Agglomeration, productivity and the role of transport system improvements

Maria Börjesson (VTI and KTH)
Gunnar Isacsson (Swedish Transport Administration)
Matts Andersson (WSP Analysis & Strategy)
Christer Anderstig (WSP Analysis & Strategy)

CTS Working Paper 2018:16

Abstract

We explore how the transport improvements impact agglomeration and thereby productivity in mid-Sweden including Stockholm 1995-2006. We measure agglomeration, and changes in agglomeration in response to transport improvements, based on travel times. This is a more accurate measure of agglomeration than previously used and also necessary for understanding how governments can impact agglomeration, and thereby productivity, by transport investments. We regress temporal changes in wages on temporal changes in agglomeration applying a FE estimator. We deal with the potential endogeneity using a novel instrument variable. Our best estimates of the agglomeration elasticity on productivity lie within the interval 0.028-0.035.

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

1   INTRODUCTION

Many countries spend substantial public resources on road and rail investments. A recurrent argument used by those favoring spending on transport infrastructure is its positive effects on agglomeration and productivity. There is also a growing number of empirical studies identifying agglomeration effects on productivity (for reviews see Combes and Gobillon (2015), Melo et al. (2013), Melo et al. (2009), and Rosenthal and Strange (2008)). However, these studies are less helpful in advising governments on how to impact agglomeration by transport system improvements. This paper contributes to the literature by estimating to what extent the Government increased agglomeration by transport system improvements, and the resulting effect on productivity, in mid-Sweden including the greater Stockholm metropolitan area in the period 1995-2006.

The vast majority of studies on the returns to agglomeration use density measures based on distance, and only a few use simple measures based on travel time.1 However, as pointed out by Glaeser (2004), it is not agglomeration per se that generates benefits. It is the number of jobs that can be reached within a given travel time that is relevant for the generation of agglomeration benefits. Glaeser further points out that agglomeration advantages seem to be as large in sprawling cities like Silicon Valley as in the denser downtowns like Detroit, which is not surprising given that it is the travel time with any mode of transport that should be relevant for the returns to agglomeration. When measuring agglomeration by travel time, the transport system plays a key role. In this paper we contribute to the literature by measuring agglomeration more accurately than previous studies, based on the travel times by all travel modes. This is essential not only for measuring agglomeration accurately, but also for analyzing the impact on agglomeration and productivity of transport system improvements.

Graham and Van Dender (2011) argue that measures of agglomeration based on travel times would impose endogeneity problems because locations with higher economic activity tend to be congested. However, such endogeneity is not a problem in our study; our measure of agglomeration is constructed such that it only picks up changes in travel times resulting from transport system improvements and not changes in congestion levels arising from changes in job and population densities.

We use data on travel time changes caused by transport system improvements combined with employee-establishment-linked micro data, including wage earnings and other socio-economic characteristics, for two points in time, 1995 and 2006. Like Gibbons et al. (2017) our data includes detailed geographic information on economic activity for all workers, including the coordinates of the place of residence and the place of employment (establishment). Hence, our

---

1 Rice et al. (2006) and Graham (2007) use the car travel times implied by speed limits. Holl (2011) and Gibbons et al. (2017) use travel times in the road network. However, other modes of transport e.g. rail and public transport, constitutes a large share of trips, primarily in metropolitan areas. Moreover, also pecuniary costs impact the travel behaviour.

measure of agglomeration accounts for the local spatial dimension of the agglomeration, following for example Rosenthal and Strange (2003) and Desmet and Fafchamps (2005), as suggested by Holl (2011, p. 20). We regress temporal changes in wage earnings on temporal changes in agglomeration, controlling for observed changes in socio-economic characteristics. We use wage earnings to measure productivity.[2] The estimator controls for time-invariant individual fixed effects (following Glaeser and Maré (2001), e.g.), and thereby takes care of most of the endogeneity.

We deal with possible time-varying unobserved variables using an instrument based on changes in agglomeration over time arising from the transport system improvements only. As discussed in Section 3.3, there are strong reasons to believe that transport system improvements are exogenous. Our instrumental variable (IV) is thus based on temporal changes rather than on historical levels, e.g. population density or total population (Ciccone and Hall, 1996; Combes et al., 2008; Matano and Naticchioni, 2012; Mion, 2004; Mion and Naticchioni, 2005; Rice et al., 2006), geography like the total land area (Ciccone, 2002; Ciccone and Hall, 1996; Combes et al., 2008; Hanson, 2005; Mion, 2004; Mion and Naticchioni, 2005; Redding and Venables, 2004; Rice et al., 2006)) or geology (Combes et al., 2010; Di Addario and Patacchini, 2008; Rosenthal and Strange, 2008; Rosenthal and Strange, 2004).[3] Graham and van Dender (2011) conclude that the literature in general gives little support for the validity of level instruments based on history and geology. Combes et al. (2011) note that level instruments based on history, geology or geography are weak when estimating the returns to agglomeration within one metropolitan area, which is the purpose of this paper, as opposed to most of the previous studies working on the metropolitan scale (examples are Foster and Stehrer (2009), Brülhart and Mathys (2008), Marrocu et al., (2013), Wheeler (2006) and Yankow (2006)).

Combes et al. (2011) note that a transport system improvement within a metropolitan area can increase agglomeration both directly (by easing interactions within the city) and indirectly (through population and employment growth). The analysis in the present paper focuses on the direct effect; we leave out the net effect on population and employment in the region in response to transport improvement, by only including workers that are residing and employed in the region both years. Now, the agglomeration measure that we use (reflecting the ease of interactions within the city) changes over time for two reasons: first, because the spatial distribution of jobs changes, and second, because transport system improvements reduce travel times. We focus on the effect of transport system improvements because this is the most

---

[2] Wage earnings are a perfect substitute for productivity assuming that the labor market is perfectly competitive. This might, however, not hold in a monopsonistic labor market (where workers receive many job offers) with job search frictions arising from generalized commuting costs. Then workers with high generalized commuting cost will negotiate higher wages because their reservation wage rises with the commuting cost (opportunity cost of the employment is lower than for employees with lower commuting costs) (Manning, 2003; Mulalic et al., 2014).

[3] Most studies report small differences between IV results and the corresponding FE estimators (Graham et al., 2010).

important policy instrument that a Government has to increase agglomeration advantages.

Our identification strategy resembles the one proposed by Gibbons et al. (2017), finding that area level employment, productivity per worker and number of plants increase with new road infrastructure. Holl (2012) uses a measure of agglomeration based on travel times in the Spanish road network and shows that firm-level productivity increases with agglomeration. However, in contrast to Holl (2012) and Gibbons (2017), using only free-flow car travel times, we use a detailed representation of the transport system. Our study is also related to some of the earlier studies estimating the returns to agglomeration by exploiting natural experiments (Hanson (1997), Redding and Sturm (2008), Davis and Weinstein (2008) and Greenstone et al. (2010)).

There are several recent studies estimating the effect of transport infrastructure in China and India. Analyzing Chinese cities from 1990 to 2010, Baum-Snow et al. (2017) show that infrastructure investments have had a substantial impact on the urban form and on some types of economic activities. Faber (2014) shows that China's National Trunk Highway System has led to a reduction in GDP growth for counties in the periphery, left outside of the system. Donaldson (forthcoming) finds that India's vast railroad network enhanced trade and welfare in the regions accessing the railroad network, but harmed those without access. However, like many metropolitan areas in the west, the Stockholm region has a more mature and well-developed transport system, compared to its population, than most metropolitan areas in China and India have. We therefore expect the impact of transport system improvements to be smaller in our study than in the studies from China and India.

The rest of this paper is organized as follows. In section 2 we describe our measures of agglomeration. In section 3, we present the empirical models and outline some of the identification problems and how we have tackled them. Data are presented in section 4, and section 5 contains the results with a discussion. Section 6 concludes.

2   MEASURING AGGLOMERATION

The measure of agglomeration that we use is sometimes referred to as (job) accessibility (Gibbons et al., 2017) or effective density (Graham, 2007). However, in contrast to earlier studies, we can construct two different measures of agglomeration because our data reveal not only the workers' job location but also their location of residence.

Our first measure is defined at the worker's location of residence and is based on the travel time between the workers' location of residence and the location of all jobs in the metropolitan area. Our second measure is defined at the worker's job location and is based on the travel time between the worker's job location and all other jobs in the metropolitan area. The latter is thus not directly related to the workers' commuting travel time. Hence, the first measure of agglomeration reflects the proximity to jobs from the workers' location of

residence and the second measure reflects the proximity to other jobs at the worker's job location.

Now, the contribution of different mechanisms to the estimated returns to agglomeration is likely to differ between our measures of agglomeration. The micro-foundations of agglomeration economies, including job match quality, human capital accumulation, and local infrastructure endowment are reviewed in Duranton and Puga (2004), classifying them as matching, learning, and sharing.[4] Glaeser and Maré (2001) and Baum-Snow and Pavan (2012) find that the main source of the wage premium in cities is greater accumulation of human capital, facilitating the exchange of ideas, creativity and innovations. Wheeler (2001) shows that the firm-worker matching process implies a higher expected return to skill acquisition. However, few empirical studies distinguish between the mechanisms of agglomeration economies.

