TI 2009-014/3
Tinbergen Institute Discussion Paper

Are Workers with A Long Commute Less Productive? An Empirical Analysis of Absenteeism

*Jos van Ommeren**
Eva Gutièrrez-i-Puigarnau

Faculty of Economics & Business Administration, VU University Amsterdam.

** Tinbergen Institute.*

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam, and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Roetersstraat 31
1018 WB Amsterdam
The Netherlands
Tel.: +31(0)20 551 3500
Fax: +31(0)20 551 3555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Most TI discussion papers can be downloaded at http://www.tinbergen.nl.

Are workers with a long commute less productive?

An empirical analysis of absenteeism

Jos van Ommeren

Eva Gutiérrez-i-Puigarnau

VU University, Amsterdam, The Netherlands

03-02-2009

VU University, FEWEB, De Boelelaan, 1081 HV Amsterdam, the Netherlands; Email: jommeren@feweb.vu.nl
We would like to thank Marcel Hoogzaad for valuable assistance and the Frisch Centre, Oslo, Norway for hospitality.

Abstract. We hypothesize, and test for, a negative effect of the length of the commute on worker's productivity, by examining whether the commute has a positive effect on worker's absenteeism. Our estimates for Germany indicate that commuting distance induces absenteeism with an elasticity of about 0.07. On average, absenteeism would be about 16 percent less if all workers would have a negligible commute. These results are consistent with urban efficiency wage models.

Keywords: absenteeism, commuting, productivity. JEL code R23; J22; J24

1. Introduction

In a number of recent papers, Zenou and co-authors have argued that urban efficiency wage models imply that shirking and the length of the commute are positively related (Zenou and Smith, 1995; Zenou, 2002; Brueckner and Zenou, 2003; Ross and Zenou, 2008; Zenou, 2008).[1] The main implication is that workers with longer commutes are less productive. This result contrasts with a large literature in labour and urban economics which assumes that the productivity of workers is independent of the commute.[2]

The urban efficiency wage literature incorporates commuting costs in a shirking model setting and assumes that workers' work effort depends negatively on commuting costs. For example, it is argued that a longer commuting time may induce workers to arrive late at work, or leave earlier, which reduces productivity. These reductions in productivity may be observed by an employer, and the employer may take this into account when hiring (or firing) a certain worker. This situation becomes more complicated however, when a worker with a

[1] In this literature, the focus is on shirking, and frequently, it is assumed that workers either shirk of do not shirk. This implies that, in equilibrium, employers set wages such that workers never shirk. In a more realistic setting with a continuous shirking decision, workers shirk *more* intensive given a longer commute.

[2] Arguably, one may dismiss the importance of this result. In particular, if wages are downward adjusted when commutes increase, so employers are then neutral regarding the residence location of workers. As there is no evidence that wages negatively related to distance is the case (in fact, all evidence supports a positive relationship between wages and the length of the commute), this implies that workers' net productivity – the productivity minus the wage – is a negative function of workers' commuting costs. One implication may be then that firms will redline workers with a long commute (Zenou, 2002).

long commute may become less productive *by reducing effort levels,* so the worker *shirks* and shirking cannot be observed by the employer (or is costly to monitor). One of the main results of the urban efficiency wage model is then that, in equilibrium, the length of the commute negatively affects productivity of workers through shirking. As far as we know, there are no empirical tests of the underlying, but fundamental, assumption that the length of the commute makes workers less productive. One obvious measure of worker's productivity is to use absenteeism, the number of days absent from work for sickness reasons during a certain period.[3] Hence, an empirical test of the assumptions of the urban efficiency wage model is that the length of the commute positively affects the number of days absent.

