WORKING PAPERS IN ECONOMICS

No. 0400

JAN ERIK ASKILDSEN, ESPEN BRATBERG AND ØYVIND ANTI NILSEN

SICKNESS ABSENCE OVER THE BUSINESS CYCLE

Department of Economics

UNIVERSITY OF BERGEN

Department of Economics
University of Bergen
Fosswinckels gate 6
N-5007 Bergen, Norway
Phone: +47 55 58 92 00
Telefax: +47 55 58 92 10
http://www.svf.uib.no/econ

Sickness Absence over the Business Cycle

By

Jan Erik Askildsen

Espen Bratberg

Øivind Anti Nilsen*

Department of Economics, University of Bergen, Fosswinckelsgt. 6, N-5007 Bergen, Norway.

March 2000

Abstract

Sick leaves may vary over the business cycle due to disciplining effects or changes in labour force composition. The latter hypothesis maintains that sickness may be pro-cyclical due to employment of "marginal" workers with poorer health when demand increases. Using individual records of labour force participants in Norway, we investigate the explanatory factors behind differing spells of work absence at different stages of the business cycle. We find no indication that new entrants into the labour force explain increases in absence. On the other hand, insiders increase their absences when the economic conditions improve. Thus, a disciplining effect seems to explain most of the changes in work absence behaviour.

JEL classification number: J22, J33, J38.
Keywords: Absenteeism, sickness benefits, unemployment.

* Corresponding author. Postal address: Department of Economics, University of Bergen, Fosswinckelsgt. 6, N-5007 Bergen, Norway. Phone: +47 55 58 92 00. Fax: +47 55 58 92 10. E-mail: Oivind.Nilsen@econ.uib.no

1. Introduction

Expenditures associated with sickness absences have led to serious concerns in many countries. In European welfare states, an increase in sick payments causes fiscal problems for often highly indebted governments. Simultaneously, industry suffers losses when workers are absent. This situation has triggered a debate about what are the causes of work absences, and what measures that should be initiated to combat an alleged negative trend. Some countries, like the Netherlands, Sweden, and Germany, have initiated changes in the sick payment schemes, which reduce the economic compensation to be received during sick leaves. Most economists will agree that this will reduce the incentives to stay away from work, at least to the degree that some sick leaves are to be termed illegitimate. On the other hand, it is interesting to observe that for most countries, irrespective of generosity of sick payment systems, absenteeism seems to follow the same pattern over the business cycles and between sectors of the economy.

The overall work absence over the business cycle will be affected by the behaviour of each worker, and their individual background characteristics. As labour market conditions vary, the individual costs and benefits from staying away from work are affected. In particular the indirect individual costs of absence may be considered as being lower when the economy is booming than in situations with labour market slackness. Then, for any degree of health conditions, the propensity to absent is higher when unemployment is low. This may be true for valid sickness absences as well as for more invalid reasons for absenting. Such patterns are confirmed in several studies for different countries, see e.g. Allen (1981a), Kenyon and Dawkins (1989), Drago and Wooden (1992), Johansson and Palme (1996) and Dyrstad (1997). Particularly, it is frequently argued that an explanation for the pro-cyclicality of sick leaves can be found with reference to a more lenient worker morale when alternative jobs are

abundant. The so called efficiency wage theory, Shapiro and Stiglitz (1984), is commonly used as a background for explaining such fluctuations in short-term sick leaves, Barmby, Sessions and Treble (1994). According to this theory, shirking is more frequent for low wages and in booming periods. However, also long-term sick leaves seem to follow the same pattern. The shirking explanation does not seem equally justifiable for longer spells of absences which in most countries require an approved medical diagnosis. Nevertheless, the individual costs in terms of loss of future income and career due to a prolonged sick leave may well vary counter-cyclically. A diagnosis that would be considered as sufficient reason for staying away from job for a longer period, may in an adverse labour market state make the worker return earlier to work since the worker may fear that absences may endanger future worker prospects. There is an alternative to this worker *behaviour* related explanation. Variations in absenteeism over the business cycles may be due to labour market demand effects, Barmby and Treble (1991). One important aspect of this is related to the changing composition of the labour force. As pointed out by Allen (1981b), employers may find it profitable to screen workers and first offer contracts that are preferable for the healthiest workers. Thus, the most able and healthy workers are basically kept continuously employed. When economic conditions worsen, the less able or sick prone workers are the first to loose their jobs. Consequently, these so-called 'marginal workers' may serve as a reserve, to be called on when workers become scarce. But when these workers are employed, sickness related absences will increase. This latter effect, often termed a *composition* effect or marginalisation of the labour force, will therefore lead to a higher incidence of work absence in booming periods. Many argue that this is a main reason for variation in sick leaves, and in particular for long-term sickness which is the focus of this study.

Empirical analyses of these competing hypotheses are important, because they may give valuable information concerning the design of sick leave schemes. It is also important for

considering whether firms should invest in sickness prevention measures for less able workers, or use more incentives to improve on work effort. The relevant measures will affect the overall expenditures associated with work absenteeism. In this paper, we will use sickness absence data from Norway to distinguish between these two competing hypotheses, and investigate whether *behaviour* or labour force *composition* explains long-term sick leaves[1]. Different from most other countries, Norway has a 100% replacement ratio during absence caused by sickness. There is an income ceiling, at approximately NOK 250 000, above which no additional benefit is paid[2]. However, most larger firms and the public sector will compensate the workers earning above the income ceiling, so that the 100% replacement ratio is relevant for the bulk of the work force. Thus, the Norwegian system is such that most workers will receive full wage for a sickness spell lasting up to one year. This high replacement ratio may give incentives to stay away from work even though it may not be strictly necessary, and to stay away for longer periods.

Expenditures associated with absenteeism are significant in Norway, and have therefore led to serious concerns about the Norwegian sick leave scheme and its incentives. According to the Norwegian Employers' Association, Norway has more than 50% higher sickness absence than the other Scandinavian countries. Sickness benefits were estimated to cost NOK 28 billion in 1997. This amounts to 2.6% of GDP. Sickness benefit expenditures in Norway are shared between the employer and the Social Insurance system (National Insurance Administration – NIA). The employers pay sickness benefits for the first 14 days, and NIA covers sickness benefits exceeding two weeks, and up to one year.[3] Sickness benefits

[1] It is often alleged that job protection rights in Scandinavian countries are such that unemployment disciplining effects are irrelevant. However, firms will often find ways to avoid strict seniority rules etc. when downsizing. Furthermore, according to OECD (1994) Norway is ranked in the middle among OECD countries in terms of labour market strictness, see also Salvanes (1997).
[2] NOK 8˜ $1
[3] As of 1998, the employer's period is increased to 16 days.

paid from the Social Insurance in 1997 amounted to NOK 15 billion. The remaining part, NOK 13 billion, was the employers' share.

In Norway, both short- and long-run sick leaves, together with sickness leave payments, have increased considerably during the latter part of the nineties. This increase has followed an increase in economic growth and reduction in unemployment. It is notable that after surging in 1993, unemployment has decreased, while sickness absences did not start to increase before 1995. Such observations are consistent with a theory stipulating that costs of staying away from work are smaller when unemployment is low. Higher absences in the public sector with better job protection draws in the same direction. However, this pattern is also consistent with the alternative hypothesis that sicker and more marginal workers are employed when economic conditions improve, and in more protected sectors. A Norwegian study, NIA (1998), gives some evidence that at least marginal male workers have a higher tendency for sick leaves than more established insider workers. There are, however, no studies that have used a broad panel data set with detailed data on individuals to test what explains this work related sickness. Comprehensive background information about several individuals for several years provides a good opportunity to confront the competing hypotheses that changes in sick leaves are either explained by variation in behaviour, or by the composition of the work force. Furthermore, we ask what drives the alleged business cycle dependence of sickness absenteeism.

In our study, for data availability reasons, only sick leaves covered by the NIA are considered. These are comprised of all sickness absences exceeding two weeks, i.e. more long-term sick leaves. We compare two years at different stages in the business cycle: 1992 when unemployment was high, and 1995, a year when unemployment was clearly on its way down. The composition effects are isolated by comparing sickness probabilities for workers with a low participation in the labour market two years prior the year of investigation, to the

rest of the workers. Workers who were under education, or too young to participate in the labour force, are excluded from both these latter sub-samples. We utilise the KIRUT database, which contains detailed individual information for a random 10% sample of the Norwegian population aged 16-67. From KIRUT we extracted more than 80 000 workers in 1992 and in 1995. The data set contains job-related information like income and absenteeism above two weeks. There is also a lot of background information about the participants, e.g. gender, age and former employment history, in addition to information about the sector for which the individuals are working.

The paper is organised as follows. In the next section, we use a standard model to derive some theoretical results of relevance for this study. In Section 3, we give an institutional overview, with emphasis on the sickness benefit system and development of unemployment and absenteeism during the nineties. Then in Section 4, we present the data and how they are prepared for this particular purpose. The empirical specifications are discussed in Section 5. In Section 6, we report the results, and some concluding remarks are offered in the final section.

2. Theoretical background

It is common in economics to consider the decision of being absent as resulting from utility maximising. A simple model based on Allen (1981a)[4] illustrates the point, and can be used to illustrate how the two competing hypotheses explain cyclicality in sickness absenteeism. A worker i maximises a utility function with consumption, C^i, and leisure, L^i, as control variables. Utility is in addition affected by the worker's health state, $?^i$, and other individual characteristics, Z^i. Health state is assumed continuously variable within an interval, and such

[4] See also Brown and Sessions (1996) and Kenyon and Dawkins (1989) for related applications of the model. An interpretation is that a worker makes a daily decision whether to work or stay away from their job.

that a higher value indicates better health. The individual characteristics include relevant information like gender, age, nationality, family situation, children, level of education, and work experience.

