphs) probabilities for different subsamples of the population. *Density:* the sample is restricted to areas with a GP density of less than 0.63 doctors per 100,000 inhabitants at the community level; *no cities:* the sample is restricted to areas with fewer than 18,705 inhabitants; *distance:* the sample is restricted to patient-GP matches where the geographical distance is less than 10 kilometers; *no inside movers:* observations that change GP but do not change their living place are dropped; *same GP:* only observations that do not change their GP during the observational period are kept. Each coefficient for $t_1, \ldots, t_{24}$ is estimated from a separate regression based on the model in (1). The dashed line shows the baseline results from Figure 1.

FIGURE 5 — Robustness checks for subsample of healthy workers.



Notes: These figures plot the estimated local average treatment effects, $\hat{\rho}_m$, $m = 1, \ldots, 24$, of a marginal day of sick leave on both employment (left graph) and unemployment (right graph) probabilities for a subsample of the population with zero days of hospitalization and less than 330 euros of aggregate medical expenses two years prior to $t_{-n_{ik}}$. Each coefficient for $t_1, \ldots, t_{24}$ is estimated from a separate regression based on the model in (1). The dashed line shows the baseline results from Figure 1.

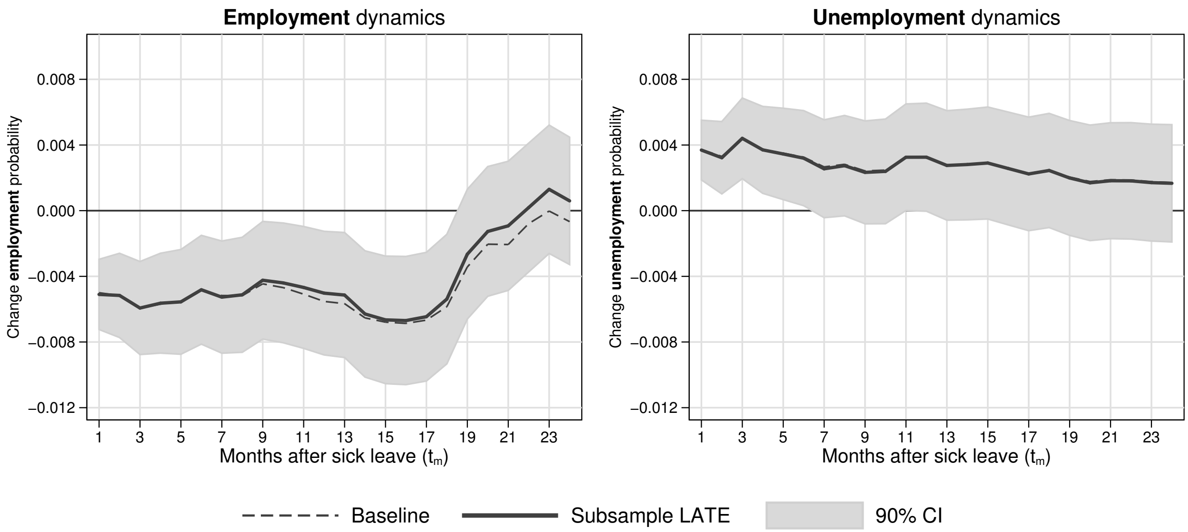
As another robustness check, I drop the patients who do not change their area of residence at the same time. This eliminates the "doctor shoppers" from the analysis; that is, patients who switch between doctors until they meet one who provides them with the treatment they demand. The results shown in Figure 4 indicate that the effects are larger in magnitude compared to the baseline and retain their statistical significance. Three months after the sick leave, the LATE for employment probability is estimated as –0.0075 ($p < 0.01$).

I conclude that sorting between patients and doctors is not a significant problem for my empirical analysis. This is perhaps not surprising. Although patients are free to select from the set of available GPs, 73.7% of Upper Austrians choose a GP from their zip code area (Hackl *et al.*, 2015). Thus, patients presumably tend to select the nearest GP in terms of geographic proximity, rather than one whose prescription behavior fits them best. This impression is confirmed by Ahammer and Schober (2017), who perform tests on the exogenous mobility assumption proposed in the empirical labor literature, to find no evidence of sorting on observables. A similar conclusion has been made by Markussen *et al.* (2011) for Norway.

Besides sorting, another important condition for my results to be valid is that the health status should be properly controlled for. Similar to Halla *et al.* (2016), I therefore restrict the sample to patients who have not been admitted to a hospital and who incur less than 330 euros of medical expenses on aggregate two years prior to the start of sick leave. This leaves me with a sample of 667,706 observations. Here, the estimates are rather imprecise because of the comparatively low sample size, resulting in non-significant coefficients whenever the effects are small (for instance, in months one or nine). However, because the estimates are almost uniformly higher in magnitude compared to the full sample, statistical non-significance should not be overemphasized.

Finally, I present model (1) with additional controls for the type of diagnosis the GP declared to the insurance when issuing the sick leave certificate. The results of this exercise are plotted in Figure 6. Controlling for type of diagnosis does not change the results at all. The point estimates even become slightly larger for employment probabilities, while for unemployment probabilities the difference compared to the

FIGURE 6 — Robustness checks where the type of diagnosis is controlled for.



Notes: These figures plot the estimated local average treatment effects, $\hat{\rho}_m$, $m = 1, \ldots, 24$, of a marginal day of sick leave on both employment (left graph) and unemployment (right graph) probabilities for the full sample (N = 3,125,759). Each coefficient for $t_1, \ldots, t_{24}$ is estimated from a separate regression based on the model in (1), but includes also a set of diagnosis dummies (150 categories based on the letter as well as the first digit of the ICD-10 code). The dashed line shows the baseline results from Figure 1.

baseline is so small that the two lines visually coincide in the graph. Both analyses provide strong support for the conjecture that my results are not driven by differences in initial health status of workers. Moreover, because most robustness checks discussed here yield effects larger in magnitude than the baseline, the initial estimate of the LATE is presumably a lower bound of the actual effect.

5 Conclusion

I quantify the impact physicians have on patients' employment prospects by influencing their sick leave duration. I isolate this channel by establishing a LATE framework that uses the supply-side variation in sick leave certifications to instrument for actual sick leave durations. The effect on employment probabilities is identified solely through workers whose sick leave duration is extended owing to consulting a GP with an above-average propensity to certify sick leaves. Thus, I estimate the effect of a *marginal* day of sick leave, namely, one that is granted only because the doctor has a preference for certifying longer leaves. I find that this marginal day of sick leave has a persistent negative effect on employment and a positive effect on unemployment probabilities, especially for workers with low job tenure. Crucial for the identification is that sorting between patients and GPs be conditionally exogenous. I devote a substantial part of the study to sensitivity analyses, which suggest that my results hold also in subsamples where patient-GP matches are exogenous.

Three potential mechanisms underlie these effects: Workers are penalized by employers because their longer sick leaves are interpreted as signals of shirking, employers statistically discriminate against sick employees because they expect persistent negative effects on productivity caused by the illness, or the sick leave itself entails negative health effects (e.g., via preventing the worker from engaging in regular activity) that lead to lower employment prospects. Empirically, I find evidence for all the three mechanisms: Shirking seems to be penalized immediately after sick leave, while statistical discrimination entails lower yet more persistent negative employment effects. Finally, also the sick leave itself appears to affect subsequent health directly, which indeed could affect future employment as well: Taking disability pension as an outcome in my framework (this is possible because it is granted by independent public health officers, not GPs), I find that high-certification propensity GPs *increase* the likelihood of the patient going into invalidity pension by granting one additional day of sick leave by 0.17%, but this effect is rather imprecisely estimated ($p = 0.186$). The probability of going into disability pension is largely independent of the employer herself or any previous signals of decreasing productivity, instead it is a mere reflection of the health status of the individual. Further research in necessary to assess the relative importance of these mechanisms.

My results raise one important recommendation for doctors: In case of doubt, it may be beneficial to certify shorter sick leaves whenever it is medically justifiable. Additionally, policy makers may consider introducing upper bounds of possible absence spell durations for certain groups of diagnoses.