Even if one cannot rule out contributions from learning and sharing, job matching is obviously important for the returns to agglomeration at the workers' location of residence (but it is for instance certainly possible that being better matched leads to more accumulation of human capital on the job). Lower commuting costs would encourage workers to search for and take jobs within larger geographical areas, accepting longer commuting distances (Pilegaard and Fosgerau, 2008).[5] Now, a better match quality can only increase earnings if accompanied by changes in employer, industry, or occupation. However, shorter commuting time may also increase the earnings because the workers can use the freed-up time to work longer hours.

On the other hand, spillovers should be an important contributor to the returns to agglomeration facing a worker at her job location. However, matching and sharing could also contribute to the wage effect, since they depend on the work environment. For instance, proximity to other workers is likely to increase interactions in relevant business networks, which could improve the job-match quality through referral-based job search networks (Dustmann et al., 2016).

Employment density used by most other studies corresponds most closely to our measure of agglomeration defined at the job location.[6] This measure of agglomeration resembles the market potential suggested by Harris (1954), based on the size of own location, the size of other locations and the distance to

---

[4] Higher wage growth in cities can also reflect changes in prices of human capital driven by changes in local prices (for instance land prices) or amenities (Glaeser, 2008). However, as pointed out by Glaeser (ibid.), the demand for labor in large cities would not exist if the workers were not productive enough to offset the higher wages.

[5] Employers also search for workers. However, the commuting or relocation costs usually fall on the worker. Thus, commuting costs defined at the worker's current place of residence is probably more relevant to the matching process than commuting costs defined at the location of the establishment

[6] The related empirical literature started with Rauch (1993), aiming at estimating human capital spillovers by including a measure of local average educational attainment together with the individual's own educational attainment in a standard Mincerian wage equation (Acemoglu and Angrist (2000), Moretti (2004) and Ciccone and Peri (2006) are more recent studies). Isacsson (2005) presents related evidence for Sweden, *inter alia*, including measures of generalized travel cost to access educated workers.

them on the road network. The standard finding is a rapid spatial decay of agglomeration advantages (Head and Mayer, 2006; Holl, 2006). Baum-Snow (2007) shows that highways fueled suburbanization.

We cannot be conclusive about the precise mechanisms contributing to the returns to agglomeration measured at the workers' location of job and location of residence. However, differences in the returns to agglomeration at the workers' location of job and location of residence could be important information to governments aiming at increasing agglomeration advantages by transport system improvements. A key question for a government could be whether to improve commuting travel connections or to improve the connections between job locations, to generate the biggest returns to agglomeration. As further discussed in section 5.4, distinguishing the returns to agglomeration at the workers location of residence from the returns to agglomeration at the workers' job location is also important in the context of cost-benefit analysis of transport investments.

3   THE MODEL

3.1   Method

We use a micro-level database comprising all individuals of age 20-64 residing in mid-Sweden (including the metropolitan area of Stockholm). Information on place of residence is geocoded with high spatial resolution. The database also includes detailed socio-economic characteristics e.g. wage earnings and educational attainment. This database is linked to another database including all establishments (work places) in the same area. The location of each establishment is also geocoded with high spatial resolution. Furthermore, the establishment database contains additional information *inter alia* on industry. This employee-establishment linked data set is longitudinal so that individuals, establishments and employees can be identified across years.

Since we focus on the effect on agglomeration of transport improvements, a precise representation of the transport system is essential. Therefore, we derive the travel times (or more precisely generalized transport cost) by all modes using the national transport model system, constituting the best practice state-of-the-art of large-scale transport models (described in Section 4.1.). The Swedish government uses this model system for cost-benefit analyses of transport investments. The transport model operates at a zonal level. All employees and establishments in the employee-establishment linked data set are assigned to a zone using the geocoding such that all workers are linked to a zone of residence r and a zone of employment r'.

We compute a measure of agglomeration for all workers (age 20-64) residing in the study area, at two points in time, 1995 and 2006. The measures of agglomeration for each of the two years are computed from three different sources: the micro-level database of individuals, the micro-level database of all

establishments[7], and the generalized transport cost between zones, simulated by the national transport model system. The generalized transport cost is the sum of the pecuniary and non-pecuniary costs of a trip. The non-pecuniary cost of a trip is usually (including the present study) the cost of travel time.[8]

The transport model simulates travel time, travel distance, travel cost and passenger volume by mode for car, public transport, cycling and walking for all pairs of origin and destination zones. From this output, we compute the generalized transport costs for all pairs of origin and destination zones. The generalized transport costs simulated by the transport model differ between the two years 1995 and 2006 because the transport infrastructure network in the latter year is extended by the transport infrastructure investments and possible changes in the train and public transport timetables made between the two years.

To simulate the generalized transport costs of 2006, the transport model is fed with the number of individuals residing in each zone, the number of workers by industry employed in each zone, and the economic input parameters: GDP, fuel prices, and public transport fares prevailing in 2006. To simulate the generalized transport costs of 1995, the transport model is fed with the 2006 population by residential zone, the 2006 number of employed workers by zone, and the economic input parameters prevailing in 2006. Thus, it is only the transport network that differs between the simulations of the generalized transport costs for 1995 and for 2006. In other words, it is only differences in the transport network that impact the simulated travel times by origin and destination pair. However, the spatial distribution of residents and workers remain unchanged. This means that the changes in the simulated generalized transport costs are not affected by increased congestion levels resulting from changes in the agglomeration. This also makes the agglomeration measure more relevant for governments aiming at increasing agglomeration by transport improvements.

We then regress temporal changes in wage earnings (between 1995 and 2006) on temporal changes in our measure of agglomeration (between 1995 and 2006) for each individual worker while controlling for a set of other variables detailed below. We use an IV approach based on changes in the transportation network to deal with remaining problems of endogeneity. In addition, we outline cross-sectional models and related results for expository purposes.

3.2 The wage equation

Assume that the study area is divided into zones, and in each zone a number of workers reside and a number of workers are employed. Remember that we have two measures of agglomeration. The first measure of agglomeration, A_{rt}, is

---

[7] Including establishments in the public and private sector.

[8] Appendix 1 describes how the generalized transport cost is derived from the output from the transport model.

defined for the residents of zone r ($r=1, 2, ..., R_{r,t}$) at time t ($t=1, 2$), and is defined as

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

where $gc_{t,r,r'}$ is the average generalized transport cost at time t between zone r and r'; $n_{t,r'}$ is the number of workers employed in zone r' at time t. $\tilde{R}_{r,t}$ is the set of zones r' to which there is a strictly positive number of travelers from zone r at time t.

The decay parameter -0.028 is estimated (in the transport model) from the travelers' observed behavior. The decay parameter reflects how sensitive travelers are to the generalized transport cost when making their mode and destination choices. If the decay parameter approaches zero, the number of full-time jobs in zone r' is multiplied by one. A decay parameter of -0.028 implies that the number of full-time jobs in zone r' is multiplied by 0.25 if the average generalized transport cost between r and r' is EUR 5 and by 0.004 if the average generalized transport cost is EUR 20. The average generalized transport cost is defined as

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

where $w(m)_{t,r,r'}$ is the share of travelers 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 transport cost 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 price a traveler would be willing to pay to save one unit of travel time). In transport economics, it is well known that the perceived non-pecuniary cost differs between different components of the public transport travel time (Wardman, 2004): in-vehicle time, average waiting time and walk access time. The value of time, therefore, differs between the components of the public transport travel time.

Let the annual gross wage earnings y_{irt} of worker i ($i=1, 2, ..., I$) residing in zone r at time t be given by

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

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 industry of employment. θ_i is an individual-specific fixed effect capturing time-invariant unobserved productivity differences between workers (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.

8

The transport model assumes that all trips 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. The zones are 0.1-1 km^2 in built-up areas, and if the zone is large, this will introduce an approximation error in travel times and travel cost simulated by the transport model. To account for this, the controls (X) also include the (log) distance from the coordinate of the workers' coordinate of residence to the centroid of the transport demand model.

Since we observe the agglomeration measure (A) and wage earnings (y) at two points in time, we may construct a fixed effect estimator canceling out some of the fixed effects. We construct the first difference fixed effect estimator

$$lny_{ir^*t+1} - lny_{irt} = (lnA_{r^*t+1} - lnA_{rt})\alpha + (X'_{it+1} - X'_{it})\beta + (\theta_{t+1} - \theta_t) +$$
$$(\theta_{r^*} - \theta_r) + (\varepsilon_{ir^*t+1} - \varepsilon_{irt}). \tag{4}$$

where r^* is the residence zone of individual i at time $t + 1$. This estimator cancels out the individual-specific fixed effect θ_i, but not the zone-specific fixed effect θ_r. The endogeneity that is not controlled for by the fixed effects for this reason is taken care of by the IV. The zonal fixed effects can be swept away if we restrict the sample to individuals who do not change zone of residence between t and $t + 1$ (stayers) but this may introduce sample selection problems in the analysis (see Section 5.2). For this reason, we present separate results for stayers and movers in Section 5.2 only for expository purposes.

