Economics of Transportation 33 (2023) 100287

Contents lists available at ScienceDirect

Economics of Transportation

journal homepage: www.elsevier.com/locate/ecotra

Urban commuting time and sick-leave medical license use: An empirical study of Santiago, Chile[☆]

Andrés Gómez-Lobo [*], Alejandro Micco

Department of Economics, University of Chile, Chile

ARTICLE INFO

Keywords:
Sick-leave behavior
Transport project appraisal
Commuting costs
Urban mobility

ABSTRACT

We use a large dataset from the Chilean unemployment insurance program covering 20% of all formal sector workers to study the impact of commuting time on the likelihood of sick-leave in Santiago, Chile. Our empirical results indicate that longer commuting times are associated with an increase in the probability of sick-leave work absence. A 20% decrease in commuting times would generate close to 36 million dollars per year in productivity benefits. Our results also suggest that commuting travel time improvements targeted to women, lower paid workers and relatively older workers would provide the highest benefits in terms of lowering sick-leave behavior. We also find evidence that mobility infrastructure investments, such as metro and commuter rail expansions, reduce the probability of sick-leave. The results of this paper have implications for measuring the social costs of congestion and for the estimation of the wider economic benefits of transport projects.

1. Introduction

The impact of commuting time and distance on absenteeism or sick-leave has not been profusely studied in the academic literature. The growing realization that transport infrastructure generates 'wider economic benefits' not taken into account by traditional cost–benefit analysis has generated a new area of research and is changing the way transport authorities appraise projects.[1] This literature has considered the impact of transport projects on general agglomeration economies (Venables, 2007), imperfect competition (Kanemoto, 2013b,a), spatial mismatch in the labor market (Eliasson and Fosgerau, 2017), among others. To date however, the impact of commuting on health and absenteeism has not been considered as part of the wider economic benefits of transport projects.

In this paper we take up the issue of commuting and sick-leave. There are two strands to the existing literature. One attempts to measure the effects of commuting time and distance on subjective and objective health outcomes, assuming that there is a direct or indirect link between commuting and health, or what may be called *involuntary* absenteeism.[2] Künn-Nelen (2016) is an example of this approach. The other strand emphasizes the endogeneity of the decision to call in sick and may be termed the *voluntary* absenteeism approach as in Ross and Zenou (2008) and Van Ommeren and Gutiérrez-i Puigarnau (2011).

We follow the above papers and use an individual unbalanced panel of formal sector workers in Santiago, Chile, to study the effects of commuting time on sick-leave behavior between the first quarter of 2006 and the first quarter of 2019. This data comes from the Chilean unemployment insurance scheme and contains information on wages, type of labor contract, medical licenses, education, marital status, age, gender, residential location as well as the identity, location, economic sector and firm size of the employer, among other variables. Our data is a publicly available random sample of 20% of all workers of the unemployment insurance scheme as of March 2019.

Observing a worker with a medical license in a given month is our indicator of sick-leave behavior. These licenses are granted by doctors if a worker is ill, had a work or non-work related accident or, in the

☆ This research was undertaken with funding from the Inter-American Development Bank. Powered@NLHPC: This research was partially supported by the supercomputing infrastructure of the NLHPC (ECM-02). We would like to thank Giovanni Villa for his excellent research assistance. We also thank Daniel Perez Jaramillo, Santiago Sanchez, Damian Clarke, as well as two anonymous referees for their useful comments. Neither them nor the Inter-American Development Bank are responsible for the opinions, errors or omissions contained in the present paper.

* Corresponding author.
E-mail addresses: agomezlo@fen.uchile.cl (A. Gómez-Lobo), amicco@fen.uchile.cl (A. Micco).

1 See Venables et al. (2014) for a review of these wider economic benefits and the UK Department for Transport (2018) for a concrete application to project appraisal.

2 By direct link we refer to impacts of commuting on health outcomes due to higher levels of stress, for example. By indirect link we refer to the impacts of commuting on certain activities, such as leisure, physical activities or nutritional habits that then affect health outcomes. This literature makes a distinction between active commuting (walking, bicycle) and passive commuting (car, public transport) given that the health impacts may be very different.

https://doi.org/10.1016/j.ecotra.2022.100287
Received 12 January 2022; Received in revised form 11 July 2022; Accepted 28 October 2022
Available online 25 November 2022

case of women, also for maternal leave, a pregnancy pathology or if she has a child under one year of age with a serious illness.[3] The medical license will prescribe the number of days a worker can legally be absent from his or her employment due to health reasons.

Compared to previous studies, our data has several advantages. First, it is monthly rather than annual data. Second, it is a very large random sample, with over 39 million observations from over 500 thousand individual workers, reducing sampling bias, at least as formal sector workers are concerned. Third, previous studies use a measure of sick-leave that is self-reported and often refers to the number of days absent from employment during the last year. In contrast, in this paper, we have administrative data on whether a worker took sick-leave for each month, thus reducing measurement error considerably. Fourth, although locational data for workers and employers are at the municipal level (44 urban and sub-urban counties close to Santiago) we use a grid of 718 zones from the 2012 Santiago mobility survey (ESTRAUS zones) to estimate average commuting time for each municipal origin–destination pair, and do not rely on recall survey questions as is the case with most of the received literature (however, we are not able to identify travel mode in our data).

Another feature of our paper is that we analyze the effects of major transport investments in Santiago (the extension of the underground Metro system and one sub-urban commuter train) on sick-leave behavior. Tsivanidis (2018) uses a general equilibrium model to evaluate the impact of the Transmilenio BRT reform in Bogota, Colombia, on labor and land market outcomes. Likewise, Zárate (2021) evaluates the impact of a new metro line on labor market informality in Mexico City. However, neither of them take up the issue of sick-leave. As far as we are aware, our research is the only study of commuting time, transit investments and sick-leave in a developing country context.

Our empirical results indicate that longer commuting times to the workplace (highly correlated with distance) increase the probability of a worker taking some sick-leave in a given month. This result holds even when individual worker fixed effects and firm level controls are included in the model. We also find that the effect of travel time on sick-leave behavior is three times stronger for female workers compared to male workers. The effect is also higher for workers that have below median labor incomes or are relatively older.

An empirical concern is that job location choice could be endogenous and correlated with health status or other circumstances related to sick-leave behavior. To tackle this issue we use the approach proposed by Andersson et al. (2018). It consists of identifying workers who have suffered from a mass layoff in their firm and track their behavior in their next job. In essence, the assumption is that the endogeneity problem will be more severe for workers who do on-the-job search. A mass lay-off, on the other hand, is an exogenous shock to workers who are then forced to find a new job. This will reduce the problem of an endogenous job change due to health reasons or other individual circumstances.

The positive correlation between commuting travel time and sick-leave remains after applying the approach suggested by Andersson et al. (2018). However, there may still be a residual endogeneity issue if laid-off workers take the opportunity to find a new job with a location better suited to their particular circumstances. We argue further below that this potential endogeneity problem would most probably bias our results towards zero. Therefore, these results are conservative in the sense that they may underestimate the true effect of travel times on sick-leave behavior.

To estimate the economic importance of the relationship between commuting times and sick-leave, we simulate a 20% fall in average commuting travel-time (around 4.9 min per trip). This would reduce medical related work absence and generate an economic benefit of close to 36 million dollars per year.

With respect to infrastructure investments, results indicate that metro and commuter rail extensions during our sample period reduced sick-leave for those workers who potentially benefited from these projects. This suggests that there are additional benefits from these transit investment beyond conventional travel time savings.

To date few studies have analyzed the relationship between commuting time and absenteeism. Our literature review indicates that only one study looks into this issue directly (Van Ommeren and Gutiérrez-i Puigarnau, 2011). Our paper then contributes to bridge this gap in the literature and hopefully will motivate more research in this area.

The paper is organized as follows. The next section presents a literature review of the relation between absenteeism and commuting time or distance. We explain the main features of our paper and how it fits into this literature. Following that, we present the data used, how variables were constructed and a descriptive analysis of sick-leave absenteeism for Santiago. We then present the model and estimation results, followed by an analysis of the impact of metro and rail extensions. We then present a discussion of the policy implications of our findings in the context of the 'wider economic benefits' of transport investments. The paper ends with a summary of our main conclusions and a discussion of potential avenues for future research.

2. Literature review

As mentioned above, there are two separate strands in the literature on sick-leave and absenteeism. The first emphasizes the direct or indirect effects of commuting-time and mode on health outcomes. This literature emphasizes the potential effect of longer commuting times on physical activity, sleeping time, stress, and objective and subjective health outcomes (Christian, 2012; Hoehner et al., 2012; Hansson et al., 2011); and references contained therein). These early studies find a negative relationship between commuting time (or distance) and sleeping time, physical activity, time spent on food preparation, stress and health outcomes. Hansson et al. (2011), using Swedish data, find that longer commuting time is associated with higher sickness absence, confirming results from previous studies in Europe (Costa et al., 1988). However, their results are sensitive to the length of the sickness absence; commuting time seems to affect short-period absenteeism, but not long-term absenteeism that may be more related to actual disease.

All of the above studies use cross-sectional data and thus suffer from the potential problems of individual idiosyncratic effects that may be correlated with commuting times and health outcomes. One study that uses panel data is Künn-Nelen (2016). She uses data from the British Household Panel Survey (BHPS) from 1991 to 2008 to study the relationship between commuting time and objective and subjective health outcomes. The results indicate that longer commuting times are associated with lower subjective (self-evaluated) health status but not to objective health outcomes. More to the point of the current paper, Künn-Nelen (2016) does not find a statistically significant relationship between commuting time and sickness absence or in-patient hospital stay.[4] Therefore, the results using panel-data, at least in the case of Britain, does not support the results from earlier studies using cross-sectional data.

