Economic Incentives and Gender Differences in Work Absence Behavior *

Göran Broström[†], Per Johansson[‡] and Mårten Palme[§]

June 27, 2002

Abstract

We estimate a labor supply model on a random sample of Swedish male and female blue collar workers to study the effect of economic incentives on work absence behavior. We observe work absence for each day during 1990 and 1991 for each worker in the sample. We use non-parametric (Kaplan-Meier) techniques; semi-parametric stratified models, where individual effects are removed; and fully parametric Cox regression models, where observed characteristics are used to control for heterogeneity. An exogenous change in the cost of being absent due to a reform of the sickness insurance, which took place during the time period covered by the data, is used as identifying information. The empirical analysis is focused on explaining gender differences in work absence behavior. We find that about one third of this difference in our sample can be attributed to differences in costs of being absent.

*We thank Bertil Holmlund, Peter Skogman Thoursie as well as participants in seminars given at Uppsala University and at the Research Department of Statistics Norway in Oslo for useful comments. Johansson and Palme acknowledge financial support from the Swedish Council for Social Research.

†Department of Statistics, Umeå University, SE-901 87 Umeå, Sweden. E-mail: gb@stat.umu.se.

‡IFAU (Institute for Labour Market Policy Evaluation), SE-751 20 Uppsala, Sweden. E-mail: Per.Johansson@ifau.uu.se.

§Department of Economics, Stockholm University, SE-106 91 Stockholm, Sweden. E-mail: Marten.Palme@ne.su.se.

1. Introduction and Motivation

The large fluctuations in the work absence rate in most European countries during recent decades have attracted considerable attention. Most workers, in most countries, are covered by some form of sickness insurance, which is regulated by labor market legislation.[1] As a result, the coinsurance in these insurance schemes,[2] and the extent to which economic incentives affect work absence in general, have been scrutinized in the public policy debate. The impact of the unemployment rate and the business cycle in general on health and work-absence behavior have also been discussed extensively. A further issue that has been considered in the literature is that females on average have a higher work-absence rate as compared to men. This is regarded as more or less a "stylized fact" in empirical studies on worker absenteeism (see e.g. VandenHeuvel and Wooden, 1995 or Vistnes, 1997). Can this gender difference in observed work-absence behavior be explained by differences in economic incentives for being absent, preferences or differences in health and work environment?

To gain understanding on this issue, we examine microdata from a sample of $1,396$ blue-collar workers obtained from the Swedish Level of Living Survey (SLLS), where information on work absence for each day during the years 1990 and 1991 has been matched with this sample. The data on work absence were obtained from registers of actual transactions compiled by the National Social Insurance Board.

During the time period covered by the data, two important policy changes,

[1] See Kangas (1991) for an international overview.

[2] See Lantto (1991) for a theoretical analysis of the coinsurance in the sickness insurance.

which radically altered the cost and virtual income[3] underlying worker's work absence decision, took place. First, as of March 1, 1991, the replacement rate was decreased from 90 percent of labor earnings below the social security ceiling to 65 percent for the first three days of a sickness spell, to 80 percent from day 4 to day 89, and remained at 90 percent after day 90. Second, an income tax reform, whereby marginal tax rates were drastically reduced, was implemented in January 1, 1991.

The usual empirical evidence of more prevailing work absence among the females is also found for this data (see Tables A7.1 and A7.2 in Appendix 2). The mean difference in number of days absent in 1990 and 1991 are respectively 4.5 and 6 days. Kaplan-Meier estimates, presented in Section 3, shows that this difference can be attributed to more frequent, rather than longer, work-absence spells for the females.

The effect of economic incentives on work absence has been studied in a number of papers (see e.g. Allen, 1981a, or Barmby, Orme and Treble, 1996). In particular, two previous papers, Johansson and Palme (1996 and 2002) analyze the effect of economic incentives generated by the Swedish compulsory sickness insurance on work absence. As in this study, a labor supply model is used to derive the empirical specification of work absence. There are, however, several differences, both in the empirical focus and methodology, between these papers and the present study. Here we focus on gender differences in work absence behavior. This, however, requires that the effect of economic incentives - and also work environment, health status and macroeconomic conditions - on work absence are measured in a first step.

[3]The income received when absent from work.

Johansson and Palme (1996) uses individual data on work absence behavior aggregated over one year. The between individual variation in virtual income and the cost of being absent is used to estimate - using a semiparametric count data regression model - the effect of economic incentives on work absence. Johansson and Palme (2002) uses the same data as in the present study and binomial logit models are used to model the choice between being absent and working. Unobserved heterogeneity is controlled for by fixed effects. Different specifications for duration dependence are applied in the work and work absence spells respectively.

The empirical analysis in this paper is divided into three parts. First, we use Kaplan-Meier estimates (Kaplan and Meier, 1958) to non-parametrically study gender as well as pre- and post-reform differences in work absence behavior. Second, we use discrete time Cox regressions where heterogeneity is controlled for by a large set of individual characteristics (primarily on health status and work place characteristics). Third, we use a stratified analysis (see e.g. Lancaster, 1990, Chapter 9). This analysis is feasible since there are repeated spells of work absence and since the reform of the sickness insurance radically altered the cost of being absent from work for the group of workers that we examine.

For all three estimators we find that economic incentives, through the cost of being absent, matter for work-absence behavior. About one third of the male-female difference in work absence can be attributed to differences in economic incentives to be present at work. The remaining two thirds cannot be explained by observable characteristics. Our interpretation of this result is that most of the gender difference in work absence behavior are due to intrinsic gender behavioral differences. This supports previous findings by e.g. Paringer (1983) and Nilsson (2001). Paringer (1983) uses the observation from Sindelar (1982) that women invest more in their health, to explain more frequent female work absence spells.

The paper is organized as follows. In Section 2 we briefly describe the sickness insurance and income taxes in Sweden. Section 3 specifies the economic model for work absence behavior and Section 4 presents the estimation procedures. The data are described and analyzed non-parametrically in Section 5. Section 6 reports the results. Section 7 concludes.

2. Sickness Insurance and Income Taxes in Sweden

Sweden has a compulsory sickness insurance scheme.[4] It is financed through a proportional tax rate levied on wages and replaces forgone earnings due to temporary health problems that prevent the insured worker from doing his regular job. Sickness insurance is administrated by local insurance offices. Since it is very hard to judge whether or not a worker is able to perform his regular job, monitoring against abuse is very light during the first six days in a sickness spell. However, a certificate from a physician is required to be entitled to sickness insurance payments as of the seventh day in a spell.

The replacement level, the share of labor earnings paid to the worker by the insurance, has changed on several occasions in recent years. In the major reform covered by our longitudinal data - in March 1, 1991 - the replacement level was decreased from being 90 percent of labor earnings below the social security ceiling[5] from the first day in a sickness spell, to 65 percent in the first three days in a spell and to 80 percent from day four to day 89.

[4]For a more detailed description of the sickness insurance, see Johansson and Palme (2002).

[5]In 1995, about 6.7 percent of all insured workers had labor earning above the social security ceiling (see, National Social Insurance Board, 1997). For a description on the construction and indexation of the social insurance ceiling see Palme and Svensson (1998).

An insured worker's economic incentives for being absent from work are also affected by income taxes. Sweden has an integrated income tax system. Taxes are paid to both the national and local governments. The national government determines the tax base for both of these taxes. The local tax is proportional and is determined by each of Sweden's 288 local governments, although some income redistribution does take place between high- and low-income municipalities. In 1991, the local government tax rate varied between 26.87 and 33.48 percent.

Sweden's income tax system underwent a radical change after the tax reform in 1991, i.e., the first year of our data covers the pre-reform system and the second year pertains to the post-reform system. This tax reform encompassed three fundamental changes. First, from being unified in the pre-reform tax system, the tax base was divided into *earned* and *capital* income. Second, marginal tax rates were reduced substantially. Figure 2.1 shows the relation between taxable income and marginal tax rates under the pre- and post-reform income tax regimes, respectively. For the calculations underlying the figure, the local government tax rate is set at 31 percent. As can be seen, the highest marginal tax rate was reduced from the local government tax rate plus a 42 percent national tax rate (with a maximum set at 75 percent in a combined marginal tax rate) to a 20 percent national tax rate in addition to the local government tax. It is also evident from Figure 2.1 that most full-time wage earners, in the income interval between 70 000 and 170 000 Swedish kronor (SEK), did receive substantial reductions in their marginal tax rates. The marginal tax rate decreased in some income intervals in the post-reform regime because the basic deduction was made income dependent, i.e., it rises with income in some intervals and decreases with income in other intervals.

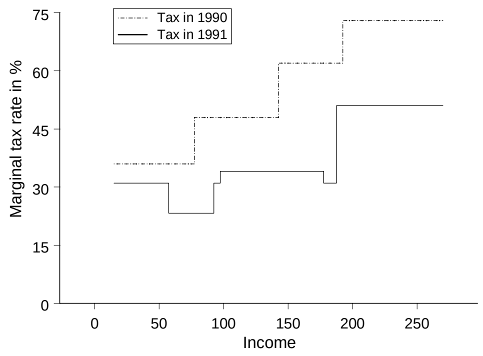

Figure 2.1: Percentage marginal tax rates in the pre- and post-1991 Swedish income tax reform regimes, respectively; taxable income in thousands of SEK.

