ELSEVIER Journal of Public Economics 59 (1996) 195–218

JOURNAL OF PUBLIC ECONOMICS

Do economic incentives affect work absence? Empirical evidence using Swedish micro data

Per Johansson[a], Mårten Palme[b,*],

[a]*Department of Economics, University of Umeå, S-901 87 Umeå, Sweden*
[b]*Department of Economic Statistics, Stockholm School of Economics, Box 6501, S-113 83 Stockholm, Sweden*

Received November 1993; final version received November 1994

Abstract

Using a linear demand function, frequently used in labour supply studies, absenteeism is modelled as an individual day-to-day decision. The parameters in the econometric model are consistently estimated, using the (time-aggregated) number of days absent in 1981 as the dependent variable for a sample of Swedish blue-collar workers (both men and women), under some assumptions on unobserved heterogeneity and serial correlation. Implications of compensating wage differentials and efficiency wage hypotheses are discussed. The results for the male subsample reveal a negative effect on work absence of the direct cost of being absent. However, for the female subsample, the Slutsky condition is rejected.

Keywords: Sickness insurance; Unobserved heterogeneity; Mixture distribution; Semiparametric estimation

JEL classification: C14; J22; J29

1. Introduction

Absenteeism has attracted increased attention in several of the social sciences (cf. Goodman and Atkin, 1984, for an overview). As this literature is primarily associated with business administration and applied psychology,

* Present address: Labour Studies, National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge, MA 02138, USA.

SSDI 0047-2727(95)01501-9

the main interest has been to assess the effects on work absence of different organizational and psycho-social conditions in the workplace. Very few studies have, as is the aim of this study, investigated whether individual economic incentives affect work absence, and if they do, to what extent. There are at least two reasons why it is relevant to analyze this issue. First, it is of fundamental interest to determine empirically whether economic analysis could be extended to an area that traditionally has been considered to be almost entirely determined by the employees' health and working conditions. Secondly, as sickness insurance is compulsory, or in other respects controlled by the state, in most industralized economies (see Kangas, 1991, for an overview of the institutional systems), the government is able to affect the cost of the individual for being absent form work. If the individuals are affected by economic incentives, the government can also influence the frequency of work absence. Analyzing the impact of economic incentives on work absence can, thus, be seen as an extension of the existing labour supply literature, which has almost exclusively dealt with the problem of the impact of income support programs and income taxes on the desired number of hours at work.

In order to explore these issues, let us first investigate to what extent the institutional setting for this study, Swedish sickness insurance, may allow for the insured employee to be affected by economic incentives. Sweden has compulsory sickness insurance as a part of the national social insurance system. Insured individuals are entitled to benefits if their perception of their state of health is such that they consider that 'it does not permit them to do their regular work'. The regulations allow the insured person to be absent from work for up to eight days without a certificate from a physician. As a person's state of health can be difficult to monitor even for a qualified doctor, and as the decision to work or to be on sick leave is left primarily to the insured individual, it seems realistic to believe that economic incentives influence the everyday choice of whether to work or to be absent from work. However, it is fair to say that many individuals who receive sickness insurance benefits are, depending on their state of health, expected to be insensitive to economic incentives.

Work absence is defined as time when the employee is absent from work, which cannot be referred to as statutory leisure time or absence agreed upon in advance with the employer. A recent time-use study (SAF, 1986) shows that 97.1% of the work absence for blue collar workers in Sweden is covered by the sickness insurance. Work absence and utilization of the sickness insurance can thus be seen as almost synonymous.

The recent developments regarding the average number of days spent on sickness insurance, shown in Fig. 1, reveal some marked fluctuations. It seems unlikely that these fluctuations are solely caused by changes in the state of health of the Swedish population. There have, however, been

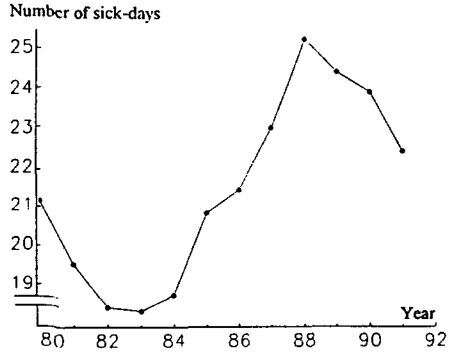

Fig. 1. The average annual number of days on sickness insurance per worker in Sweden, 1980–1991. All insured individuals. *Source:* The Swedish National Social Insurance Board.

macroeconomic fluctuations and alterations in the contribution levels, that deserve further investigation.

In the empirical analysis of this study we use micro data, i.e. individual data from a sample of Swedish blue-collar workers, both men and women. The econometric models are derived from standard economic utility theory, i.e. a theoretical model for the day-to-day decision of whether to be absent from work or not is formulated. We only have information on how many days each individual has been absent from work over a time period of one year. Under the assumption of no serial correlation and no unobserved heterogeneity, the parameters in the day-to-day model can be efficiently estimated with binomial maximum likelihood. If serial correlation is present, the parameters can be estimated using the binomial distribution as a quasi-likelihood estimator.

Unobserved heterogeneity is very likely to be present when analyzing absenteeism behavior empirically. Individuals, primarily due to different states of health, are likely to differ in their sensitivity to economic incentives in their work absence behaviour. Obviously, it would have been advantageous if individuals who are suffering from bad health could be singled out. Unfortunately, this requires more information than can be obtained from our (or any) data set. The data do permit that individuals who have been absent less then a certain number of days are selected and analyzed separately. However, sample selection problems will arise following such a strategy. The approach we adopt is to consider information on the individual's state of health that can be obtained from the data set and to use an estimation method that can handle the possibility of unobserved heterogeneity. More specifically, a mixture distribution model is estimated (cf. Titterington et al., 1987) using a semiparametric estimator.

Few other empirical studies have used economic theory to explain differences in absence from work.[1] Most of these studies share the methodological problem of time aggregation. The present study extends two aspects of this literature. The everyday economic choice of being absent from work is explicitly modelled contingent on occupation and personal characteristics that may influence the choice. Given the kind of data in this study, i.e. the number of days of work absence aggregated over the time period of one year, theoretically consistent estimation methods are used.

The data are collected from the 1981 *Swedish Level of Living Survey* (SLLS). In order to obtain information on individual work absence, this sample is matched with the National Social Insurance Board register on the annual number of days for which sickness insurance for each individual is paid. Individual income data are collected from tax registers. The data set consists of 1967 individuals for the year 1981.

The paper is organized as follows. In Section 2, we derive the empirical specification of the choice between work and work absence and discuss how hypotheses previously used in labour economics can be used to explain the relationship between occupation and absenteeism. Section 3 describes the data set and the Swedish sickness insurance and tax systems. Section 4 discusses the econometric model specification, and addresses estimation and testing issues. The empirical results are given in Section 5 and the final section contains a discussion of the results.