6 Bibliography

Aakvik, A., Holmås, T. H., and Islam, M. K. (2010). Does Variation in General Practitioner (GP) Practice Matter for the Length of Sick Leave? A Multilevel Analysis Based on Norwegian GP-Patient Data. *Social Science & Medicine*, **70**(10), 1590 – 1598.

Abadie, A. (2003). Semiparametric Instrumental Variable Estimation of Treatment Response Models. *Journal of Econometrics*, **113**(2), 231–263.

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High Wage Workers and High Wage Firms. *Econometrica*, **67**(2), 251–333.

Ahammer, A., Horvath, T. G., and Winter-Ebmer, R. (2017). The Effect of Income on Mortality – New Evidence for the Absence of a Causal Link. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **180**(3), 793–816.

Ahammer, A. and Schober, T. (2017). Explaining Variations in Health Care Expenditures – What is the Role of Practice Styles? Working Paper 1705, Johannes Kepler University Linz, Department of Economics.

Amilon, A. and Wallette, M. (2009). Work Absence – A Signalling Factor for Temporary Workers? *Labour*, **23**(1), 171–194.

Andersen, S. H. (2010). The Cost of Sickness: On the Effect of the Duration of Sick Leave on Post-Sick Leave Earnings. *Social Science and Medicine*, **70**(10), 1581–1589.

Angrist, J. D. and Fernández-Val, I. (2013). ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework. In Acemoğlu, D., Arellano, M., and Dekel, E. (editors), *Advances in Economics and Econometrics: Volume 3, Econometrics: Tenth World Congress*, volume 51, page 401. Cambridge University Press.

Angrist, J. D. and Imbens, G. W. (1995). Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity. *Journal of the American Statistical Association*, **90**(430), 431–442.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Clogg, C. C., Petkova, E., and Haritou, A. (1995). Statistical Methods for Comparing Regression Coefficients Between Models. *American Journal of Sociology*, **100**(5), 1261–1293.

Correia, S. (2015). Singletons, Cluster-Robust Standard Errors and Fixed Effects: A Bad Mix. Mimeo, Duke University.

Federal Ministry of Social Affairs (2014). *Social Protection in Austria 2014*. Available online at https://broschuerenservice.sozialministerium.at, accessed: Wednesday 17th January, 2018.

Grytten, J. and Sørensen, R. (2003). Practice Variation and Physician-Specific Effects. *Journal of Health Economics*, **22**(3), 403–418.

Hackl, F., Hummer, M., and Pruckner, G. J. (2015). Old Boys' Network in General Practitioners' Referral Behavior? *Journal of Health Economics*, **43**, 56–73.

Halla, M., Mayr, H., Pruckner, G. J., and García-Gómez, P. (2016). Cutting Fertility? The Causal Effect of Supply-

Determined Cesarean Section on Fertility. Unpublished manuscript, Johannes Kepler University Linz, Department of Economics.

Hansen, J. (2000). The Effect of Work Absence on Wages and Wage Gaps in Sweden. *Journal of Population Economics*, **13**(1), 45–55.

Hesselius, P. (2007). Does Sickness Absence Increase the Risk of Unemployment? *Journal of Socio-Economics*, **36**(2), 288–310.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, **62**(2), 467–475.

Li, J., Fine, J., and Brookhart, A. (2015). Instrumental Variable Additive Hazards Models. *Biometrics*, **71**(1), 122–130.

Markussen, S. (2012). The Individual Cost of Sick Leave. *Journal of Population Economics*, **25**(4), 1287–1306.

Markussen, S., Røed, K., Røgeberg, O. J., and Gaure, S. (2011). The Anatomy of Absenteeism. *Journal of Health Economics*, **30**(2), 277–292.

Nilsen, S., Werner, E. L., Maeland, S., Eriksen, H. R., and Magnussen, L. H. (2011). Considerations made by the general practitioner when dealing with sick-listing of patients suffering from subjective and composite health complaints. *Scandinavian Journal of Primary Health Care*, **29**(1), 7.

Scoppa, V. and Vuri, D. (2014). Absenteeism, Unemployment and Employment Protection Legislation: Evidence from Italy. *IZA Journal of Labor Economics*, **3**(1), 3.

Zweimüller, J., Winter-Ebmer, R., Lalive, R., Kuhn, A., Wuellrich, J.-P., Ruf, O., and Büchi, S. (2009). Austrian Social Security Database. Working Paper 0903, NRN: The Austrian Center for Labor Economics and the Analysis of the Welfare State.

ONLINE APPENDIX

Physicians, Sick Leave Certificates, and Patients' Subsequent Employment Outcomes

This is an online appendix (not intended for publication) providing additional material for the paper *"Physicians, Sick Leave Certificates, and Patients' Subsequent Employment Outcomes"* by Alexander Ahammer.

A Additional Content

A.1 Arrangement of spells in the data

I construct a panel where each observation is a single sick leave spell. Let $S_{ik} = [t_{-n_{ik}}, t_0]$ denote a sick leave spell of individual $i = 1, \ldots, N$ with length n_{ik}, and let $E_{ik} = [t_0, t_{e_{ik}}]$ be the total remaining employment spell after t_0 with length e_{ik}. Each worker i may have $k = 1, \ldots, K_i$ different non-intersecting sick leaves ordered by $t_{-n_{ik}}$, followed by another k different remaining employment spells. Furthermore, let $E^* = \max\{e_{ik}\}$ be the longest employment spell experienced by any individual in the sample. Sick leave spells are recorded as the total time an employee was absent at work, not the actual time the GP certified (these may not necessarily coincide, for example if the worker decides to return to work earlier).

Figure A.1 provides an exemplary sketch of how spells are arranged in the data and outcomes are generated. The data are centered around t_0, where I measure the length of the completed sick leave spell as well as the employment duration which follows afterwards. In order to ease notation, I denote each subinterval by its endpoint, for example, $t_m = (t_{m-1}, t_m]$ for some $t_m \subset \tilde{E}_k$. Sick leave $k = 1$, for example, is assumed to last $n_{11} = 6$ days, and is followed by an employment spell of $e_{11} = 4$ months. The first four binary outcome variables $(y^e_{111}, \ldots, y^e_{114})$ are then equal to one, whereas the remaining 20 equal zero. Unemployment probabilities are calculated according to whether the employment spells E_{ik} result in unemployment or in another employment spell.

A.2 Descriptives

Descriptive statistics are provided in Table A.1. The average sick leave spell lasts around 6 days (the median is 5 days), while the average employment spell lasts 8 years. Figure A.2 depicts their distributions, which both are right-skewed. After a sick leave, the average *remaining* employment spell lasts 2.69 years (here, the median is 2.1 years). Surprisingly, a small yet negative reduced-form relationship can be observed in the raw data: Sick leaves certified by below-average propensity-to-certify doctors are followed by employment spells that last around 0.059 years ($\approx$ 22 days) longer than those following absences certified by more lenient doctors (this difference is statistically significant at the 5% level).

With a probability of 27%, the subsequent spell after the sick leave is an unemployment spell rather than a firm-to-firm transition. As expected, sick leaves certified by physicians with an above-average propensity to certify are on average 0.597 days longer ($p < 0.05$). After two years, 51% of all workers still belong to the same firm in which they worked in at t_0, while 22% registered at the unemployment office and 27% transitioned to a different firm.

Lenient GPs appear to be more often consulted by females, older patients, migrants (as compared to Austrian citizens), and lower income workers. Average levels of both tenure and experience in the sample are at 15.25 and 5.30 years, respectively. This can be interpreted as a sign that more sick leaves are taken towards the end of one's career, which is reasonable because workers health status decreases with age. Another reason, however, is simply that I dropped all apprentices from the sample, who indeed account for a large share of the young workforce in Austria.

Health proxies such as the amount of drug expenses aggregated over two years prior to the start of the sick leave, as well as aggregate days of hospitalization two years prior to the sick leave both seem to be lower for patients who consult more lenient doctors. Also, there seems to be a negative relationship between physician density and doctors' certification propensities. Finally, it is worthwhile to note that patients consulting high-propensity doctors tend to live in areas with higher unemployment rates.

