# Agglomeration, Transportation and the Quality of Life

**Pedersen, Jesper Hybel**

*Publication date:*
2020

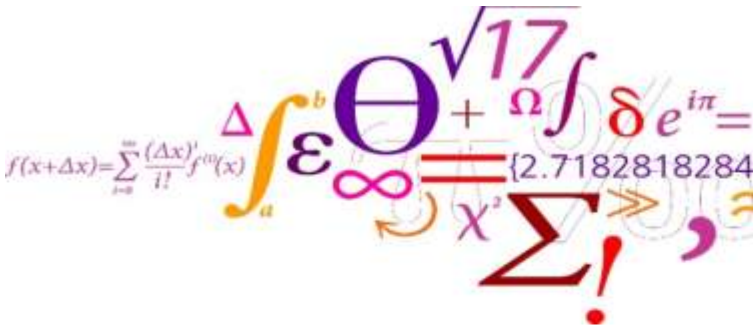
*Document Version*
Publisher's PDF, also known as Version of record

# Agglomeration, Transportation and the Quality of Life

Jesper Hybel Pedersen

PhD Thesis
05-2020

DTU Management
Technical University of Denmark



| | |
|---|---|
| Title: | Agglomeration, Transportation and the Quality of Life |
| Type: | PhD thesis |
| Date: | 14-05-2020 |
| | |
| Author: | Jesper Hybel Pedersen |
| | |
| Supervisors: | Ismir Mulalic and Ninette Pilegaard |
| | |
| University: | Technical University of Denmark |
| Department: | DTU Management |
| Division: | |
| Address: | Akademivej Building 358<br>DK-2800 Kgs. Lyngby<br>www.man.dtu.dk |
| Cover picture: | Colourbox.com |

# Summary (English)

The main objective of this thesis is to investigate the economic consequences of the agglomeration of economic activity. The concentration of economic activities is important because of its many costs and benefits that shape the location choices of economic actors and thereby the shape of modern cities.

The concentration of economic activity is associated with gains in productivity. This creates high rewarding labour markets, access to which are particularly important for workers. The changes in labour market outcomes, associated with workers choices to relocate, from one local labour market to another, reveal the potential magnitude of the productivity gains of agglomeration. This thesis aims to quantify the size and scope of these static productivity gains of agglomeration, that accrue to workers immediately upon relocation.

High density areas usually offer workers high quality carrier opportunities and access to institutions of higher education. While the static part of the gains of agglomeration can be gained by workers immediately upon relocation, part of the gains are only achieved over time. The value of the experience workers acquire is therefore likely to vary depending on the density of the area where the experience is acquired. The other objective of this thesis is to quantify the magnitude and explore the heterogeneity of these dynamic agglomeration effects.

To gain easy access to high rewarding and dense labour markets workers often prefer to live close to these industrious areas. The high demand for housing in the areas of dense economic activity results in high housing prices, which deters further concentration. The workers who choose to live outside the urban

centers can usually enjoy lower housing prices but must either suffer a longer daily commute to the urban centers or accept lower wage in the periphery. An important part of this dual location choice of where to live and where to work are the local amenities often pictured as access to beaches, clean air or high quality public services. To quantify the value of these amenities this thesis constructs the Quality of Life (QOL) index that measure the representative household's willingness-to-pay for the local amenities and use the index to investigate the importance of the commuting costs and transportation for households quality of life.

Urban centers do not only function as centers of employment but also as providers of large markets with a high variety of goods and services. Access to such markets are important for consumers and often requires a trip by car to the city. Access to the market involves the cost of parking which consist of both observed monetary price as well as the often unobserved costs of searching for a parking spot. This thesis provides estimates of the parking demand elasticity for Copenhagen while taking into account the presence of unobserved search costs.

# Summary (Danish)

Målet for denne afhandling er at undersøge de økonomiske konsekvenser af agglomerationen af økonomisk aktivitet. Koncentrationen af økonomisk aktivitet er vigtig grundet sine mange fordele og ulemper, der præger økonomiske aktørers lokalitetsvalg og dermed formen af moderne byer.

Koncentrationen af økonomisk aktivitet er blandt andet forbundet med produktivitetsgevinster. Dette skaber lokale arbejdsmarker med høje lønninger og adgangen til disse er særligt vigtig for lønarbejdere. Forandringer i løn forbundet med skift af ansættelse på ét lokalt arbejdsmarked til et andet, afslører den potentielle størrelse af de underliggende agglomerationsgevinster. Denne afhandling kvantificerer størrelsen og omfanget af disse statiske produktivitetsgevinster, der tilkommer lønarbejdere umiddelbart ved skift i ansættelse.

Områder med høj befolkningstæthed tilbyder karrieremuligheder af høj kvalitet samt adgang til højere læreanstalter. Dette betyder, at selvom den statiske del af agglomerationsgevinsterne tilkommer lønarbejdere umiddelbart når de skifter ansættelse til et område med høj befolkningstæthed, så tilfalder en vigtig del af agglomerationsgevinsterne kun lønarbejderne over tid. Værdien af den erfaring lønarbejderne akkumulerer, afhænger potentielt af befolkningstætheden de steder, hvor de arbejder. Det andet mål for denne afhandling er at estimere størrelsen af disse dynamiske agglomerationsgevinster samt analysere deres heterogenitet.

For at opnå let adgang til arbejdsmarkederne med høj løn foretrækker lønarbejdere som regel at bo tæt på, hvor de arbejder. Den høje efterspørgsel på boliger i områderne med høj erhvervsaktivitet resulterer i høje boligpriser, hvillket af-

skrækker yderligere koncentration af den økonomiske aktivitet. De lønarbejdere, der vælger at bosætte sig udenfor de urbane centre, nyder som regel godt af lavere boligpriser, men må som følge af deres bosætningsvalg acceptere enten længere pendlingstider til det urbane center eller lavere løn forbundet med at arbejde på det lokale arbejdsmarked. En vigtig faktor i dette duale lokalitetsvalg om, hvor man skal bosætte sig og hvor man skal arbejde, er de ikke markedsførte lokale goder såsom adgang til strand, ren luft samt niveauet af offentligt leverede goder og services. For at kvantificere værdien af disse ikke markedsførte lokale goder udarbejdes i denne afhandling et livskavlitetsindeks, der måler den repræsentative husholdnings marginale betalingsvillighed for de ikke markedsførte lokale gode. Indekset anvendes endvidere til at undersøge betydningen af transport og pendlingsomkostninger for husholdsningers livskvalitet.