One may argue that the analogy between workers' 'absenteeism' and 'effort', which we will use in our paper, is misleading, because an essential assumption of the efficiency wage model is that the workers' effort level and therefore productivity cannot be observed (at zero costs with probability one).[4] In contrast, absenteeism, our measure of not being productive, is fully observed by an employer. So, one may argue that an analysis of absenteeism cannot be interpreted as a test of the underlying assumptions. Our response to this –potentially valid– criticism is that in many countries, including Germany, the country we focus on, employers are not allowed by law to fully reduce wages in accordance to the number of days absent. In Germany, the law is even stricter than in other European countries: workers who are absent for less than six weeks (30 uninterrupted working days) keep the same wage. These long durations of absenteeism occur only infrequently.[5] More fundamentally, employers cannot observe whether a worker is really sick (involuntary

[3] One exception is the study by Allen (1981), which uses a cross-section data set. The use of a cross-section data set is however highly problematic. One expects, and this is later on confirmed by our (panel data) analysis, that a cross-section analysis generates negatively biased estimates of the effect of distance on absenteeism. This is likely so because individuals who derive a higher utility of employment are more likely to accept long commuting distances and are less likely to shirk.

[4] In efficiency wage models, it is frequently assumed that monitoring is costless, but shirking is detected by the employer with a positive probability smaller than one.

[5] In our data, only 3% of all workers are absent uninterruptedly for six weeks.

absenteeism) or shirks (voluntary absenteeism). In other words, shirking through voluntary absenteeism cannot be observed at zero costs.

Let us suppose, for example, that workers choose the optimal number of days that they aim to shirk, by being absent at work (consistent with the seminal paper by Ehrenberg, 1970). Shirking has its costs for the worker (e.g. the worker may be caught shirking and be fired, the worker may miss a promotion when the number of days absent exceeds a certain, maybe unknown, threshold) and its benefits. The costs of shirking are independent of the length of the commute. The benefits consist of two parts: the increase in leisure time due to the reduction in the duration at work (e.g. 8 hours for each day reported absent) as well as an increase in leisure time as the worker saves on travel time, which depends on the length of the commute.[6] Hence, the worker's benefits of shirking positively depend on the length of the commute.

In the current paper, we use absenteeism as a measure of shirking. Arguably, there exist *two* explanations why the length of the commute may increase absenteeism. The first explanation is that the benefit of an additional day absent is an increasing function of the length of the commute because workers not only gain in leisure time while being absent, but workers with a longer commute also enjoy a larger reduction in commuting time. This explanation is consistent with voluntary absenteeism and therefore shirking. The second explanation is that the workers' length of the commute decreases the workers' health which induces absenteeism. For example, according to Koslowsky et al. (1995), long commutes cause a lot of stress. This explanation is consistent with involuntary absenteeism.[7] By controlling for a number of subjective and objective health indicators, we aim to identify the effect of distance on voluntary absenteeism.

[6] The worker also saves monetary costs by not travelling. Note however that the marginal monetary costs of commuting one day are small, as a large part of the monetary costs are fixed (e.g. purchase of car, rail discount cards, etc).
[7] This form of absenteeism may also be labelled as voluntary, when workers voluntary choose a long commuting distance and realize that they will be more absent.

One statistical issue we address in the current paper is that in the survey analysed here (as is usual) absenteeism refers to the number of days absent during a predefined *period*, whereas commuting distance is reported only at the beginning and the end of the period. The effect of distance on absenteeism will be shown to be (downward) biased if this issue is not accounted for.[8] Another issue we address is commuting distance may be endogenously chosen with respect to absenteeism. We solve this endogeneity issue by using a worker fixed-effects approach where changes in commuting distance are employer-induced, and therefore exogenous.

2. Empirical approach

2.1 Data and method

In the current study, we use seven waves of the 1999–2007 German Socio-Economic Panel (GSOEP) survey (which includes about 48,000 observations) to study the effect of the commute on absenteeism.[9] Absenteeism refers to the number of days absent during the year before the interview date. Commuting distance, our other main explanatory variable of interest, is measured at the interview date.[10] In case that the commuting distance has changed during the year before the interview, then the commuting distance reported will only be applicable for a *part* of the year for which absenteeism is reported. This will induce a bias in

[8] We address and solve this statistical issue by using a sample of observations of workers who did not change commuting distance during the previous period (e.g. last year). This guarantees that worker's commuting distance and absenteeism are consistently measured.