The most common diagnoses for sick leaves are given in Table A.2 in the web appendix. Typical flus (for instance, J06.9, "acute upper respiratory infection") make up a considerable portion of all sick leaves. In total, 62% of all diagnoses can be attributed loosely to this category. The means of sick leave durations for such diagnoses lie between three and six days. Musculoskeletal diseases (indexed by the letter M), including conditions involving acute pain, account for another 13.6% of all diagnoses with mean sick leave durations being higher at seven to ten days. Potentially stress-related conditions, such as headaches (R51), migraine (G43), and major depressive disorders (F32), make up for 2.9% of cases. Burn-outs (Z73.0) are diagnosed 1,489 times during the observation period.

A.3 First-stage regression results

Figure A.3 depicts the first-stage relationship graphically, plotting the duration of individual sick leaves against estimated GP fixed effects from the AKM model in equation (2) in the main paper. As expected we see a positive relationship between these two variables in the raw data: The higher a GP's propensity to certify sick leaves, the longer their actual durations. The results from estimating the first-stage are summarized in Table A.3. Judging from the coefficient $\hat{\delta}$, consulting a GP who has an above-average propensity to certify sick leaves increases the duration of a single sick leave spell by roughly half a day ($p < 0.01$). Using other cutoff points, e.g., the median or the 90$^{\text{th}}$ percentile of the fixed effect distribution, yields almost identical results, with F-statistics being far beyond the conventional rule-of-thumb level of 10. Likewise, second-stage estimates change only little with the choice of the cutoff point as well. All 2SLS estimations reported in the remainder of this paper are available using one of the other IVs specified in Table A.3 upon request.

A.4 Full regression results for month 3 after the sick leave

Main estimates for the effect of a marginal day of sick leave on employment outcomes are provided in section 4 in the main paper. In order to gain a more comprehensive picture—especially with regard to coefficients of control variables and test statistics—I show full regression results for t_3 (where the LATE is strongest in magnitude for both employment and unemployment probabilities) in Table A.4. Notice that F-statistics of the excluded instrument are well above 1,000 in all specifications. The outcome variable in columns E.1—E.6 is the probability of still being employed in the same firm three months after the sick leave, whereas the outcome in columns U.1—U.6 is the probability of becoming unemployed during the first three months after the sick leave. The table is organized such that the model is extended in various steps.

Column E.3 show the estimated coefficients for a model without any covariates. Here, the LATE is estimated to be –0.0063 ($p < 0.01$), indicating that a marginal day of sick leave decreases employment probability during the first three months by roughly 0.63 pps. In column E.4, the model is augmented with worker-level characteristics such as age, wages, and occupational characteristics. All of them have a significant impact on the employment probability: The age effect is inverted U-shaped, higher initial income and being a part-time worker lead to higher employment probabilities; while tenure, experience, and being a blue collar worker are associated with lower employment prospects. After inclusion of these covariates, the LATE decreases slightly to –0.0048 ($p < 0.01$). Incorporating health proxies does not change the estimated coefficients by much. In contrast, firm size and macroeconomic conditions have a sizable effect on employment probability. Competition amongst doctors (measured through the GP density at the community level) does not appear to have a significant effect on employment.

My preferred specification is the full model in column E.6, where the LATE is –0.0059 ($p < 0.01$) suggesting that the employment probability three months after the absence spell is reduced by 0.59 pps. for each marginal day of sick leave. OLS effects are similar to the IV estimates across specifications. The results for unemployment probabilities largely mirror the ones for employment probabilities. In my preferred specification (column U.6), the LATE is estimated as 0.0044 ($p < 0.05$), implying that a marginal day of sick leave increases the probability of becoming unemployed after three months by 0.44 pps.

A.5 Complier analysis

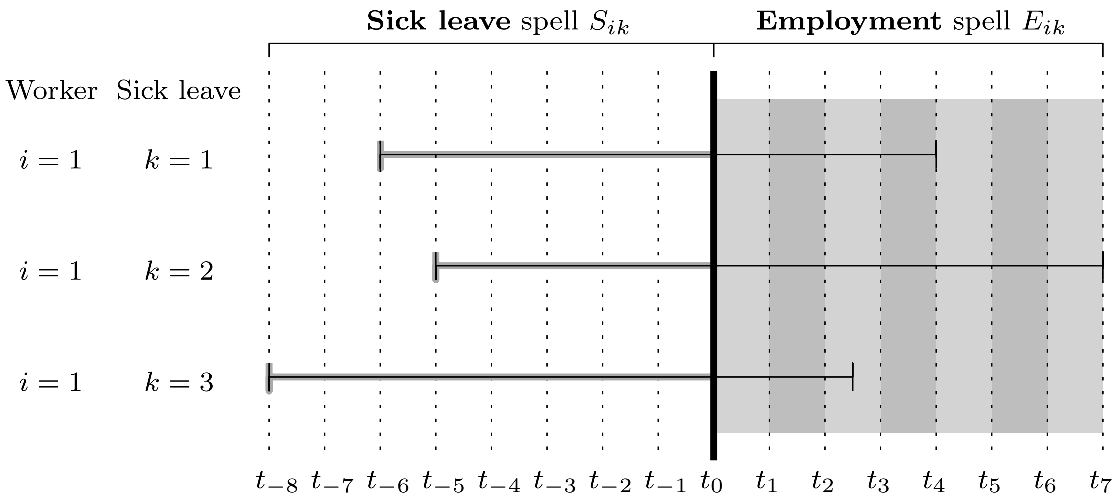
Following Angrist and Imbens (1995), I examine compliance (which crucially determines unit causal responses) by comparing the cumulative distribution function (CDF) of sick leave duration n_{ik} when the instrument Λ_i is switched on and off. Figure A.4 plots these CDFs in the left-hand graph, their difference (i.e., the difference in probabilities that n_i is greater or equal to the respective level on the horizontal axis when $\Lambda_i = 0$ and $\Lambda_i = 1$) is illustrated in the right-hand graph. Compliers are located almost exclusively between 3 and 9 days of sick leave along the support of n_{ik}, with the maximum being at 5 days. Thus, the $\hat{\rho}_m$'s are identified primarily through patients whose counterfactual sick leave duration of 5 to 9 days is extended by consulting a lenient physician.

Listing the five most common ICD-10 codes whose average sick leave duration in the sample lies between 4 and 6 days (which is ± 1 day around the maximum, see Table A.5) shows an interesting pattern: These are outright diseases of the respiratory system. Thus, it is mostly acute colds and flus for which doctors who have a high certification propensity grant an additional day of sick leave.

Although compliers cannot be identified individually, we can learn about distributional features of their demographic and occupational characteristics applying simple calculations proposed by Angrist and Fernández-Val (2013) and Abadie (2003). Results of these calculations are reported in Table A.6 (web appendix). Compliers are 18.8% more likely to be migrants, 11.1% less likely to have at least an A-level degree, 6.3% less likely to be part-time workers, and 11.1% more likely to be blue-collar workers. In terms of gender, compliers are equally likely male or female. Furthermore, compliers are on average 35.7 years old, earn 26,715 Euros, have around 15.3 years of experience, and roughly 5.3 years of tenure. It appears as though compliers are largely located near the means of the independent variables in the model, which is highly beneficial in terms of external validity.