Urbane centre fungerer ikke kun som lokale arbejdsmarkeder, men også som markeder, hvor forbrugere kan få adgang til en bred vifte af forbrugsgoder. Adgang til de urbane markeder kræver ofte en rejse med bil til den nærmeste storby. Prisen for denne adgang inkluderer derfor parkeringsomkostninger, hvori indgår både en observeret parkeringsafgift, samt en uobserveret omkostning i form af tid anvendt på at søge efter en parkeringsplads. Denne afhandling estimerer elasticiteten af efterspørgslen på parkering i København, hvor der tages højde for forekomsten af uobserverede søgeomkostninger.

# Acknowledgements

I would like to thank my master and former supervisor Ismir Mulalic whos will to knowledge made this thesis possible. I would also like to thank my current supervisor Ninette Pilegaard for all her help. I would like to thank Mogens Fosgerau for accepting me as a Ph.D.-student on the Urban Project and for all the good sessions on mathematical topics. I would like to thank my coauthors Malthe Borghorst, Albert Gragera and Edith Madsen. Finally offcourse I would like to thank my son and girlfriend for being there and being supportive.

I would also like to thank the Innovation Fund Denmark for the financial support via the Urban Project.

# Contents

# 1 Introduction and Summary

Denmark and many other countries experience a process of urbanization. Part of this process is the spatial clustering of economic activities. To explain the concentration of activities from an economic perspective, the existence of agglomeration benefits has been hypothesized (Duranton and Puga 2004). Agglomeration benefits increase the economic efficiency when firms and households choose to locate near to each other. The colocation choices concentrates the economic activities and causes cities to grow. The growth of cities are however counterbalanced by disintegrative forces such as increasing housing prices and transportation costs (Ahlfeldt and Pietrostefani 2019). Due to the importance of these economic costs and benefits, many government policies aims to control the location choices of firms and households. Such policies include among others zoning policies to restrict land use, location-specific tax incentives, and transportation subsidies to foster further concentration of economic activities.

My Ph.D. project aims to contribute to this policy making process by providing a better understanding of certain aspects of the consequence of the concentration of economic activities. The thesis consist of four seperate papers. The first two papers quantify respectively the static and dynamic benefits of agglomeration. The third paper quantify the spatial differences in the quality of life. The fourth and final paper estimates the parking demand elasticity to the full cost of parking. The Ph.D. project is part of the Innovation Fund Denmark project URBAN that studies urbanization, productivity and congestion.

Transportation investments are standardly evaluated using the methods of conventional Cost Benefit Analysis. These methods assume perfectly competitive markets such that all costs and benefits of transportion investments are captured on the markets for transportation. Using a stylized monocentric city model, Venables (2007) argues that transportation investments are an important driver of increases in agglomeration economies. Given the existence of agglomeration economies, the transportation investments may trigger increases in economic effeciency not captured directly on the market for transportation. Such externalities are standardly referred to as the Wider Economic Impacts (WEI) and because they are additional to the costs and benefits apparent on the markets for transportion, they represent a challenge for the conventional CBA-methods (Graham and Gibbons

2019). Quantification of the agglomeration benefits are therefore important for urban planners and politicians. They add to the correct evaluation of the transporation investments. The standard measures of the agglomeration economies are the elasticity of agglomeration and the spatial decay of agglomeration.

In the first paper of my Ph.D. I estimate the elasticity of agglomeration and the spatial decay of agglomeration using a panel data set for the full population of workers in Denmark for the years 2008-2016. The basic identification strategy is based on the assumption that as workers change their employment area, the effect of the area specific level of employment density on the productivity manifests itself in the workers wages. The elasticity of agglomeration can therefore be measured as the elasticity of wage with respect to the employment density. It is well known from the litterature on sorting that individual workers are not randomly distributed across space (Kuminoff, Smith, and Timmins 2013). Workers with different characteristics relevant for labour market outcomes such as wage choose to live and work in certain areas. I control for sorting on observable and unobservable worker heterogeneity. Using the panel data methods I find the elasticity of agglomeration estimates to be a 0.01 log point increase in wage for a log point increase in the employment density. Moreover the strength of the spillover declines with 50% for every 6 minutes of travel time.

The gains in wages achieved by workers immediately when changing their location of employment are called static agglomeration benefits. Removing the effects of worker characteristics remains important to identify the productivity effects of locations. However, to assume that the total gains in productivity are captured solely by the static productivity effects of locations, ignores agglomeration benefits that arise over time. These dynamic benefits are generated by the interaction between the location and the individual worker. To allow for such dynamic effects, De La Roca & Puga (2017) estimate an econometric model that allows for the experience of workers to be location specific. Using a sample of male worker in Spain they find that the experience accumulated in cities have a higher wage payoff anywhere. High density areas provide better learning environments for workers. Hirsch et. al. (2013) argue theoretically that in more dense labour markets competition is stronger and this constrain employers' ability to discriminate against women. Using data for workers in western Germany they show that the gender wage gap is lower in large metropolitan than in rural areas.

In the second paper of my Ph.D. I use the econometric model suggested by De La Roca & Puga (2017) to estimate the dynamic wage effects of location for both the male and the female workers. I find that the dynamic effects of location are important for both the male and the female workers. I also find that the male workers compared to female workers have a higher additional wage benefit of the experience used and accumulated in the high density areas. Finally I find that the dynamic gains of location are important for the estimate of the elasticity of agglomeration.

Areas of high employment density provide both static and dynamic wage benefits to workers. Access to these areas are therefore very important for workers. However, workers who choose to live in the city might not be better off in terms of the quality of life because they have to pay a higher cost of living. The higher cost of living are primarily composed of higher levels of housing prices. Workers who decide to live outside the urban centers must either accept the longer commute to the work in the urban center or accept the lower wages of the rural area. The dual choice of where to live and where to work is therefore a complicated exercise in balancing these tradeoffs. The litterature on sorting further points to the existence of local amenities as being important for the quality of life achieved by the residents of a given area (Kuminoff, Smith, and Timmins 2013). In the spatial equilibrium a high level of locally accessible earnings and low housing prices compensate individuals for a low quality level of remaining local amenities (Rosen 1979 & Roback 1982). The transportation infrastructure play a key role in making it possible for individuals to live in one place and work in another, albeit at a certain cost.