Finally, the third main component of the tax reform was a substantial increase in child and housing allowances. The child allowance, which is the same amount per child irrespective of the parents' income, was increased by about one third. For example, the child allowance for the first child in a family was increased from SEK $6,720$ to SEK $9,000$ per year. The housing allowance is means tested. The amount is determined by the individual's earnings two years before he or she actually receives the allowance payment, as well as by his or her housing costs. The magnitude of the increase in the housing allowance was about the same as that for the child allowance.

3. Modelling Absence from Work

3.1. General Specification of the Hazard Functions

The effect of economic incentives on work-absence behavior has been examined in several empirical and theoretical studies (see e.g. Barmby, Orme and Treble, 1996, Barmby and Sibly, 1999, Johansson and Palme, 1996 and 2002). These studies analyze different costs of absence from work. In Sweden, as in most other industrialized countries with a compulsory sickness insurance, the direct cost of being absent corresponds to the share of daily earnings not covered by the sickness insurance.

In order to define the cost and virtual income variables, let us first define the worker's daily budget set. Let L_j represents leisure time at day j. L_j consists of two components: contracted leisure time, t_j^l and time in work absence t_j^a (that is $L_j = t_j^l + t_j^a$). Denote the duration in the present spell (in either work or work absence) Dur_j and assume that the contracted leisure time is fixed over the time period studied (two years), i.e., $t_j^l \equiv t^l$. The daily budget constraint can then be defined as

$$x_j + (1 - \delta_j)\, w_j t_j^a = w_j t^c + R_j,$$

where x_j is daily consumption, t^c is the contracted number of daily working hours, R_j is income from sources other than labor, w_j is net hourly wage and $(1-\delta_j)$ is the share of the income the worker receives when absent. Assuming that the worker maximize a conditional (on the duration in the present spell) utility function which includes weakly separable consumption of goods, services and leisure time, it is straight forward to obtain the following general demand function for time absent

$$\tau_j^a = f\left(t^c, w_j(1-\delta_j), R_j + t^c w \delta_j, \mathbf{s}_j, \varepsilon_j, Dur_j\right) = f\left(t^c, c_j, \mu_j, \mathbf{s}_j, \varepsilon_j, Dur_j\right), \tag{3.1}$$

where $c_j = w_j(1-\delta_j)$ and $\mu_j = R_j + t^c w_j \delta_j$ are the cost and virtual income of being absent, respectively; $\mathbf{s}_j$ is observable and ε_j unobservable personal characteristics and random errors.

As a local approximation, let the conditional demand function at day j in a spell of absence or work be linear, hence

$$\tau_j^a = \mathbf{q}_j'\boldsymbol{\alpha} + \gamma_j + \theta + \eta_j, \tag{3.2}$$

where $\mathbf{q}_j = (t^c, c_j, \mu_j, \mathbf{s}_j')'$, $\boldsymbol{\alpha}$ is a parameter vector, γ_j is a duration parameter, θ is unobservables and η_j is a random error. Assume η_j to be complementary log-log then the probability to be absent (or in work) at day j is

$$\lambda_j = 1 - \exp\left(-e^{\gamma_j + \mathbf{q}_j'\boldsymbol{\alpha} + \theta}\right), j = 1, .., T.$$

This probability to be absent, conditional being absent (or in work), is the discrete time Cox regression model (see e.g. Kalbfleich and Prentice, 1980, Chapter 4).

In the continuous time (t), this model is the proportional hazard model

$$\lambda(t; \mathbf{q}(t)) = h_0(t)\overline{\theta}\exp(\mathbf{q}(t)'\boldsymbol{\alpha}), \tag{3.3}$$

where $h_0(t)$ is the baseline density, $\overline{\theta} = \exp(\theta)$ and $\gamma_j = \ln(\int_{t_{j-1}}^{t_j} h_0(t))$, $t_0 = 0$ and $t_J = \infty$ (see e.g. Kalbfleich and Prentice, 1980, Chapter 4). The continuous time specification is useful since we have the possibility, due to repeated spells, to remove the unobserved heterogeneity term θ using a stratified approach (see below).

Differences in preferences for for absence and difficulties in measuring secondary costs[6] of being absent are two possible causes of unobserved heterogeneity, represented by θ in the model. This heterogeneity may, in turn, be correlated with both c and μ.

Efficiency wage theory predicts that an employer may pay an employee somewhat more than the market wage in order to elicit the employee not to shirk. Some work absence may be interpreted as a form of shirking. Jobs differ in terms of the cost of absenteeism for the employer (see e.g. Weiss, 1985). That is, it may be profitable for an employer to pay some employees more in order to give them incentives which prevent them from being absent from work.

There might also be compensating wage differentials for the option of being absent from work. Jobs that enable a worker to be absent will, other things equal, have a lower wage rate.[7]

Preferences for work-absence are most likely affected by a worker's health status. It is an empirical fact that workers with bad health on average have a higher work-absence rate than workers without health problems. For some jobs, it is reasonable to assume that workers with bad health are less productive and therefore earn less than those with good health status.

Some differences in preferences for work absence may not be driven primarily by health differences. An individual with strong preferences for work absence will, on average during his or her career, be absent more hours. If there are economic returns to on-the-job training, such individuals will, everything else equal, earn less.

[6] This is the cost of being absent in addition to the direct cost; such as a lower probability of being promoted, an increased probability of losing one's job and foregone on-the-job training

[7] Allen (1981b) examines, and finds some support for, this hypothesis empirically.

All four of the above hypotheses will create an correlation between the heterogeneity and the wage; in the hazards to work and work absence we expect, respectively, a negative and a positive relation. Since both the cost of being absent and virtual income are related to the wage rate, unobserved heterogeneity is not likely to be independent of these variables.

3.2. Empirical Specification

Since the females and males may differ in their work absence behavior whether in work or in work absence we allow for state dependent utilities. Hence, we thus specify the hazard in continuous time hazards as

$$\lambda^W(t;\mathbf{z}_{it}) = h_0^W(t)e^{\mathbf{z}_{it}\boldsymbol{\beta}^W+\boldsymbol{\theta}_i^W} = h_0^W(t)\overline{\theta}_i^W e^{\mathbf{z}_{it}\boldsymbol{\beta}^W},\ i=1,\dots,N, \tag{3.4}$$

and

$$\lambda^{WA}(t;\mathbf{z}_{it}) = h_0^{WA}(t)e^{\mathbf{z}_{it}\boldsymbol{\beta}^{WA}+\boldsymbol{\theta}_i^{WA}} = h_0^{WA}(t)\overline{\theta}_i^{WA} e^{\mathbf{z}_{it}\boldsymbol{\beta}^{WA}},\ i=1,\dots,N, \tag{3.5}$$

where W and WA indicate work and work-absence states, respectively; i is an index for the N individuals included in the sample; $\mathbf{z}_{it} = (c_{it}, \mu_{it}, UNEMP_{it}, \mathbf{DAY}'_{it})'$ are time invariant covariates.

As a measure of the secondary cost of a work absence monthly unemployment rate in the county, $UNEMP$, is included in the model. If the unemployment rate is relatively high on the local labor market where the worker is active, the worker's cost of losing his job is likely to be relatively high; the search cost of finding a new job is on average higher in labor markets with high unemployment rate.

For several institutional reasons, the work-absence rate differ between different days of the years. Therefore, the vector **DAY** contains several different indicator variables. Since weekends are not included in the regular work schedule for most workers there is a clear “weekday-pattern” when work-absence spells begins and ends. Therefore, a weekday factor (Mon, Tue, ...Sun) and an indicator variable for public holidays ($Holi$)[8] are included in the specification.

There is anecdotal evidence suggesting that work absence is higher on days between public holidays and Saturdays or Sundays if there happens to be only one day between the public holiday and the weekend. To allow for such an effect, and indeed to test if it is true, we include a dummy variable, BH, which is one for days between public holidays and weekends. Finally, although an insured worker is entitled to compensation on holidays, it is an empirical fact that most workers do not use that possibility. To control for that, an indicator for the month during which most industrial workers are on vacation, $July$, is included.

The heterogeneity is modelled as $\theta_i^k = \mathbf{x}_i' \boldsymbol{\varphi}^k, k = WA, W$, where $\mathbf{x}_i = (\mathbf{CIV}_i',$ AGE_i, $\mathbf{HEALTH}_i', \mathbf{WENV}_i', \mathbf{CONTR}_i')'$; $\mathbf{CIV}$ is a vector of indicators for marital status and number of dependent children; AGE is a the individual's age;[9] **HEALTH** is a vector of indicator variables measuring different aspects of the individual's health status; **WENV** is a vector of indicator variables measuring the individual's work environment; and **CONTR** is a vector of variables measuring employer monitoring. The employer has several means of contributing to lower

[8] Seven days for each year.

[9] One way of measuring differences in the cost of forgone on-the-job training, is to use the well-known result from human capital theory (see Willis, 1986) that the benefits of on-the-job training are higher if such training takes place relatively early in the worker's career, as the wage increase due to improved skills is earned for a longer period of time. The cost of work absence owing to forgone on-the-job training is thus likely to be inversely related to a worker's age. This result cannot be used empirically, however, since a worker's health is also likely to depend on age, which, in turn, affect his preferences for work absence. Therefore, it is not possible to identify the differences in costs of work absence owing to general health depreciation by age.

frequency of work absence. These include direct monitoring as well as pay schemes that provide incentives for the worker to be present. Our data set contains some information that can be used to measure such differences in employer's level of monitoring which are included in the **CONTR** vector. These include whether there is a time-clock at the workplace, $CLOCK$, whether the worker has flexible working hours, $FLEX$, and, finally, whether it is important to be on time, $INTIME$.