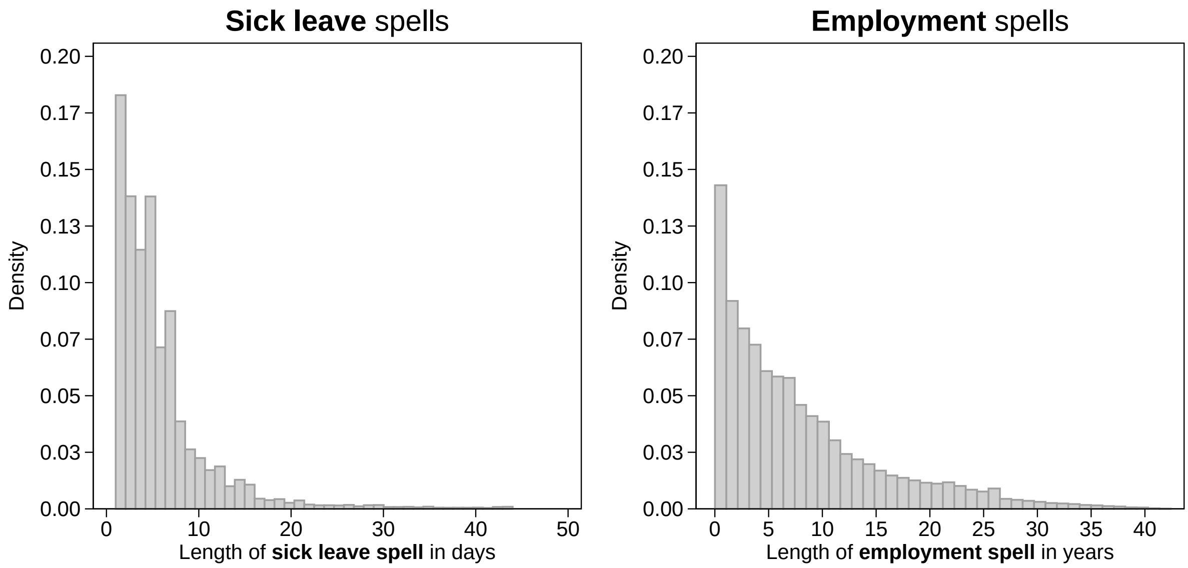
Tables and Figures

Figure A.1 — Exemplary sketch of spell arrangement.



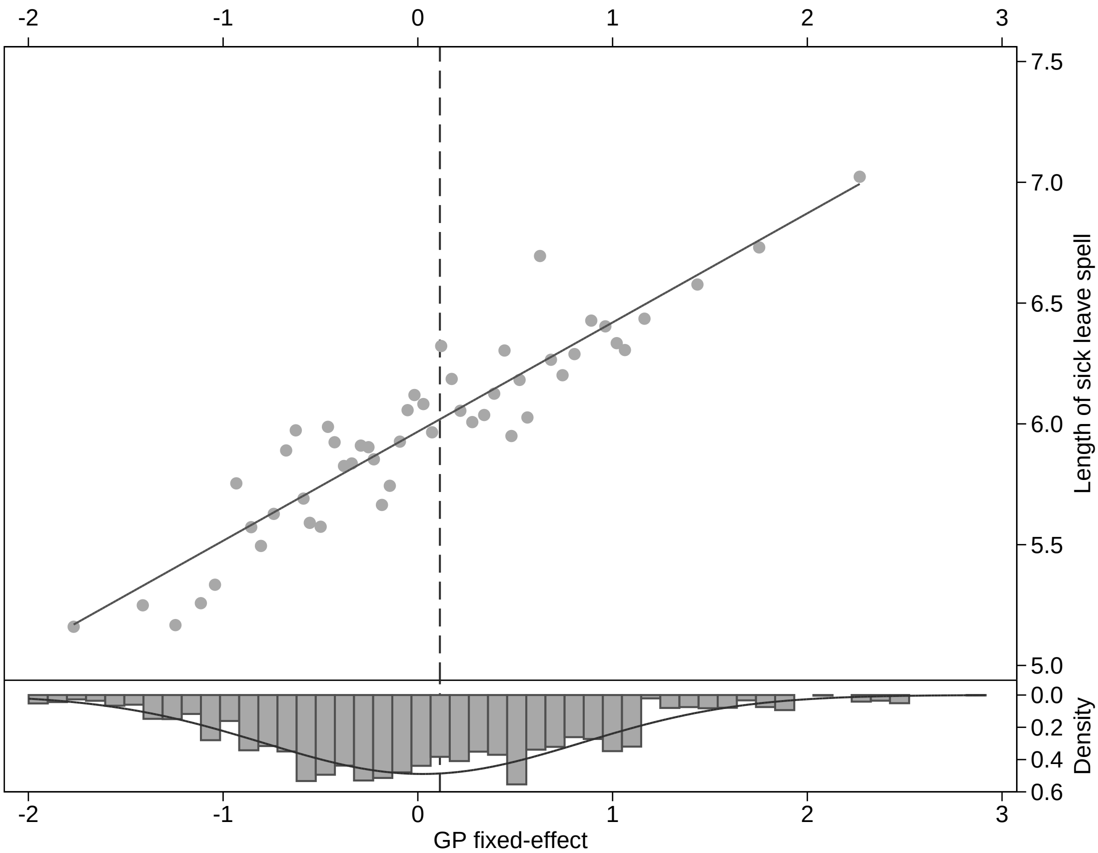
Notes: This graph provides a sketch of how spells are arranged in the data. It displays three sick leaves of an individual $i = 1$, whose first sick leave S_{11} lasts $n_{11} = 6$ days and is employed for four months afterwards ($e_{11} = 4$). If the employment spell is split into monthly subintervals, the first four employment dummies for intervals t_1, t_2, t_3, and t_4 would be one, while the remaining ones would be zero. The second sick leave spells then lasts $n_{12} = 5$ days, and the employment spell afterwards $e_{12} = 7$ months. Finally, the third sick leave spell lasts $n_{13} = 8$ days with a posterior employment spell of $e_{13} = 7$ months.

Figure A.2 — Spell length distributions.



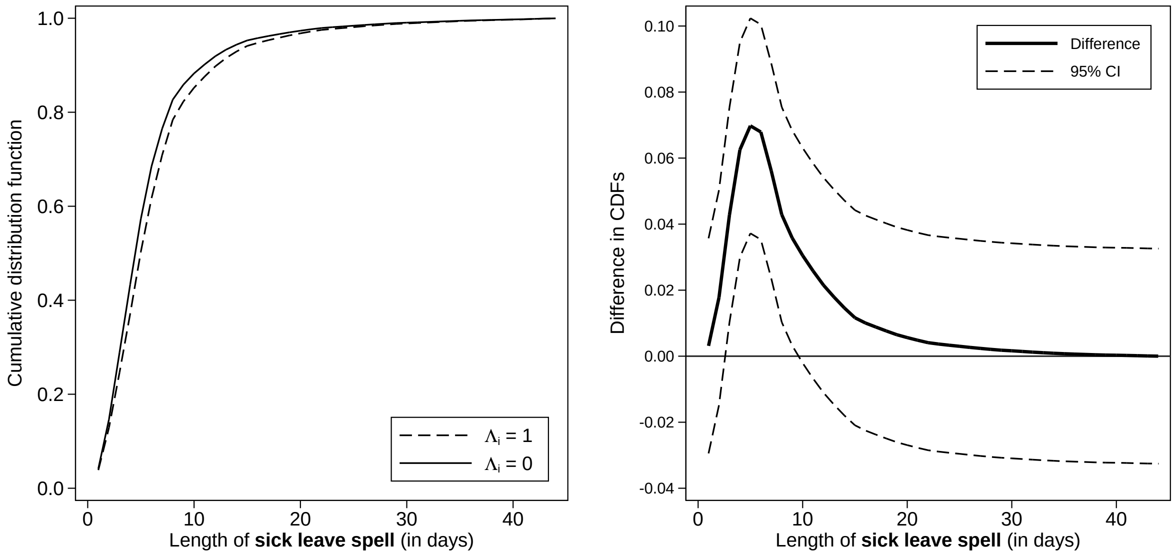
Notes: These graphs depict the distribution of sick leave spell durations (n_k, left figure) and total employment spell durations (right figure) in the data.

FIGURE A.3 — Length of sick leave spells versus estimated GP fixed effects.



