Centre for Transport Studies
STOCKHOLM

The Impact of Accessibility on Labor Earnings

Gunnar Isacsson
Maria Börjesson
Matts Andersson
Christer Anderstig

CTS Working Paper 2015:18

Abstract
We estimate the impact of job accessibility on wage earnings using micro-level data including all workers in the greater Stockholm region at two points in time 11 years apart. We control for both zone-specific and individual-specific fixed effects by separating workers who have changed zone of residence and those who have stayed. The accessibility is derived from the national transport model, taking into account consumer behavior and preferences for all travel modes and travel time components. A novel instrumental variable based on the temporal changes in job accessibility resulting from transport system improvements over the 11 years is applied. The elasticity of accessibility defined from the worker's place of residence is estimated at 0.007. The elasticity of wage earnings with respect to job accessibility at the work place is only significant for workers moving work place and for those estimated at 0.036.

Keywords: Cost-Benefit Analysis, Accessibility, Agglomeration effects, Wider Economic Benefits

JEL Codes: R41, R42, R48, R12

Acknowledgments: Funding from the National Transport Administration and VINNOVA is gratefully acknowledged. We are indebted to Jos van Ommeren, Ismir Mulalic, Stef Proost and Jonas Eliasson for discussions, insights and helpful comments. We are also thankful to Christian Nilsson for extensive help with the data from SAMPERS. Any shortcomings are, of course, our own responsibility.

Centre for Transport Studies
SE-100 44 Stockholm
Sweden
www.cts.kth.se

1 INTRODUCTION

A positive statistical relationship between some measure of labor market accessibility and productivity is now well established in a growing body of research (See Rosenthal and Strange (2004), Melo et al. (2013) and Combes and Gobillon (2014) for reviews). Most empirical studies have used either employment density (e.g. Ciccone and Hall 1996), or market potential based on the Euclidean distances to existing work places (e.g. Graham and van Dender (2011), Mion and Naticchioni (2005), Rice et al. (2006), and Combes et al. (2008)). These measures, however, neglect the fact that accessibility is a function of the transport infrastructure and the generalized travel costs of different modes.[1] Measurement errors may therefore have been introduced into the estimations (Graham, 2007b). Further approximation errors might be introduced when assessing the impact of transport investments on the economy.

This paper estimates the impact of job accessibility on labor earnings in the greater Stockholm region. The job accessibility for all workers (age 20-64) is computed based on output from the national travel demand forecasting tool Sampers, including information on travel time, travel distance, travel cost and volume by mode for car, public transport, cycling and walking (Beser and Algers, 2002). Using historical transport networks, accessibility is computed for two years, 1995 and 2006. In addition, we use micro-level data from all workers and these two years. This data has high spatial resolution and includes detailed socio-economic characteristics including wage earnings for all individuals. It is linked to an establishment database including number of jobs in different sectors that is also coded with high spatial resolution.

We regress temporal changes in wage earnings on temporal changes in accessibility for each individual worker, controlling for observed socio-economic characteristics and fixed effects at the individual and zonal level. Socioeconomic variables control for spatial sorting on observable characteristics. The individual-specific fixed effect controls for any arbitrary time-invariant unobserved specific variables influencing wage earnings (e.g. ambition, ability and skills). The zone-specific fixed effect may arise from spatial differences in productivity related to differences in non-human endowments and local interaction across neighborhoods (Combes et al., 2008). To control for zone-specific fixed effect, we distinguish between workers who have changed zone of residence (movers) between the study years and those who have not (stayers).

[1] Rice et al. (2006) and Graham (2007b) use the car travel times implied by speed limits. We also note that Graham and Van Dender (2011, p. 412, footnote 4), argue that more accurate measures of transport accessibility would impose endogeneity problems because congestion tends to be higher in areas with higher economic activity. Travel distances are exogenous according to the same line of reasoning. However, if transportation costs are the relevant metric in this context, using distances to avoid endogeneity problems may instead introduce measurement error biases in the analysis.

Endogeneity, in the sense that higher local wages attract more workers and jobs, or that external shock simultaneously affects the number of jobs and wages, may introduce biases in our analysis. To reduce endogeneity problems we use an instrumental variable approach (IV) that isolates the effect of changes in the transportation infrastructure on changes in accessibility. Specifically, in our first stage equation, we regress the change in total accessibility on the change in accessibility resulting from changes in the transportation infrastructure only while keeping the spatial distribution of jobs constant. The long time lags between planning for and actual completion of changes in the transport infrastructure provide a motivation for our instrument being exogenous in the wage equation. In addition, this instrument variable is found to be strong by most standards (Stock et al., 2002).

Previous studies have used different types of instruments, but it has been hard to find relevant and exogenous instruments for agglomeration (Graham and Van Dender, 2011). Some authors have used lags on population density or total population (Ciccone and Hall, 1996; Combes et al., 2008; Mion, 2004; Mion and Naticchioni, 2005; Rice et al., 2006). Others have used instruments based on the total land area of the analyzed region (Ciccone, 2002; Combes et al., 2008) and geological features (Combes et al., 2010; Di Addario and Patacchini, 2008; Rosenthal and Strange, 2003). Most studies report small differences between IV results and the corresponding FE estimators (Graham et al., 2010).

This paper makes several contributions to the literature: i) we control for both zone- and individual-specific fixed effects, ii) the accessibility measure takes the transport system and travelers preferences into account, iii) we use a novel and relevant instrument, and iv) in a sensitivity analysis, we compare the effect of accessibility defined from the worker's place of residence on the wage earnings with the effect of the accessibility defined from the worker's work place, and *v*) we compute how much of the total increase in job accessibility over the 11 years that is attributable to the transport infrastructure improvements.

