Title: Estimating the effect of accessibility on wages

Authors: Elias Stapput Knudsen, Katrine Hjorth, Ninette Pilegaard, DTU Management

Extended abstract

In this paper, we estimate the elasticity of wages in Denmark with respect to changes in generalised travel costs. The estimation of the elasticity is done on Danish register data from year 2002 and 2010. This elasticity is used as an estimate for the productivity effects of transport infrastructure changes and is to be applied in calculations of the wider economic benefits of transport infrastructure improvements in Denmark.

Agglomeration, the spatial concentration of activity, leads to increased productivity through a number of mechanisms, including the better accessibility of firms to both their markets and supply of specialized labor. It is a long-standing debate to which extent transport projects lead to such so-called wider economic benefits and how such benefits should be quantified and incorporated into policy assessment. The link between transport and agglomeration has been studied intensively (see e.g. Graham & Gibbons 2018, Venables 2007). Now many agree that the effect exists and there is a focus on how to quantify it and include this in transport policy appraisals (see e.g. Graham & Gibbons 2018).

There has been many attempts to estimate the agglomeration effects and determine the causal effect of transport and accessibility on productivity. Many analyses focus on the effect on wages as data for wages are typically more accessible and robust. Other approaches look at total factor productivity or land/house-price changes. It is typically found, that estimates are sensitive to the specific environment and it is therefore generally recommended that local estimates are used in applications.

In a Danish context, both researchers and politicians recognize the need to develop new economic appraisal methods that take into account the effects of agglomeration on productivity, wages and employment, and to estimate elasticities that can be used in Danish project appraisal. The purpose of this paper is to estimate such an elasticity. We use real wages as a measure of productivity and job-to-job accessibility (a distance-weighted sum of jobs) as a measure of agglomeration. We thereby estimate the elasticity on wages from job-to-job accessibility, which we interpret as the elasticity on productivity from agglomeration.

More or less all of the previously used approaches can be questioned due to endogeneity issues or bias (see e.g. Graham & Gibbons 2018). The firms' location choices and the workers' job choices may very well depend on their productivity, implying a reverse causality. We acknowledge this problem but also recognize that neglecting the effects can be a bias. We use a combination of three different approaches to control for selection bias and test the robustness of the results. The first approach (our benchmark) is simply to analyse all people who change work zone in the period of analysis. However, as the decision to change job is likely endogenous, our second approach focuses on the subsample who stay in the same firm, but change job location due to a relocation of the workplace. This controls for some of the endogeneity issues, but not all, since workers still make the decision of staying in the job or leaving, and since the moving decision of the firm may also depend on its productivity. Our last approach is similar to the two-step estimation strategy from Combes et al. (2008).

We use a combination of Danish register data and data from the Danish National Transport Model (NTM). From the register data, we have information about individual wages, employment types, employment locations and socioeconomic characteristics for the entire Danish population in the years 2002 and 2010. From NTM, we have information about travel times and travel costs at a rather detailed level.

We use data for full time wage earners in non-subsidised jobs in 2002 and 2010. We leave out people working in Southern Jutland close to the Danish-German border, and people working on the island of Bornholm: For the first group holds that the labour market close to the border is affected not only by the accessibility to Danish jobs but also to German jobs, that are not possible to include in the analysis; for the second group holds that the island of Bornholm can be considered a separate labour market due to its distance from the rest of the country. Finally, we exclude wage earners less than 18 years old in 2002. The remaining sample consists of 612,814 people.

Our geographical units are the 907 zones of the Danish National Transport Model, which covers all of Denmark, and this is the most detailed level for which we have information about travel times and travel costs. Our main explanatory variable, the accessibility to jobs, is computed for zone i as:

$$A_i(\eta) = \sum_j N_j \cdot \exp\left(-\eta \cdot GTC_{ij}\right)$$

where N_j is the number of jobs in zone j (full time equivalents), GTC_{ij} is the generalised car travel cost from zone i to zone j, and $\eta > 0$ is the decay parameter that determines the distance sensitivity, i.e. how jobs close by are weighted relative to jobs further away. The value of η is not directly observable, and must be assumed or estimated, as we discuss in the following section.

We have chosen to use car travel costs to weight jobs in different zones. In principle, it may be more correct to apply an average over car and public transport, as the public modes constitute a quite large share of the traffic in the urban areas. However, the available information about travel times and ticket costs for 2002 is of quite poor quality.