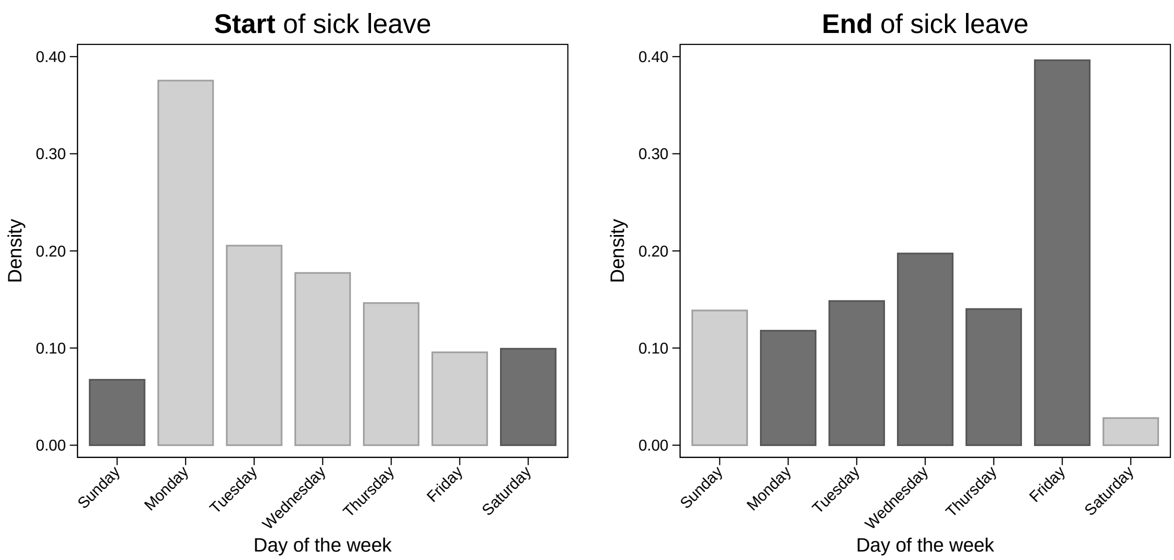
Notes: This graph illustrates the relationship between estimated GP fixed effects $\hat{\psi}_d$ on the horizontal axis and sick leave durations n_k on the vertical axis. Due to the large sample size (N = 3,125,759), observations are grouped into 100 equally sized bins. Within each bin, means of E_k and n_k are calculated and then plotted in the upper graph. The solid line indicating fitted values is calculated based upon all observations in the data. Furthermore, the distribution of estimated GP fixed effects $\hat{\psi}_d$ along with a hypothetical normal distribution are plotted underneath the graph (65,401 observations whose fixed effect lies outside the interval $[-2, 3]$ are not shown for presentational reasons). The dashed line indicates the sample mean of $\hat{\psi}_d$.

Figure A.4 — Complier location across the sick leave spell length distribution.



Notes: The left graph illustrates the cumulative distribution functions (CDF) of the sick leave duration for both realizations of the instrument ($\Lambda_i = 1$ and $\Lambda_i = 0$). The right graph plots the difference in those CDFs, i.e., the differences in the probability that sick leave duration is greater or equal than the respective level on the horizontal axis. Additionally, the 95% confidence interval for the difference function is provided.

Figure A.5 — Starting and ending days of sick leaves.



Notes: These graphs illustrate the distribution of both the first week day of the sick leave (left graph) and the last week day of the sick leave (right graph).

Table A.1 — Descriptive statistics.

	Entire sample		$\hat{\psi}_{d(it)} > \bar{\bar{\psi}}_d$		$\hat{\psi}_{d(it)} \leq \bar{\bar{\psi}}_d$		
	Mean (1)	Std. dev. (2)	Mean (3)	Std. dev. (4)	Mean (5)	Std. dev. (6)	Difference (7)
Spells							
Length of sick leave spell (in days)	5.99	(5.34)	6.33	(5.51)	5.73	(5.18)	-0.597***
Length of total employment spell (in years)	8.00	(7.62)	7.90	(7.60)	8.07	(7.63)	0.164***
Length of remaining employment spell (in years)	2.69	(2.42)	2.65	(2.43)	2.71	(2.41)	0.059***
Subsequent spell is unemployment spell	0.27	(0.45)	0.28	(0.45)	0.27	(0.44)	-0.016***
Outcome variables							
$\mathbb{P}[i$ is still employed at the end of $t_{24}]$	0.51		0.51		0.52		0.015***
$\mathbb{P}[i$ became unemployed between t_0 and $t_{24}]$	0.22		0.23		0.21		-0.014***
Instrumental Variable							
Estimated GP fixed-effect ($\hat{\psi}_{d(it)}$)	0.11	(1.27)	0.98	(1.41)	-0.57	(0.50)	-1.550***
Binary instrument ($\Lambda_i \equiv \mathbf{1}\{\hat{\psi}_{d(it)} > \bar{\bar{\psi}}_d\}$)	0.44						
Control variables							
Age	36.78	(11.19)	36.89	(11.16)	36.70	(11.21)	-0.188***
Part-time worker	0.17		0.17		0.17		-0.007***
Female	0.40		0.40		0.39		-0.013***
Migrant	0.20		0.21		0.19		-0.022***
At least A-level degree	0.34		0.35		0.34		-0.006***
log(annual wage prior to sick leave)	9.94	(0.84)	9.92	(0.86)	9.96	(0.83)	0.032***
Experience until start of sick leave (in years)	15.25	(5.78)	15.20	(5.79)	15.28	(5.78)	0.084***
Tenure until start of sick leave (in years)	5.30	(6.57)	5.24	(6.53)	5.34	(6.60)	0.107***
Part-time worker	0.17		0.17		0.17		-0.007***
Blue collar worker	0.61		0.61		0.62		0.009***
log(drug expenses 2 years prior to leave)	4.94	(1.99)	4.99	(1.98)	4.89	(2.00)	-0.103***
Days of hospitalization 2 years prior to leave	3.42	(2.76)	3.47	(2.76)	3.39	(2.76)	-0.074***
log(firm size)	0.73	(6.20)	0.73	(6.28)	0.72	(6.13)	-0.005
Physician density within community[a]	0.85	(0.35)	0.81	(0.35)	0.88	(0.35)	0.063***
Unemployment rate at industry sector level[b]	8.28	(4.17)	8.39	(4.26)	8.18	(4.09)	-0.206***
Number of observations (N^*)	3,125,759		1,373,340		1,752,419		
Number of different workers (N)	423,352		250,976		317,404		
Number of different firms (J)	43,297		31,373		36,293		
Number of different general practitioners (D)	1,078		350		728		

Notes: This table reports descriptive statistics for all variables used throughout the empirical analysis. The unit of observation is an individual sick leave spell. In columns (3) to (6) the sample is split into sick leaves certified by physicians having an above-average propensity to certify sick leaves [(3) and (4)] and those certified by physicians having a below-average propensity [(5) and (6)]. In column (7) the differences in means between (3) and (5) are tested for statistical significance using Welch's t-test. *** denotes statistical significance at the 5% level ($p < 0.05$).

[a] Number of GPs per 10,000 inhabitants within a community.
[b] Number of unemployed workers divided by the total work force for each NACE95 two-digit industry sector.

Table A.2 — Most common medical conditions, only diagnoses with more than 20,000 cases in the data.

		Occurences		Sick leave durations	
ICD-10 code	Description	No. of cases	in %	Mean of n_k	Std. dev.
J06.9	Acute upper respiratory infection	957665	31.42%	5.39	(3.32)
A09	Infectious gastroenteritis and colitis	329143	10.80%	3.83	(2.86)
M53	Other and unspecified dorsopathies	169131	5.55%	8.18	(6.94)
J40	Bronchitis, not specified as acute or chronic	111004	3.64%	6.10	(3.87)
J02	Streptococcal pharyngitis	81368	2.67%	4.90	(3.04)
J01	Acute sinusitis	67769	2.22%	5.75	(3.65)
J20	Acute bronchitis	59194	1.94%	6.06	(3.80)
M54.5	Low back pain	58922	1.93%	7.16	(6.28)
B34.8	Other viral infections of unspecified site	51999	1.71%	4.50	(2.95)
J03	Acute tonsillitis	40135	1.32%	5.34	(3.15)
M54.4	Lumbago with sciatica	37977	1.25%	9.60	(7.85)
J06	Acute upper respiratory infections	37363	1.23%	5.63	(3.62)
A08.5	Other specified intestinal infections	37190	1.22%	3.47	(2.69)
K29	Gastritis and duodenitis	27216	0.89%	4.63	(4.49)
R51	Headache	26752	0.88%	3.95	(4.42)
G43	Migraine	26216	0.86%	2.50	(2.96)
F32	Major depressive disorder, single episode	24228	0.79%	11.62	(9.80)
J04	Acute laryngitis and tracheitis	23862	0.78%	5.45	(3.45)
N39.0	Urinary tract infection, site not specified	20710	0.68%	4.65	(4.11)

Notes: This table presents all ICD-10 codes with more than 20,000 cases in the data, including sample means and standard deviations of sick leaves certified based on these diagnoses.