In sum, the involuntary approach to absenteeism is somewhat inconclusive as to the effects of commuting time or distance on absenteeism. However, this research, although still very limited in scope and size, does seem to imply that the possible effect would be on short-term absenteeism, not on objective health outcomes, long-term sickness or

[3] Therefore, we expect female workers to have a higher rate of sick-leave compared to male workers, something confirmed by the data as discussed further below.

[4] Interestingly, she finds that longer commuting time is associated with a higher number of visits to a general practitioner. Also, if part-time workers are included in the sample then a statistically significant relationship is found between commuting time and absenteeism.

the use of health services (hospital stay). These conclusions suggest that the second strand in the literature, the endogenous or voluntary approach to absenteeism, may have some merit.

The endogenous absenteeism approach is studied by Ross and Zenou (2008). They develop a standard theoretical urban economic model including commuting distance, land market and wages. Workers endogenously decide where to live and all work in the Central Business District (CBD). Workers' effort determines productivity (shirking) in their model. Individuals that spend higher amounts of time commuting have less time for leisure, affecting the trade-off between leisure (shirking) and effort at work. If firms can observe workers' residential location and can condition wages accordingly, then an efficiency wage model implies that wages will be higher for those with longer commutes. In this scenario Ross and Zenou (2008) show that wage discrimination based on commuting time will avoid shirking in equilibrium. However, if firms cannot observe workers' residential location or for legal reasons cannot wage discriminate based on commuting time, then in equilibrium one would observe higher levels of shirking and higher unemployment spells for those workers that reside further away from their jobs.

Ross and Zenou (2008) test the predictions of their model using census data for the year 2000 in the United States.[5] They do not test shirking directly but do find that higher commuting times are associated with higher wages and higher levels of unemployment for heavily supervised occupations (presumably those occupations where shirking has a large impact on productivity). Therefore, in these occupations, the results would seem to suggest that firms do not eliminate all shirking through higher wages. For lightly supervised occupations, there is no robust statistical relationship between commuting time and either employment or wages.

Van Ommeren and Gutiérrez-i Puigarnau (2011) use the German Socio-Economic Panel (GSOEP) data from 1999 to 2008 to study the predictions of the Ross and Zenou (2008) model. They equate shirking with absenteeism arguing that they are closely related (Barmby et al., 1994). The panel nature of their data allows them to use individual fixed effects, residence fixed effects and job fixed effects. In order to identify the impact of commuting time on absenteeism, they use employer induced changes in commuting distance (relocation of workplace maintaining residence and job) on changes in absenteeism behavior. They find a positive relationship between commuting distance and absenteeism. Workers that commute 50 kilometers on average would have 15% more days absent compared to a worker with a 10 kilometer commute. They also argue that their results are probably predominantly due to voluntary absenteeism given that they are not sensitive to the inclusion of workers' health indicators.

Our paper builds on these earlier contributions using panel data. However, our data is novel in several respects. First, unlike Künn-Nelen (2016) and Van Ommeren and Gutiérrez-i Puigarnau (2011) we have monthly data rather than yearly data. Second, it is a random sample of 20% formal sector workers as of March 2019, generating a large dataset that reduces sampling bias. Third, previous studies use a measure of sick-leave that is self-reported in survey interviews. For example, the measure of absenteeism used by Künn-Nelen (2016) is whether an individual reports to have called in sick at least one-day during the last year. Van Ommeren and Gutiérrez-i Puigarnau (2011) define absenteeism as the number of days reported absent during the year before the interview. In both cases, they are recall questions over a long period of time (one year) and may be subject to measurement error. In contrast, we use administrative information to determine whether a worker had a sick-leave episode in a given month. Fourth, unlike previous studies we have information on the location of the employer. This allows us to calculate commuting times in a more objective way compared to self-reported answers used in the papers reviewed above (however, unlike previous research we are not able to identify travel mode in our data).

3. Data

Our study uses administrative records of the labor histories of workers participating in the Chilean unemployment benefit program (UI). These records are maintained by the *Superintendencia de Pensiones*, the agency in charge of the regulation and supervision of the UI scheme. We use a publicly available random sample of labor histories for 20% of workers drawn from the administrative records of March 2019.[6]

The UI system became operational in October 2002. It covers all individuals over 18 years old who are employed in the private sector and have a formal contract, either fixed term or indefinite. Workers excluded from UI are employees in domestic services (house-keepers)[7], public sector employees (regulated by the Public Service Act), the self-employed, workers under 18 years of age, and those hired as trainees and retirees. Contribution to the UI system is mandatory since October 2002 for all new private sector contracts. Permanent workers hired before October 2002 who have not switched jobs can opt-in to the scheme. Since this is voluntary, there is a fraction of permanent workers that are not affiliated to the system. This fraction is negligible by the end of our sample period.

In October 2010, there were 3,446,000 contributors in the UI system. Using aggregate employment survey information from the national statistical office (INE) we can estimate that 79% of workers that satisfied the enrollment conditions at that date were effectively registered in the system. For the remaining 21%, informality accounted for 14% and 7% were formal permanent workers who where not obliged to enroll and had not done so voluntarily. By February 2017, there were 4,364,000 contributors in the system, or 87% of potential users. Informality accounted for 11% of those not in the UI by this date, and only 2% were formal permanent worker that were still not enrolled. By the end of our sample, March 2019, these percentages were even smaller.

Workers can legally take sick-leave if they have a medical license. The system works as follows. Individuals who are sick must see a doctor who may then issue a medical license if the worker is incapacitated to go to work. For illnesses or accidents that require more than ten days absence from work, the scheme provides workers with an insurance subsidy to cover all the lost labor earnings. If the illness or accident requires 10 days or less of absence, the insurance scheme does not cover lost earnings from the first 3 days. Therefore, a worker does not receive a full income coverage for less extended illnesses. As mentioned above, medical licenses are also granted for maternal leave, serious illness of a child under one year of age or for a pathology during pregnancy. Therefore, we expect female workers to have a higher rate of sick-leave, particularly during fertile age. Unfortunately, with our data we cannot identify the underlying cause of the medical license, since we only observe whether an individual had a license in a given month and the amount of the insurance wage subsidy received. We use the first of these last two variables as our indicator of sick-leave behavior.

[5] They use an instrumental variable approach based on variation across metropolitan areas in certain physical attributes that exogenously determine commuting times in those localities.

[6] The data and accompanying documentation was downloaded from https://www.spensiones.cl/apps/bdp/index.php.

[7] These workers have a special unemployment insurance scheme.

We focus on Santiago, the Chilean capital.[8] Santiago has nearly 6 million inhabitants and 42% of total contributors to the UI System in 2013. Besides information on having a medical licenses and the wage subsidy received each month, we have data on labor earnings, gender, age, education, marital status, and municipality of residence of each employee, as well as the economic sector, size (in terms of number of workers), average wage bill and the municipality location of their employer.

As just noted, we have the municipality of a worker's residence and the municipality of his or her employer, for each worker-month. A first problem with the locational data is that only one residential location per worker is registered, thus changes in these locations are not observed in the database.[9] Consultation with the *Superintendencia de Pensiones* revealed that the residential address is updated each time a worker applies for unemployment insurance or retires. Thus, for workers who had recently received some unemployment benefit, their residential location should be up to date. However, for those that have not been unemployed, their registered residential location may be outdated if the worker changed homes during the sample period.[10] Further below we explain how we test the robustness of the results to this possible measurement error.[11]

A second problem is that our commuting time variable (whose construction we explain below) is based on municipal location, which may be an imprecise estimate of workers' true commuting travel times. Furthermore, we do not observe the travel mode of each worker. On the other hand, there are 44 municipalities in the sample and thus they give a rough approximation to travel times.

Commuting travel times were obtained using the 'georoute' STATA command, which uses the HERE application programming interface (API) to retrieve this information.[12][13] To obtain the value for each municipality origin–destination pair, we first calculated travel-times for a matrix of 718 zones defined in the 2012 Santiago mobility survey (also known as the 2012 Origin–Destination survey). We then aggregated these zones by taking the simple average of travel times for zones within each municipality. Initial estimates using these last results show that travel times calculated using this approach are strongly correlated with travel times reported in the 2012 Mobility survey, both for transit as well as private car travel.[14] This provides some evidence that our travel time measure is a good approximation to actual travel times.

To study the effects of major mobility investments in the city, we also used information on metro and commuter rail expansion. The Santiago Metro had a network extension of 67.4 kilometers in 2005, prior to our sample period. During our sample period it expanded to 140 kilometers with the inauguration of a final segment to Line 4 (March 2006), the opening of Line 4 A (August 2006), the extension of L1 (December 2009), the first stage of Line 5 (January 2010), the full operation of L5 (January 2011), the opening of Line 6 (November 2017) and Line 3 (January 2019). In addition, a 20 kilometer commuter rail service was inaugurated in March 2017, further extending the mass transit system in Santiago.[15]

The metro and rail extensions were the only important mobility investments in the city during the period of our data. All the major high-speed urban toll roads were already in operation at the start of our sample period.[16] On a much smaller scale were the inauguration of bus priority infrastructure in some avenues. However, these were limited in scale relative to the aggregate bus service kilometers.[17]

For the metro and rail extensions we define two variables. First, for each municipality origin–destination (OD) pair we calculated an indicator variable whether such an OD pair could be traveled using metro or rail with at most one transfer between different lines. Second, we multiplied the previous variable by the average flow of passengers through stations in the origin municipality in the morning peak and the flow of passengers through stations in the destination municipality in the evening peak, all divided by the population of the origin municipality. In other words, if we define $Metro^{ij}$ as the indicator variable that takes the value of 1 if the origin–destination pair ij can be reached by metro or rail with at most one transfer between lines, and f^{i}_{AM} as the flow of passengers through stations in the origin municipality i in the morning peak period, and f^{j}_{PM} as the flow of passengers through the stations in the destination municipality j in the evening peak period, then the second variable we use is defined as $Metro^{ij}_{pob} = Metro^{ij} \cdot \left(\frac{f^{i}_{AM}+f^{j}_{PM}}{2 \cdot Pob^{i}}\right)$, where Pob^{i} is the population of municipality i. This last variable measures accessibility to the metro and commuter rail services but weighted by the importance of ridership to the population of the origin municipality.