The second agglomeration measure, A_{et}, is defined from the workers' zone of employment, e (e=1, 2, ..., $R_{e,t}$), at time t (t=1, 2) but is otherwise equivalent to (1)

$$A_{et} = \sum_{r' \in \tilde{R}_{e,t}} exp(-0.028gc_{t,e,r'}) n_{t,r'}. \tag{5}$$

where $gc_{t,e,r'}$ is the average generalized transport cost at time t between zone e and r'; $\tilde{R}_{e,t}$ is the set of zones r' to which there is a strictly positive number of travelers from zone e at time t. When applying the second measure of agglomeration, A_{et}, let the annual gross wage earnings y_{iet} of worker i (i=1, 2, ..., I) employed in zone e at time t be given by

$$lny_{iet} = lnA_{et}\gamma + X'_{it}\delta + \theta_i + \theta_e + \theta_t + \varepsilon_{iet}. \tag{6}$$

The corresponding fixed effect estimator is

$$lny_{ie^*t+1} - lny_{iet} = (lnA_{e^*t+1} - lnA_{et})\gamma + (X'_{it+1} - X'_{it})\delta + (\theta_{t+1} - \theta_t) +$$
$$(\theta_{e^*} - \theta_e) + (\varepsilon_{ie^*t+1} - \varepsilon_{iet}). \tag{7}$$

Again, this estimator cancels out the individual-specific fixed effect θ_i, but not the zone-specific fixed effect θ_e. The wage effect estimated by (4) is assumed to mainly arise from improved job match quality and the effect estimated from (7) is assumed to arise mainly from knowledge spillovers.

9

We do not include both measures in the same regression model, because they are both endogenous. Including the two endogenous variables in the same model causes identification problems, since they are strongly correlated.[9] Using an endogenous variable as a control variable may be a "bad control" discussed by Angrist and Pischke (2009 pp. 64-68). Since changes in the two different measures of agglomeration will probably often be correlated, we advise against adding the effect of them in evaluations of transport improvements. However, as we will see, the returns to agglomeration measured from the job location are approximately nine times larger than the returns to agglomeration measured from the location of residence, so this is in practice not a major problem.

3.3 Instrumental variable

Agglomeration is a function of the number of workers employed in each zone and the generalized transport cost to access them. The fixed effects estimator (FE) allows us to control for all time-invariant unobserved variables where there may still be endogeneity in the wage equations (4) and (7) for at least two reasons. First, external shocks may simultaneously influence the firms' demand for labor and wages in a given zone. Second, time-varying unobserved characteristics in wages due to job match quality or spillovers may correlate with changes in agglomeration. Moreover, a relatively larger increase in local wages in a zone may attract workers to this zone, increasing the supply of labor. We therefore apply an IV that controls for the change in the number of workers employed in each zone. For the fixed effect estimator (4), we define the instrument

$$
\begin{aligned}
ln\tilde{A}_{rt+1} - lnA_{rt} = \\
ln\big(\textstyle\sum_{r'\in\tilde{R}_{r,t+1}} exp\left(-0.028gc_{t+1,r,r'}\right) n_{t,r'}\big) - \\
ln\big(\textstyle\sum_{r'\in\tilde{R}_{r,t}} exp\left(-0.028gc_{t,r,r'}\right) n_{t,r'}\big),
\end{aligned}
\tag{8}
$$

where $\tilde{A}_{rt+1}$ measures agglomeration assuming the transport network of *t+1* (i.e. the generalized transport costs $gc_{t+1,r,r'}$) and the number of workers employed in each zone of time t ($n_{t,r'}$). For the fixed effect estimator (7), we define the equivalent instrument

$$
\begin{aligned}
ln\tilde{A}_{et+1} - lnA_{et} = \\
ln\big(\textstyle\sum_{r'\in\tilde{R}_{e,t+1}} exp\left(-0.028gc_{t+1,e,r'}\right) n_{t,r'}\big) - \\
ln\big(\textstyle\sum_{r'\in\tilde{R}_{e,t}} exp\left(-0.028gc_{t,e,r'}\right) n_{t,r'}\big).
\end{aligned}
\tag{9}
$$

This instrument captures the change in agglomeration between *t* and *t+1* driven by changes in the generalized transport costs arising from transport network changes only. It is thus strongly correlated with the change in agglomeration in (4) and (7), respectively, and therefore a relevant instrument.

---

[9] We tested to include both agglomeration measures in the same equation, but the result could not be interpreted: the estimated elasticities changed signs because of correlation between the two endogenous variables.

The instruments do not capture changes in transport costs originating from changes in congestion levels, which in turn would originate from changes in the number of residents or workers employed in each zone. Hence, endogeneity arising from congestion is controlled for as described in Section 2.

Now, if transport infrastructure improvements for political reasons were targeted towards areas with systematically higher or lower changes in incomes (Baum-Snow and Ferreira, 2015; Ozbay et al., 2006) our instrument could be correlated with the error term in the wage questions (4) and (7). However, previous studies show that this is not the case in Sweden. Studying the Swedish infrastructure investments plans from 1990 and 2000, Nilsson (1991) concludes that road transport projects in areas with weak economic growth and low population density are no more likely to be included in the plan. A more recent study (Eliasson et al., 2014) analyzing the factors influencing project selection for the Swedish (and Norwegian) national transport plan, shows that rural projects do not have a higher probability of being selected, which would be the case if project were selected on the basis of the employment trends in the region. Nor have projects in metropolitan areas, where there is congestion, a greater probability of being selected.

Hence, previous studies clearly demonstrate that the transport infrastructure investments in Sweden (and Norway) are not targeted to regions with higher or lower employment density or income growth, which would invalidate our identification strategy. Instead, the selection of infrastructure investments seems to be driven purely by political considerations. This might be the case in many countries, but in Norway and Sweden there are data available on all *suggested* infrastructure investments, making it possible to analyze how they are selected to the national plan.

Moreover, even if transport infrastructure investments were targeted towards regions with higher or lower productivity growth (at the time when they are decided on), this would still not induce endogeneity because of the long planning and building processes. It takes at least a decade, normally longer, for a transport infrastructure investment to open for traffic from the time it was decided. During this time, the relative economic growth in the region might have shifted. In addition, note that since we are estimating how *changes* in agglomeration impact *changes* in productivity, the instrument would still be exogenous even if transport infrastructure investments were prioritized in regions with generally higher or lower *levels* of productivity than the average.

However, the IVs in (8) and (9) may still be correlated with the error terms in equations (4) and (7), respectively, if the level of employment in 1995 correlates with residual wage changes between 1995 and 2006. If, for example, zones with high employment at time t experience a larger residual wage growth than zones with low employment, the IV estimator using (8) and (9) will be biased upwards. Therefore, we also apply the following IV to (4)

$$lnZ_{rt+1} - lnZ_{rt} = ln\left(R_{r,t+1}^{-1} \sum_{r' \in \tilde{R}_{rt+1}} exp\left(-0.028gc_{t+1,r,r'}\right)\right)$$
$$-ln\left(R_{r,t}^{-1} \sum_{r' \in \tilde{R}_{r,t}} exp\left(-0.028gc_{t,r,r'}\right)\right), \qquad (10)$$

where $R_{r,t}$ is the number of zones to which there is a strictly positive number of travelers from zone r at time t. This instrument captures the change in agglomeration between t and $t+1$ caused by changes in the transport network but it does not depend on the level of employment at time t. It therefore controls for a possible correlation between unobserved individual-specific wage growth and the employment level at time t. Still, we conjecture that it is strongly correlated with the change in agglomeration in $(lnA_{r't+1} - lnA_{rt})$ and therefore is a relevant instrument. We define the IV for (7) correspondingly. In the following, we refer to the IV estimator applied to (4) and (7) using the instrumental variables presented here as the fixed effects by instrumental variables estimator (FE by IV).

In (3) and in (6) we assume that the parameter of the agglomeration measure is a constant, although it might vary across individuals, regions, municipalities, or zones.[10] We therefore estimate the average of this parameter. Imbens and Angrist (1994) show that the estimated average depends on the choice of instrument: it recovers the local average treatment effect (LATE) and not the average treatment effect (ATE). They show that the IV estimate (under weak conditions) equals the weighted average of local average treatment effects, where the weights depend on the elasticity of the instrument on the endogenous regressor. By choosing the temporal increase in agglomeration arising from infrastructure investments as the instrument, we thus estimate the wage effect of workers in the zones where agglomeration is increased by infrastructure improvements. This is exactly the wage effect that we want to estimate and that is relevant for governments. We are interested in the effect of agglomeration caused by transport improvements, but less so of agglomeration caused by other factors; transport improvement is the most important policy that governments can enforce to induce agglomeration (for instance, comparing New York with a metro and Los Angeles without does not help the policy-maker).