We assume that all workers have the same utility function but that they differ in health and background characteristics. Their utility function, which is concave and twice differentiable, and increasing in consumption and leisure, is defined as

(1) $V^i = V(C^i, L^i; ?^i, Z^i)$

The consumption possibilities are restricted by wage income, with W^i as the wage rate, and other monetary sources, R^i, in addition to a possible penalty resulting from work absences, $P(A^i, u, W^i)$, where A^i denotes work absence and u is the labour market state as measured by the relevant unemployment rate. The penalty function is increasing in all its arguments, and convex in A^i. Long and/or frequent absences will be costly for the worker, and it is more costly to absent when labour market conditions are adverse and for individuals with high wages. The upward shifts in the penalty function for increased wages and unemployment can be considered as capturing the idea from the efficiency wage theory, Shapiro and Stiglitz (1984), that high wage and unemployment have disciplining effects. It is based on asymmetric information between firm and employee in the absence decision. We argue that it may be relevant also for long-term absences, and even with a 100% replacement ratio. The main reason is that it is always costly for an employer when workers absent. Examples are that substitute workers of equal skills may be hard to find, or that tasks have to be postponed. The penalty function can also be given a more general interpretation. Individual costs of duration and frequency of absences can be explained by workers fearing that their career can be adversely affected, and that this is worse for high-income earners and more acute when

unemployment is high. Lastly, observed pattern of absences may be used by firms if they are to decide who is to leave the firm in case of downscaling.[5] Each worker is supposed to work an individually contracted number of hours, $\overline{H}^i$. If absent due to sickness, the worker receives a compensation ratio of k. Then the budget restriction can be written as

$$(2) \qquad C^i = R^i + W^i\left[\overline{H}^i - (1-k)A^i\right] - P\left(A^i, u, W^i\right)$$

Lastly there is the time restriction with total time available as T:

$$(3) \qquad L^i = T - \overline{H}^i + A^i$$

With subscripts denoting partial differentials, the standard first order to this problem is written as

$$(4) \qquad \frac{V_L(.)}{V_c(.)} = W(1-k) + P_A(A^i, u, W^i)$$

The marginal rate of substitution between leisure and consumption depends on the individual's health state. The absence decision is furthermore influenced by the compensation the person receives when sick. A 100% replacement ratio, k = 1, gives a strong incentive for workers to stay away when sick, and also to be absent for invalid reasons ('shirking'). This moral hazard problem is well know within social insurance situations, Whinston (1983), since it is generally hard to observe whether a person claiming that an insurance situation has arisen, really fulfils the stated conditions. The moral hazard problem may be smaller when

[5] Employers may find ways to avoid potential seniority rules.

dealing with long-term sickness than short-term due to the requirement of a doctor's certification. Still, many sickness claims are difficult to monitor also for physicians. Simultaneously, when economic conditions improve, it is reasonable that more marginal workers are attracted to the labour force, and firms are willing to employ these workers when labour is scarce. Firms may e.g. recruit people based on former sickness records which are observable to the firms' experience rating. Thus, the firms will try primarily to hire workers with few and short spells of absence. However, when activity in the economy increases, together with firms' demand for more workers, it may be hard to find the workers with the best records. The firms will then have to recruit workers with a poorer experience in absenteeism. These may be workers that are more prone to shirking but it is as likely that it is workers with poorer health and thereby a higher probability of becoming sick and being eligible for sickness benefits.

Some assumptions are useful when discussing the comparative statics of (4). Leisure is a normal good, and we assume, as indicated above, that the marginal utility of leisure is decreasing in the health state. If we allow for interpersonal comparisons of individuals with the same functional form of the utility function, this implies that a sick worker values increased leisure time relatively higher than a more healthy worker. Marginal utility of consumption is non-decreasing in health. Then we readily compute the comparative static results

(5a) $$\frac{dA}{dq^i} < 0, \quad \frac{dA}{du} < 0$$

Employment of less healthy workers with a lower $?^i$ will tend to increase absences, and the same happens for each individual worker if unemployment, u, decreases. Thus, from (5a) we

must conclude that only an empirical analysis can reveal the relative importance of the two explanations for cyclical absence variations.

We may similarly compute the effects of changes in contracted work hours, $\overline{H}^i$, wages, W^i, and income from other sources, like spouse income and wealth, R^i. The comparative statics are given by[6]

$$\text{(5b)} \quad \begin{aligned} &\frac{dA}{dW^i} > 0 \text{ if } k = 1 \text{ and } \frac{\partial P(.)}{\partial W^i} = 0 \\ &\frac{dA}{dW^i} = ? \text{ if } k < 1 \text{ and/or } \frac{\partial P(.)}{\partial W^i} > 0 \\ &\frac{dA}{dR^i} > 0 \\ &\frac{dA}{d\overline{H}^i} > 0 \end{aligned}$$

In the Norwegian system, $k = 1$ for the majority of income earners. Thus, changes in wages would imply a pure income effect, which is positive, save for the substitution effect through the penalty function. If this is existent, it becomes costlier to stay away from work when income increases for an individual, and high income earners will more reluctantly absent due to this particular effect. In general the effect of changes in income should be considered ambiguous.

The signs of changes in the individual characteristics, Z^i, are, as is reasonable, ambiguous. The relevant expression is

$$\text{(5c)} \quad \frac{dA}{dZ^i} = -\frac{1}{D}\left\{V_{LZ}W - V_{CZ}\left[W(1-k) + P_A^i\left(A^i, u\right)\right]\right\} = ?$$

[6] Some unproblematic assumptions are necessary to ambiguously sign the effect of changes in contracted work hours.

In (5c), D = $\frac{\partial^2 V}{\partial A^2} < 0$ from second order conditions. The sign of $\frac{dA}{dZ^i}$ depends on how individual characteristics like gender, education, number of children etc. affect the marginal utility of leisure and consumption. It is reasonable to assume that number of children will affect marginal utility of leisure positively, which contributes to making absences positively related to children. For education it is common to assume the opposite. It is harder to say how marginal consumption is affected but we will expect the former effect to be of most importance when evaluating the comparative statics in (5c). The same will hold true for other variables of relevance. We return to this discussion in Section 4.2.

3. Institutional Background

The sickness benefit system in Norway is organised under the National Insurance Administration (NIA), which also encompasses unemployment insurance, disability insurance, and old age pensions. All employees who have been with the same employer for at least two weeks, are covered by sickness insurance. Once this requirement is filled, coverage is 100 per cent from the first day. There is an upper limit but for typical employees there is little variation in replacement ratios. The upper limit is 6*G*, where *G* is the basic unit used in the pension system, NOK 35 500 in 1992 and NOK 39 230 in 1995. The after tax replacement ratio is 100% for average wages in manufacturing, and some 90% for wages 25% above that average (NOSOSCO (1992)). A medical certificate is necessary for absences of more than three days. For sickness spells that last for more than eight weeks the physician is obliged to provide a more detailed certificate to the Social Insurance authorities, stating diagnosis and a prognosis assessment.

Sickness benefits are paid by the employer for the first two weeks, and then by the NIA for a maximum of 50 weeks. If unable to return to work after one year, the options are to apply

for (permanent) disability benefits or for (temporary) rehabilitation benefits. These benefits are comparable to old age pensions and considerably lower than sickness benefits. Payment of premia for sickness insurance is part of the Social Insurance system, and based on a pay-as-you-go system. Workers pay a given share of income as a 'sickness insurance' tax, and employers contribute through a payroll tax on the total wage bill. There is no experience rating, and firms are not allowed to use sickness as reasons for laying off employees.

NIA expenses are sizeable. In 1995, the last year covered in this analysis, NIA social insurance payments totalled NOK 126.2 billion, about 13% of GDP. Sickness benefits contributed 9% of NIA outlays. From Figure 1, we see that sickness absences have moved countercyclically during the eighties and the nineties. The average number of sickness days per employee per year are calculated by counting the overall number of sickness days exceeding the first 14 days, beyond the 'employer's period', and divide this number by the number of employees. Note that this figure does not give any information about the development of shorter sickness spells, only those exceeding 14 days. Furthermore, state employees are excluded.

(Figure 1 in about here)

We note that for the period 1992-1995, the unemployment rate started to decrease one year before the number of sickness days started to rise. This may be due to the way data are registered by the NIA. However, the significant lags between the unemployment rates and the sickness absence is also consistent with the hypothesis that increased absence is due to composition effects. In addition, it is possible that the pattern comes from more stress and increased speed at the work place. If the shirking hypothesis dominates the composition effects, we would expect unemployment and absenteeism to mirror each other more

immediately. On the other hand, it may be that workers are more careful during the early stage of an upturn.

4. Data

4.1 Data Sources

The analysis draws on data from the KIRUT database.[7] The data contain detailed individual information on socio-economic background, labour market participation, and social insurance payments for a random 10% sample of the Norwegian population aged 16-67 (the total sample exceeds 300 000 individuals).

Our sample includes a 50% selection of the individuals that occupy a job from January 1, 1992 until December 31, 1992, and a 50% selection of the individuals that occupy a job from January 1, 1995 until December 31, 1995.[8] The drawings are done by selecting the 50% lowest identification-code of the individuals the relevant years. Note, however, that the id-code is unrelated to any characteristics of the individuals. All state employees are excluded from our sample since there is no information about sickness periods or payments for this category of workers. After excluding individuals with missing variables and self-employed, we end up with final samples of 82 349 individuals in 1992, and 88 354 individuals in 1995. We construct several sub-samples. The first one covers those individuals that are included in both the 1992 and 1995 yearly samples. We call this the *common sample*. This sample consists of 69 131 individuals. Next, we distinguish between what we term *marginal workers* and *non-marginal workers*. Marginal workers are individuals who had a loose relation to the labour force two years ahead of the year of investigation, 1992 and 1995 respectively. We

[7] KIRUT is a Norwegian acronym that roughly translates into "Clients into, through and out of the Social Insurance System".

[8] This information is based on the registrations in the employers' register. Employers are obliged to report to this register all new employees who are expected to stay in the job for at least three days.

require that they did not work more than 500 hours. However, we are not interested in those individuals who did not work due to education or age. Therefore we exclude individuals with less than two years potential experience (age – years of education – 7). We also exclude workers with a seniority within a firm of more than two years, and those who are younger than 18 years old at time *t*, the year of investigation. The sub-sample of marginal workers covers 4 028 individuals in 1992, and 4 783 in 1995. This sample is compared to a sample of "non-marginal", the full sample except those with less than two years potential experience.