[9] For one wave only, the GSOEP also contains information about commuting *time*. Because commuting time is endogenously chosen by choosing the optimal mode of transport (see e.g. Van Ommeren and Fosgerau, 2009), our preferred measure is distance.

[10] For the first two waves, information about commuting distance is only available if the workplace municipality differs from the residence municipality, so the exact commuting distance is unknown for workers who commute to a workplace location within the residence municipality. This is unproblematic as the distances of workers who live and work in the same municipality do not vary much. Hence, for the first three waves, we have imputed a value of 5 km for workers who live and work in the same municipality. A sensitivity analysis shows that the results presented later on are insensitive to the imputed value (e.g. 0 or 6 km). This makes sense as the imputation refers to only 15% of the observations, and the difference between the (unobserved) distance and the imputed distance is small (less than 10% of the mean commuting distance).

the estimates. The magnitude of the bias depends on the type of estimation method used.[11] This can be easily understood by comparing standard cross-section regression and fixed-effects panel data analyses.

Let us suppose now that commuting distance is exogenous, there is no correlation between distance and the error term, and the worker's annual change in commuting distance is random. The estimated effect of commuting distance (reported at the interview) on absenteeism is then downward biased when standard regression analysis is used, because it is ignored that the reported distance is only applicable for a part of the observation period to which absenteeism refers, so commuting distance is measured with random error (see Verbeek, 2003). Yet, as for most workers the commuting distance does not change during the year before the interview, this bias will be (negligibly) small. Unfortunately, standard regression analysis is not the preferred estimation method due to unobserved worker heterogeneity (Dionne and Dostie, 2007).[12] In this literature, it is argued to use worker fixed-effects.

Using a fixed-effects method, the effect of commuting distance on absenteeism is identified using the worker's *change* in reported commuting distance on the change in absenteeism. In this case, the change in reported commuting distance has a *systematic* measurement error, so the bias in the estimates may not be so small. In Appendix A, we show that the bias in the estimated effect of distance is indeed substantial, and given (reasonable) assumptions, is equal to (approximately) 50%.

In the current paper, we will therefore proceed by using worker's fixed-effects estimation approaches and we will remove the systematic measurement-error bias identified

[11] The type of bias discussed in the current paper is not only relevant in the case of absenteeism, but is generally a problem when the dependent variable is measured over a period and explanatory variables change within this period. Examples which come to mind are annual labour supply and annual income.

[12] For example, it seems plausible that workers who have a higher value of leisure time (which is not observed) will only accept jobs with a short commute and are also more likely to report that they are sick and therefore absent.

above by only selecting observations of workers for which it is known that the commuting distance did not change during the year before the interview date. We are then left with 16,762 annual observations for 7,104 employees. Note that a worker's fixed-effects estimation approach is based only on annual changes in the worker's dependent and independent variables over time.[13] As we exclude observations of workers for which the commuting distance changes between the previous and current year of observation, the effect of commuting distance on absenteeism is identified using observations of workers that are at least two years apart.[14]

Let us first focus on the dependent variable of interest: the number of days absent. The mean number of days absent is 7.85, with a standard deviation of 19.14.[15] For about 43% of the observations, the worker has not been absent at all during the whole year. The number of days absent is a *count variable* (0, 1, 2, 3, etc), and we therefore estimate worker's fixed-effects negative binomial regression models. We include a large number of (time-varying) explanatory variables including commuting distance, year dummies, weekly working hours, presence of children, wage, region dummies, firm size and industry dummies (see also Barmby et al., 1991; Barmby, 2002, and Barmby et al., 2002).

Furthermore, we control for a number of subjective and objective health indicators. By controlling for health, we control for involuntary absence due to sickness. These indicators include a self-reported description of current health (very good, good, satisfactory, poor, very bad), number of trips to the doctor in the last three months before the interview date, as well as number of nights admitted to a hospital in the year before the interview. For ease of interpretation, we have annualised the doctor trip data.