3.4 Standard errors

The models contain a mixture of variables defined at the worker and zonal level, implying that the agglomeration measure is constant across many workers. Hence, standard errors need to be corrected for clustering of observations at the zonal level. In addition, some workers change zone of residence and zone of employment between the years so the natural question arises of how to define clusters for correcting the standard errors. We have chosen to apply the procedure outlined in Cameron and Miller (2015 pp. 336-337) when computing cluster-robust standard errors of the parameters. The unique number of

---

[10] For instance, it is likely that the returns to agglomeration are higher in regions where the workforce is more skilled. Many studies (Abel et al., 2012; Bacolod et al., 2009; Groot et al., 2014; Lindley and Machin, 2014) show that the returns to education are higher in cities.

combinations of zones in 1995 and 2006 is 161 872. The corresponding numbers of unique zones in 1995 and 2006 are 2 333 and 2 322, respectively. We use standard econometric software to implement the procedure of Cameron and Miller (2015). However, the standard errors of the FE by IV models are obtained by using a block bootstrap (Cameron and Trivedi, 2005, p. 845) where we resample the original sample 30 times. We tested resampling the original data 50 times for one of the models but the resulting standard errors did not differ substantially from those obtained with a resampling rate of 30.

4   DATA

4.1   Population, employment and transport system data

As explained in Section 3.1, the data used in the present paper are derived from the administrative registers 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. Locations 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 is the output from the national transport model system, constituting the best practice state-of-the-art of large-scale transport models. Its structure – logit demand models linked to a network assignment model – is similar to most operational, large-scale multi-modal transport models. It has been carefully estimated and calibrated using state-of-the-art methods, and validated against in-sample data.

The transport 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) in mid-Sweden including Stockholm.

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

4.2   Descriptive statistics

We restrict the sample to individuals who are employed in both years and who reside and work in mid-Sweden including the Greater Stockholm metropolitan area (approximate size 200*200 km). 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 location. Since we apply FEs on two points in time, we use a balanced panel; i.e. we restrict the sample to individuals who meet the sample restrictions both in 1995 and in 2006.[11] Our final sample used in the estimation where we measure

---

[11] 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).

agglomeration at the zone of residence includes 598 771 workers. When measuring agglomeration from the zone of employment, we restrict the sample in the same way. The final sample then includes 447 043 workers.[12]

Table 1 and Table 2 present descriptive statistics of key variables for the two samples used in the analysis. The real average yearly earnings increased by just over 50% over the eleven years. Since this is a balanced panel, the increase in earnings reflects the increased productivity and changes in hours worked of these workers.

There is a modest increase in agglomeration between the two years: A_{rt} has increased by 9 percent and A_{et} by 14 percent. The agglomeration measure A_{rt} facing the workers who have not changed zone of residence between t and $t+1$ (stayers, see Section 5.2) has changed for two reasons: changes in generalized transport cost and changes in the spatial distribution of jobs. Decomposing the average change in A_{rt} for stayers between 1995 and 2006 into one component arising from changes in the transport system and one component arising from changes in the spatial distribution of jobs indicates the contribution of transport system improvements to the increased agglomeration. From (1) we have

$$A_{rt+1} - A_{rt} =$$

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

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

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

The same computation for the change in A_{et} between 1995 and 2006 for workers not changing zone of employment shows that 37 percent of the total change in agglomeration comes from changes in the transport system. Hence, although transport system improvements have been an important source of

---

[12] The latter sample is smaller because there are some employment zones between which no individuals travel according to the transport model. The generalized costs of travel can therefore not be computed since they require modal shares of travel between zones. This is not a major problem in the analysis simply because the lack of travelers indicates that the destination is not attractive to the travelers. Still, we do undertake sensitivity analysis exploring whether different sample sizes affect the differences between the models (4) and (7).

increased agglomeration over time, they have contributed less to the increases in agglomeration than the changes in the spatial distribution of jobs.

Tables 1 and 2 show that A_{rt} and A_{et} vary substantially across zones. Mid-Sweden is partly dense (central Stockholm), but also in some parts rural. Hence, the spatial variation in the agglomeration is fair enough, enabling a good identification. Tables 1 and 2 also show that the average distance to the centroid in the zone, for residents and workers, increases between the two years. 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 samples 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.[13] The average educational attainment increases slightly between 1995 and 2006.

Table 1: Descriptive statistic; sample used for estimating the effect of agglomeration defined at zone of residence. Earnings are given in price level[14] 2010

	1995		2006	
Variable	Mean	StDev	Mean	Stdev
Yearly earnings (EUR)	23027.30	14134.61	35389.45	27550.00
lnA_{rt}	9.095	1.302	9.185	1.317
Distance to centroid (km) from location of residence	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	

---

[13] This is, more formally, any kind of research degree that we call Ph.D. for short.
[14] Throughout the paper, we use the conversion rate of 10 SEK ≈ 1 EUR.

Table 2: Descriptive statistics sample used for estimating the effect of agglomeration defined at zone of employment. Earnings are given in price level 2010

Variable	1995		2006	
	Mean	*StDev*	*Mean*	*Stdev*
Yearly earnings (EUR)	22148.82	13237.02	33398.46	24237.03
lnA_{et}	9.646	1.299	9.802	1.334
Distance to centroid (km) from location of employment	0.417	0.594	0.422	0.598
Male	0.464	0.499	0.494	0.499
Age	37.632	9.084	48.632	9.084
Primary school (< 9 y)	0.034	0.182	0.033	0.178
Primary school (9 -10 y)	0.101	0.301	0.087	0.282
Secondary school (<3 y)	0.336	0.472	0.310	0.462
Secondary school (>=3 y)	0.158	0.365	0.156	0.363
University (<= 3 y)	0.183	0.387	0.178	0.382
University (> 3 y)	0.177	0.382	0.219	0.413
Ph.D.	0.011	0.103	0.018	0.133
Married	0.476	0.499	0.541	0.498
Children aged 0-3	0.211	0.487	0.109	0.362
Children aged 4-6	0.178	0.432	0.105	0.337
Children aged 7-10	0.207	0.479	0.158	0.425
Children aged 11-15	0.232	0.513	0.252	0.551
Children aged 16-17	0.086	0.291	0.111	0.328
Number of observations	447 043		447 043	

5   RESULTS

In Section 5.1, we present the result of the fixed effects (FE) estimator (4), using the measure of agglomeration defined at the workers' zone of residence. We also report the results obtained with the between estimator (BE) applied to equation (3) with and without the control variables for expository purposes. In Section 5.2 we present the corresponding results from the BE and the FE of (6) and (7), respectively, using the measure of agglomeration defined at the workers' zone of employment. In Section 5.3 we split the sample used to estimate (4) into stayers (workers who have not changed zone of residence) and movers (workers who have changed zone of residence). This is not done for the estimator (7), where we measure agglomeration from the workers' zone of employment because we expect substantial selection problems (the decision to change job is often related to changes in wage). Section 5.4 applies the model to the construction of the Stockholm Metro.

5.1   Measuring agglomeration from zone of residence

The first column of Table 3 shows the result of the BE without the socio-economic controls; the estimated elasticity is 0.033. This is well in line with elasticities reported in previous literature. Control variables are added in the model in the second column, addressing spatial sorting on observables. The

elasticity falls to 0.004, indicating substantial sorting with respect to the controls, i.e. more skilled workers tend to be residing in more productive locations in the cities. This is consistent with Bacolod et al. (2010), Combes et al. (2008), D'Costa and Overman (2014), De La Roca and Puga (2017), Eeckhout et al. (2014) and Mion and Naticchioni (2009), finding that skills are sorted across cities and that this accounts for a large portion of the spatial variation in productivity.

Model 3 is the FE defined by (4). This estimator controls for individual-specific fixed effects. 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 zones where A_{rt} is higher. Model 4 is the FE with the stronger instrument (8); i.e. the FE by IV. The estimates are lower in this model than those obtained with the fixed effects model, indicating some endogeneity in the FE model.

Table A2 in Appendix 3 shows that the partial F-statistics for the first-stage regression of model 4 is high, indicating a strong correlation between the IV and the endogenous variable. This may be a consequence of including the 1995 zonal employment levels in the instrument. As argued in Section 3.3, this may call for the exclusion of employment levels from the instrument, producing a weaker IV. Model 5 in Table 3 is estimated applying this weaker IV and the first-stage regression of model 5 is shown in appendix Table A2. It shows that the partial F-statistics are indeed lower with the instrument (10), indicating a weaker correlation between the instrument and the endogenous variable. Table 3 shows that the estimated elasticity falls from 0.004 to 0.002, and the latter estimate is not statistically different from zero.

In all the models we refrain from adding the establishment size. The reason is that establishment size may be an effect of accessibility (see, for example, Manning, 2010). Thus, if we are interested in estimating the effect of accessibility on wage earnings, establishment size should not be included in the model since part of the accessibility effect on wage earnings would wrongfully be attributed to establishment size. In other words, establishment size in these models may be an example of a 'bad control' (Angrist and Pischke, 2009, pp. 64-68).

5.2   Measuring agglomeration from zone of employment

Table 4 presents the corresponding estimates of (7), applying a measure of agglomeration defined at the worker's zone of employment and IV (9). The elasticity is substantially higher in all models, but follows the same pattern as the models in Section 5.1: it is highest for the BE and lowest for the FE by IV estimator. The effect of the instrument is small, indicating that the FE controls for most of the endogeneity.