4.2 Variables

The number of days with sickness benefits paid by NIA two years back, FORM_ABS, is used as a proxy for the health of the individuals. All the other explanatory variables are measured the relevant year with the exception of income which is measured at year *t-1*. The relevant family variables are: marital status, (UNMARR), divorced or widowed (PRE_MARR), gender (WOMAN), and number of children less than 11 years old (CHILDREN). Further background variables included in Z^i above are origin of birth (NONSCAND = 1 if non-Scandinavian), years of education (EDUCATION), age (AGE) and age squared (AGE_SQR), where the latter is included to take potential non-linearities into consideration. EDUCATION reflects the worker's education and/or skills. EDUCATION will thereby also to some degree proxy for the individual wage rates, since our income variable (see below) is total wage income. As a pure education variable, we expect its sign to be negative. However, if representing wage rates, it should be remembered from Section 2 that wages affect absences ambiguously. Thus, there is potential for some ambiguity, although we find a negative sign most likely. Also experience and seniority will reflect wage rates. Labour market experience (EXPERIENCE) is measured as the number of years with registered pension points (income above 1 G). Seniority (SENIORITY) is the number of years a worker has been within a firm since the beginning of the running contract period.

Again, if disciplining effects are not present, their signs will be positive, since we hold effects from $k < 1$ to be of minor importance. Otherwise they should be positively signed. PARTTIME indicates whether an individual works less than 20 hours a week. The dummy is equal to one for part time work. Since longer contracted work hours according to the comparative statics affect absences positively, its sign should be negative. Income previous year (INCOME) reflects wage rates as well as hours of work, and also the individual's income potential the current year if sick. Sign is assumed to be the same as for EXPERINCE and SENIORITY. To account for non-linearities, we include squared income (INC_SQR). Household wealth (WEALTH) and spouse's income (SPOU_INC) should increase the propensity to absent through income effects. Spouse's income may also be given an alternative explanation in our model. So-called 'assortative mating' indicates that those with a preference for working (low marginal valuation of leisure) will find a spouse of similar property. If this were the case, spouse's income would affect absenteeism negatively. All income variables (spouse's income, own income, and wealth) are measured in NOK 10 000.

Finally, we control for the current labour market situation of each municipality, location, and industry sectors. We use unemployment (UNEMP) in the municipality where the individual lives. It measures registered unemployment as a percentage of the registered labour force (employed plus unemployed). It is hard to say anything definite about its sign. According to unemployment disciplining effect it should be negative. However, it will also capture other labour market differences between municipalities, which are not captured by the broader seven regional dummies representing region of residence. Lastly, we include six industry sector dummies. For ease of exposition, we do not report the results for the regional and sector dummies in the regressions.

(Table 1 about here)

Summary statistics are presented in Table 1. We observe that for the total sample the sickness probability is lower in 1995 than in 1992. This may seem surprising, since unemployment is lower in 1995. It is first and foremost for the non-marginal workers and within the private sector of the economy that absenteeism is reduced[9]. For the common sample it has increased. The other surprising result is that marginal workers for both years have a lower absence probability than the non-marginal workers. We note that the sample of marginal workers is younger than the non-marginal sample, and that they earn considerably less. We also note that the marginal workers are slightly older in 1995 than in 1992. Furthermore, there are more non-Scandinavians in this group but this share decreases from 1992 to 1995.

(Table 2 about here)

Table 2 shows the distributions of absence days (calendar days) for those who have any. We remind the reader that our records only include absences lasting more than two weeks. When that threshold is passed, roughly a quarter have less than 14 additional days of absence, and about 50 % have less than 42 days.

5. Empirical specification

The number of absence days is an integer variable, suitable for count data modelling. For each individual we have information on whether s/he had any sickness spells exceeding two weeks (paid for by the NIA) in a given year, and the length of these spells, if any. This introduces a censoring problem in our dependent variable, as a zero count may mean anything between 0

[9] Note from Figure 1 that absences are higher in 1995 than in 1994.

and 14 days. We address the problem by using a two-part, or "hurdle" count model (Mullahy 1986). The first part consists of a binary response model, and the second of a model for the number of absence days once the hurdle has been passed. The hurdle model is a generalisation of ordinary censoring models, where it is assumed that the same process governs the probability of passing the threshold and the outcome conditional on having passed it.

Consider first the Poisson regression model. Let Y denote a positive integer random variable which follows the Poisson distribution, i.e.

$$\text{(6)} \qquad \Pr(Y = y) = f(y) = \frac{\exp(-l)\, l^{y}}{y!}.$$

Introducing covariates, $\mathbf{z}_i$, and coefficients, g, and specifying $\ln l_i = g'\mathbf{z}_i$, the Poisson regression model results. A restrictive property of the Poisson model is equidispersion, i.e. the mean equals the variance: $E(Y) = \mathrm{var}(Y) = l$. The negative binomial (negbin) model is a generalisation of the Poisson model, which allows the variance to exceed the mean. The probability density function is

$$\text{(7)} \qquad f(y) = \frac{\Gamma(y + \frac{1}{a})}{\Gamma(y+1)\Gamma(\frac{1}{a})} \left(\frac{\frac{1}{a}}{\frac{1}{a} + m} \right)^{\frac{1}{a}} \left(\frac{m}{\frac{1}{a} + m} \right)^{y}, \; a > 0,$$

where G denotes the gamma function. For the limiting case $a = 0$, this reduces to the Poisson density. Equation (7) may be derived from the Poisson by letting $l = mv$, where v is a random variable. Assuming that v is gamma distributed with mean 1 and variance a, it may be integrated out of the distribution of y and v, and (7) results. This model has $E(Y) = m$ and

$\text{var}(Y) = m(1+am)$. In the regression context, $\ln l_i = \ln m_i + \ln v_i = g'\mathbf{z}_i + e_i$, where $\exp(e_i) = v_i$.[10]

Hurdle models allow for excess numbers of zeros (or values at some other truncation point). It is assumed that the probability of having counts greater than zero results from one process, and positive counts from another process. Consider a hurdle model where both parts are negative binomial but with different parameters. Let $f_1(y)$ be the negbin density with parameters (m_1, a_1) that governs the probability of having a zero count, Pr(y = 0). Using (7), we have that

$$\text{Pr}(Y=0) = f_1(0) = \left(\frac{\frac{1}{a_1}}{\frac{1}{a_1}+m_1}\right)^{\frac{1}{a_1}} = (1+a_1 m_1)^{-\frac{1}{a_1}}$$

(8)

$$\text{Pr}(Y \geq 1) = 1 - f_1(0) = 1-(1+a_1 m_1)^{-\frac{1}{a_1}}.$$

Furthermore, let $f_2(y)$ be the density function of the second process, which is also a negative binomial model with parameters (m_2, a_2). We obtain Pr(Y = $y|Y \geq 1$) by conditioning on $1 - f_2(0) = 1-(1+a_2 m_2)^{-\frac{1}{a_2}}$, yielding the truncated negbin model:

$$f(y \mid y \geq 1) = \frac{\Gamma(y+\frac{1}{a})}{\Gamma(y+1)\Gamma(\frac{1}{a})}\left(\frac{m}{\frac{1}{a}+m}\right)^{y}\left((1+am)^{\frac{1}{a}}-1\right). \tag{9}$$

The hurdle model may be estimated by maximum likelihood. Because (m_1, a_1) and (m_2, a_2) are independent by assumption, the ML estimates may be obtained by estimating each part of the model separately. If we impose the restriction $a_1 = 1$, equation (8) reduces to

[10] See Cameron and Trivedi (1998) for a comprehensive review of count data models.

the logit model. That is the approach taken here: First we estimate Pr(Y_i – 14 > 0) with a logit model, and then we use the zero-truncated negative binomial model to estimate Pr(Y_i –14 = y_i – 14 | y_i – 14 > 0).

Note that both parts of the hurdle model may be given a behavioural interpretation: in the first part a latent variable measures the utility of having at least one long-term absence. What is observed is the binary outcome. In the second part another latent variable measures the utility of an additional absence day. Each time a threshold is passed, a new day is added to the number of absence days.

To discriminate between the 'behaviour' and 'composition' hypotheses, we perform a decomposition analysis. As a shorthand, let us formulate the expected outcome of each part of the process above as

(10) $$E(y) = F\left(X, b\right),$$

where E(y) is either the probability of absence (logit) or the expected number of absence days (negative binomial), and X and b denote vectors of variables and coefficients. The $b-s$ measure how individuals behave or respond when their characteristics vary, while the X vector represents the individual characteristics. Consider two individuals indexed 0 and 1, who differ in characteristics as well as response. The difference in outcome is

(11) $$\Delta = E(y^0) - E(y^1) = F(X^0, b^0) - F(X^1, b^1)$$

For a *linear* model, i.e. $E(y) = bX$, (11) may be decomposed as**.**

(12) $\Delta = b^0 X^0 - b^1 X^1 = b^1 (X^0 - X^1) + X^0 (b^0 - b^1)$

This decomposition, well known from the wage discrimination literature, Blinder (1973) and Oaxaca (1973), has an interesting interpretation: the first part on the r.h.s. expression is the part of the difference which is caused by difference in characteristics, whereas the second part is due to difference in behaviour, or response to the characteristics. By estimating the model separately for "0-type" and "1-type" individuals, (12) may be evaluated at the sample means.

Equation (12) does not hold for the non-linear models used in this paper. However, we shall perform decompositions as follows,

(13) $\Delta = F(X^0, b^0) - F(X^1, b^1) = \{F(X^0, b^1) - F(X^1, b^1)\} + \{F(X^0, b^0) - F(X^0, b^1)\}.$

We interpret these expressions as outlined above. Equation (13) is evaluated by computing the necessary components for each individual and then averaging.

We proceed as follows. At the outset we estimate both stages of the hurdle model, logit and negative binomial, for the *full 1992* and *1995 samples*, and compare the estimated parameter vectors. The hypothesis that the coefficients are equal, i.e., that the behaviour of the individuals in the two samples is the same, may be tested by a likelihood ratio test. If the null hypothesis is rejected, we can use the decomposition in equation (13) to appreciate how much of the differences in outcome may be attributed to behaviour and characteristics, respectively. We then proceed with the same exercise for the *common sample*, those individuals who are present both in the 1992-sample and the 1995-sample. In doing so, we investigate whether those with a stable labour force attachment changed behaviour during the economic upturn. Finally, we compare the 'marginal' to the 'non-marginal' workers, as defined in Section 4.1. This comparison is performed for the 1992 and the 1995-samples.

