Applied Economics Letters, 1999, **6**, 337–341

Wages, firm size and absenteeism

RAINER WINKELMANN

Department of Economics, University of Canterbury, Christchurch, New Zealand
E-mail: r.winkelmann@econ.canterbury.ac.nz.

Received 19 May 1997

This paper examines two competing explanations for workers' absenteeism, the shirking hypothesis and the adjustment-to-equilibrium hypothesis. Data on German workers for 1985–88 from the German SocioEconomic Panel are used in order to estimate the determinants of workers' absenteeism. The results indicate that firm size matters after wage effects are controlled for. This evidence supports the shirking hypothesis.

I. INTRODUCTION

Workers' absences from the workplace cause substantial losses in production in most developed economies. For instance, in Germany the average worker misses eight working days per year as a result of absence. Assuming an average number of working days for a full-time worker of about 220 days per year, the effectively available amount of labour input is about 3.6% below the contracted amount. While the rate of under-utilization of labour is below common unemployment rates, the comparison illustrates the magnitude of the problem and the need for a better understanding of the factors that determine absenteeism.

The recent literature on absenteeism has developed around two broad themes. The first involves improvements in the microeconomic foundations that underpin the empirical analysis. Instead of correlating absence rates and socio-economic characteristics, these studies explicitly develop a labour supply model of absenteeism. An example in this area is the recent paper by Barmby *et al.* (1995). Their paper shows both theoretically and empirically how contract structures affect absence rates. Other papers falling in this category are Barmby *et al.* (1993), Johansson and Palme (1996), and, to some extent, Johansson and Brännäs (1996). The second theme are improved econometric methods, in particular if counts of absent days are used as a dependent variable. Econometric studies developing count data models for the analysis of workplace absenteeism include Delgado and Kniesner (1996).

In this paper, we derive and estimate a simple labour supply model of worker absenteeism that distinguishes between two potential explanations of absences, adjustment-to-equilibrium versus shrinking. The difference between the two explanations lies in their predicted effect of firm size on work absences. In the adjustment model, firm size affects absence rates only through its effect on wages, whereas in the shirking model firm size has a direct effect on absence rates, for given wages, by affecting the probability of being detected.

These two alternative hypotheses are tested using data from the German Socio-Economic Panel (GSOEP) on absence histories of German full-time workers over the period 1985–88. The econometric analysis is based on count data models that are appropriate whenever the dependent variable is a nonnegative integer. The empirical evidence provides support for the shirking hypothesis rather than the adjustment model.

II. A SIMPLE ECONOMIC MODEL OF WORKER ABSENTEEISM

The basic adjustment-to-equilibrium model of worker absenteeism goes back to Dunn and Youngblood (1986). It models absenteeism as a reaction of workers who would like to work less than full-time (40 hours per week, say) but are 'forced' into a full-time contract. The situation is depicted in Fig. 1. Point A represents the optimal choice for the worker. The corresponding hours of work H_A are less than H_F and hence the worker is 'overworked'. An alternative way to characterize overworked workers is to compare the marginal rate of substitution between income and leisure to the wage rate at point *F*. If the marginal rate of substitution exceeds the wage rate, the worker would be better off to work less.

How realistic is such a model? In the context of Germany, most workers are covered by union contracts that define standard working hours of 40 hours per week for full-time workers.

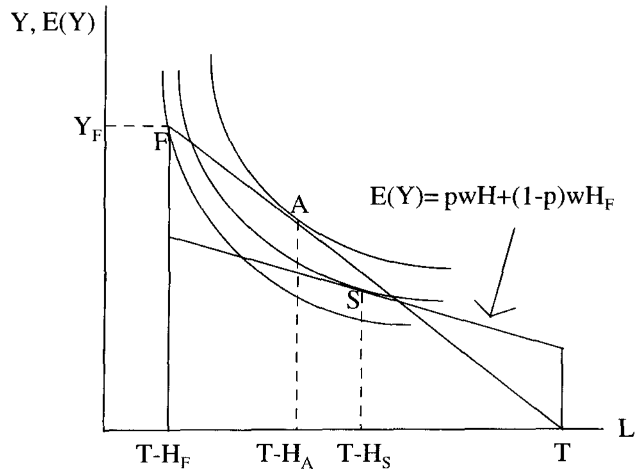

Fig. 1 *Absenteeism as adjustment-to-equilibrium versus shirking*

Actual working time frequently exceeds these nominal hours by overtime work, but downward adjustments are the exception. The 1985 wave of the GSOEP contained a question on the desired hours of work, taking into account a proportional adjustment of wages. In that year, the average full-time worker worked 44 hours per week, whereas average desired working hours were only 38 hours per week. Moreover, correlating the number of actual absent days during that year with an indicator variable for full-time workers wishing to work 35 hours or less, i.e., an 'overworked' employee, we find that such an employee has an extra two absent days per year. This finding is compatible with the adjustment towards equilibrium hypothesis.