The elasticity obtained when applying the weaker IV corresponding to (10) instead of the IV (9) again falls from 0.035 to 0.028. Moreover, the first-stage regressions of model 9 and model 10 in Table A3 in the appendix again shows that the partial F-statistics are indeed lower with the instrument corresponding

17

to (10). Hence, the correlation between the instrument and the endogenous variable is weaker.

The substantially higher parameter estimate for $ln\ A_{et}$ than for $ln\ A_{rt}$ indicates that high job accessibility at the job location has substantially larger impact on wages than high job accessibility at the workers' location of residence. The greater differences between the estimates for $ln\ A_{et}$ than for $ln\ A_{rt}$ are not an effect of different samples sizes; Table A4 compares estimates of models 1-5 with those of model 6-10 applying identical samples and shows that the results are robust.

Table 3: Estimation result, agglomeration defined at zone of residence. 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*	*5. FE by IV weaker IV*
$ln\ A_{rt}$	0.033 (0.003)	0.004 (0.002)	0.007 (0.002)	0.004 (0.001)	0.002 (0.002)
ln (Distance to centroid)	-	-0.011 (0.002)	-0.005 (0.002)	-0.007 (0.002)	-0.008 (0.002)
Male	-	0.299 (0.003)	-	-	-
Age	-	0.082 (0.001)	-	-	-
Age-squared/100	-	-0.083 (0.001)	-0.088 (0.001)	-0.088 (0.001)	-0.088 (0.001)
Primary school (9 -10 y)	-	0.106 (0.004)	0.147 (0.029)	0.147 (0.024)	0.147 (0.025)
Secondary school (<3 y)	-	0.174 (0.004)	0.137 (0.027)	0.137 (0.023)	0.137 (0.028)
Secondary school (>=3 y)	-	0.255 (0.005)	0.083 (0.028)	0.083 (0.028)	0.083 (0.029)
University (<= 3 y)	-	0.364 (0.005)	-0.006 (0.029)	-0.006 (0.026)	-0.006 (0.028)
University (>3 y)	-	0.557 (0.006)	0.410 (0.029)	0.410 (0.026)	0.410 (0.027)
Ph.D.	-	0.845 (0.010)	0.676 (0.030)	0.676 (0.024)	0.677 (0.029)
Married	-	0.052 (0.002)	0.020 (0.003)	0.020 (0.001)	0.019 (0.001)
Children aged 0-3	-	-0.215 (0.003)	-0.243 (0.003)	-0.243 (0.003)	-0.244 (0.002)
Children aged 4-6	-	0.034 (0.003)	-0.034 (0.002)	-0.034 (0.003)	-0.034 (0.001)
Children aged 7-10	-	0.012 (0.003)	-0.045 (0.002)	-0.045 (0.001)	-0.046 (0.001)
Children aged 11-15	-	-0.034 (0.002)	0.006 (0.002)	0.006 (0.002)	0.006 (0.002)
Children aged 16-17	-	-0.019 (0.004)	-0.022 (0.003)	-0.023 (0.003)	-0.023 (0.003)
Number of observations	598 771	598 771	598 771	598 771	598 771

Note: 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 of employment but the corresponding parameters are not reported in the table.

Table 4: Estimation results, agglomeration defined at zone of employment. BE refers to the between estimator and FE refers to the fixed effects estimator

Model	6. BE	7. BE	8. FE	9. FE by IV	10. FE by IV weaker IV
ln A$_{et}$	0.086 (0.006)	0.046 (0.005)	0.038 (0.003)	0.035 (0.003)	0.028 (0.005)
Ln (Distance to centroid)	-	0.001 (0.008)	0.001 (0.004)	-0.001 (0.003)	-0.004 (0.004)
Male	-	0.300 (0.005)	-	-	-
Age	-	0.073 (0.003)	-	-	-
Age-squared/100	-	-0.074 (0.003)	-0.082 (0.002)	-0.082 (0.002)	-0.082 (0.002)
Primary school (9 -10 y)	-	0.095 (0.006)	0.145 (0.033)	0.145 (0.038)	0.144 (0.030)
Secondary school (<3 y)	-	0.157 (0.006)	0.146 (0.032)	0.146 (0.034)	0.145 (0.029)
Secondary school (>=3 y)	-	0.222 (0.009)	0.102 (0.033)	0.102 (0.034)	0.101 (0.037)
University (<= 3 y)	-	0.330 (0.010)	0.035 (0.035)	0.035 (0.039)	0.034 (0.028)
University (> 3 y)	-	0.502 (0.010)	0.421 (0.036)	0.421 (0.038)	0.420 (0.035)
Ph.D.	-	0.788 (0.049)	0.689 (0.041)	0.689 (0.042)	0.688 (0.041)
Married	-	0.045 (0.003)	0.015 (0.003)	0.015 (0.002)	0.015 (0.004)
Children aged 0-3	-	-0.240 (0.009)	-0.264 (0.008)	-0.264 (0.007)	-0.263 (0.008)
Children aged 4-6	-	0.027 (0.005)	-0.039 (0.003)	-0.039 (0.003)	-0.039 (0.002)
Children aged 7-10	-	0.008 (0.003)	-0.047 (0.002)	-0.048 (0.001)	-0.048 (0.002)
Children aged 11-15	-	-0.025 (0.003)	0.003 (0.002)	0.003 (0.001)	0.003 (0.002)
Children aged 16-17	-	-0.006 (0.005)	-0.023 (0.003)	-0.023 (0.003)	-0.023 (0.003)
Number of observations	447 043	447 043	447 043	447 043	447 043

Note: 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 of employment but the corresponding parameters are not reported in the table.

5.3 Stayers and movers

Considering the estimation when the measure of agglomeration is defined at the workers' zone of residence as in Section 5.1, we noted previously that for workers that stay in the same zone of residence between *t* and *t+1*, the zone-

specific fixed effect ($\theta_{r'} - \theta_r$), representing non-human endowments and local interaction in the neighborhoods, cancels out in (4). We refer to this group as *stayers*, and to the workers changing zone of residence between *t* and *t+1* as *movers*. Hence, by splitting the sample into stayers and movers, we can provide some tentative results regarding the zone-specific fixed effects (distinguishing stayers and movers in this way is similar to the analysis of Combes and Gobillon (2015), but they refer to movers between cities).

Distinguishing between stayers and movers in the modeling framework introduces a potential sample selection problem into the analysis. The standard model of sample selection (Heckman, 1979) shows that if the error terms of the selection equation and the wage earnings are correlated, the OLS estimator applied to the wage earnings equations (4) will be biased and inconsistent if this correlation is ignored. For this reason, we do not estimate separate models for stayers and movers when estimating (7), measuring the agglomeration from the workers' zone of employment. The decision to change zone of employment is probably highly correlated with the error term in the wage equation.

Our data suggest, however, that changing zone of residence is mostly related to life choices such as marriage and size of household and not to wage earnings. Table 1A in Appendix 2 shows that movers are younger and more likely to get married between the two years. 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, we therefore address most of the sample selection related to such decisions and the related change in wage earnings.[15]

The models defined by (4), estimated separately for stayers and movers, are reported in Table 5. The FE estimate of the elasticity is 0.020 for stayers and lower, 0.009, for movers. For stayers, the FE by IV model elasticity is lower, 0.007, than in the FE model. This indicates some endogeneity in the FE model, controlled for by the instrument. However, the standard error is high so the elasticity is barely significantly different from zero at conventional levels of significance (t-ratio 1.75). The finding that the FE estimate in the sample of movers is similar to the FE by IV suggests that the former solves much of the endogeneity problem.

The FE by IV estimate of the elasticity of wage earnings with respect to A_{rt} is 0.007 for both stayers and movers. The similar results of the FE by IV estimator for stayers and movers indicate that the residential zone-specific fixed effects on wage earnings are small when using the IV.

---

[15] 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).

Table 5: Estimation results for the subsamples of stayers and movers, agglomeration defined at zone of residence

	11. FE (Stayers)	12. FE (Movers)	13. FE by IV (Stayers)	14. FE by IV (Movers)
ln A_{rt}	0.020 (0.005)	0.009 (0.002)	0.007 (0.004)	0.007 (0.001)
Ln (Distance to centroid)	0.003 (0.004)	-0.008 (0.002)	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 (0.042)	0.127 (0.033)	0.147 (0.034)
Secondary school (<3 y)	0.148 (0.035)	0.121 (0.040)	0.147 (0.036)	0.121 (0.029)
Secondary school (>=3 y)	0.148 (0.037)	0.043 (0.041)	0.147 (0.039)	0.043 (0.033)
University (<= 3 y)	0.122 (0.038)	-0.076 (0.042)	0.121 (0.040)	-0.076 (0.034)
University (> 3 y)	0.403 (0.038)	0.383 (0.041)	0.402 (0.039)	0.383 (0.032)
Ph.D.	0.651 (0.042)	0.645 (0.043)	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 (0.003)
Children aged 4-6	-0.059 (0.003)	-0.031 (0.003)	-0.059 (0.003)	-0.031 (0.004)
Children aged 7-10	-0.047 (0.002)	-0.045 (0.002)	-0.047 (0.002)	-0.045 (0.002)
Children aged 11-15	-0.014 (0.002)	0.022 (0.003)	-0.014 (0.003)	0.021 (0.004)
Children aged 16-17	-0.030 (0.003)	-0.020 (0.004)	-0.030 (0.004)	-0.020 (0.004)
Number of observations	295 420	303 351	295 420	303 351

Note: 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 of employment, but the corresponding parameters are not reported in the table.