2. Theoretical specification

To model the everyday choice of whether or not to be absent from work, we adopt an approach that has been widely use for labour supply models. We assume that each individual has a utility function:

$$u = U(x, L; s) , \tag{1}$$

where x is a composite consumption good, L is leisure and s is a vector of personal characteristics. Leisure consists of contracted leisure time, t^{l}, and time absent, t^{a}. The utility function is maximized subject to a budget constraint:

$$x = w(h^* - (1 - \delta)t^{a}) + R , \tag{2}$$

where h^* is the contracted number of working hours, R is income from sources other than labour, w is net wage and δ is the share of the income the worker receives when absent. The price of the composite consumption

[1] See Allen (1981a,b), Barmby et al. (1991, 1993), Delgado and Kneisner (1992) and Dunn and Youngblood (1986).

good, x, is normalized to be one. h^* can be divided into desired number of working hours and time absent, hence $h^* = h + t^a$. This gives the identity $T \equiv h + t^a + t^l$, where T is total time available.

The following functional form for the direct utility function is chosen (cf. Hausman, 1980):

$$U(x, t^a) = \exp\left\{ -\left(1 + \frac{\beta(x + \bar{s})}{b - T + (t^l + t^a)}\right)\right\}\left(\frac{T - (t^l + t^a) - b}{\beta}\right), \quad (3)$$

where, $\bar{s} = s/\beta - \alpha/\beta^2$, $b = \alpha/\beta$ and α and β are parameters. Maximizing the utility function (3) subject to the budget constraint (2) we obtain the demand function for work absence:

$$t^a = h^* - \alpha w(1 - \delta) - \beta(R + h^* w\delta) - s = h^* + \alpha w^* + \beta y - s . \quad (4)$$

This is a linear function of the cost for the individual of being absent from work (net earnings not covered by the sickness insurance, i.e. the relative cost between absence and consumption), w^*, and virtual income when the individual is absent, y. It is easy to show that the demand for absence equation, Eq. (4), is the dual to the supply function for labour given in Hausman (1980) and Blomquist (1983).

It can be argued that men and women may have different arguments in their utility function, there are several empirical evidences on this in the labour supply literature (see Blomquist and Hansson-Brusewitz, 1990, for a recent study using using Swedish data). Also, as women in most families have the main responsibility for the children in the household, child illnesses, and thus the number of children in the household, may affect the female work absence behaviour more than the male. In order to allow for these differences, a female dummy variable, f, is included in s as well as interaction terms between the dummy variable and the main economic variables, cost and income. This gives us the expression, with the notations to be used, for the demand for absence:

$$t^a = h^* + \alpha w^* + \beta y_i + \alpha_f f w^* + \beta_f f y - s . \quad (5)$$

If the demand for absence equation, Eq. (5), is to be consistent with a well-behaved utility function, the Slutsky condition must be satisfied. For this functional form it is sufficient that $(a + a_f f) < 0$ and $(\beta + \beta_f f) \geq 0$, for every $t^a \geq 0$, thus for the male subsample, $\alpha < 0$ and $\beta \geq 0$.

It is important to emphasize that the contracted leisure time is exogenous in this model. Every day the individuals are assumed to choose between attending work or not, *conditional* on their perception of their health, the contracted number of work hours and the costs of the alternatives. We thus assume that the contract specifying the hours of work is made between the employee and the employer in advance. An *alternative* specification is to assume that contracted leisure time and time absent are substitutes and that the individual chooses a combination of these two. An important component of such a model would be to consider the costs of abuse of the sickness insurance, i.e. a moral hazard problem.

In the vector of socio-economic variables, s, variables related to health are included. It is also important to consider, and in the empirical analysis control for, individual characteristics that are likely to be correlated with work absence and the cost and income variables. Since the cost variable partly consists of the wage rate, we use the compensating wage differential (e.g. Rosen, 1986) and the efficiency wage (e.g. Shapiro and Stiglitz, 1984) hypothesis to get some idea about which variables should be included in s.

2.1. Compensating wage differentials

When a person accepts a job offer, he/she not only accepts a wage rate and a contracted number of work hours, but also a number of non-wage characteristics. The worker may accept higher risks at the workplace in exchange for a higher wage rate. If the workers chooses a higher risk level, he is likely to have a higher rate of work absence for two reasons. First, since the worker faces a higher level of risk at the workplace, he will strive to minimize the exposure time to these risks, i.e. he has, ceteris paribus, greater incentives to be absent from work. The fact that less risk-averse workers may self select into higher risk jobs may partially offset this effect. Secondly, workers who are exposed to higher risks also have a higher probability of being absent due to work-related illnesses or injuries. Ideally, one would like to separate these causes of differences in absenteeism. This requires, however, that work absence can be separated according to cause of absence. Thus, omitting the risk may lead to biased estimates of the relationship between wage rate and absence.

2.2. Efficiency wage hypothesis

The employer may influence the level of absence by several means; for example, the controls on the employees could be increased (see Henrekson et al., 1992, p. 79, for a review of an empirical study on the firm level). However, some jobs are by their nature very difficult to monitor. Shapiro and Stiglitz's (1984) shirking model predicts that when the possibilities of monitoring workers's job performance are poor, employers may pay wages above the market clearing level in order to elicit adequate effort from their employees. That is, the employer pays a wage higher than the market wage to induce the worker not to be absent from work. This implies that we, in addition to the indirect effect of wages through the cost variable, w^*, will get a separate (direct) effect of wages on work absence. Furthermore, higher monitoring will imply lower wages and lower absence and, therefore, a positive correlation between wages and absence will be induced. Thus, if we do not control for the employer's monitoring level our estimates of α and α_f may be positively biased as daily earnings are included in the variable measuring the cost of being absent from work. Another implication is that when unemployment increases, shirking or absenteeism decreases. The potential cost of being caught shirking and eventually losing one's job is greater if the unemployment rate is high and it is harder to find a new job. Wages will, because of excess supply on the labour market, ceteris paribus, be lower in times of (areas with) high unemployment. Thus, if we do not control for unemployment we may get a negative bias on the estimates of α and α_f.

3. Insurance system and data

In this section we present the variables included in the models (tables containing all the variables are given in Appendices 1 and 2). The data are obtained from the 1981 SLLS (see Eriksson and Åberg, 1987). Some selections are made for the present study. First, the sample is restricted to individuals aged 20 to 64 years. Secondly, individuals not in the labour force are excluded from the sample. We also exclude self-employed, students, military personnel and white-collar workers, i.e. the study is restricted to blue-collar workers. The reason for excluding these other groups is to limit heterogeneity arising from differences in sickness insurance systems that cannot be obtained from the available data. After these exclusions are made, the sample consists of 1967 individuals, 1045 women and 922 men.