[13] Within-worker variation in absenteeism is quite large and refers to 35% of all variation.
[14] For example, if the original sample includes observations for workers in the years $t-1$, t and $t+1$, and the commuting distance has changed between $t-1$ and t, then we only include the observations for the years $t-1$ and $t+1$.
[15] For the original sample of 48,000 observations and the selective sample of 16,762 observations, the sample descriptives are almost identical, demonstrating that sample selection is not an issue. Note that in the multivariate analysis later on we include worker fixed-effects, so worker-specific time-invariant selection effects are controlled for.

For the negative binomial model, for each worker i holds that $\log E(A_{it}) = \beta X_{it} + \alpha_i$,

where $\log E(A_{it})$ denotes the logarithm of the expected number of days absent, and β, α_i are parameters to be estimated. We report the coefficients for β, which refer to the effects of regressors on the logarithm of the expected number of days absent.

3.2 Empirical results: worker fixed-effects

As stated in the introduction, our main interest is to estimate the effect of commuting distance on absenteeism. We have experimented with several functional forms for commuting distance. The main results are hardly sensitive to the exact form chosen, because the elasticities of distance (evaluated at the mean distance) are close to each other.[16] In the current paper, we report log-linear specifications of commuting distance. As can be seen from Table 1, the effect of commuting distance on the number of days absent is positive and statistically significant (at the 1% level). The point estimate, and therefore the elasticity, is 0.077 (s.e. is 0.0014). This indicates, for example, that the (expected) number of days absent is 12% higher for workers with a (one-way) commuting distance of 40 km for those workers with a distance of 10 km.

To understand the magnitude of the effect, let us focus now on a hypothetical firm that actively starts to redline all workers who do not live within 1 km of the workplace, such that after a certain time all workers will live at about 1 km from the firm.[17] In this case, the average logarithm of commuting distance falls from 2.12 to about 0 km and absenteeism within this firm will fall by about 16% (0.077x2.12), so, on average, by 1.28 working days per year. Clearly, the results are not only statistically but also economically significant.

[16] For example, we have estimated a model using a quadratic distance specification (so we use distance as well as distance squared). The estimated elasticity is identical to the ones reported here. Furthermore, for this specification, the marginal effect of distance is decreasing in distance in line with the log-linear specification.
[17] Note that such a recruitment rule is only hypothetical as it is counterproductive to the firm, because it strongly reduces the supply of workers.

3.3 Worker fixed effects and employer-induced changes in commuting distance

By estimating worker fixed-effects models, we have avoided bias in the estimates related to *time-invariant* heterogeneity. However, one may argue that *time-varying* worker heterogeneity potentially biases our results, so the estimated effect of commuting distance may be affected by omitted-variable bias (see, similarly, MaCurdy, 1981; Altonji, 1986; Lee, 2001).[18] We therefore go one step further and use only employer-induced changes in commuting distance. Note that a change in the worker's commuting distance may be either due to a residence move, a job move or due to a workplace relocation (a workplace location move while staying with the same employer). The change in commuting distance as a result of the latter type of move is most likely exogenous.[19] This is particularly the case if the firm moves the whole establishment to another location, as the firm ignores idiosyncratic preferences of workers (see e.g. Zax, 1991).[20]

Formally, using only employer-induced changes in commuting distance is the same as estimating a group fixed-effects model, where a group is defined such that each group only includes observations of *one* worker who does not move residence or employer. Hence, observations of the same worker belong to different groups if the worker moves residence or employer. In our data set, we have 7,104 workers and 563 (residential or job) moves. So, the total number of groups is 7,667 (7,104 + 563). The results of the model are shown in Table 2. The results are essentially unaltered (compared to Table 1). The point estimate of log distance

[18] For example, one may argue that a worker who expects a deterioration in health (and therefore an increase in absenteeism), may decide to move residence closer to the workplace.

