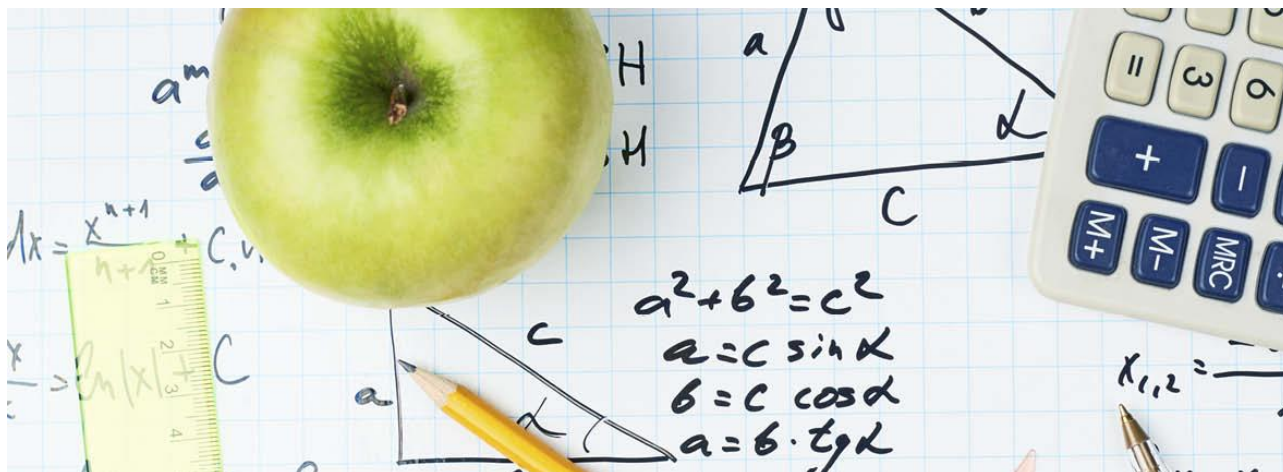


COHERE - Centre of Health Economics Research



Assessing the Propensity for Presenteeism with Sickness Absence Data

By

Sébastien RICHARD, Department of Business and Economics, University of Lille, France – CLERSE
Kristian Skagen, Department of Business and Economics, University of Southern Denmark – COHERE
Kjeld Møller Pedersen, Department of Business and Economics, University of Southern Denmark – COHERE
Benjamin Hover, Department of Business and Economics, University of Lille, France – CLERSE

COHERE discussion paper No. 1/2017

FURTHER INFORMATION
Department of Business and Economics
Faculty of Business and Social Sciences
University of Southern Denmark
Campusvej 55,
DK-5230 Odense M Denmark
www.cohere.dk
ISSN 2246-3097

Assessing the Propensity for Presenteeism with Sickness Absence Data

Sébastien RICHARD*

Department of Business and Economics,
University of Lille, France – CLERSE

<http://clerse.univ-lille1.fr/spip.php?article186&lang=fr>

sebastien.richard@univ-lille1.fr

Kristian SKAGEN

Department of Business and Economics,
University of Southern Denmark – COHERE

Kjeld Møller PEDERSEN

Department of Business and Economics,
University of Southern Denmark – COHERE

Benjamin HOVER

Department of Business and Economics,
University of Lille, France – CLERSE

Abstract

Presenteeism occurs when an employee attends work while sick or unwell. It is a major Human Resource and organizational issue: in addition to productivity losses, presenteeism is believed to increase sickness absence and decrease self-rated health. However, by its very nature, presenteeism cannot be monitored in the same manner as sickness absence. We show how the probability of presenteeism can be estimated from simple absence data by means of a zero-inflated binomial regression analysis (ZINB). The approach is validated on a Danish data set that contains self-reported sickness absence and presenteeism, whereas causality and reliability are verified by conducting Monte-Carlo simulations.

The objective of paper was to explore how far the traditional but costly tool used to assess presenteeism behaviour, a questionnaire, could advantageously be replaced by a statistical approach that relies on easily available information on sickness. We show that the ZINB model captures presenteeism well via the inflation process and delivers insight on both absenteeism and presenteeism. Using Monte Carlo simulations, we further highlight that the model can be used to compute a global indicator, propensity for presenteeism, even when important assumptions are violated.

Key words: Presenteeism, sickness absence, ZINB

JEL classifications: I10 J22 J28

1. INTRODUCTION

Over the last fifteen years, presenteeism has emerged as an important organizational phenomenon. Presenteeism occurs when an employee attends work while sick or unwell. In other words, the employee renounces sickness absence to go to work. Many studies emphasize the potentially negative outcomes of this kind of behaviour: deteriorated self-reported health (Bergström et al., 2009a; Kivimäki et al., 2005), subsequent increased sickness absence (Bergström et al., 2009b; Hansen & Andersen, 2009), contagion in the workplace (Kumar et al., 2013) and increased health-related costs (Goetzel et al., 2004; Schultz et al., 2009).

Absenteeism remains the main concern for companies; nevertheless, owing to its potentially negative consequences, much more focus should be on presenteeism. However, the phenomenon remains nearly unexplored scientifically. Although absenteeism is measured routinely by employers, the same does not hold for presenteeism, probably because costs, i.e. decreased productivity, are not directly visible. Measuring presenteeism requires surveys that are too expensive for firms to be used on a regular basis.

These observations are the starting point for this study. Our objective is to develop a quantitative tool designed to assess the propensity for presenteeism simply based on individual sickness absence records. To our knowledge, such a method has not been developed in the academic literature thus far; however, we show that a “Zero-Inflated Negative Binomial” (ZINB) model is well suited to provide an accurate estimate of presenteeism.

After a presentation of its methodological underpinnings, we propose an empirical validation application using a Danish representative dataset. This dataset contains relevant information for a direct test of the proposed methodology because it has self-reported information on both

presenteeism and sickness absence. Thus, it is possible to provide strong evidence on the relevance of our tool to assess presenteeism.

2. MEASURING PRESENTEEISM

A paper by Johns (2010) summarizes the state of knowledge regarding presenteeism and shows how the definition of presenteeism adopted here, namely, attending work during an illness event that may have justified sickness absence, progressively has become a standard definition. We do not return to this definition. Instead, focus will be on how presenteeism is measured in research articles.

Micro-data is often used in research on absenteeism, while studies on presenteeism, by their nature, nearly always rely on self-reported interview data. This reliance is to such an extent that certain authors, e.g. Claes (2011), claim that only self-reported data allow presenteeism to be measured because only employees know whether they have actually worked while sick.

To our knowledge, extremely few studies exploit absence data to study presenteeism. Kivimäki et al. (2005) identify presenteeism by combining health information and absence data; individuals who claim to be unhealthy and, furthermore, have not been absent for three years are assumed to exhibit presenteeism behaviour. This identification is of course presenteeism by definition, but it is not validated by empirical observation. McKevitt et al. (1997) also suggested approaching presenteeism by an “artificially low” absence level. It is in view of this very sparse literature that the method taken here represents a novel approach to the identification of presenteeism.

The most widely used method to collect information on presenteeism remains surveys. Employees’ presenteeism is evaluated through questions similar to the following:

- “How many days did you go to work in the past six months even though you were sick or not feeling well?” (Johns, 2011)

Direct questions such as the one above include at least three elements: a description of presenteeism behaviour (working through illness), a recall period (usually 12 months, occasionally less) and a “unit of measurement” (usually a number of episodes, or occasionally a number of days). The response generally takes a categorical form: from “never” to “11 times or more” (e.g. Hansen and Andersen, 2009).

Many psychometric scales have been used to measure presenteeism (Lofland et al., 2004), among which are the “Work Limitations Questionnaire” (Lerner et al., 2001) or the “Stanford Presenteeism Scale” (Koopman et al., 2002). These scales often originate from the medical field or occupational psychology and adopt a specific approach to presenteeism, in terms of lost productivity related to health problems.

In all cases, these self-reporting tools are difficult to implement in companies. The development of an alternative statistical estimation strategy is the goal of this article.