However, the basic adjustment model has several restrictive features. Absenteeism is modelled as a pure labour supply phenomenon while labour demand is ignored. Similarly, the model assumes that the opportunity costs of absenteeism are equal to the wage rate. In practice, many absences are covered by sick leave provisions where the full wage or part of the wage is paid for a certain amount of time. This would rotate the budget constraint *FT* outwards around *F*. In the extreme, a horizontal line through *F* represents the case of no opportunity costs at all.

One such case would be if a worker could shirk without being detected. However, the shirking model needs to be modified if the detection is uncertain. This situation is depicted by the second, kinked budget constraint. Here, it is assumed that the absence is detected with probability p. Hence, with probability $1-p$ the worker receives the full-time wage $Y_F = wH_F$. We assume for simplicity that there is no penalty associated with being detected. Rather, if detected the worker receives an income wH where H is the actual hours of work. The relevant budget constraint depends on the wage rate and the probability of being detected:

$$E(y) = pwH + (1-p)wH_F$$

In Fig. 1, a (risk neutral) worker will maximize utility by working H_S hours. An important implication of this model is that the probability of detection has an unambiguous effect on the optimal amount of absenteeism $H_F - H_S$. An increase in p will increase the slope of the budget constraint while at the same time decreasing the expected shirking income $(1-p)Y_F$. Therefore workers will substitute leisure for expected income and shirk less. In the empirical analysis, the probability of detection is modelled as a function of firm size. In particular, it is assumed that firm size and p are inversely related: the larger the firm, the lower the probability of being detected. While firm size matters in the shirking model we would expect to see no direct effect of firm size in the pure adjustment-to-equilibrium model. There may be an indirect effect if firm size is correlated with wages.

In the adjustment model, an increase in the wage will reduce the number of absences if the substitution effect dominates the income effect. In the shirking model, the situation is more complicated since there is an additional income effect through the increase in shirking income $Y_F = wH_F$. The effect of this increase on expected income will be small if p is large as might reasonably be assumed. In this situation, the effect of wages on absence rates again depends on the relative size of income and substitution effects and is predicted to be negative if the labour supply curve is upwardly sloped. Therefore, the shirking model gives firms wanting to achieve a given absence rate a choice along the trade-off between p and w.

III. EMPIRICAL METHODS

The following empirical analysis models the annual number of absent days by a worker using count data regressions. Introductory surveys of the count data literature are Winkelmann and Zimmermann (1995) and Winkelmann (1997). Absences can be thought of as resulting from a sequence of Bernoulli trials. In a given year, there are n trials (where n is the number of working days, say 250) and on any day, a worker is absent with probability p and at work with probability $1-p$. The total number of absent days Y is the sum of the n Bernoulli variables X_n

$$Y = \sum_{n=1}^{N} X_n \tag{1}$$

Under the assumption that the trials are independent and that the probability $p = \lambda/n$ is constant it can be shown that the resulting number of absent days tends to a Poisson distribution with expected value λ as n tends to infinity. As in a log–linear model, explanatory variables are introduced by letting

$$\lambda_{it} = \exp(x'_{it}\beta) \tag{2}$$

where x_{it} is a $(k \times 1)$-vector of independent variables and β a conformable vector of covariates. The Poisson model makes two assumptions that are likely to be violated in an application to workplace absenteeism. First, it postulates independence of the process generating daily absences (Winkelmann, 1995). The probability of an absence in t does not depend on whether or not the worker was absent in $t-1$. However, Barmby *et al.* (1991) show directly through duration analysis that absence spells are subject to duration dependence, i.e. do not occur randomly over time. Evidence for state dependence is also found in Barmby *et al.* (1995). Secondly, the Poisson model postulates that all factors relevant for establishing the expected number of absent days are controlled for. In the practice of absent studies, there will always be at least one important unmeasurable determinant of absence behavior, namely an accurate measure of health. But unobserved heterogeneity invalidates the basic Poisson model.