The dependent variable is the number of days for which each individual receives compensation from the sickness insurance, aggregated over the time period of one year. Data for this variable are obtained from the National Social Insurance Board registers, by matching with the SLLS sample. As a measure of work absence, these data contain three deficiencies. First, as mentioned in the introduction, the share of absence from work that is covered by sickness insurance is estimated at 97.1%. Secondly, during the time period studied, the first day in each sickness spell was not compensated by the insurance and consequently was not reported. Third, those days that the individual does not regularly work (e.g. Saturdays and Sundays), are not covered by the sickness insurance for the first ten compensated days in one spell. After that time period, however, they are covered by the insurance as well and are therefore recorded in our data. The advantage of using data from the register is that they generally contain very few, if any, measurement errors.

Data to construct the cost and income variables, i.e. the different components of the individual's income, are obtained from tax registers that are matched with the SLI.S. As was shown in Section 2, the cost and virtual income variables are influenced by the compensation level in the sickness insurance and by the income tax. In the studied time period, 90% of the insured income below the social insurance ceiling was compensated by the insurance for all days covered by the insurance, i.e. for the days that we have recorded in the data set. Very few individuals in our data set, only about 3%, have labour incomes above the ceiling. The Swedish income tax system consists of two parts: a proportional tax imposed by the local authorities, and a progressive tax imposed by central government. THelocal tax rate varies between sweden's 286 communes but, as there is a dense clustering around the mean tax rate, we obtain a simplification by assuming that all individuals pay the mean local tax rate. the income tax is based on the individual's gross income from labour, capital and social security.

as labour income is not fully compensated by the sickness insurance, the marginal tax rate is dependent on how many days the individual is absent from work, i.e. it is not independent of the individual decision whether or not to attend work. this means that if the marginal tax rate is changed for the individual, the budget set will vary between days within the time period)a year) that we have aggregated. furthermore, it means that the marginal tax rate is endogenous for at least one day during that time period. this implies that we have to approximate these separate budget sets with one for each individual during the aggregated time period. the approximation used is the marginal tax rate each individual actually pays for the year studied. this marginal tax rate is then used to calculate a linear budget set, i.e. the cost variable and the variable for average daily virtual non-labour income.

the cost variable, w^*, is calculated as 10%)the share of labour income not covered by the sickness insurance) of the net marginal hourly wage rate, i.e. the hourly wage rate multiplied by one minus the individual's marginal income tax rate, for labour incomes below the social security ceiling. for incomes above the ceiling, the cost is counted as 100% of the hourly wage rate. the hourly wage rate is obtained by dividing the potential annual labour income by the number of hours of work stated in the survey. to calculate the potential annual income from labour, we have added 10% for each day recorded as the individual having been compensated by the sickness insurance. for individuals with insured incomes above the social security ceiling, 100% is added for the part of their incomes exceeding the social security ceiling.

The virtual income, y, consists of two parts. The first one is the average daily income from capital and child and housing allowances. If the individual is married or cohabiting, only 50% of these allowances are incorporated. The second part is net daily income from sickness insurance.

Contracted working hours, h^*, is based on the self-reported number of contracted weekly working hours. The self-reported contracted working hours are classified into four levels to obtain h^*. h^* takes the values 2, 4, 6 and 8, if the self-reported contracted weekly working hours are ≤ 10, >10 and ≤ 20, >20 and ≤ 30 and >30, respectively.

The specification of the vector of the socio-economic variables was discussed in Section 2. The conclusion from that discussion was that, in addition to variables related to health, variables relating to working conditions, the level of monitoring of workers, the unemployment level and family status should be considered.

The variables relating to the state of health are obtained from the SLLS survey. Questions on some 60 different illnesses are put forward. The questions are of the following kind: Have you during the last twelve months had a headache, a cold, a heart attack, etc.? Because the survey was conducted during the spring and summer of 1981, it is likely that some variables are endogenous, depending on the number of days spent on sickness insurance. Since we include working conditions, some types of illness that may be symptoms of bad working conditions, e.g. headache, pain in the chest, etc. are not included. A subset of 11 variables, for which a physician has most likely made a diagnosis, are chosen (see Appendix B for a description of the variables). The illness variables chosen can be considered as exogenous and are hopefully not severely affected by bad working conditions. Since we are only interested in controlling for differences in the state of health we choose to reduce the dimension by performing a principal component (PC) analysis (see Appendix B). The three first PCs, *HS*1, *HS*2 and *HS*3 are included, they account for almost 60% of the variation in the illness variables. A dichotomous variable, *DISAB*, taking the value one if a person is in some way disabled and prevented from doing his work properly, is also used.

We use two types of working condition variable: occupation-specific risk indexes and self-reported information from the survey.

The occupation-specific task indexes, consisting of two SIR (Standard Incidence Ratio) indexes, measure reported accidents at work and work-related diseases for the period 1981–1983 per 100,000 worked hours, respectively, and are provided by the National Board of Occupational Safety and Health; 195 different occupations are considered and the indices are matched with the SLLS. As these two indices are highly correlated and measure the same expected influence on absence, we perform a PC analysis (see Appendix B). The first PC, *RISK*, which takes care of 87% of the variation in the risk variables, is included in the model.

The SLLS survey contains questions on several job characteristics, some of which we use in this study. The number of such variables is large and they are highly correlated. Furthermore, they have, in most cases, the same expected influence on work absence. Therefore, a PC analysis is performed (see Appendix B). We choose to include the first three PCs that take care of about 40% of the variation in the variables. The first PC, *DISS1*, can be interpreted as a physically demanding factor. The second PC, *DISS2*, can be interpreted as a mentally demanding factor. The third PC, *DISS3*, cannot be given any firm interpretation.

The risk indexes and the self-reported job characteristics variables can be interpreted as measuring different aspects of higher costs for actually attending a job; the costs for risk exposure and disutility, respectively. On the other hand, all the job characteristic variables, e.g. contact with poison and monotonous body movements, are likely to cause work-related diseases and accidents. The advantage of using the risk indexes is that they give a more objective measure of risk. When an individual characterizes his job as physically demanding, this could be generated by his perception of his health rather than by the nature of his work. The advantage of using self-reported characteristics is that they are more directly tied to the particular individual's workplace.

Two dummy variables, if it is 'important to arrive on time at the work place', *ITIME*, and 'if it is required to clock in', *PUNCH*, are used to measure the monitoring level at the individual's workplace.[2] The unemployment variable, *UNEMP*, is the county-specific annual average unemployment rates that have been matched with the original data set. In Section 2 we concluded that there is a separate effect of wages, or, more precisely, the wage paid in excess of the market clearing level on work absence. There are, however, problems that prevent us from trying to measure that effect in our model. As we are not able to measure fully the productivity of each worker, it is impossible to measure this additional amount on the wage rate. If we simply use the hourly wage rate as a independent variable, we will get problems with colinearity with our cost variable, w^*.