4. Descriptive statistics

Table 1 shows descriptive statistics for our sample. There are over 39 million observations from 585,600 individual workers' labor histories. The average age of workers is 37.9 years, 61% are male, and 75% have a permanent (indefinite) labor contract. Average travel-time is 24.0 min. This is lower than the 30.4 min average travel times in private transport registered in the 2012 Santiago mobility survey. However, the average from this survey includes all trips, not just labor related trips.

Fig. 1 shows the share of workers taking some sick-leave in a given month by contract type.[18] Permanent workers are those with an indefinite contract while temporary workers are those with a fixed term contract. The proportion of workers taking sick-leave fluctuates

[8] Specifically, we use the data where the residential location of workers and firms are in one of 44 municipalities of the Santiago Metropolitan Region. The Province of Santiago has 32 municipalities, while the whole region has 52. Our data base includes information from the 32 municipalities of the Santiago Province, 2 municipalities that, while not in the Province, are an integral part of the city (the municipalities of San Bernardo and Puente Alto) plus 10 nearby sub-urban municipalities where commuting travel times are available from the 2012 Santiago mobility survey.

[9] We would expect employer location to be more time invariant.

[10] As discussed futher below, this would bias the absolute value of the coefficient of the commuting time variable downwards.

[11] Basically, the strategy is to limit the sample to those workers that applied for unemployment benefit in the recent past.

[12] See https://developer.here.com.

[13] We also calculated travel distance. However, commuting travel time has a correlation of 0.95 with commuting distance so we only discuss travel times in what follows. Also, the correlation coefficient of commuting travel time with Euclidean distance is 0.93.

[14] Results of these correlations are available upon request. The reason we did not use the travel times recorded in the Origin–Destination study is that not all travel patterns (i.e. municipality to municipality pairs) from our sample are contained in this mobility survey.

[15] This train service is integrated to the Santiago public transport system, that includes metro and buses.

[16] The exception being a 1.9 kilometer tunnel, with 2.2 kilometers of highway (San Cristobal Tunnel concession) inaugurated in mid 2008, and a highway leaving the city through a northern suburb (Radial Nor-Oriente) between 2008 and 2009.

[17] What we are not explicitly considering is the Transantiago reform in February 2007 that introduced an integrated fare between metro and buses. This reduced the cost of multimodal trips using the metro and passengers in this last mode more than doubled as a consequence.

[18] There is sharp fall in the proportion of sick-leave in September 2007. This may be related to the major transit reform implemented in Santiago in February 2007 (Transantiago). This reform was initially problematic and created a mobility crisis in the city that may account for some of the rise in sick-leave behavior during the first two quarters of 2007. Eventually, higher transit supply and the fare integration between buses and metro may have reduced travel times generating the fall in sick-leave behavior towards the end of that year. In any case, the results reported below are unchanged if the first two years are excluded from the estimations.

Table 1
Descriptive statistics, Santiago, monthly data from January 2006 to March 2019.

Variable	Obs.	Average	s.d.	Min	Max
Medical license (binary)	39,448,422	0.046	0.21	0	1
Contract type (2 Temporary)	39,448,422	1.25	0.43	1	2
Age (Year)	39,448,422	37.9	11.2	18	70
Gender (1 Male)	39,448,422	0.61	0.49	0	1
Tenure (Months)	39,448,422	29.8	31.9	1	198
Travel time (minutes)	39,448,422	24.0	10.8	4.76	92.7

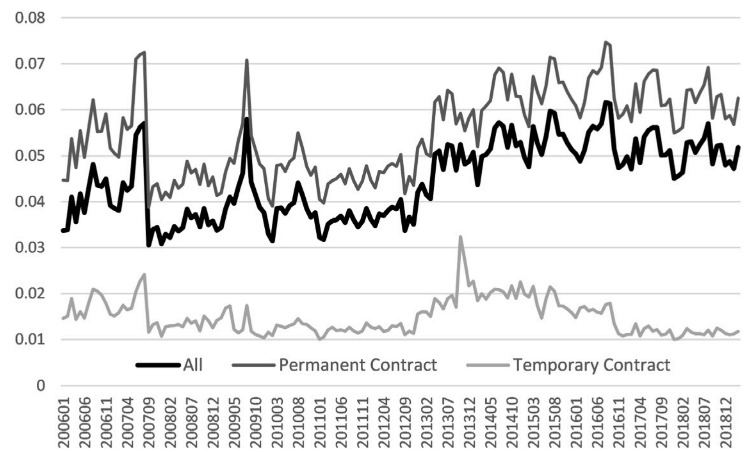

Fig. 1. Share of workers that take some sick-leave in a given month, data from January 2006 to March 2019.

Table 2
Medical licenses, number of workers and share by type of contract and gender, average January 2006 to March 2019.

Contract	All		Permanent				Temporary			
			Female		Male		Female		Male	
Month	Med. lic.	Number workers	Med. lic.	Share workers	Med. lic.	Share workers	Med. lic.	Share workers	Med. lic.	Share workers
Feb.	4.0%	235,909	7.9%	29.9%	2.9%	43.7%	2.5%	8.5%	0.9%	17.9%
May	4.6%	256,795	9.1%	30.2%	3.3%	43.8%	2.6%	8.5%	0.9%	17.6%
Aug.	5.2%	260,164	10.1%	30.3%	3.8%	43.9%	3.0%	8.3%	1.1%	17.4%
Nov.	4.4%	263,947	8.7%	30.4%	3.2%	43.8%	2.7%	8.5%	0.9%	17.3%

between 3.1% and 6.2% at any point in time. It is interesting to note that the rate of sick-leave for workers with indefinite contracts is four to five times higher than that among workers with temporary contracts.

There seems to be a structural change starting in 2012, with an increase in the rate of sick-leave after that date. From January 2013 onward, the proportion of workers taking some sick-leave in a given month never drops below 4.1%.[19] This change is related to workers with permanent contracts which account for the largest share of the employed workforce and is probably explained by the extension of maternity leave from 12 to 24 weeks that came into effect at the end of 2011.

There is also some seasonality in sick-leave behavior, with spikes in August, the mid winter month, particularly marked during the first few years of the sample. The lowest proportion is evidenced during the summer months of January and February, rising until reaching a peak in the middle to end of winter.[20]

Table 2 present some other interesting stylized facts. Although all groups of workers increase the proportion of sick-leave during the winter, the likelihood among female workers is higher than for men, as expected.[21] However, there are differences by type of contract. For both genders, the likelihood of taking sick-leave is much lower among workers with temporary contracts, although there is still a gap between male and female workers.

We now turn to the relation between sick-leave and commuting travel times. Fig. 2 shows a positive correlation between the rate of sick-leave and commuting travel time (natural logarithm of minutes).

5. Estimation approach

To study the effects of commuting travel time on sick-leave behavior we estimate a linear probability model. Given its linear structure, this model is easy to estimate in a panel data setting.

The linear probability model may generate predicted probabilities that lie outside the unit interval. If the relative proportion of these abnormal predicted probabilities is large, it may signal that coefficient estimates are biased and inconsistent (Horrace and Oaxaca, 2006). Therefore, as a robustness check in Appendix A we also present results using a probit panel data model estimated with the algorithm proposed by Stammann et al. (2016). The results are very similar to those presented further below and thus we can be fairly confident that the

[19] The average share is 3.9% between January 2006 and December 2012, rising to 5.2% between January 2013 and March 2019.

[20] According to Fonasa, Superintendencia de Seguridad Social y Superintendencia de Salud (2017), respiratory diseases represent around 15% of total medical licenses in Chile. These become more prevalent during the winter months.

[21] Pezoa (2010) indicates that medical licenses related to maternity leave and diseases of children under one year explain 26.8% of total medical license expenditure in 2009. Pezoa (2010) also shows that women are twice as likely to take sick-leave related to all other pathologies.

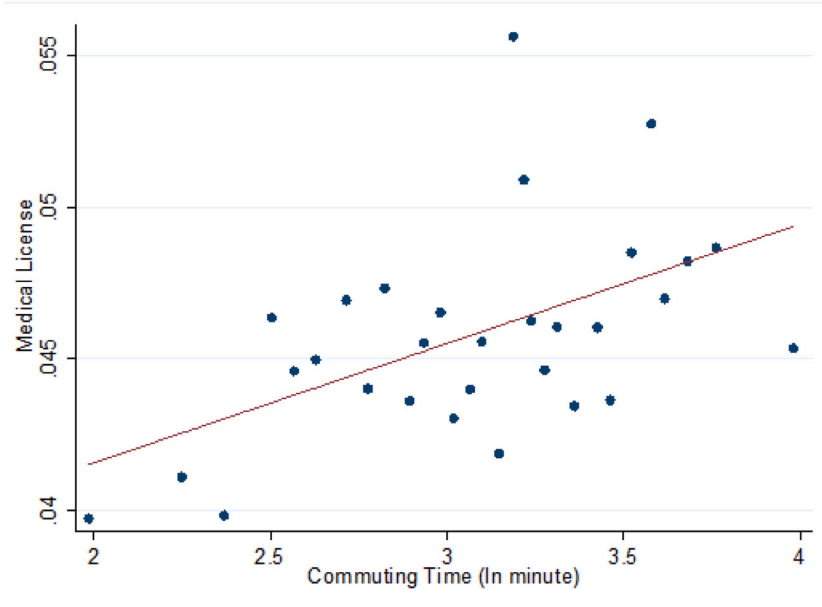

Fig. 2. Sick-leave and commuting travel times (ln min).

linear probability model does not generate biased results in the context of our study.[22]

One important issue is how to control for possible endogeneity problems. A worker who has certain unobservable characteristics may choose his residential location or employer based on these characteristics. To the extent that these unobservable characteristics also affect the probability of taking sick-leave, they may generate a classical endogeneity problem through the travel time variable.