In both cases the data are likely to display overdispersion, a situation in which the conditional mean exceeds the conditional variance of the number of absent days. A more general model count data model that allows for overdispersion is the negative binomial model (Negbin). It can be shown that the Negbin model can arise if the Bernoulli process is characterized by occurrence dependence. Alternatively, it arises if $\lambda_{it} = \exp(x_{it}'\beta)u_{it}$ where u_{it} is identically and independently gamma distributed. A test of the Poisson against the Negbin model is whether $\sigma_u^2 = 0$.

An estimator that explicitly takes into account the panel data structure of the data is provided in Hausman *et al.* (1984). They derive a Poisson model with gamma distributed individual specific effect $\exp(u_i)$, i.e. a model individual specific unobserved heterogeneity. The resulting model is of the negative binomial form with

$$P(y_{i1}, \ldots, y_{iT}) = \frac{\Gamma(\Sigma_t y_{it} + \alpha)}{\Gamma(\alpha)}\left(\frac{\alpha}{\alpha + \Sigma_t \lambda_{it}}\right)^{\alpha} \times \frac{1}{(\alpha + \Sigma_t \lambda_{it})^{\Sigma_t y_{it}}} \prod_{t=1}^{T}\left(\frac{\lambda_{it}^{y_{it}}}{y_{it}!}\right) \tag{3}$$

Hausman *et al.* generalize this model further by specifying a random effects negative binomial model where $\alpha/(1+\alpha)$ is distributed as $\beta(a, b)$. Under this assumption, α can be integrated out and the resulting joint probability function for individual i is

$$P(y_{i1}, \ldots, y_{iT}) = \frac{\Gamma(a+b)\Gamma(a + \Sigma_t \exp(x_{it}'\beta))\Gamma(b + \Sigma_t y_{it})}{\Gamma(a)\Gamma(b)\Gamma(a + b + \Sigma_t \exp(x_{it}'\beta) + \Sigma_t y_{it})} \times \Pi_t \frac{\Gamma(\exp(x_{it}'\beta + y_{it})}{\Gamma(\exp(x_{it}'\beta) y_{it}!} \tag{4}$$

Under the assumption of no random effects, $a = b = 0$. This restriction can be tested with a standard Wald test or a likelihood ratio test.

IV. EMPIRICAL RESULTS

The data are from the German Socio-Economic Panel (see Wagner *et al.* 1993). Workers are observed over the period 1985–88. The retrospective question on the number of absent days in the previous year (in $t+1$) is matched with information on current workplace and personal characteristics (in t), in particular firm size, hourly wage, employment status (blue collar or white collar, public sector or private sector), health status, tenure, age, sex, family status and contract type: limited or unlimited contract. The final sample of male workers aged 18 to 64 has 7373 observations. 1217 workers are present in all four years. As in Barmby *et al.* (1995) absence is defined as any non-attendance from work for reasons other than hospitalization. The fraction of absent days qualifying for sick leave benefits is unobserved but most likely large. Hospital days are excluded since arguably they are conditioned less than other absent days by individual labour supply optimization.

The dependent variable is the annual number of absent days. It has a lower limit of 0 and a theoretical upper limit of about 250 (the maximum number of workdays). In the final sample, 54% of all workers report no absence day and the largest number of absent days is 180. The average numbers of working days lost to absenteeism is eight with standard deviation 16. Hence, absence rates vary substantially across individuals, and the goal of this empirical analysis is to uncover whether differences in absence rates are systematically related to wages as predicted by the two models and to firm size as predicted by the shirking model, while controlling for various individual characteristics that might affect preferences. Table 1 provides a first insight into the nature and size of the correlations in our sample. For instance, workers on a non-continuing contract on average miss 6.3 days, while workers on continuing contracts miss 7.8 days. The t-test statistic of 2.25 shows that the hypothesis of equal absence rates for the two groups of workers can be rejected: workers on continuing contracts have significantly higher absence rates. Economically, though, the difference of 1.5 days per year is not very large. Differences, that are both statistically and economically significant, are found for the following categories: blue collar workers have five more absent days than white collar workers; and workers in good health have about six fewer absent days than workers in bad health. All other effects are below two days, notably a reduction in absent days for those who are satisfied with their job, who work in small firms; who are prime aged and who have tenure of five years or less. As found in other datasources married men are absent more frequently, and children at home reduce absenteeism for men (see Barmby *et al.*, 1991).