We find that the wage earnings elasticity with respect to job accessibility is just below 0.01. This is below the lower end of the interval found in the previous literature (reviewed above) 0.01-0.20, which might be due to the strong instrument reducing the elasticity substantially. We find no evidence for zone-specific effects from the place of residence (the elasticity does not differ between the samples of stayers and movers). Applying job accessibility at the work place results in an elasticity not significantly different from zero for workers who have not changed work place in the intervening 11 years. For workers who have changed work place, however, the elasticity 0.036 is obtained, suggesting that zone-specific fixed effects at the work place influence wages, and that this effect is positively correlated with accessibility change. An alternative interpretation of the higher elasticity for workers having changed work place is that dynamic agglomeration economies are more important than static.

The rest of this paper is organized as follows. In section 2, we present the empirical models and outline some of the identification problems and how we have tackled them. Data are presented in section 3, and section 4 contains the results with a discussion. Section 5 presents some concluding remarks.

2 THE MODEL

The basic idea of our model is that individuals' labor earnings are higher because of agglomeration economies. The underpinning of the theory is that job accessibility (our measure of agglomeration) increases labor productivity due to improved possibilities of sharing, matching and learning (Duranton and Puga, 2004).

2.1 The wage equation

Our measure of job accessibility is closely related to the concept of market potential (Harris (1954); Fujita et al. (1999); Hanson (2005)), and the accessibility measure derived in standard transport models, the logsum.[2] The job accessibility, A_{rt}, for residents in zone r ($r=1, 2, \ldots, R$), at time t ($t=1, 2$) is

$$A_{rt} = \sum_{r' \in \tilde{R}_r} exp\,(-0.028 gc_{t,r,r'})\, n_{t,r'}, \tag{1}$$

where $gc_{t,r,r'}$ is the average generalized commuting cost at time t between zone r and r'; $n_{t,r'}$ is the number of jobs (the number of employed individuals) located in zone r' at time t; and -0.028 is the scale parameter estimated in the travel demand model, measuring how sensitive commuters are to the generalized travel cost.

The average generalized commuting cost is

$$gc_{t,r,r'} = \sum_{m=1}^{4} w(m)_{t,r,r'} \left[c(m)_{t,r,r'} + d(m)_{t,r,r'} v(m)\right], \tag{2}$$

where $w(m)_{t,r,r'}$ is the share of commuters from r to r' in t using mode m ($\sum_{m=1}^{4} w(m)_{t,r,r'} = 1$); $c(m)_{t,r,r'} + d(m)_{t,r,r'} v(m)$ is the generalized cost of commuting with mode m, where $c(m)_{t,r,r'}$ is the pecuniary cost, $d(m)_{t,r,r'}$ is the travel time, and $v(m)$ is the value of time. The travel times for public transport are a weighted sum of in-vehicle-time, waiting time and access time. The value of waiting time is a piecewise linear function of the headway – the longer the headway, the lower its valuation per minute. We describe the implementation of the generalized cost of travel further in the data section.

The generalized travel costs $gc_{t,r,r'}$ for time 1995 are simulated by feeding the transport model the population, work places, and all other input parameters such as GDP and fuel prices at the 2006 level but the historical network of 1995. Hence, these generalized travel costs do not pick up changes in congestion levels that result from changes in the spatial distribution of jobs and workers. This also makes the accessibility measure more relevant in appraisal, where

[2] It is also related to the "effective density" (Graham (2007a), Graham (2007b) and Graham and van Dender (2011)).

typically two scenarios are compared, and where only the transport network differs between them.

Let the annual pre-tax wage earnings y_{irt} of worker *i* (*i=1, 2, …, I*) residing in zone *r* (*r=1, 2, …, R*) at time *t* be given by

$$lny_{irt} = lnA_{rt}\delta + X'_{it}\beta + \theta_i + \theta_r + \theta_t + \varepsilon_{irt}, \tag{3}$$

where A_{rt} is job accessibility of zone *r* at time *t*. The vector X'_{it} includes the individual characteristics: indicator variable for being male, age, age-squared indicator variables for educational attainment, number of children in different age classes, an indicator variable for marital status and two sets of dummy variables for sector and industry of employment. The control variables correct for related variations in labor demand and supply. θ_i is an individual-specific fixed effect capturing time-invariant unobserved productivity differences between individuals (e.g. ambition or skills); θ_r is a zonal-specific fixed effect capturing time-invariant unobserved productivity differences between residents of different zones arising from such factors as zone-specific non-human endowments and local interactions (Combes et al., 2008). θ_t is a time effect capturing general business cycle effects and ε_{irt} is the error term.

The place of residence of all workers and all establishments is geocoded (see the data section for further details). However, the transport model operates at a zonal level. The zones are 0.1-1 km^2 in built-up areas. All trips are assumed to depart from and arrive at a given point within each zone, called the centroid, rather than to and from the precise coordinates where the individual resides and work. If the zone is large, this will introduce an approximation error in travel times and travel cost computed by the transport model. To account for this, the controls (X) also include the (log) distance from the coordinate of the workers' place of residence to the centroid of the transport demand model.

Since we observe the accessibility (A) and wage earnings (y) at two points in time, we may construct a fixed effect estimator canceling out some of the fixed effects. For workers that don't change place of residence between *t* and *t+1*, referred to as *stayers*, we estimate the equation by using the within-worker differences of (3)

$$lny_{irt+1} - lny_{irt} = (lnA_{rt+1} - lnA_{rt})\delta + (X'_{it+1} - X'_{it})\beta + (\theta_{t+1} - \theta_t) + (\varepsilon_{irt+1} - \varepsilon_{irt}). \tag{4}$$

This estimator cancels out the two fixed effects θ_i and θ_r. For workers that change place of residence between t and *t+1*, referred to as movers, we construct the fixed effect estimator

$$lny_{ir't+1} - lny_{irt} = (lnA_{r't+1} - lnA_{rt})\delta + (X'_{it+1} - X'_{it})\beta + (\theta_{t+1} - \theta_t) + (\theta_{r'} - \theta_r) + (\varepsilon_{ir't+1} - \varepsilon_{irt}). \tag{5}$$

This estimator cancels out the individual-specific fixed effect θ_i, but not the zone-specific fixed effect θ_r. We also provide estimates of a model defined by (5) based on the pooled sample of stayers and movers. Note also that job accessibility is a characteristic of the zone of residence rather than a characteristic of the individual. Thus, equation 5 does not measure how a change in job accessibility of a specific zone affects the change in wage earnings for a resident in that zone. Instead, it measures what happens to an individual's wage earnings when he changes zone of residence and the related change in job accessibility.