In the third paper of my Ph.D. I follow Albouy and Lue (2015) and construct a transport adjusted quality of life index for the 98 municipalities covering Denmark in order to quantify the importance of transportation costs for the quality of life. The quality of life index is typically high in the Greater Copenhagen Area and other large cities in Denmark. My empirical findings suggest that the quality of public transport system is important for households quality of life and that households prefer to separate the workplace locations from the residence locations that induces commuting.

Local governernments around the world regulate city access and levels of urban congestion using parking policies. A common component of these policies are price regulations. It is therefore important to quantify the sensitivity of the demand for on-street parking to the cost of parking. The full cost of parking includes however the cost of searching for a vacant parking space in addition to a parking fee. The search cost of cruising is usually unobserved and ignoring this leads to bias when estimating the elasticity of the demand for on-street parking (Inci, Ommeren, and Kobus 2017 & Zakharenko 2016).

In the fourth and final paper of my Ph.D. I demonstrate that, even when the cost of cruising is unobserved, the demand elasticity can be identified by extending an econometric model to include the spatial interaction between the parking facilities. The research design and method used is based on the working paper by Madsen et. al (2013). I estimate the model and the parking demand elasticity using data from Copenhagen to test if there exist a significant cruising bias. I find a 55% larger parking demand elasticity to total cost than to parking fees. This result suggests that there is a significant cruising bias in the usually reported estimates of the parking demand elasticity.

## References

Ahlfeldt, Gabriel M. and Elisabetta Pietrostefani (2019). "The Economic Effects of Density: A Synthesis". In: *Journal of Urban Economics* 111, pp. 93–107.

Albouy, David and Bert Lue (2015). "Driving to Opportunity: Local Rents, Wages, Commuting, and Sub-Metropolitan Quality of Life". In: *Journal of Urban Economics* 89, pp. 74–92.

Duranton, Gilles and Diego Puga (2004). "Chapter 48 Micro-Foundations of Urban Agglomeration Economies". In: *Handbook of Regional and Urban Economics.* Vol. 4. Elsevier, pp. 2063–2117.

Graham, Daniel J. and Stephen Gibbons (2019). "Quantifying Wider Economic Impacts of Agglomeration for Transport Appraisal: Existing Evidence and Future Directions". In: *Economics of Transportation* 19, p. 100121.

Hirsch, Boris, Marion König, and Joachim Möller (2013). "Is There a Gap in the Gap? Regional Differences in the Gender Pay Gap". In: *Scottish Journal of Political Economy* 60.4, pp. 412–439.

Inci, Eren, Jos van Ommeren, and Martijn Kobus (2017). "The External Cruising Costs of Parking". In: *Journal of Economic Geography* 17.6, pp. 1301–1323.

Kuminoff, Nicolai V., V. Kerry Smith, and Christopher Timmins (2013). "The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets". In: *Journal of Economic Literature* 51.4, pp. 1007–1062.

Madsen, Edith, Ismir Mulalic, and Ninette Pilegaard (2013). *A Model for Estimation of the Demand for On-Street Parking*. MPRA Working Paper 52301.

Roback, Jennifer (1982). "Wages, Rents, and the Quality of Life". In: *Journal of Political Economy* 90.6, pp. 1257–1278.

Roca, Jorge De La and Diego Puga (2017). "Learning by Working in Big Cities". In: *The Review of Economic Studies* 84.1, pp. 106–142.

Rosen, Sherwin (1979). "Wages-Based Indexes of Urban Quality of Life". In: *Current Issues in Urban Economics*. Ed. by P. Straszheim and M. Mieszkowski. Baltimore: John Hopkins University Press.

Venables, Anthony J. (2007). "Evaluating Urban Transport Improvements: Cost-Benefit Analysis in the Presence of Agglomeration and Income Taxation". In: *Journal of Transport Economics and Policy* 41.2, pp. 173–188.

Zakharenko, Roman (2016). "The Time Dimension of Parking Economics". In: *Transportation Research Part B: Methodological* 91, pp. 211–228.

# The strength and scope of agglomeration in Denmark

Jesper Hybel*

**Abstract**

I estimate the strength and scope of agglomeration in Denmark using the full population of workers for the years 2008-2016 controlling for observed and unobserved worker heterogeneity including education. The elasticity of agglomeration estimates to a 0.01 log point increase in wage for a log point increase in the employment density and the strength of the spillover declines with 50% every 6 minutes of travel time. This implies that doubling labour market's employment density raises hourly earnings by nearly 0.7%.

**Keywords:** Agglomeration, Wages, Productivity, Transportation.

**JEL codes:** R12, R4.

*Technical University of Denmark, DTU Management, Bygningstorvet, Building 116, DK-2800 Kgs. Lyngby, Denmark, email: jeshyb@dtu.dk. Jesper Hybel is supported by the Innovation Fund Denmark via the Urban Project.

# 1 Introduction

Firms located in areas with a high level of economic activity are on average more productive (Combes et al. 2012b). But how far reaching are the effects of density on productivity? Quantifying the productive advantages of high density areas involves both measuring how much firms on average gain in productivity when locating in areas of high economic activity, as well as how close firms must be to a high density area in order to reap the gains in productivity (Graham, Gibbons, and Martin 2010). The productivity effect is in general assumed to decrease with distance and is therefore referred to as the *decay* of agglomeration. This paper estimates the agglomeration elasticity and the decay of agglomeration using a panel data set for the full population of workers in Denmark in the period 2008-2016.

The exact quantification of the effects of density is important because many local public policy initiatives, such as for example transportation investments, attempts to foster the agglomeration of economic activity in order to bring about efficiency gains or create *compact city form* (Venables 2007 & Holman et al. 2015). While the theoretical and empirical literature seems to have reached a consensus about the existence of productivity gains from agglomeration the exact policy implications are far less one-sided. The concentration of economic activity have complex consequences, including not only the effect of increasing productivity. Combining these effects, in a complete welfare calculation, are at present stage of research an exercise in "admittedly imperfect accounting" (Ahlfeldt and Pietrostefani 2019, p. 94). Specifically for the field of transport economics the *wider economic benefits* of transport investments include the agglomeration effects but the "extent to which direct user and wider impacts can be calculated separately in practice is open to question." (Graham and Gibbons 2019, p. 4).

