The European Journal of Health Economics

UNOBSERVED HETEROGENEITY IN WORK ABSENCE

--Manuscript Draft--

Manuscript Number:	EJHE-D-17-00218R1
Full Title:	UNOBSERVED HETEROGENEITY IN WORK ABSENCE
Article Type:	Original Research
Keywords:	Absenteeism; sick leave; unobserved heterogeneity; finite mixture model
Corresponding Author:	María José Suárez, Ph.D. Universidad de Oviedo Oviedo, Asturias SPAIN
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Universidad de Oviedo
Corresponding Author's Secondary Institution:	
First Author:	María José Suárez, Ph.D.
First Author Secondary Information:	
Order of Authors:	María José Suárez, Ph.D.
	Cristina Muñiz
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Labour absenteeism may be detrimental to firms and society because of the economic costs, organizational problems and production cuts that it involves. Although involuntary absenteeism due to accident or illness that prevents workers from performing their work is unavoidable, avoidable voluntary absenteeism may also emerge due to asymmetric information given that neither employers nor doctors have perfect information about workers' health status. Assuming that there is heterogeneity in individual's behaviour and thus some workers are more likely to take sick leave than others due to differences in observable and unobservable characteristics, we specify a Finite Mixture Model to analyse sick leave days per year using a sample of employees from the 2014 European Health Survey in Spain. This specification accounts for unobserved heterogeneity in a discrete way assuming that there are two types of workers even though the data do not allow us to identify which group any individual belongs to. Our results reveal that, although health indicators have the greatest impact on the proportional change in days of absenteeism, there is heterogeneity in sick leave decisions and individual and job characteristics have different effect on the absenteeism of each group.
Response to Reviewers:	See attachment

Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation

Title Page

UNOBSERVED HETEROGENEITY IN WORK ABSENCE

María José Suárez and Cristina Muñiz

(University of Oviedo, Spain)

Address:
Department of Economics
University of Oviedo
Avenida del Cristo, s/n
33006 Oviedo, Spain

Corresponding author:
María José Suárez
Email: msuarezf@uniovi.es
Tel: +(34) 985104886

ORCID:
Cristina Muñiz: 0000-0001-9935-8149
María José Suárez: 0000-0002-8152-1435

ABSTRACT

Labour absenteeism may be detrimental to firms and society because of the economic costs, organizational problems and production cuts that it involves. Although involuntary absenteeism due to accident or illness that prevents workers from performing their work is unavoidable, avoidable voluntary absenteeism may also emerge due to asymmetric information given that neither employers nor doctors have perfect information about workers' health status. Assuming that there is heterogeneity in individual's behaviour and thus some workers are more likely to take sick leave than others due to differences in observable and unobservable characteristics, we specify a Finite Mixture Model to analyse sick leave days per year using a sample of employees from the 2014 European Health Survey in Spain. This specification accounts for unobserved heterogeneity in a discrete way assuming that there are two types of workers even though the data do not allow us to identify which group any individual belongs to. Our results reveal that, although health indicators have the greatest impact on the proportional change in days of absenteeism, there is heterogeneity in sick leave decisions and individual and job characteristics have different effect on the absenteeism of each group.

JEL codes: J01, J22, I1

Keywords: Absenteeism, sick leave, unobserved heterogeneity, finite mixture model.

Click here to view linked References

UNOBSERVED HETEROGENEITY IN WORK ABSENCE

1. Introduction

Absence from work is necessary for the employees to recover from an illness or accident and/or to avoid infecting other co-workers or customers. But it may also be the result of fraudulent or opportunistic behaviour by workers because of the lack of perfect information by both employers and doctors about employee's health status. Thus, Brown and Sessions [1] state that absenteeism may be classified into two types: involuntary or unavoidable absenteeism, and voluntary or avoidable absenteeism. The first refers to those situations where employees must be on sick leave because an illness or accident prevent them from properly performing their work or because the risk of contagion. Conversely, voluntary absenteeism exists when workers do not go to work giving an excuse of illness, or having a noncontagious disease that actually would allow them to normally perform their tasks.

Absenteeism entails significant costs for workers, companies and society. For workers, absenteeism may result in income losses, reduce promotion opportunities, or increase the risk of dismissal. The economic costs of absenteeism for governments and firms consist of monetary costs by means of sick benefit payments and/or health expenditures, as well as opportunity costs arising from productivity losses.

Absenteeism figures vary among countries. Livanos and Zangelidis [2] compute the sickness absence rates for 28 European countries in the period 1992-2008 using the European's Union Labour Force Survey. The figures show that Spain ranks in an intermediate position, being the Scandinavian countries the ones with the highest rates, whereas Eastern European and Balkan countries present low absence rates. More recently, the Sixth European Working Conditions Survey shows that absenteeism in Spain is well below the EU28 average in 2015 (Eurofound [3]). The divergence among countries may be partly due to differences in the sickness benefit systems and employment protection legislation. Frick and Malo [4] elaborated two measures of the sickness benefit generosity in 15 European countries, and Spain occupies an intermediate position in their ranking.

In Spain, public health care is almost universal and sick leave must be medically certified by the public health service or by a mutual insurance company from the fourth day in a row. The benefit eligibility requirements and the amount of benefits vary according to the reason for the sick leave. If the sick leave is due to non-work-related illness, the worker should have payed Social Security contributions for at least 180 days in the previous five years and s/he also has to be registered in the Social Security system. In the case of accident or occupational disease, no contribution period is required. In each sick leave spell, the Spanish system guarantees 60% of the reference wage, which is a function of the contribution base of the preceding month, from the 4th day to the 20th day, and 75% afterwards, in the case of non-work-related illness or accident.[1] However, when the sick leave is due to occupational disease or work accident, employees receive 75% of the reference wage all along. Nevertheless, collective agreements usually increase these percentages up to 100%

[1] Companies are in charge of the benefit between the 4th and the 15th days.

from the first day. The maximum period of temporary incapacity is 12 months, but an extension of 6 months is possible under certain conditions.[2] Several reforms have been approved in last decades with the aim of reducing expenditures and preventing fraud. The latest regulatory change is the Royal Decree 625/2014 that came into force in 2015. It regulates the management and control of worker's temporary incapacity during the first year.

Oliva-Moreno [6] provides quantitative estimates of the loss of labour productivity in Spain due to health problems (premature deaths, permanent and temporary disabilities) based on the human capital theory. According to his calculations, it amounts to 37,969 million euros in 2005, of which 10,255 million correspond to temporary disabilities - equivalent to 1.13% of Spanish GDP.

Given the relevance of absenteeism, it is important to deepen our understanding of its determinants. Health issues are obviously one of the leading causes of absenteeism, but other factors such as attitudes to risk, lifestyles, valuation of time devoted to both work and leisure may affect the decision to be off sick.[3] The international economic literature on this subject usually assumes the same behavioural equation for all individuals but it would be more reasonable to consider that the motivations and the degree of response of individual decisions to the correlates of absenteeism may vary among people. However, the applied methodologies do not generally take into account this heterogeneity.

Our paper tries to fill this gap by choosing an econometric specification that allows for unobserved heterogeneity in cross-sectional data. In particular, we estimate a two-component finite mixture model (FMM) of absenteeism that represents unobserved population heterogeneity in a discrete way assuming that there are two subpopulations, so that the correlates may have different influence on the absenteeism of each group. In this specification, any individual in the sample may be a draw from any group and the researcher does not know which subpopulation a particular observation belongs to. To our knowledge, there is only one previous paper by Johansson and Palme [8] that applies this methodology to the study of absenteeism, but we use a more flexible approach than theirs.

In our empirical specification we use the European Health Survey in Spain 2014 (EESE-2014), which provides information about absenteeism as well as other personal, family and job-related characteristics. The results suggest that there is evidence of two separate groups of workers when analysing absenteeism, although involuntary absenteeism exists in both of them since health variables are the most significant covariates in explaining days of absenteeism regardless the subpopulation. The rest of the article is organized as follows. In Section 2 we review the previous research on absenteeism, focusing on the field of economics. In Section 3 we describe the main characteristics of the data base and the econometric specification. In Section 4 we present the main results and finally we report the conclusions in Section 5.