Making the distinction between stayers and movers in the modeling framework introduces a potential sample selection problem into the analysis. From the standard model of sample selection (Heckman (1979)), we know that if the error terms of the selection equation and the wage earnings equation are correlated, the OLS estimator applied to the wage earnings equations (4) and (5) will be biased and inconsistent if this correlation is ignored.

In our data, however, there is evidence that the decision to move seems mostly related to life choices such as marriage and size of household. Movers are younger on average and they are much more likely than stayers to become married between *t* and *t+1*. In addition, for movers, the average number of young children in the household increases between *t* and *t+1* whereas it decreases for stayers. By controlling for marital status and number of children in the analysis, we therefore address most of the sample selection related to such decisions and the related change in wage earnings.[3]

The models contain a mixture of variables defined at the worker and zonal level, implying that the accessibility is constant across a large number of workers. To avoid overestimating the precision of the estimators for this reason, we use a block bootstrap (Cameron & Trivedi, 2005, p. 845) and resample the original sample 30 times. The standard errors presented for all models in this paper are the standard deviation of the resulting 30 estimates of each parameter.

2.2 Instrumental variable

Labor market accessibility is a function of the number of jobs in each zone. $\Delta \mathrm{A} = \ln \mathrm{A}_{r\prime t+1} - \ln \mathrm{A}_{rt}$ may, therefore, be endogenous in the wage equation (4) for at least two reasons: higher local wages may attract more jobs and workers, and external shocks may simultaneously affect the number of jobs and wages. We therefore apply an instrumental variable that controls for the change in the number of and spatial distribution of jobs defined by

$$ln\tilde{A}_{rt+1} - lnA_{rt} = \tag{6}$$
$$ln\left(\sum_{r' \in \tilde{R}_r} exp\left(-0.028 gc_{t+1,r,r'}\right) n_{t,r'}\right) - ln\left(\sum_{r' \in \tilde{R}_r} exp\left(-0.028 gc_{t,r,r'}\right) n_{t,r'}\right),$$

[3] It may also be a poor empirical strategy to use restrictions on functional forms of the control variables in the wage equation to identify the selection effect (see Cameron and Trivedi, 2005, pp. 551-552 for further discussion).

where $\tilde{A}_{rt+1}$ is the job accessibility computed by assuming the transport system of *t+1* (i.e. the generalized travel costs $gc_{t+1,r,r'}$) and the number and spatial distribution of jobs in t ($n_{t,r'}$).

This instrument captures the change in accessibility between t and *t+1* driven by changes in the generalized travel costs arising from transport system improvements. It is thus strongly correlated with the change in job accessibility in (4) ($lnA_{r\prime t+1} - lnA_{rt}$) and therefore a relevant instrument.

For stayers, this instrument does not capture the change in accessibility arising from changes in the number and spatial distribution of jobs. Specifically, it is not influenced by changes in congestion arising from changes in the number and spatial distribution of jobs. Hence, the instrument should be exogenous in the wage equation and consequently valid. One could argue that if infrastructure investments were prioritized in regions experiencing declining or increasing wages to a larger extent than the average, the instrument would not be exogenous. This is, however, unlikely, given the considerable time it takes from the time an infrastructure investment is suggested to the time when it opens for traffic - usually a decade or more. Since we are only modeling *changes* in wages earnings, the instrument would still be exogenous even if transport investments were prioritized in regions with generally high or low *levels* of productivity than the average. Hence, for stayers, there is no reason to believe that this instrument is correlated with the error term in the wage question (4) and should therefore be a valid instrument.

For movers, however, the instrument captures the changes in the number of accessible jobs and the changes in the accessible transport system due to the move. However, the instrument should still be exogenous because the number of accessible jobs pertains to the 1995 level when computing the instrument. Hence, the instrument is not affected by endogeneity caused by wage increases attracting more jobs and workers and by external shocks simultaneously affecting the number of jobs and wages.

2.3 Data sources

To estimate the above models, we use data for two years eleven years apart: 1995 and 2006. The data used in the present paper are derived from two different data sets. One is the administrative register of the Swedish population in the range 20-64 years of age and all establishments (Statistics Sweden, 2011). Workers are linked to the establishments where they work. Places of residence and establishment are geocoded on a grid which is 1000 by 1000 meters in rural areas and 250 by 250 meters in urban areas. The other data source consists of output from the national travel demand forecasting tool Sampers.

The origin and destination zones in the transport model are represented by one coordinate, a centroid. The worker's place of residence is linked to the closest centroid using the Euclidean distances. Likewise, each establishment is linked to the closest centroid using the Euclidean distance. The Sampers model system consists of the five regional sub-models. In this paper, we use the largest region only since historical transport systems are available for this region only. It includes approximately a third of the Swedish workforce (in 1995) and consists of the counties Stockholm, Uppsala, Sodermanland, Vastmanland, and Orebro.

We analyze gross annual wage earnings (converted to the price level of 2010 using the consumer price index) in this paper, i.e. we estimate the combined effect on wage rates and labor supply (Gutiérrez-i-Puigarnau and van Ommeren, 2010).

2.4 The generalized cost

To compute the generalized cost of transport, we use the following output from the transport model system:

- (i) in-vehicle travel time by car at peak hours,
- (ii) travel distance in the road network,
- (iii) in-vehicle travel time by public transport,
- (iv) first waiting time for public transport
- (v) total waiting time for public transport
- (vi) auxiliary time for public transport
- (vii) cost of a commuter's card adjusted to reflect the cost per trip
- (viii) number of commuters by mode: walk, bicycle, car and public transport

Appendix 1 describes in detail how we have used this information to compute the average generalized cost of travel $gc_{t,r,r'}$.