Now, since θ_i is modelled by including time-invariant covariates the hazard is the discrete time Cox proportional hazard

$$\lambda_{ij}^{(k)}(\mathbf{z}_i) = 1 - \exp\left(-e^{\gamma_j^k + \mathbf{z}_{ij}'\beta^k + \mathbf{x}_i'\varphi^k}\right), \; j = 1, \dots, T-1, \; i = 1, ..., N \text{ and } k = WA \text{ or } W. \quad (3.6)$$

Since we have multiple spells in both work and work absence it is possible using a stratified approach to test whether or not the included covarites $\mathbf{x}_i$ in (3.6) are sufficient to remove the confounding heterogeneity.

4. Estimation

4.1. Proportional Hazards for Discrete Time Data

The log-likelihood function for the discrete time Cox regression with the parameter vectors $\boldsymbol{\gamma}^k = (\gamma_1^k, ..., \gamma_{T-1}^k)'$, $\boldsymbol{\beta}^k$ and $\boldsymbol{\varphi}^k$is given by

$$\ell(\boldsymbol{\beta}^k, \boldsymbol{\lambda}^k, \boldsymbol{\gamma}) = \sum_{j=1}^{T-1} \left\{ \sum_{l \in D_j} \ln(1 - \exp(-\exp(\gamma_j^k + \mathbf{z}_l'\boldsymbol{\beta}^k + \mathbf{x}_l'\boldsymbol{\varphi}^k)) - \sum_{l \in R_j \setminus D_j} \exp(\gamma_j^k + \mathbf{z}_l'\boldsymbol{\beta}^k + \mathbf{x}_l'\boldsymbol{\varphi}^k) \right\}, \quad k = W, WA,$$

where R_j is the set of observations with spell lengths longer than or equal to j and D_j is the set of observations with spell length j, $j = 1, \ldots, (T-1)$. The variance-covariance matrix of the parameter estimates is, as usual, estimated by the negative of the inverse of the Hessian matrix, inserting the estimates $\hat{\boldsymbol{\gamma}}^k, \hat{\boldsymbol{\beta}}^k$ and $\hat{\boldsymbol{\varphi}}^k$.

A statistically correct method for treating the spells which begin before the start of the time period under study and continue during this period, is to condition on the spell length obtained on January 1, 1990. This could not be done here, however, because we only have information about the state (W or WA) an individual is in at the beginning of the period, not how long he or she has been in that state. Instead, we used two methods for handling this problem: first, a "generous" method where the interval in which a spell starts is counted fully; and second, a "restrictive" approach where a spell is accounted for only in intervals which contain the start of the interval.

Out of a total of 8,145 work spells, as many as 1,319 individuals or 16.2 percent, are left censored for the incidence of work absence. Note that by throwing out these 1,319 spells, we also disregard all individuals who, during the two years studied, are never absent from work. In order to include the work spells starting before January 1, 1990 it is necessary to assume that the intensity of being absent from work on a particular day is independent of the duration in work. Based on the Nelson-Aalen estimates[10] of the cumulative hazards functions for work spells (see Figure 4.1) we think it is reasonable to use the non-restricted sample in the analysis of incidence of work absence.

[10]The cumulative hazard is estimated under the restriction $\boldsymbol{\beta} = \mathbf{0}$. Hence $\hat{\gamma}_j(\mathbf{0}) = \ln(-\ln(1 - D_j/n_j))$,where n_j is the number of observations at risk in R_j.

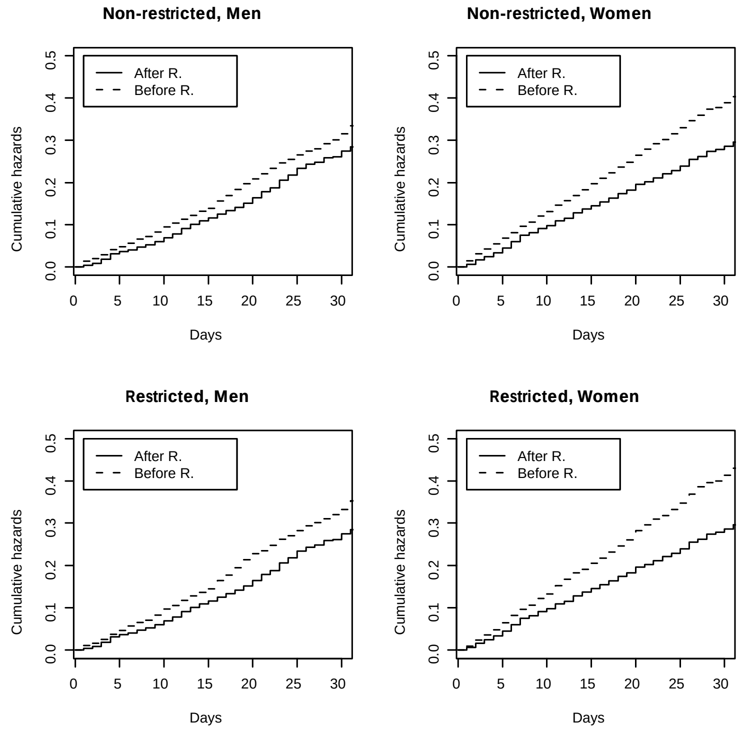

Figure 4.1: The Nelson-Aalen estimator of the cumulative hazard function before and after the reform (R) for the male and female samples, respectively. The estimates in the upper panels are based on the full sample and the estimates in the lower panels are based on the restricted sample, i.e., including only those spells which started after January 1, 1990.

It is quite clear from Figure 5.1 that it is not appropriate to include the left censored observations for the duration in work absence. However, the left truncation problem is not likely to be very severe because out of 6,911 spells, only 71 individuals entered as absent from work and these spells can safely be neglected since there are still 6,840 work-absence spells to consider.

4.2. Stratified Analysis

The effect of the time-varying covariates may be examined using a stratified analysis (see e.g. Lancaster, 1990, Chapter 9). The partial likelihood for the m_i uncensored spells of individual i is

$$L_i(\boldsymbol{\beta}^k) = \prod_{\kappa=1}^{m_i} \frac{\lambda^k(t_{i(\kappa)}; z_{ij_\kappa})}{\sum_{l \in R_{ij_\kappa}} \lambda^k(t_{i(\kappa)}; z_{il})}, k = W, WA \tag{4.2}$$

where $\lambda^k(t_{i(\kappa)}; z_{ij_k})$ is given in (3.4) or (3.5), R_{ij_k} is the risk set at $t_{i(\kappa)}$ for individual i and j_κ labels the $\kappa : th$ largest spells, $t_{i(\kappa)}$. Note that $\overline{\theta}_i^k$ and $h_0^k(t)$, $k = WA, W$ cancels from expression (4.2). The total partial likelihood is simply

$$L(\boldsymbol{\beta}^k) = \prod_{i=1}^{N} L_i(\boldsymbol{\beta}^k), k = W, WA. \tag{4.3}$$

Standard errors were estimated as if we had an ordinary likelihood function, using an asymptotic maximum likelihood approach.

5. Data and Descriptive Statistics

5.1. Data Sources and Measurement

We use the 1991 Swedish Level of Living Survey (SLLS). The SLLS is a microdata set that contains information compiled from interviews as well as official public registers for a random sample of about 6,000 individuals. This survey is described in detail in Fritzell and Lundberg (1994). Data on the dependent variable, absence from work, were obtained from the National Social Insurance Board by matching with the SLLS sample.

The definition of work absence is that an individual is compensated by the compulsory sickness insurance system a particulat day. As the data were collected from registers of actual payments to insured individuals, there are likely to be much less measurement errors as compared to self-assessed data. However, if we define work absence as time during which an employee is absent from work without prior agreement with the employer (such as holidays), then a small fraction of work absence is not likely to be included in the sickness insurance data.[11]

We restricted the sample to blue-collar workers aged between 20 and 64 who were employed during 1991 (the year of the survey). The final sample consisted of 1,396 individuals (738 males and 658 females). The motive for restricting the population to blue-collar workers was to limit heterogeneity arising from differences in sickness insurance schemes. Swedish white-collar workers often have negotiated schemes whose rules cannot be obtained from the available data. Tables A7.1 and A7.2 in Appendix 2 provide descriptive statistics on all variables included in our analysis.

Measuring the two variables for economic incentives encompassed by the econometric model, the cost of being absent (c) and virtual daily income from being absent (μ), involves several steps. We began by calculating the hourly real wage rate. First, we computed the income from labor a worker would have received if he or she had not been absent from work during 1990 and 1991 (potential income from labor), i.e., we added the share of income not covered by sickness insurance for each day the worker was absent during the year. Data on income from labor were compiled from tax registers matched with the SLLS survey. It was then straightforward to calculate the cost of being absent from work using the pre- and post-reform replacement levels in the insurance schemes, respectively. We then calculated and deducted income taxes from the potential income. Finally, we used

[11] According to one survey, the amount was 2.9 percent in 1986 (SAF, 1986).

the number of hours of work stated by the worker in the 1991 SLLS pertaining to 1990 to obtain the hourly wage rate.[12]

We calculated virtual income as the daily income received from sickness insurance when a worker is absent from work. We also added observed labor income for the spouse if a worker is married as well as family income from capital, child and housing allowances. Data on all of these components were obtained from tax registers matched with the SLLS survey.