Theoretically there are good reasons to believe that economies of agglomeration must exist because the increasing returns to scale form a central role in the economic explanation of the existence of cities. The core argument offered within Urban Economics and New Economic Geography is due to Starrett (1978) and the *spatial impossibility theorem.* This theorem states that the existence of cities in a competitive equilibrium is incompatible with the joint assumption of homogenous space, abscence of indivisibilities, abscence of increasing returns to scale and positive transport costs.

Space is clearly not homogenous and therefore cities may exist simply due to exogenous variation in the occurence of natural endownments such as minerals, navigable rivers, amicable weather etc. While the natural endowments are accepted as playing a causal role in bringing about cities, space is nevertheless assumed to be homogenous in the theoretical literature of New Economic Geography (Ottaviano and Thisse 2004, p. 2571). This assumption is due to the theoretical interest of identifying economic mechanisms that explain agglomeration without appealing to the physical attributes of locations. The empirical literature on economies of agglomeration has largely treated natural endownments as unobserved factors to be controlled for in the empirical models when testing for the existence of gains of agglomeration (Combes and Gobillon 2015). The empirical approach applied in this study will use a similar strategy.

While spatial heterogeneity is partly a matter of natural endowments a second source is arguable manmade. City development involves large investments in *shared* infrastructure facilities to provide goods such as water, electricity, garbage removal, public transportation, public health services, etc. Because production involves sharing of costly indivisibilities, production for a larger number of consumers requires less than proportional increase in production costs (Scotchmer 2002). While useful as a modelling device, the assumption of such large indivisibilities is however critizable for not tackling the issue of what gives rise to increasing returns (Duranton and Puga 2004). The theoretical ambition must according to Puga & Duranton (2004) be to provide micro founded theories showing how increasing returns to scale emerge at the level of the city due to mechanisms working on the micro level. They suggest that these mechanisms are either *sharing*, *matching* or *learning* giving rise to outcomes that are in most respects observationally equivalent. Specifically they all explain why a high concentration of economic activity goes hand in hand with higher levels of productivity. This paper focuses in particular on these mechanisms, but does not aim to distinguish between them.

An extensive part of the empirical work on agglomeration has largely focused on testing for the existence as well as quantifying the size of the impact of agglomeration on productivity by estimating the elasticity of productivity with respect to agglomeration. Early applications arrived at relatively high estimates of 5.6% for the U.S. and a slightly lower 5.0% for European countries (Ciccone and Hall 1996 & Ciccone 2002). These early studies used aggregated data and argued that unobserved natural endownments at the aggregate level should be assumed correlated with the density due to the feedback effect. The natural endownments would increase wages and higher wages would in return attract more workers therefore increasing density (Ciccone and Hall 1996, p. 61). To consistently estimate the agglomeration elasticity density was instrumented using historical instruments in Ciconne and Hall (1996) and later with the land area of the geographical units in Ciconne (2002). The argument offered in support of the historical instruments was that the natural endownments affecting employment densities of the past are no longer relevant for production, while the past densities remain correlated with present day densities due to inertia. The particular geographical instruments suggested in Ciconne (2002) seems not to have found much acceptance perhaps due to reasons mentioned in Combes et. al. (2015, p. 300). Nevertheless both geographical and historical instruments are today part of the methodological toolbox commonly applied in estimating the elasticity of agglomeration to control for the feedback effect (see for example Combes et. al. (2010) and Roca and Puga (2017)). The prescence of such a feedback effect has been tested by Graham et. al (2010) finding that agglomeration economies are not strictly unidirectional. On the other hand, Melo and Graham (2010) conclude, based on a meta-analysis, that correcting for reverse causality is of minor importance for the size of urban agglomeration estimates. This paper uses historical instruments to control for the endogeneity of employment density due to reverse causality.

To measure productivity the empirical literature has primarily either used data on firms or workers. For firms productivity is commonly measured as output value or value added while wages have been used for workers (Combes and Gobillon 2015). The use of wages as a measure of productivity is often justified by the assumption of competitive markets labor in which case labour is paid the value of its marginal product. However even if labor markets are not perfectly competitive, firms must be capable of paying the higher wages. If more dense areas have higher wages but no gains in productivity, then the firms in tradable sectors, would choose to relocate to less dense areas with lower wages (Enrico 2011, p. 1249). Hence even in the abscence of perfectly competetive labour markets the higher wages in dense urban areas can still be seen as evidence of higher productivity (Puga 2010).

Improving on the use of aggregated data later empirical studies turned to the use panel data in order to take into account firm or worker heterogeneity. Given that firms and workers are heterogenous with respect to some individual level productivity enhancing features, some areas may be more productive than others, simply because they have a larger share of highly productive workers or firms. Glaeser and Maré (2001) argued that high skill workers have stronger preferences for the aminities offered by big cities to explain the larger share of highly educated individuals in the more dense areas. Using a panel of workers to control for both observed and unobserved individual level skills Glaeser and Maré (2001) found gains in productivity for large cities, but they did not estimate the agglomeration elasticity because of data limitations. Using French wage data Combes (2008) estimated the agglomeration elasticity taking individual fixed effects into account. The elasticity was measured to be 2.1%, half of that obtained when individual heterogeneity is not controlled for. Puga and De La Roca (2017) estimate the elasticity for Spain to 1.7% also using wage data and controlling for sorting of individuals. Overman and D'Costa (2014) find an elasticity of 1.1% for the United Kingdom again using wage data. DeBorger et. al. (2019) find an elasticity of 0.3% for Denmark using a quasi-natural experiment. This paper will uses a panel of workers and controls for workers observed and especially their unobserved skills, i.e. worker sorting.

The empirical literature is large and with the estimates of the agglomeration elasticity differing across countries and with the exact econometric specification (see Melo, Graham, and Noland 2009 & Rosenthal and Strange 2004). The estimates are nevertheless predominantly positive with the unweighted mean elasticity from 47 international empirical studies being 3.2% (Graham and Gibbons 2018). Based on these studies it seems that a consensus concerning the existence of agglomeration effects have emerged.

This paper estimates the elasticity of agglomeration for Denmark to be 0.0107 log point increase in wage for a log point increase in effective density and the scope is estimated to have a half-time of approximately 6 minutes of transportation time. The elasticity of agglomeration is small in comparison to estimates from other European countries and the US. This paper also finds that controlling for unobserved worker heterogeneity is of primary importance for the estimate of the agglomeration elasticity while less important for the estimates of the scope of the agglomeration effects. While of secondary importance, controlling for education is found to be important, for the estimation of the elasticity of agglomeration even when using individual fixed effects.