The generalised car travel cost from zone i to zone j, is a weighted sum of the monetary travel cost (tolls, parking costs, marginal costs of fuel and tyres) and travel time costs. The generalised travel costs changes from 2002 to 2010 because of differences in the road network and ferry services, but also because of differences in the localisation of jobs and residences, and the implied differences in travel behaviour (destination, mode and route choice), which in turn leads to differences in congestion levels.

The number of jobs (in fulltime equivalents) N_j in zone j is computed based on Danish register data of occupation and employment (a database containing all jobs in the country, including self-employed and assisting spouses). There are approximately 2.6-2.7 million working people in the population in each of the two years. Each working person in the population is assigned to the zone of his main occupation in November. Government-subsidised jobs do not count (58,558 in 2002, 81,754 in 2010). Approximately half a million jobs in each year cannot be not matched to a zone, either because the workplace address is missing or because it cannot be matched to the zone system. We distribute these jobs on zones based on information about the workplace municipality, which is (almost) always available.[1]

[1] Only for 930 jobs in 2010, there is no information available about the workplace municipality. Hence, these jobs are not counted.

In line with Combes et al. (2008) we assume that changes in wages reflect changes in worker productivity. The argument is that in a competitive equilibrium wages should equal marginal productivity. We apply the following first difference linear regression of the logarithm of gross hourly wages on the logarithm of the accessibility defined at the zone of employment:

$$\ln(w_{nz\,t+1}) - \ln(w_{nzt}) = \delta \cdot (\ln(A_{z\,t+1}) - \ln((A_{zt})) + \beta'(x_{z\,t+1} - x_{nt}) + (\theta_{t+1} - \theta_t) + (\varepsilon_{zit+1} - \varepsilon_{zit})$$

w_{nzt} is the gross hourly wage for worker n in zone z at time t. A_{zt} is the accessibility for zone z at time t. x_{nt} is a vector of individual specific characteristics for worker n at time t. u_n is the worker fixed effect for worker n. θ_t is a time dummy, and ε_{nzt} is the error term. As we measure the accessibility at the zone level, we need to cluster the standard errors at the level of $(work\ zone_{nt}, work\ zone_{n\,t+1})$. With a first difference model all time invariant variables are controlled out, whether they are observable or not. Due to sorting effects, a pooled regression gives substantially higher elasticities.

The accessibility for a zone measured in 2002 is highly correlated with the measured accessibility in 2010. To obtain the required variation in the data, it is necessary to narrow the sample down to individuals working in two different work zones in 2002 and 2010. This is our first approach. The selection introduces a potential matching effect; those who are most likely to accept a job offer may well be those who get the largest wage increase.

To mitigate the matching effect and the sorting bias from firms, we restrict our sample to those who are changing jobs due to a plant relocation. With this second approach, we limit the potential bias from more productive workers getting better wage offers. While bias from the workers is reduced, we are instead exposing us to a selection bias stemming from more productive firms moving into denser areas. In order to limit this potential bias, we add an additional vector of changes in observable firm characteristics in the wage regression.

Finally, we use a third approach, where we estimate the agglomeration effect in a two-step approach following Combes et al. (2008). The first stage is similar to the estimated wage regression above except that we do not include the agglomeration measure. Instead, we include zone dummies with the aim of regressing these on the agglomeration measures in the second stage.

We find that our results are surprisingly robust to the different estimation strategies. The three approaches yield similar results in terms of the estimated elasticity of wages w.r.t. agglomeration, which is around 2%. Based on this, we argue that it is reasonable to use an elasticity of 2% for the accessibility effect on wages. The three different estimation strategies expose us to three different potential biases, so the fact that the results are robust make us conclude that we with a first difference estimator control for most of the potential bias.

To better understand and describe the agglomeration effects, we then investigate how the elasticity differ for different sectors and sociodemographic groups. The estimated elasticity of productivity w.r.t. agglomeration is somewhat larger in the service sector than in manufacturing. The corresponding decay parameter is also found to differ between the respective sectors.

References

Combes, P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: Sorting matters! Journal of Urban Economics 63, 723-742.

Graham, D., Gibbons, S., 2018. Quantifying Wider Economic Impacts of Agglomeration for Transport Appraisal: Existing Evidence and Future Directions. Technical report for the Department for Transport.

Venables, A. J., 2007. Evaluating urban transport improvements: cost-beneit analysis in the presence of agglomeration and income taxation. Journal of Transport Economics and Policy 41 (2), 173-188.