[19] Firm relocation as a source of exogenous change in commuting distance is quite common (see e.g. Zax, 1991; Zax and Kain, 1996). For example, about 7–8% of firms in the Netherlands are each year involved in relocation decisions (Weltevreden et al., 2007). In Great Britain, in each year 0.5% of workers state that they change residence because of an employer-induced workplace move, suggesting that workplace moves are quite important (National Statistics, 2002). Note that in the survey analysed here, there is no information whether firms move. However, by keeping employer and residence given, we infer that all changes in commuting distance are caused by a (exogenous) change in commuting distance as a result of a relocation of the workplace by the firm. We estimate that, in our sample, about 10% of changes in commuting distance are employer-induced.

[20] Note that Ose (2005) estimates the effect of firm relocations on absenteeism, whereas our study focuses on the effect of the change in the commuting distance induced by workplace relocation as the effect of interest.

is slightly higher (0.0742 instead of 0.0641). More formally, we have applied a Hausman *t*-test, which tests whether the estimated coefficient of log distance in Table 2 is statistically different from the one reported in Table 1. This test is valid given the assumption that the estimator reported in Table 2 is consistent but less efficient than the one reported in Table 1. For details, see Wooldridge, 2002, p. 290. We find that the *t*-statistic is equal to 1.15, far below 1.96, indicating that the estimated effect of log distance cannot be rejected against the alternative estimate reported in Table 2. Apparently, in this context, time-varying unobserved heterogeneity does *not* affect the consistency of the estimated effect of distance.

3.4 Sensitivity analyses

We have subjected the results to a number of sensitivity analyses. For example, we have estimated the same models as discussed above for males and females separately as the effect of distance may be gender-specific (see Vistnes, 1997). We find that the estimated effects for distance are almost exactly the same for males and females, indicating that the distance-effect identified is not gender specific. Further, we have examined the effect of possible 'outliers' of the dependent variable. This may be important for two reasons. First, a well-known feature of count models is that estimates are *not* consistent given random measurement error in the *dependent* variable (Winkelman, 2003). One can imagine that measurement error may be particularly large for workers with a large number of days absent. Second, workers who are absent for more than 30 working days may receive a wage reduction (this applies to about 3% of the workers). Hence, we have estimated the same models selecting only observations for which absenteeism is less than 20 days. For this sample, which contains about 95% of the original sample, measurement error in the number of days absent is strongly minimised and none of the workers have received a wage reduction because of long absenteeism. The results are almost identical to the ones reported above. This indicates that the results are robust, and not due to a few outliers, and that unobserved wage reductions due to long absenteeism do not affect our estimates.

Furthermore, we have estimated the same models *not* controlling for subjective and objective health indicators (see last columns of Tables 1 and 2). Essentially, the results remain unaltered, although the results demonstrate that the point estimates not controlling for health are slightly higher.[21]

We have also analysed the effect of *interactions* of distance with health indicators. This is relevant as one may imagine that unhealthy workers or for workers that visit doctors more frequently, the marginal costs of the commute are higher. We do *not* find any evidence that the interactions of distance with health indicators have an effect on absenteeism. This strongly suggests that the workers' marginal costs of commuting do not depend on the workers' health.

We have also estimated the negative binomial model *without* worker fixed effects. We find a much lower estimate of distance (0.0227) which is even statistically insignificant at a common significance level of 5% (s.e. is 0.0134). Hence, cross-section estimation of the effect of commuting on absenteeism negatively biases the results. The most plausible explanation for the bias is that workers with unobserved positive attitudes to work are more likely to accept jobs at long distances and are also less likely to be absent. Fixed-effects estimators address this issue.

Finally, recall that we have estimated models on a selective sample of workers to avoid a bias that may occur as absenteeism is measured over a period, whereas distance is measured *at* a point in time. To see the importance of this selection, we have also estimated models on the full sample. For this sample the coefficient of commuting distance is indeed

[21] In fact, controlling for any other time-varying variable does not appear to be essential for the estimated effect of distance. As an aside, we have also investigated the effect of distance on the full range of health indicators. For all health indicators, given worker's fixed effects, we find a positive, but statistically insignificant effect of distance, indicating that a longer commute does *not* lead to a deterioration of health. This is in contrast to claims by Koslowsky et al. (1995).

about 30 to 60% lower than the ones reported here (the exact percentage depends on the specification of distance), consistent with our theoretical claim that the bias is about 50% (see Appendix A).