---

[2] See Villaplana [5] for more details about the Spanish sickness benefit system.

[3] Deb and Trivedi [7] also discuss these reasons to explain the distinction between groups with high average demand for medical care and low average demand.

2. Economic literature on absenteeism

There is no single definition of absenteeism. The most general one includes any type of non-attendance at work, either by legally justified reasons or not. Labour absenteeism has been the object of study of different disciplines such as sociology, psychology, management, medicine, law or economics. Kaiser [9] and Johns [10] offer reviews of different approaches and conclude that, although there may be complementarities among them, they have separately evolved and there are no interdisciplinary analyses.

The explanation of absenteeism in economics is usually based in one of the following three theories: the labour supply model, the efficiency wage model, and the theory of compensating wage differentials [1, 9]. As stated by Brown and Sessions [1], the main theoretical approach in the economic literature on absenteeism is the labour supply model, whereas labour demand or dynamic considerations are less frequent.

Following the neoclassical approach, the basic labour supply model assumes that individuals maximize their utility, which depends on consumption and leisure time, subject to budget and time constraints. Absenteeism is part of leisure time and individuals may use it as a way to approach their desired number of working hours when the labour contract sets a working time higher than their optimal value.[4] Under these assumptions, an increase in the wage rate will have an indeterminate effect on absenteeism, whereas an increase in non-labour income or in the contractual working time will raise it. Moreover, penalties for absenteeism are expected to have a negative effect. The costs of absenteeism for workers may include an increase in the probability of dismissal and a decrease in the probability of promotion or wage rise, in addition to income losses.[5]

The neoclassical work-leisure model is considered in several studies such as Allen [11], Chaudhury and Ng [12], Johansson and Palme [8, 13] and Broström et al. [14]. Avdic [15] specifies a version of this model extended to the family, assuming a household utility that is a function of consumption, leisure time and absenteeism of both spouses. In this case, the demand for absence depends not only on variables related to the individual but also to the spouse. All these models could help to explain voluntary absenteeism, i.e. absences from work that could be avoided. Conversely, Grossman's health production model incorporates involuntary absenteeism due to sickness or accident (Grossman [16, 17]. In his approach, utility depends on commodities produced at home, one of which is health. Bad health reduces the time available for work or leisure activities, thus causing a decrease in wellbeing. The stock of health depreciates along time but can be improved by investing time and goods.

Demand considerations may also be relevant in explaining absenteeism. The employer's tolerance for absenteeism will depend on the marginal benefits and costs of enforcing the contract. Those firms where the absence of a worker implies serious damages to the productive activity may enforce attendance, through monitoring, threat of dismissal,

---

[4] Workers may be willing to accept a job, even if they will not work the number of desired hours, if it is the best option among the available alternatives.

[5] See Brown and Sessions [1] and Allen [11].

or by offering a higher wage to induce workers to fulfil contract conditions. Therefore, the efficiency wage hypothesis can be applied to the analysis of voluntary absence [18, 1] and it implies a negative correlation between wages and absenteeism and between unemployment rates and absenteeism.[6] Otherwise, the theory of compensating wage differentials concludes that wages and involuntary absenteeism are positively associated since it suggests that those jobs with more risk of illness or accident will offer higher wages to attract workers [8]. However, it also implies a negative relationship between wages and voluntary absenteeism since a lower salary can compensate for a higher level of voluntary absenteeism [14].[7]

It should also be mentioned that some recent economic papers analyse the opposite behaviour to absenteeism, known as presenteeism, that is, employees who go to work when they should not. This phenomenon may also generate losses in both production and earnings to the companies, as well as health risks to workers. Some reasons to explain presenteeism are the increase in the probability of dismissal, the decrease in promotion opportunities or the reduction in income in case of taking sick leave. In addition, working conditions such as work overload and more demanding jobs can induce presenteeism. Pichler and Ziebarth [20] and Hirsch et al. [21] provide an analytical framework to explain presenteeism and absenteeism.

Turning now to the empirical analysis on absenteeism, different methodologies have been applied depending on the objective of the studies and the data available. In some cases, absenteeism is defined in absolute or proportional terms and it is considered as a continuous variable. In particular, Barmby et al. [22] estimate the absenteeism rates across countries using the Ordinary Least Squares (OLS) method. More recently, Scoppa [23], Scoppa and Vuri [24], Avdic [15] and Goerke and Pannenberg [25] also apply the same methodology, to Italian data in the first two papers, and to Swedish and German data in the third and fourth papers respectively. In other cases, absenteeism is defined as a binary variable. Leontaridi and Ward [26] apply probit models to study the probability of absenteeism in 15 OECD countries while Böckerman and Ilmakunnas [27] focus on Finnish data. Moreover, the previously mentioned papers by Scoppa [23], Scoppa and Vuri [24] and Goerke and Pannenberg [25] also use probit or Linear Probability Models (LPM). Instead, Howard and Potter [28] estimate logit models with US data.

Several works analyse transitions between work and absenteeism applying hazard rate models. This is the case of Johansson and Palme [13], Broström et al. [14] and Avdic [15] for Sweden, or Markussen et al. [29] for Norway. Other authors address absenteeism as a count variable that can take a discrete number of values. Specifically, Delgado and Kniesner [30] compare different count data specifications to study short absenteeism spells of London bus drivers; Frick and Malo [4] apply Zero Inflated Negative Binomial (ZINB) models to explain absenteeism in 14 European countries, whereas Lechmann and Schnabel [31] and Lorenz

[6] Barmby et al. [18] develop a theoretical model assuming that preferences are a positive function of leisure, consumption and health, and leisure valuation of workers is higher when they are in bad health. They also model the firm's decisions and obtain a positive relationship between wages and monitoring costs, supporting the efficiency wage hypothesis.

[7] A different approach can be found in Kahana and Weiss [19], who apply game theory to the explanation of unjustified absenteeism in labour-managed firms and profit-maximizing firms.

and Goerke [32] analyse German data specifying hurdle count data models, or pooled and fixed effect negative binomial regressions respectively. Finally, Johansson and Palme [8] estimate FMM to explain absenteeism of Swedish blue-collar workers assuming a binomial distribution. Moreover, the authors consider that there are two subpopulations but unobserved heterogeneity only affects the intercept.

In Spain, Jimeno and Toharia [33] and Blázquez [34] use the Labour Force Survey (LFS) to estimate the probability of absenteeism by applying probit and logit models respectively. In particular, Jimeno and Toharia [33] focus on the effect of the type of contract. García-Mainar et al. [35] also study the effect of permanent contracts on absenteeism with longitudinal data from the LFS and applying instrumental variable methodologies. Instead, García-Serrano and Malo [36] estimate the probability and frequency of voluntary and involuntary absences using a Spanish panel data of large firms in order to check the influence of union voice. Finally, Murcia et al. [37] use Cox proportional hazards models to compare absenteeism duration before and after the start of the Spanish economic crisis.

Other Spanish papers analyse specific types of absenteeism. This is the case of Catalina-Romero et al. [38], who apply ZINB models to explain the likelihood and the length of non-work-related sickness absenteeism to check the effect of work-related psychosocial factors. Conversely, Bande and López-Mourelo [39] study the impact of age on absenteeism due to occupational accidents by estimating duration models with administrative data from the Spanish Statistics on Accidents at Work. Also, Martín-Román and Moral [40, 41] focus on work-related accidents and use the same administrative data. On the one hand, Martín-Román and Moral [40] specify a theoretical model to explain why there are more work accidents on Mondays than on other weekdays and estimate probit models to check the Monday gap. More recently, Martín-Román and Moral [41] apply stochastic cost frontiers to disentangle which part of the sick leave duration may be attributed to medical reasons and which part to opportunistic behaviour.

In previous studies, the set of covariates included as determinants of absenteeism usually comprise personal and family factors, characteristics of the employment contract, occupation, firm-related variables, job satisfaction, income or labour earnings, health indicators, geographical variables, and/or local or regional unemployment rates. Most studies conclude that absenteeism is positively correlated with age, being female, being single, blue-collar occupation, permanent contract and shift work. In addition, some studies obtain that workers with higher educational level show lower levels of absenteeism. Finally, several studies pose that absenteeism is also related with job stress factors [34].