We tackle this potential endogeneity issue through several approaches. First, by including workers' fixed effects we are controlling for time-invariant observable and unobservable characteristics. Since we only have one residential address for each worker, this amounts to identifying the impact of travel time on workers' sick-leave behavior through workers who switch jobs.[23]

However, it could be argued that individuals may change jobs because they expect to take more sick-leave in the future (due to changing health conditions or after a child is born) and so including workers' fixed effects may not control for all the potential endogeneity of the travel time variable.

A second approach then is to use the idea of Andersson et al. (2018) who estimate a model of labor market outcomes using those workers who suffer a mass layoff in their firm. We expect that a mass layoff is an exogenous shock not correlated with workers' individual characteristics. We identify workers who have suffered a mass layoff as those who become unemployed after their firm lays-off 50% or more of their workforce. We then follow these workers in their next job.[24] Analyzing the sick-leave behavior of these workers as travel times change with their new job should then be less prone to endogeneity problems since the decision to leave the previous job was exogenous and not related to the individual's particular circumstances.

The approach of Andersson et al. (2018), however, may not solve all the endogeneity issues since there might be a residual effect operating through a laid off worker's new employment choice. Although the job change is exogenous, a worker might take the opportunity to find a new employment location more suitable to his/her particular circumstances, including possible health status and sick-leave behavior.[25] Although we cannot control for this residual endogeneity effect, it is reasonable to assume that sickly workers would prefer a new job closer to their residential location. If this is the case, our estimated commuting time coefficient would be biased towards zero. Therefore, if this residual endogeneity effect is present, our results would be a conservative estimate of the effects of commuting time on sick-leave behavior.

An additional problem that was mentioned above is that workers who have not applied for unemployment insurance may have an outdated residential location registered in the database. Therefore, in some estimations we restrict the sample to workers who have applied for unemployment insurance in the recent past (last two years) and therefore have an up-dated residential address.

6. Results

Table 3 presents the first set of results. To have an initial benchmark, the specifications in all columns of this table were estimated without workers' fixed effects. They do include a seasonal calendar month effect plus workers' educational level fixed effects.[26] The estimated model in column (2) to (5) also include firm level controls. To avoid too many leading zeros in the estimated coefficients, the dependent variable was multiplied by 100. Therefore, estimated coefficients should be read as percentage points.[27]

[22] The potential pitfalls of the linear probability model vis a vis a non-linear binary response model have been greatly exaggerated. Binary response models require strong assumption on the error term and they often give estimated partial effects that are very similar to those using the linear probability model. See the discussion in Angrist and Pischke (2008), Chapter 3, or Wooldridge (2010), Chapter 15.

[23] It could also be that the employer changes location, but this is much less likely.

[24] Mass layoff for a worker is defined as follows. Define t as the month that a worker becomes unemployed. This event is due to a mass layoff if the worker's firm reduced its total employment by 50% between $t-2$ and t, $t-1$ and t, or t and $t+2$.

[25] However, a laid-off worker may have more pressure to accept any available job offer and be less picky regarding the new job location compared to a worker who undertakes on-the-job search and who can remain in their current employment if they do not find a better located alternative.

[26] There are four categories of educational attainment.

[27] Standard errors were clustered at the municipal origin–destination pair. Spatial correlation of errors among neighboring municipalities may also be

Table 3
Linear probability model of observing a worker taking sick-leave in a given month.

Sample	(1) All	(2) All	(3) Mass layoff	(4) Unemployed recently	(5) OD in different municipalities
Unemployment rate	0.0147	0.0128	−0.0322	−0.105***	0.0109
	(0.0110)	(0.0106)	(0.0203)	(0.0122)	(0.0116)
Gender	−4.874***	−4.930***	−4.593***	−3.487***	−5.067***
	(0.0870)	(0.0853)	(0.0985)	(0.0743)	(0.0876)
Contract type	−2.192***	−2.100***	−2.262***	−3.074***	−2.133***
	(0.0345)	(0.0342)	(0.0465)	(0.0450)	(0.0343)
ln(tenure)	1.389***	1.395***	1.012***	1.502***	1.424***
	(0.0192)	(0.0181)	(0.0216)	(0.0266)	(0.0187)
ln(age)	20.98***	20.78***	7.171***	1.349	21.35***
	(1.040)	(0.993)	(1.656)	(0.962)	(1.018)
$\ln(\text{age})^2$	−3.281***	−3.254***	−1.312***	−0.378***	−3.333***
	(0.151)	(0.144)	(0.233)	(0.136)	(0.148)
ln(average firm wage)		−0.432***	−0.129***	−0.410***	−0.494***
		(0.0414)	(0.0497)	(0.0403)	(0.0382)
Maternity reform	1.087***	1.186***	1.138***	0.893***	1.229***
	(0.0336)	(0.0332)	(0.0435)	(0.0317)	(0.0339)
Maternity reform × Female	0.0105	0.0200	0.456***	−0.0930	−0.0151
	(0.0596)	(0.0588)	(0.0996)	(0.0702)	(0.0611)
ln(travel time)	0.720***	0.753***	0.687***	0.401***	0.623***
	(0.0808)	(0.0687)	(0.0725)	(0.0524)	(0.0758)
Constant	−25.94***	−22.92***	−1.909	12.00***	−22.64***
	(1.611)	(1.558)	(2.867)	(1.712)	(1.636)
Month seasonal controls	Yes	Yes	Yes	Yes	Yes
Workers education controls	Yes	Yes	Yes	Yes	Yes
Firm location controls	No	Yes	Yes	Yes	Yes
Firm economic sector controls	No	Yes	Yes	Yes	Yes
Observations	32,556,397	32,556,397	7,083,229	9,392,695	28,597,296
R^2	0.028	0.029	0.027	0.032	0.030

Notes: Clustered (by origin–destination municipal pairs) standard errors in parenthesis. Dependent variable multiplied by 100 to reduce number of leading zeros of estimated coefficients. Thus, coefficients represent changes in percentage points. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}\ p < 0.01$.

The first column of Table 3 was estimated without any firm level controls as a starting point. Our main variable of interest is commuting travel time (expressed as the logarithm of minutes), labeled as *ln(travel time)*. It can be seen that this variable has a statistically positive relationship with the probability of a worker taking sick-leave in a given month. This semi-elasticity implies that a 1% increase in travel time increases the probability of sick-leave by 0.72 percentage points.

As for the other control variables it can be seen that the national unemployment rate is not significant. The gender variable has the expected sign with men being less likely to take sick-leave (males are coded 1 and females 0). Contract type has the effect showed above when describing the data; workers with fixed term contracts are less prone to take sick-leave compared to those with an indefinite labor contract. Tenure and age increase the probability of sick-leave, although in the case of age, the relationship is non-linear with an inverted U shape. In order to control for the additional maternity leave licenses after the reform in October 2012, we include a discrete variable (*Maternity reform*) taking a value of one after this date. We can see there is a statistically significant positive effect from this reform that extended maternity leave from 12 to 24 weeks. When interacted with a female dummy (1 minus the gender dummy), we can see that this impact is in general higher for women, as expected. Except for the *Unemployment rate* and the interaction of *Maternity reform* with the *Female* dummy in some specifications, all other coefficients are statistically significant at a 1% confidence level.

Column (2) to (5) introduce employer level controls, including locational (44 municipality) and economic sector (22 sectors) fixed effects as well as the average wage paid by each firm (highly correlated with firm size). The results are very similar to those of column (1) with the coefficient on commuting travel time increasing slightly. It is interesting to note that the coefficient on average firm wage is negative, perhaps because firms that pay better wages hire workers with more unobservable characteristics correlated with health or higher wages imply lower shirking as in the efficiency wage model discussed in the literature review above.

Column (3) shows the results of restricting the sample to workers who suffered a mass layoff in their previous job and are observed only in their following job. The sample size decreases substantially in this restricted sample. We can see that the travel time variable is still highly significant. The coefficient of the travel time variable is somewhat smaller than in the previous two columns but this difference is not statistically significant.[28]

Column (4) restricts the sample to those workers who applied for unemployment insurance in the past 24 months and so their residential address has been recently updated. The sample is also smaller than those in column (1) to (2), albeit larger than the mass layoff sample. The results are qualitatively similar to those estimated in the previous models. However, the coefficient of the travel time variable is lower compared to those estimated in column (1) to (3) but still highly significant. Therefore, we do not find evidence that restricting the sample to workers who have applied for unemployment insurance during the last two years changes the qualitative results.

Finally, column (5) estimates the model of column (2) but excluding workers who reside and work in the same municipality. This was undertaken to test the sensitivity of results to possible differences in measurement errors between short and long commuting distances. The

relevant. However, given the statistical significance of most coefficients it is doubtful whether this would change our inference results.