5.4 Model application

A better understanding of the effect of transport investments on the agglomeration advantages is important for cost-benefit analysis (CBA) (SACTRA, 1999; Venables, 2007). Specifically, distinguishing the returns to the two measures of agglomeration is important in the context of transport CBA. Most of the wage effect from better matching due to lower commuting costs is

already included in the consumer surplus (Eliasson and Fosgerau, 2017).[16] However, transport system improvements increasing agglomeration at the workers' job location is not directly related to the commuting cost. For this reason, the wage effect of the agglomeration at the workers' job locations is not included in the consumer surplus in the standard CBA (and should therefore be added). As discussed in section 2, such agglomeration advantages should result from spillovers, but also matching (referral-based job search) and sharing could contribute to this wage effect, since they impact the work environment.

Our study suggests that the agglomeration advantages arising from better connections between workers' location of job is substantially larger than the agglomeration advantages resulting from reduced commuting costs. This implies that a larger share of the wage effect arising from transport system improvements is not included in standard transport CBA. Still, commuting connections are important because firms need workers and lower generalized commuting costs reduce unemployment through reducing reservation wages. For instance, Norman et al. (2017) find a negative relationship between generalized commuting costs and unemployment on Swedish data, which is more pronounced for low-educated workers (the most important mechanism should be that, for an unemployed worker with a given reservation wage, a reduction in the generalized commuting cost extends the radius of the job search area, increasing the probability of finding a job (Åslund et al., 2010; Brueckner and Martin, 1997; Brueckner and Zenou, 2003; Pilegaard and Fosgerau, 2008).

To illustrate the magnitude of the effect on wages predicted by our model, in relation to a standard cost benefit analysis, we apply models 9 and 10 (the latter based on the weaker instrument), to compute the wage effect of the Stockholm Metro system. We disregard the wage effect of reduced commuting cost estimated by model 4, because we advise against adding the effects of the two measures of agglomeration, because the estimated effects in model 4 are small, and because most of the wage effect from reduced commuting cost is picked up by the standard consumer surplus.

We first compute A_{e2006} for 2006 for all workers in the Stockholm County, assuming the actual 2006 transport system including the Metro. Second, we compute A_{e2006}', which is computed assuming the actual 2006 transport system except for the Metro (but the buses, trams and commuting trains remain in the public transport network). On average we then have $\frac{A_{e2006}}{A_{e2006}'}$ =1.32.

Total wage earnings in the County in 2006 were EUR 48.4756 billion. Applying the elasticities 0.035 and 0.028 from models 9 and 10, total wage earnings would have been EUR 48.10-48.01 billion had the Metro not been built. This means that the increase in productivity due to the Metro is EUR 0.3754-0.4688 billion in 2006. The standard consumer surplus is EUR 0.9648 billion for 2006

---

[16] However, in the case of an income tax, the workers do not receive the full value of their higher income. For this reason, increases in income tax are external and should be added to the consumer surplus.

(Börjesson et al., 2014). Thus, the consumer surplus should be multiplied by a factor of 1.4-1.5.

6 CONCLUDING REMARKS

We have empirically estimated the returns to agglomeration arising from transport improvements in the Greater Stockholm metropolitan area, using geocoded micro data covering all workers and establishment with high spatial resolution. We use measures of agglomeration with high spatial resolution, taking the transport system and travel times into account. We deal with endogeneity by using temporal changes over 11 years, controlling for all time-invariant individual fixed effects in wage earnings. We also construct novel IVs based on temporal changes, namely the increase in agglomeration arising from transport system improvements only. As opposed to previous studies, this analysis therefore indicates to what extent governments can increase agglomeration through transport system improvements, and the corresponding effect on productivity.

We find that the estimated elasticity is 0.002-0.004 when the agglomeration measure is defined at the location of residence. The corresponding elasticity for the agglomeration measure defined at the location of employment is substantially higher: 0.028-0.035. The results obtained with our weaker IV when we measure agglomeration at the place of residence is 0.002 compared to 0.004 obtained with the stronger IV. The corresponding figures are 0.028 compared to 0.035 when we measure agglomeration at the place of the establishment. The lower estimates obtained with our weaker IV suggest that the level of employment in the base year correlates with the residual of the corresponding FE model.

We find a substantially higher estimate when the measure of agglomeration is defined at the location of employment than at location of residence. This suggests that governments that aim to increase agglomerations, and thereby productivity, by transport improvements should focus on connections between jobs and not only on commuting connections, at least in countries with an already well-developed transport system like Sweden. Still, commuting connections are important because firms need workers, and lower generalized commuting costs reduce unemployment by reducing reservation wages (Norman et al., 2017).

Distinguishing the returns to the two measures of agglomeration is also important in the context of transport CBA. Most of the wage effect from better matching due to lower commuting costs arising from transport system improvements are already included in the consumer surplus (Eliasson and Fosgerau, 2017) and should therefore not be added in a CBA. However, the wage effects from agglomeration advantages in response to transport system improvements at the workers' job location is not directly related to the commuting distance and does not therefore overlap with the consumer surplus in the standard CBA.

Our elasticity for agglomeration defined at the location of employment is within, but at the lower end of, the range of the elasticities reported in the previous literature. Models that ignore individual unobserved effects and reverse causality typically produce elasticities in the range of 0.04-0.07 (Combes and Gobillon, 2015). However, Combes and Gobillon report that studies taking individual fixed effects into account typically find an elasticity of around 0.02. Moreover, previous studies have not focused on the extent to which transport system improvement can induce agglomeration. Furthermore, they have typically measured agglomeration at a metropolitan scale as opposed to our study, which uses variation in the agglomeration measure both within and between cities and rural areas.

Moreover, the previous literature, using cross-section, estimates the long-run effect. Another possible reason for the lower elasticity compared to some other studies is the difficulty of finding a relevant and exogenous instrument. The transport system in the Greater Stockholm metropolitan area is already well developed. The effects could be larger in metropolitan areas with a less well-developed transport system. Still, the elasticity we estimate implies that the consumer surplus for work trips should be multiplied by a factor of 1.4-1.5, which is similar to the current British appraisal guidelines (Department for Transport, 2016, 2014).

7 REFERENCES

Abel, J.R., Dey, I., Gabe, T.M., 2012. Productivity and the Density of Human Capital*. J. Reg. Sci. 52, 562–586. https://doi.org/10.1111/j.1467-9787.2011.00742.x

Acemoglu, D., Angrist, J., 2000. How Large Are Human-Capital Externalities? Evidence from Compulsory Schooling Laws. NBER Macroecon. Annu. 15, 9–59. https://doi.org/10.1086/654403

Angrist, J.D., Pischke, J.-S., 2009. Instrumental variables in action: sometimes you get what you need. Most. Harmless Econom. Empiricist's Companion 113–220.

Åslund, O., Östh, J., Zenou, Y., 2010. How important is access to jobs? Old question—improved answer. J. Econ. Geogr. 10, 389–422. https://doi.org/10.1093/jeg/lbp040

Bacolod, M., Blum, B.S., Strange, W.C., 2010. Elements of Skill: Traits, Intelligences, Education, and Agglomeration*. J. Reg. Sci. 50, 245–280. https://doi.org/10.1111/j.1467-9787.2009.00650.x

Bacolod, M., Blum, B.S., Strange, W.C., 2009. Urban interactions: soft skills versus specialization. J. Econ. Geogr. 9, 227–262. https://doi.org/10.1093/jeg/lbn057

Baum-Snow, N., 2007. Did Highways Cause Suburbanization? Q. J. Econ. 122, 775–805. https://doi.org/10.1162/qjec.122.2.775

Baum-Snow, N., Brandt, L., Henderson, J.V., Turner, M.A., Zhang, Q., 2017. Roads, Railroads, and Decentralization of Chinese Cities. Rev. Econ. Stat. 99, 435–448. https://doi.org/10.1162/REST_a_00660

Baum-Snow, N., Ferreira, F., 2015. Causal Inference in Urban and Regional Economics, in: Handbook of Regional and Urban Economics. Elsevier, pp. 3–68.

Baum-Snow, N., Pavan, R., 2012. Understanding the City Size Wage Gap. Rev. Econ. Stud. 79, 88–127. https://doi.org/10.1093/restud/rdr022

Börjesson, M., 2012. Valuing perceived insecurity associated with use of and access to public transport. Transp. Policy 22, 1–10. https://doi.org/10.1016/j.tranpol.2012.04.004

Börjesson, M., Eliasson, J., 2014. Experiences from the Swedish Value of Time study. Transp. Res. Part A 59, 144–158.