3. Data and empirical specification

The database used in this paper is the 2014 European Health Survey in Spain (EESE-2014), carried out on a sample of individuals aged 15 and over and residing in family homes throughout the Spanish territory. This survey was conducted from January 2014 to January 2015 and its objective was to provide information on the health of the population living in

Spain, following the criteria established by Eurostat. Therefore, the information is comparable with that of other European countries.[8]

The information about absenteeism comes from the following two questions: In the past 12 months, have you been absent from work due to health problems? How many days in total? The questionnaire indicates that respondents must take into account all kind of diseases, injuries and other kind of health problems that they had and which resulted in their absence from work. Thus, the dependent variable in the empirical specification is the number of absenteeism days in past year. Regarding the covariates, in first place, absenteeism should be correlated with health. Therefore, we include a set of health indicators: dummies of self-perceived health status, another binary variable that identifies those individuals suffering from chronic illness or pain and a dummy equal to one if the individual has had psychological problems. We also include two dummies, equal to one if the person has been hospitalized at least one night or admitted to hospital as a day patient respectively. All these variables refer to the last 12 months prior to the interview. In second place, personal characteristics may reflect different preferences towards leisure time and absenteeism. Thus, we incorporate gender, age (in quadratic form), marital status and number of children. In third place, family income may have an influence on absenteeism decisions because, according to the labour supply model, higher income will imply more absenteeism if it is a normal good. In fourth place, working conditions can be determinants of the probability of suffering from illness or accident and also the culture of absenteeism in the workplace may affect individual decisions. The employment characteristics included are white collar occupation,[9] type of contract (a binary variable for permanent contract), a dummy equal to one if the work does not require physical effort and variables related to working time (part-time and daily split shift dummies).[10]

Our empirical analysis is restricted to employees who provide information about all variables used in the estimation. The total sample size is 6,289 observations and the summary statistics are displayed in Table 1. 58% of the sample reports that their health is good, but 59% had an illness or physical health problem in last 12 months. However, a very small proportion of the sample had psychological problems in that period or were admitted to a hospital. Regarding personal and family variables, 51% are males, 60% are married and the average age is near 43 years. Turning to job characteristics, 80% have a permanent job, 67% of the employees are white-collar workers, 16% work part-time, a third have a daily split shift, and most of them have a sedentary work, i.e. they are sitting or standing most of the workday without large movements or physical efforts.

[8] See INE [42] for more details. The EESE has been carried out every 5 years since 2009.

[9] White collar occupations include managerial, professional, technical, clerical, sales, service and military occupations.

[10] In initial estimates, we also considered regional unemployment rates and activity sector but these variables were never significant, and the Bayesian Information Criteria took higher values when including them.

Table 1. Summary Statistics

	Mean	St. Dev.	Min.	Max.
Dummy absenteeism	0.241	0.428	0	1
Days of absenteeism	7.872	36.068	0	365
Income1 (<970 euros)	0.110	0.313	0	1
Income2 (970-1399 euros)	0.206	0.404	0	1
Income3 (1400-2039 euros)	0.307	0.461	0	1
Income4 (2040-3279 euros)	0.258	0.438	0	1
Income5 (≥3280 euros)	0.119	0.324	0	1
Male	0.511	0.500	0	1
Age	42.923	10.007	17	74
Married	0.599	0.490	0	1
#children	0.606	0.844	0	6
Permanent job	0.804	0.397	0	1
Part-time job	0.156	0.363	0	1
Daily split shift	0.333	0.471	0	1
White-collar worker	0.671	0.470	0	1
Sedentary job	0.729	0.444	0	1
Bad or very bad health	0.028	0.166	0	1
Fair health	0.145	0.352	0	1
Good health	0.583	0.493	0	1
Very good health	0.244	0.429	0	1
Physical illness	0.587	0.492	0	1
Psychological illness	0.068	0.252	0	1
Hospital admission	0.054	0.226	0	1
Day hospital admission	0.064	0.245	0	1
# observations	6289			

Concerning the econometric specification, given the information about absenteeism in the data set, it is reasonable to apply a count data model. Another characteristic of the dependent variable is the high proportion of zeros (i.e. no absenteeism). Some authors have dealt with this issue by specifying zero inflated regression e.g. [4] or hurdle models e.g. [31]. Zero inflated models assume that there may be two reasons to explain no absenteeism: some individuals are not off sick under any circumstances, whereas other workers might have taken sick leave but they did not in the studied period. We think that this specification would be appropriate to the analysis of voluntary absenteeism, i.e. when individuals are able to choose to be absent, but we cannot distinguish between voluntary and involuntary absenteeism in our data, so that any employee may have a positive probability of being absent. On the other hand, the double hurdle model assumes a two-stage decision-making process: firstly, there is the individual's decision about taking time off work or not, and secondly the length of the absenteeism period. In the first step all observations are taken into account, while in the second step the estimation is only applied to those who have been absent. This last specification is also questionable when there is involuntary absenteeism, i.e. individuals who do not choose to be absent but they cannot work due to health problems.

In our empirical analysis we apply a FMM to explain absenteeism because we believe that not all individuals respond equally to changes in the correlates. The FMM assumes that any observation in the sample is a draw from a population that is a mixture of J subpopulations, i.e. there are different groups with different behaviour but it is not possible to distinguish the group to which each observation belongs to, because there is no prior information to classify them in advance [43]. Thus, this model incorporates unobserved heterogeneity in a discrete way, so that the density of the dependent variable is modelled as a linear combination of *J* different densities. In this formulation, the total number of components is not subject to estimation but it has to be established. In addition, all observations may have a positive probability of belonging to any group and any individual can have positive or zero absenteeism in the previous year, regardless the group s/he belongs to.[11]

In our specification, the number of absenteeism days in past year (y_i) follows a negative binomial 1 distribution -which assumes a linear variance function- and we consider two subpopulations or components, which may have different attitudes towards absenteeism, although there may be voluntary and involuntary absenteeism in any group.

The density of an observation *i* is modelled as:

$$f(y_i/\theta, x_i, \pi) = \pi f_1(y_i/\theta_1, x_i) + (1-\pi) f_2(y_i/\theta_{2,}\, x_i)$$

In the previous equation f is called the mixture distribution, $f_j(.)$ is the *jth* subpopulation density or component distribution (j=1, 2), and π is the probability of belonging to group 1 (mixing distribution). Finally, x_i is a vector of covariates and θ_j refers to the parameters in subpopulation j. Given that we do not make any assumption about the mixing distribution π, this is a semi-parametric model and it is regressed by maximum likelihood. All observations contribute to the estimation of both subpopulation parameters and the results provide us estimates of the component coefficients and the mixing distribution.

4. Results

In this section, we present and describe the main results of the FMM estimates. Before discussing the relationship between the covariates included and absenteeism, we split the sample into two groups according to the posterior probabilities of belonging to each component, and make a descriptive analysis of each one. The posterior probabilities are computed from the estimated coefficients as follows:

$$\Pr(y_i \in subpopulation\ 1\ /\ x_i, y_i, \theta) = \frac{\pi\, f_1(y_i\ /\ \theta_1, x_i)}{\pi\, f_1(y_i\ /\ \theta_1, x_i) + (1-\pi)\, f_2(y_i\ /\ \theta_2, x_i)}$$

[11] The zero inflated negative binomial model (ZINB) is a particular case of the negative binomial finite mixture model. ZINB just allows mixing with respect to zeros whereas the second model allows mixing in both zero and positive values of absenteeism [7].

$$\Pr(y_i \in subpopulation\ 2\ /\ x_i, y_i, \theta) = \frac{(1-\pi)\ f_2(y_i\ /\ \theta_2, x_i)}{\pi f_1(y_i\ /\ \theta_1, x_i) + (1-\pi)\ f_2(y_i\ /\ \theta_2, x_i)}$$

Table 2 presents the means and standard deviations of the posterior probabilities, as well as the predicted days of absenteeism in each possible scenario, i.e. the weighted value of both components, and the values that would correspond if the whole sample would behave like group 1 and 2 respectively. For the total sample the predicted mean is 7.7 days of absence in a year –next to the mean observed value that is 7.9 (see Table 1). However significant differences exist when comparing the length of absenteeism predicted if all workers behave as group 1 or 2. In the first case, mean days of sick leave is around one and, in the second case, the figure reaches almost 10. This result may be an indication of more voluntary absenteeism in the second component.