[28] Since we have a sample of 20% of workers, on average we only identify 20% of workers of each firm. Therefore, for small firms with less than 50 workers, we only observe on average less than 10 of its workers in the sample. If a small firm lays-off one or two workers, we are unsure whether this is a mass layoff or not. If we estimate the model of column (3) but restricting the sample to firms with 50 or more workers, the coefficient of the travel time variable is still positive and statistically significant.

Table 4
Linear probability model of observing a worker taking sick-leave in a given month including worker fixed effects.

Sample	(1) All	(2) Female	(3) Male	(4) All	(5) All
Unemployment rate	0.0409***	0.130***	−0.0183***	0.0271***	
	(0.00879)	(0.0205)	(0.00657)	(0.00885)	
Contract type	−2.171***	−2.443***	−1.818***	−1.992***	−1.980***
	(0.0301)	(0.0517)	(0.0266)	(0.0337)	(0.0334)
ln(tenure)	1.768***	3.236***	0.892***	1.895***	1.882***
	(0.0230)	(0.0444)	(0.0161)	(0.0269)	(0.0267)
ln(age)	38.54***	121.3***	−5.215***	30.55***	183.1***
	(1.559)	(2.735)	(1.149)	(1.821)	(5.932)
$\ln(\text{age})^2$	−5.631***	−17.68***	0.580***	−3.729***	−34.17***
	(0.224)	(0.391)	(0.171)	(0.275)	(1.195)
ln(average firm wage)	−0.629***	−1.260***	−0.189***	−3.037***	−3.098***
	(0.0315)	(0.0625)	(0.0202)	(0.0637)	(0.0644)
Maternity reform	1.166***	1.853***	1.360***	1.050***	
	(0.0346)	(0.0602)	(0.0260)	(0.0354)	
Maternity reform × Female	0.996***			1.242***	1.301***
	(0.0506)			(0.0503)	(0.0493)
ln(travel time)	0.469***	0.800***	0.252***	0.178***	0.176***
	(0.0410)	(0.0882)	(0.0294)	(0.0339)	(0.0336)
Constant	−57.16***	−192.7***	14.89***	−21.10***	−172.3***
	(2.615)	(4.783)	(1.890)	(3.309)	(5.850)
Month seasonal controls	Yes	Yes	Yes	Yes	No
Worker fixed effects	Yes	Yes	Yes	Yes	Yes
Firm location controls	Yes	Yes	Yes	No	No
Firm economic sector controls	Yes	Yes	Yes	No	No
Firm fixed effects	No	No	No	Yes	Yes
Month-year effects	No	No	No	No	Yes
Observations	39,448,422	15,324,505	24,123,917	39,438,555	39,438,555
R^2	0.131	0.139	0.094	0.145	0.146
# Changes	2.036e+06	660564	1.375e+06	2.026e+06	2.026e+06

Notes: Clustered (by origin–destination municipal pairs) standard errors in parenthesis. Dependent variable multiplied by 100 to reduce number of leading zeros of estimated coefficients. Thus, they represent changes in percentage points. $^{*}p < 0.10$; $^{**}p < 0.05$; *** $p < 0.01$.

estimated coefficient on travel time is slightly lower than that estimated with the full sample shown in column (2) but the qualitative results are the same.

Table A.1 of Appendix A presents the marginal effects estimated using a non-linear probit panel data model applied to specification (1) to (4) of Table 3. The marginal effects are very similar in magnitude to the coefficients of this last table and highly significant.[29] This last result implies that in the current application, the non-linearity of a true binary response model does not have a large effect on the marginal impact of the commuting time variable compared to the linear model. Therefore, there does not seem to be a bias from using the linear probability model in our context, a result that is quite common in applied work according to Angrist and Pischke (2008).

Table 4 presents the results including worker fixed effects.[30] Specifications (1) to (3) include a calendar month seasonal control, firm location controls (municipality) and firm economic sector controls. Model (1) uses the whole sample, while model (2) and (3) restrict the sample to female and male workers, respectively.

The results of column (1) indicate that including worker fixed effects reduces the impact of travel time on the occurrence of sick-leave. According to these results a 1% increase in travel time increases the probability of sick-leave by 0.47 percentage points. Column (2) and (3) show that the effect is much stronger for females (0.80) than for males (0.25).[31]

Column (4) estimates the model with the full sample but including firm specific fixed effects in addition to worker fixed effects (and monthly seasonality controls). In this double fixed effects estimation, the coefficient on travel time is still significant, although less than half in value compared to the results reported in column (1). If we add month-year fixed effects to this last model (column (5)) the results are unchanged.

The models estimates in the two previous tables include only one coefficient for the travel time variable. However, the effect may well be non-linear. To examine this issue, model (1) of Table 4 was estimated using categories of travel times instead of this variable in levels. To this end, observations were grouped into travel time deciles and a dummy variables was defined for each one.[32] These dummies were used in the estimation of the model.

The results are presented in Fig. 3 where the mean log travel time of each decile is graphed in the horizontal axis. It can be seen that the parameter estimate for each decile rises almost linearly with the logarithm of travel time. This would indicate that the semi-elasticity of travel time on sick-leave behavior is approximately constant.

From Table 4 we know that the effect of travel time on sick-leave differs by gender; females are more sensitive to travel time compared to males. We now explore whether there are heterogeneous impacts along other dimensions. Table 5 presents results from estimating the model of Column (1) of Table 4 but interacting the travel time variable with an indicator of low wage, age and contract type, respectively.

The first column of Table 5 interacts the time variable with a dummy variable (D_{lw}) denoting for each observation whether monthly labor earnings are below the median earnings in each month of the sample. It can be seen that the interaction parameter is positive and statistically significant. This implies that workers with relatively low wages

[29] The probit model standard errors are clustered at the worker level so are not directly comparable to those of Table 3.

[30] With individual fixed effects the coefficients on the travel time variable is identified by workers' employment changes that imply a change in commuting time. At the bottom of this table we report the number of observations with a transition (change in employment).

[31] Table A.2 of Appendix A shows results of estimating models (1) to (3) of Table 4 using the probit panel data model. As in the previous case, the average partial effects are similar to those estimated using the linear probability model.

[32] The mean and upper and lower limits of these deciles are presented in Table A.3 of Appendix B.

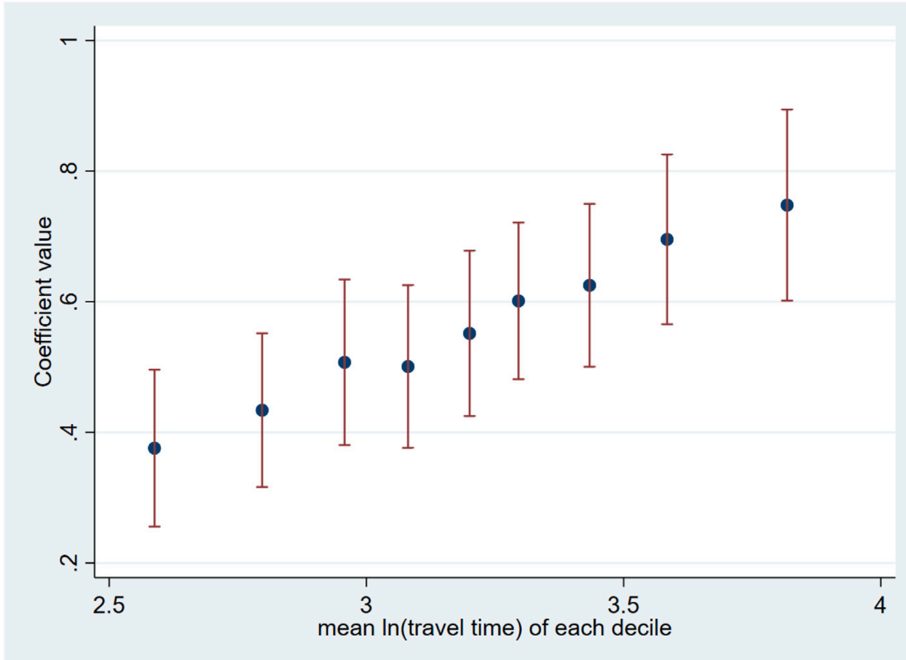

Fig. 3. Sick-leave and commuting travel time deciles.

are the ones most affected by commuting times. This could reflect that lower income workers use more public transportation and thus have more interaction with other people, increasing the probability of contracting infectious diseases.

Column (2) interacts the travel time variable with a dummy variable (H_{age}) indicating whether the worker is above the median age of the sample (36 years). The interaction coefficient is positive and statistically significant, so relatively older workers have a higher propensity to take sick-leave as travel times increase compared to younger workers. This may reflect that relatively older workers are more vulnerable to sickness.

The last column interacts the travel time variable with the contract type dummy. It shows that having a fixed term or temporary contract reduces the effect of travel time on sick-leave behavior. A possible explanation for this last result is that temporary workers may be more conservative with respect to sick-leave given their more fragile contractual relationship with their firm.

In summary, we can conclude that women, lower wage earners, older workers and workers with an indefinite contract are more likely to take sick-leave as commuting travel times increase.

Next we analyze the impact of metro and rail expansions. Table 6 presents the results. Column (1) includes a dummy variable indicating whether the municipality of residence and the municipality of employer's location are connected by a direct metro or rail line or by making only one transfer between different lines. We can see that the impact is a reduction in the probability of sick-leave.[33] Notice that in this model, we do not include the travel time variable since fixed effects for each origin–destination pair are included in the regression (in addition to month seasonality effects, workers' educational level effects and firms' economic sector effects). Also, since origin–destination fixed effects are included, the identification of the metro and rail expansion variables is due to changes in this variable through time for a given origin–destination pair.