3 METHODOLOGICAL FRAMEWORK

Statistically, presenteeism is invisible for firms with only a possible indication through lower work performance, which remains difficult or impossible to monitor. However, presenteeism may be indirectly present in the individual sickness absence records in which it may manifest itself through a higher frequency of nil values: people who are never absent. Therefore, the challenge is to develop a statistical strategy to sort nil values and to distinguish presenteeism in the distribution of the sickness absence variable.

3.1 Modelling absence data

To begin modelling, we need a relevant statistical distribution for sickness absence. Our dependent variable, the number of days absent in a year, is a discrete count variable. Thus, a Poisson distribution would be suitable because it is commonly used to “count” relatively rare events that occur during a given period. However, the Poisson assumption of equality between

mean and variance is often not tenable. Inter-individual heterogeneity results in statistical over-dispersion and variance greater than the mean. When the sources of this over-dispersion are known and observed, e.g. job status or demographic variables, the Poisson model continues to remain an option. However, when they are not observed, for instance psychological attitudes, or health status, the Poisson distribution needs to be improved.

The traditional means to improve the model is to modify the Poisson model by including a heterogeneity parameter to capture over-dispersion. For reasons described in the statistical/econometric literature (Winkelmann, 2008; Hilbe, 2011), this parameter is assumed to follow a Gamma distribution. The resulting Poisson-Gamma mixture is the Negative Binomial distribution. This distribution allows the data to be over-dispersed, which is a very useful property when fitting classical absence data in empirical applications (Barnby et al., 2001; Frick and Malo, 2008; Jensen and MacIntosh, 2007; Winkelmann, 1999).

However, there is a second form of heterogeneity that characterizes sickness absence data. When they are sick, employees can behave in one of two different ways: take days off to recover or decide to attend work. Therefore, the over-representation of nil values in the sickness absence distribution may, to a certain extent, be a manifestation of presenteeism.

Of course, this supposition does not mean that presenteeism is solely characterized by nil values for sickness absence. Certain employees, who usually attend work while sick, may decide to remain at home during a more severe episode of sickness. However, these individuals generally behave similar to those who exhibit nil values; their respective profiles are similar and could be determined by focusing on those with nil values.

3.2 Modelling excess zeros

In sum, the nil values in the sickness absence distribution are generated by two different processes. In the first case, zeros result from a “normal” count data process, which produces both nil and positive values. These values are so-called “incidental” zeros (Winkelmann, 2008, p.189); the employee was not absent because he was not sick. However, if sickness had occurred, he would have exhibited a positive value. In the second case, zeros are generated by a specific process that solely produces values. Despite sickness, the employee voluntarily decided to attend work, hence generating a zero absence value. These nil values are called “strategic” because they result from a strategic behaviour by employees.

Zero-Inflated count data models enable the analysis of these two situations. To our knowledge, very few papers that study sickness absence have exploited them without clearly calling presenteeism by its name. Among them, Frick and Malo (2008, p.517-18) assume that the second part of the Zero-Inflated model describes “some individuals [that] have zero absence days because they follow an absolute rule of no voluntary absenteeism”. In this paper, we will show that these individuals exhibit presenteeism behaviour.

Two equations are jointly estimated in the Zero-Inflated Negative Binomial (hereafter ZINB) model: a counting equation, modelling the multi-value process (“incidental” nil and positive values), and an inflation equation, modelling “strategic” nil values. The standard Negative Binomial model is used to estimate the counting equation and allows us to describe the predictors of sickness absence and to compute an expected absence value. However, a joint process is introduced, which allows a specific sorting of nil values. The estimation of this second equation is based on a logistic model, whose explanatory variables can be different from the first set (Winkelmann 2008). These variables provide information on the causes of presenteeism and allow the computation of an individual presenteeism probability.

The probability distribution of a Zero-Inflated model is:

$$(1) P(Y_i = y_i) = \begin{cases} p_i + (1 - p_i)f(0) & \text{if } y_i = 0 \\ (1 - p_i)f(y_i) & \text{if } y_i > 0 \end{cases}$$

with y_i is the observed number of days absent for individual i .

If $y_i = 0$, two situations are considered. p_i is the probability for individual i to be in a “perfect state” in which only zero values are generated (presenteeism). p_i depends on a specific vector of independent variables and is estimated with a logistic function (“inflation equation”). We consider that this set of variables includes the explanatory factors for presenteeism, and that p_i is the presenteeism probability for individual i . $(1 - p_i)f(0)$ covers the situation in which zeros are “incidental”; it is estimated with a Negative Binomial function.

3.3 Presenteeism Determinants

Because presenteeism may be a response to (potentially implicit) organizational demands, the work context the individual is embedded in can play an important role in the decision to attend work while sick. As organizational variables, we include occupational sector (Aronsson et al. 2000, 2005), team responsibility (Hansen and Andersen, 2008; Gosselin et al., 2013) and job satisfaction (Krohne and Magnussen, 2011; Johansen et al., 2014) in the regression.

Considering presenteeism as an individual decision that favours organizational demands to the detriment of private demands, one would also expect that employees could be differently affected according to their individual characteristics. Gender (Aronsson and Gustafsson, 2005; Johansen, 2012), age, education (Taloyan et al. 2012) and number of young children at home (Hansen and Andersen, 2008) were also put into the model. An attitudinal variable, ‘pride in presenteeism’ is also included as a novelty. For the expected signs and reasoning for inclusion of both organizational and individual variables, consult the references provided.

4 ESTIMATION OF ZINB

To show the practical relevance of the ZINB model, it is applied to a Danish dataset. This unique dataset contains information on employees (demographics, work-related variables, and sickness absence records) that are commonly available in HR departments and that are considered as predictors of presenteeism. The innovative aspect here is that we also have self-reported data on health, presenteeism and presenteeism attitudes. Thus, we will be able to cross this information with our estimated presenteeism probability and assess its reliability.

4.1 Data

The Danish Presenteeism Questionnaire was used for a cross-sectional survey of the occupationally active Danish population. The questionnaire was collected in December 2010 through an Internet-based survey aimed specifically at presenteeism, sickness absence, and health insurance. The effective realized sample size was 4,060. Of the respondents, 1,257 were excluded because they had not held a job for at least the previous 12 months.

The questionnaire consisted of 56 questions: 10 questions on socioeconomic variables; 14 questions on workplace, type of work and satisfaction; 5 questions on health and illness; 14 questions on health insurance and employer-paid health schemes; and 13 questions on attitudes towards absence, number of presenteeism and sickness absence days.

The study sample was representative of the Danish working population. Approximately 86% of the adult Danish population has Internet access at home. The remaining 14% are pensioners. Because the sample solely included occupationally active adults aged 65 or less, the use of the Internet should not influence representativity. Compared to the employed Danish background population, respondents from the capital region are slightly overrepresented, and respondents in the 18–25 age bracket are slightly underrepresented (7 percentage points).

4.2 Construction of the sickness absence and presenteeism variables

The data regarding sickness absence and presenteeism originates from two questions:

- The first question asks how many **times** the respondents were absent or worked while sick during the last 12 months. The answer is categorical (0, 1, 2-3, 4-5, 6-10, 10 and more);
- A second question, solely for people who declared a positive value, requests **number of days** in the last 12 months (3 months for presenteeism).

From these two answers, we develop a self-reported presenteeism variable: zero if the answer is “none” in the first question, otherwise we use the value of the second question. We removed 69 individuals who did not answer the question concerning sickness absence because this is our dependent variable. Based on the 3,981 remaining individuals, Figure 1 presents the distributions of sickness absence and presenteeism.

Figure 1 to be inserted here

Concerning the sickness absence distribution, we observe that the zero frequency is low (31.2%), and that frequency peaks are observed for values (2, 3, 5, 10, 20 and 30). This result is due to a common rounding effect (refer to section 4.4). On average, employees were absent 6.37 days in the previous year; the figure is higher for women and for employees over 50 years.