4. Econometric models and estimation

Demand for time absent, t^a, is a latent variable that is not directly observable. Instead, we can observe absence for two days or more. If we partition the vector of socio-economic variables, s, into γ, z and θ, where z is a vector of exogenous variables, γ is the corresponding parameter vector and θ is a random or fixed variable taking into account non-observable variables. For individual i at day h, Eq. (5) may be written:

[2] These two variables have previously been used by Arai (1994) to measure the monitoring level.

$$\begin{aligned} t_{ih}^{a} &= h_{ih}^{*} + \alpha w_{ih}^{*} + \beta y_{ih} + \alpha_f f_i w_{ih}^{*} + \beta_f f_i y_{ih} + \gamma z_{ih} + \theta_i + \zeta_{ih} \\ &= h_i^{*} + \alpha w_i^{*} + \beta y_i + \alpha_f f_i w_i^{*} + \beta_f f_i y_i + \gamma z_i + \varepsilon_{ih} + \zeta_{ih} , \\ &= h_i^{*} + \boldsymbol{x}_i \omega + \varepsilon_{ih} + \zeta_{ih} , \end{aligned} \tag{6}$$

where a subscript i indicates an annual average variable, $\omega = (\alpha, \beta, \alpha_f, \beta_f, \gamma')'$, ζ_{ih} are identically and independently distributed (iid) random variables and $\varepsilon_{ih} = \theta_i + \varepsilon_h$, with $\varepsilon_h = h_{ih}^{*} - h_i^{*} + (\alpha + \alpha_f f_i)(w_{ih}^{*} - w_i^{*}) + (\beta + \beta_f f_i)(y_{ih} - y_i) + \gamma(z_{ih} - z_i)$. If all explanatory variables as well as h_{ih}^{*} are constant over the year, we have that $\varepsilon_h = 0$. This is a situation that is unlikely. Thus, even if no unobserved heterogeneity is present, it is likely that we have systematic errors due to time aggregation of the explanatory variables.

Let I_{ih} be an indicator variable for absenteeism such that

$$I_{ih} = \begin{cases} 1 , & \text{if } t_{ih}^{a} > k_{ih} , \\ 0 , & \text{if } t_{ih}^{a} \leq k_{ih} , \end{cases}$$

where k_{ih} is a threshold value for day h If demand for absence exceeds k_{ih}, individual i will be absent. Under the assumption that ζ_{ih} has a logistic distribution and that $k_i = k_{ih}$, the probability to be at work on day h is

$$\Pr(I_{ih} = 0) = \Pr(t_{ih}^{a} - k_i \leq 0) = 1/[1 + \exp(h_i^{*} + \boldsymbol{x}_i \omega + \varepsilon_{ih} - k_i)] ,$$

with the probability of being absent

$$\Pr(I_{ih} = 1) = 1 - \Pr(I_{ih} = 0) = 1/[1 + \exp(k_i - h_i^{*} - \boldsymbol{x}_i \omega - \varepsilon_{ih})].$$

4.1. Estimation

4.1.1. The binomial model

Let V_i be the number of days person i is absent during a year. Under the assumption of no state dependence, i.e. that ε_{ih} is independent of ε_{ig}, for every h and g, V_i is binomially (Bin) distributed with parameters $\pi_i = \Pr(I_{ih} = 1)$ and N, the number of days in the year, i.e.

$$V_i = \sum_{h=1}^{N} I_{ih} \sim \text{Bin}(N, \pi_i) .$$

If V_i is independent of V_j, for every i and j and if there is no unobserved heterogeneity present (i.e. $\theta_i = 0$, for every i), the log-likelihood function can be written:

$$\ell(\omega; V_1, \ldots, V_n) = \sum_{i=1}^{n} \ln\binom{N}{V_i} + V_i \ln \pi_i + (N - V_i) \ln(1 - \pi_i) , \tag{7}$$

where n is the number of individuals. The maximum likelihood (ML) estimates of the parameter vector ω are obtained by maximizing ℓ in (7).

If there is serial correlation (i.e. ε_{ih} is not independent of ε_{ig}, for every h and g) and/or there is individual unobserved heterogeneity (i.e. $\theta_i \neq 0$), the ML estimator of the Bin model will generally yield inconsistent estimates. Unobserved heterogeneity is a main cause for overdispersion (i.e. the predicted variance is greater than the one expected from the theoretical distribution). Depending on the assumption of the cause for overdispersion, different tests for overdispersion and different estimation procedures are applicable.

4.1.2. Overdispersed binomial model

When absence is serially correlated over days, V_i is not binomially distributed. Under the assumption of a correct mean function, the estimation can, e.g. be performed by regarding ℓ in (7) as a quasi-likelihood function. This estimator is suggested by McCullagh and Nelder (1983) along with the additional assumption $\mathrm{Var}(V_i) = \sigma^2 N\pi_i(1-\pi_i)$, where $\sigma^2 > 1$. The σ^2 is a measure of overdispersion. The same parameter estimates as for the original binomial model are obtained, but the covariance matrix is adjusted with an estimate of σ^2.

4.1.3. The mixture distribution model

Unobserved heterogeneity can be modelled if we let the stochastic parameter $\varepsilon_{ih} = \theta_j$ in the demand for absence function, hence

$$t^{\mathrm{a}}_{ihj} = h^*_i + x_i\omega + \theta_j + \zeta_{ih} \, .$$

Heterogeneity is individual if $i = j$ and group-specific if the number of groups, K, is less than n. A semiparametric estimator may be used (e.g. Andersson and Brännäs, 1992) to estimate the unknown parameters, ω, and the distribution function $H(\theta)$ of θ. The parameters and the distribution function are estimated jointly. The non-parametric ML estimator of $H(\theta)$ take the form of a discrete distribution function. By this, the probability for a person to be absent on day h equals

$$\Pr(I_{ih} = 1) = \sum_{j=1}^{K} q_j \Pr(I_{ih} = 1|\theta_j) \, ,$$

where $\Pr(I_{ih} = 1|\theta_j) = 1/[1 + \exp(k_i - (h^*_i + x_i\omega + \theta_j))]$. This yields the log-likelihood function:

$$\ell \propto \sum_{i=1}^{n} V_i \ln \sum_{j=1}^{K} q_j \Pr(I_{ih}|\theta_j) + (N - V_i) \ln\left(1 - \sum_{j=1}^{K} q_j \Pr(I_{ih}|\theta_j)\right) , \qquad (8)$$

that is maximized with respect to ω, the probabilities $q_1, \ldots, q_K$, the mass points $\theta_1, \ldots, \theta_K$ and the finite number K. Leroux (1992) shows that the number of groups may be consistently estimated using the Akaike information criteria (AIC, Akaike, 1973) or the Baysian information criteria (BIC, Schwarz, 1978).