It is well-established that the value of time depends on the wage rate, implying that it also increases over time as wages increase. However, since we only want to capture changes in accessibility arising from changes in the transport system to keep this variable exogenous, we have kept the values of time constant across years. By the same line of reasoning, we do not adjust travel costs over time. Fuel prices and public transport fares have risen at a similar rate. A second reason for keeping the values of time, GDP, and transport costs constant over time is that when analyzing the effect of a transport investment in appraisal, these parameters do not differ between the do-nothing and the investment scenario.

2.5 Descriptive statistics

We restrict the sample to individuals who are employed in both years and who live and work in the aforementioned counties. We also restrict the sample to only include individuals with commuting distances of 200 km or less. Workers having longer commutes may have a second dwelling closer to the work place. Since we apply fixed effects estimators we use a balanced panel; i.e. we restrict the sample to individuals who meet the sample restrictions both in 1995 and in 2006.[4]

[4] The results obtained with the estimators that do not address fixed effects produce similar results for the unbalanced panel and the balanced panel (cf. equation 3 excluding the fixed effects).

Table 1 presents descriptive statistics of key variables used in the analysis. The average yearly earnings were 231 000 SEK in 1995 and 354 000 SEK eleven years later. Since this is a balanced panel, the increase in earnings partly reflects general wage increase due to economic growth but also the increase in productivity of the workers in the sample.

There is a modest increase in average log job accessibility between the two years. This corresponds to an increase by some 14 percent in average job accessibility between 1995 and 2006. In the appendix, we present the corresponding table separately for the two subsamples: stayers and movers. In the former group, the average job accessibility increases by 24 percent and in the latter by 8 percent.

Job accessibility for stayers has changed for two reasons: because of changes in generalized cost of transport and changes in the spatial distribution of jobs. Decomposing the average change in job accessibility for stayers between 1995 and 2006 into one component arising from changes in the transport system (generalized costs of transport) and one component arising from changes in the spatial distribution and number of jobs indicates how relatively important transport system improvements are for the increased accessibility. From (1) we have

$$A_{rt+1} - A_{rt} = \left(\sum_{r' \in \tilde{R}_r} \left[exp\left(-0.028 gc_{t+1,r,r'}\right) - exp\left(-0.028 gc_{t,r,r'}\right) \right] \left(\frac{n_{t+1,r'} + n_{t,r'}}{2} \right) \right) + \left(\sum_{r' \in \tilde{R}_r} \left[n_{t+1,r'} - n_{t,r'} \right] \frac{\left(exp\left(-0.028 gc_{t+1,r,r'}\right) + exp\left(-0.028 gc_{t,r,r'}\right) \right)}{2} \right),$$

where the first expression on the right-hand side of the equal sign measures the impact of changes in the generalized cost of transportation on the change in job accessibility and the second expression measures the impact of changes in the number and spatial distribution of jobs. Taking the average of the two expressions over all stayers shows that approximately 20 percent of the total change in accessibility arises from changes in the transport system.

Turning back to Table 1, we see that the average distance to the closest centroid also increases between the two years from around 500 meters to almost 600 meters. This increase is driven by the movers (see Table A1 in appendix). Approximately half of the sample is male and the average age in 1995 is 37 years. Finally, some 13 percent of the workers in the sample have an educational attainment corresponding to primary school in 1995, almost 50 percent have at most attained secondary school, some 36 percent have a university degree and one percent of the workers have a Ph.D.[5] The average educational attainment increases slightly between 1995 and 2006. Table 1A in

[5] This is, more formally, any kind of research degree that we call Ph.D. for short.

appendix shows that movers are younger, are more likely to get married, and have more children between the two years than stayers.

Table 1 Descriptive statistics

	1995		*2006*	
Variable	*Mean*	*StDev*	*Mean*	*Stdev*
Earnings (kSEK)	2302.730	1413.461	3538.945	2755.000
Ln (Accessibility)	9.095	1.302	9.185	1.317
Distance to centroid (km)	0.531	0.803	0.600	0.884
Male	0.496	0.500	0.496	0.500
Age	37.275	9.086	48.275	9.086
Primary school (< 9 y)	0.032	0.176	0.031	0.172
Primary school (9 -10 y)	0.101	0.301	0.088	0.283
Secondary school (<3 y)	0.327	0.469	0.303	0.460
Secondary school (>=3 y)	0.168	0.374	0.164	0.371
University (< 3 years)	0.183	0.387	0.176	0.381
University (<= 3 years)	0.178	0.382	0.220	0.414
Ph.D.	0.011	0.104	0.018	0.132
Married	0.466	0.499	0.541	0.498
Children aged 0-3	0.216	0.493	0.119	0.377
Children aged 4-6	0.176	0.430	0.112	0.347
Children aged 7-10	0.200	0.472	0.164	0.433
Children aged 11-15	0.219	0.501	0.256	0.555
Children aged 16-17	0.083	0.283	0.110	0.326
Number of observations	598771		598771	

3 RESULTS

3.1 Stayers and movers pooled

Table 2 presents the results of estimating (5) on the pooled sample of stayers and movers. The first two columns show the estimation results applying a between estimator (BE). The first column shows the result of the BE estimator without the socio-economic controls, by which the elasticity is estimated at 0.033. This is well in line with the elasticity estimated in previous literature. Control variables are added in the second model, to address spatial sorting on observables. The elasticity falls to 0.004, indicating that sorting with respect to the controls is substantial.

The controls include dummy variables for being male, being married and educational attainment. They also include age, age-squared, the number of children aged 0-3 years in the household, the number of children aged 4-6 years, the number of children aged 7-10 years, the number of children aged 11-15 years, and the number of children aged 16-17 years. All models in this paper, except the first BE model in Table 2, include control for industry and sector, but these are not presented in the tables to save space. The impact of the controls is assumed to stay constant across years.