5.2. Description of Spells of Work and Work Absence

Figures 5.1-5.4 show Kaplan-Meier estimates of the survival function for work-absence spells as well as work spells by different classifications. Figure 5.1 shows the overall male-female difference in exit rates from work and work-absence spells. The first panel reveals male-female differences in work spells. Although the difference is very small, this panel discloses that the graph for the male subsample exceeds the graph for the female subsample. This difference reflects the fact that women have a higher incidence of work absence.

The second panel in Figure 5.1 displays the work-absence survival function for males and females. Both survival functions in this panel show a similar pattern: a steep decrease until day seven, followed by a relatively flat segment. This clear-cut pattern is due to legislation whereby a certificate from a physician is required after day seven in a work-absence spell. Moreover, the hazard for females is higher during spells of up to five days as compared with males. For longer spells, the survival functions are quite similar.

[12]We thus assume that the worker do not change his or her regular hours of work between 1990 and 1991.

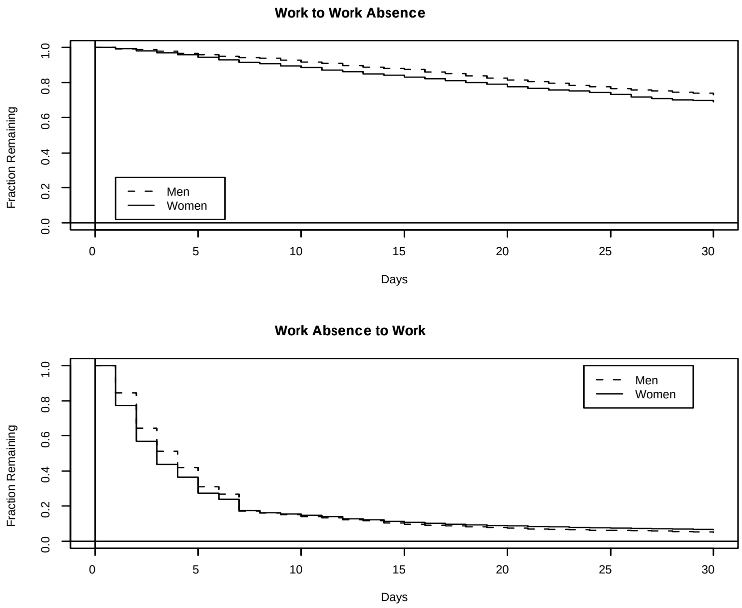

Figure 5.1: Kaplan-Meier non-parametric estimates of duration of work spells (panel 1) and work-absence spells (panel 2); men and women, respectively.

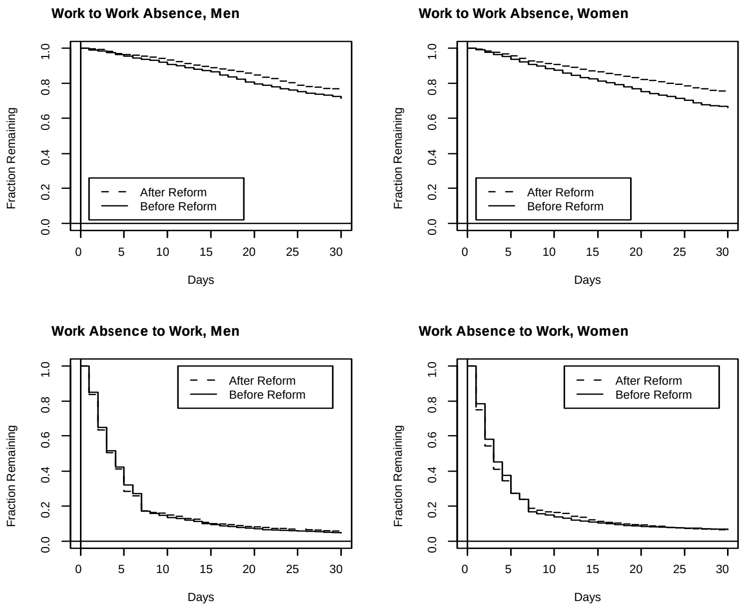

Figure 5.2: Kaplan-Meier non-parametric estimates of the effect of the March 1991 reform of the sickness insurance system on work-absence behavior; duration of work spells for males (panel 1); duration of work spells for females (panel 2); duration of work-absence spells for males (panel 3); duration of work-absence spells for females (panel 4).

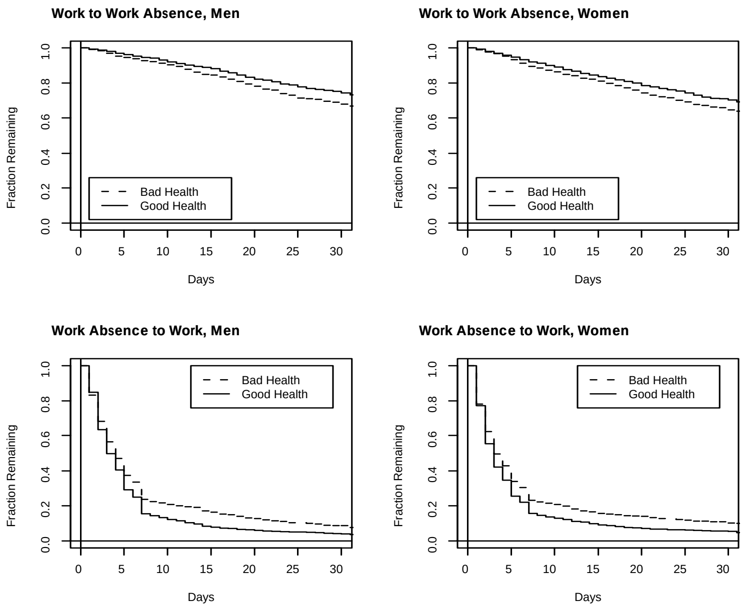

Figure 5.3: Kaplan-Meier non-parametric estimates of the effect of health status on work-absence behavior; duration of work spells for males (panel 1); duration of work spells for females (panel 2); duration of work-absence spells for males (panel 3); duration of work-absence spells for females (panel 4).

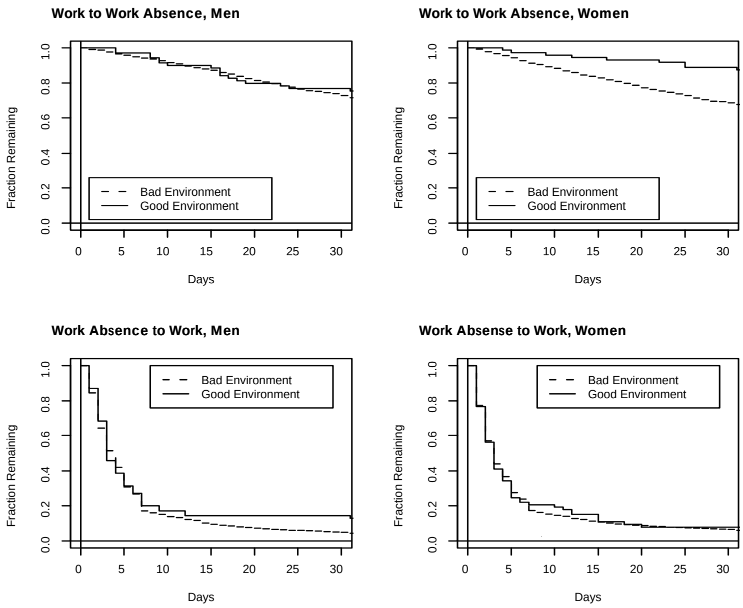

Figure 5.4: Kaplan-Meier non-parametric estimates of the effect of work environment on work-absence behavior; duration of work spells for males (panel 1); duration of work spells for females (panel 2); duration of work-absence spells for males (panel 3); duration of work-absence spells for females (panel 4).

Figure 5.2 shows the effect of the 1991 reform of the sickness insurance system on the work-absence behavior. The first two panels in Figure 5.2 display the effect on survival in work for men and women, respectively. The incidence of work absence decreased markedly after the reform. It can also be seen that women changed their behavior somewhat more than men.

The third and fourth panels of Figure 5.2 show that the hazard increases somewhat during the first five or six days for both men and women. After that, the relation is reversed; the work-absence spells tend to become longer. To some extent, these changes in the shape of the survival function correspond to the changes in economic incentives implied by the reform, in the sense that the largest decrease in the replacement level of sickness insurance, i.e., the greatest increase in the cost of being absent, pertains to the first three days of an absence spell. Both genders seem to react very similarly to the reform.

The survival functions of individuals with different health status are shown in Figure 5.3. Those with bad health are defined as having at least one indication of a health problem among 13 indicator variables used to characterize health differences among the individuals in the sample (see Tables A7.1 and A7.2 for descriptions as well as descriptive statistics of these variables). About 21.1 percent if the individuals in the sample meet the definition of bad health (19.8 percent among the women and 22.4 percent among the men).

The first and second panels of Figure 5.3 show that workers with bad health status have a somewhat higher exit rate from work spells. The difference is very similar for men and women. The third and fourth panels indicate that the exit rates for individuals with good health are always higher than the exit rates from work-absence for those who are in bad health. Men with bad health exhibit a much lower exit rate than men with good health. The difference in exit rates is smaller for women. As regards long absence spells, however, the difference between men and women is very small.