The probability that a given observation is associated with a particular group j ($j = 1, \ldots, K$) may be obtained by the Bayes rule as:

$$d_j = \Pr(j|V_j) = q_j \Pr(V_i|j) \Big/ \sum_j^K q_j \Pr(V_i|j) . \tag{9}$$

The d_j is a logistic discrimination probability.

4.2. Overdispersion tests

It is important to test for overdispersion in qualitative dependent variable models because generally, if present, this leads to inconsistent estimates (see Godfrey, 1990, ch. 6, and Yatchew and Griliches, 1985). However, if the overdispersion is on the borderline of detectability and the mean function is correctly chosen, the ML estimator without overdispersion is asymptotically efficient (Cox, 1983).

Under the assumption that $\text{Var}(V_i) = \sigma^2 N\pi_i(1 - \pi_i)$, tests for overdispersion can be made using, e.g. the Pearson chi-squared statistic:

$$\chi^2 = \sum_{i=1}^{n} (V_i - N\hat{\pi}_i)/[N\hat{\pi}_i(1 - \hat{\pi}_i)] ,$$

this is asymptotically $\chi^2(n - p)$ distributed and where p is the number of estimated parameters. A satisfactory (when $n \to \infty$, see McCullagh and Nelder, 1983) estimator for σ^2 is

$$\hat{\sigma}^2 = \chi^2/(n - p) .$$

Dean (1992) proposed a score test for extra-binomial variation under the assumption that $\Pr(I_{ih} = 1|\theta_i) = 1/[1 + \exp(k_i - (h_i^* + x_i\omega + \theta_i))]$, where θ_i are iid with $\text{E}(\theta_i) = 0$ and $\text{Var}(\theta_i) = \sigma_\theta^2$. The score test for testing $\sigma_\theta^2 = 0$, i.e. to test if the model (7) is adequate, is based on the log-likelihood function under the alternative hypothesis:

$$\partial\ell/\partial\sigma_\theta^2|_{\sigma_\theta^2=0} = T(\hat{\omega}) ,$$

where $\hat{\omega}$ is the ML estimate when $\sigma_\theta^2 = 0$ and $T_i(\hat{\omega}) = (V_i - N\hat{\pi}_i)^2 - N\hat{\pi}_i(1 - \hat{\pi}_i)$. The Fisher information matrix, $I(\hat{\omega}, \sigma_\theta^2)$, evaluated at $\sigma_\theta^2 = 0$ is used to obtain an estimator of the variance of $T(\hat{\omega})$. If the information matrix is partitioned as

$$I(\hat{\omega}, \sigma_\theta^2) = \begin{bmatrix} I_{\omega\omega} & I_{\omega\sigma_\theta^2} \\ I'_{\omega\sigma_\theta^2} & I_{\sigma_\theta^2\sigma_\theta^2} \end{bmatrix},$$

the score test takes the form:

$$S = T(\hat{\omega})/\hat{W}\,, \tag{10}$$

where $\hat{W}^2 = I_{\sigma_\theta^2\sigma_\theta^2} - I_{\omega\sigma_\theta^2} I_{\omega\omega}^{-1} I'_{\omega\sigma_\theta^2}$. S is asymptotically normally distributed with expectation zero and variance one.

5. Results

In this section, we will discuss the results. First, we give a general description and comparison between the different estimation methods used. We then discuss the parameter estimates from the preferred model.

5.1. Model selection

From the binomial ML, we find that overdispersion is present, since $\chi^2 = 186{,}091.28$, $S = 3108.62$ and $\hat{\sigma}^2 = 95.82$. The coefficient of determination is, as expected, quite small, $R^2 = 0.13$.

The mixture model is semiparametrically estimated (the BFGS algorithm in GAUSS) with $k = 1, 2, \ldots, K$ treated as fixed. The AIC and BIC for $k = 1, 2, 3$ are given in Table 1. The binomial model ($k = 1$) is rejected since the minimum AIC and BIC are obtained for $k = 2$. The signs of the intercepts for the two groups, given in Table 2, reveal that the probability to be absent is larger for the second group.

By using Bayes rule, or the logistic discrimination probability (9), we classify the individuals into group one if the probability d_1 is larger than 0.5, and into group two if not. The first group consists of 98.7% of the sample and the rest, naturally, belongs to the second group. The second group may be interpreted as primarily consisting of the long-term sick and thus less sensitive to economic incentives to be absent from work. The classifying rule above (i.e. using the mean prediction, $\hat{V}_{ij} = N\hat{p}_{ij}$, for individual i belonging to group j) is used to estimate the coefficient of determination, $R^2 = 0.53$.

Table 1
Information criteria values for different values of k

k	1	2	3
AIC	151,930.12	151,837.96	151,913.35
BIC	151,999.91	151,839.66	151,920.63

Table 2
Parameter estimates and standard errors for the mixture distribution

Variable/parameter	Estimate	S.E.
θ_1	−4.5711	0.1019
θ_2	7.5464	1.7539
h^*	0.1950	0.0081
w^*/α	−0.4658	0.0393
y/β	−0.0107	0.0004
f^*w^*/α_f	1.2477	0.0518
f^*y/β_f	0.0167	0.0006
AGE	0.0312	0.0007
DISAB	1.0818	0.0175
HS1	−0.1861	0.0239
HS2	−0.8004	0.0478
HS3	0.7497	0.0903
RISK	0.1159	0.0062
DISS1	0.2822	0.0087
DISS2	−0.3788	0.0143
DISS3	0.0862	0.0112
TIME	0.2312	0.0168
CLOCK	0.1968	0.0137
UNEMP	−0.1975	0.0079
f	−3.1136	0.1019
KIDS6	−0.3157	0.0251
KIDS16	−0.2040	0.0153
f^**KIDS6*	0.3772	0.0338
f^**KIDS16*	0.0677	0.0201
SINGLE	0.1415	0.0187
DIV	0.1209	0.0185
q_1	0.9902	0.0002

Note: Let $q_1 = 1/(1 + e^{-\phi})$ and maximize with respect to the unconstrained ϕ. On convergence, we get $\hat{q}_1$. The variance is estimated from the Gauss approximation $V(\hat{q}_1) = \hat{\sigma}^2_{\phi} e^{-2\hat{\phi}}/(1 + e^{-\hat{\phi}})^4$.

Compared with the binomial model, this is a major improvement in goodness of fit.

5.2. Results for the mixture distribution model

As can be seen in Table 2, all parameter estimates are significantly different from zero.[3,4] The discussion of the parameter estimates and

[3] The parameter estimates from the binomial and overdispersed binomial models are presented in Johansson and Palme (1993) and can be sent to the reader upon request.

[4] It must be noted that the estimated parameters are standardized by the standard error in the logistic distribution. Thus, the estimates cannot be interpreted as the structural ones and it would not be correct to restrict the parameter of h^* to be one. Consequently, it is not possible to test if this estimated parameter is different from one.

calculated elasticities are divided into six groups, starting with the estimated parameters for the cost and income variables.