4. Conclusion

A common assumption in the labour and urban economics literature is that private costs of commuting are fully borne by the worker and do not affect the worker's productivity. This assumption is challenged by Zenou and co-authors who assume that worker's work effort is negatively affected by the length of the commute. We are not aware of empirical tests of this assumption. In the current paper, we focus on the relationship between commuting distance and absenteeism. Our results indicate that commuting distance has a strong positive effect on absenteeism, with an elasticity of about 0.07. In the hypothetical case that all workers in the economy have a negligible commute, absenteeism would be about 16% lower, roughly one day per year. This implies that worker's productivity is negatively affected by the length of the commute, in line with the theoretical studies by Zenou and co-authors.

In the current paper, we have emphasised the importance of the econometric specification of the absenteeism model to be estimated. In particular, it seems fundamental to address (time-invariant) unobserved worker heterogeneity, which is standard in the panel data literature, but also to address the issue that absenteeism is measured over a period, whereas commuting is measured at a certain period in time. When this technical detail is ignored, fixed-effects estimation generates a downward bias of about 50%. Time-varying unobserved heterogeneity, which we address by examining changes in commuting distance induced by the employer, is shown to be a less relevant issue.

The consequences of our empirical results for our understanding of the urban labour market will mostly depend on how worker's wages vary with absenteeism, and how worker's wages depend on the commuting distance. Given the stylised facts that for few jobs wages are fully reduced in accordance with the number of days absent and that wages are not a negative function of distance, employers will have an incentive not to hire workers with a long commute, see Zenou (2002).[22]

[22] The redlining result by Zenou should maybe be interpreted more broadly than having only implications for employer recruitment. Urban efficiency wage models that also allow for random shocks to worker's productivity predict that *involuntary* job moves are a positive function of commuting distance. Our suggestion is that such a hypothesis is interesting to test.

References

Allen, S.G. (1981), An empirical model of work attendance, *Review of Economics and Statistics*, 63, 1, 77–87

Altonji, J.G. (1986), Intertemporal substitution in labor supply: evidence from micro data, *The Journal of Political Economy*, 94, 3, 176–215

Barmby, T. (2002), Worker absenteeism: a discrete hazard model with bivariate heterogeneity, *Labour Economics*, 9, 4, 469–476

Barmby, T., M. Ercolani and J.G. Treble (2002), Sickness absence: an international comparison, *Economic Journal*, 112, 480, F315–F331

Barmby, T., C. Orme and J.G. Treble (1991), Worker absenteeism: an analysis using microdata, *Economic Journal*, 101, 405, 214–229

Brueckner, J.K. and Y. Zenou (2003), Space and unemployment: the labor-market effects of spatial mismatch, *Journal of Labor Economics*, 21, 1, 242–266

Dionne, G. and B. Dostie (2007), New evidence on the determinants of absenteeism using linked employer-employee data, *Industrial and Labor Relations Review*, 61, 1, 108–120

Ehrenberg, R.G. (1970), Absenteeism and the overtime decision, *American Economic Review*, 60, 3, 352–357

Koslowsky, M., A.N. Kluger and M. Reich (1995), Commuting stress: causes, effects and methods of coping, New York: Plenum Press

Lee, C. (2001), Finite sample bias in IV estimation of intertemporal labor supply models: is the intertemporal substitution elasticity really small?, *The Review of Economics and Statistics*, 83, 4, 638–646

MaCurdy, T.E. (1981), An empirical model of labor supply in a life-cycle setting, *The Journal of Political Economy*, 89, 6, 1059–1085

National Statistics (2002), *Labour Force Survey LFS 2002*, Newport, Great Britain