The effects of poor working conditions are explored in Figure 5.4. The strategy used to define bad health was also applied poor working conditions. In the case of work environment, 97.2 percent of the sample (96.8 percent among the women and 98 percent among the men) are defined as having poor working conditions, i.e., with at least one indication of poor working conditions among the 13 indicator variables used to define individual differences in work environment.[13]

The survival functions with respect to work spells for the females and males are given in panel 1 and 2. As expected, the exit rates for women who work in a poor environment are larger than for those in a good environment. However, no effect was found in the male sample; the third and fourth panels of Figure 5.4 show that the exits rate are very similar for both good and poor working conditions.

6. Results

The results from the discrete time Cox regression models (henceforth: discrete time model) are shown in Tables 6.1 and 6.2. Table 6.1 reports the results for the duration of the work spells, i.e., the incidence of the work absence periods, and Table 6.2 the corresponding ones for the duration of the work absence spells. Two different specifications were estimated: one with the full set of covariates and another, restricted version, with only time-varying covariates (the same as those used in the stratified analysis for the sake of comparison). The results from

[13]Figure 5.4 should be interpreted with caution since only 39 individuals (21 females and 18 males) had no indication poor working conditions.

the stratified analysis, for both the work and work absence spells, are reported in Table 6.3.

The discussion of the results are divided into three sub-sections. In the first one, Section 6.1, we discuss the estimates for the time-varying covariates, i.e., the economic incentives, the unemployment rate, indicators for holidays and weekdays. This sub-section applies to both the discrete time models and the stratified analysis. Section 6.2 discusses the estimates for the time invariant covariates which are included in the discrete time models to control for heterogeneity: contracted hours of work, personal characteristics, health status indicators and measures of work environment. Finally, Section 6.3 discusses gender differences in work absence behavior from all the results obtained in this study.

6.1. Time-Varying Covariates

The results show that the cost of absence from work has a significant negative impact on the incidence of work absence for both men and women in all three models. The magnitude of the estimates are also very similar. This result implies that we cannot reject the hypothesis that the cost variable is uncorrelated with variables included in the first specification but not the second and with unobserved variables.

The effect of cost on the duration in work absence is much weaker. The only significant coefficient estimate (at the 10 percent level) is for females in the stratified model. The estimate in this model is, as expected, positive. It is noteworthy that increasing the control for heterogeneity leads to larger positive coefficients for both the female and male samples. Since the stratified model is less efficient than the discrete time models, it is plausible that unobserved heterogeneity does affect the estimates in the discrete time model.

Contrary to what is predicted by the economic model, the estimates for the virtual income variable are significantly negative in both specifications of the discrete time models for the male subsample. In the stratified model, however, they are insignificant. That is, the results from the discrete time models could have been caused by unobserved heterogeneity. The same pattern emerge in the female subsample, although the estimates for virtual income are only significant in the specification without controls for observed heterogeneity. For the hazard from work absence, the virtual income variable is insignificant in all specifications except for women in the discrete time model with only time-varying covariates, where the estimated coefficients are (as expected) negative.

According to the discussion in Section 3, we expect a positive (negative) coefficient for the unemployment rate on the hazard from (incidence of) work absence. The only significant estimate we obtained from this parameter is for the female subsample in the stratified model, where - contrary to expectations - the estimate in the model for the hazard from the work absence state is negative.

These results differ from those generally obtained on aggregate data (see e.g. Lantto and Lindblom, 1987)[14] as well as in previous studies where no attempts were made to control for correlation between unobserved heterogeneity and regressors. There are, however, explanations as to why we may observe a spurious relation between work absence and unemployment. It is common knowledge that work absence is lower in firms with few employees (see e.g. SAF, 1986) and in small communities, due to more extensive social control. The unemployment rate in Sweden, and in most other countries, is in general higher in small communities.

[14]This result is, however, questioned by Bäckman (1998).

This may account for the previously observed negative correlation between work absence and unemployment. Furthermore, according to the results obtained from aggregate data, the labor force changes when the unemployment rate increases: those who become unemployed may have previously had a higher rate of work absence than the rest of the labor force.

There are several conceivable explanations to the observed result of a positive correlation between the unemployment rate and the incidence of work-absence spells. When aggregate demand decreases at the beginning of a recession, firms will try to lay off their workers rather than permanently dismiss them in order to avoid hiring, training and firing costs, if they expect demand to increase again later on. It is much cheaper, although not legal, for a firm if laid-off workers claim sickness benefits rather than get paid by the firm. There is anecdotal evidence that employers may in fact urge their employees to do so. If this were the case, we may also observe a positive correlation between the unemployment rate and work absence at the beginning of a recession, when the unemployment rate is increasing. Moreover, a rising unemployment rate may be financially stressful and thereby detrimental to health (see e.g. Vahtera, Kivimäki and Pentti, 1997). This, in turn, may lead to a higher rate of work absence.

The effects of weekdays (with Monday as the reference), $Holi$, BH and $July$ are very similar for all estimators. The incidence of absence from work is the highest for Sundays, i.e., most sickness spells begin on a Monday. Thereafter, the incidence is (almost) monotonously decreasing until Saturday, which is exactly as expected. The results for the durations of work-absence spells show that an individual are less likely to leave a spell of absence on a Saturday. This is because most workers do not work on Sundays.

The results also indicate that work-absence spells are less likely to begin during public holidays, days between work-free days and during July (*Holi*, *BH* and *July*). The results for the *BH* coefficient provide no support for the anecdotal evidences that work absence is higher for such days. However, the fact that some workers use vacation days and that some employers give their employees an extra day off on days between holidays and/or work-free days may counteract an increased rate of work absence due to abuse of the insurance schemes that may also be inherent in the data.

6.2. Time-Invariant Covariates

The expected effect of contracted number of hours of work, t^c, is that more contracted hours of work lead to higher rate of work-absence. This is also found in our female subsample: the estimate is significantly positive for the incidence of work absence. The result indicates a difference between the gender groups in this respect.

Two sets of variables were used to describe differences in family composition: indicators for being unmarried, single or divorced ("married" is the omitted category) and indicators for number of children ("no children" is the omitted category). In interpreting the negative coefficient estimates for several of the "number of children" indicators, it should be kept in mind that, in Sweden, care of dependent children while ill is covered by a separate insurance, with a somewhat higher replacement level for most insured workers, i.e., the result can be driven by abuse of this scheme.

The coefficient estimates of the age variable are significantly negative for both men and women in the work-absence state, indicating that older workers on average have longer work-absence spells. In the work state, the coefficient estimates for the age variable are still negative for both men and women, indicating that older workers on average have fewer work-absence spells. One interpretation of these results is that they simply reflect differences in preferences between older and younger workers. They could, however, also be related to selection over time: workers with high preferences for being absent either exit the labor force or remain in long work-absence spells, as time evolves. Workers with low preferences for being absent will then constitute a larger share of the older workers in the work state.

We used 14 different health indicators. Each of them measures a specific health problem except $DISAB$, which indicates whether or not a permanent physical handicap prevents a worker from taking all possible jobs. Descriptive statistics and a short description of each of these variables are given in Tables A7.1 and A7.2 in Appendix 2. The health indicators are jointly significant (at all reasonable levels of risk) for both states and for both genders. In addition, several health indicators are individually significant. This tells us that health status is, as expected, an important determinant of differences in preferences for absence from work.

We have included two different types of measures of work environment: 13 subjective measures of workplace characteristics and two occupation-specific measures of risk exposure (SIR - standardized incidence ratios) for work accidents and work-related diseases, respectively. Tables A7.1 and A7.2 contains descriptive statistics along with a brief description of each of these indicators.

When interpreting the effects of the work environment variables on absence from work, it should be kept in mind that these results are likely to be affected by selection of physically (or mentally) strong workers into demanding jobs - the so-called "healthy worker effect" (see e.g. Östlin, 1989). The negative estimate for work spells for "jobs with heavy lifting" ($LIFT$) and "mentally exhaustive jobs" ($EXHM$) for men can be interpreted as a result of selection. However, several work characteristics, such as jobs with "unpleasant body positions" (UBP) for both men and women and "contact with smoke" ($SMOKE$) for women, seem to increase work absence, i.e. they are significantly positive in the work state.

The coefficient estimates for the $CLOCK$ and $INTIME$ indicator variables have the opposite sign from what was expected according to the discussion in Section 4. $INTIME$ is significant for men in work spells and $CLOCK$ is significant for women in work absence. A possible explanation for these results is that timeclocks are used primarily when other forms of monitoring are not feasible, e.g. in large firms. These results indicate that other types of monitoring are likely to be more efficient in decreasing the rate of work absence. They might also reflect the fact that unplanned work absence may be recorded more easily when a timeclock is punched. The parameter estimates may therefore indicate that the measures of work absence do not in fact include all forms of unscheduled absence from work.[15]

6.3. Overall Male-Female Differences

As noted in the Introduction, the male-female difference in work absence behavior has emerged as a "stylized fact" from empirical research on work absence. The estimated models can be used to analyze to what extent the differences in observable characteristics can explain the observed behavior. This analysis is based on

[15]In a survey of time allocation (SAF, 1986) about 97.1 % of all unscheduled work absence for blue-collar workers was found to be covered by sickness insurance.

the comparison of the effect of subsets of variables included in the model on the predicted mean duration of work or work absence (Appendix 1 gives the details of the method used to estimate the mean given a censored sample). In addition, this exercise gives a measure of the economic significance of the estimated effects, since gender differences in observed characteristics can easily be studied by the reader.