5.2.1. Cost and income variables

The estimates of α and β are -0.4658 and -0.0107, respectively. It can be noted that the same signs as above are obtained in Allen (1981a) with respect to wage rate and disposable income. The parameter estimates of α_f and β_f take the values 1.2477 and 0.0167. Thus, the total effects of the cost and income variables for the female subsample are 0.7819 and 0.006, respectively. As the parameter estimate for the cost variable for females is positive, the Slutsky condition is not satisfied for the female subsample. Even though the parameter estimate for the income variable is negative, the Slutsky condition is satisfied for all male individuals in the sample.

By using the parameter estimates of the cost and income variables it is possible to calculate the elasticity of the mean number of days absent with respect to the contribution level, δ, i.e. the parameter that could be controlled by the government. Using the estimates for males (the results where the Slutsky condition is satisfied), this elasticity could be calculated for the two groups separately. This yields a mean value of 4.602 and 0.0003 for groups one and two respectively. Thus a reduction of δ by 1% would decrease the mean level of days absent by about 4.6% in the male group one.

5.2.2. State of health

The parameter estimate on *DISAB* is positive and strongly significant. The same is true for *AGE*, which can also be interpreted as a measure of health status. The parameter estimates for the PCs (*HS1–HS3*) are somewhat hard to interpret. The estimate for the first PC is of a unanticipated sign (see the loading in Appendix B). The most likely explanation is that the variables relating to health are highly correlated with *DISAB*.

5.2.3. Risk and disutility of work

The risk variable has a positive coefficient. This effect on work absence should be taken into account in, for example, cost-benefit analysis of increased safety at work. The PCs (*DISS1–DISS3*) for the self-reported variables measuring the disutility of work have a significant effect on work absence. The effect, however, is not uniform. The second PC, the contrast between repetitive, monotonous (negative loadings) and outdoor (positive loadings) job characteristics, have a negative coefficient estimate. Thus, individuals with the work profile of low stress and outdoor work, have on average a lower absence rate.

As the risk and disutility components are highly correlated, it is, unfortunately, not possible to separate these two effects on work absence.

5.2.4. Monitoring level

The parameter estimates for the two dummy variables used to measure the monitoring level, whether or not the individual uses a punch clock (*CLOCK*) and whether or not it is important to be on time (*TIME*), have positive signs. Following the discussion in Section 2, this result was not expected.[5] There are, however, several possible explanations. First, a small fraction of work absence is not reported to the sickness insurance (see SAF, 1986), so it is likely that the reporting rate is higher in firms with a higher monitoring level. Secondly, lower work time flexibility will probably induce higher absence as, for example, visits to the dentist or the bank must be reported if the work time flexibility is low. Thirdly, it is likely that our two measures of monitoring level are correlated with e.g. monotonous working conditions that increase work absence (see previous section).

5.2.5. Unemployment rate

The unemployment rate has a significant inverse effect on work absence. The elasticity with respect to the unemployment rate is -0.446 for group one and -0.00003 for group two. Thus, an increase of 1% in the unemployment level would decrease the number of days absent by 0.45% in group one. This is a fairly strong effect and could very well be an important reason for the decline in work absence that has been observed recently.

5.2.6. Male–female differences and personal characteristics

Considering the five variables measuring male–female differences, the dummy variable for females and the four interaction terms, the mean elasticity for women is calculated to 0.20 in the first group and almost zero in the second, i.e. work absence is, on average, almost 0.2% higher for females in the sample. However, the female dummy taken separately is negative. Thus, controlling for all the interaction terms, women have a lower absence rate.

The dummy variables for the occurrence of children below age 6 living in the household (*KIDS6*) and children in the interval 6–16 years (*KIDS16*) both have negative signs. These estimates, taken together with the interaction between the female dummy, give a positive sign to the coefficient for *KIDS6* and negative for *KIDS16* for the females. The results are, in part, explained by the fact that there is a separate insurance within the Swedish social insurance system that allows one of the parents to take care of dependent children with illnesses in the home, with the same contribution level as the sickness insurance. In 1981, the usage of this insurance was

[5] As the sign of this coefficient took the opposite as was expected, the bias of omitting this variable on the estimates of the cost and income variables would be positive rather than negative as was expected.

limited to 60 days per child and year. The positive effect of the *KIDS6* variable for the female subsample supports the specialization within the household hypothesis. Women have the main responsibility for taking care of children with illnesses for more than 60 days per year, which is more likely for small children than for children between age 6 and 16. The explanation of the negative signs for the male sub-sample with children in both age categories, and for females with children between age 6 and 16, is probably due to the fact that these variables also catch up other personal characteristics promoting low work absence, such as good health and regular life habits. The positive signs of the dummy variables for divorced and unmarried individuals may reflect that these variables catch up, opposite to households with children, bad health and irregular life habits. This results on personal characteristics are fairly consistent with those obtained in similar studies (see, for example, Allen, 1981a, or Björklund, 1991).

6. Conclusions and further research

What is the contribution of this study to the existing empirical knowledge of how economic incentives affect work absence? We have chosen to give economic incentive a broad definition. The effects of unemployment, risk exposure and control at the workplace have been given an economic interpretation and empirical results are commented on in the previous section. As reported above, the female sub-sample failed to meet the Slutsky conditions. Thus, our model was, in some respect, misspecified for this subsample. To evaluate the estimates of the male subsample and to give an illustration of what our results tell us about the interpretation of changes in work absence seen recently in Sweden, we will predict the change in male work absence between 1990 and 1991 and compare the result with the actual outcome on aggregate data. The reason for analyzing the change between these two particular years is because two interesting changes took place in 1991. First, on 1 March the level of the sickness insurance paid was reduced from 90% to 60% of daily earnings for the first three days and from 90% to 80% for days 4–89. Secondly, the unemployment level increased from an annual average of 1.5% to 2.7%. Table 3 shows the change in the average number of days compensated by sickness insurance, divided into different lengths of spell, between these years. As can be seen in the table, there is a sharp decrease in the number of days compensated by sickness insurance in the kind of spells affected by the change in the compensation level (1–89 days), a decrease of, on average, 16.09%.

Under the assumption that 10/12 of the work absence occured after 1 March, the level of the sickness insurance, δ, has decreased by 8.1, on average, between 1990 and 1991.