6. Results

The results from the hurdle regressions on the different samples are reported in Tables 3-5. The upper panels include the results from the logit model giving the probability of having a sickness spell, whereas the lower panel results are from the negative binomial model explaining duration of sickness absences. We have also run the regressions by gender and report those results in the Appendix. In the discussion we focus on the regressions in Table 3-5. In Table 6 we show decompositions as explained in the previous section, equation (13), for both genders together and for males and females separately.

Table 3 contains estimates of absenteeism for the full samples in 1992 and 1995. As was seen from summary statistics and in accordance with aggregate numbers of absences, the predicted absence probability decreases from 1992 to 1995. Since unemployment is lower in 1995 than in 1992 the result is contrary to what is expected from theory. The result may be due to a delay in reaction to changes in economic conditions, or that long-term sickness absences are not explained by behaviour or composition effects. We question this latter explanation. Looking at the coefficients for the logit part, we see that there are small differences between the two years. The coefficient for local unemployment, UNEMP, changes sign but it is not significant. As argued above, in this model the variable's main role is to take care of labour market differences among municipalities, and it is less important for explaining unemployment disciplining effects as in e.g. Johansson and Palme (1996). Former sickness, FORM_ABS, explains absences as expected, with coefficients of 0.0077 in 1992 and 0.0074 in 1995. This means that if an individual in the 1995 sample had an increase of 10 absence days in period *t*-2, evaluated at the sample means, this would increase the probability of having a sick leave period from 0.119 to 0.123. Thus, the effect of former sickness is only marginal. The income coefficient, INCOME, is positive and significant for the absence

probability, with an income elasticity based on the 1995-results at 1.25. Interestingly, the effect on the number of absence days once absent (Part 2 of the model) is negative. This implies that if an unemployment disciplining effect plays a role for long-term sickness absences through income, cf. (5b), then this effect is more apparent for the length of an absence stay than for the probability of having an absence. But note that a positive sign does not preclude behaviour or disciplining effects from being present. The income effect may dominate the behaviour effect and the substitution effect when $k<1$, for some high-income earners. Spouse's income and wealth have opposite signs to own income in Part 1 of the process but the coefficients are smaller. The variables EDUCATION, SENIORITY and PARTTIME have negative signs, whereas EXPERIENCE is positive and significant in 1992 for Part 1 of the model. As noted above, there are several possible interpretations for these variables but based on theory at least the signs for EDUCATION and EXPERIENCE reported here are not unreasonable. Individuals with an additional year of education have a probability of a long-term absence spell at 10.3% lower than individuals with an education equal to the average. Since signs of variables proxying for a wage rate are negative, the results indicate that disciplining effects are present. The negative sign of the part-time variable may reflect that it does not properly measure contracted hours. The coefficient for number of children, CHILDREN, is positive as was expected. More children induce more long-term absences. However, when considering males and females separately, we found that the coefficient was actually negative and significant for men in 1992 (Part 1), and insignificant in 1995. With the exception of income, the results for Part 2 are similar when considering sign and significance of the main variables.[11]

[11] For sake of brevity we have not reported the industry coefficients. Absences are actually lower in local public administration and education than in manufacturing industry. This is surprising based on an assumed lack of disciplining effects due to job protection in the public sector. However, individuals working in the health sector have a higher absence propensity than in manufacturing industry.

(Table 3 in about here)

In Table 6 we see that when we decompose the changes from 1992 to 1995 in Part 1 into a composition (characteristics) effect and a behaviour effect, approximately 60% of the change is due to changes in characteristics.

To investigate the decomposition further, we turn to the *common sample*, Table 4. The results are very similar to what we found using the total samples for each year, so the coefficient estimates need no further comments. The likelihood ratio tests, reported in Table A1, show that the coefficients for the 1995 sample are significantly different from the coefficients in 1992 (both parts). Of most interest are the overall sickness probabilities, which are 0.111 in 1992 and 0.122 in 1995. For the sub-sample of workers who participate both in 1992 and 1995, the sickness probability has increased! This is as expected when economic conditions improve. Are the changes due to different behaviour or differences in the individuals' background characteristics? Behaviour, or disciplining effects, matters. From Table 6 we see that more than 70% of the change in absence probability (0.0082 out of 0.0109) is due to changes in the workers' behaviour. This is even more pronounced for Part 2. We find that once sick, the mean number of sickness days increase by 12.1, with 11.2 days (91%) being due to changes in behaviour. Thus, a significant share of the increase in sickness absence from 1992 to 1995 for established workers may be explained by a behaviour effect. The insiders tend to be more sick and absent from work when the economy is booming compared to a situation with high unemployment. One interpretation of this is that there are unemployment disciplining effects present for the insiders in the labour market.

(Table 4 in about here)

We now turn to the so-called marginal workers, and compare them to the non-marginal workers. The results are reported in Table 5. The marginal workers are those who had a loose connection to the labour market two years back. They were either working very few hours, were unemployed, or were outside of the work force but not under education. Contrary to the composition effects argument, we find that this group has a *lower* probability of being absent than the reference group. The coefficients indicate somewhat different behaviour between marginal and non-marginal workers (Part 1). When decomposing, Table 6, we find that different characteristics explain some 110% (1992) and 165% (1995) of the absence probability difference. In other words, given the characteristics of the non-marginal workers, the marginal workers would have had *more* absences. For the number of sickness days, the picture is less clear. In 1992, marginal workers have less absence days than the non-marginal but there is no difference in 1995. The likelihood ratio tests do not reject the null hypotheses of equal behaviour between marginal and non-marginal workers. Comparing marginal workers in 1992 and 1995 in Table 5 (Part 2), we find an increase in absence days from 1992 to 1995 for marginal workers, from 64.2 days to 68.5 days, contrary to what is found for non-marginal workers. The decompositions in Table 6 indicates that the relative change between the two groups over time can be ascribed to the marginal workers changing their behaviour in the direction of longer absences, given their characteristics. A disciplining effect seems to be working also for the marginal workers: they are more prone to long absences in booming periods.

(Table 5 in about here)

(Table 6 in about here)

As mentioned, we have also divided the sample into sub-samples of males and females. These results are found in the Appendix. The overall conclusion from this exercise is that the differences in the sickness absence behaviour are only marginally different between the two genders. It should be noted that male more than female marginal workers account for their change in behaviour from 1992 to 1995, see Tables 6, A6 and A7.

7. Concluding remarks.

We have investigated factors explaining sickness absences in Norway. The study is limited to sickness spells lasting more than two weeks, i.e. spells that are paid for by the Social Insurance. The data, drawn from the KIRUT database, include extensive individual background information, so we can control for individual characteristics when analysing sickness absence. Background variables explain sickness behaviour as indicated by a theoretical model. We note in particular that former sickness increases probability and duration of absences in a given period. It is noteworthy, however, that income increases the probability of absences, while duration is negatively affected. The latter is in accordance with the prediction from an efficiency wage model.

A prime objective was to investigate whether sickness absences are explained primarily by a behaviour effect notably among insiders, or by composition effects within the labour force. Behaviour or disciplinary effects of unemployment are analysed by comparing sickness absences in 1992 and 1995. Unemployment was high in the first of these years, lower and on its way down during the latter. Sickness absence was slightly lower in 1995 than in 1992, but increasing. The behaviour of the workers who participated in both 1992 and 1995 changed, and such that they for given characteristics have more absences in 1995. Interestingly, the same holds true for marginal workers. The so-called marginal workers

change behaviour but it is not their characteristics that account for more absences. This gives support to the hypothesis that sickness absence increases when economic conditions improve, both for insiders and the marginal workers, and thus disciplining effects seem to explain the work absence behaviour. The marginalisation hypothesis is further questioned since former sickness plays only a minor role in explaining current absence. The composition effects are analysed by comparing absenteeism between workers with a longer lasting attachment to the labour force, with marginal workers. We find that the marginal workers actually have the lowest absence probability.

In aggregate, our analysis gives no support to the claim that cyclical variations in sickness absence can be explained by composition effects, or by 'marginalised' workers entering the labour force during an economic upturn. To the contrary, we find that the "stable" workers – the insiders who were in the labour force in both periods under study, are the ones who change behaviour and increase absences.

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*, 207-18.

Barmby, T.A., and J.G. Treble, 1991, "Absenteeism in a Medium Sized Manufacturing Plant", *Applied Economics 23*, 161-6.

Barmby T, J.G. Sessions, and J.G. Treble, 1994, "Absenteeism, Efficiency Wages and Shirking", *Scandinavian Journal of Economics 96*, 561-566.

Blinder, A.S. 1973, "Wage Discrimination: Reduced Form and Structural Estimates," *Journal of Human Resources 18*, 436-455.

Brown S, and J.G. Sessions (1996), "The Economics of Absence: Theory and Evidence", *Journal of Economic Surveys 10*, 23-53.

Cameron, C.A., and P.K. Trivedi, 1998, *Regression Analysis of Count Data*, Cambridge, Cambridge University Press.

Chaudhury, M., and I. Ng, 1992, "Absenteeism Predictors: Leas Squares, Rank Regression, and Model Selection Results", *Canadian Journal of Economics 3*, 615-34.

Drago, R. And M Wooden, 1992, "The Determinants of Labour Absence: Economic Factors and Workgroup Norms Across Countries", *Industrial and Labor Relations Review 45*, 764-78

Dyrstad, J.M., 1997, "Absence: Discipline versus Composition Effects", mimeo, Department of Economics, Norwegian University of Science and Technology, Trondheim, Norway.

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

Kenyon, P. and P. Dawkins, 1989, "A Time Series Analysis of Labour Absence in Australia", *Review of Economics and Statistics 71*, 232-39.

Mullahy, J., 1986, "Specification and Testing of Some Modified Count Data Models," *Journal of Econometrics 33*, 341-365.

NOSOSCO, 1992, "Statistical Reports of the Nordic Countries", no. 58, Oslo, Norway.

OECD, 1994, "The OECD Jobs Study: Evidence and Explanations. Part II", Paris.