These patterns are confirmed by formal multivariate count data regressions. Regression results are given in Table 2. All

Table 1. *Number of absent days by personal characteristics*

	No		Yes		*t*-test
Continuing contract	6.320	(390)	7.752	(6983)	−2.25
Public sector	7.255	(5378)	8.811	(1995)	−3.21
Blue collar	5.750	(4333)	10.422	(3040)	−11.42
Married	6.342	(1010)	7.888	(6363)	−2.90
20–200 employees	7.808	(5646)	7.246	(1727)	1.26
201–2000 employees	7.273	(5822)	9.188	(1551)	−3.72
> 2000 employees	7.250	(4940)	8.542	(2433)	−3.04
Children at home	8.173	(4253)	6.999	(3120)	3.12
Age < 30	7.693	(5691)	7.618	(1682)	0.18
30 ≤ age < 45	8.426	(4467)	6.52	(2906)	5.10
Age ≥ 45	6.924	(4588)	8.914	(2785)	−4.61
Tenure < 5	7.964	(5966)	6.455	(1407)	3.71
5 ≤ tenure < 15	7.723	(6066)	7.456	(1307)	0.56
Tenure ≥ 45	6.937	(2714)	8.106	(4659)	−3.11
Satisfied with job	9.805	(1169)	7.275	(6204)	4.52
No handicap	12.494	(2012)	5.868	(5361)	12.28
No chronic condition	11.884	(1919)	6.195	(5454)	10.60

Notes: Number of observations in parentheses. *t*-test statistic is for H_0: means are equal.

regressions include log hourly wages as a control for the opportunity costs of absenteeism, three dummies for firm size, plus various characteristics of the worker. Poisson, negative binomial, and random effects negative binomial models are estimated each for a sample of 7373 observations. Estimated standard errors are in parentheses.

A comparison of the models reveals the inappropriateness of the Poisson model. The hypothesis $H_0 : \sigma_u^2 = 1/\alpha = 0$ is on the boundary of the parameter space and the standard likelihood ratio statistic does not have the usual distribution. However, it can be shown that under the null hypothesis, the likelihood ratio test statistic has a probability mass of 0.5 at zero and 0.5 chi-squared(1) for positive values. Accordingly, the Poisson model is rejected against the negative binomial model. The log-likelihoods of the negative binomial and random effects negative binomial models are not directly comparable. Based on any information criterion, the random effects model emerges as the model with the superior fit. A pervasive aspect of the regression results is the robustness of the main effects with respect to distributional assumptions. In fact it can be shown that the Poisson estimator is consistent even when the negative binomial model is the true model. In nonlinear regression models as estimated here, the estimated coefficients do not represent marginal effects and are not necessarily comparable across models. Therefore, Table 3 gives the estimated marginal effects evaluated at the means of the regressors. For instance, workers in the public sector are predicted to have around two more absent days per year than workers in the private sector. Similarly, workers with potentially higher relative values of non-market activities (married individuals) are predicted to have one additional day of absence.

What are the effects of wages and firm size? Based on the negative binomial model, a 10% increase in wages reduces the number of absent days by 0.3. The observed wage differential between larger firms and smaller firms during the period is about 6% (see, for instance, Winkelmann, 1996). Hence, based on the wage effect alone, we would expect that workers in larger firms have 0.2 fewer absent days than workers in smaller firms. However, the direct effect of firm size is of opposite direction and much larger: workers in large firms (201–2000 employees) are predicted to have 1.8 more absent days than workers in smaller firms (20–2000 employees) after

Table 2. *Regression results for number of absent days*

	Poisson	Negbin	Panel Negbin
Log wage	−0.412 (0.013)	−0.508 (0.094)	−0.124 (0.042)
Public sector	0.314 (0.010)	0.375 (0.077)	0.283 (0.031)
Blue collar	0.562 (0.009)	0.625 (0.073)	0.407 (0.029)
Married	0.196 (0.013)	0.241 (0.092)	0.122 (0.041)
Children at home	−0.026 (0.009)	−0.027 (0.065)	0.062 (0.029)
20–200 employees	0.203 (0.014)	0.174 (0.088)	0.178 (0.042)
201–2000 employees	0.471 (0.014)	0.449 (0.097)	0.406 (0.042)
> 2000 employees	0.425 (0.013)	0.377 (0.093)	0.359 (0.041)
Continuing contract	0.318 (0.021)	0.339 (0.162)	0.140 (0.066)
No chronic condition	−0.308 (0.010)	−0.325 (0.081)	−0.189 (0.037)
No handicap	−0.482 (0.010)	−0.504 (0.080)	−0.378 (0.036)
30 ≤ age < 45	−0.103 (0.012)	−0.089 (0.099)	−0.261 (0.038)
Age ≥ 45	−0.071 (0.013)	−0.015 (0.100)	−0.454 (0.042)
5 ≤ tenure < 15	0.087 (0.015)	0.006 (0.112)	0.037 (0.049)
Tenure ≥ 45	0.171 (0.012)	0.095 (0.083)	−0.013 (0.043)
Satisfied with job	−0.032 (0.002)	−0.040 (0.016)	−0.064 (0.006)
Constant	3.132 (0.041)	3.478 (0.325)	−0.768 (0.139)
Variance parameter			
$\sigma_u^2 = 1/\alpha$	–	5.274 (0.108)	6.262 (1.708)
Scale parameter τ	–	–	201.1 (65.8)
Observations	7373	7373	7373
Log–likelihood	−71665.6	−17994.9	−17892.1