Table 3
Average number of days on sickness insurance divided into different kinds of spells. 1990 and 1991. All insured males

Spell length (days)	1990	1991	Difference (%)
1–3	1.88	1.56	−17.02
4–89	8.50	7.15	−15.88
89–365	4.83	4.85	0.41
365–	5.31	5.76	8.47
All	20.53	19.31	−5.94

Using these presumptions and our estimates, we have predicted the change in percentage between 1990 and 1991. These predictions are revealed in Table 4. The results are somewhat disappointing: the decrease in work absence between 1990 and 1991 is substantially overestimated, or, more precisely, the estimates of the decrease due to the change in the compensation level in sickness insurance is overestimated, while the predictions made of the effects of the change in the unemployment level are more reasonable. However, there are several explanations for these results. First, the actual outcome is based on all insured individuals, while our estimates are for a subsample of blue-collar workers. As most white-collar workers did not experience any, or experienced a much smaller, decrease in the level of their sickness insurance, the actual change for our subsample is probably larger. Secondly, the estimates originate from data from ten years before the predictions. One major change in sickness insurance was realised in 1987: the first day on sick-leave is compensated and the days that individuals do not work regularly (e.g.weekends) are covered fo the first 10 days. Thirdly, other alterations of potential importance for the changes in the average number of days on sickness may have either not been identified in this study, or have been identified but the change between 1990 and 1991 could not be measured, e.g. changes in working conditions.

The main methodological lesson from this study is that estimation methods that are consistent with a model for day-to-day choice can be used when analyzing time-aggregated data on work absence, even if unobserved heterogeneity is present. In our view, the next step in this research area is to develop an estimation technique that allows us to relax the assumptions of

Table 4
Predicted changes (%) in work absence between 1990 and 1991

	Group one	Group two	All
Compensation level, δ	−33.05	−0.00	−32.64
Unemployment	−0.53	−0.00	−0.53
Total	−33.58	−0.00	−33.27

no state dependence and unobserved heterogeneity, simultaneously.[6] To fully investigate the effects of state and duration dependence, and the efficiency loss of time aggregation, we need access to day-to-day data.

Acknowledgements

We are indebted primarily to Kurt Brännäs and two anonymous referees for valuable comments on earlier versions of this paper. We have also received comments from Thomas Aronsson, Anders Björklund, Per-Anders Edin, Pravin Trivedi, Magnus Wikström, participants in the seminar at the 1993 ESPE Conference in Budapest as well as participants at the seminars at the Economics Departments at the Universities of Uppsala and Umeå. Financial support from the Swedish Council for Social Research is gratefully acknowledged. The authors take full responsibility for any remaining errors and shortcomings.

[6] In Johansson and Palme (1993), estimations with a Markov model that relax the assumption of no state dependence (while the assumption of no unobserved heterogeneity has to be maintained) are presented. A problem with this model was that it could not be proved to be globally identified.

Appendix A: Description

Descriptive statistics for variables in the regression model

		Mean	St. dev.	Min	Max
Dependent variable					
V	Number of days absent	22.31	53.43	0	365
	Economic variables				
*h**	Contracted labour time	7.25	1.38	2.00	8.00
*w**	Cost of being absent	1.40	0.94	0.14	25.03
y	Virtual income	111.73	29.15	−3.56	199.73
UNEMP	Unemployment rate	2.37	0.92	1.00	5.50
	Dangerous or unhealthy work environments				
DISS1	First PC (see Appendix B)	1.18	0.81	0.00	3.36
DISS2	Second PC (see Appendix B)	−0.41	0.59	−1.92	1.22
DISS3	Third PC (see Appendix B)	−0.33	0.56	−1.84	1.19
RISK	First PC (see Appendix B)	−0.00	1.32	−1.59	6.48
	Control at the workplace				
TIME	Punctuality is important	0.74	0.44	0.00	1.00
PUNCH	Use of punch clock is required	0.33	0.47	0.00	1.00

Descriptive statistics for variables in the regression model (*Continued*)

		Mean	St. dev.	Min	Max
	Personal characteristics				
SINGLE	Single	0.20	0.40	0.00	1.00
DIVORCED	Divorced	0.08	0.27	0.00	1.00
FEMALE	Female	0.53	0.50	0.00	1.00
AGE	Age of individual	38.37	11.70	20.00	61.00
KIDS6	Number of children below 6 years	0.27	0.58	0.00	3.00
KIDS16	Number of children below 16 years	0.47	0.79	0.00	6.00
	State of health				
DISAB	Disabled	0.18	0.27	0.00	1.00
HS1	First PC (see Appendix B)	0.09	0.32	−0.02	2.28
HS2	Second PC (see Appendix B)	−2.15	0.26	−2.14	0.99
HS3	Third PC (see Appendix B)	−0.04	0.24	−0.18	0.50

Appendix B: Principal component analysis

Loadings for the principal components, *HS1*, *HS2* and *HS3*

Variable/PC	*HS1*	*HS2*	*HS3*
INJURED	0.14	−0.88	−0.41
STRUMA	0.04	0.01	0.03
TB	0.00	0.00	0.00
HEARTH	0.01	0.00	0.01
STOMACH	0.09	−0.19	0.02
HEMORR	0.96	0.07	−0.21
PREGNANT	0.17	0.42	−0.88
BROCK	0.01	−0.01	0.02
MENTAL	−0.00	−0.06	0.00
CANCER	−0.00	0.00	0.01
DIABETIC	0.00	0.01	−0.01
NEURO	0.00	−0.03	−0.00
Variance/total variance in %	26.22	16.56	14.23

Loadings for the principal components (PC), *RISK*

Variable/PC	*RISK*
RISK1	0.71
RISK2	0.71
Variance/total variance in %	87.15

Loadings for the principal components for (un)healthy conditions, *DISS1*, *DISS2* and *DISS3*

Variable/PC	*DISS1*	*DISS2*	*DISS3*
DIRTY	0.19	0.19	0.04
NOISE1	0.13	−0.02	0.15
NOISE2	0.13	0.18	−0.14
OUTSIDE	0.17	0.45	−0.16
TEMP	0.16	−0.04	0.09
DRAFT	0.27	0.09	0.02
SMOKE	0.24	0.02	0.21
SHAKE	0.10	0.02	0.03
POISON	0.11	0.04	0.01
LIFT	0.26	0.22	−0.17
OTHPHY	0.41	−0.02	−0.07
SWEAT	0.39	0.04	0.00
PHYEXH	0.26	−0.24	−0.05
TIRED	0.02	−0.21	−0.23
MENTEXH	0.01	−0.30	−0.54
STRESS	0.12	−0.39	−0.49
REPET	0.10	−0.27	0.33
MONBODY	0.21	−0.50	0.38
UNPBODY	0.44	0.06	−0.04
Variance/total variance in %	19.21	10.03	9.01