Oaxaca, R., 1973, “Male Female Wage Differentials in Urban Labor Markets,” *International Economic Review 14*, 693-709.

Riphahn, R.T., and A. Thalmaier, 1999, “Absenteeism and Employment Probation”, mimeo, The Institute for the Study of Labor (IZA), Bonn.

RTV, 1998, “Basisrapport 1997: Sammendrag (Basis Report 1998: Summary)”, RTV (‘Rikstrygdeverket’ – National Insurance Administration) Report 04/98.

Salvanes, K.G., 1997, “Market Rigidities and Labour Market Flexibility: An International Comparison”, *Scandinavian Journal of Economics 99*, 315-333.

Shapiro, C. and J.E. Stiglitz, 1984, “Equilibrium Unemployment as a Worker Discipline Device”, *American Economic Review 74*, 433-444.

Whinston, M.D., 1983, ”Moral hazard, Adverse Selection, and the Optimal Provision of Social Insurance”, *Journal of Public Economics 22*, 49-71.

Figure 1. Sickness absence and unemployment

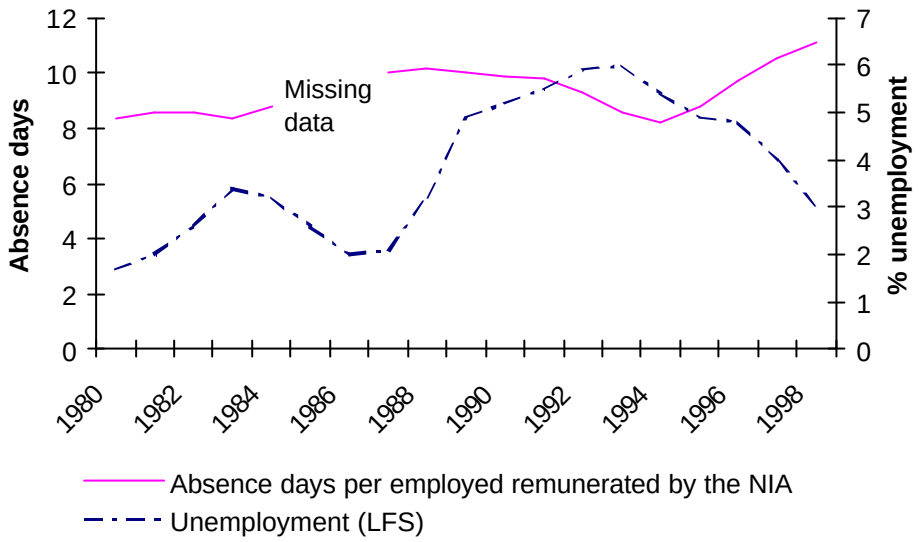

Table 1 Sample means

	Full		Males		Females		Common		Marginal		Non-marginal	
	1992	1995	1992	1995	1992	1995	1992	1995	1992	1995	1992	1995
Form_abs	6.61	5.59	5.43	4.50	7.86	6.74	5.98	6.27	2.23	1.77	7.15	6.07
Married	0.56	0.53	0.56	0.52	0.56	0.54	0.56	0.58	0.41	0.37	0.61	0.57
Pre_marr	0.11	0.12	0.09	0.10	0.13	0.14	0.11	0.13	0.09	0.09	0.12	0.13
Woman	0.49	0.49					0.49	0.49	0.58	0.53	0.48	0.48
Age	38.08	38.24	38.06	38.12	38.09	38.37	37.38	40.38	31.21	31.72	39.82	39.91
Children	0.39	0.41	0.36	0.36	0.42	0.46	0.41	0.41	0.48	0.46	0.41	0.43
Education	11.39	11.66	11.58	11.80	11.19	11.52	11.47	11.64	11.50	11.68	11.33	11.60
NonScand	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.04	0.03	0.02	0.01
Income	16.88	17.11	20.78	20.72	12.77	13.28	17.32	18.90	10.83	11.77	18.13	18.32
Spou_inc	8.86	8.94	5.95	6.02	11.93	12.04	9.16	9.85	7.39	6.97	9.56	9.66
Wealth	14.60	16.21	14.00	15.36	15.24	17.12	12.92	17.85	8.23	7.86	15.88	17.70
Experience	13.86	14.89	15.71	16.54	11.93	13.14	13.85	16.73	8.36	9.25	15.10	16.19
Seniority	5.66	6.01	6.12	6.38	5.16	5.62	5.66	7.23	1.10	1.02	6.32	6.73
Parttime	0.16	0.16	0.07	0.07	0.25	0.25	0.15	0.13	0.31	0.29	0.13	0.13
Unemp	6.36	5.49	6.36	5.50	6.36	5.48	6.36	5.48	6.43	5.55	6.37	5.49
Agrifish[1]	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.02	0.02	0.01	0.01
Salhottr[1]	0.26	0.25	0.26	0.25	0.26	0.24	0.25	0.24	0.27	0.24	0.25	0.24
Dwelfina[1]	0.08	0.08	0.08	0.08	0.08	0.07	0.08	0.08	0.07	0.08	0.08	0.08
Gover_hl[1]	0.21	0.21	0.19	0.19	0.23	0.23	0.21	0.22	0.22	0.20	0.21	0.22
Healtsoc[1]	0.13	0.14	0.04	0.04	0.23	0.25	0.13	0.14	0.20	0.20	0.13	0.14
Norway_a[1]	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
Norway_e[1]	0.15	0.15	0.16	0.16	0.15	0.15	0.15	0.16	0.14	0.15	0.16	0.15
Norway_s[1]	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.09	0.10	0.09	0.09	0.09
Norway_w[1]	0.20	0.21	0.21	0.21	0.20	0.20	0.21	0.20	0.22	0.20	0.20	0.20
Norway_m[1]	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.13	0.13	0.14	0.14
Norway_n[1]	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.11	0.11	0.10	0.10
≥1 absence	0.12	0.12	0.10	0.09	0.14	0.14	0.11	0.12	0.11	0.11	0.13	0.12
Absence days	10.6	9.6	9.1	7.6	12.2	11.6	8.0	10.3	8.6	9.0	11.4	10.1
Absence days if ≥1 absence	85.4	82.2	87.2	80.3	84.0	83.6	71.2	83.3	78.3	82.2	86.0	82.7
N	82349	88354	42234	45456	40115	42898	69131	69131	4028	4783	72966	78169

[1]Dummy variables used in the regressions but not reported

Table 2 Distribution of absence days net of 14 days (max(absence days-14,0))

	1992						1995					
	All		Males		Females		All		Males		Females	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
≤14	2628	25.8	1182	27.02	1446	24.88	2758	27.05	1275	29.85	1483	25.03
15-28	1663	16.33	707	16.16	956	16.45	1634	16.02	703	16.46	931	15.71
29-42	1134	11.13	428	9.79	706	12.15	1148	11.26	455	10.65	693	11.7
43-56	803	7.88	326	7.45	477	8.21	792	7.77	306	7.16	486	8.20
57-70	559	5.49	230	5.26	329	5.66	565	5.54	206	4.82	359	6.06
71-84	478	4.69	194	4.44	284	4.89	532	5.22	184	4.31	348	5.87
85-98	341	3.35	131	2.99	210	3.61	349	3.42	137	3.21	212	3.58
99-112	358	3.51	148	3.38	210	3.61	344	3.37	142	3.32	202	3.41
113-126	234	2.30	104	2.38	130	2.24	263	2.58	117	2.74	146	2.46
127-140	251	2.46	128	2.93	123	2.12	235	2.30	99	2.32	136	2.30
141-210	801	7.86	371	8.48	430	7.40	752	7.37	304	7.12	448	7.56
211-280	549	5.39	248	5.67	301	5.18	480	4.71	200	4.68	280	4.73
281-365	386	3.79	177	4.05	209	3.60	345	3.38	144	3.37	201	3.39
Total	10185	100	4374	100	5811	100	10197	100	4272	100	5925	100

Note: See Table 1 for fraction of sample with absences

Table 3 Hurdle regressions for 1992 and 1995 (full samples)

	1992		1995	
	Part 1: Pr(Y>0) (Logit)			
	Coef.	z-value	Coef.	z-value
Form_abs	0.0077	27.049	0.0074	25.342
Married	0.2221	5.310	0.1152	2.817
Pre_marr	0.3584	8.682	0.2901	7.338
Woman	0.4600	16.161	0.4668	16.863
Age	0.0129	7.736	0.0152	8.285
Children	0.0481	2.816	0.0987	6.139
Education	-0.1327	-25.179	-0.1170	-22.783
NonScand	0.2036	2.342	0.1588	1.770
Income	0.0503	11.668	0.0827	18.125
Inc_sqr	-0.0010	-10.546	-0.0018	-16.665
Spou_inc	-0.0071	-5.117	-0.0066	-4.928
Wealth	-0.0017	-4.112	-0.0041	-9.387
Experience	0.0028	1.079	-0.0028	-1.120
Seniority	-0.0125	-5.643	-0.0047	-2.234
Parttime	-0.2378	-6.574	-0.1700	-4.879
Unemp	0.0065	0.941	-0.0093	-1.192
Constant	-1.7636	-17.891	-2.1352	-21.670
Loglikelihood	-28820.3		-29595.7	
N	82349		88354	
Mean predicted	0.1189		0.1154	
	Part 2: E(Y|Y³ 1) (Negative binomial)			
	Coef.	z-value	Coef.	z-value
Form_abs	0.0023	7.640	0.0026	8.218
Married	0.0437	0.978	0.0116	0.266
Pre_marr	0.0264	0.613	0.0270	0.648
Woman	-0.0746	-2.366	0.0099	0.320
Age	0.0196	11.023	0.0182	9.115
Children	0.0246	1.311	0.0457	2.521
Education	-0.0241	-4.263	-0.0150	-2.634
NonScand	0.1152	1.217	-0.0854	-0.863
Income	-0.0118	-4.201	-0.0105	-2.780
Inc_sqr	0.0001	2.031	0.0001	1.276
Spou_inc	-0.0050	-3.323	-0.0031	-2.244
Wealth	0.0000	-0.011	0.0004	0.821
Experience	-0.0076	-2.587	-0.0066	-2.317
Seniority	-0.0001	-0.021	-0.0024	-1.038
Parttime	-0.0177	-0.466	0.0076	0.201
Unemp	-0.0048	-0.638	-0.0111	-1.253
Constant	3.9755	39.803	3.8314	37.641
Alpha	1.409		1.446	
Loglikelihood	-52972.16		-52612.9	
N	10185		10197	
Mean predicted	69.5		68.2	