The rest of this paper is organized as follows. In Section 2 a stylized model is set up to motivate an econometric specification and guide the discussion of endogeneity. Section 3 then introduces the data used for estimation and central descriptive statistics for the spatial units. Section 4 presents the estimation technique and the identification strategy. Section 5 reports the estimation results and finally section 6 concludes.

# 2 The theoretical model

In this section I present a stylized model and derive from it the Mincerian wage equation used for estimation of the agglomeration effect. The stylized model is presented in subsection 2.1 and in subsection 2.2 I discuss the endogeneity concerns related to worker skills, unobserved local endowments and feed back effects.

## 2.1 Stylized model

To motivate the reduced form empirical approach and support interpretation I introduce a stylized model. Firms at each location $c$ are assumed to use a Cobb-Douglas production technology. The units produced $Y_c$ are given as

$$Y_c = A_c(N_c)\left(\frac{\sum_{i\in c} s_i l_i}{b}\right)^b \left(\frac{K_c}{1-b}\right)^{1-b}, \tag{1}$$

where $s_i$ represents the skills and $l_i$ represents the hours of labour of the individual worker $i$, $A_c$ is the city $c$ specific local productivity shifter and $K_c$ is capital or more generally simply other factors entering production. Under perfect competition the profit can then be written as

$$\Pi_c = P_c Y_c - \sum_{i\in c} w_i l_i - R_c K_c, \tag{2}$$

where $P_c$ is the output price, $R_c$ is the price of capital and $w_i$ is the wage. Firm maximize profit with respect to $l_i$ and $K_c$ implying that the wage of individual worker $i$ is given as

$$w_i = s_i \left( \frac{P_c A_c(N)}{R_c^{1-b}} \right)^{1/b} = s_i B_c \tag{3}$$

where $B_c := (P_c A_c / R_c^{1-b})^{1/b}$. This simple model does not incorporate the micro founding mechanisms of sharing, matching and learning as more complex models, for example by Duranton and Puga (2004). However it leads to equilibrium wage equations of the same form, with the local wages depending on the local productivity shifter $A_c(N_c)$, which is an increasing function of the local employment population $N_c$. Because the different micro mechanisms are observationally equivalent with respect to the higher density $N_c$ increasing the productivity $A_c(N_c)$, the model does not distinguish between to what extent one mechanism rather than another are driving the increases in the productivity.

To operationalize the model I allow for variation over time and assume that individual skills are given as

$$s_{it} = \exp(\mathbf{x}_{it}^\top \beta + \mu_i + \epsilon_{it}), \tag{4}$$

where $\mathbf{x}_{it}$ are observable skills such as experience, tenure, education and occupation, while $\mu_i$ is unobservable individual specific time constant skills and $\epsilon_{it}$ is measurement error. The productivity shifter $B_c$ is allowed to vary by time $t$ and sector $k$ assuming that

$$B_{kct} = \exp(\mathbf{z}_{ct}^\top \lambda + \eta_{ct} + \phi_k + \delta_t), \tag{5}$$

where the increasing returns to scale is captured by allowing for one of the area specific covariates $\mathbf{z}_{ct}$ to be a measure the local employment density. While $\mathbf{z}_{ct}$ is assumed observed, the factors $\eta_{ct}$ capture all factors relevant for production that are heterogenous across space and time. I include such unobserved factors $\eta_{ct}$ because it is unlikely that the area specific covariates $\mathbf{z}_{ct}$ succesfully control for all factor relevant for production and varying across locations.

Finally to allow for spatial spill-over effects, the *effective density* of area $c$ denoted $z_{ct}$ is assumed to take the market potential form (Harris 1954) being an exponentially weighted average

$$z_{ct}(\alpha) := \sum_{a \in \mathcal{C}} d_{at} \exp(-\alpha \cdot T(c,a)), \tag{6}$$

where $d_{at} = L_{at}/S_a$ is the local employment density measured as the number of local workers $L_{at}$ divided by the area $S_a$ of the local unit and $T(a,b)$ is the transportation time from area $a$ to $b$ and back. While there is no a priori argument for this functional form it does enjoy popularity in the theoretical (Fujita and Ogawa 1982 & Jr. and Esteban Rossi-Hansberg 2002) and empircal (H. Hanson 2005 & Ahlfeldt et al. 2015) literature.

Inserting Equation (4) and (5) into Equation (3) and taking logs implies that log wage of the invidual worker at time $t$ is given as

$$\log w_{it} = \mu_i + \mathbf{x}_{it}^\top \beta + \mathbf{z}_{ct}^\top \lambda + \eta_{ct} + \phi_k + \delta_t + \epsilon_{it}, \tag{7}$$

where it is ignored for notational convenience that one of the local variables $\mathbf{z}_{ct}$ is a function of the parameter $\alpha$ if the effective density is included.

I am primarily interested in the agglomeration elasticity $\lambda_l$. Estimation of the wage equation is however challenging due to potential endogeneity which I discuss and offer a solution to in the next section.

## 2.2 Endogeneity

The sources of endogeneity can best be illustrated by considering estimation of a simple model that includes only employment density and a global time trend $\delta_t$. The estimation equation is then given as

$$\log w_{it} = z_{ct}\lambda + \delta_t + u_{it}.$$

Because the estimation equation ignores individual skills $\mu_i + \mathbf{x}_{it}^\top \beta$, sector fixed effects $\phi_k$ and unobserved productivity enhancing factors $\eta_{ct}$, it follows that the error term contains these factors, i.e. $u_{it} = \mu_i + \mathbf{x}_{it}^\top \beta + \eta_{ct} + \phi_k + \delta_t + \epsilon_{it}$. The Ordinary Least Squares estimator will therefore be inconsistent in case the employment density $z_{ct}$ is correlated with any of these factors

$$Cov(z_{ct}, u_{it}) = Cov(z_{ct}, \mathbf{x}_{it}^{\top}\beta + \mu_i + \eta_{ct} + \phi_k) \neq 0. \tag{8}$$