Table 2. Posterior probabilities and predicted days of absenteeism (total sample)

	Mean	Std. Dev.
Posterior prob. Є group 1	0.2556	0.1811
Posterior prob. Є group 2	0.7444	0.1811
Predicted days of absenteeism[a]	7.6576	19.2768
Predicted days of absenteeism (if comp. 1)[b]	1.3665	6.1725
Predicted days of absenteeism (if comp. 2)[c]	9.8179	25.0902
# Observations	6289	

Notes:
[a] The predicted days of absenteeism are computed as: $\hat{\pi} f_1(y_i/\widehat{\theta_1}, x_i) + (1-\hat{\pi}) f_2(y_i/\widehat{\theta_2}\ x_i)$
[b] The predicted days of absenteeism (if comp. 1) are computed as follows: $f_1(y_i/\widehat{\theta_1}, x_i)$
[c] The predicted days of absenteeism (if comp. 2) are calculated as: $f_2(y_i/\widehat{\theta_2}, x_i)$

If we split up the sample assigning each individual to the subpopulation to which they are more likely to belong -according to the posterior probabilities-, about 10% of the sample are assigned to the first group and 90% to the second group. The upper part of Table 3 shows the mean values of all the variables included in our estimates for each group. There are interesting differences between components. People belonging to the first component have a higher absenteeism rate but of shorter duration than group 2.[12] Health variables seem to indicate that individuals in group 1 have poorer health status. There is a higher proportion of females and the average age and family commitments are lower in the first component than in the second one. Regarding job characteristics, there are greater proportions of white-collar workers, with permanent contracts, split working time and sedentary jobs in the first component.

[12] The classification obtained by Johansson and Palme [8] distinguishes between a group primarily consisting of the long-term sick and the rest of the sample.

Table 3. Mean values of the variables and the predicted days of absenteeism by subpopulation

	Subpopulation 1	Subpopulation 2
Summary statistics (mean values)		
Dummy absenteeism	0.849	0.175
Days of absenteeism	3.775	8.321
Income1 (<970 euros)	0.074	0.114
Income2 (970-1399 euros)	0.164	0.210
Income3 (1400-2039 euros)	0.296	0.308
Income4 (2040-3279 euros)	0.314	0.252
Income5 (≥3280 euros)	0.151	0.116
Male	0.435	0.520
Age	39.403	43.308
Married	0.502	0.610
#children	0.583	0.608
Permanent job	0.849	0.799
Part-time job	0.155	0.156
Daily split shift	0.428	0.323
White-collar worker	0.836	0.653
Sedentary job	0.820	0.719
Bad or very bad health	0.056	0.025
Fair health	0.127	0.147
Good health	0.568	0.584
Very good health	0.248	0.243
Physical illness	0.675	0.577
Psychological illness	0.097	0.065
Hospital admission	0.124	0.046
Day hospital admission	0.122	0.058
Predicted absenteeism (mean values)		
Predicted days of absenteeism (if comp. 1)	3.197	1.166
Predicted days of absenteeism (if comp. 2)	15.833	9.159
# Observations	621	5668

In the lower part of Table 3 we include the predicted days of absenteeism for each subsample according to their behaviour equation and that of the other component. The predicted absence days when applying the appropriate behavioural equation are quite similar to the average sample values shown at the top of the table. It is also interesting to note that if group 1 behaved as group 2, their average predicted days of absenteeism would be higher than when we apply their parameters (15.8 *versus* 3.2 days). Conversely, if group 2 behaved as group 1 their average days of absenteeism would be lower than when applying the parameters corresponding to this group (1.2 *versus* 9.2). These results seem to indicate more voluntary or opportunistic absenteeism in group 2 because, with the same socio-economic, health and labour characteristics, the mean predicted days of absenteeism are higher when applying group 2 coefficients.

Turning now to the estimation results, Table 4 shows two sets of coefficients which correspond to each subgroup, as well as the prior probability of belonging to subpopulation 1 (π). Although there is a non-linear relationship between the dependent variable and the coefficients, the expected value of the number of absenteeism days in component j is:

$$E_j(y_i \,/\, x_i, \theta_j) = e^{x_i \theta_j} \qquad\qquad j = 1, 2$$

Therefore, the proportional change in the expected value as a result of a unitary change in an independent variable, x_k, is equal to the coefficient of the variable:

$$\frac{\partial E_j(y_i \,/\, x_i, \theta_j)}{\partial x_k} \frac{1}{E_j(y_i \,/\, x_i, \theta_j)} = \theta_k \qquad\qquad j = 1, 2$$

In conclusion, the coefficients in Table 4 can be interpreted as semielasticities, i.e. they show the proportional change in days of absenteeism when the independent variable changes in one unit.[13]

As expected, health indicators are the most significant factors in explaining absenteeism in both groups. Furthermore, according to our results, self-reported health and overnight hospitalizations are the covariates with the highest effect on the proportional change in days of absenteeism (in absolute value). Comparing between groups, the impact of suffering longstanding physical health problems and hospitalizations as an inpatient is greater in the first component. With regard to the rest of socio-economic variables, income has a positive relationship with absenteeism in both groups but the effect is again higher in group 1. The positive sign of income suggests that absenteeism is a normal good. Gender and marital status are not significant in group 1 whereas females and unmarried employees have higher absenteeism in group 2. In the case of age, it has an increasing effect on the expected days of absenteeism in group 1, reaching a maximum at about 30 years of age. By contrast, it has a U-shaped effect on absenteeism in group 2 being the minimum value of absenteeism at 43 years of age. Therefore, in group 2 the older and younger employees have higher absenteeism rates than middle-aged workers and the contrary happens in group 1. A plausible explanation for the increasing absenteeism of young people in component 1 may be the mismatch between qualifications and employment, which may generate job dissatisfaction and favour absenteeism.[14] However, absenteeism in component 2 increases at the end of the working life.

[13] In the case of dummy covariates it would be more accurate to compute the difference in absenteeism when the variable takes the value 1 and 0 respectively. However, the coefficients give an approximation of their impact.

[14] Young Spaniards tend to be overqualified for the jobs they perform and a third of Spanish university graduates are employed in jobs below their qualifications [44].

Table 4. Days of absenteeism in previous year: Two-component FMM

	Component 1			Component 2		
	Coefficient		t-Student	Coefficient		t-Student
Income2	0.259		0.63	0.250		1.64
Income3	0.371		1.43	0.430	***	3.16
Income4	0.684	***	2.59	0.370	**	2.54
Income5	0.758	**	2.49	0.413	**	2.45
Male	0.083		0.69	-0.185	**	-2.43
Age	0.167	***	2.72	-0.050	*	-1.89
$Age^2/100$	-0.285	***	-3.73	0.058	**	1.96
Married	-0.004		-0.03	-0.273	***	-3.34
#children	-0.225	***	-2.83	0.099	**	2.02
Permanent job	0.612	***	3.48	0.363	***	3.31
Part-time job	0.446	**	2.26	-0.214	*	-1.93
Daily split shift	0.452	***	3.30	-0.182	**	-2.21
White-collar worker	0.842	***	3.92	-0.207	***	-2.61
Sedentary job	0.318	*	1.76	-0.191	**	-2.34
Fair health	-4.095	***	-7.74	-0.743	***	-6.38
Good health	-0.499	***	-2.79	-1.868	***	-14.54
Very good health	-0.838	***	-3.55	-2.304	***	-13.37
Physical illness	0.568	***	3.56	0.188	*	1.96
Psychological illness	-0.028		-0.13	0.442	***	4.56
Hospital admission	3.212	***	20.62	1.463	***	14.78
Day hospital admission	0.418	***	2.84	0.795	***	8.51
Constant	-4.044	***	-3.36	4.083	***	6.84
Π	0.256 (st. dev.: 0.025)					
Log L	-8613.663					
BIC	17638.41					
# observations	6289					

Note: * p<0.1, ** p<0.05, *** p<0.01

Permanent workers present higher absenteeism in both components. The effect of the type of contract has been the focus of study of some previous articles in Spain and our result is in line with them (e.g. Jimeno and Toharia [33] and García-Mainar et al. [35]).[15] Finally, children and most job characteristics have an opposite effect on the absenteeism of each group. In group 1, absenteeism is positively associated with white-collar, part-time and sedentary jobs or daily split shifts, contrary to what happens in group 2.