Among the respondents, 46.9% reported that they had never experienced presenteeism in the previous three months, and 24.7% reported no presenteeism at all during **the previous year**. Employees reported in average 2.47 days of presenteeism (2.67 for women) in the last three months.

Table I to be inserted here

4.3 Model estimation

The usual approach when working with micro-data is to use ZINB regressions on a truncated sickness absence distribution (Bierla et al., 2013; Huver et al., 2014) by excluding individuals who had the longest spells of sickness absence. Mainly because presenteeism becomes irrelevant beyond a given illness threshold for severe diseases, there is at least an incompressible part of absence. Usually the threshold is 45 days (also used hereafter); however, the results do not fluctuate very much according to the truncation level, particularly concerning the inflation equation. This finding leaves a total 3,898 individuals with a zero frequency of 31.8%. The 83 excluded individuals (2.1%) cumulated 9,010 days of sickness absence (35.5%). Not surprisingly, 66% of these individuals perceive their own health to be fair, poor or very poor.

We computed four models with step-by-step specifications. Model 1 solely includes HR-like explanatory variables; this specification is close to the regressions classically run on firm micro-data. We progressively add different variables that are commonly not available for firms: perceived health (model 2), job satisfaction (model 3) and attitudes towards presenteeism variables (model 4). The comparison of these four models will allow us to assess our methodology.

The estimation outputs are presented in Table II. The first block is the counting equation, which contains explanatory variables for sickness absence. The second block is the inflation equation, expressing the presenteeism determinants. Only the significant variables are retained.

Table II to be inserted here

The dispersion parameter directly indicates the degree of over-dispersion in the model (Hilbe, 2011). When the parameter tends towards zero (mean and variance converge), the model returns to a Zero-Inflated Poisson distribution. Although significantly above 0, the parameter is very low and decreases at each step. The decline between model 1 and model 2 is large and

supports the idea that over-dispersion is due to a significant degree to health status. The difference between model 3 and model 4 is also noticeable and supports the idea that the phenomenon estimated in the inflation equation is sensitive to judgement regarding how one must behave when sick.

4.4 Goodness of fit

Figure 2 shows the four models to be well adjusted to empirical data. A closer analysis reveals some differences, which can be explained by classical features of absence data. The overestimation on values 1 and 9 (and potentially 4) is a common outcome. For a minor health event, doctors rarely prescribe a sick note for a single day, but a note is typical for 2 or 3 days. Doctors also often prescribe 10 days off (2 weeks) instead of 9, and 5 days off (a week) instead of 4. Consequently, for values 1, 4 and 9, frequencies are generally over-estimated, whereas for values 2-3, 5 and 10, they are generally under-estimated. Of course, the theoretical distribution cannot explain these specific details; this leads to slight differences between the empirical and the estimated distributions, which, however, never exceeds 0.04.

Figure 2 to be inserted here

To assess the explanatory power of the four models, their respective log-likelihood were compared to the log-likelihood of the model without explanatory variables (LL=-9677). All the models score better, of course, and the explanatory power increases with the provision of additional information (increased log-likelihood and decreased AIC). The differences between predicted and observed frequencies are very low for the first model and tend to slightly decrease. The residual differences appear to be mainly due to the previously discussed features of absence data.

Moreover, the results of Vuong and Schwarz tests on these four models consistently show that a ZINB regression should be preferred to a standard negative binomial regression, despite the

low zero frequency. The models score better when considering that a part of the nil values are generated by a “perfect state”, where employees always attend.

4.5 Presenteeism probabilities

In addition to an assessment of the presenteeism factors, the ZINB model allows the computation of an individual presenteeism probability, which is most importantly in the present context. The global distribution of these probabilities is presented in Figure 3; there are no significant differences between the four models (refer to Table III). One can observe that presenteeism probabilities are low, probably because of the low zero frequency. In Model 1, the average probability is 0.143, and the median is 0.1. However, the inclusion of attitudinal variables enhances the values in the last quartile.

Figure 3 to be inserted here

Table III to be inserted here

As noted previously, the fact that the presenteeism profile computation is based on nil values does not imply that only individuals with no absence exhibit high presenteeism probabilities. We find that among individuals with probabilities higher than the 75th percentile in Model 1 (0.196), 53.24% have also a positive absence value.

5 FINDINGS AND RELEVANCE OF ANALYSIS

In this section, we discuss the results obtained using a ZINB model to assess presenteeism. Generally, they are consistent with previous findings but allow a deeper analysis of the phenomenon owing to the method deployed and the variables available in the dataset.

Comments on estimation results

Gender. Whatever the model, when controlling for health status, we find that women experience longer spells of absence than men. Models also predict that, despite having a

slightly larger number of presenteeism days (2.67 vs. 2.29), women are less likely to experience presenteeism, *ceteris paribus*.

It appears that women attempt to restrain organizational demands, to reconcile their vocational and private lives and to preserve their health; this may be because they commonly take over a larger portion of household responsibilities, particularly childcare.

Age. Older employees are more likely to exhibit presenteeism, although the length of their sick leave is commonly longer when absence actually occurs. There are several explanations for this. One could think that older employees are more worried about finding a new job if they become unemployed (Dew et al., 2005). To avoid that worry, the employees could attempt to convince their employer of their stamina. Considering that presenteeism conveys a positive image, older employees attempt to compensate for these longer absence periods by coming to work when sickness is milder. Presenteeism behaviour of older employees could also be explained from a generational perspective; they make a point of always being on the job and appearing as trustworthy workers.

General health. As indicated above, in the Negative Binomial model, a random variable is used to model unobserved inter-individual heterogeneity. Our assumption was that these unobserved differences were partly due to the lack of information regarding health; this calls for the inclusion of health information (“PERCEIVED HEALTH” and “LONG-TERM SICKNESS”) solely in the counting equation. Interestingly, when we do so, the coefficient of the age variable decreases sharply and becomes non-significant. The logical conclusion is that, in Model 1, the health effect was largely captured by age. When a better proxy of health is included, the relation between age and sickness absence is much weaker. Of course, the perceived health variable (ranging from 1, very good health, to 5, very poor) is positively correlated with absence: the poorer the health level, the longer the duration of absence. Accordingly, people who declare long-term sickness also have, as expected, longer absences.

Education. We found that persons with more than four years' higher education show both lower absence durations and lower presenteeism propensities. The effect on absence was expected; higher education leads to better and less health-deteriorating jobs. Conversely, the effect on presenteeism was unexpected. Higher educated employees may more frequently develop a balance between vocational and private lives or may be more aware of the risks on health of presenteeism. Organizational demands can also be lower for these employees, particularly because they more frequently occupy jobs in which the workload can be postponed in case of absence.

Team responsibility. We find shorter absences and a larger presenteeism probability for persons with team responsibility. However, it varies with the size of the team. Those in charge of 26 subordinates or more exhibit very high presenteeism level. However, the effect decreases when the number of subordinates decreases. The fact that other employees depend on them is an incentive for supervisors to be present at work despite being sick.

When controlling for presenteeism attitudes (model 4), these coefficients significantly decrease. Exhibiting a presenteeism attitude is very common for managers and is probably expected; the effect of being a manager, *ceteris paribus*, decreases but remains significant, particularly for large teams.

Occupational context. Despite the fact that the effect is small, we found that people working in the private sector, *a fortiori* in small firms (the effects are potentially cumulative), have shorter absence spells. They also have a slightly higher propensity for presenteeism but the effect disappears when controlling for attitudes. It suggests that these employees exhibit higher presenteeism levels because they feel more frequently "pride in presenteeism", most notably when they know (in firms with 1 to 9 employees) that their absence will increase the workload of the colleagues.

Self-employed workers work more frequently through illness (but do not record shorter absences); they are often alone when conducting their activity and therefore encounter strong external demands, e.g. from customers. We could also expect an endogenous effect; people who were accustomed to work sick (when they were an employee) are more likely to become self-employed.