The first potential bias arises due to correlation between skills $\mathbf{x}_{it}^{\top}\beta + \mu_i$ and the employment density $z_{ct}$. If high skilled workers sort into high density areas, wage will be higher due to their higher skill. Because skill level is not controlled for, the estimate of $\lambda$ will falsely attribute the skill-effect to density and therefore be positively biased. This hypothesis finds support in workhorse theories of urban economics (Roback 1982 & Brueckner and Zenou 1999). They argue that the urban space is heterogenous due to the uneven distribution of consumer amenities. These consumer amenities are important for households residential choices. Firstly because areas with higher levels of amenities attract more workers and thereby increase density and productivity. This is not in itself a problem because the effect from amenities on productivity here is assumed to be channeled via density (Combes, Duranton, and Gobillon 2011). However there is a wide literature on residential sorting documenting that households have different preferences for the consumer amenities (see Kuminoff, Smith, and Timmins 2013 for a review). In extension of this literature Glaeser and Maré (2001) suggest that individual workers with high levels of skill particularly favor the amenities offered by high density areas. As a consequence, part of the explanation for the high wages of high density areas may therefore be the purely compositional effect arising because the high skilled workers to a larger extent choose to live in these areas and therefore work in the surrounding areas to minimize transportation costs. Formally, the problem is that $Cov(z_{ct}, \mathbf{x}_{it}^{\top}\beta + \mu_i) > 0$ underlining the importance to control for both observed and unobserved abilities as suggested Glaeser and Maré (2001).

The second source of endogeneity is a non-zero correlation $Cov(z_{ct}, \eta_{ct}) \neq 0$ between the unobserved city effects $\eta_{ct}$ and the employment density $z_{ct}$. Such a correlation could occur if consumer amenities also affect producer amenities entering $\eta_{ct}$ directly, besides affecting employment density by attracting labour. The prescence of an airport could both attract residents due to easier access to vacational destinations, as well as affect the productivity of firms by increasing access to international labour or by decreasing input prices. Since the consumer amenities affect both the employment density as well as the unobserved productive factors, it is likely that $Cov(z_{ct}, \eta_{ct}) \neq 0$. The general direction of bias is however unclear. It is imaginable that $Cov(z_{ct}, \eta_{ct}) > 0$ as with for example crime, where higher levels of crime assumedly is both a unattractive for residents as well as for local producers. But it is also imaginable that $Cov(z_{ct}, \eta_{ct}) < 0$ as with high levels of pollution being unattractive for residents but perhaps enabling firms to produce with lower costs (Combes, Duranton, and Gobillon 2011, p. 256 & Storper and Scott 2008).

The third source of endogeneity in the simple specification arises due to the lacking control for sector specific effects $\phi_k$. High density areas could have a particulary large share of plants in a specific sector that for different reasons could be more productive. This type of endogenity could also be related to firms choice of location where particularly productive firms choose to locate in high density areas. Moreover Combes et. al. (2012) show that firm selection cannot explain spatial productivity differences for the case of France.

A final and fourth source of endogeneity arises due to the feedback effect wages could have on density. High wage areas are more attractive places to work and it is therefore likely that there is a positive feedback from high wages to density reinforcing agglomeration. To put it differently, agglomeration economies improve economic performance and economic performance reinforces agglomeration, making the latter endogenous. The prescence of such a feedback effect has been tested for by Graham et. al (2010) finding that agglomeration economies are not strictly unidirectional.

# 3 Data and descriptives

I use an unbalanced panel data set with a total of 7.2 million observations for the full population of 1.2 million employees in Denmark during the period 2008-2016. The observations are registered on a yearly basis and for each individual worker the dataset contains information on the *nominal hourly wage* along with covariates such as *experience*, *jobtenure*, *education* and *occupation* as well as the workers place of work identified on a spatial scale of the 98 municipalities covering Denmark. The micro panel data set is combined with a dataset for the municipalities that includes observations on *geographical size*, as well as *historical population counts* for the year 1801 and 1834 and finally *transportation time* between municipalities anno 2010.

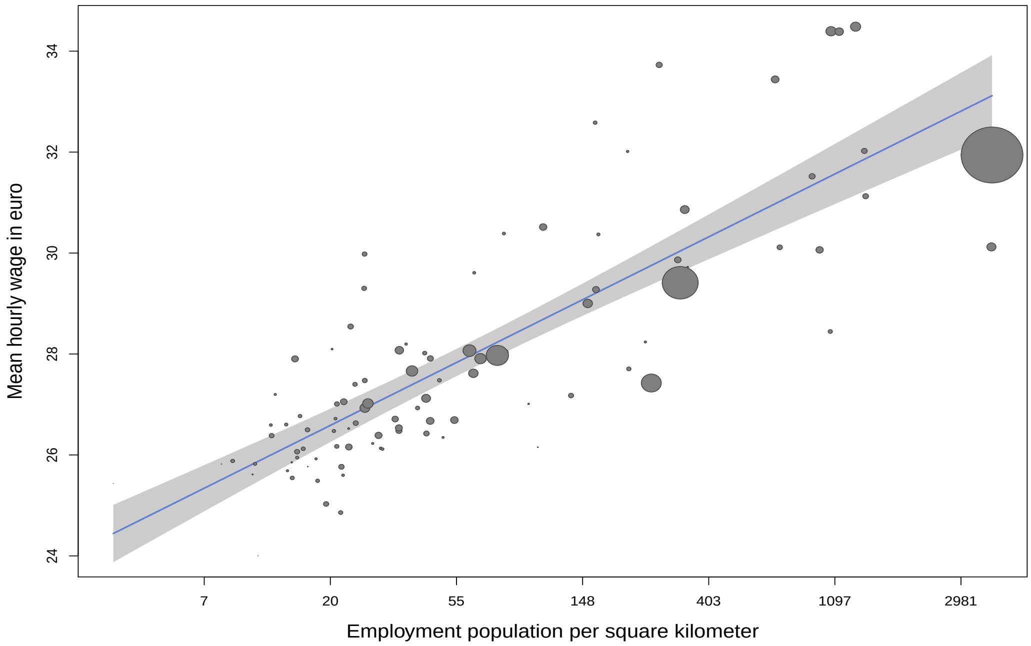
The spatial units used in this paper are the 98 municipalities covering Denmark. The Danish municipalities are relatively small with a median size of 360 km$^2$ and an average size of 438 km$^2$. The median size is equivalent to a circle with 10.2 km radius which is very close to the median commuting distance of 9.8 km anno 2013.