The differences found in the significance and the effect of the covariates on each component, corroborate that this specification seems more appropriate than those that do not account for unobserved heterogeneity, or that assume that heterogeneity does not affect the sensibility of absenteeism to changes in the covariates [8]. Nevertheless, we have made some additional estimates to check the robustness of our specification. First, we regressed

[15] Although we cannot rule out presenteeism in our country, this problem is more serious when there are no sick leave benefits or they are very low. This is not the case of the Spanish system, which covers from 60 to 75% of the reference wage from the fourth day -and many collective agreements increase these percentages up to 100%. However, the high rate of temporary employment in Spain could justify presenteeism in this group for fear of losing the job. In fact, absenteeism is higher among workers with permanent contracts.

the model using a negative binomial 2 distribution, which assumes a quadratic variance function, but the negative binomial 1 distribution is preferred, according to the Bayesian Information Criterion (BIC). Second, we considered three components and one component instead of two, but the BIC values allow us to state that two components are more appropriate for our data. We also tried to estimate the model with more than three components but we found convergence problems. Third, we imposed the constraint that the components only differed in the intercept, but the likelihood ratio test led us to reject the null hypothesis of equality of slope coefficients across groups. Fourth, we added regional variables and, according to the likelihood ratio test, we could not reject the hypothesis that regional coefficients were jointly zero.[16] All these estimates are shown in Tables A.1 and A.2 of the Appendix.

In conclusion, we have robust results suggesting that employees' behaviour is not homogeneous, but there are unobserved attitudes towards absenteeism or health indicators not included in the set of covariates that explain the heterogeneity in absenteeism decisions. Besides health covariates, some job characteristics such as job qualification and type of contract, as well as household income level have a relatively high impact on absenteeism.

It is difficult to determine whether voluntary absenteeism is concentrated on one of the two groups. On the one hand, the summary statistics in Table 3 show that workers assigned to group 1 are more likely to take sick leave. What is more, the marginal effects of most of the socio-economic and labour covariates on the proportional change in absenteeism are higher in group 1. On the other hand, the mean absenteeism duration is less than half in group 1 compared to group 2 and they also have worse health status (greater rates of hospital admissions, physical or psychological diseases). Moreover, our model predicts more days of absenteeism when applying the behavioural equation of group 2, thus we conclude that there is more voluntary or opportunistic absenteeism in this group.

5. Concluding remarks

Labour absenteeism causes damages to the companies because of the fall in production or organization problems that it may entail, as well as economic costs for governments, but it is justified when health problems prevent employees from performing their tasks successfully or there is risk of contagion to other colleagues or customers. However, there is also certain degree of discretion on the part of workers when asking for sick leave because individuals have more information about their health status than doctors or employers. Given that individuals differ in risk attitudes, opportunistic behaviour and/or reactions to health problems, some employees are off sick even though they should go to work, whereas others only take sick leave when it is strictly necessary or even do not take it when they should.

Therefore, there is heterogeneity in workers' behaviour but this heterogeneity is not observed and the applied methodologies generally assume the same behaviour of workers

[16] In the last two estimates mentioned in the text, we apply likelihood ratio tests because the models are nested.

towards absenteeism. Given that we have a cross-sectional database, we have taken into account unobserved heterogeneity by specifying a finite mixture model (FMM) which explains the number of days of absenteeism in the previous year. Our specification assumes that any observation in the sample may be a draw from two different groups or subpopulations with different absence behaviour. This model is applied to a sample of employees from the 2014 European Health Survey in Spain, and we include as covariates personal and family variables, job-related characteristics and health indicators. It is worth noting that although FMM has been previously applied to the analysis of absenteeism by Johansson and Palme [8], we contribute to the literature by estimating a more flexible model than theirs, by allowing different responses by group to changes in the covariates. In fact, our results corroborate this assumption.

Our estimates reveal that health indicators are the main determinants of absenteeism in both groups. Moreover, the model divides workers into two subpopulations, which differ in the duration and probability of absence: The first group is more likely to be absent from work but absences tend to be short, whereas the contrary happens with the other group. The estimates reveal that personal and job characteristics usually have a greater effect, in absolute value, on the proportional change in days of absenteeism in the first group than in the second one, when significant. In addition, the predicted days of absence when using the behavioural equation of the second group are higher than when using that of the first component, thus opportunistic behaviour is probably more present in the second group.

The results obtained can be useful for policy makers. Firstly, health-related policies are central because health variables are the main determinants of absenteeism. Occupational risk prevention policies as well as preventive health policies are necessary to reduce involuntary absenteeism. As regards to opportunistic absenteeism, our estimates suggest that greater monitoring over permanent workers can reduce it, since they have higher absenteeism in both subpopulations, perhaps because their risk of losing employment is lower than that of temporary workers. However, the rest of job characteristics as well as other variables have an opposite effect on absenteeism of each group, so that policies that do not take into account the heterogeneity of workers may fail to achieve the aim of reducing absenteeism.

The positive effect of children and females in the second component may be reflecting problems of finding a balance between work and family, so that policies aimed at balancing work and home may reduce absenteeism. In particular, greater flexibility in work schedules may reduce incentives to improperly take time off work. Also, greater control over sick leaves of young employees or policies that facilitate a rapid integration into jobs that match the qualifications acquired by youth could reduce voluntary absenteeism given that absenteeism increases with age until 30 years in group 1. However, absenteeism increases at the end of the working life in group 2, so that public measures related to delaying retirement age could exacerbate the problem of absenteeism in this group and increase healthcare costs and sickness benefits.

Finally, it is worth noting that, although opportunistic absenteeism may be partly reduced with greater control in the workplace, control is costly and some firms may be willing to tolerate some level of shirking in exchange for lower wages, or they may offer higher wages to reduce the probability of absenteeism following the theories of compensating wage differentials and efficiency wages.

This study has some limitations. We would like to include other variables in the set of covariates such as dummies for public/private sector, moonlighting or job satisfaction as they could be relevant in explaining worker absenteeism. For further extensions of this work, it would be interesting to have more detailed information on the causes of absenteeism as well as data about the timing and duration of spells in the analysed time interval. Panel data would also allow a more flexible control of unobserved heterogeneity.

References

1. Brown, S., Sessions, J.G.: The economics of absence: theory and evidence. J. Econ. Surv. 10(1), 23-53 (1996)
2. Livanos, I., Zangelidis, A.: Unemployment, labor market flexibility, and absenteeism: a pan-european study. Ind. Relat. 52(2), 492-515 (2013)
3. Eurofound: Sixth European Working Conditions Survey – Overview report, Publications Office of the European Union, Luxembourg. https://www.eurofound.europa.eu/sites/default/files/ef_publication/field_ef_document/ef1634en.pdf (2016). Accessed 10 July 2017
4. Frick, B., Malo, M.Á.: Labor market institutions and individual absenteeism in the European Union: the relative importance of sickness benefit systems and employment protection legislation. Ind. Relat. 47(4), 505-529 (2008)
5. Villaplana, M.D.: Absentismo e incapacidad laboral. Promover organizaciones saludables como garantía de excelencia y efectividad organizativa. Ed. Consejo Económico y Social del Principado de Asturias. Colección de Estudios, nº 19 (2015)
6. Oliva-Moreno, J.: Loss of labour productivity caused by disease and health problems: what is the magnitude of its effect on Spain's Economy? Eur. J. Health Econ. 13(5), 605-614 (2012)
7. Deb, P., Trivedi, P.K.: Demand for medical care by the elderly: a finite mixture approach, J. Appl. Econom. 12(3), 313-336 (1997)
8. Johansson, P., Palme, M.: Do economic incentives affect work absence? Empirical evidence using Swedish micro data. J. Public Econ. 59(2), 195-218 (1996)
9. Kaiser, C.P.: What do we know about employee absence behavior? An interdisciplinary interpretation. J. Socio-Econ. 27(1), 79-96(1998)
10. Johns, G.: How methodological diversity has improved our understanding of absenteeism from work. Hum. Resour. Manage. Rev. 13(2), 157-184 (2003)
11. Allen, S.G.: An empirical model of work attendance. Rev. Econ. Stat. 63(1), 77-87 (1981)
12. Chaudhury, M., Ng, I.: Absenteeism predictors: least squares, rank regression, and model selection results. Can. J. Econ. 25(3), 615-635 (1992)
13. Johansson, P., Palme, M.: Assessing the effect of public policy on worker absenteeism. J. Hum. Resour. 37(2), 381-409 (2002)
14. Broström, G., Johansson, P., Palme, M.: Economic incentives and gender differences in work absence behaviour. Swed. Econ. Policy. 11(1), 33-63 (2004)
15. Avdic, D.: Microeconometric Analyses of individual behavior in public welfare systems. Applications in health and education economics, Thesis, Economic Studies 143. Uppsala University, Department of Economics. ISBN: 978-91-85519-50-7(2014)