Model 3 is the fixed effect estimator (FE) defined by (5). This estimator controls for individual-specific fixed effects by regressing the within-individual differences in earnings on the within-individual differences in accessibility and the controls. The elasticity increases slightly to 0.007, indicating that conditional on the controls, workers predicted to have lower productivity based on the unobserved factors tend to reside in places with higher job accessibility. Age and gender are excluded in the FE model; gender stays constant across the years and age differs by 11 years for all workers.

In the final column, we present the results obtained with the fixed effect estimator and the instrument (6), capturing the change in total accessibility arising from changes in the transport system. The estimates are similar to those obtained with the fixed effects model. We discuss this result further in the next section, distinguishing between stayers and movers.

Table 2 Estimation results of pooled observations of stayers and movers. BE refers to the between estimator and FE refers to the fixed effects estimator.

Model	*1. BE*	*2. BE*	*3. FE*	*4. FE by IV*
Ln (Accessibility)	0.033 (0.002)	0.004 (0.001)	0.007 (0.001)	0.004 (0.001)
Ln (Distance to centroid)	-	-0.011 (0.001)	-0.005 (0.001)	-0.007 (0.001)
Male	-	0.299 (0.001)	-	-
Age	-	0.082 (0.001)	-	-
Age-squared/100	-	-0.083 (0.001)	-0.088 (0.001)	-0.088 (0.001)
Primary school (9 -10 y)	-	0.106 (0.004)	0.147 (0.027)	0.147 (0.028)
Secondary school (<3 y)	-	0.174 (0.004)	0.137 (0.026)	0.137 (0.024)
Secondary school (>=3 y)	-	0.256 (0.004)	0.083 (0.027)	0.083 (0.026)
University (< 3 y)	-	0.364 (0.004)	-0.006 (0.027)	-0.006 (0.025)
University (<= 3 y)	-	0.557 (0.005)	0.410 (0.027)	0.410 (0.026)
Ph.D.	-	0.845 (0.008)	0.676 (0.029)	0.676 (0.027)
Married	-	0.052 (0.002)	0.020 (0.003)	0.020 (0.002)
Children aged 0-3	-	-0.215 (0.003)	-0.243 (0.002)	-0.243 (0.002)
Children aged 4-6	-	0.034 (0.003)	-0.034 (0.002)	-0.034 (0.002)
Children aged 7-10	-	0.012 (0.002)	-0.045 (0.002)	-0.045 (0.002)
Children aged 11-15	-	-0.034 (0.002)	0.006 (0.002)	0.006 (0.002)
Children aged 16-17	-	-0.019 (0.003)	-0.022 (0.002)	-0.023 (0.002)
Intercept	7.470 (0.017)	4.662 (0.034)	1.181 (0.006)	1.182 (0.007)
Number of observations	598 771	598 771	598 771	598 771

Note: Bootstrapped standard errors robust for clustering at the zone of residence in parentheses. Models 2-4 also include a full set of dummy variables for industry and sector of employment but the corresponding parameters are not reported in the table.

3.2 Stayers and movers

The zone-specific fixed effect, representing non-human endowments and local interaction in the neighborhoods, is present in the equation for movers but canceled out for stayers. Therefore, the elasticity estimated for stayers is more relevant when assessing the impact of transport improvement on the productivity of workers -- for the movers, the change in job accessibility will depend on the characteristic of the zones of residence before and after the move rather than on the transport infrastructure. Moreover, there are good reasons to believe that our instrument (6) is more valid for stayers, according to the discussion in Section 2.2.

The models defined by (4) and (5), estimated separately for stayers and movers, are reported in Table 3. The FE estimate of the elasticity of wage earnings with respect to job accessibility is 0.020 for stayers and lower, 0.009, for movers.

For stayers, the FE by IV model is reported in column three. The elasticity is lower, 0.007, than in the FE model. This indicates some substantial endogeneity in the FE model, controlled for by the instrument. However, the standard error is relatively high so the elasticity is barely significantly (t-ratio 1.75) different from zero at conventional levels of significance. For movers, the IV approach changes the estimated elasticity marginally.

The FE by IV estimate of the elasticity of wage earnings with respect to job accessibility is 0.007 for both stayers and movers. The similar results of the FE by IV estimator for stayers and movers indicate that there are no zone-specific fixed effects on wage earnings. Moreover, the finding that the FE estimate in the sample of movers is similar to the FE by IV results in the sample of stayers suggests that the FE estimator in the sample of movers solves much of the endogeneity problem in the estimation equation.

Table 3. Estimation results for the subsamples of stayers and movers.

	5. FE (Stayers)	*6. FE (Movers)*	*7. FE by IV (Stayers)*	*8. FE by IV (Movers)*
Ln (Accessibility)	0.020 (0.005)	0.009 (0.001)	0.007 (0.004)	0.007 (0.001)
Ln (Distance to centroid)	0.003 (0.004)	-0.008 (0.001)	0.003 (0.004)	-0.009 (0.001)
Age-squared/100	-0.075 (0.001)	-0.093 (0.001)	-0.075 (0.001)	-0.093 (0.001)
Primary school (9 -10 y)	0.128 (0.037)	0.147 (0.036)	0.127 (0.033)	0.147 (0.036)
Secondary school (<3 y)	0.148 (0.035)	0.121 (0.033)	0.147 (0.036)	0.121 (0.032)
Secondary school (>=3 y)	0.148 (0.037)	0.043 (0.034)	0.147 (0.039)	0.043 (0.034)
University (< 3 y)	0.122 (0.038)	-0.076 (0.033)	0.121 (0.040)	-0.076 (0.034)
University (<= 3 y)	0.403 (0.038)	0.383 (0.032)	0.402 (0.039)	0.383 (0.033)
Ph.D.	0.651 (0.042)	0.645 (0.037)	0.651 (0.041)	0.645 (0.036)
Married	-0.004 (0.004)	0.025 (0.003)	-0.004 (0.006)	0.025 (0.003)
Children aged 0-3	-0.285 (0.004)	-0.238 (0.003)	-0.285 (0.004)	-0.238 (0.004)
Children aged 4-6	-0.059 (0.003)	-0.031 (0.004)	-0.059 (0.003)	-0.031 (0.002)
Children aged 7-10	-0.047 (0.002)	-0.045 (0.003)	-0.047 (0.002)	-0.045 (0.002)
Children aged 11-15	-0.014 (0.002)	0.022 (0.002)	-0.014 (0.003)	0.021 (0.003)
Children aged 16-17	-0.030 (0.003)	-0.020 (0.004)	-0.030 (0.004)	-0.020 (0.004)
Intercept	1.021 (0.010)	1.233 (0.008)	1.023 (0.011)	1.233 (0.006)
Number of observations	295 420	303 351	295 420	303 351

Note: Bootstrapped standard errors robust for clustering at the zone of residence are given in parentheses. All models also include a full set of dummy variables for industry and sector of employment but the corresponding parameters are not reported in the table.