Job satisfaction. The model predicts that, as job satisfaction increases, the propensity for presenteeism decreases. No suitable explanation is available. One may speculate that people who encounter organizational demands are more likely to experience presenteeism and feel less satisfied.

5.1 Presenteeism attitudes

Presenteeism attitudes. As stated previously, the inflation equation directly estimates the presenteeism probability, and we clearly expect attitudes towards presenteeism to be significant determinants. To test this, three presenteeism “attitude” variables were included in model 4. The first one (“PRIDE IN PRESENTEEISM”) is a dummy indicating that respondents agree (or strongly agree) with the following statement: “I take pride in coming to work no matter how I feel”. This question is very differentiating because it is directly related to an individual action (coming to work); it could be viewed as a strong validation of the model relevance to assess presenteeism. Similarly, the second item (“ATTEND IF A BIT SICK”) indicates that the individuals think that “it is okay to go to work, although [(s)he is] a bit sick”. The third item (“ATTEND WITH 38.2°”) is a dummy that is marked with a 1 if the interviewee thinks it is unfair to take days off when “F. has a temperature of 38.2° and feels a bit uncomfortable.”

As expected, all three variables are positively correlated with the captured phenomenon. This provides us robust evidence that the phenomenon we measure is at least strongly related to

presenteeism; when people do approve of working through illness, their propensity for presenteeism generally increases.

6. PREDICTING PRESENTEEISM

Because it both contains information on sickness absence and presenteeism, the Danish Presenteeism Questionnaire data are relevant for a direct evaluation of our model. Sickness absence records are required for an assessment of the phenomenon via the ZINB model. Presenteeism self-reported data allow us to confront different measurement approaches and to establish how accurately our model identifies presenteeism at minimum cost and information.

If the presenteeism probability is as an accurate predictor of the likelihood an employee would work through illness, we expect this indicator to be able to predict episodes of presenteeism. As with sickness absence, presenteeism is a discrete count variable, which is over-dispersed. Therefore, a Negative Binomial distribution should fit the data well.

However, a problem should be anticipated for employees reporting no presenteeism episodes.

A zero value can be explained by two situations.

1. An individual who was not sick and, thus, had no opportunity to exhibit presenteeism;
or
2. An individual who was sick and did not work during illness.

Ideally, observations corresponding to the first case should be excluded because we do not really know if people would have exhibited presenteeism behaviour if they had been sick. Therefore, we expect our model to score better if nil values are previously sorted. Unfortunately there is no means to do that; however, we can implement different strategies.

We can indiscriminately use all values, including nil values, and thus disregard the problem (scenario 1). We can also take only positive values, and thus exclude all nil values (scenario

2). However, in this awkward case, we would not be able to check the ability our model to identify employees who do not work through illness. Finally, we can build a crude filter by excluding individuals with a good general health level who also have a nil value of absence (scenario 3). Indeed, they self-reported a good health status and no absence. Hence, it is more likely that they do not experience illness and do not have the opportunity to exhibit presenteeism.

We estimated 12 econometric regressions, one for each scenario and for each ZINB model we previously estimated. As propensity for presenteeism appears as a global predictor of the behaviour, we only estimated univariate regressions with presenteeism as the dependent variable. Owing to the severe drawbacks of scenario 2, we mainly focus on scenarios 1 and 3.

Table IV to be inserted here

Regardless of the selected scenario and the selected model, the relation is always positive and (strongly) significant; the higher the estimated probability, the higher the self-reported presenteeism. Placing a crude filter on nil values enhances the significance of the variable but does not affect the estimation (apart from the intercept). However, when providing lower information level (model 1) and when retaining the simplest strategy (scenario 1), the quality of the model is wholly satisfactory (refer to Figure 4). Indeed, despite the traditional features of sickness absence data that also occurs for presenteeism (over-estimation of values 1, 4 and 9; under-estimation of values 2-3, 5 and 10), it is clear that the model performs well when attempting to predict presenteeism.

Figure 4 to be inserted here

7. MONTE-CARLO SIMULATIONS

Another way to verify that the presenteeism probabilities really express a propensity for presenteeism is to conduct Monte-Carlo simulations with the main purpose of analysing the relevance of the ZINB model when its fundamental assumptions are not fully satisfied.

The simulations are developed in three steps. The first is the simplest case (0); the objective is to observe how the model addresses unobserved heterogeneity when individuals who adopt presenteeism always exhibit nil values of absence. We then successively introduce two modifications making the estimation less straightforward. First, individuals experiencing severe sicknesses are not allowed to exhibit presenteeism (0). Secondly, presenteeism does not exclusively result in nil values (0).

We simulate a dataset with 10,000 agents. For each of them, we randomly generate three independent variables β_i (two quantitative and one dummy). Based on these variables, we set two equations generating an individual absence mean (λ_i) and a presenteeism propensity (p_i):

$$\lambda_i = \exp(-1.3 - 0.6\beta_1 + 0.04\beta_2 + 0.045\beta_3)$$

$$p_i = \frac{\eta_i}{1 + \eta_i} \text{ with } \eta_i = \exp(-1.5 - 0.3\beta_1 + 0.04\beta_2 + 0.04\beta_3)$$

The coefficients of these equations, close to common findings in the literature, fluctuate randomly (for each agent) in a plus or minus 5% range.

Next, according to the “theoretical” presenteeism probability p_i , the individuals are sorted into two categories: those exhibiting a “presenteeism attitude” and others. To sort, the presenteeism probability is compared to a random value following a uniform distribution (ranging from 0 to 1); if the probability is higher than the random value, the agent exhibits presenteeism. We initially assume that these individuals always exhibit a nil absence value, whereas for others, a Poisson distributed absence value (with mean λ_i) is generated.

Each step of the simulation is repeated 100 times. In this section, we present the average outputs of the ZINB estimations (case of strong unobserved heterogeneity illustrated). As an example, Figure 5 shows how the model performs in the three coming steps, with the coefficient of parameter β_2 (set at 0.04) in the inflation equation. This is a quantitative variable simulating the effect of age in an empirical dataset (following a uniform distribution and ranging between 20 and 65). One can observe that the coefficient is, on average, perfectly estimated in the first step and slightly understated when the assumptions are relaxed (step 2 and 3).

Figure 5 to be inserted here.

7.1 Step 1: estimation with unobserved heterogeneity

When studying sickness absence, some individual features, such as health status, should obviously be taken into account. Unfortunately, this information is not easily available, and the key issue is to know if the model works efficiently when sources of heterogeneity remain unobserved.

To model health in our simulation we multiply the individual absence mean by a positive Gamma-distributed random value (with mean equal to one). Thus, when the random value (which remains of course unobserved in the estimation) is small, the individual is considered as healthy and his absence value is lower. When the random value is large, his health status is poor and his absence value is higher.

The model was able to give an accurate estimation of the presenteeism probability. As shown in Figure 6, the estimated probabilities from the ZINB model perfectly overlap the theoretical probabilities. The model is built to address unobserved heterogeneity, and it manages to do so as expected.

Figure 6 to be inserted here.

7.2 Step 2: effects of severe sicknesses

We now extend the standard model by adding a new feature. People who are in very poor health are considered unable to work sick. At this point, we focused on the 20% of the population whose health is the poorest. These individuals were not allowed to “choose” presenteeism, regardless of their theoretical probability; their absence was instead generated with the common Poisson process. This process leads the estimated probability distribution to remain perfectly ordered (Figure 7) but also to deviate slightly and to be under-estimated, mainly for larger values. However, the conclusions of the model are relevant, and the diagnosis remains accurate when the unobserved heterogeneity is large.

Figure 7 to be inserted here.

7.3 Step 3: presenteeism and positive absence values

In the first step, individuals who performed presenteeism systematically exhibited a nil value. Zero-Inflated models exploit these “strategic” nil values to build a statistical profile of presenteeism so that they appear as the key issue of our methodology. However, one could actually concede that presenteeism does not systematically lead to a zero absence but only shortens the absence period (more or less, depending on the presenteeism propensity). Therefore, the question is if the model remains relevant in this case.