Column (2) includes an alternative metro variable, defined as the original metro variable multiplied by the ratio of the average flow of passengers in stations in the origin and destination municipalities (the first for the morning peak and the second for the evening peak) divided by the population of the origin municipality.[34] This variable is an access variable but weighted by the importance of metro or rail ridership in the origin municipality. Once again, the coefficient is negative and statistically significant.

Column (3) is the same model of column (2) but including individual worker fixed effects. The coefficient of the metro variable is negative and statistically significant. Column (4) adds firm level fixed effects to the previous specification. In this double fixed effect specification, the coefficient of the metro variable is still negative and statistically significant.

In sum, using the results that include worker fixed effects we find evidence that the expansion of the Santiago Metro and commuter rail reduced the frequency of sick-leave by 0.47 to 0.54 percentage point. This is a reduction of close to 10% in the average number of sick-leave absences during the sample period.

Finally, we also explore possible heterogeneity in the impact of metro expansion on sick-leave by gender, age, contract type and low wage indicators. The results are presented in Table 7. All models include worker fixed effects, monthly seasonal controls, firm controls and fixed effects for each municipality origin–destination pair.

Column (1) interacts the gender variable with the $Metro_{pob}$ variable to test whether there is a difference in the impact of metro expansion by gender. We also add a dummy variable that takes a value of one if the origin–destination pair will have a metro connection in the future times the gender variable ($D_{Metro}\, x\, gender$). This is included to control for differences in the sick leave behavior by gender in those zones prior to the metro expansion.[35]

[33] Because metro expansion may be endogenous to municipal characteristics which may affect also residential choice, in some estimations we included a variable to indicate whether the origin–destination municipal pair would someday have a direct metro or rail link as an additional control variable. However, this variable was not statistically significant and did not change the results shown here.

[34] See the end of Section 3 for details.

[35] Since the model includes worker fixed effects the *gender* variable is not included in this regression.

Table 5
Linear probability model with heterogeneous effects.

Sample	(1) All	(2) All	(3) All
Unemployment rate	0.0398***	0.0393***	0.0410***
	(0.00881)	(0.00878)	(0.00879)
Contract type	−2.192***	−2.167***	−0.202
	(0.0304)	(0.0301)	(0.222)
ln(tenure)	1.790***	1.764***	1.768***
	(0.0237)	(0.0230)	(0.0230)
ln(age)	42.59***	34.48***	38.47***
	(1.523)	(1.541)	(1.559)
ln(age)^2	−6.270***	−4.894***	−5.620***
	(0.218)	(0.220)	(0.224)
ln(average firm wage)	−0.448***	−0.643***	−0.629***
	(0.0316)	(0.0316)	(0.0316)
Maternity reform	1.169***	1.136***	1.165***
	(0.0347)	(0.0346)	(0.0347)
Maternity reform × Female	0.988***	1.017***	0.999***
	(0.0506)	(0.0505)	(0.0506)
ln(travel time)	0.330***	0.347***	0.696***
	(0.0497)	(0.0474)	(0.0469)
ln(travel time) × D_{lw}	0.259***		
	(0.0474)		
D_{lw}	−0.165		
	(0.157)		
ln(travel time) × H_{age}		0.270***	
		(0.0599)	
H_{age}		−1.641***	
		(0.196)	
ln(travel time) × contract type			−0.637***
			(0.0708)
Constant	−65.65***	−51.12***	−59.70***
	(2.535)	(2.606)	(2.650)
Observations	39,448,422	39,448,422	39,448,422
R-squared	0.131	0.131	0.131

Notes: Clustered (by origin–destination municipal pairs) standard errors in parenthesis. Dependent variable multiplied by 100 to reduce number of leading zeros of estimated coefficients. Thus, they represent changes in percentage points. All specifications include worker fixed effects, firm controls and seasonal controls. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}\ p < 0.01$.

The results indicate that the coefficient on the interaction variable of metro expansion and gender ($Metro_{pob}\, x\, gender$) is negative. Since gender takes the value of one for males, this implies that male workers reduce their sick-leave behavior more than females after the infrastructure expansion.

Column (2) presents the results of interacting the metro expansion variable with contract type. Once again, we control for potential differences in the relation of contract type and sick leave behavior in those municipalities that will eventually have a metro connection. In this case, the coefficient on the interaction variable ($Metro_{pob}\, x\, contract\, type$) is not statistically significant, indicating that the impact of metro expansion does not differ over this dimension.

The same was done with the older age dummy, H_{old} (column (3)) and low wage dummy variable, D_{lw} (column (4)). In the first case, metro expansion reduces sick-leave behavior more for relatively older workers, although the coefficient is only marginally significant. In the second case, the coefficient on the interaction variable is not statistically significant.

In sum, we do not find much evidence of a heterogeneous impact of metro and rail expansion on sick-leave behavior, except for the case of male workers and marginally for relatively older workers.

7. Policy implications

The econometric results point to a significant effect of commuting travel time on the probability of a worker taking sick-leave in a given month. From Table 4 we estimate the average marginal effect with respect to the logarithm of travel time to be 0.469 overall, 0.252 for men and 0.800 for women.

To estimate the economic impact of the relationship between travel time and sick-leave, we simulate a 20% reduction in commuting time for all workers (about 4.8 min). At the end of 2018 there were 3,383,070 occupied workers in the municipalities of the Province of Santiago, earning on average US$1,288/month.[36] Using the average number of days per sick-leave license (20) and the average marginal effect of 0.469, the annual benefit to the economy would be close to 36 million dollars.[37]

It is also relevant to note that the above benefit arises irrespective of the explanation given to the link between travel times and sick-leave as discussed in the literature review. If this link is due to real health impacts, then this productivity benefit is direct. In this case, the economic costs should also include the cost of health services. But if part of the explanation is related to endogenous factors as in the shirking model discussed in Section 2, then the productivity benefit from a reduction in travel times would still be relevant. If mobility improvements reduce travel times, then shirking would also fall as workers do not need to take as much sick-leave in order to have more time for other activities.

The policy implications of our results are also important. First, they suggest that the benefits from reduced travel times are stronger for women (three times larger than for men) and for lower paid workers. Therefore, policies and programs aimed to improve mobility specifically for female workers will have a higher proportional benefit in terms of lower sick-leave than for men. The same is true for policies aimed at lowering travel times for lower paid or older workers.

[36] INE, Encuesta Suplementaria de Ingresos, octubre-diciembre, 2018 and Encuesta Nacional de Empleo, 2018. We used the average exchange rate for that year (650 CLP$/US$) to convert the monetary figure to US dollars.

[37] This is just the productivity effect and does not consider medical expenses related to the treatment received by workers.

Table 6
Linear probability model with Metro expansion variables.

	(1)	(2)	(3)	(4)
Unemployment rate	0.0173	0.0219*	0.0490***	0.0322***
	(0.0121)	(0.0119)	(0.0100)	(0.0103)
Gender	−5.006***	−4.998***		
	(0.103)	(0.104)		
contract type	−2.099***	−2.096***	−2.140***	−1.934***
	(0.0371)	(0.0372)	(0.0346)	(0.0380)
ln(tenure)	1.395***	1.396***	1.799***	1.934***
	(0.0225)	(0.0225)	(0.0288)	(0.0337)
ln(age)	22.67***	22.69***	37.37***	29.61***
	(1.123)	(1.121)	(1.827)	(2.154)
$\ln(\text{age})^2$	−3.514***	−3.517***	−5.408***	−3.534***
	(0.162)	(0.162)	(0.261)	(0.325)
ln(average firm wage)	−0.373***	−0.363***	−0.713***	−3.176***
	(0.0478)	(0.0478)	(0.0378)	(0.0735)
Maternity reform	1.239***	1.281***	1.181***	1.061***
	(0.0410)	(0.0409)	(0.0447)	(0.0464)
Maternity reform × Female	−0.132**	−0.115*	0.933***	1.188***
	(0.0671)	(0.0678)	(0.0623)	(0.0625)
Metro	−0.532***			
	(0.0701)			
Metro_{pob}		−0.826***	−0.473***	−0.536***
		(0.0876)	(0.0801)	(0.0862)
Constant	−24.55***	−24.54***	−53.17***	−17.76***
	(1.703)	(1.704)	(3.129)	(3.969)
Month seasonal controls	Yes	Yes	Yes	Yes
OD pair controls	Yes	Yes	Yes	Yes
Worker education controls	Yes	Yes	No	No
Worker fixed effect	No	No	Yes	Yes
Firm sector controls	Yes	No	Yes	No
Firm fixed effect	No	No	No	Yes
Observations	24,698,650	24,698,650	30,134,094	30,125,762
R^2	0.029	0.029	0.134	0.149
# Changes	1.736e+06	1.736e+06	1.491e+06	1.483e+06

Notes: Clustered (by origin–destination municipal pairs) standard errors in parenthesis. Dependent variable multiplied by 100 to reduce number of leading zeros of estimated coefficients. Thus, they represent changes in percentage points. ***Metro*** is a dummy variable indicating whether the municipality of residence and the municipality of employer are connected by a direct metro or rail line or by making only one transfer between different lines. Metro_{pob} is the ***Metro*** variable multiplied by the ratio of the average flow of passengers in stations in the origin and destination municipalities. The number of observations with a transition in the $Metro$ or $Metro_{pob}$ variable is presented in the last row of the table. $^{*}p < 0.10$; $^{**}p < 0.05$; $^{***}\ p < 0.01$.