The choice of spatial units are important because using the same econometric specification with different designs of spatial units can change the estimates. Using French data Briant et. al (2010) show that when the units are small, the choice of size is of secondary importance to the econometric specification and the choice of the shape of units is of tertiary importance. In their paper *small* units refer to areas with a size comparable to the 341 French employment areas having an average size of 1500 km$^2$ (Combes et al. 2012a, p. 914). This suggests that that using municipalities as spatial units should not be expected to affect the results more than changes in the econometric specification. This should be kept in mind when comparing the different estimates of the agglomeration elasticity across specifications.

The areas are grouped according to their economic activity measured as employment density. Employment population is usually preferred to residential population because it better reflects the magnitude of local economic activity. Since Ciccone and Hall (1996) the employment density rather than employment population simpliciter has often been used. The measure of density used is either simply the density $d_{ct} = N_{ct}/S_c$ being the number of workers $N_{ct}$ divided by the time invariant geographical size of the units $S_c$ or the effective density as defined in Equation (6) allowing for economic spillovers between areas. Figure (1) displays how workers in areas with higher density on average have higher wages than workers in areas with lower density. The size of the bubbles represent the employment population sizes. The largest bubble being the municipality of Copenhagen which is also the core of the greatest urban area in Denmark called the Greater Copenhagen area. Regressing the log of mean hourly wages by area on the log of the employment population density results in an elasticity of 0.043 which amounts to a 3% increase in the when doubling the employment density of the area.

The higher average wages in the high density areas could also be related to workforce composition. Figure (2) shows that the share of local workers with a university degree is positively correlated with the density and that the share of high

Figure 1: Wages and employment density

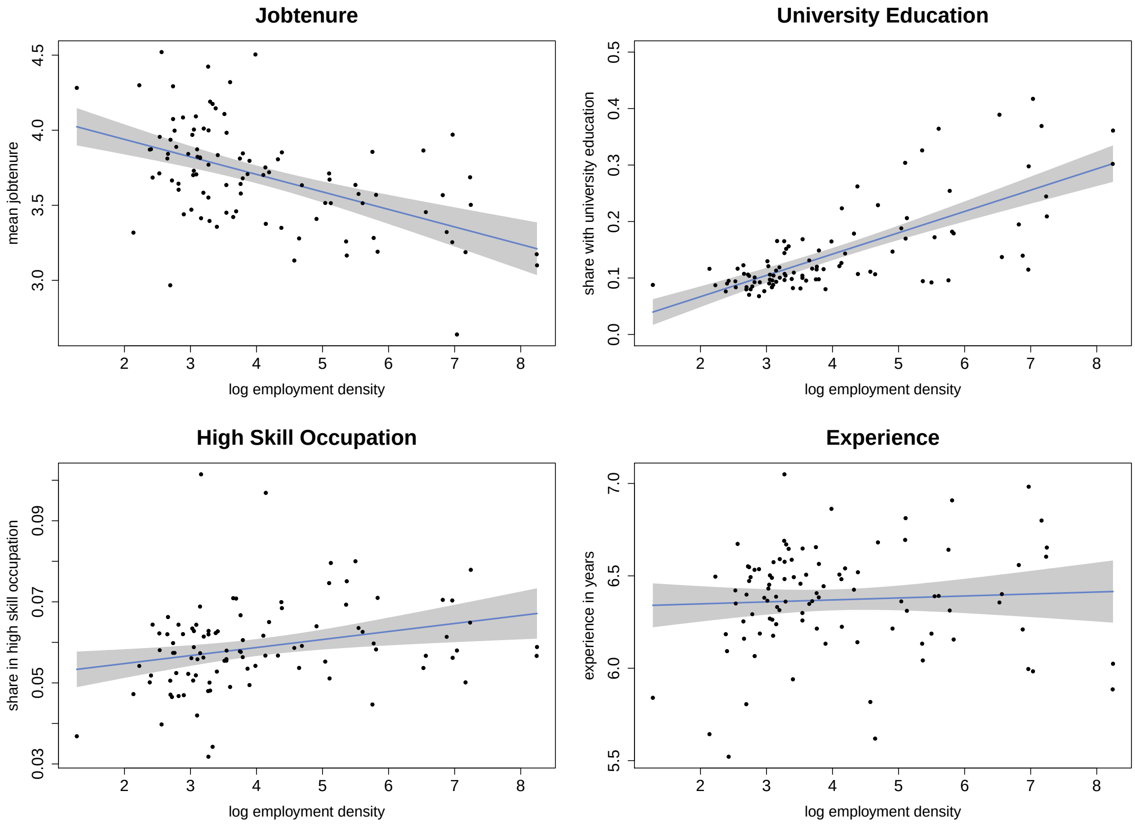



skill occupation is weakly positively correlated with the density. This is consistent with self selection in the sense that high skill workers have stronger preferences for the larger variety of consumer goods offered in big cities. But it is also consistent with e.g. the prescence of high quality educational institutions in big cities.

Jobtenure - the length of time a worker has been with an employer - is negatively correlated with density. This is consistent with workers in high density areas more frequently change employment perhaps due to the more dense labour market. The higher frequency of job changes could thus be an important part of the agglomeration gains by better matching.

Experience - work activity accumulated over time from the year 2003 - has a relatively high variation ranging from 5.5 to 7 at the area level but is uncorrelated with density. Hence to the extent that workers sort according to experience - also closely related to age - this does seem to be systematically correlated with the employment density.

Figure 2: Density and observable skills




# 4 Estimation and identification

The estimation of the main log-wage equation (7) is carried out using two distinct methods. The first method estimates the wage regression

$$\log w_{it} = \mu_i + \mathbf{x}_{it}^\top \beta + z_{ct}^\top \lambda + \eta_{ct} + \phi_k + \delta_t + \epsilon_{it}, \tag{9}$$

using fixed effects for the unobserved factors: indivdual $\mu_i$, sector $\phi_k$ and year $\delta_t$ hence leaving $\eta_{ct} + \epsilon_{it}$ in the error term. The only observed city specific characteristic is $z_{ct}$, which is either simply the observed employment density $d_{ct}$, or else the effective density $z_{ct}(\alpha)$ as given by Equation (6). Following Ciccone and Hall (1996) the density is instrumented using historically lagged population counts from the years 1801 and 1834. The assumption guiding this choice of instrument is that historical population counts are correlated with current employment densities. This is very likely the case because the local housing stock, office buildings and factories last over time creating persistence in local population and economic activity. If the lags are long enough, outcomes are very likely unrelated to the current unobserved local drivers $\eta_{ct}$ of productivity.