Brueckner, J.K., Martin, R.W., 1997. Spatial mismatch: An equilibrium analysis. Reg. Sci. Urban Econ., Policy Issues in the American and Japanese Space 27, 693–714. https://doi.org/10.1016/S0166-0462(97)00004-5

Brueckner, J.K., Zenou, Y., 2003. Space and Unemployment: The Labor-Market Effects of Spatial Mismatch. J. Labor Econ. 21, 242–262. https://doi.org/10.1086/344129

Brülhart, M., Mathys, N.A., 2008. Sectoral agglomeration economies in a panel of European regions. Reg. Sci. Urban Econ. 38, 348–362. https://doi.org/10.1016/j.regsciurbeco.2008.03.003

Cameron, A.C., Miller, D.L., 2015. A Practitioner's Guide to Cluster-Robust Inference. J. Hum. Resour. 50, 317–372. https://doi.org/10.3368/jhr.50.2.317

Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics: Methods and Applications. Cambridge University Press.

Ciccone, A., 2002. Agglomeration effects in Europe. Eur. Econ. Rev. 46, 213–227. https://doi.org/10.1016/S0014-2921(00)00099-4

Ciccone, A., Hall, R.E., 1996. Productivity and the Density of Economic Activity. Natl. Bur. Econ. Res. Work. Pap. Ser. No. 4313.

Ciccone, A., Peri, G., 2006. Identifying Human-Capital Externalities: Theory with Applications. Rev. Econ. Stud. 73, 381–412. https://doi.org/10.1111/j.1467-937X.2006.00380.x

Combes, P.-P., Duranton, G., Gobillon, L., 2011. The identification of agglomeration economies. J. Econ. Geogr. 11, 253–266. https://doi.org/10.1093/jeg/lbq038

Combes, P.-P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: Sorting matters! J. Urban Econ. 63, 723–742. https://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. University of Chicago Press, pp. 15–66.

Combes, P.-P., Gobillon, L., 2015. Chapter 5 - The Empirics of Agglomeration Economies, in: Duranton, G., Henderson, J.V., Strange, W.C. (Eds.), Handbook of Regional and Urban Economics, Handbook of Regional and Urban Economics. Elsevier, pp. 247–348. https://doi.org/10.1016/B978-0-444-59517-1.00005-2

Davis, D.R., Weinstein, D.E., 2008. A Search for Multiple Equilibria in Urban Industrial Structure*. J. Reg. Sci. 48, 29–65. https://doi.org/10.1111/j.1467-9787.2008.00545.x

D'Costa, S., Overman, H.G., 2014. The urban wage growth premium: Sorting or learning? Reg. Sci. Urban Econ. 48, 168–179. https://doi.org/10.1016/j.regsciurbeco.2014.06.006

De La Roca, J., Puga, D., 2017. Learning by Working in Big Cities. Rev. Econ. Stud. 84, 106–142. https://doi.org/10.1093/restud/rdw031

Department for Transport, 2016. WebTAG: Understanding and valuing impacts of transport investment: updating wider economic impacts guidance, 〈https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/554783/transport-appraisal-guidance-webtag-consultation-document.pdf〉.

Department for Transport, 2014. WebTAG: TAG unit A2-1 wider impacts, 〈https://www.gov.uk/government/publications/webtag-tag-unit-a2-1-wider-impacts〉.

Desmet, K., Fafchamps, M., 2005. Changes in the spatial concentration of employment across US counties: a sectoral analysis 1972–2000. J. Econ. Geogr. 5, 261–284. https://doi.org/10.1093/jnlecg/lbh046

Di Addario, S., Patacchini, E., 2008. Wages and the City. Evidence from Italy. Labour Econ. 15, 1040–1061. https://doi.org/10.1016/j.labeco.2007.09.003

Donaldson, D., forthcoming. Forthcoming."Railroads of the Raj: Estimating the impact of transportation infrastructure." Am. Econ. Rev.

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. Elsevier, pp. 2063–2117.

Dustmann, C., Glitz, A., Schönberg, U., Brücker, H., 2016. Referral-based Job Search Networks. Rev. Econ. Stud. 83, 514–546. https://doi.org/10.1093/restud/rdv045

Eeckhout, J., Pinheiro, R., Schmidheiny, K., 2014. Spatial Sorting. J. Polit. Econ. 122, 554–620. https://doi.org/10.1086/676141

Eliasson, J., Börjesson, M., Odeck, J., Welde, M., 2014. Does benefit/cost-efficiency influence transport investment decisions? (No. 2014:6), CTS Working Paper.

Eliasson, J., Fosgerau, M., 2017. Cost-benefit analysis of transport improvements in the presence of spillovers, matching and an income tax. MPRA Paper.

Faber, B., 2014. Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System. Rev. Econ. Stud. 81, 1046–1070. https://doi.org/10.1093/restud/rdu010

Foster, N., Stehrer, R., 2009. Sectoral Productivity, Density and Agglomeration in the Wider Europe. Spat. Econ. Anal. 4, 427–446. https://doi.org/10.1080/17421770903317629

Gibbons, S., Lyytikäinen, T., Overman, H., Sanchis-Guarner, R., 2017. New road infrastructure: the effects on firms. SERC Discussion Paper 214.

Glaeser, E.L., 2008. Cities, Agglomeration, and Spatial Equilibrium. Oxford University Press.

Glaeser, E.L., Kahn, M.E., 2004. Sprawl and Urban Growth. Handb. Reg. Urban Econ., Cities and Geography 4, 2481–2527. https://doi.org/10.1016/S1574-0080(04)80013-0

Glaeser, E.L., Mare, D., 2001. Cities and skills. J. Labor Econ. 19, 316–342.

Graham, D.J., 2007. Variable returns to agglomeration and the effect of road traffic congestion. J. Urban Econ. 62, 103–120. https://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. J. Reg. Sci. 50, 935–951. https://doi.org/10.1111/j.1467-9787.2010.00676.x

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

Greenstone, M., Hornbeck, R., Moretti, E., 2010. Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings. J. Polit. Econ. 118.

Groot, S.P.T., Groot, D., F, H.L., 2014. Estimating the Skill Bias in Agglomeration Externalities and Social Returns to Education: Evidence from Dutch Matched Worker-Firm Micro-Data (SSRN Scholarly Paper No. ID 2464995). Social Science Research Network, Rochester, NY.

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).

Hanson, G., 2005. Market potential, increasing returns and geographic concentration. J. Int. Econ. 67, 1–24. https://doi.org/10.1016/j.jinteco.2004.09.008

Hanson, G., 1997. Increasing Returns, Trade and the Regional Structure of Wages. Econ. J. 107, 113–133. https://doi.org/10.2307/2235274

Harris, C.D., 1954. The, Market as a Factor in the Localization of Industry in the United States. Ann. Assoc. Am. Geogr. 44, 315–348. https://doi.org/10.1080/00045605409352140

Head, K., Mayer, T., 2006. Regional wage and employment responses to market potential in the EU. Reg. Sci. Urban Econ., New Economic Geography: Closing the Gap between Theory and Empirics 36, 573–594. https://doi.org/10.1016/j.regsciurbeco.2006.06.002

Heckman, J.J., 1979. Statistical models for discrete panel data. Department of Economics and Graduate School of Business, University of Chicago Chicago, IL.

Holl, A., 2012. Market potential and firm-level productivity in Spain. J. Econ. Geogr. 12, 1191–1215. https://doi.org/10.1093/jeg/lbr030

Holl, A., 2006. A Review of the Firm-Level Role of Transport Infrastructure with Implications for Transport Project Evaluation. J. Plan. Lit. 21, 3–14. https://doi.org/10.1177/0885412206288905

Imbens, G.W., Angrist, J.D., 1994. Identification and Estimation of Local Average Treatment Effects. Econometrica 62, 467–475. https://doi.org/10.2307/2951620

Isacsson, G., 2005. External effects of education on earnings: Swedish evidence using matched employee-establishment data (Working Paper No. 2005:10). Working Paper, IFAU - Institute for Labour Market Policy Evaluation.

K., O., D., O., J., B., 2006. Modeling and Analysis of the Link between Accessibility and Employment Growth. J. Transp. Eng. 132, 385–393. https://doi.org/10.1061/(ASCE)0733-947X(2006)132:5(385)

Lindley, J., Machin, S., 2014. Spatial changes in labour market inequality. J. Urban Econ., Spatial Dimensions of Labor Markets 79, 121–138. https://doi.org/10.1016/j.jue.2013.07.001

Manning, A., 2003. The real thin theory: monopsony in modern labour markets. Labour Econ., European Association of Labour Economists, 14th Annual Conference, 2002 10, 105–131. https://doi.org/10.1016/S0927-5371(03)00018-6

Marrocu, E., Paci, R., Usai, S., 2013. Productivity Growth in the Old and New Europe: The Role of Agglomeration Externalities. J. Reg. Sci. 53, 418–442. https://doi.org/10.1111/jors.12000

Matano, A., Naticchioni, P., 2012. Wage distribution and the spatial sorting of workers. J. Econ. Geogr. 12, 379–408. https://doi.org/10.1093/jeg/lbr013

Melo, P.C., Graham, D.J., Brage-Ardao, R., 2013. The productivity of transport infrastructure investment: A meta-analysis of empirical evidence. Reg. Sci. Urban Econ. 43, 695–706. https://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. Reg. Sci. Urban Econ. 39, 332–342. https://doi.org/10.1016/j.regsciurbeco.2008.12.002

Mion, G., 2004. Spatial externalities and empirical analysis: the case of Italy. J. Urban Econ. 56, 97–118. https://doi.org/10.1016/j.jue.2004.03.004

Mion, G., Naticchioni, P., 2009. The spatial sorting and matching of skills and firms. Can. J. Econ. Can. Déconomique 42, 28–55. https://doi.org/10.1111/j.1540-5982.2008.01498.x

Mion, G., Naticchioni, P., 2005. Urbanization Externalities, Market Potential and Spatial Sorting of Skills and Firms (SSRN Scholarly Paper No. ID 821105). Social Science Research Network, Rochester, NY.