Second, the link between travel times and sick-leave implies that the social cost of congestion is probably higher than what is conventionally estimated. This is worrisome for Latin America. Higher motorization rates are producing higher congestion levels in the major cities of the region. The social costs of this phenomenon should include the productivity losses from higher sick-leave behavior.

Third, Venables (2007) shows that transport investments that reduce travel times and make cities denser or larger, have benefits beyond those conventionally measured by Cost–Benefit Analysis. These additional benefits have spawned a growing literature on the wider economic benefits of transport investments. These additional benefits are already being considered in formal project appraisal methodologies, for example in the UK and Australia.[38] Research has identified output changes in imperfectly competitive markets, labor supply changes, and the benefits of static clustering (including better matching between workers and firms) as the major wider economic benefits of transport projects.

This literature has yet to consider health impacts. In this respect, the results of the present paper are novel. The lower sick-leave behavior resulting from better accessibility and lower commuting times will generate productivity benefits for the economy not measured in the conventional approach to transport investment appraisal. This is a new and additional impact beyond those being considered to date in expanded project appraisal methodologies.

We also find evidence that metro and commuter rail expansion lowered sick-leave behavior. This is in line with other studies in the region that identify positive impacts of metro investments. For example, Zárate (2021) estimates the positive casual impact of metro expansion in reducing labor market informality in Mexico City. We find that the benefits of these infrastructure investments may also include a reduction in sick-leave behavior.

Two questions remain. First, how do our results compare to previous research on this topic? There is only one paper that analyzes absenteeism directly, Van Ommeren and Gutiérrez-i Puigarnau (2011) using German data. The rest of the literature analyzes health related variables or considers the issue of sick-leave only briefly (e.g. Künn-Nelen, 2016). Ross and Zenou (2008) analyze the relationship between commuting times and shirking but indirectly, through the predictions of their model on wages and unemployment. Therefore, with one exception, our results are not comparable to previous studies. This suggests a void in the literature relating commuting times with sick-leave behavior. This is an issue where more research is warranted, as we argue in the conclusions.

In the case of Van Ommeren and Gutiérrez-i Puigarnau (2011), they use a count model of the number of days absent in the last year of the interviews (6.62 on average in their sample). Since we model the probability of taking some sick-leave but not the number of days of sick-leave (which is not directly observable in our data), we cannot compare our results directly with Van Ommeren and Gutiérrez-i Puigarnau (2011).

Given the lack of research just noted on the relationship between commuting times and sick-leave or absenteeism, the second question is then whether we can somehow extrapolate our results to other experiences? One alternative would be to apply our coefficient estimates

[38] For the case of the UK, see UK Department for Transport (2018).

Table 7
Linear probability model with heterogeneous impacts.

	(1)	(2)	(3)	(4)
Unemployment rate	0.0490***	0.0490***	0.0482***	0.0481***
	(0.0100)	(0.0100)	(0.00996)	(0.0100)
Contract type	−2.138***	−2.507***	−2.136***	−2.162***
	(0.0347)	(0.0820)	(0.0346)	(0.0351)
ln(tenure)	1.799***	1.798***	1.794***	1.819***
	(0.0288)	(0.0287)	(0.0288)	(0.0298)
ln(age)	37.36***	37.39***	33.32***	41.39***
	(1.827)	(1.829)	(1.818)	(1.791)
ln(age)^2	−5.407***	−5.412***	−4.685***	−6.042***
	(0.261)	(0.262)	(0.260)	(0.256)
ln(average firm wage)	−0.712***	−0.713***	−0.727***	−0.534***
	(0.0379)	(0.0379)	(0.0378)	(0.0379)
Maternity reform	1.200***	1.181***	1.156***	1.182***
	(0.0445)	(0.0447)	(0.0441)	(0.0447)
Maternity reform × gender	0.883***	0.933***	0.951***	0.925***
	(0.0634)	(0.0626)	(0.0620)	(0.0623)
Metro$_{pob}$	−0.266***	−0.462***	−0.401***	−0.457***
	(0.0847)	(0.0802)	(0.0752)	(0.0814)
Metro$_{pob}$ × gender	−0.361***			
	(0.0624)			
D$_{Metro}$ × gender	0.428***			
	(0.135)			
Metro$_{pob}$ × contract type		−0.0150		
		(0.0825)		
D$_{Metro}$ × contract type		0.580***		
		(0.120)		
Metro$_{pob}$ × H$_{age}$			−0.0725*	
			(0.0427)	
D$_{Metro}$ × H$_{age}$			−0.165*	
			(0.0907)	
H$_{age}$			−0.630***	
			(0.0688)	
Metro$_{pob}$ × D$_{lw}$				−0.0515
				(0.0325)
D$_{Metro}$ × D$_{lw}$				−0.111*
				(0.0646)
D$_{lw}$				0.732***
				(0.0479)
Constant	−53.33***	−52.84***	−47.46***	−62.05***
	(3.120)	(3.122)	(3.123)	(3.044)
Observations	30,134,094	30,134,094	30,134,094	30,134,094
R-squared	0.134	0.134	0.134	0.134

Notes: Clustered (by origin–destination municipal pairs) standard errors in parenthesis. Dependent variable multiplied by 100 to reduce number of leading zeros of estimated coefficients. Thus, they represent changes in percentage points. D_{Metro} is an OD dummy for two municipalities that will eventually be connected by Metro multiplied by the gender dummy. $Metro_{pob}$ is the OD dummy for two municipalities connected by Metro multiplied by the ratio of the average flow of passengers in stations in the origin and destination municipalities. $D_l w$ is a dummy for workers with labor earnings below the median in each month of the sample. Workers' fixed effects, monthly seasonal dummies, firm sector controls and origin–destination pair controls included in all regressions. *$p < 0.10$; **$p < 0.05$; *** $p < 0.01$.

to the average travel times of workers in another city. However, simulating the impact for another city using our results would be perilous. Average commuting times may not be the only relevant variable affecting sick-leave behavior. The dispersion of labor earnings, female labor market participation rates, the age structure of the workforce, modal shares, as well as the institutional particularities of labor contracts and health institutions (such as the medical license scheme used in Chile) may also be relevant. What our results do indicate is that this may be a fruitful topic for further research in transportation economics, both in Santiago (to ascertain what are the causal mechanisms behind our results) as well as in other cities.

8. Conclusions

In this paper we have presented evidence from Santiago, Chile, that indicates that longer commuting times is associated with higher sick-leave behavior by workers. This relationship is stronger for women than for men, for lower income workers and for relatively older workers. We also find evidence that metro and commuter rail expansion reduces sick-leave behavior by workers.

As discussed in the previous section, our results have implications for labor productivity in a large city such as Santiago. It also provides another benefit of urban infrastructure projects that is not considered in the conventional appraisal of these projects.

The most important limitation of our research is the inability to distinguish between the two potential underlying causes behind our results. As discussed in the literature review, longer commuting times may have health effects in terms of stress in traffic, viral exposure in public transport, and accidents (involuntary absenteeism). On the other hand, longer commuting times may increase shirking (voluntary absenteeism).

This is an open empirical question that should be addressed by future research. It would be ideal to have health diagnosis information for individual workers together with the labor market outcomes. Unfortunately, the unemployment insurance database does not contain this information and there is no readily available database on individual health outcomes to complement our data. It may be feasible to pair individual information from the public health system with the unemployment insurance database. But this proposal goes beyond the scope of this paper.

Table A.1
Probit average partial effects of observing a worker with a medical license.

Sample	(1) All	(2) All	(3) Mass layoff	(4) Unemployed recently
Unemployment rate	0.0025	0.0035	−0.0481**	−0.1461**
	(0.0048)	(0.0048)	(0.0116)	(0.0215)
Gender	−5.431**	−5.448**	−5.622**	−3.765**
	(0.1065)	(0.3530)	(0.3922)	(0.3490)
Contract type	−3.385**	−3.216**	−3.392**	−3.216**
	(0.1108)	(0.2741)	(0.3334)	(0.43891)
ln(tenure)	1.498**	1.516**	1.069**	1.594**
	(0.04886)	(0.1292)	(0.1050)	(0.2175)
ln(age)	17.67**	19.43**	3.692**	0.750
	(0.645)	(1.681)	(0.7414)	(0.5217)
ln(age)^2	−2.821**	−3.067**	−0.824**	−0.286**
	(0.1006)	(0.2644)	(0.1210)	(0.0816)
ln(average firm wage)		−0.4687**	−0.0869**	−0.2926**
		(0.0407)	(0.0174)	(0.0416)
Maternity reform	1.510**	1.739**	1.862**	1.106**
	(0.0472)	(0.1501)	(0.1870)	(0.1526)
Maternity reform × gender	−1.011**	−1.004**	−1.141**	−0.793**
	(0.0385)	(0.0897)	(0.1215)	(0.1164)
ln(travel time)	0.6591**	0.6858**	0.6145**	0.3508**
	(0.0228)	(0.05899)	(0.06258)	(0.04976)
Month seasonal controls	Yes	Yes	Yes	Yes
Workers education controls	Yes	Yes	Yes	Yes
Firm location controls	No	Yes	Yes	Yes
Firm economic sector controls	No	Yes	Yes	Yes
Observations	32,556,404	32,556,397	7,083,229	9,392,695
Pseudo R^2	0.08	0.08	0.08	0.11

Notes: Probit model estimates using the R-package Alpaca written by Amrei Stammann and Daniel Czarnowske: https://amrei-stammann.github.io/talk/alpaca/ Clustered (by worker) standard errors in parenthesis. Dependent variable multiplied by 100 to reduce number of leading zeros of estimated coefficients. Thus, they represent changes in percentage points. All models include the incidental parameter bias correction proposed by Hahn and Newey (2004). * $p < 0.05$; ** $p < 0.01$. The pseudo R_2 is calculated as one minus the ratio of the residual deviance over the null deviance.