Descriptive statistics for the variables included in the PC analysis

		Mean	St. dev.	Min.	Max.
	Dangerous or unhealthy work environment				
DIRTY	Work is dirty	0.15	0.36	0.00	1.00
NOISE1	Noisy environment	0.14	0.35	0.00	1.00
NOISE2	Noisy environment	0.31	0.46	0.00	1.00
OUTSIDE	Work is outside	0.28	0.45	0.00	1.00
TEMP	Exposed to non-normal temperatures	0.17	0.38	0.00	1.00
DRAFT	Exposed to strong drafts	0.23	0.42	0.00	1.00
SMOKE	Exposed to gas, dust or smoke	0.23	0.42	0.00	1.00
SHAKE	Exposed to strong shakes or vibrations	0.06	0.24	0.00	1.00
POISON	Exposed to poisons, acids or explosives	0.11	0.32	0.00	1.00
	Physically demanding work				
LIFT	Heavy lifting	0.25	0.43	0.00	1.00
OTHPHY	Otherwise physically demanding	0.51	0.50	0.00	1.00
SWEAT	Work causing daily sweating	0.32	0.47	0.00	1.00
PHYEXH	Work is physically exhausting	0.26	0.44	0.00	1.00
TIRED	Feel very tired at end of day	0.14	0.35	0.00	1.00

Descriptive statistics for the variables included in the PC analysis

		Mean	St. dev.	Min.	Max.
	Stressful or monotonous work				
MENTEXH	Work is mentally exhausting	0.34	0.47	0.00	1.00
STRESS	Work is stressful	0.57	0.49	0.00	1.00
REPET	Work is repetitive	0.23	0.42	0.00	1.00
MONBODY	Monotonous movements	0.49	0.50	0.00	1.00
UNPBODY	Unpleasant body positions	0.47	0.50	0.00	1.00
	Risk indexes				
RISK1	SIR, work accidents	1213.88	984.30	50.00	6000.0
RISK2	SIR, work-related diseases	1105.49	999.02	50.00	6000.0
	State of health variables				
INJURED	Persistent injury	0.05	0.26	0.00	2.00
STRUMA	Struma	0.02	0.15	0.00	2.00
TB	Tuberculosis	0.003	0.06	0.00	2.00
HEARTH	Heart attack	0.005	0.08	0.00	2.00
STOMACH	Gastric ulcer	0.03	0.21	0.00	2.00
HAEMORR	Haemorrhoids	0.08	0.33	0.00	2.00
PREGNANT	Pregnant or pregnancy difficulty	0.04	0.25	0.00	2.00
BROCK	Inguinal hernia	0.01	0.14	0.00	2.00
MENTAL	Mentally sick	0.02	0.16	0.00	2.00
CANCER	Cancer	0.01	0.12	0.00	2.00
DIABETIC	Diabetic	0.02	0.15	0.00	2.00
NEURO	Neurological illness, e.g. Polio	0.01	0.13	0.00	2.00

References

Akaike, H., 1973, Information theory and an extension of the maximum likelihood principle, in: B.N. Petrov and F. Csaki, eds., Second International Symposium on Information Theory (Akademiai Kiado, Budapest) 267–281.

Allen, S.G., 1981a, An empirical model of work attendance, Review of Economics and Statistics 63, 77–87.

Allen, S.G., 1981b, Compensation, safety, and absenteeism: Evidence from the paper industry, Industrial and Labor Relations Review 34, 207–218.

Andersson, T. and K. Brännäs, 1992, Explaining cross-country variation in nationalization frequencies, International Review of Economics and Business 40, 47–62.

Arai, M., 1994, Compensating wage differentials versus efficiency wages. An empirical study of job autonomy and wages, Industrial Relations 33, 249–262.

Barmby, T.A., C.D. Orme and J.G. Treble, 1991, Worker absenteeism: An analysis using micro data, The Economic Journal 101, 214–229.

Barmby, T.A., C.D. Orme and J.G. Treble, 1993, Worker absence histories: A panel data study, Newcastle Discussion Papers in Economics 93/02, University of Newcastle upon Tyne.

Björklund, A., 1991, Vem får sjukpenning? En empirisk analys av sjukfrånvarons bestämningsfaktorer, In: Arbetskraft, arbetsmarknad och produktivitet (Produktivitetsdelegationen, Stockholm) 289–299.

Blomquist, N.S., 1983, The effect of income taxation on the labor supply of married men in Sweden, Journal of Public Economics 22, 169–197.

Blomquist, N.S. and U. Hansson-Brusewitz, 1990, The effect of taxes on male and female labour supply in Sweden, Journal of Human Resources 25, 317–357.

Cox, D.R., 1983, Some remarks on overdispersion, Biometrica 70, 269–274.

Dean, C.P., 1992, Testing for overdispersion in Poisson and binomial regression models, Journal of the American Statistical Association 87, 451–457.

Delgado, M.A. and T.J. Kniesner, 1992, Semiparametric versus parametric count-data models – econometric considerations and estimates of a hedonic-equilibrium model of worker absenteeism, mimeo, Indiana University, Bloomington.

Dunn, L.F. and S.A. Youngblood, 1986, Absenteeism as a mechanism for approaching an optimal labor market equilibrium: An empirical study, The Review of Economics and Statistics 68, 668–674.

Eriksson, R. and R. Åberg, 1987, Welfare in transition – Living conditions in Sweden 1968–1981, Clarendon Press, Oxford.

Godfrey, L.G., 1990, Misspecification tests in econometrics (Cambridge University Press, Cambridge).

Goodman, P.S. and R.S. Atkin, eds., 1984, Absenteeism (Jossey-Bass Inc., Publisher, San Francisco).

Hausman, J.A., 1980, The effect of wages, taxes, and fixed costs on women's labor force participation, Journal of Public Economics 14, 161–194.

Henrekson, M., K. Lantto and M. Persson, 1992, Bruk och missbruk av sjukförsäkringen (SNS Förlag, Stockholm).

Johansson, P. and M. Palme, 1993, The effect of economic incentives on worker absenteeism: An empirical study using Swedish micro data, Umeå Economic Studies 314, University of Umeå.

Kangas, O., 1991, The politics of social rights. Studies on the dimensions of sickness insurance in OECD countries, Ph.D. thesis at the Swedish Institute for Social Research.

Leroux, B.G., 1992, Consistent estimation of a mixing distribution, The Annals of Statistics 20, 1350–1360.

McCullagh, P. and J.A. Nelder, 1983, Generalized liner models (Chapman and Hall, London).

Rosen, S., 1986, The theory of equalising differences, in: O. Ashenfelter and R. Layard, eds., Handbook of labour economics, 1 (Elsevier, Amsterdam).

SAF, 1986, Tidsanvändningsstatistik (Svenska Arbetsgivareföreningen, Stockholm).

Schwarz, G., 1978, Estimating the dimension of a model, The Annals of Statistics 6, 461–464.

Shapiro, C. and J.E. Stiglitz, 1984, Equilibrium unemployment as a worker discipline device, American Economic Review 74, 433–444.

Titterington, D.M., A.F.M. Smith and U.E. Makov, 1987, Finite mixture distributions (John Wiley, Chichester).

Yatchew, A. and Z. Griliches, 1985, Specification error in probit models, The Review of Economics and Statistics 63, 134–139.