Table A.3 — Summary of first-stage regression results for different choices of Λ_d

Instrumental variable	Explanation	$\hat{\delta}$	Std. err.	F-statistic[a]	partial R^2
$\Lambda_d \equiv \hat{\psi}_d$	$\hat{\psi}_d$ is continuous	0.2444	(0.005)***	2275.3	0.00084
$\Lambda_d \equiv \mathbf{1}\{\hat{\psi}_d > \bar{\hat{\psi}}_d\}$	$\hat{\psi}_d$ is above its sample mean	0.4583	(0.011)***	1750.8	0.00065
$\Lambda_d \equiv \mathbf{1}\{\hat{\psi}_d > \hat{\psi}_{d,50}\}$	$\hat{\psi}_d$ is above its sample median	0.4701	(0.011)***	1939.0	0.00072
$\Lambda_d \equiv \mathbf{1}\{\hat{\psi}_d > \hat{\psi}_{d,90}\}$	$\hat{\psi}_d$ is above its 90$^{\text{th}}$ percentile	0.5833	(0.020)***	892.4	0.00033

Notes: This table summarizes results of estimating the first-stage equation in (1) with different IVs. Each row represents a separate regression where sick leave duration n_{ik} is regressed on the IV $\Lambda_{d(it)}$, a vector $\mathbf{x}_{ik}$ of control variables described in Section 3, and a full set of worker-level fixed effects. The number of observations in all regressions is 3,125,759. Standard errors are heteroskedasticity-robust and clustered on the worker-level. *** indicates statistical significance at the 1% level ($p < 0.01$).

[a] Kleinbergen Paap *rk* F-statistic.

TABLE A.4 — The effect of a marginal day of sick leave on employment and unemployment at t_3 (three months after the end of the sick leave).

	Dependent variable: Pr[i is still employed at the end of t_3]						Dependent variable: Pr[i became unemployed between t_0 and t_3]					
	OLS		IV-LATE (instrument: $\Lambda_{d(it)}$)				OLS		IV-LATE (instrument: $\Lambda_{d(it)}$)			
	(E.1)	(E.2)	(E.3)	(E.4)	(E.5)	(E.6)	(U.1)	(U.2)	(U.3)	(U.4)	(U.5)	(U.6)
Length of sick leave spell (in days)	-0.0051 (0.000)***	-0.0045 (0.000)***	-0.0063 (0.001)***	-0.0048 (0.001)***	-0.0049 (0.001)***	-0.0059 (0.001)***	0.0035 (0.000)***	0.0033 (0.000)***	0.0039 (0.001)***	0.0036 (0.001)***	0.0036 (0.001)***	0.0044 (0.001)***
Age		-0.0492 (0.001)***		0.0038 (0.000)***	0.0037 (0.000)***	-0.0504 (0.001)***		0.0434 (0.001)***		0.0041 (0.000)***	0.0041 (0.000)***	0.0443 (0.001)***
Age^2		-0.0001 (0.000)***		-0.0001 (0.000)***	-0.0001 (0.000)***	-0.0001 (0.000)***		-0.0000 (0.000)***		-0.0000 (0.000)	-0.0000 (0.000)	-0.0000 (0.000)***
log(annual wage at $t_{-n_{ik}}$)		0.1860 (0.000)***		0.1874 (0.001)***	0.1873 (0.001)***	0.1855 (0.001)***		-0.0954 (0.000)***		-0.0965 (0.000)***	-0.0965 (0.000)***	-0.0950 (0.000)***
Part-time worker		0.0485 (0.001)***		0.0465 (0.001)***	0.0465 (0.001)***	0.0483 (0.001)***		-0.0261 (0.001)***		-0.0253 (0.001)***	-0.0253 (0.001)***	-0.0260 (0.001)***
Tenure until $t_{-n_{ik}}$ (in years)		-0.0100 (0.000)***		-0.0095 (0.000)***	-0.0095 (0.000)***	-0.0099 (0.000)***		0.0030 (0.000)***		0.0026 (0.000)***	0.0026 (0.000)***	0.0029 (0.000)***
Experience until $t_{-n_{ik}}$ (in years)		-0.0002 (0.000)***		-0.0002 (0.000)***	-0.0002 (0.000)***	-0.0002 (0.000)***		0.0001 (0.000)**		0.0001 (0.000)**	0.0001 (0.000)**	0.0001 (0.000)**
Blue collar worker		-0.0402 (0.001)***		-0.0491 (0.001)***	-0.0491 (0.001)***	-0.0401 (0.001)***		0.0230 (0.001)***		0.0277 (0.001)***	0.0277 (0.001)***	0.0229 (0.001)***
log(drug expenses two years prior to t_{-n_k})		0.0003 (0.000)***			0.0005 (0.000)**	0.0006 (0.000)**		-0.0001 (0.000)*			-0.0003 (0.000)	-0.0003 (0.000)
Days of hospitalization two years prior to t_{n_k}		-0.0001 (0.000)**			-0.0000 (0.000)	0.0000 (0.000)		-0.0000 (0.000)			-0.0001 (0.000)	-0.0001 (0.000)
log(firm size)		0.0029 (0.000)***				0.0029 (0.000)***		-0.0053 (0.000)***				-0.0053 (0.000)***
GP density at community level[a]		-0.0012 (0.001)				-0.0012 (0.001)		0.0006 (0.001)				0.0006 (0.001)
Unemployment rate at industry level[b]		(0.001) (0.000)***				(0.001) (0.000)***		(0.001) (0.000)***				(0.001) (0.000)***
Region dummies	No	Yes	No	No	No	Yes	No	Yes	No	No	No	Yes
Year dummies	No	Yes	No	No	No	Yes	No	Yes	No	No	No	Yes
N	3,125,759	3,125,759	3,125,759	3,125,759	3,125,759	3,125,759	3,125,759	3,125,759	3,125,759	3,125,759	3,125,759	3,125,759
Mean of outcome	0.86	0.86	0.86	0.86	0.86	0.86	0.08	0.08	0.08	0.08	0.08	0.08
First-stage F-statistic			2,100.64	1,854.19	1,776.94	1,750.84			2,100.64	1,854.19	1,776.94	1,750.84

Notes: Columns E.1, E.2, U.1, and U.2 are estimated via ordinary least squares (OLS), columns E.3 through E.6 as well as columns U.3 through U.6 are estimated via two-stage least squares (2SLS) where the IV is defined in equation (3). All regressions incorporating worker-level fixed effects as well. The outcome is a binary variable equal to unity if the worker is still employed in the same firm as in t_0 (i.e., the end of the sick leave) after three months for columns E.1—E.6, and a binary variable equal to unity if the worker became unemployed at some point of time between t_0 (i.e., the end of the sick leave) and t_3 (i.e., three months after the sick leave) for columns U.1—U.6. The coefficient on "length of sick leave spell" represents the local average treatment effect (LATE) of a marginal day of sick leave. Heteroskedasticity-robust and worker-level clustered standard errors are given in parentheses below coefficients, stars indicate significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

[a] Measured as number of GPs per 10,000 inhabitants within a community.

[b] Measured as number of unemployed workers divided by the total work force for each NACE95 two-digit industry sector.