3.3 First stage regressions

Table 4 presents the first stage regressions of Models 4, 7 and 8. The partial F-test included in the table shows that the instrument is very relevant for both stayers and movers.

Table 4. First stage regressions for the pooled sample and the separate subsamples of stayers and movers (dependent variable is the change in total job accessibility)

	4. FE by IV (Stayers + Movers)	*7. FE by IV (Stayers)*	*8. FE by IV (Movers)*
Ln (IV Accessibility)	0.999 (0.001)	1.031 (0.014)	0.997 (0.000)
Ln (Distance to centroid)	-0.003 (0.000)	0.001 (0.002)	-0.005 (0.000)
Age-squared/100	0.001 (0.000)	0.002 (0.000)	0.000 (0.000)
Primary school (9 -10 y)	-0.006 (0.005)	-0.009 (0.008)	0.004 (0.008)
Secondary school (<3 y)	-0.017 (0.005)	-0.024 (0.007)	-0.006 (0.007)
Secondary school (>=3 y)	-0.035 (0.005)	-0.046 (0.007)	-0.021 (0.007)
University (< 3 y)	-0.032 (0.005)	-0.041 (0.007)	-0.019 (0.007)
University (<= 3 y)	-0.017 (0.005)	-0.028 (0.007)	-0.006 (0.007)
Ph.D.	0.005 (0.005)	0.006 (0.008)	0.007 (0.008)
Married	-0.001 (0.001)	-0.003 (0.001)	0.002 (0.000)
Children aged 0-3	0.007 (0.001)	0.002 (0.001)	0.005 (0.000)
Children aged 4-6	0.006 (0.001)	0.006 (0.001)	0.002 (0.001)
Children aged 7-10	0.003 (0.000)	0.002 (0.001)	0.001 (0.001)
Children aged 11-15	0.005 (0.000)	0.005 (0.001)	0.003 (0.000)
Children aged 16-17	0.004 (0.001)	0.008 (0.001)	0.000 (0.001)
Intercept	0.147 (0.002)	0.104 (0.009)	0.150 (0.001)
Partial F-test	2 681 439	5 297	9 267 456
Number of observations	598 771	295 420	303 351

Note: Bootstrapped standard errors robust for clustering at the zone of residence in parentheses. All models also include a full set of dummy variables for industry and sector of employment but the corresponding parameters are not reported in the table.

3.4 Sensitivity tests

We have undertaken a number of sensitivity tests to explore the robustness of the previous results. Most important, we have performed an analysis similar to the main analysis but applied job accessibility at the place of work rather than from the place of residence. Hence, A_{rt} is still defined by (1) but r now refers to the zone of the work place.

Job accessibility at the place of residence (the main analysis) influences the matching of workers and jobs. A natural hypothesis is also that the accessibility at the place of residence influences learning and knowledge spillovers due to interactions with neighbors, but we found evidence for this in our main analysis. However, job accessibility at the place of work should have a larger influence on learning and knowledge spillovers around the work place arising in industrial clusters, for instance. In addition, job accessibility at the place of work would better capture agglomeration economies that result from increasing returns to scale in production. Accessibility at the place of work also corresponds better to the accessibility measure in some previous papers in the literature; e.g. Rosenthal & Strange (2003), who estimate the effect of employment density.

Table 5 presents the estimation results obtained from equations (4) and (5), the FE and FE by IV models in the subsamples of stayers and movers. Note that "stayers" now refers to workers who kept working in the same zone between the years, while "movers" refers to workers who have changed zone of work place. Hence, the zone-specific fixed effect now pertains to the zone of work rather than the zone of residence. There is a risk that the distinction between stayers and movers in this context introduces bias due to sample selection, if the change of work place is related to change in wage earnings. The instrument, however, is still likely to be both relevant and exogenous for stayers and movers along the line argument in Section 2.2

The FE estimate for stayers in Table 5 is slightly higher than the FE estimate in Table 3: 0.025 compared to 0.020. However, the FE by IV estimate in the sample of stayers is not significantly different from zero.

For movers, the elasticity is higher than in the main analysis, indicating that job accessibility defined at the place of work has a stronger impact on wage earnings than job accessibility defined at the place of residence. Furthermore, in Table 5 we see again that the FE and the FE by IV estimators result in almost the same elasticity in the sample of movers. Again, just as in the main analysis, this suggests that the FE solves the problem of endogeneity in the estimation equation for movers, i.e. that the change in job accessibility resulting from the change of work place is exogenous.

The estimates in Table 5 are not significantly different between stayers and movers for the FE models, but are significantly different for the FE by IV model. The higher elasticity for movers than for stayers indicates that the work place zone-specific fixed effects are positively correlated with changes in job accessibility for movers. Another possible interpretation of the larger effect for movers is that dynamic agglomeration effects are larger than static effects: more productive firms tend to grow faster and locate in more accessible zones than less productive firms. Workers moving work place are part of, and therefore capture, this dynamic effect, whereas stayers capture only the static agglomeration effects. This interpretation would also explain why the effect of job accessibility at the place of work for movers is higher than that at place of residence estimated in the main analysis.