16. Grossman, M.: On the concept of health capital and the demand for health. J. Polit. Econ. 80(2), 223-255 (1972)
17. Grossman, M.: The human capital model. In: Culyer, A.J., Newhouse, J.P. (eds.), Handbook of Health Economics, vol. 1, Part A, Chapter 7, pp. 347-408. Elsevier, Amsterdam (2000)
18. Barmby, T.A., Sessions, J., Treble, J.G.: Absenteeism, efficiency wages and shirking. Scand. J. Econ. 96(4), 561-566 (1994)
19. Kahana, N., Weiss, A.: Absenteeism: a comparison of incentives in alternative organizations. J. Comp. Econ. 16(4), 573-595 (1992)
20. Pichler, S., Ziebarth, N.R.: The pros and cons of sick pay schemes: testing for contagious presenteeism and noncontagious absenteeism behaviour. J. Public Econ. 156, 14-33 (2017)
21. Hirsch, B., Lechmann, D.S., Schnabel, C.: Coming to work while sick: An economic theory of presenteeism with an application to German data. Oxford Econ. Pap. 69(4), 1010-1031 (2017)
22. Barmby, T.A., Ercolani, M.G., Treble, J.G.: Sickness absence: an international comparison. Econ. J. 112(480), 315-331(2002)
23. Scoppa, V.: Worker absenteeism and incentives: evidence from Italy. Manage. Decis. Econ. 31, 503-515 (2010)
24. Scoppa, V., Vuri, D.: Abseenteeism, unemployment and employment protection legislation: evidence from Italy, IZA DP No. 7091(2012)
25. Goerke, L., Pannenberg, M.: Trade union membership and sickness absence: Evidence from a sick pay reform. Labour Econ. 33, 13-25(2015)
26. Leontaridi, R.M., Ward, M.E.: Work-related stress, quitting intentions and absenteeism, IZA DP No. 493 (2002)
27. Böckerman, P., Ilmakunnas, P.: Interaction of working conditions, job satisfaction, and sickness absence: Evidence from a representative sample of employees. Soc. Sci. Med. 67, 520-528(2008)
28. Howard, J.T., Potter, L.B.: An assessment of the relationships between overweight, obesity, related chronic health conditions and worker absenteeism. Obes. Res. Clin. Pract. 8, e1-e15 (2014)
29. Markussen, S., Røed, K., Røgeberg, O.J., Gaure, S.: The anatomy of absenteeism. J. Health Econ. 30(2), 277-292 (2011)
30. Delgado, M.A., Kniesner, T.J.: Variance of unknown form: An application to a hedonic model of worker absenteeism. Rev. Econ. Stat. 79(1), 41-49 (1997)
31. Lechmann, D.S., Schnabel, C.: Absence from work of the self-employed: A comparison with paid employees. Kyklos, 67(3), 368-390 (2014)
32. Lorenz, O., Goerke, L.: Commuting and sickness absence. Beitrage zur Jahrestagung des Vereins fur Socialpolitik 2015: Okonomische Entwicklung Theorie und Politik-Session: Labor- Empirical Studies 5. No. G16-V4 (2015)
33. Jimeno, J.F., Toharia, L.: Effort, absenteeism, and fixed term employment contracts. Revista Española de Economía, 13(1), 105-119 (1996)
34. Blázquez, M. B.: Sickness absence rates in Spain− Evidence for the period 1996–2004. Cuadernos de economía, 35(97), 1-8 (2012)
35. García Mainar, I., Green, C. P., Navarro Paniagua, M.: The effect of permanent employment on absenteeism: Evidence from labor reform in Spain. ILR Rev. (2017). DOI: 10.1177/0019793917717226
36. García-Serrano, C., Malo, M. A.: The impact of union direct voice on voluntary and involuntary absenteeism. J. Socio-Econ. 38(2), 372-383 (2009)

37. Murcia, G., Delclós, J., Ubalde, M., Calvo, E., Benavides, F. G.: Has the Spanish economic crisis affected the duration of sickness absence episodes?. Soc. Sci. Med. 160, 29-34 (2016)
38. Catalina-Romero, C., Sainz, J. C., Pastrana-Jiménez, J. I., García-Diéguez, N., Irízar-Muñoz, I., Aleixandre-Chiva, J. L., González-Quintela, A.; Calvo-Bonacho, E.: The impact of poor psychosocial work environment on non-work-related sickness absence. Soc. Sci. Med. 138, 210-216 (2015)
39. Bande, R., López-Mourelo, E.: The impact of worker's age on the consequences of occupational accidents: Empirical evidence using Spanish data. J. Labor Res. 36(2), 129-174 (2015)
40. Martín-Román, A.L., Moral, A.: Moral hazard in Monday claim filing: Evidence from Spanish sick leave insurance. BE J. Econ. Anal. Policy. 16(1), 437-476 (2016)
41. Martín-Román, A.L., Moral, A.: A methodological proposal to evaluate the cost of duration moral hazard in workplace accident insurance. Eur. J. Health Econ. (2017). DOI: 10.1007/s10198-017-0878-6
42. INE: Encuesta Europea de Salud en España 2014 (EESE-2014). Metodología http://www.ine.es/metodologia/t15/t153042014.pdf (2014). Accessed 9 November 2016
43. Cameron, A.C., Trivedi, P.K.: Microeconometrics. Methods and applications. Cambridge University Press, New York (2005)
44. OECD. Technical report of the Survey of Adult Skills (PIAAC). OECD Publishing, Paris (2013)

Appendix

Table A.1. Days of absenteeism in previous year: Alternative estimates (I)

	FMM – 2 components - Negative binomial 2		FMM – 3 components -		
	Component 1	Component 2	Comp. 1	Comp. 2	Comp. 3
Income2	0.455*	-0.289	0.640***	0.002	3.802***
Income3	0.595**	-0.042	-0.481***	0.333**	4.811***
Income4	0.861***	0.100	0.211	0.251*	4.569***
Income5	0.934***	-0.097	-0.181	0.299*	5.017***
Male	-0.434***	-0.465**	-0.401***	-0.206**	0.442***
Age	-0.110**	-0.106	0.297***	-0.108***	0.046
$Age^2/100$	0.082	0.129	-0.421***	0.121***	-0.114**
Married	-0.334**	0.092	0.208	-0.205**	-0.154
#children	-0.030	0.030	-0.352***	0.122**	-0.141**
Permanent job	0.954***	0.185	0.765***	0.475***	-0.294*
Part-time job	-0.262	-0.144	0.060	-0.207*	-0.168
Daily split shift	-0.084	-0.214	0.452***	-0.169**	-0.076
White-collar worker	-0.035	-0.189	1.286***	-0.237***	-0.003
Sedentary job	0.053	0.052	0.578***	-0.303***	1.063***
Fair health	-2.061***	-0.540	1.846***	-1.144***	-6.639***
Good health	-3.427***	-1.567***	-0.020	-2.057***	0.590***
Very good health	-4.258***	-1.788***	-0.515*	-2.477***	0.266
Physical illness	0.546***	0.078	0.759***	0.213**	0.086
Psychological illness	1.383***	0.540**	0.117	0.589***	-0.278*
Hospital admission	3.626***	0.244	3.621***	1.501***	2.938***
Day hospital admission	2.692***	0.432	0.414***	0.555***	4.387***
Constant	4.247***	7.033***	-7.652***	5.536***	-6.023***
Π_1	0.9271 (st. dev.: 0.016)		0.1316 (st. dev.: 0.017)		
Π_2			0.7246 (st. dev.: 0.023)		
Log L	-8738.344		-8543.006		
BIC	17887.78		17707.02		
# observations	6289		6289		