In this third step, we focus on the subpopulation that presents a presenteeism profile. For 20% of these individuals with the poorest health status, the absence value is no longer zero, but the Poisson generated absence value. However, this absence value is reduced (divided by 3) to simulate an early return. Once again, the probabilities remain perfectly ordered but are slightly underestimated by the ZINB model. However, the adjustment remains relevant (Figure 8) and provides us a meaningful (although moderately underrated) assessment of the employees’ presenteeism probability.

Figure 8 to be inserted here.

These three sets of simulations emphasize the relevance of a ZINB model to describe presenteeism behaviour. In the presence of strong unobserved heterogeneity, the model remains very accurate. Additionally, when the model's fundamental assumptions are relaxed, a slight underestimation will be noticed for the highest values.

8. CONCLUSION

The objective of this article was to explore how far the traditional but costly tool used to assess presenteeism behaviour, a questionnaire, could advantageously be replaced by a statistical approach that relies on easily available information on sickness. We show that the ZINB model captures presenteeism well via the inflation process and delivers insight on both absenteeism and presenteeism. Using Monte Carlo simulations, we further highlight that the model can be used to compute a global indicator, propensity for presenteeism, even when important assumptions are violated.

Due to data from the Danish Presenteeism Questionnaire, we were able to confront this indicator with actual data on presenteeism while controlling for a very wide range of variables, most notably health status and attitudes towards presenteeism. The conclusion is that the propensities are reliable and can be used to impute presenteeism; this is important for two reasons. First, companies can obtain estimates of presenteeism from absence data in a fairly simple manner, including when only HR variables are available. Second, costs of presenteeism may be computed (Skagen, 2015).

This work allows for more massive and reproducible studies on presenteeism. As presenteeism becomes a major organizational issue, it appears useful to conduct large-scale research and explore new hypotheses on its determinants and consequences.

REFERENCES

- Agudelo-Suárez AA, Benavides FG, Felt E, Ronda-Pérez E, Vives-Cases C, García AM. 2010. Sickness presenteeism in Spanish-born and immigrant workers in Spain. *BMC Public Health* **10**: 1-7.
- Aronsson G, Gustafsson K. 2005. Sickness presenteeism: prevalence, attendance pressure factors, and an outline of a model for research. *Journal of Occupational and Environmental Medicine* **47**(9): 958-966.
- Aronsson G, Gustafsson K, Dallner M. 2000. Sick but yet at work: an empirical study of sickness presenteeism. *Journal of Epidemiology and Community Health* **54**: 502-509.
- Baker-McCleary D, Greasley K, Dale J, Griffith F. 2010. Absence management and presenteeism: the pressures on employees to attend work and the impact of attendance on performance. *Human Resource Management Journal* **20**(3): 311-328.
- Barmby T, Nolan M, Winkelmann R. 2001. Contracted workdays and absence. *The Manchester School* **69**(3): 269-275.
- Bergström G, Bodin L, Hagberg J, Aronsson G, Josephson M. 2009a. Does sickness presenteeism have an impact on future general health? *International Archives of Occupational and Environmental Health* **82**(10): 1179-1190.
- Bergström G, Bodin L, Hagberg J, Aronsson G, Josephson M. 2009b. Sickness presenteeism today, sickness absenteeism tomorrow? A prospective study on sickness presenteeism and future sickness absenteeism. *Journal of Occupational and Environmental Medicine* **51**(6): 629-638.
- Bierla I, Huver B, Richard S. 2013. New evidence on absenteeism and presenteeism. *International Journal of Human Resource Management* **24**(7): 1536-1550.
- Böckerman P, Laukkanen E. 2009. What makes you work while you are sick? Evidence from a survey of workers. *European Journal of Public Health* **20**(1): 43-46.
- Caverley N, Cunningham JB, MacGregor JN. 2007. Sickness presenteeism, sickness absenteeism, and health following restructuring in a public service organization. *Journal of Management Studies* **44**(2): 304-319.
- Claes R. 2011. Employee correlates of presenteeism: a study across four European countries. *Work and Stress* **25**(3): 224-242.
- Demerouti E, Le Blanc P, Bakker AB, Schaufeli WB, Hox J. 2009. Present but sick: a three-wave study on job demands, presenteeism and burnout. *Career Development International* **14**(1): 50-68.
- Dew K, Keefe V, Small K. 2005. 'Choosing' to work when sick: workplace presenteeism, *Social Science and Medicine* **60**(10): 2273-2282.
- Elstad JI, Vabø M. 2008. Job stress, sickness absence and sickness presenteeism in Nordic elderly care. *Scandinavian Journal of Public Health* **36**(5): 467-474.
- Frick B, Malo MA. 2008. Labor market institutions and individual absenteeism in the European Union: the relative importance of sickness benefit systems and employment protection legislation. *Industrial Relations* **47**(4): 505-529.

- Goetzel RZ, Long SR, Ozminkowski RJ, Hawkins K, Wang S, Lynch W. 2004. Health, absence, disability, and presenteeism cost estimates of certain physical and mental health conditions affecting U.S. employers. *Journal of Occupational and Environmental Medicine* **46**(4): 398-411.
- Gosselin E, Lemyre L, Corneil W. 2013. Presenteeism and absenteeism: differentiated understanding of related phenomena. *Journal of Occupational Health Psychology* **18**(1): 75-86.
- Hansen CD, Andersen JH. 2008. Going ill to work: what personal circumstances, attitudes and work-related factors are associated with sickness presenteeism? *Social Science and Medicine* **67**: 956-964.
- Hansen CD, Andersen JH. 2009. Sick at work: a risk factor for long-term sickness absence at a later date? *Journal of Epidemiology and Community Health* **63**(5): 397-402.
- Hilbe JM. 2011. *Negative Binomial Regression - Second Edition*. Cambridge University Press: New York.
- Huver B, Richard S, Vaneecloo N, Bierla I. 2014. Âge, absence-maladie et présentéisme au travail: le cas d'un établissement de santé régional. *Management & Avenir* **70**(4): 97-114.
- Jensen S, MacIntosh J. 2007. Absenteeism in the workplace: results from Danish sample survey data, *Empirical Economics* **32**: 125-139.
- Johansen V. 2012. Sickness presenteeism in Norway and Sweden. *Nordic Journal of Social Research* **3**: 1-15.
- Johansen V, Aronsson G, Marklund S. 2014. Positive and negative reasons for sickness presenteeism in Norway and Sweden: a cross-sectional survey. *BMJ Open* **4**, e004123.
- Johns G. 2010. Presenteeism in the workplace: a review and research agenda. *Journal of Organizational Behavior* **31**(4): 519-542.
- Johns G. 2011. Attendance dynamics at work: the antecedents and correlates of presenteeism, absenteeism, and productivity loss. *Journal of Occupational Health Psychology* **16**(4): 483-500.
- Kivimäki M, Head J, Ferrie JE, Hemingway H, Shipley MJ, Vahtera J, Marmot MG. 2005. Working while ill as a risk factor for serious coronary events: the Whitehall II study. *American Journal of Public Health* **95**: 98-102.
- Koopman C, Pelletier KR, Murray JF, Sharda CE, Berger ML, Turpin RS, Hackleman P, Gibson P, Holmes DM, Bendel T. 2002. Stanford presenteeism scale: health status and employee productivity. *Journal of Occupational and Environmental Medicine* **44**(1): 14-20.
- Krohne K, Magnussen LH. 2011. Go to work or report sick? A focus group study on decisions of presenteeism among offshore catering section workers. *BMC Research Notes* **4**: 1-7.
- Kumar S, Grefenstette JJ, Galloway D, Albert SM, Burke DS. 2013. Policies to reduce influenza in the workplace: impact assessments using an agent-based model. *American Journal of Public Health* **103**(8): 1406-1411.