Table 5. Sensitivity test: job accessibility measured from place of work

	11. FE (Stayers)	*12. FE (Movers)*	*13. FE by IV (Stayers)*	*14. FE by IV (Movers)*
Ln (Accessibility)	0.025 (0.014)	0.037 (0.001)	-0.024 (0.020)	0.036 (0.002)
Number of observations	154 887	292 156	154 887	292 156

Note: Bootstrapped standard errors robust for clustering at the zone of work place in parentheses. All models also include the same control variables used in the models of the main analysis except distance to the closest centroid of the individual.

We have also conducted sensitivity tests with respect to potential changes in the parameters of the control variables by relaxing the restriction that the parameters of the controls stay constant across the years. This resulted in slightly lower estimates of the elasticity than in the main analysis.

The final sensitivity analysis explored how the level of the scale parameter estimated in the transport model (the constant -0.028 in equation (1)) affects the results. We have applied a scale parameter that was twice -0.028 and one only half of -0.028. This did not have any substantial effect on the results obtained with the BE and FE estimators in the total sample (cf. Models 1, 2 and 3 in Table 2).

4 CONCLUDING REMARKS

This paper has empirically estimated the effect of job accessibility on wage earnings in the greater Stockholm area, using geocoded micro data on workers and establishment with high spatial resolution. We use measures of accessibility with high spatial resolution, taking into account commuters' actual travel behavior and preferences for different modes. This also allows us to construct a relevant and exogenous instrumental variable previously never used in this research, namely the increase in accessibility arising from transport system improvements only.

Approximately 20 percent of the total change in accessibility between the 11 years for workers staying in the same zone arises from transport system improvements. Although transport system improvements are an important source of increased accessibility to the labor market for workers– it is not as large as the temporal changes in the number and spatial distribution of jobs.

The estimated elasticity of wage earnings with respect to the job accessibility from place of residence is just below 0.01, when controlling for both individual-specific and zone-specific fixed effects. This is lower than the previous estimates, which might be due to the instrument effectively controlling for endogeneity. We find, moreover, that elasticity is similar for stayers and movers, indicating that the zone-specific fixed effects at the place of residence are small or non-existent.

The corresponding elasticity of wage earnings with respect to job accessibility at the place of work is not significantly different from zero in the sample of stayers. However, the corresponding elasticity in the sample of movers is estimated at 0.036. The higher effect of accessibility at the work place for movers than for stayers suggests that work place zone-specific fixed effects are important and positively correlated with accessibility change for workers changing work place. This is consistent with the result of Glaeser and Mare (2001), showing that workers become more productive when moving to cities (i.e. more accessible work places). A related but slightly different interpretation is that dynamic agglomeration economies are more important than static agglomeration economies: more productive firms tend to grow faster and locate in more accessible zones than less productive firms.

The low elasticity of the accessibility defined at the zone of residence may partly be due to small wage effects arising from matching between workers (as opposed to effects of dynamic agglomeration economies or knowledge spillovers when workers move to more accessible work places). This has some important implications for appraisal: Infrastructure investments improving the accessibility for commuters are not likely to result in any major economic benefits not included in a standard cost-benefit analysis.

5 REFERENCES

Beser, M., & Algers, S. (2002). SAMPERS - The New Swedish National Travel Demand Forecasting Tool. In *In: Lundqvist, L., Mattsson, L.-G. (Eds.), National transport models: recent developments and prospects*. Springer.

Börjesson, M., & Eliasson, J. (2014). Experiences from the Swedish Value of Time study. *Transportation Research A*, *59*, 144–158.

Ciccone, A. (2002). Agglomeration effects in Europe. *European Economic Review*, *46*(2), 213–227. http://doi.org/10.1016/S0014-2921(00)00099-4

Ciccone, A., & Hall, R. E. (1996). Productivity and the Density of Economic Activity. *National Bureau of Economic Research Working Paper Series*, *No. 4313*. Retrieved from http://www.nber.org/papers/w4313

Combes, P.-P., Duranton, G., & Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, *63*(2), 723–742. http://doi.org/10.1016/j.jue.2007.04.004

Combes, P.-P., Duranton, G., Gobillon, L., & Roux, S. (2010). Estimating agglomeration economies with history, geology, and worker effects. In *Agglomeration Economics* (pp. 15–66). University of Chicago Press.

Combes, P.-P., & Gobillon, L. (2014). *The Empirics of Agglomeration Economies* (SSRN Scholarly Paper No. ID 2503404). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2503404

Di Addario, S., & Patacchini, E. (2008). Wages and the City. Evidence from Italy. *Labour Economics*, *15*(5), 1040–1061. http://doi.org/10.1016/j.labeco.2007.09.003

Duranton, G., & Puga, D. (2004). Chapter 48 Micro-foundations of urban agglomeration economies. In J. Vernon Henderson and Jacques-François Thisse (Ed.), *Handbook of Regional and Urban Economics* (Vol. Volume 4, pp. 2063–2117). Elsevier.

Glaeser, E. L., & Mare, D. (2001). Cities and skills. *Journal of Labor Economics*, *19*(2), 316–342.