Note: * $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table A.2. Days of absenteeism in previous year: Alternative estimates (II)

	FMM – 2 components – Adding regional variables		FMM – 2 components- Comp. only differ in the intercept	
	Component 1	Component 2	Component 1	Component 2
Income2	1.214***	0.055	0.199*	0.199*
Income3	1.099***	0.291**	0.350***	0.350***
Income4	1.142***	0.291**	0.343***	0.343***
Income5	1.670***	0.206	0.349***	0.349***
Male	0.290***	-0.299***	-0.105*	-0.105*
Age	0.001	-0.029	-0.030	-0.030
Age²/100	-0.045	0.026	0.019	0.019
Married	-0.156*	-0.181**	-0.145**	-0.145**
#children	-0.333***	0.099**	0.012	0.012
Permanent job	1.003***	0.351***	0.469***	0.469***
Part-time job	1.065***	-0.353***	-0.047	-0.047
Daily split shift	0.686***	-0.287***	-0.063	-0.063
White-collar worker	0.530***	-0.195**	-0.046	-0.046
Sedentary job	0.612***	-0.167**	-0.059	-0.059
Fair health	-1.878***	-0.925***	-1.061***	-1.061***
Good health	-2.218***	-1.603***	-1.661***	-1.661***
Very good health	-2.626***	-1.939***	-2.043***	-2.043***
Physical illness	0.392***	0.237***	0.248***	0.248***
Psychological illness	-1.119***	0.736***	0.392***	0.392***
Hospital admission	4.308***	1.120***	1.637***	1.637***
Day hospital admission	-0.413***	1.074***	0.742***	0.742***
Aragón	0.969***	-0.303		
Asturias	0.256	0.184		
Baleares	1.420***	-0.259		
Canarias	-0.149	0.091		
Cantabria	1.269***	-0.632**		
Castilla y León	1.222***	-0.551**		
Castilla-La Mancha	-0.452	0.069		
Cataluña	0.681***	-0.012		
Valencia	0.418**	-0.189		
Extremadura	0.001	-0.242		
Galicia	-2.035***	-0.037		
Madrid	0.487*	-0.036		
Murcia	0.116	0.222		
Navarra	-0.864**	0.607***		
País Vasco	-1.854***	0.442***		
La Rioja	1.104***	-0.105		
Ceuta & Melilla	0.296	-0.040		
Constant	-1.552**	3.808***	3.422***	2.443***
Π	0.2796 (st. dev.: 0.026)		0.8870 (st. dev.: 0.013)	
Log L	-8591.0036		-8665.288	
LR test	45.32		103.25	
# observations	6289		6289	

Note: * p<0.1, ** p<0.05, *** p<0.01

Answers to Reviewer #1

We are very grateful for your comments and suggestions. In the second version of the paper we tried to follow your advice and those of the other reviewer and we think that the paper has substantially improved.

Below, we answer the major comments you raised in your report point by point.

Point 1: In the new version of the paper, we rewrote the Introduction and Concluding Remarks sections following your suggestions. We hope that now our contribution is clearer than in the first version.

Point 2: We dropped the equations about the labour/leisure choice.

Point 3: We realized that the description of Table 2 contents was a bit confusing in the first version. Table 2 shows the mean observed and predicted days of absence for the whole sample. The predicted days of absenteeism included in row 3 are computed as:

$$\hat{y_i} = \hat{\pi} f_1\left(y_i/\widehat{\theta_1}, x_i\right) + (1-\hat{\pi}) f_2\left(y_i/\widehat{\theta_2}\ x_i\right)$$

whereas the predicted days of absenteeism if component 1 or 2 (rows 4 and 5 of the Table) are computed as follows respectively:

$$\hat{y_i} = f_1\left(y_i/\widehat{\theta_1}, x_i\right)$$

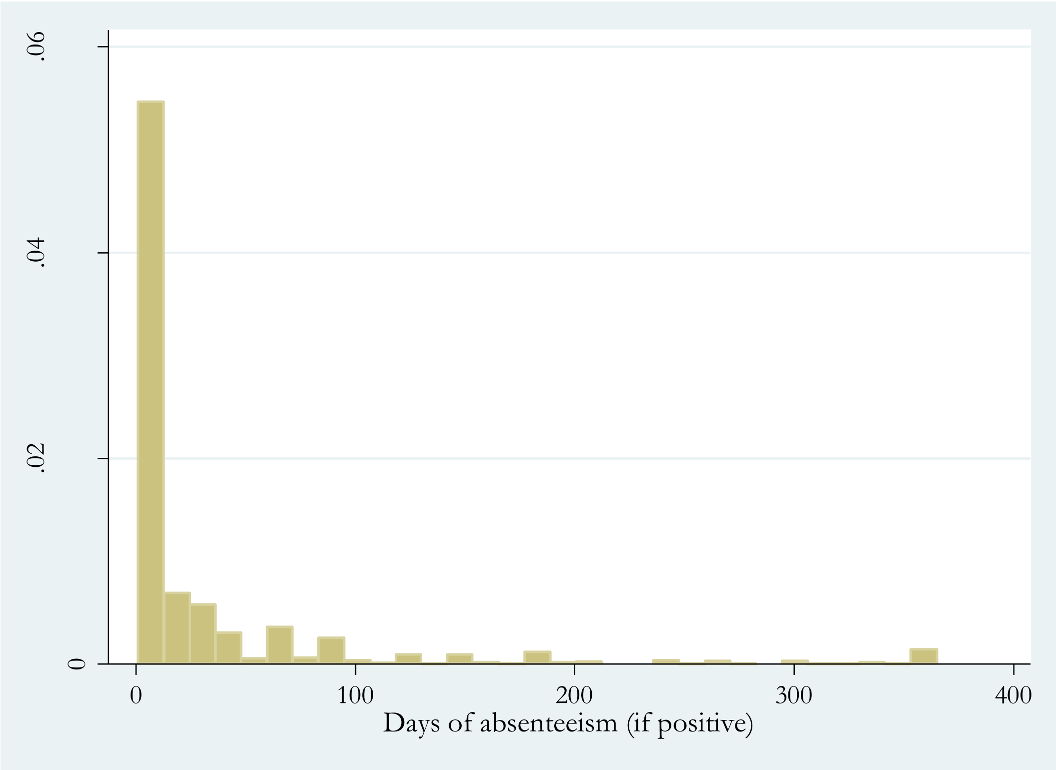
$$\hat{y_i} = f_2\left(y_i/\widehat{\theta_2}, x_i\right)$$

We have rewritten the paragraph preceding Table 2 and added a clarification in the title of the table as well as a caption to the table to explain the information included.