Notes: There are 4014 zeros in the sample. The Poisson model predicts 10 zeros; The negative binomial model predicts 3752 zeros. Random Effects Negative Binomial Regression: Unbalanced panel has 2432 individuals.

Table 3. *Marginal effects*

	Poisson	Negbin
Log wage	−2.715	−3.307
Public sector	2.069	2.440
Blue collar	3.705	4.067
Married	1.291	1.567
Children at home	−0.176	−0.180
20–200 employees	1.340	1.132
201–2000 employees	3.103	2.922
> 2000 employees	2.805	2.451
Continuing contract	2.098	2.205
No chronic condition	−2.033	−2.113
No handicap	−3.176	−3.280
30 < age < 45	−0.685	−0.580
Age > 45	−0.473	−0.101
5 ≤ tenure < 15	0.577	0.042
Tenure ≥ 45	1.129	0.623
Satisfied with job	−0.215	−0.266

Notes: Partial derivatives of expected value with respect to the vector of characteristics. The partial derivatives are computed at the means of the regressors. The option for computing marginal effects is not implemented in the LIMDEP 7.0 routine for the random effects negative binomial model.

controlling for wage effects. Taken together, the indirect wage effect and the direct size effect suggest that workers in larger firms have about 1.6 more absent days per year than workers in smaller firms. To put it differently, while larger firms pay higher wages than smaller firms, possibly in order to reduce their inherently higher absenteeism, this effect is negligible compared to the large direct, positive effect of firm size on absence rates. This result constitutes evidence against the pure adjustment model and for the shirking model where workers' uncertain prospects of being caught are a decreasing function of firm size.

ACKNOWLEDGEMENTS

Tim Barmby and one anonymous referee provided very valuable comments.

REFERENCES

Barmby, T.A., Orme, C.D. and Treble, J.G. (1991) Worker absenteeism: an analysis using micro data, *Economic Journal*, **101,** 214–29.

Barmby, T.A., Sessions, J. and Treble, J.G. (1993) Absenteeism, efficiency wages, and shirking, *Scandinavian Journal of Economics* (forthcoming).

Barmby, T.A., Orme, C.D. and Treble, J.G. (1995) Worker absence histories: a panel data study, *Labour Economics*, **2,** 53–66.

Delgado, M.A. and Kniesner, T.J. (1996) Count data models with variance of unknown form: an application to a hedonic model of worker absenteeism, *Review of Economics and Statistics*.

Dunn, L.F. and Youngblood, S.A. (1986) Absenteeism as a mechanism for approaching an optimal labor market equilibrium: An empirical study, *Review of Economics and Statistics*, **68,** 668–74.

Hausman, J., Hall, B.H. and Griliches, Z. (1984) Econometric models for count data with an application to the patents – R&D relationship, *Econometrica*, **52,** 909–38.

Johansson, P. and Palme, M. (1996) Do economic incentives affect work absence? Empirical evidence using Swedish micro data, *Journal of Public Economics*, **59**(2), 195–218.

Johansson, P. and Brännäs, (1996) A household model for work absence, *Journal of Applied Econometrics* (forthcoming).

Wagner, G.G., Burkhauser, R.V. and Behringer, F. (1993) The English language public use file of the German socio-economic panel, *Journal of Human Resources*, **28,** 429–33.

Winkelmann, R. (1996) Training, earnings and mobility in Germany, *Konjunkturpolitik*, **42**(4), 275–98.

Winkelmann, R. (1997) *Econometric Analysis of Court Data* 2nd rev. & enl. edn, Springer: Heidelberg, New York.

Winkelmann, R. (1995) Duration dependence and dispersion in count data models, *Journal of Business and Economic Statistics*, **13,** 467–74.

Winkelmann, R. and Zimmermann, K.F. (1995) Recent developments in count data modeling: theory and application, *Journal of Economic Surveys*, **9,** 1–24.