Another avenue for future research would be to replicate our results in other cities. Since our data is exclusively from Santiago, it would be useful to examine this issue in other countries to evaluate to what extent the relationship between commuting times and sick-leave is a general phenomenon. Although the academic literature has empirically analyzed this topic for a few European countries, more research should be undertaken. This is particularly important in developing countries where for many workers commuting times are long and in uncomfortable and crowded transit systems. The main message from this paper is that this is an issue worth examining.

CRediT authorship contribution statement

Andrés Gómez-Lobo: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Visualization, Supervision. **Alejandro Micco:** Conceptualization, Methodology, Software, Data curation, Formal analysis, Investigation, Validation.

Data availability

Data will be made available on request.

Appendix A. Probit panel data model results

As a robustness check, in this Appendix we present results estimated using the panel data probit model proposed by Stammann et al. (2016) and Stammann (2018).[39]

Given its linear structure, the linear probability model is easy to estimate in a panel data setting. However, it is well known that this model may not be a good approximation to a true binary response model (e.g. logit or probit) when the outcomes are close to 0 or 1. However, estimating a non-linear binary response model with panel data is not trivial since it is usually not possible to factor out individual fixed effects. Even when this is possible, individual fixed effects are not estimated and thus marginal effects of changes in the independent variables of the model cannot be calculated.

In order to obtain marginal effects, all individual fixed effects must be estimated as parameters of the model. However, the number of these parameters increases with the number of individuals (N) in the data. This raises two problems. First, the computational difficulty of estimating such a large number of parameters. Second, the incidental parameter problem implying that coefficient estimates are not consistent as the number of individuals in the panel increases for a fixed T.

To tackle both problems, Stammann et al. (2016) propose a pseudo-demeaning algorithm and incidental parameter bias correction to estimate logit or probit panel data models. Their algorithm exploits the sparseness of the Hessian matrix when fixed effects have to be estimated and is computationally more efficient by orders of magnitude compared to alternative algorithms for large N and/or T. As for the incidental parameter problem, Stammann et al. (2016) use the bias correction proposed by Hahn and Newey (2004).

Appendix B. Travel time deciles

See Table A.3.

[39] We used the R-package Alpaca written by Amrei Stammann and Daniel Czarnowske: https://amrei-stammann.github.io/talk/alpaca/. All reported estimates include the incidental parameter bias correction proposed by Hahn and Newey (2004).

Table A.2
Average partial effects from probit model of the probability of observing a worker with a medical license including worker fixed-effects.

Sample	(1) All	(2) Female	(3) Male
Unemployment rate	0.0118*	0.0892**	−0.0341**
	(0.0045)	(0.0117)	(0.0059)
Contract type	−2.807**	−3.676**	−2.158**
	(0.3266)	(0.3317)	(0.2470)
ln(tenure)	2.064**	3.767**	1.022**
	(0.2399)	(0.3386)	(0.1169)
ln(age)	37.71**	107.07**	1.241
	(4.419)	(9.6956)	(0.5375)
ln(age)^2	−5.743**	−15.979**	−0.4063**
	(0.6726)	(1.4469)	(0.0897)
ln(average firm wage)	−0.7654**	−1.4082**	−0.3519**
	(0.0898)	(0.1291)	(0.0420)
Maternity reform	2.305**	2.040**	1.457**
	(0.2689)	(0.1858)	(0.1670)
Maternity ref × gender	−1.006**		
	(0.1193)		
ln(travel time)	0.4326**	0.7806**	0.2301**
	(0.0539)	(0.0810)	(0.0324)
Month seasonal controls	Yes	Yes	Yes
Worker fixed effects	Yes	Yes	Yes
Firm location controls	Yes	Yes	Yes
Firm sector controls	Yes	Yes	Yes
Observations	26,147,407	11,704,234	14,443,173
Pseudo R^2	0.19	0.18	0.17

Notes: Stammann et al. (2016) probit model estimates using the R-package Alpaca written by Amrei Stammann and Daniel Czarnowske: https://amrei-stammann.github.io/talk/alpaca/ Clustered (by worker) standard errors in parenthesis. Dependent variable multiplied by 100 to reduce number of leading zeros of estimated coefficients. Thus, they represent changes in percentage points. All models include the incidental parameter bias correction proposed by Hahn and Newey (2004). * $p < 0.05$; ** $p < 0.01$. The pseudo R^2 is calculated as one minus the ratio of the residual deviance over the null deviance.

Table A.3
Commuting travel time deciles (minutes).
Source: Own calculations using the data.

Deciles	Mean	Minimum	Maximum
1	8.86	4.76	11.86
2	13.31	11.92	15.10
3	16.40	15.21	18.15
4	19.25	18.15	20.35
5	21.78	20.36	23.19
6	24.55	23.20	25.49
7	27.00	25.50	28.61
8	31.00	28.61	33.20
9	36.04	33.20	38.09
10	45.51	38.09	92.73

References

Andersson, F., Haltiwanger, J.C., Kutzbach, M.J., Pollakowski, H.O., Weinberg, D.H., 2018. Job displacement and the duration of joblessness: The role of spatial mismatch. Rev. Econ. Stat. 100 (2), 203–218.

Angrist, J., Pischke, J.-S., 2008. Mostly Harmless Econometrics, An Empiricist's Companion. Princeton University Press.

Barmby, T., Sessions, J., Treble, J.G., 1994. Absenteeism, efficiency wages and shirking. Scand. J. Econ. 96 (4), 561–566.

Christian, T.J., 2012. Trade-offs between commuting time and health-related activities. J. Urban Health, Bull. New York Acad. Med. 89 (5), 746–757.

Costa, G., Pickup, L., Di Martino, V., 1988. Commuting – a further stress factor for the working people, evidence from the European community I. A review. Int. Arch. Occup. Environ. Health 60, 371–376.

Eliasson, J., Fosgerau, M., 2017. Cost-benefit analysis of transport imporvements in the presence of spillovers, matching and an income tax. Munich Personal RePEc Archive, MPRA Paper No. 76526, September.

Fonasa, Superintendencia de Seguridad Social y Superintendencia de Salud, 2017. Estadísticas de Licencias Médicas de Origen Común Y Subsidio Por Incapacidad Laboral 2016. Mimeo.

Hahn, J., Newey, W., 2004. Jackknife and analytical bias reduction for nonlinear panel models. Econometrica 72 (4), 1295–1319.

Hansson, E., Mattisson, K., Björk, J., Östergren, P-O., Jakobsson, Kristina, 2011. Relationship between commuting and health outcomes in a cross-sectional population survey in Southern Sweden. BMC Public Health 11 (834).

Hoehner, C.M., Barlow, C.E., Allen, P.A., Schootman, M., 2012. Commuting distance, cardiorespiratory fitness, and metabolic risk. Am. J. Prev. Med. 42 (6), 571–578.

Horrace, C.H., Oaxaca, R.L., 2006. Results on the bias and inconsistency of ordinary least squares for the linear probability model. Econ. Lett. 90 (3), 321–327.

Kanemoto, Y., 2013a. Second-best cost–benefit analysis in monopolistic competition models of urban agglomeration. J. Urb. Econ. 76, 83–92.

Kanemoto, Y., 2013b. Evaluating benefits of transportation in models of new economic geography. Econ. of Transp. 2, 53–62.

Künn-Nelen, A., 2016. Does commuting affect health? Health Econ. 25, 984–1004.

Pezoa, M., 2010. Licencias Médicas – Gasto Por Subsidio Incapacidad Laboral (S.I.L.). Working Paper, Superintendencia de Salud, Chile.

Ross, S.L., Zenou, Y., 2008. Are shirking and leisure substitutable? An empirical test of efficiency wages based on urban economic theory. Reg. Sci. Urban Econ. 38 (5), 498–517.

Stammann, A., 2018. Fast and Feasible Estimation of Generalized Linear Models with High-Dimensional K-Way Fixed Effects. Heinrich-Heine University Düsseldorf, July 24.

Stammann, A., Heiß, F., McFadden, D., 2016. Estimating fixed effects logit models with large panel data. Beiträge zur Jahrestagung des Vereins für Socialpolitik 2016: Demographischer Wandel - Session: Microeconometrics, No. G01-V3, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel und Hamburg.

Tsivanidis, N., 2018. The Aggregate and Distributional Effects of Urban Transit Infrastructure: Evidence from Bogota's Transmilenio. University of Chicago Booth School of Business, Job Market Paper, January 14th.

UK Department for Transport, 2018. Tag Unit A2.1: Wider Impacts (No. A2.1), Transport Analysis Guidance. UK Department for Transport.

Van Ommeren, J.N., Gutiérrez-i Puigarnau, E., 2011. Are workers with long commute less productive? An empirical analysis of absenteeism. Reg. Sci. Urban Econ. 41, 1–8.

Venables, A.J., 2007. Evaluating urban transport improvements—cost–benefit analysis in the presence of agglomeration and income taxation. J. Transp. Econ. Policy 41, 173–188.

Venables, A., Laird, J., Overman, H., 2014. Transport investment and economic performance: implications for project appraisal, October, paper commissioned by UK department for transport.

Wooldridge, J.M., 2010. Econometric Analysis of Cross-Section and Panel Data. MIT Press.

Zárate, R.D., 2021. Spatial Misallocation, Informality, and Transit Improvements: Evidence from Mexico City. Mimeo, The World Bank, January.