- Lambert D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**(1): 1-14.
- Leineweber C, Westerlund H, Hagberg J, Svedberg P, Luukkala M, Alexanderson K. 2010. Sickness presenteeism among Swedish police officers. *Journal of Occupational Rehabilitation* **21**: 17-22.
- Lerner D, Amick BC, Rogers WH, Malpeis S, Bungay K, Cynn D. 2001. The work limitations questionnaire. *Medical Care* **39**(1): 72-85.
- Lofland JH, Pizzi L, Frick KD. 2004. A review of health-related workplace productivity loss instruments. *Pharmacoeconomics* **22**(3): 165-184.
- McKevitt C, Morgan M, Dundas R, Holland WW. 1997. Sickness absence and 'working through' illness: a comparison of two professional groups. *Journal of Public Health* **19**(3): 295-300.
- Nielsen A-KL. 2008. Determinants of absenteeism in public organizations: a unit-level analysis of work absence in a large Danish municipality. *International Journal of Human Resource Management* **19**(7): 1330-1348.
- Rantanen I, Tuominen R. 2011. Relative magnitude of presenteeism and absenteeism and work-related factors affecting them among health care professionals. *International Archives of Occupational and Environmental Health* **84**(2): 225-230.
- Schultz AB, Chen C-Y, Edington DW. 2009. The cost and impact of health conditions on presenteeism to employers. *Pharmacoeconomics* **27**(5): 365-378.
- Shmueli A. 2003. Socio-economic and demographic variation in health and in heterogeneity: the issue of reporting heterogeneity. *Social Science and Medicine* **57**: 125-134.
- Simpson R. 1998. Presenteeism, power and organizational change: long hours as a career barrier and the impact on the working lives of women managers. *British Journal of Management* **9**: 37-50.
- Skagen, K: The costs of presenteeism, manuscript (part of submitted Ph.d. thesis, January 2016)
- Szymczak JE, Smathers S, Hoegg C, Klieger S, Coffin SE, Sammons JS. 2015. Reasons why physicians and advanced practice clinicians work while sick: A mixed-methods analysis. *JAMA Pediatrics* **169**(9): 815-821.
- Taloyan M, Aronsson G, Leineweber C, Hanson LM, Alexanderson K, Westerlund H. 2012. Sickness presenteeism predicts suboptimal self-rated health and sickness absence: a nationally representative study of the Swedish working population. *PlosOne* **7**(9), e44721.
- Winkelmann R. 1999. Wages, firm size and absenteeism. *Applied Economics Letters* **6**(6): 337-341.
- Winkelmann R. 2008. *Econometric Analysis of Count Data - Fifth Edition*. Springer-Verlag: Berlin Heidelberg.

FIGURES

Figure 1 – Sickness absence and presenteeism in observed distributions

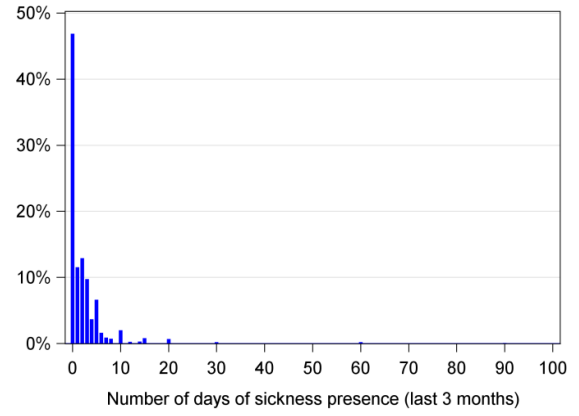
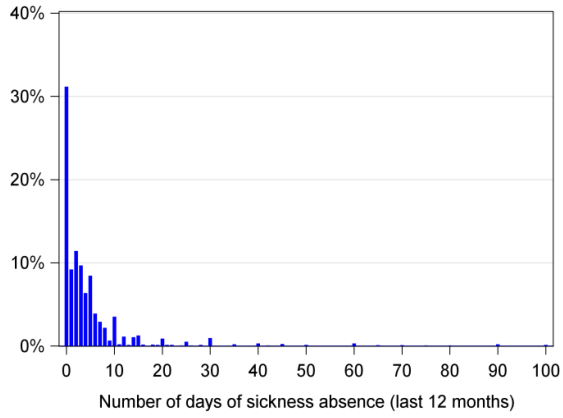


Figure 2 – Observed vs. estimated distributions (ZINB models)

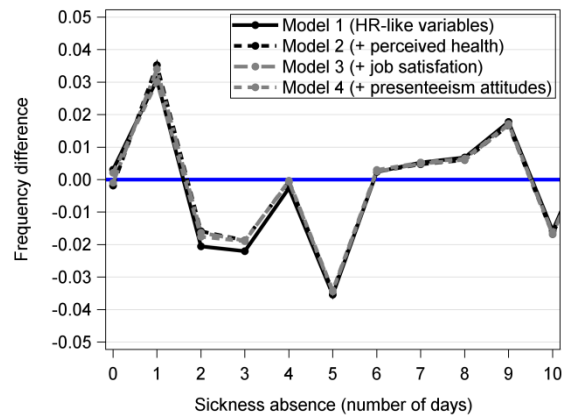
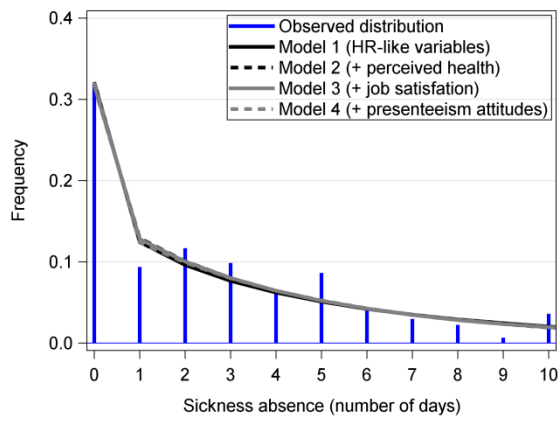


Figure 3 – Presenteeism probability distribution (model 1)

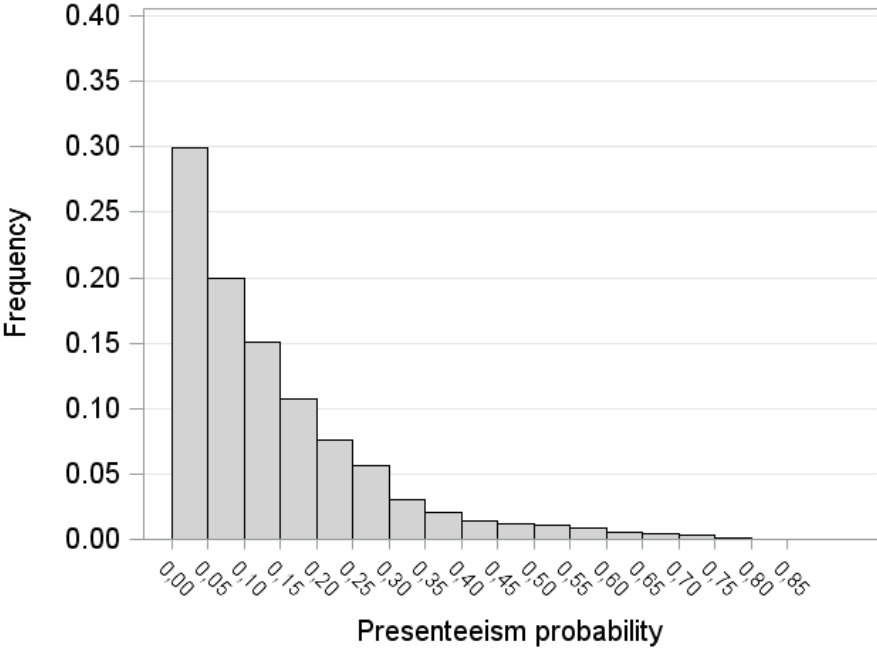


Figure 4 – Observed vs. estimated distributions (Simple Negative Binomial regressions)