Ose, S.O. (2005), Working conditions, compensation and absenteeism, *Journal of Health Economics*, 24, 1, 161–188

Ross, S.L. and Y. Zenou (2008), Are shirking and leisure substitutable? An empirical test of efficiency wages based on urban economics theory, *Regional Science and Urban Economics*, 38, 5, 498–517

Verbeek, M. (2003) *Modern Econometrics*, Second edition, Wiley

Vistnes, J.P. (1997), Gender differences in days lost from work due to illness, *Industrial and Labor Relations Review*, 50, 2, 304–323

Van Ommeren, J.N. and M. Fosgerau (2009), Workers' marginal costs of commuting, *Journal of Urban Economics*, 65, 1, 38–47

Weltevreden, J.W.J., F.G. van Oort, J. van Vliet, P.H. Pellenbarg, H. van Amsterdam and M.R.M.J. Traa (2007), Firm relocation and regional employment development in the Netherlands (1999-2006), European Regional Science Association

Winkelman, R. (2003), *Econometric Analysis of Count Data*, Fourth edition, Springer

Wooldridge, J.M. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT press, Cambridge, Massachusetts

Zax, J.S. (1991), The substitution between moves and quits, *The Economic Journal*, 101, 409, 1510–1521

Zax, J.S. and J.F. Kain (1996), Moving to the suburbs: do relocating companies leave their black employees behind?, *Journal of Labor Economics*, 14, 3, 472–504

Zenou, Y. (2002), How do firms redline workers?, *Journal of Urban Economics*, 52, 3, 391–408

Zenou, Y. (2008), Urban Labor Economic Theory, in *A Companion to Urban Economics*, eds. Arnott, R. and D.P. McMillen, Blackwell Publishing

Zenou, Y. and T.E. Smith (1995), Efficiency wages, involuntary unemployment and urban spatial structure, *Regional Science and Urban Economics*, 25, 4, 821–845

Table 1. Estimates of *Worker* Fixed-Effects Negative Binomial Models: Number of Days Absent

log (distance)	0.0641	(0.0122)	0.0773	(0.0119)
hours per week	0.0018	(0.0014)	–0.0010	(0.0014)
number of children	0.0515	(0.0151)	0.0406	(0.0153)
wage (in log)	0.3082	(0.0372)	0.2682	(0.0359)
very good health				
good health	0.1650	(0.0492)		
satisfactory health	0.3596	(0.0508)		
poor health	0.4108	(0.0565)		
very bad health	0.5545	(0.0992)		
doctor visits	0.0068	(0.0003)		
hospital visits	0.7258	(0.0083)		
industry dummies (6)	included	included	included	included
region dummies (5)	included	included	included	included
year dummies (8)	included	included	included	included
firm size dummies (6)	included	included	included	included
number of observations	16,762		16,762	
number of workers	7,104		7,104	

Note: standard errors in parentheses.

Table 2. Estimates of *Group* Fixed-Effects Negative Binomial Models: Number of Days Absent

log (distance)	0.0742	(0.0127)	0.0875	(0.0124)
hours per week	0.0014	(0.0014)	–0.0012	(0.0014)
number of children	0.0612	(0.0157)	0.0426	(0.0157)
wage (in log)	0.2620	(0.0390)	0.2482	(0.0369)
very good health				
good health	0.1508	(0.0508)		
satisfactory health	0.3402	(0.0523)		
poor health	0.3592	(0.0579)		
very bad health	0.4096	(0.1059)		
doctor visits	0.0063	(0.0003)		
hospital visits	0.8261	(0.0113)		
industry dummies (6)	included	included	included	included
region dummies (5)	included	included	included	included
year dummies (8)	included	included	included	included
firm size dummies (6)	included	included	included	included
number of observations	16,762		16,762	
number of groups	7,667		7,667	

Note: standard errors in parentheses. Groups are defined (in section 3.3) such that a group includes only observations of the one worker with the same residence and employer.