We calculate the mean durations in work and work absence for males ($E(\overline{\mathbf{w}}'_m\widehat{\boldsymbol{\theta}}_m)$) and females ($E(\overline{\mathbf{w}}'_f\widehat{\boldsymbol{\theta}}_f)$), where $\overline{\mathbf{w}}_q$ and $\widehat{\boldsymbol{\theta}}_q$, $q = m, f$ are the mean values of the co-variates and the estimated parameters from the discrete time model, respectively. The mean difference in duration for a subgroup k of variables is then estimated as

$$\Delta^q(k) = E(\overline{\mathbf{w}}_q(k)'\widehat{\boldsymbol{\theta}}_q) - E(\overline{\mathbf{w}}'_q\widehat{\boldsymbol{\theta}}_q), \quad q = m, f, \tag{6.1}$$

where $\overline{\mathbf{w}}_m(k) = (\overline{\mathbf{w}}_f^{k'}, \mathbf{w}_m^{j'})'$ and $\overline{\mathbf{w}}_f(k) = (\overline{\mathbf{w}}_m^{k'}, \overline{\mathbf{w}}_f^{j'})'$, $\overline{\mathbf{w}}_q^{k'}, q = m, f$ are the mean of the subgroup k of variables that are taken from the opposite gender and $\overline{\mathbf{w}}_q^{j}, q = m, f$ are the means for the sub-group of explanatory variables that are constant. If $\Delta^q(k) < 0, q = m, f$, this implies an increasing hazard to work absence or work from the change in mean values from $\overline{\mathbf{w}}_m^{k'}{}'$ to $\overline{\mathbf{w}}_f^{k'}{}'$, or from $\overline{\mathbf{w}}_f^{k'}{}'$ to $\overline{\mathbf{w}}_m^{k'}{}'$.

Males are predicted to have on average 11.5 days longer work spells than females while only 0.3 days shorter work absence spells. Hence, the result that females on average have a higher work absence rate (see Tables A7.1 and A7.2) can be attributed to more frequent, rather than longer, work absence spells. We will therefore concentrate the analysis on the gender differences in the frequencies of the work absence spells.

Table 6.4 shows the results of the comparison for six different groups of variables. The second row shows that if the females would have had the male mean cost of being absent, then the mean duration of the work spell would increase by 3 days (from 85 to 88 days). If, on the other hand, the males would have had the females mean cost, the males mean duration would decrease by 4 days (from 96.5 to 92.5 days). Since the estimates of the parameters for the cost of being absent are very similar for males and females,[16] this difference is due to the on average higher cost for males of being absent.

Row 3 shows that a decrease in contracted labour time for the males from 38.94 to 33.92 would decrease the hazard to work absence by one day[17] while an increase of the contracted labour time for the females from 33.92 to 38.94 would increase the hazard to work absence by 1.5 day. That is, interpreting this as a contrafactual result, the male-female difference in work absence behavior would have been exaggerated if female workers would have had the same average number of hours of work as the male workers.

The results on work environment works in the same direction: men are on average exposed to inferior work environment compared to women and using male covariates in the female equation again exaggerates the differences in work absence. One should, however, be careful in making a causal interpretation of this result. It is likely that workers with a strong health select themselves into physically demanding jobs (the so called "healthy worker effect" which is discussed in Section 6.3). If such an effect is present, it would result in a negative bias of the causal effect of work environment on work absence.

Finally, Table 6.4 shows that male-female differences in health status seem to

[16]This is true also from the result of the stratified model.

[17]Note, however, that this parameter is insignificant for the males subsample.

have very little effect on observed behavior.

To sum up, of the total difference in predicted number of days of 11.5 days we found that about 4 of those days can be explained by differences in cost of being absent. All other observable characteristics works in the other direction, i.e., using the male characteristics in the female equation exaggerates the difference. This means that most of the observed difference in male-female work absence behavior can be attributed to gender differences in unobserved characteristics, i.e., intrinsic differences in work absence behavior.

Table 6.1: Results from the discrete time hazard model for the incidence of work absence.

	Males				Females			
	Coef	p-val	Coef	p-val	Coef	p-val	Coef	p-val
c	-0.031	0.000	-0.033	0.000	-0.023	0.000	-0.029	0.000
μ	-0.063	0.000	-0.108	0.000	0.023	0.136	-0.041	0.001
UNEMP	-0.007	0.760	-0.016	0.451	-0.021	0.360	-0.029	0.189
Tue	-0.020	0.745	-0.020	0.756	0.101	0.112	0.101	0.110
Wed	-0.272	0.000	-0.273	0.000	-0.006	0.926	-0.005	0.934
Thu	-0.634	0.000	-0.633	0.000	-0.315	0.000	-0.317	0.000
Fri	-1.969	0.000	-1.970	0.000	-1.303	0.000	-1.304	0.000
Sat	-2.515	0.000	-2.513	0.000	-2.103	0.000	-2.103	0.000
Sun	0.417	0.000	0.419	0.000	0.388	0.000	0.392	0.000
Holi	-0.210	0.051	-0.210	0.051	-0.221	0.033	-0.209	0.044
BH	-1.059	0.003	-1.055	0.003	-0.606	0.030	-0.587	0.036
July	-0.616	0.000	-0.598	0.000	-0.754	0.000	-0.715	0.000
t^c	-0.002	0.779			0.007	0.036		
Unmarried	-0.099	0.105			-0.172	0.005		
Divorced	0.255	0.000			0.231	0.004		
One child	-0.126	0.041			-0.047	0.354		
Two children	-0.270	0.000			-0.168	0.006		
Three children	-0.117	0.276			-0.258	0.012		
Four children	-0.447	0.035			-0.278	0.205		
Five children	-0.376	0.460			-1.012	0.084		
Six children	-0.138	0.726						
Age	-0.010	0.000			-0.012	0.000		
DISAB	0.765	0.000			0.896	0.000		
NOISE1	-0.036	0.575			0.299	0.000		
NOISE2	0.110	0.023			0.049	0.306		
SMOKE	0.059	0.218			0.284	0.000		
SHAKE	0.090	0.117			-0.117	0.416		
POISON	0.044	0.413			-0.020	0.815		
LIFT	-0.127	0.005			0.138	0.012		
HARD	-0.012	0.803			0.009	0.856		
SWEAT	0.275	0.000			0.036	0.440		
EXHM	-0.136	0.002			-0.013	0.768		
STRESS	0.032	0.460			-0.061	0.168		
REP	0.246	0.000			0.059	0.216		
MOM	0.047	0.288			-0.061	0.188		
UBP	0.142	0.003			0.176	0.000		
RISK1	0.000	0.013			-0.000	0.296		
RISK2	-0.000	0.024			-0.000	0.429		

Table 6.1 continued

FLEX	0.024	0.587	-0.067	0.148
CLOCK	0.071	0.091	0.051	0.324
INTIME	0.131	0.008	0.038	0.482
STRUMA			0.319	0.006
TBC	0.908	0.003	0.500	0.318
HEART	0.312	0.060	0.753	0.000
HBLOOD	0.131	0.091	0.161	0.056
ULCER	0.099	0.381	0.321	0.011
HEMORR	0.309	0.001	0.106	0.267
PREGNANT			0.211	0.048
HERNIA	-0.167	0.478	-0.165	0.715
VAV	0.346	0.001	0.083	0.313
MENTAL	0.018	0.945	-1.291	0.029
CANCER	0.072	0.706	-0.031	0.878
DIABETIC	0.022	0.900	0.509	0.001
NEURO	0.129	0.539	-0.622	0.050

7. Conclusions

Like a number of previous papers on work absence, this study supports the view that economic incentives affect work absence behavior. It is shown that differences in costs for being absent can explain about one third of the observed male-female difference in the frequency of work absence spells. This, in turn, implies that a smaller gender wage gap will decrease the differences in the observed work absence behavior between men and women.

Another interesting result is that women seem to be more sensitive to exposure to bad work conditions in their work absence behavior. This was seen directly in the Kaplan-Meier survival graphs, where the difference between those exposed and not exposed to bad work condition were larger for the females than for the males, as well as when predicting the change in work absence behavior from using the males workplace attributes in the female equations. A somewhat related result,

Table 6.2: Results from the discrete time hazard model for the hazard from work absence.