Table A.5 — Most common medical conditions with average sick leave durations between 4 and 6 days.

		Occurences		Sick leave durations	
ICD-10 code	Description	No. of cases	in %	Mean of n_k	Std. dev.
J06.9	Acute upper respiratory infection	957,665	31.42%	5.39	(3.32)
J02	Acute pharyngitis	81,368	2.67%	4.90	(3.04)
J01	Acute sinusitis	67,769	2.22%	5.75	(3.65)
B34.8	Other viral infections of unspecified site	51,999	1.71%	4.50	(2.95)
J03	Acute tonsillitis	40,135	1.32%	5.34	(3.15)

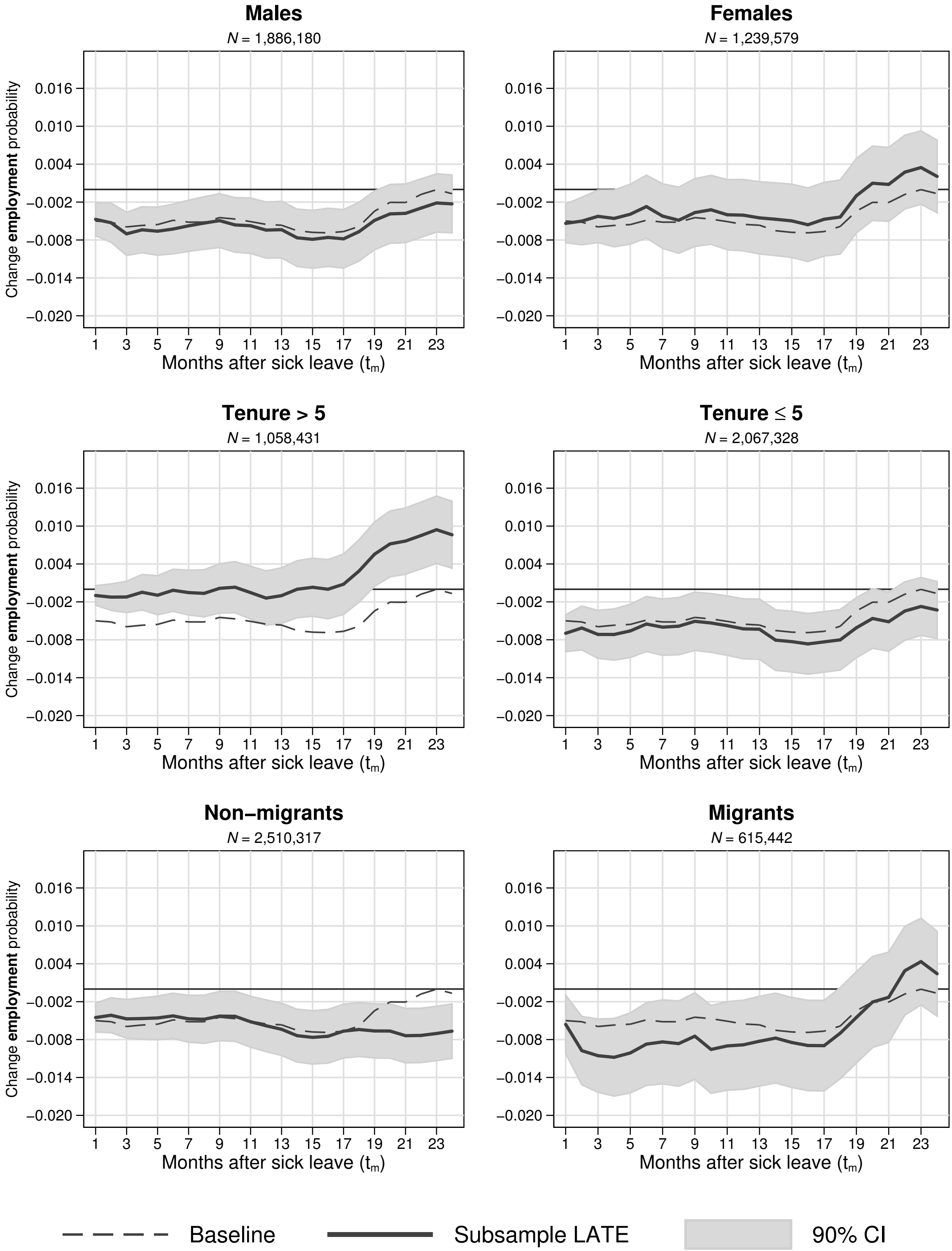
Notes: This table presents the five most common ICD-10 codes whose average sick leave duration in the sample is between 4 and 6 days.

Table A.6 — Characteristics of compliers.

Binary covariates	Female	Migrant	High educ.	Part-time	Blue-collar
Sample average	0.397	0.197	0.342	0.169	0.612
Relative likelihood of being a complier	1.000	1.188	0.889	0.937	1.111
Continuous covariates	Age	Wage	Experience	Tenure	
Sample average	36.8	26,077.5	15.2	5.3	
Average among compliers	35.7	26,715.0	15.3	5.3	

Notes: This table reports characteristics of compliers based on calculations derived in Angrist and Fernández-Val (2013). All covariates are measured at the beginning of the sick leave. High education is a binary variable indicating whether the observation has at least an A-level degree. Age, wage, and tenure are given in years, wage is given in Euros. The number of observations is 3,125,759 in all cells.

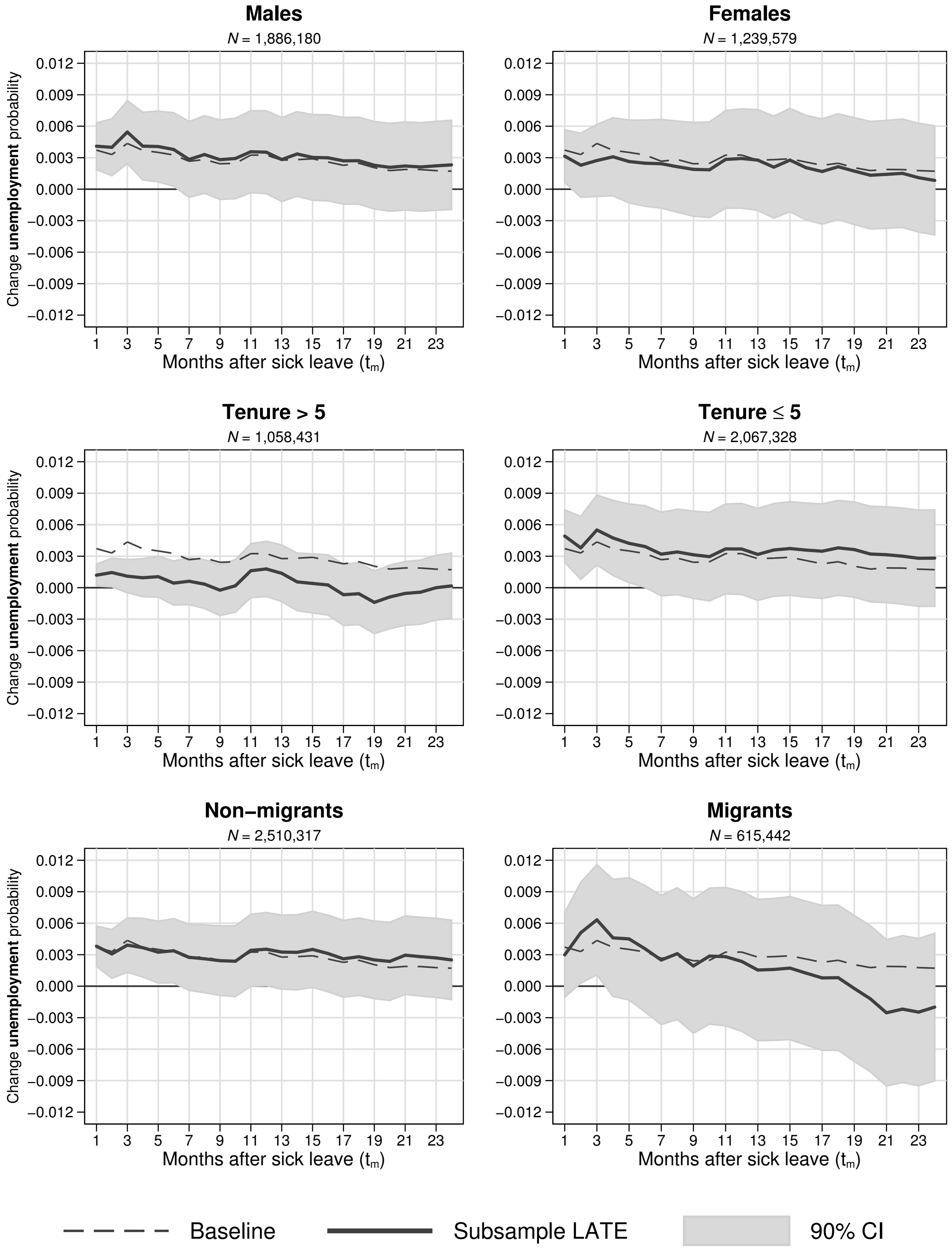
Figure B.1 — Estimated employment dynamics for different subsamples of the population.