Note: Location and sector dummies are used in the regressions but not reported
LR tests $b^{92} = b^{95}$: Part 1: 71.56 (DF=28) Part 2: 35.22 (DF=29)

Table 4 Hurdle regressions for common sample

	1992		1995	
	Part 1: Pr(Y>0) (Logit)			
Abs15yr1	Coef.	z-value	Coef.	z-value
Form_abs	0.0087	25.493	0.0075	23.767
Married	0.2372	4.917	0.1581	3.518
Pre_marr	0.3861	8.181	0.2872	6.623
Woman	0.4958	15.156	0.4822	15.500
Age	0.0074	3.584	0.0137	6.795
Children	0.0378	1.962	0.0857	4.747
Education	-0.1314	-21.723	-0.1146	-20.366
NonScand	0.1359	1.286	0.3018	2.818
Income	0.0464	9.212	0.0739	13.360
Inc_sqr	-0.0009	-8.403	-0.0016	-13.224
Spou_inc	-0.0068	-4.245	-0.0076	-5.210
Wealth	-0.0017	-3.364	-0.0041	-8.691
Experience	0.0058	1.869	-0.0027	-0.954
Seniority	-0.0121	-4.583	-0.0034	-1.505
Parttime	-0.2302	-5.497	-0.1236	-3.109
Unemp	0.0083	1.048	-0.0088	-1.015
Constant	-1.7083	-14.989	-2.0583	-17.821
Loglikelihood	-22657.8		-24129.8	
N	69131		69131	
Mean predicted	0.1113		0.1223	
	Part 2: E(Y|Y³ 1) (Negative binomial)			
	Coef.	z-value	Coef.	z-value
	0.0025	6.979	0.0027	8.086
Married	0.0531	1.026	0.0344	0.719
Pre_marr	0.0658	1.319	0.0562	1.222
Woman	-0.0135	-0.366	0.0150	0.435
Age	0.0118	5.209	0.0179	8.249
Children	0.0314	1.461	0.0402	1.992
Education	-0.0190	-2.910	-0.0105	-1.691
NonScand	0.1511	1.293	-0.0974	-0.836
Income	-0.0066	-2.120	-0.0146	-2.558
Inc_sqr	0.0000	1.458	0.0001	1.029
Spou_inc	-0.0045	-2.543	-0.0038	-2.605
Wealth	-0.0003	-0.890	0.0004	0.847
Experience	-0.0035	-1.037	-0.0071	-2.241
Seniority	-0.0009	-0.303	-0.0032	-1.300
Parttime	-0.0913	-2.051	0.0037	0.086
Unemp	-0.0060	-0.679	-0.0132	-1.360
Constant	3.9048	33.892	3.8633	31.939
Alpha	1.4258		1.4433	
Loglikelihood	-38455.3		-43740.3	
N	7697		8452	
Mean predicted	57.3		69.4	

Note: Location and sector dummies are used in the regressions but not reported

LR tests $b^{92} = b^{95}$: Part 1: 64.27 (DF=28) Part 2: 103.93 (DF=29)

Table 5 Hurdle regressions for marginal and non-marginal samples

	1992				1995			
	Non-marginal		Marginal		Non-marginal		Marginal	
	Part 1: Pr(Y>0) (Logit)							
	Coef.	z-value	Coef.	z-value	Coef.	z-value	Coef.	z-value
Form_abs	0.0076	26.079	0.0170	4.793	0.0073	24.414	0.0209	5.660
Married	0.2064	4.838	0.3892	1.970	0.0914	2.178	0.5241	2.923
Pre_marr	0.3204	7.612	0.6544	3.348	0.2569	6.341	0.6448	3.720
Woman	0.4410	14.914	0.4450	3.240	0.4717	16.306	0.3114	2.562
Age	0.0105	6.117	0.0100	1.116	0.0143	7.553	-0.0081	-0.849
Children	0.0115	0.643	0.1123	1.602	0.0816	4.836	0.0259	0.385
Education	-0.1259	-23.156	-0.1285	-4.921	-0.1109	-20.857	-0.1133	-4.794
NonScand	0.1965	2.143	-0.0399	-0.132	0.1940	2.071	-0.4249	-1.215
Income	0.0333	7.498	0.1357	5.413	0.0759	15.372	0.0219	2.224
Inc_sqr	-0.0007	-7.461	-0.0035	-4.424	-0.0017	-14.917	-0.0001	-0.743
Spou_inc	-0.0083	-5.799	-0.0022	-0.340	-0.0071	-5.148	-0.0080	-1.219
Wealth	-0.0014	-3.602	-0.0099	-2.809	-0.0040	-8.976	-0.0081	-2.641
Experience	0.0037	1.377	0.0026	0.200	-0.0025	-0.947	0.0146	1.209
Seniority	-0.0121	-5.400	0.0424	0.545	-0.0041	-1.922	0.2366	3.304
Parttime	-0.2062	-5.346	-0.2979	-2.231	-0.1517	-4.067	-0.3063	-2.447
Unemp	0.0054	0.748	-0.0197	-0.632	-0.0119	-1.459	-0.0021	-0.063
Constant	-1.5055	-14.451	-2.0704	-4.398	-2.0521	-19.525	-1.3641	-3.182
Loglikelihood	-26747.8		-1284.1		-27221.6		-1549.6	
N	72966		4028		78169		4783	
Mean predicted	0.1312		0.1095		0.1213		0.1089	
	Part 2: E(Y|Y³ 1) (Negative binomial)							
	Coef.	z-value	Coef.	z-value	Coef.	z-value	Coef.	z-value
Form_abs	0.0024	7.692	-0.0029	-0.920	0.0026	7.994	0.0022	0.619
Married	0.0397	0.857	0.0450	0.238	0.0041	0.092	0.2329	1.140
Pre_marr	0.0210	0.475	0.1364	0.638	0.0231	0.535	0.1592	0.878
Woman	-0.0826	-2.532	0.1079	0.670	0.0063	0.196	-0.0507	-0.361
Age	0.0203	11.156	-0.0038	-0.407	0.0179	8.730	0.0136	1.166
Children	0.0244	1.245	-0.0340	-0.440	0.0432	2.294	0.0169	0.203
Education	-0.0227	-3.902	-0.0310	-1.046	-0.0124	-2.114	-0.0378	-1.293
NonScand	0.1027	1.034	0.2749	0.814	-0.0400	-0.388	-0.8970	-2.181
Income	-0.0134	-4.432	-0.0212	-0.698	-0.0145	-2.881	0.0047	0.364
Inc_sqr	0.0001	2.132	0.0005	0.532	0.0001	1.249	0.0000	-0.029
Spou_inc	-0.0051	-3.257	-0.0043	-0.804	-0.0031	-2.212	-0.0088	-1.047
Wealth	0.0000	-0.113	0.0007	0.371	0.0005	0.974	-0.0001	-0.049
Experience	-0.0083	-2.755	0.0292	2.047	-0.0062	-2.090	-0.0163	-1.119
Seniority	-0.0003	-0.114	0.0814	0.862	-0.0020	-0.852	0.1147	1.368
Parttime	-0.0236	-0.587	0.0559	0.366	-0.0042	-0.105	0.0529	0.382
Unemp	-0.0065	-0.824	0.0236	0.700	-0.0127	-1.392	0.0284	0.671
Constant	3.9882	38.349	4.1648	7.843	3.8845	35.084	3.9269	8.259
Alpha	1.4031		1.3236		1.4444		1.3495	
Loglikelihood	-49861.0		-2242.7		-49004.8		-2690.5	
N	9570		441		9485		521	
Mean predicted	72.1		64.2		68.7		68.5	

Note: Location and sector dummies are used in the regressions but not reported

LR tests $b^{non\text{-}marg} = b^{marg}$: Part 1, 1992: 67.00 (DF=28) Part 2, 1992: 30.22 (DF=29)
Part 1, 1995: 78.53 (DF=28) Part 2, 1995: 27.65 (DF=29)

Table 6 Decompositions of mean differences in outcomes

	All			Males			Females		
	Total	Charac-teristics	Behavi-our	Total	Charac-teristics	Behavi-our	Total	Charac-teristics	Behavi-our
Full sample 95 vs. 92, Part 1	-0.0083	-0.0053	-0.0030	-0.0096	-0-0096	(-)	-0.0067	-0.0041	-0.0027
Full sample 95 vs. 92, Part 2	-3.2	-3.2	(-)	-6.9	-6.9	(-)	-0.5	-0.5	(-)
Common sample 95 vs. 92, Part 1	0.0109	0.0028	0.0082	0.0059	0.0002	0.0058	0.0162	0.0058	0.0104
Common sample 95 vs. 92, Part 2	12.1	0.9	11.2	9.0	1.0	8.0	14.3	1.3	12.0
Marg. vs. non-marg, 92, Part 1	-0.0217	-0.0240	0.0023	-0.0217	-0.0192	-0.0025	-0.0296	-0.0296	(-)
Marg. vs. non-marg, 92, Part 2	-7.9	-7.9	(-)	-7.3	-7.3	(-)	-7.4	-7.4	(-)
Marg. vs. non-marg, 95, Part 1	-0.0124	-0.0205	0.0081	0.0024	-0.0109	0.0133	-0.0302	-0.0328	0.0025
Marg. vs. non-marg, 95, Part 2	-0.2	-0.2	(-)	1.6	1.6	(-)	-1.0	-1.0	(-)

Notes: Part 1: Pr(Y>0) Part 2: E(Y|Y>0)

Non-rejection of $b^0 = b^1$ is interpreted as Total difference = Difference due to characteristics

Table A1 LR test for equal coefficients males and females

	Part 1 (logit)	Part 2 (negbin)
Full sample 1992	206.0	50.6
1995	205.4	64.4
Common sample 1992	188.8	26.4
1995	140.2	63.6
Non-marginal 1992	171.0	55.2
1995	25.6	24.4
Marginal 1992	158.6	61.8
1995	42.8	29.0