Moretti, E., 2004. Estimating the social return to higher education: evidence from longitudinal and repeated cross-sectional data. J. Econom., Higher education (Annals issue) 121, 175–212. https://doi.org/10.1016/j.jeconom.2003.10.015

Mulalic, I., Van Ommeren, J.N., Pilegaard, N., 2014. Wages and Commuting: Quasi-natural Experiments' Evidence from Firms that Relocate. Econ. J. 124, 1086–1105. https://doi.org/10.1111/ecoj.12074

Nilsson, J.-E., 1991. Investment Decisions in a Public Bureaucracy: A Case Study of Swedish Road Planning Practices. J. Transp. Econ. Policy 25, 163–175.

Norman, T., Börjesson, M., Anderstig, C., 2017. Labour Market Accessibility and Unemployment. J. Transp. Econ. Policy JTEP 51, 47–73.

Pilegaard, N., Fosgerau, M., 2008. Cost Benefit Analysis of a Transport Improvement in the Case of Search Unemployment. J. Transp. Econ. Policy JTEP 42, 23–42.

Rauch, J.E., 1993. Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities. J. Urban Econ. 34, 380–400. https://doi.org/10.1006/juec.1993.1042

Redding, S., Venables, A.J., 2004. Economic geography and international inequality. J. Int. Econ. 62, 53–82. https://doi.org/10.1016/j.jinteco.2003.07.001

Redding, S.J., Sturm, D.M., 2008. The Costs of Remoteness: Evidence from German Division and Reunification. Am. Econ. Rev. 98, 1766–1797. https://doi.org/10.2307/29730152

Rice, P., Venables, A.J., Patacchini, E., 2006. Spatial determinants of productivity: Analysis for the regions of Great Britain. Reg. Sci. Urban Econ. 36, 727–752. https://doi.org/10.1016/j.regsciurbeco.2006.03.006

Rosenthal, S.S., Strange, W.C., 2008. The attenuation of human capital spillovers. J. Urban Econ. 64, 373–389. https://doi.org/10.1016/j.jue.2008.02.006

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. Elsevier, pp. 2119–2171.

Rosenthal, S.S., Strange, W.C., 2003. Geography, Industrial Organization, and Agglomeration. Rev. Econ. Stat. 85, 377–393. https://doi.org/10.1162/003465303765299882

SACTRA, 1999. Transport and the Economy. Standing Advisory Committee on Trunk Road Assessment, London.

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

Venables, A.J., 2007. Evaluating Urban Transport Improvements: Cost-Benefit Analysis in the Presence of Agglomeration and Income Taxation. J. Transp. Econ. Policy 41, 173–188.

Wardman, M., 2004. Public transport values of time. Transp. Policy 11, 363–377. https://doi.org/10.1016/j.tranpol.2004.05.001

Wheeler, C.H., 2006. Cities and the growth of wages among young workers: Evidence from the NLSY. J. Urban Econ. 60, 162–184. https://doi.org/10.1016/j.jue.2006.02.004

Wheeler, C.H., 2001. Search, sorting, and urban agglomeration. J. Labor Econ. 19, 879–899.

Yankow, J.J., 2006. Why do cities pay more? An empirical examination of some competing theories of the urban wage premium. J. Urban Econ. 60, 139–161. https://doi.org/10.1016/j.jue.2006.03.004

Appendix 1

To compute the generalized transport cost $gc_{t,r,r'}$, 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 monthly ticket adjusted to reflect the cost per trip
(viii) number of travelers by mode: walk, bicycle, car and public transport

The generalized transport cost is computed as follows, based on the national value of time study and all prices are given in price level 2010 (Börjesson, 2012; Börjesson and Eliasson, 2014). In-vehicle travel time by car is valued at 8.7 EUR/h. Car km cost is assumed to be 0.13 EUR. In-vehicle travel time and auxiliary time for public transport time are both valued at 6.9 EUR/h. The average waiting time is inversely proportional to the service frequency. 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 8.0 EUR/h, the next 20 minutes (10-30 minutes) at 6.5 EUR/h, the next 30 minutes (30-60 minutes) at 3.2 EUR/h, the next 60 minutes (60-120 minutes) at 1.9 EUR/h, and the next 360 minutes (120-480 minutes) are valued at 1.0 EUR/h. Transfer time, the difference between total waiting time and first waiting time, is valued at 17.3 EUR/h.

The generalized transport cost 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 8.1 EUR/h for walking and 23.1 EUR/h for cycling.

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 agglomeration arising from changes in the transport network 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 system improvements in appraisal, these parameters do not differ between the do-nothing and the investment scenario.

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 A_{rt} among movers suggest that they tended to move to zones with lower A_{rt} to afford a larger house.

Table A1: Descriptive statistics – movers and stayers

| | Movers | | | |
| | 1995 | | 2006 | |
Variable	Mean	StDev	Mean	Stdev
Yearly earnings (EUR)	22013.51	13571.61	35920.97	28968.78
lnA_{rt}	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			
Yearly earnings (EUR)	24068.31	14617.33	34843.67	26002.16
lnA_{et}	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	

Appendix 3

Table A2: First-stage regressions for the main and the alternative instrument (the measure of agglomeration is defined at the zone of residence)

	4. FE by IV	*5. FE by IV weaker IV*
ln IV A_{rt}	0.999	1.364
	(0.002)	(0.013)
ln (Distance to centroid)	-0.003	-0.051
	(0.001)	(0.005)
Age-squared/100	0.001	0.002
	(0.000)	(0.001)
Primary school (9 -10 y)	-0.006	-0.013
	(0.005)	(0.012)
Secondary school (<3 y)	-0.017	-0.030
	(0.006)	(0.012)
Secondary school (>=3 y)	-0.035	-0.052
	(0.006)	(0.014)
University (<= 3 y)	-0.032	-0.032
	(0.006)	(0.014)
University (> 3 y)	-0.017	0.004
	(0.006)	(0.013)
Ph.D.	0.005	0.043
	(0.006)	(0.017)
Married	-0.001	-0.023
	(0.001)	(0.003)
Children aged 0-3	0.007	-0.009
	(0.001)	(0.003)
Children aged 4-6	0.006	-0.018
	(0.001)	(0.003)
Children aged 7-10	0.003	-0.009
	(0.000)	(0.002)
Children aged 11-15	0.005	-0.006
	(0.001)	(0.002)
Children aged 16-17	0.004	-0.010
	(0.001)	(0.003)
Partial F-test	258 463	11 903
Number of observations	598 771	598 771

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

Table A3: First-stage regressions for the main and the alternative instrument (the measure of agglomeration is defined at the zone of employment)

	9. FE by IV	*10. FE by IV weaker IV*
ln IV A_{et}	1.002	1.191
	(0.003)	(0.019)
Ln (Distance to centroid)	-0.000	-0.052
	(0.004)	(0.010)
Age-squared/100	-0.001	-0.002
	(0.000)	(0.002)
Primary school (9 -10 y)	-0.009	0.009
	(0.005)	(0.017)
Secondary school (<3 y)	-0.014	-0.004
	(0.006)	(0.017)
Secondary school (>=3 y)	-0.027	-0.018
	(0.008)	(0.020)
University (<= 3 y)	-0.023	-0.010
	(0.008)	(0.020)
University (> 3 y)	-0.005	0.022
	(0.009)	(0.024)
Ph.D.	0.022	0.112
	(0.014)	(0.062)
Married	0.001	-0.005
	(0.001)	(0.002)
Children aged 0-3	0.005	0.012
	(0.001)	(0.003)
Children aged 4-6	0.005	0.001
	(0.001)	(0.002)
Children aged 7-10	0.002	-0.005
	(0.000)	(0.002)
Children aged 11-15	0.004	-0.000
	(0.001)	(0.002)
Children aged 16-17	0.004	-0.001
	(0.001)	(0.002)
Partial F-test	155 233	3 963
Number of observations	447 043	447 043

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

Table A4: Estimation results - sample with non-missing information on both $\ln A_{rt}$ and $\ln A_{et}$. 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	5. FE by IV weaker IV
$\ln A_{rt}$	0.033 (0.004)	0.004 (0.002)	0.009 (0.002)	0.006 (0.001)	0.003 (0.002)
$\ln A_{et}$	0.086 (0.006)	0.046 (0.005)	0.038 (0.003)	0.035 (0.003)	0.029 (0.004)
Number of observations	433468	433468	433468	433468	433468

Note: Standard errors robust for clustering at the zone of residence (row 1) and establishment (row 2) in parentheses. Models 2-4 also include a full set of dummy variables for industry of employment but the corresponding parameters are not reported in the table.