Notes: These figures plot the estimated local average treatment effects, $\hat{\rho}_m$, $m = 1, \ldots, 24$, of a marginal day of sick leave on employment probabilities for different subsamples of the population. Each coefficient for $t_1, \ldots, t_{24}$ is estimated from a separate regression based on the model in (1). The dashed line plots the baseline results from Figure 1.

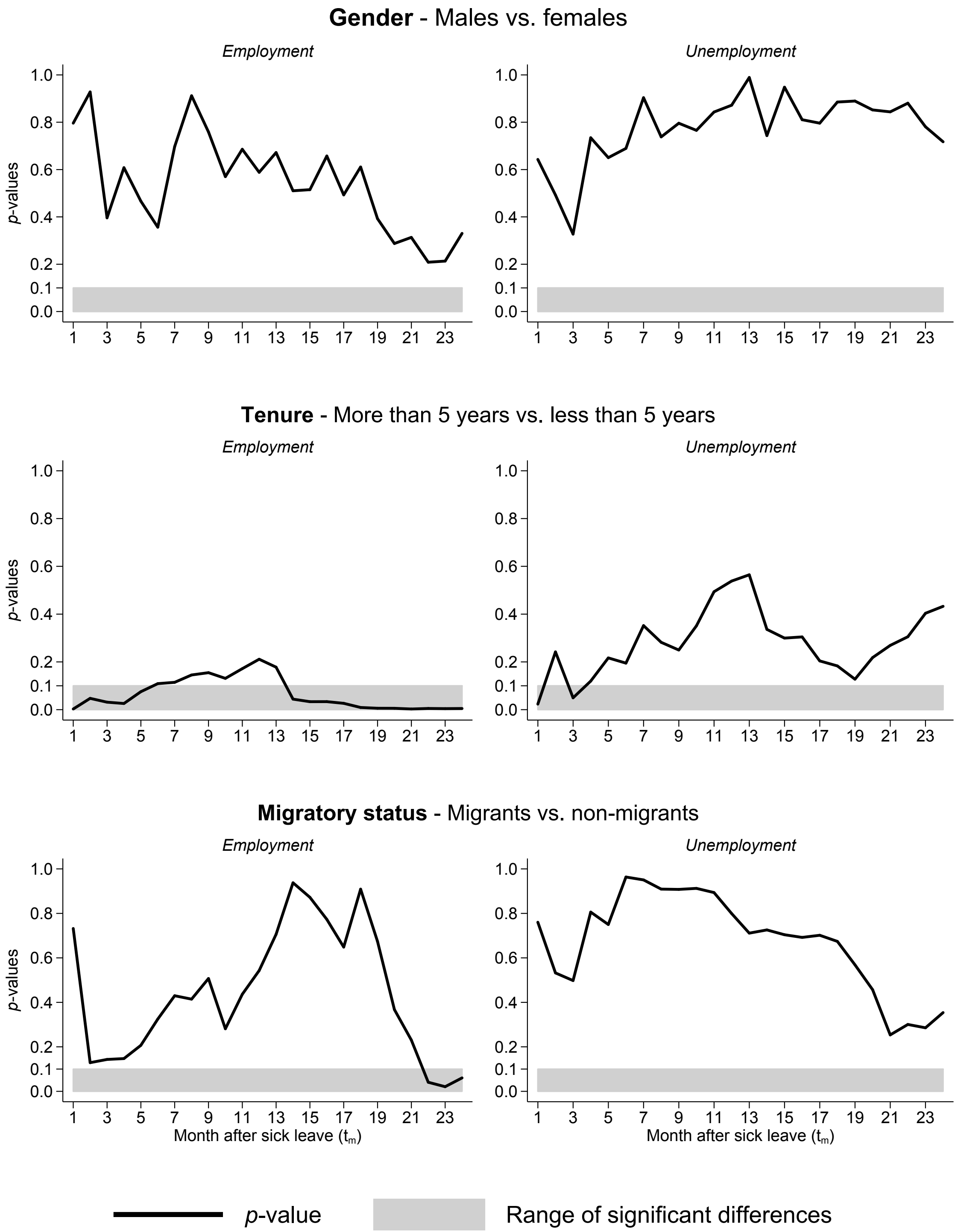
B Additional Tables and Figures

Figure B.2 — Estimated unemployment dynamics for different subsamples of the population.



Notes: These figures plot the estimated local average treatment effects, $\hat{\rho}_m$, $m = 1, \ldots, 24$, of a marginal day of sick leave on unemployment probabilities for different subsamples of the population. Each coefficient for $t_1, \ldots, t_{24}$ is estimated from a separate regression based on the model in (1). The dashed line plots the baseline results from Figure 1.

Figure B.3 — Statistical significance of differences in heterogeneous effects between subgroups.



Notes: In this graph I test whether the subsample LATE coefficients estimated in section 4.1 (in particular Figures B.1 and B.2) are significantly different from each other. I follow Clogg *et al.* (1995) and calculate the z-statistic $\frac{\hat{\rho}^a_m - \hat{\rho}^b_m}{\sqrt{(\hat{\sigma}^a_{\rho m})^2 + (\hat{\sigma}^b_{\rho m})^2}}$, where a and b are the groups which are tested against each other (e.g., males and females), $\hat{\rho}_m$ is the 2SLS coefficient from equation (1), and $\hat{\sigma}_{\rho m}$ is its standard error. In the figure above, we plot the corresponding p-values for this z-statistic, for each month m after the end of the sick leave. If the p-value is below 0.5,

Table B.1 — Summary statistics comparing observations whose fitted values are inside the unit interval vs. observations who are not in month three after the sick leave (N = 3,125,759).

	Not inside		Inside			
	Mean	N	Mean	N	Diff.	p-value
Length of sick leave	5.68	1800292	6.41	1325467	-0.73***	0.00
Posterior employment duration	2.30	1800292	3.21	1325467	-0.91***	0.00
Posterior unemployment probability	0.29	1800292	0.25	1325467	0.04***	0.00
Female	0.39	1800292	0.41	1325467	-0.02***	0.00
log(wage)	9.88	1800292	10.02	1325467	-0.14***	0.00
Part-time worker	0.00	1800292	0.00	1325467	0.00***	0.00
Experience	14.10	1800292	16.80	1325467	-2.69***	0.00
Tenure	4.30	1800292	6.64	1325467	-2.34***	0.00
Blue collar	0.61	1800292	0.62	1325467	-0.01***	0.00
White collar	0.39	1800292	0.38	1325467	0.01***	0.00
log(firm size)	4.86	1800292	5.03	1325467	-0.17***	0.00
GP fixed effect	0.11	1800292	0.12	1325467	-0.01***	0.00
Binary instrument	0.44	1800292	0.44	1325467	-0.01***	0.00
Hospital history	0.65	1800292	0.82	1325467	-0.17***	0.00
log(medication history)	3.21	1800292	3.72	1325467	-0.52***	0.00