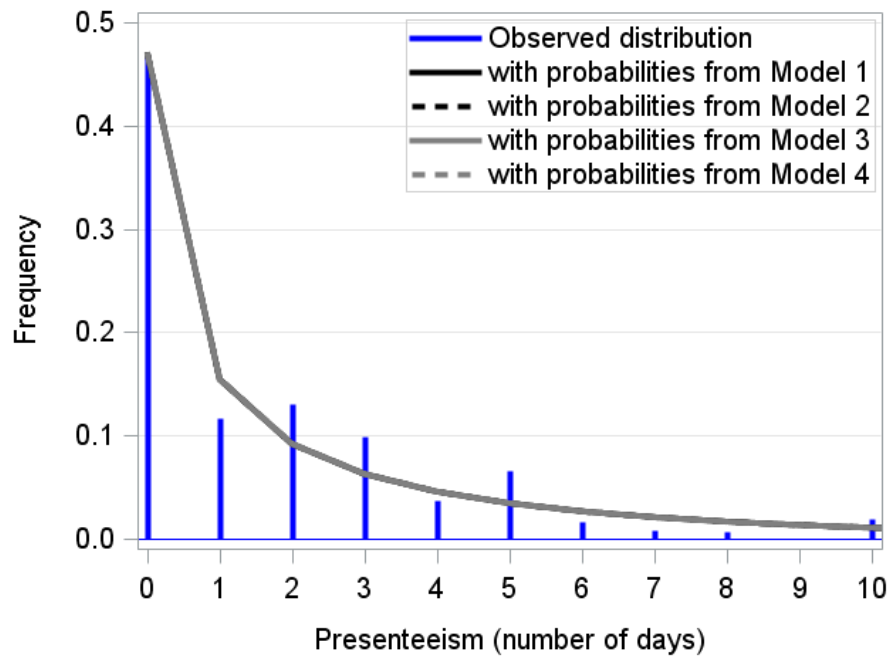


Figure 5 – Estimation of coefficient β_2 (100 simulations)

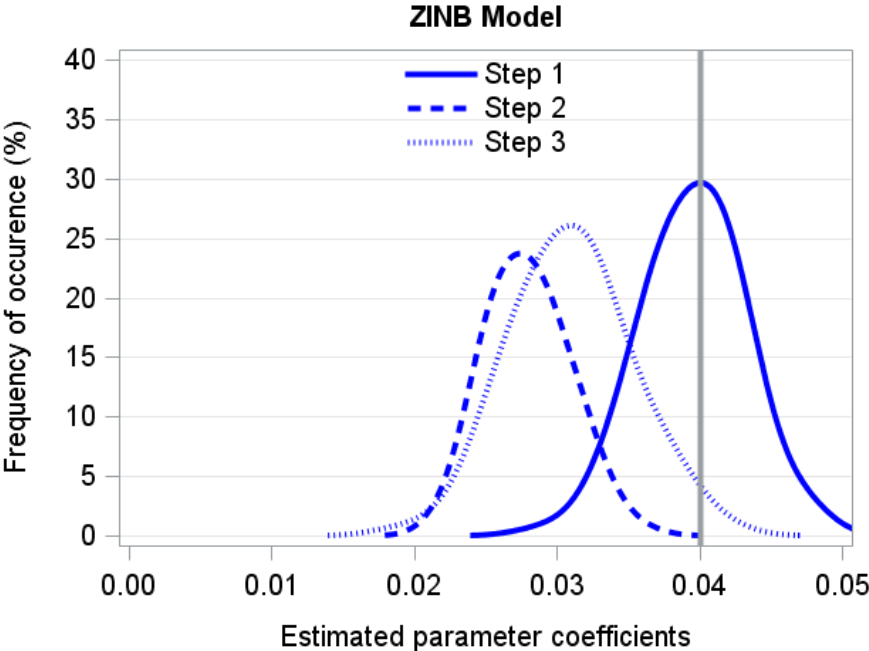


Figure 6 – Step 1: theoretical vs. ZINB-estimated probabilities

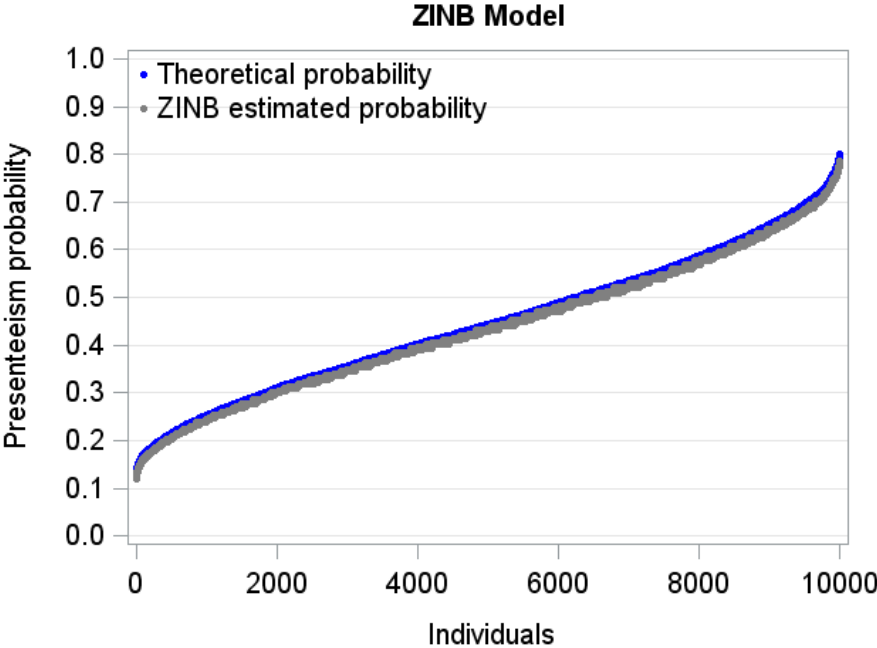


Figure 7 – Step 2: theoretical vs. ZINB-estimated probabilities

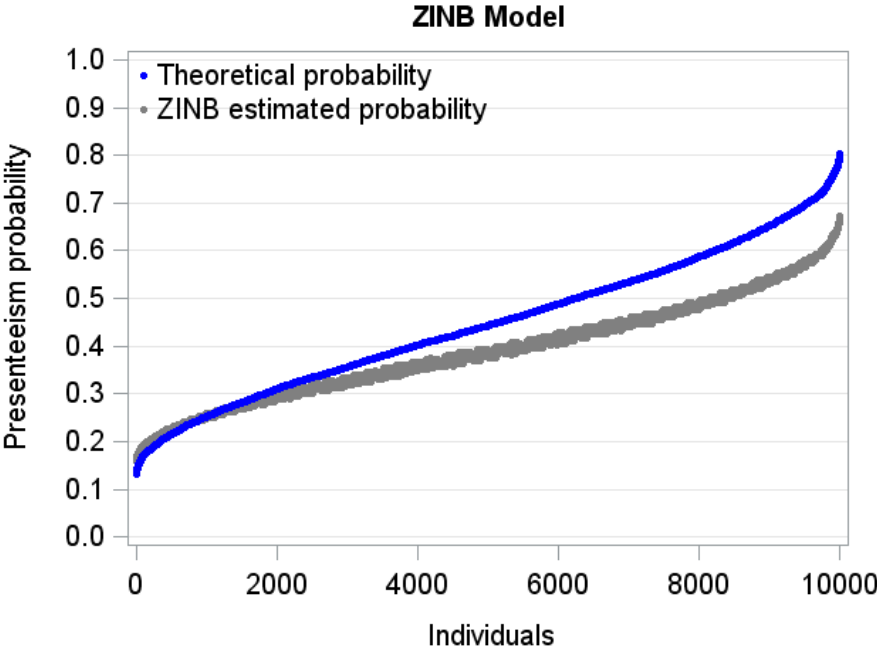
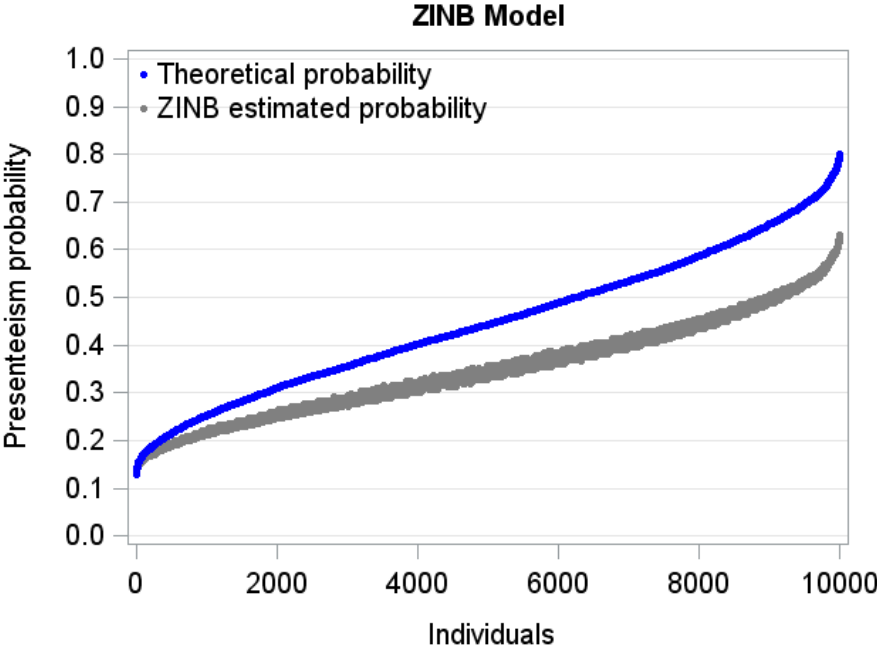