	Males				Females			
	Coef	p-val	Coef	p-val	Coef	p-val	Coef	p-val
c	0.005	0.185	0.005	0.217	0.001	0.850	-0.002	0.544
μ	0.017	0.284	-0.002	0.892	-0.032	0.015	-0.054	0.000
UNEMP	0.011	0.572	0.011	0.556	0.005	0.818	0.010	0.605
Tue	0.039	0.583	0.038	0.596	0.193	0.007	0.198	0.006
Wed	0.027	0.710	0.026	0.717	0.222	0.002	0.227	0.001
Thu	-0.168	0.025	-0.165	0.028	0.119	0.099	0.120	0.095
Fri	1.019	0.000	1.013	0.000	1.023	0.000	1.017	0.000
Sat	-0.953	0.000	-0.967	0.000	-0.714	0.000	-0.720	0.000
Sun	0.300	0.000	0.296	0.000	0.403	0.000	0.401	0.000
Holi	-0.058	0.613	-0.099	0.382	-0.109	0.355	-0.085	0.472
BH	0.037	0.858	-0.013	0.949	0.005	0.981	-0.014	0.946
July	-0.213	0.009	-0.193	0.016	-0.234	0.011	-0.218	0.017
t^c	-0.002	0.683			0.003	0.274		
Unmarried	1.041	0.488			-0.005	0.927		
Divorced	0.997	0.977			-0.057	0.452		
One child	-0.013	0.824			-0.039	0.408		
Two children	-0.037	0.558			-0.065	0.249		
Three children	0.158	0.121			-0.143	0.125		
Four children	0.161	0.393			-0.486	0.015		
Five children	0.531	0.247			-1.044	0.075		
Six children	-0.226	0.541						
Age	-0.013	0.000			-0.009	0.000		
Disab	-0.589	0.000			-0.358	0.004		
NOISE1	0.183	0.003			-0.069	0.374		
NOISE2	0.069	0.119			-0.050	0.263		
SMOKE	-0.034	0.449			0.024	0.623		
SHAKE	-0.184	0.001			-0.135	0.323		
POISON	-0.041	0.409			0.221	0.008		
LIFT	0.051	0.225			0.019	0.716		
HARD	0.008	0.866			0.013	0.759		
SWEAT	-0.068	0.113			-0.111	0.012		
EXHM	-0.006	0.886			0.108	0.011		
STRESS	0.036	0.368			-0.072	0.074		
REP	0.031	0.470			-0.063	0.160		
MOM	-0.164	0.000			-0.048	0.261		
UBP	0.029	0.518			-0.073	0.113		
RISK1	0.000	0.914			0.000	0.635		
RISK2	-0.000	0.559			-0.000	0.495		

Table 6.2 continued

FLEX	-0.027	0.502	-0.137	0.002
CLOCK	-0.036	0.356	-0.207	0.000
INTIME	-0.006	0.903	-0.033	0.510
STRUMA	0.788	0.308	-0.015	0.895
TBC	0.475	0.114	-0.458	0.356
HEART	-0.197	0.211	-0.466	0.011
HBLOOD	-0.112	0.132	-0.297	0.000
ULCER	-0.013	0.897	-0.109	0.357
HEMORR	0.118	0.209	0.078	0.380
PREGNANT			-0.257	0.008
HERNIA	-0.231	0.267	-0.708	0.074
VAV	-0.142	0.134	-0.023	0.764
MENTAL	-0.356	0.159	-0.010	0.982
CANCER	-0.682	0.000	-0.275	0.153
DIABETIC	-0.168	0.310	-0.311	0.026
NEURO	0.133	0.512	0.072	0.799

Table 6.3: Results from the stratified analysis.

	Males				Females			
	Hazard from		Incidence		Hazard from		Incidence	
	Coef	p-val	Coef	p-val	Coef	p-val	Coef	p-val
c	0.007	0.350	-0.026	0.000	0.013	0.061	-0.033	0.000
μ	-0.003	0.946	0.004	0.924	-0.006	0.839	0.033	0.271
UNEMP	-0.044	0.291	0.003	0.945	-0.117	0.007	0.064	0.176
Tue	0.019	0.842	0.005	0.951	0.189	0.040	0.110	0.189
Wed	-0.038	0.692	-0.205	0.018	0.239	0.008	-0.125	0.142
Thu	-0.233	0.020	-0.579	0.000	0.146	0.115	-0.327	0.000
Fri	1.271	0.000	-2.060	0.000	1.237	0.000	-1.308	0.000
Sat	-1.142	0.000	-2.765	0.000	-0.630	0.000	-2.136	0.000
Sun	0.031	0.768	0.418	0.000	0.273	0.012	0.360	0.000
Holi	-0.143	0.389	-0.198	0.140	-0.329	0.055	-0.294	0.020
BH	-0.226	0.448	-0.544	0.156	-0.116	0.682	-0.544	0.090
July	-0.217	0.055	-0.817	0.000	-0.225	0.060	-0.864	0.000

Table 6.4: The mean difference (in days) hazards to work absence when performing the experiment of using the means of the observed variables for females in the male equation and the mean of the male characteristics in the female equation respectively.

Group of variables	$\Delta^m(k)$	$\Delta^f(k)$
Personal characteristics	2	-2
c	-4	3
Contracted hours of work (t^c)	1*	-1.5
Work environment	9	-11.5
Health status	-0.5	-0.5
Secondary cost	-0.0	-0.5

* This parameter was not significant in the Cox regression model

which is suported in the theoretical model, is that the contracted number of hours of work affects work absence. The results of the predictions indicate that the gender differences in work absence would have been greater if women on average would have worked the same number of hours as males.

Finally, our results also show that most of the male-female difference in work absence behavior cannot be explained by differences in characteristics included in the estimation. Given the detailed characteristics included, the background to this result is likely to be intrinsical differences in male-female work absence behavior. This implies that the compulsory sickness insurance with a premium which is more or less proportional to the insured income will redistribute income from men to women compared to an insurance market, where the insurer is able to price discriminate between male and female workers. This applies even if the economic incentives to be present at work would have been the same for men and women.

References

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.

Bäckman, O. (1998). "Longitudinal Studies on Sickness Absence in Sweden", Ph.D. Thesis, Swedish Institute for Social Research, Stockholm.

Barmby, T., C. Orme and J. Treble (1996). "Worker Absence Histories: A Panel Data Study", *Labour Economics* **2**, 53-65.

Barmby, T. and S. Sibly (1999). "A Markov Model of Worker Absenteeism", mimeo, University of Newcastle upon Tyne.

Fritzell, J. and O. Lundberg (1994). *Vardagens Villkor. Levnadsförhållanden i Sverige under Tre Decennier (*Everyday Situations. Living Conditions in Sweden during Three Decades*)*, Stockholm: Brombergs Förlag.

Johansson, P. and M. Palme (1996). "Do Economic Incentives Affect Worker Absence? Empirical Evidence Using Swedish Data", *Journal of Public Economics* **59**, 195-218.

Johansson, P. and M. Palme (2002). "Assessing the Effect of Public Policy on Worker Absenteeism", *Journal of Human Resources*, **37**, 381-409.

Kalbfleich, J. D. and R. L. Prentice (1980). *The Statistical Analysis of Failure Time Data*, New York: Wiley.

Kangas, O. (1991). "The Politics of Social Rights. Studies in the Dimensions of Sickness Insurance in the OECD Countries", Ph.D. Thesis, Swedish Institute for Social Research, Stockholm.

Kaplan, E.L. and P. Meier (1958). "Nonparametric Estimation from Incomplete Observations", *Journal of the American Statistical Association* **53**, 457–481.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data,* Cambridge: Cambridge University Press.

Lantto, K. (1991). "Optimal Deterrents to Malingerence", Ph.D. Thesis, Department of Economics, Stockholm University.

Lantto, K. and E. Lindblom (1987). "Är Arbetslösheten Hälsosam" (Is Unemployment Healthy), *Ekonomisk Debatt* **15** 333-336.

National Social Insurance Board (1997). "Facts about Swedish Social Insurance 1995". Stockholm: RFV.

Nilsson, M. (2001) "Studies of Work Absence due to Own Illness", Thesis, Department of Economics, Lund University.

Östlin, P. (1989). *Occupational Career and Health : Methodological Considerations on the Healthy Worker Effect.* Stockholm: Almqvist and Wiksell International.

Palme, M. and I. Svensson (1998). "Social Security, Occupational Pensions, and Retirement in Sweden", in *Social Security and Retirement Around the World,* J. Gruber and D. Wise, eds. Chicago: Chicago University Press.

Paringer, L. (1983). "Women and Absenteeism: Health or Economics?", AEA

Papers and Proceedings 73, 123-127.

SAF (1986). “Tidsanvändningsstatistik” (Time Allocation Survey) Swedish Employers’ Federation, Stockholm.

Sindelar, J. (1982). “Differential Use of Medical Care by Sex”, Journal of Political Economy 90, 1003-1019 [ej citerad].

Vahtera, J., M. Kivimäki and J. Pentti. 1997. “Effect of Downsizing on Health of Employees.” *The Lancet* **350**, 1124-1128.

VandenHeuvel, A. and M. Wooden (1995). “Do Explanations of Absenteeism Differ for Men and Women”, *Human Relations* **48**, 1309-1329.

Weiss, A. (1985). “Absenteeism and Wages”, *Economics Letters* **19**, 277-279.

Willis, R. (1986). “Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions”, in O. Ashenfelter and R. Layard, eds., *Handbook of Labor Economics*, Amsterdam: North-Holland.

Vistnes, J. P. (1997). “Gender Differences in Days Lost from Work due to Illness”, *Industrial and Labor Relations Review* **50**, 304-323.

Appendix 1

The problem of estimating the mean of a random variable, given a censored sample, was addressed as follows. From the censored sample, we estimated the survival function by the method given in Kaplan and Meier (1958). This estimator is a discrete distribution function. We calculated the mean in this distribution and used it as our estimator of the mean in the original distribution.

This estimator has two obvious properties: (i) it is easy to calculate, and (ii) it coincides with the sample mean when there is no censoring. Its general large-sample properties in the case of uninformative censoring remain to be investigated. A negative bias might be expected, since the largest values tend to be censored more frequently than small values.