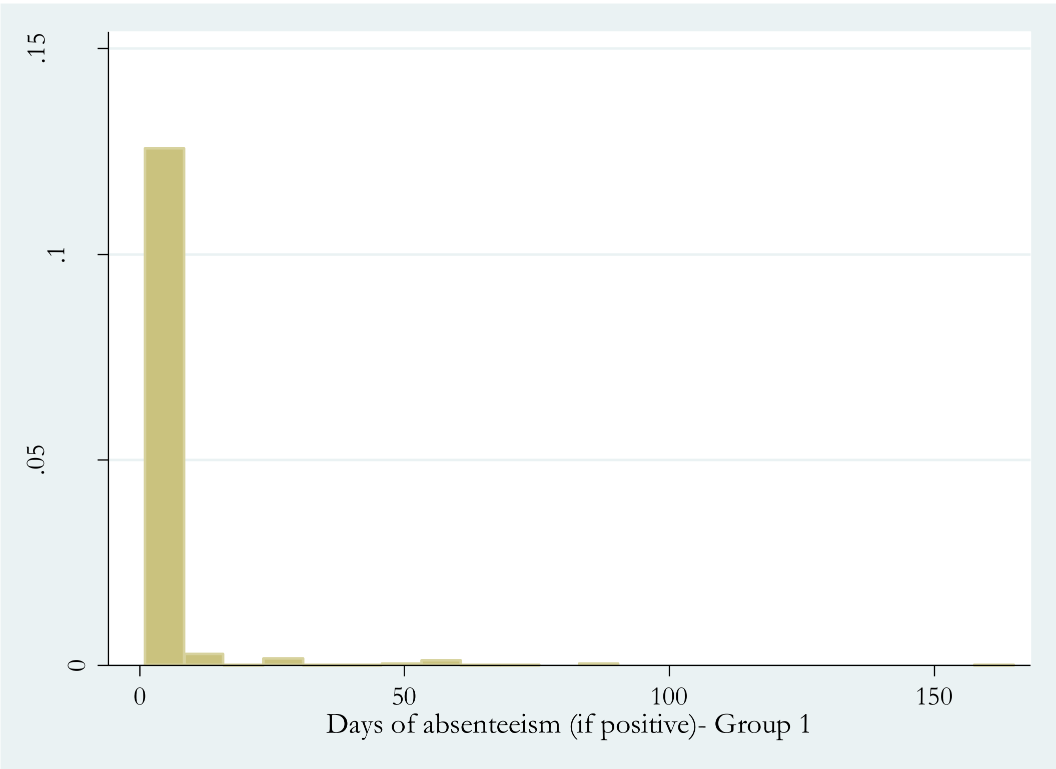
Instead, Table 3 shows the average of the observed days of absenteeism for the subsamples assigned to each group. Besides, additional calculations about the predicted days of absenteeism of each group have been placed at the end of this table. A new paragraph has been inserted after the table explaining and interpreting the added information.

Point 4: Graphs below show the distribution of days of absenteeism for the whole sample and for the two subsamples. Graph 2 illustrates that in the first group, absenteeism is concentrated in small values –about two thirds of the subsample have been absent from work between 1 and 3 days in previous year. However, in the second group the distribution is much more spread (Graph 3).

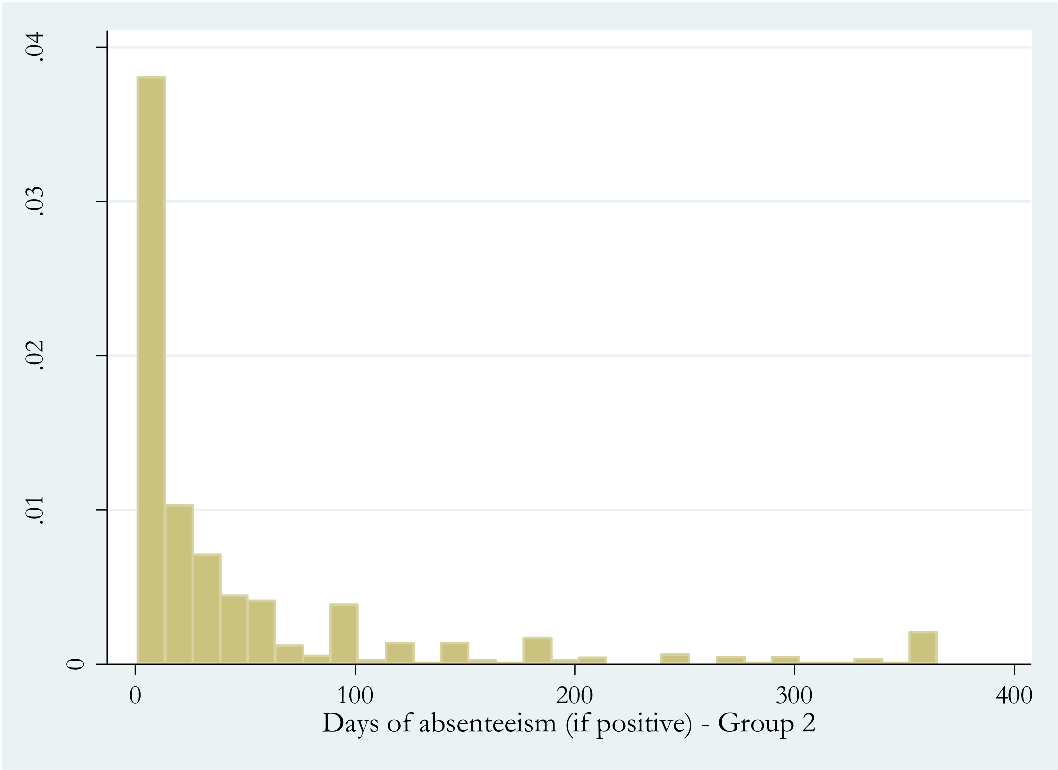
Graph 1. Histogram of days of absenteeism (for the subsample with positive values)



Graph 2. Histogram of days of absenteeism (group 1, subsample with positive values)



Graph 3. Histogram of days of absenteeism (group 2, subsample with positive values)



Point 5: In the first version of the paper, the marginal effects led us to think that voluntary absenteeism was more likely in group 1. However, after your comments and the ones of the other referee, we made some additional calculations in order to be more accurate in the interpretation of our results. In particular, we computed the predicted days of absenteeism of each group applying both their own behaviour equation and that of the other group and the results seem to indicate more voluntary absenteeism in group 2. The new information included at the end of Table 3 has therefore led us to change the interpretation of our previous results.

Point 6: Sick leave must be verified from the fourth day of absence in each spell. However, it is worth mentioning that our dependent variable measures the number of days of absenteeism during the past year, but we do not have information about the duration of each spell.

Point 7: We have expanded our comments about policy prescriptions in the Conclusion section.

If you consider any additional changes to be necessary, please do not hesitate to inform us.

Anwers to Reviewer #2

We are very grateful for your comments and suggestions. In the second version of the paper we tried to follow your advice and those of the other reviewer and we think that the paper has substantially improved.

Below, we answer to the major and minor comments you raised in your report point by point.

Major comments:

2.1: In the second version of the paper we have added an Appendix to include the estimates done to check the robustness of our specification, as well as the estimates when including regional dummies as covariates, as you suggested in point 3.3.

2.2: You are right in that there is no reason to state that there are only two groups of individuals according to their absenteeism behavior. In fact, we tried different number of components but, when comparing two *versus* three groups, the BIC led us to prefer two groups specification and, when we assume more than three groups, the model did not converge. We have added a comment in Section 4 about this issue.

2.3: We have linked our empirical results with the implementation of various potential strategies to reduce absenteeism and we have expanded our comments about relevant policy prescriptions in the Concluding Remarks section.

2.4: In the second version of the paper, we have included in Section 2 the two Spanish references you suggested.

2.5: We now mention presenteeism in the Literature Review section and added some references. We have also added a footnote in the Results section about the possible influence of presenteeism on our results.

Minor comments:

3.1: In the second version of the paper we have included a brief description of the Spanish sick leave system in the case of work-related accidents, which was missing in the first version.

3.2: We have dropped the equations related to the labour supply model.

3.3: As we have already mentioned in point 2.1, we also re-estimated the model incorporating regional dummies but the Likelihood Ratio test led us to prefer the initial specification. The results are shown in the Appendix (Table A.2).

3.4: We have added asterisk signs in Table 4.

3.5: In the first version of the paper, the marginal effects led us to think that voluntary absenteeism was more likely in group 1. However, after your comments and the ones of the other referee, we made some additional calculations in order to be more accurate in the interpretation of our results. In particular, we computed the predicted days of absenteeism of each group applying both their own behaviour equation and that of the other group and the results seem to indicate more voluntary absenteeism in group 2. The new information included at the end of Table 3 has therefore led us to change the interpretation of our previous results.

If you consider any additional changes to be necessary, please do not hesitate to inform us.