Figure 8 – Step 3: theoretical vs. ZINB-estimated probabilities



TABLES

Table I – Descriptive statistics

| Variables | Men | Women | Over 50 | Population |
|---|--------|--------|---------|------------|
| Age (mean) | 43.19 | 41.34 | 56.05 | 42.32 |
| Number of children < 12 (mean) | 0.58 | 0.50 | 0.09 | 0.54 |
| Length of higher education (> 4 years) | 23.58% | 19.45% | 14.96% | 21.63% |
| Working in the private sector | 68.22% | 49.10% | 51.79% | 59.18% |
| Working in a micro-firm | 15.20% | 15.52% | 17.07% | 15.35% |
| Self-employed | 7.58% | 4.57% | 8.46% | 6.15% |
| Manager | 35.68% | 19.45% | 32.76% | 28.01% |
| 1 to 5 subordinates | 19.72% | 11.96% | 17.97% | 16.05% |
| 6 to 25 subordinates | 12.01% | 5.58% | 10.24% | 8.97% |
| > 25 subordinates | 3.95% | 1.91% | 4.55% | 2.99% |
| Job satisfaction: very good or good | 65.65% | 64.45% | 65.61% | 65.08% |
| General health: very good or good | 72.51% | 72.85% | 66.10% | 72.67% |
| Number of days of sickness absence (mean) | 5.67 | 7.15 | 7.41 | 6.37 |
| Number of days of presenteeism (mean) | 2.29 | 2.67 | 2.30 | 2.47 |
| Individuals | 2 099 | 1 882 | 1230 | 3981 |
| % of total population | 52,73% | 47,27% | 30.90% | 100% |

Table II – Descriptive statistics

| Parameter | Model 1 | Model 2 | Model 3 | Model 4 |
|----------------------------|-------------|-------------|-------------|-------------|
| <i>Counting equation:</i> | | | | |
| Intercept | 1.1032 *** | 0.6128 *** | 0.6492 *** | 0.6882 *** |
| Female | 0.2242 *** | 0.1852 *** | 0.1809 *** | 0.1717 *** |
| Age | 0.0122 *** | 0.0030 | 0.0025 | 0.0024 |
| Higher education | -0.1736 ** | -0.1427 ** | -0.1466 ** | -0.1397 ** |
| Private sector | -0.1012 * | -0.0884 * | -0.0910 * | -0.0941 * |
| Micro-firm | -0.1458 * | -0.1482 * | -0.1456 * | -0.1485 * |
| Manager | -0.1544 ** | -0.1691 ** | -0.1665 ** | -0.1643 ** |
| Perceived health | | 0.3535 *** | 0.3502 *** | 0.3437 *** |
| Long-term sickness | | 0.3640 *** | 0.3680 *** | 0.3727 *** |
| Dispersion parameter | 1.2114 | 1.0797 | 1.0655 | 1.0141 |
| <i>Inflation equation:</i> | | | | |
| Intercept | -3.9853 *** | -3.2693 *** | -2.4431 *** | -3.2070 *** |
| Female | -0.5073 ** | -0.6348 *** | -0.6466 *** | -0.6344 *** |
| Age | 0.0474 *** | 0.0341 *** | 0.0284 ** | 0.0297 *** |
| Number of children | -0.6659 ** | -0.7298 ** | -0.7175 *** | -0.4751 *** |
| Higher education | -0.7953 ** | -0.8358 ** | -0.8167 ** | -0.6355 ** |
| Private sector | 0.3700 * | 0.3532 ° | 0.3304 ° | 0.2146 |
| Micro-firm | 0.4886 * | 0.5544 * | 0.5185 * | 0.3987 ° |
| Self-employed | 1.1159 *** | 1.1006 *** | 1.0268 *** | 0.9753 *** |
| Manager | 1.5569 *** | 1.5581 *** | 1.4901 *** | 1.1471 *** |
| 1 to 5 subordinates | -1.3536 *** | -1.3908 *** | -1.3425 *** | -1.0236 ** |
| 6 to 25 subordinates | -0.6938 * | -0.7293 * | -0.7353 * | -0.6260 ° |
| Job satisfaction | | | -0.2260 ** | -0.1608 * |
| Pride in presenteeism | | | | 0.8420 *** |
| Attend if a bit sick | | | | 0.3478 * |
| Attend with 38.2° | | | | 0.4391 ** |
| Log-likelihood | -9520 | -9385 | -9381 | -9346 |
| AIC | 19078 | 18812 | 18806 | 18742 |

Significance levels: p<0,1%=***; p<1%=**; p<5%=*; p<10%=°.

Table III – Presenteeism probabilities distributions

| Presenteeism probabilities | Minimum | 25th percentile | Median | Mean | 75th percentile | Maximum |
|----------------------------|---------|-----------------|--------|-------|-----------------|---------|
| Model 1 | 0.002 | 0.042 | 0.100 | 0.143 | 0.196 | 0.811 |
| Model 2 | 0.002 | 0.048 | 0.106 | 0.144 | 0.198 | 0.809 |
| Model 3 | 0.002 | 0.053 | 0.110 | 0.148 | 0.200 | 0.821 |
| Model 4 | 0.006 | 0.057 | 0.114 | 0.164 | 0.215 | 0.881 |

Table IV – Correlation between estimated probabilities and self-reported presenteeism days

| | Parameter | (1) All SP values | | (2) Without zeros | | (3) Filtered zeros | |
|---------|-------------|-------------------|-----|-------------------|-----|--------------------|-----|
| Model 1 | Intercept | 0.7769 | *** | 1.3291 | *** | 0.8874 | *** |
| | Pres. prob. | 0.6286 | ** | 1.1927 | *** | 0.6292 | *** |
| Model 2 | Intercept | 0.7759 | *** | 1.3257 | *** | 0.8863 | *** |
| | Pres. prob. | 0.6317 | ** | 1.2067 | *** | 0.6330 | ** |
| Model 3 | Intercept | 0.7981 | *** | 1.3449 | *** | 0.9035 | *** |
| | Pres. prob. | 0.4729 | * | 1.0527 | *** | 0.5089 | ** |
| Model 4 | Intercept | 0.7116 | *** | 1.3152 | *** | 0.8190 | *** |
| | Pres. prob. | 0.9108 | *** | 1.0684 | *** | 0.9305 | *** |