Suppose that, from (a censored) sample, we have estimated the survival function S of the random variable T by $\hat{S}$, where $\hat{S}$ has support $t = (t_1 < t_2 < \dots, < t_k)$. Then $\hat{S}$ is the survival function of a discrete random variable T^*, with

$$P(T^* = t_i) = \hat{S}(t_i) - \hat{S}(t_i+), \ i = 1, \dots, k.$$

The mean in this distribution is

$$E\left(T^*\right) = \sum_{i=1}^{k} t_i \left(\hat{S}(t_i-) - \hat{S}(t_i)\right).$$

Now we obviously define our estimator as

$$\hat{E}\left(T\right) = E\left(T^*\right). \tag{7.1}$$

Suppose for the moment that censoring is of *Type I*, i.e., the items has a common (potential) censoring time $t = T_0$. Then our estimator $\hat{E}\left(T\right)$, given in (7.1), is an unbiased estimator of the *conditional* expectation $E(T|T \leq T_0)$ which, of course, is smaller than the quantity we want to estimate. It would also be impossible to find an unbiased (nonparametric) estimator of the mean, since we have no information about the distribution above $t = T_0$.

The conclusion from this procedure is that our estimator is an estimator of the conditional expectation $E(T|T \leq t_k)$, where t_k is the largest observed event time in our sample. It should kept in mind, however, that even with this formulation, our estimator is biased under certain censoring patterns.

Appendix 2

Table A7.1: Decriptive statistics, all individuals in the data set, males and females

Variable	Mean	StdDev	Mean	StdDev
	Males (n=738)		Females (n=658)	
Number of days absent 1990	27.1526	58.0451	32.5808	65.4579
Number of days absent 1991	27.3978	65.2519	34.1905	71.9869
Personal Characteristics				
MARR (married)	0.6962	0.4602	0.7530	0.4316
DIV (divorced)	0.0490	0.2161	0.0976	0.2969
AGE	39.3324	11.8957	41.2881	11.8919
DISAB (disabled)	0.0259	0.1589	0.0274	0.1635
NRCH (number of children under 16)	0.6471	0.9963	0.7088	0.9911
Economic Incentives				
c (cost of being absent 1990)	3.9274	0.9423	3.5000	0.8768
c (cost of being absent 1991)	16.0930	4.2678	14.9043	4.4700
μ (virtual income 1990 (10^{-2}))	5.6837	1.4910	6.1463	1.6401
μ (virtual income 1991 (10^{-2}))	4.1898	1.1886	4.6563	1.5790
t^c (contracted daily working hours)	38.9414	4.1088	33.9207	7.5088
Work Environment				
NOISE1 (noisy environment)	0.1798	0.3843	0.0488	0.2156
NOISE2 (noisy environment)	0.4414	0.4969	0.2256	0.4183
SMOKE (exposed to gas, dust or smoke)	0.3460	0.4760	0.1738	0.3792
SHAKE (exposed to strong shaking or vibrations)	0.1471	0.3545	0.0183	0.1341
POISON (exposed to gas, dust or smoke)	0.1594	0.3663	0.0442	0.2057
LIFT (heavy lifting)	0.3501	0.4773	0.1570	0.3641
HARD (work is physically exhausting)	0.5858	0.4929	0.6098	0.4882
SWEAT (work causes daily sweating)	0.3815	0.4861	0.2881	0.4532
EXHM (work is metally exhausting)	0.3229	0.4679	0.4680	0.4994
STRESS (work is stressfull)	0.6049	0.4892	0.6829	0.4657
REP (work is repetitive)	0.2548	0.4360	0.2607	0.4393
MOM (monotonous movements)	0.4946	0.5003	0.5640	0.4963
UBP (unpleasent body positions)	0.6063	0.4889	0.5457	0.4983
RISK1 (SIR, work accidents)	1623.5014	1038.5457	924.6951	865.3817
RISK2 (SIR, work-related diseases)	1871.7984	1013.1511	615.7774	455.8679
Health Status				
STRUMA	0.0027	0.0522	0.0198	0.1395
TBC (tuberculosis)	0.0027	0.0522	0.0015	0.0390
HEART (heart problems)	0.0177	0.1320	0.0061	0.0779
HBLOOD (high blood presure)	0.0790	0.2700	0.0686	0.2530
ULCER (gastric ulcer)	0.0259	0.1589	0.0198	0.1395
HEMORR (hemorrhoids)	0.0381	0.1917	0.0488	0.2156
PREGNANT (difficult pregnacy)	0.0000	—–	0.9345	0.2477
HERNIA	0.0095	0.0973	0.0030	0.0552
VAV (varicose veins)	0.0300	0.1706	0.0762	0.2656
MENTAL (mentally ill)	0.0041	0.0638	0.0061	0.0779
CANCER	0.0123	0.1101	0.0107	0.1028
DIABETIC	0.0191	0.1369	0.0107	0.1028
NEURO (neurological illness)	0.0054	0.0737	0.0046	0.0675
Secondary cost				
UNEMP (Monthly municipal unemployment rate 1990)	1.6603	0.6516	1.5995	0.6037
UNEMP (Mothly muncipal unemployment rate 1991)	3.1772	0.7512	3.1012	0.7155
FLEX (flexible working schedule)	0.6499	0.4773	0.6936	0.4614
CLOCK (use of timeclock)	0.4264	0.4949	0.2759	0.4473
INTIME (important to be on time)	0.7480	0.4345	0.8262	0.3792

Table A7.2: Decriptive statistics for the individuals with at least one days of both work-absence and work-presence (i.e., individuals who change state).

Variable	Mean	StdDev	Mean	StdDev
	Males (n = 635)		Females (n = 591)	
Number of days absent 1990	28.5568	53.7334	33.1322	60.9637
Number of days absent 1991	28.8407	62.5735	34.9220	68.6455
Personal Characteristics				
MARR (married)	0.6924	0.4619	0.7576	0.4289
DIV (divorced)	0.0521	0.2223	0.1000	0.3003
AGE	38.6893	11.8715	40.7949	11.7339
DISAB (disabled)	0.0205	0.1418	0.0220	0.1469
NRCH (number of children under 16)	0.6562	0.9969	0.7458	0.9973
Economic Incentives				
c (cost of being absent 1990)	3.9264	0.9478	3.4906	0.8811
c (cost of being absent 1990)	16.0120	4.1482	14.8284	4.3133
μ (virtual income 1990 (10^{-2}))	5.6415	1.4939	6.1234	1.5830
μ (virtual income 1991 (10^{-2}))	4.1816	1.1907	4.6695	1.5970
t^c (contracted daily working hours)	39.0063	4.0420	33.9169	7.4718
Work Environment				
NOISE1 (noisy environment)	0.1767	0.3817	0.0525	0.2233
NOISE2 (noisy environment)	0.4527	0.4981	0.2254	0.4182
SMOKE (exposed to gas, dust or smoke)	0.3502	0.4774	0.1780	0.3828
SHAKE (exposed to strong shaking or vibrations)	0.1483	0.3556	0.0186	0.1354
POISON (exposed to gas, dust or smoke)	0.1562	0.3633	0.0492	0.2164
LIFT (heavy lifting)	0.3438	0.4754	0.1644	0.3710
HARD (work is physically exhausting)	0.6025	0.4898	0.6220	0.4853
SWEAT (work causes daily sweating)	0.3943	0.4891	0.2949	0.4564
EXHM (work is metally exhausting)	0.3170	0.4657	0.4797	0.5000
STRESS (work is stressfull)	0.6073	0.4887	0.6797	0.4670
REP (work is repetitive)	0.2697	0.4442	0.2678	0.4432
MMO (monotonous movements)	0.5047	0.5004	0.5576	0.4971
UBP (unpleasent body positions)	0.6246	0.4846	0.5593	0.4969
RISK1 (SIR, work accidents)	1641.5615	1057.3726	930.1695	863.0728
RISK2 (SIR, work-related diseases)	1883.5962	1032.5886	620.7627	457.4012
Health Status				
STRUMA	0.0032	0.0561	0.0220	0.1469
TBC (tuberculosis)	0.0016	0.0397	0.0017	0.0412
HEART (heart problems)	0.0158	0.1247	0.0068	0.0821
HBLOD (high blood presure)	0.0757	0.2647	0.0644	0.2457
ULCER (gastric ulcer)	0.0252	0.1570	0.0186	0.1354
HEMOR (hemorrhoids)	0.0347	0.1832	0.0475	0.2128
PREGNANT (difficult pregnacy)	0.0000	——	0.9356	0.2457
HERNIA	0.0095	0.0969	0.0034	0.0582
VAV (varicose veins)	0.0331	0.1791	0.0729	0.2602
MENTAL (mentally ill)	0.0047	0.0687	0.0051	0.0712
CANCER	0.0126	0.1117	0.0119	0.1084
DIABETIC	0.0174	0.1307	0.0119	0.1084
NEURO (neurological illness)	0.0063	0.0792	0.0051	0.0712
Secondary cost				
UNEMP (Monthly municipal unemployment rate 1990)	1.6545	0.6456	1.5953	0.6096
UNEMP (Monthly municipal unemployment rate 1991)	3.1761	0.7476	3.0932	0.7196
FLEX (flexible working schedule)	0.6562	0.4754	0.6966	0.4601
CLOCK (use of timeclock)	0.4211	0.4941	0.2797	0.4492
INTIME (important to be on time)	0.7539	0.4311	0.8339	0.3725