DF= 27 for part 1, 28 for part 2 with critical values 40.1 and 41.3 (5%)

Table A2 Hurdle regressions for 1992 and 1995 Males (full sample)

	1992		1995	
	Part 1: Pr(Y>0) (Logit)			
	Coef.	z-value	Coef.	z-value
Form_abs	0.0085	19.011	0.0077	16.806
Married	0.2440	3.772	0.0885	1.356
Pre_marr	0.3428	5.219	0.2674	4.231
Age	0.0186	6.196	0.0165	4.742
Children	-0.0871	-3.047	0.0161	0.583
Education	-0.1385	-17.552	-0.1277	-16.080
NonScand	0.2896	2.450	0.1223	0.954
Income	0.0181	3.226	0.0526	8.496
Inc_sqr	-0.0004	-3.958	-0.0012	-9.263
Spou_inc	-0.0107	-3.629	-0.0117	-3.895
Wealth	-0.0020	-3.105	-0.0028	-4.286
Experience	0.0015	0.293	-0.0002	-0.032
Seniority	-0.0157	-5.161	-0.0078	-2.678
Parttime	-0.5429	-5.741	-0.3008	-3.538
Unemp	0.0170	1.613	0.0100	0.830
Constant	-1.5268	-10.729	-1.7338	-11.752
Loglikelihood	-13091.1		-13289.0	
N	42234		45456	
Mean predicted	0.1036		0.0940	
	Part 2: E(Y|Y³ 1) (Negative binomial)			
	Coef.	z-value	Coef.	z-value
Form_abs	0.0023	4.954	0.0025	4.760
Married	0.1161	1.637	0.0921	1.254
Pre_marr	0.1017	1.446	0.1408	2.007
Age	0.0260	8.048	0.0277	7.016
Children	-0.0390	-1.252	0.0107	0.337
Education	-0.0194	-2.178	-0.0132	-1.402
NonScand	-0.0865	-0.651	-0.3199	-2.176
Income	-0.0116	-3.199	-0.0145	-2.595
Inc_sqr	0.0001	1.893	0.0001	1.625
Spou_inc	-0.0087	-2.593	-0.0078	-2.295
Wealth	0.0000	-0.070	0.0014	1.948
Experience	-0.0133	-2.328	-0.0158	-2.749
Seniority	-0.0043	-1.318	-0.0077	-2.328
Parttime	0.0071	0.068	0.0828	0.853
Unemp	-0.0057	-0.473	-0.0242	-1.728
Constant	3.7682	25.624	3.7585	23.723
Alpha	1.4581		1.5379	
Loglikelihood	-22801.0		-21828.6	
N	4374		4272	
Mean predicted	73.3		66.4	

Note: Location and sector dummies are used in the regressions but not reported

LR tests $b^{92} = b^{95}$: Part 1: 34.73 (DF=27) Part 2: 29.23 (DF=28)

Table A3 Hurdle regressions for 1992 and 1995 Females (full sample)

	1992		1995	
	Part 1: Pr(Y>0) (Logit)			
	Coef.	z-value	Coef.	z-value
Form_abs	0.0070	18.853	0.0071	18.691
Married	0.3232	5.604	0.2500	4.501
Pre_marr	0.4193	7.740	0.3421	6.658
Age	0.0121	5.744	0.0166	7.326
Children	0.1534	6.881	0.1667	8.078
Education	-0.1218	-16.559	-0.1037	-14.858
NonScand	0.0718	0.540	0.1537	1.193
Income	0.1165	13.708	0.1359	16.297
Inc_sqr	-0.0029	-10.480	-0.0034	-12.673
Spou_inc	-0.0089	-5.248	-0.0084	-5.142
Wealth	-0.0015	-2.929	-0.0052	-8.761
Experience	-0.0071	-2.118	-0.0106	-3.307
Seniority	-0.0060	-1.831	0.0012	0.383
Parttime	-0.1272	-3.182	-0.1153	-2.983
Unemp	-0.0030	-0.321	-0.0250	-2.404
Constant	-1.8379	-13.040	-2.2849	-16.492
Loglikelihood	-15626.2		-16204.7	
N	40115		42898	
Mean predicted	0.1449		0.1381	
	Part 2: E(Y|Y³ 1) (Negative binomial)			
	Coef.	z-value	Coef.	z-value
Form_abs	0.0023	5.983	0.0026	6.510
Married	0.0342	0.567	-0.0196	-0.348
Pre_marr	0.0150	0.269	-0.0254	-0.481
Age	0.0160	7.174	0.0157	6.560
Children	0.0520	2.168	0.0544	2.417
Education	-0.0332	-4.411	-0.0188	-2.546
NonScand	0.2817	2.006	0.0265	0.195
Income	-0.0149	-2.270	0.0066	0.946
Inc_sqr	0.0002	1.387	-0.0004	-2.162
Spou_inc	-0.0046	-2.588	-0.0021	-1.311
Wealth	0.0001	0.151	-0.0003	-0.489
Experience	-0.0088	-2.461	-0.0086	-2.480
Seniority	0.0060	1.662	0.0036	1.116
Parttime	-0.0208	-0.493	0.0178	0.430
Unemp	-0.0038	-0.398	-0.0025	-0.223
Constant	4.1158	29.431	3.7715	26.915
Alpha	1.3596		1.3689	
Loglikelihood	-30145.9		-30752.1	
N	5811		5925	
Mean predicted	70.0		69.6	

Note: Location and sector dummies are used in the regressions but not reported
LR tests $b^{92} = b^{95}$: Part 1: 44.98 (DF=27) Part 2: 26.35 (DF=28)

Table A4 Hurdle regressions for common sample (males)

	1992		1995	
	Part 1: Pr(Y>0) (Logit)			
	Coef.	z-value	Coef.	z-value
Form_abs	0.0096	17.574	0.0080	15.750
Married	0.2736	3.647	0.2106	2.918
Pre_marr	0.3462	4.533	0.3223	4.568
Age	0.0064	1.583	0.0161	4.129
Children	-0.0712	-2.236	-0.0192	-0.628
Education	-0.1422	-15.187	-0.1247	-14.216
NonScand	0.2110	1.450	0.2059	1.340
Income	0.0023	0.386	0.0524	6.586
Inc_sqr	-0.0001	-1.423	-0.0012	-7.764
Spou_inc	-0.0110	-3.208	-0.0138	-4.286
Wealth	-0.0013	-1.567	-0.0028	-3.886
Experience	0.0134	2.107	-0.0040	-0.687
Seniority	-0.0122	-3.332	-0.0053	-1.690
Parttime	-0.4924	-4.408	-0.1508	-1.428
Unemp	0.0246	2.020	0.0138	1.016
Constant	-1.2265	-7.369	-1.7735	-10.074
Loglikelihood	-10119.0		-10591.2	
N	35473		35473	
Mean predicted	0.0907		0.0967	
	Part 2: E(Y|Y³ 1) (Negative binomial)			
	Coef.	z-value	Coef.	z-value
Form_abs	0.0029	4.879	0.0028	4.757
Married	0.1245	1.481	0.0983	1.191
Pre_marr	0.1236	1.490	0.1872	2.368
Age	0.0121	2.677	0.0290	6.584
Children	-0.0361	-1.032	0.0285	0.806
Education	-0.0039	-0.363	-0.0107	-1.029
NonScand	0.1810	1.059	-0.3603	-2.069
Income	-0.0094	-2.185	-0.0243	-2.576
Inc_sqr	0.0001	1.672	0.0002	1.577
Spou_inc	-0.0092	-2.308	-0.0079	-2.188
Wealth	-0.0001	-0.167	0.0018	2.138
Experience	0.0005	0.070	-0.0181	-2.701
Seniority	-0.0079	-1.884	-0.0060	-1.656
Parttime	0.0032	0.026	0.1120	0.924
Unemp	-0.0057	-0.400	-0.0292	-1.831
Constant	3.7194	21.253	3.8333	19.597
Alpha	1.5011		1.5418	
Loglikelihood	-16062.9		-17511.0	
N	3218		3429	
Mean predicted	57.7		67.0	

Note: Location and sector dummies are used in the regressions but not reported
LR tests $b^{92} = b^{95}$: Part 1: 54.13 (DF=27) Part 2: 43.59 (DF=28)

Table A5 Hurdle regressions for common sample (females)

	1992		1995	
	Part 1: Pr(Y>0) (Logit)			
	Coef.	z-value	Coef.	z-value
Form_abs	0.0078	17.954	0.0071	17.714
Married	0.3111	4.694	0.2368	3.902
Pre_marr	0.4389	7.167	0.3099	5.549
Age	0.0105	4.083	0.0159	6.444
Children	0.1328	5.256	0.1687	7.208
Education	-0.1125	-13.577	-0.1050	-13.803
NonScand	0.0497	0.315	0.3365	2.197
Income	0.1351	13.044	0.1143	11.699
Inc_sqr	-0.0035	-10.387	-0.0028	-9.417
Spou_inc	-0.0081	-4.071	-0.0082	-4.677
Wealth	-0.0020	-3.144	-0.0051	-8.118
Experience	-0.0070	-1.835	-0.0099	-2.838
Seniority	-0.0088	-2.316	0.0006	0.175
Parttime	-0.1081	-2.347	-0.0948	-2.187
Unemp	-0.0046	-0.437	-0.0271	-2.376
Constant	-2.0470	-12.569	-2.0742	-12.960
Loglikelihood	-12444.4		-13468.5	
N	33658		33658	
Mean predicted	0.1331		0.1492	
	Part 2: E(Y|Y³ 1) (Negative binomial)			
	Coef.	z-value	Coef.	z-value
Form_abs	0.0021	4.866	0.0026	6.380
Married	0.0250	0.362	0.0113	0.185
Pre_marr	0.0314	0.490	0.0032	0.055
Age	0.0115	4.199	0.0149	5.786
Children	0.0806	2.878	0.0362	1.432
Education	-0.0319	-3.764	-0.0144	-1.797
NonScand	0.1391	0.832	0.0140	0.088
Income	0.0059	0.563	-0.0014	-0.164
Inc_sqr	-0.0002	-0.644	-0.0002	-0.948
Spou_inc	-0.0037	-1.780	-0.0029	-1.746
Wealth	-0.0006	-0.866	-0.0004	-0.658
Experience	-0.0081	-1.967	-0.0085	-2.276
Seniority	0.0070	1.661	0.0002	0.073
Parttime	-0.0884	-1.763	0.0159	0.342
Unemp	-0.0083	-0.744	-0.0033	-0.267
_cons	3.9501	23.811	3.8667	24.164
Alpha	1.3617		1.3633	
Loglikelihood	-22374.2		-26197.5	
N	4479		5023	
Mean predicted	57.0		71.3	