Graham, D. J. (2007). Variable returns to agglomeration and the effect of road traffic congestion. *Journal of Urban Economics*, *62*(1), 103–120. http://doi.org/10.1016/j.jue.2006.10.001

Graham, D. J., Melo, P. S., Jiwattanakulpaisarn, P., & Noland, R. B. (2010). Testing for Causality Between Productivity and Agglomeration Economies. *Journal of Regional Science*, *50*(5), 935–951. http://doi.org/10.1111/j.1467-9787.2010.00676.x

Graham, D., & van Dender, K. (2011). Estimating the agglomeration benefits of transport investments: some tests for stability. *Transportation*, *38*(3), 409–426. http://doi.org/10.1007/s11116-010-9310-0

Gutiérrez-i-Puigarnau, E., & Ommeren, J. N. van. (2010). *Start Time and Worker Compensation: Implications for Staggered-Hours Programs* (SOEPpapers on Multidisciplinary Panel Data Research No. 310). DIW Berlin, The German Socio-Economic Panel (SOEP). Retrieved from https://ideas.repec.org/p/diw/diwsop/diw_sp310.html

Melo, P. C., Graham, D. J., & Brage-Ardao, R. (2013). The productivity of transport infrastructure investment: A meta-analysis of empirical evidence. *Regional Science and Urban Economics*, *43*(5), 695–706. http://doi.org/10.1016/j.regsciurbeco.2013.05.002

Melo, P. C., Graham, D. J., & Noland, R. B. (2009). A meta-analysis of estimates of urban agglomeration economies. *Regional Science and Urban Economics*, *39*(3), 332–342. http://doi.org/10.1016/j.regsciurbeco.2008.12.002

Mion, G. (2004). Spatial externalities and empirical analysis: the case of Italy. *Journal of Urban Economics*, *56*(1), 97–118. http://doi.org/10.1016/j.jue.2004.03.004

Mion, G., & Naticchioni, P. (2005). *Urbanization Externalities, Market Potential and Spatial Sorting of Skills and Firms* (SSRN Scholarly Paper No. ID 821105). Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=821105

Rice, P., Venables, A. J., & Patacchini, E. (2006). Spatial determinants of productivity: Analysis for the regions of Great Britain. *Regional Science and Urban Economics*, *36*(6), 727–752. http://doi.org/10.1016/j.regsciurbeco.2006.03.006

Rosenthal, S. S., & Strange, W. C. (2004). Chapter 49 Evidence on the nature and sources of agglomeration economies. In J. Vernon Henderson and Jacques-François Thisse (Ed.), *Handbook of Regional and Urban Economics* (Vol. Volume 4, pp. 2119–2171). Elsevier. Retrieved from http://www.sciencedirect.com/science/article/pii/S1574008004800063

Statistics Sweden (2011), Longitudinell integrationsdatabas för Sjukförsäkrings- och Arbetsmarknadsstudier (LISA) 1990–2009, Bakgrundsfakta 2011:4.

Appendix 1

The generalized commuting cost is computed as follows based on the national value of time study and all prices are given in price level 2010 (Börjesson and Eliasson, 2014). In-vehicle travel time by car is valued at 87 SEK/h. Car travel cost is obtained by multiplying distance by 1.30 SEK. In-vehicle travel time and auxiliary time for public transport time are both valued at 69 SEK/h. The value of first waiting time decreases with time because travelers are assumed to spend a larger share of the waiting time at home the longer the first waiting time is. The first 10 minutes are valued at 80 SEK/h, the next 20 minutes (10-30 minutes) at 65 SEK/h, the next 30 minutes (30-60 minutes) at 32 SEK/h, the next 60 minutes (60-120 minutes) at 19 SEK/h, and the next 360 minutes (120-480 minutes) are valued at 10 SEK/h. Transit time, the difference between total waiting time and first waiting time, is valued at 173 SEK/h.

The generalized cost of transport for walk and bicycle is computed based on car travel distance, and by assuming a walking speed of 6 km/h and a cycling speed of 30 km/h. The value of time is 81 SEK/h for walking and 231 SEK/h for cycling.

Appendix 2

Table A1 presents descriptive statistics for movers and stayers. Movers tend to be younger than stayers and to a large extent change marital status between 1995 and 2006. The increase in number of children and reduction in job accessibility among movers suggest that they tended to move to places with lower accessibility to afford a larger house.

Table A1 Descriptive statistics – movers and stayers

	Movers			
	1995		*2006*	
Variable	*Mean*	*StDev*	*Mean*	*Stdev*
Earnings (100 SEKs)	2201.351	1357.161	3592.097	2896.878
Ln (Accessibility)	9.302	1.290	9.296	1.329
Distance to centroid (km)	0.432	0.664	0.566	0.854
Male	0.506	0.500	0.506	0.500
Age	34.073	8.940	45.073	8.940
Primary school (< 9 y)	0.019	0.137	0.018	0.132
Primary school (9 -10 y)	0.097	0.296	0.080	0.271
Secondary school (<3 y)	0.324	0.468	0.293	0.455
Secondary school (>=3 y)	0.190	0.392	0.179	0.383
University (< 3 years)	0.194	0.395	0.183	0.386
University (<= 3 years)	0.167	0.373	0.230	0.421
Ph.D.	0.009	0.096	0.018	0.134
Married	0.353	0.478	0.476	0.499
Children aged 0-3	0.213	0.490	0.204	0.479
Children aged 4-6	0.144	0.395	0.180	0.428
Children aged 7-10	0.148	0.414	0.220	0.493
Children aged 11-15	0.163	0.440	0.235	0.534
Children aged 16-17	0.064	0.253	0.083	0.288
Number of observations	303 351		303 351	
	Stayers			
Earnings (100 SEK)	2406.831	1461.733	3484.367	2600.216
Ln (Accessibility)	8.883	1.280	9.071	1.296
Distance to centroid (km)	0.633	0.913	0.634	0.912
Male	0.485	0.500	0.485	0.500
Age	40.564	7.995	51.564	7.995
Primary school (< 9 y)	0.046	0.208	0.044	0.205
Primary school (9 -10 y)	0.105	0.307	0.096	0.294
Secondary school (<3 y)	0.330	0.470	0.313	0.464
Secondary school (>=3 y)	0.146	0.353	0.150	0.357
University (< 3 years)	0.172	0.378	0.170	0.376
University (<= 3 years)	0.189	0.392	0.210	0.407
Ph.D.	0.012	0.111	0.017	0.130
Married	0.581	0.493	0.607	0.489
Children aged 0-3	0.219	0.495	0.031	0.194
Children aged 4-6	0.209	0.461	0.042	0.215
Children aged 7-10	0.252	0.519	0.107	0.353
Children aged 11-15	0.277	0.550	0.278	0.575
Children aged 16-17	0.103	0.311	0.137	0.358
Number of observations	295 420		295 420	