Appendix A: Bias in the estimate of the effect of commuting distance

We denote $A_{t,t-1}$ as the worker's number of days absent between time $t-1$ and t.[23] Between $t-1$ and t, the worker may have, for a certain time, a commuting distance of d_t and for the rest of the time a distance of d_{t-1}. The distance $D_{t,,t-1}$ denotes the weighted average of both distances d_t and d_{t-1}, so:

$$D_{t,t-1} = \lambda_t d_t + (1-\lambda_t) d_{t-1}, \tag{A1}$$

where $0 \leq \lambda_t \leq 1$. The weighting variable λ_t measures the proportion of time between $t-1$ and t that the worker's distance is d_t, so $1-\lambda_t$ measures the proportion of time that the worker's distance is equal to d_{t-1}. For example, in case that the worker does not change distance between $t-1$ and t, then $\lambda_t = 1$, so $D_{t,t-1} = d_t$.

We denote the absenteeism between $t-1$ and t as $A_{t,t-1}$. The true relationship between $A_{t,t-1}$ and average distance $D_{t,,t-1}$, is assumed to be linear, so:

$$A_{t,t-1} = \alpha + \beta D_{t,t-1} + u_t, \tag{A2}$$

where α, β are parameters to be estimated and u_t is random error. For convenience, suppose that one observes for each worker absenteeism A exactly two times in a row (at t and $t-1$), so one observes $A_{t,t-1}$ and $A_{t-1,t-2}$.

Now suppose that the worker's commuting distance has changed between $t-1$ and t (e.g. due to a residence move), but the distance has remained the same between $t-2$ and $t-1$, so $D_{t-1,t-2} = d_{t-1}$. A fixed-effects estimator which removes the fixed effect α, is then based on the following expression:

[23] For convenience, we ignore subscripts related to worker i.

$$A_{t,t-1} - A_{t-1,t-2} = \lambda_t \beta (d_t - d_{t-1}) + u_t - u_{t-1}. \tag{A3}$$

It follows that the *estimated* value of β, $\hat{\beta} = \lambda_t \beta$, so $\hat{\beta} \leq \beta$. The value of λ_t will differ per observation, but one may make assumptions about its distribution. For example, given the absence of seasonal variation in (residential/job) moving behaviour, λ_t will have a uniform distribution on the interval [0, 1], so the expected value of λ_t is 0.5. It follows that $\hat{\beta} = 0.5\beta$.

We have analysed above the situation that the commuting distance has changed between $t-1$ and t, but remained constant between $t-1$ and $t-2$. For the reversed situation (no change between $t-1$ and t, but a change between $t-1$ and $t-2$), it can be shown that $\hat{\beta} = (1 - \lambda_t)\beta$. So, if λ_t has a uniform distribution, we obtain the same result: $\hat{\beta} = 0.5\beta$. In case that a worker does not change distance between $t-2$ and t, the observation of the worker does not add information to the identification of β, so this case can be ignored. Finally, in the case of changes in commuting during both intervals, so between $t-2$ and $t-1$, as well as $t-1$ and t, it appears that:

$$A_{t,t-1} - A_{t-1,t-2} = \beta\left[\lambda_t (d_t - d_{t-1}) + (1-\lambda_{t-1})(d_{t-1} - d_{t-2})\right]. \tag{A4}$$

Only if $d_t - d_{t-1}$ is *exactly* equal to $d_{t-1} - d_{t-2}$, which is unlikely to happen, then there will be no bias in the estimates. On the other hand, suppose that $d_{t-1} - d_{t-2}$ is a random draw from a given, but arbitrarily chosen random distribution. In this case, it can be established that $\hat{\beta} < \lambda_t \beta$ (because $d_{t-1} - d_{t-2}$ can be treated as random measurement error). Note that the latter case does not occur frequently, so will have little effect on the overall bias.

In conclusion, it can be shown that $\hat{\beta} < \beta$, whereas, given reasonable assumptions, $\hat{\beta} = 0.5\beta$, so the magnitude of the bias will be about 50%.