Note: Location and sector dummies are used in the regressions but not reported
LR tests $b^{92} = b^{95}$: Part 1: 52.26 (DF=27) Part 2: 92.08 (DF=28)

Table A6 Hurdle regressions for marginal and non-marginal samples (males)

	1992				1995			
	Non-marginal		Marginal		Non-marginal		Marginal	
	Part 1: Pr(Y>0) (Logit)							
	Coef.	z-value	Coef.	z-value	Coef.	z-value	Coef.	z-value
Form_abs	0.0083	18.226	0.0179	2.794	0.0076	16.329	0.0237	4.209
Married	0.2540	3.882	0.5350	1.388	0.0914	1.366	0.2099	0.691
Pre_marr	0.3409	5.121	0.4512	1.275	0.2611	4.017	0.3854	1.420
Age	0.0184	6.054	-0.0025	-0.116	0.0158	4.424	0.0121	0.635
Children	-0.1041	-3.564	-0.0719	-0.452	0.0086	0.302	0.0082	0.057
Education	-0.1350	-16.758	-0.1309	-2.841	-0.1212	-14.827	-0.1671	-4.288
NonScand	0.2343	1.897	0.2514	0.517	0.1461	1.098	-0.5794	-1.034
Income	0.0057	1.029	0.1064	2.669	0.0491	7.217	0.0255	2.018
Inc_sqr	-0.0002	-2.217	-0.0030	-2.712	-0.0012	-8.517	-0.0001	-0.726
Spou_inc	-0.0111	-3.735	-0.0031	-0.166	-0.0117	-3.853	-0.0121	-0.709
Wealth	-0.0017	-2.703	-0.0309	-2.862	-0.0026	-4.008	-0.0057	-1.230
Experience	-0.0024	-0.459	0.0437	1.451	-0.0004	-0.071	-0.0045	-0.188
Seniority	-0.0148	-4.843	0.0220	0.165	-0.0065	-2.194	0.1881	1.684
Parttime	-0.3810	-3.663	-0.3908	-1.243	-0.2421	-2.518	-0.3362	-1.384
Unemp	0.0165	1.519	-0.0238	-0.447	0.0064	0.513	0.0133	0.257
Constant	-1.3078	-8.849	-1.9597	-2.367	-1.6947	-10.781	-1.0049	-1.479
Loglikelihood	-12314.7		-456.2		-12238.5		-673.2	
N	38154		1708		40685		2254	
Mean predicted	0.1090		0.0872		0.0974		0.0998	
	Part 2: E(Y|Y³ 1) (Negative binomial)							
	Coef.	z-value	Coef.	z-value	Coef.	z-value	Coef.	z-value
Form_abs	0.0025	5.252	-0.0048	-0.712	0.0025	4.662	0.0063	1.209
Married	0.1038	1.432	0.3892	0.961	0.0600	0.781	0.6197	1.937
Pre_marr	0.1022	1.427	-0.0984	-0.243	0.1242	1.699	0.2269	0.791
Age	0.0266	8.166	-0.0122	-0.706	0.0284	7.055	-0.0194	-0.820
Children	-0.0409	-1.287	-0.0299	-0.168	0.0209	0.642	-0.3628	-2.017
Education	-0.0171	-1.893	-0.0929	-1.417	-0.0122	-1.262	0.0567	1.112
NonScand	-0.0688	-0.499	-0.3137	-0.483	-0.2781	-1.823	-1.0666	-1.652
Income	-0.0131	-3.460	0.0014	0.029	-0.0211	-2.679	-0.0303	-1.643
Inc_sqr	0.0001	2.012	0.0000	0.001	0.0002	1.661	0.0003	1.667
Spou_inc	-0.0087	-2.587	-0.0299	-1.188	-0.0079	-2.260	0.0088	0.512
Wealth	0.0000	0.020	-0.0222	-2.058	0.0015	2.091	-0.0012	-0.264
Experience	-0.0133	-2.268	0.0454	1.671	-0.0161	-2.708	0.0179	0.588
Seniority	-0.0044	-1.333	0.2112	1.255	-0.0070	-2.085	0.0285	0.230
Parttime	-0.0016	-0.014	-0.0646	-0.163	0.0721	0.655	0.2433	0.876
Unemp	-0.0076	-0.615	0.0577	0.902	-0.0232	-1.590	-0.0085	-0.128
Constant	3.7624	24.885	4.4086	4.364	3.8233	21.931	4.3003	5.446
Alpha	1.4508		1.1535		1.5417		1.2362	
Loglikelihood	-21694.7		-753.3		-20278.4		-1155.8	
N	4157		149		3964		225	
Mean predicted	73.9		66.6		66.9		68.6	

Note: Location and sector dummies are used in the regressions but not reported

LR tests $b^{non\text{-}marg} = b^{marg}$: Part 1, 1992: 43.85 (DF=27) Part 2, 1992: 31.90 (DF=28)
Part 1, 1995: 52.79 (DF=27) Part 2, 1995: 24.65 (DF=28)

Table A7 Hurdle regressions for marginal and non-marginal samples (females)

	1992				1995			
	Non-marginal		Marginal		Non-marginal		Marginal	
	Part 1: Pr(Y>0) (Logit)							
	Coef.	z-value	Coef.	z-value	Coef.	z-value	Coef.	z-value
Form_abs	0.0069	18.349	0.0176	3.991	0.0069	17.987	0.0183	3.687
Married	0.2914	4.901	0.2914	1.183	0.1887	3.303	0.9557	3.857
Pre_marr	0.3756	6.757	0.6299	2.605	0.2926	5.556	0.9080	3.942
Age	0.0097	4.457	0.0153	1.493	0.0162	6.905	-0.0055	-0.472
Children	0.1130	4.727	0.1741	2.147	0.1464	6.654	0.0306	0.384
Education	-0.1163	-15.210	-0.1217	-3.706	-0.1001	-13.832	-0.0575	-1.821
NonScand	0.0659	0.464	-0.1193	-0.292	0.1983	1.463	-0.4303	-0.936
Income	0.0932	10.389	0.1696	4.252	0.1226	13.803	0.1469	3.701
Inc_sqr	-0.0023	-8.142	-0.0047	-2.927	-0.0030	-10.986	-0.0050	-3.157
Spou_inc	-0.0095	-5.440	-0.0008	-0.110	-0.0080	-4.808	-0.0164	-1.966
Wealth	-0.0013	-2.561	-0.0060	-1.773	-0.0050	-8.430	-0.0088	-2.173
Experience	-0.0054	-1.602	-0.0256	-1.584	-0.0106	-3.213	0.0064	0.395
Seniority	-0.0063	-1.907	0.0690	0.706	0.0008	0.259	0.2111	2.209
Parttime	-0.1116	-2.628	-0.2543	-1.687	-0.1069	-2.608	-0.2194	-1.461
Unemp	-0.0041	-0.424	-0.0216	-0.557	-0.0269	-2.498	-0.0098	-0.210
Constant	-1.5761	-10.444	-1.8177	-2.973	-2.1361	-14.452	-2.6486	-4.148
Loglikelihood	-14347.6		-815.1		-14903.8		-855.0	
N	34812		2320		37484		2529	
Mean predicted	0.1555		0.1259		0.1473		0.1170	
	Part 2: E(Y|Y³ 1) (Negative binomial)							
	Coef.	z-value	Coef.	z-value	Coef.	z-value	Coef.	z-value
Form_abs	0.0023	5.766	-0.0007	-0.178	0.0025	6.313	0.0004	0.078
married	0.0385	0.606	0.0028	0.012	-0.0078	-0.134	-0.1430	-0.474
Pre_marr	0.0030	0.051	0.3099	1.199	-0.0166	-0.304	-0.1188	-0.480
Age	0.0164	7.124	-0.0024	-0.221	0.0147	5.998	0.0210	1.475
Children	0.0544	2.133	-0.0697	-0.797	0.0470	1.984	0.1328	1.423
Education	-0.0329	-4.209	0.0104	0.288	-0.0153	-2.018	-0.0924	-2.427
NonScand	0.2614	1.750	0.7257	1.605	0.0675	0.475	-0.6590	-1.249
Income	-0.0179	-2.487	-0.0008	-0.016	0.0013	0.165	0.0314	0.756
Inc_sqr	0.0003	1.619	-0.0007	-0.388	-0.0003	-1.589	-0.0005	-0.342
Spou_inc	-0.0051	-2.703	-0.0024	-0.386	-0.0022	-1.396	-0.0072	-0.660
Wealth	0.0000	-0.105	0.0009	0.439	-0.0003	-0.482	0.0014	0.401
Experience	-0.0097	-2.652	0.0283	1.533	-0.0075	-2.123	-0.0282	-1.524
Seniority	0.0060	1.640	0.0922	0.743	0.0037	1.135	0.2141	1.862
Parttime	-0.0265	-0.597	0.1608	0.938	0.0087	0.198	-0.0697	-0.401
Unemp	-0.0057	-0.564	0.0008	0.021	-0.0063	-0.537	0.0211	0.393
Constant	4.1548	28.149	3.9548	5.796	3.8453	26.012	3.9699	5.485
Alpha	1.3516		1.2909		1.3639		1.2847	
Loglikelihood	-28138.7		-1477.2		-28695.5		-1519.8	
N	5413		292		5521		296	
Mean predicted	70.8		63.4		70.0		69.1	

Note: Location and sector dummies are used in the regressions but not reported

LR tests $b^{non\text{-}marg} = b^{marg}$: Part 1, 1992: 39.34 (DF=27) Part 2, 1992: 26.72 (DF=28)
Part 1, 1995: 43.25 (DF=27) Part 2, 1995: 29.64 (DF=28)