The agglomeration elasticity $\lambda$ is identified by the variation in wages for stayers due to changes in density over time. Formally this can be seen by the first difference $\log w_{it'} - \log w_{it} = (z_{c(i,t'),t'} - z_{c(i,t),t})\lambda$ with $c(i,t') = c(i,t)$ ignoring all other factors than density. For the moovers the change in wages is given as $\log w_{it'} - \log w_{it} = (z_{c'(i,t'),t'} - z_{c(i,t),t})\lambda$ with $c'(i,t') \neq c(i,t)$ which both include the change in density over time as well as between the areas $c'(i,t') \neq c(i,t)$.

The second estimation method use the two-step approach following Combes et. al. (2008) where it is assumed that $\sigma_{ct} = z_{ct}^\top \lambda + \eta_{ct} + \delta_t$. Equation (7) can then be rewritten as

$$\log w_{it} = \mu_i + \mathbf{x}_{it}^\top \beta + \sigma_{ct} + \phi_k + \epsilon_{it}, \tag{10}$$

and estimated in the first step with individual fixed effect for $\mu_i$, municipality-time fixed effects for $\sigma_{ct}$ and sector fixed effects for $\phi_k$. In the second step the equation for the municipality-time fixed effects

$$\sigma_{ct} = z_{ct}^\top \lambda + \eta_{ct} + \delta_t, \tag{11}$$

is estimated with the fixed effect estimates $\hat{\sigma}_{ct}$ substituted for $\sigma_{ct}$ and with time fixed effects $\delta_t$. To control for feedback effects the second step uses historical population counts to instrument the density $z_{ct}$ similarly to the first method.

The identification of the municipality-time fixed effects $\sigma_{ct}$ irequires both moovers and stayers. Again the moover experience both a time effects and a location effect when changing location of work $c$ to $c'$ from one period $t$ to another $t'$. The total effect is given as

$$\sigma_{c't'} - \sigma_{ct} = (\sigma_{c't'} - \sigma_{c't}) + (\sigma_{c't} - \sigma_{ct}). \tag{12}$$

Stayers on the other hand identify the time changes $\sigma_{c't'} - \sigma_{c't}$ such that the difference between a stayer and a moover makes it possible to identify effect of a pure location change.

It is clear that the second method unlike the first requires the prescence of both moovers and stayers in order to identify the agglomeration elasticity $\lambda$. Moreover, while the first method assumes that $Cov(\mathbf{x}_{it}, \eta_{ct}) = 0$ because $\eta_{ct}$ is left in the error term of the estimation equation, the second method has no such assumption due to using municipality-time fixed effects. Such an assumption could result in inconsistent estimates based on the first method, if high skilled workers where better capable of responding to increases in the (for the econometrician) unobserved productivity shocks $\eta_{ct}$ (Combes and Gobillon 2015, p. 11). While this added generality favors the second method the empirical literature suggests that the local endownments $\eta_{ct}$ play a weak role (Combes, Duranton, and Gobillon 2008).

For both methods the inclusion of the decay parameter $\alpha$ changes the estimation equation from a standard linear regression model to a non-linear regression. Due to the endogeneity of the effective density $z_{ct}(\alpha)$ instruments are necessary and it is suggested to estimate the models using non-linear 2SLS estimator (Amemiya 1974) defined as the minimizer of the objective function

$$Q_N(\theta) = \left(\sum_{i=1}^{N} Z_i^\top \rho(s_i, \theta)\right)^\top \left(\sum_{i=1}^{N} Z_i^\top Z_i\right)^{-1} \left(\sum_{i=1}^{N} Z_i^\top \rho(s_i, \theta)\right), \tag{13}$$

where $Z_i$ is the matrix of instruments and exogenous variables and $\rho(s_i, \theta)$ is the generalized error function with $s_i = (w_i, z_i, x_i)$. Conditional on $\alpha$ the estimator reduces to the linear 2SLS estimator with fixed effects. I therefore define $\theta = (\theta', \alpha)$ and the function

$$V(\alpha) := \max_{\theta'}\{Q_N(\theta', \alpha)\} = Q_N(\hat{\theta}'(\alpha), \alpha) \tag{14}$$

where $\hat{\theta}'(\alpha)$ is the linear 2SLS estimate of $\hat{\theta}'$ for the given value of $\alpha$. To find the $\alpha$ that minimize $Q_N(\theta)$ the function $Q_N(\hat{\theta}'(\alpha), \alpha)$ can be minimized with respect to $\alpha$ which is computationally more practical because the linear 2SLS estimate $\hat{\theta}'(\alpha)$ is easy to compute with a standard statistical software. Further details related to the estimation procedure are given in Appendix A.

# 5 Results

This section presents the estimation results. In subsection 5.1 the estimates of the one-step estimation method are presented. Subsection 5.2 presents the results of the two-step estimation method. Finally subsection 5.3 presents the estimates for the decay parameter found using the two-step method.

## 5.1 One-step estimation method

This section reports the results of the one-step estimation method as defined by equation (9). First the equation is estimated without individual fixed effects. Then the equation is estimated with individual fixed effects. In both cases the same four specifications are estimated. The specifications differ with respect to whether they include education or not and with respect to whether they use employment density or the effective density. Finally the model is estimated using historical instruments.

Table 1 column (1)-(4) reports the estimates of the four different specifications not including individual fixed effects. The estimates of the coefficient on log density or log effective density is the estimate of the agglomeration elasticity and these estimates are all within the range of 0.018-0.027. This implies that a 1% increase in employment density results in a 0.018%-0.027% increase in wages ignoring spill-over effects. Doubling the area density would lead to a percentage wage increase of $100 \cdot (2^{\hat{\lambda}} - 1)$, which given the estimates correspond to a percentage wage increase in the range of 1.26%-1.84%. While these estimates are consistent with the literature, they are somewhat lower than result presented by elsewhere. Puga and De la Roca (2017) finds an elasticity of 0.0455 for Spain, while Combes et. al. (2010) finds an elasticity of 0.051 for France.

Including education results in a reduction of the estimate of the elasticity of agglomeration. When log density is used the reduction is 16%, as is evident by comparing column (1) with column (3), and when log effective density is used the reduction is 13%, which can be seen by comparing column (2) with column (4). This reduction is consistent with higher educated workers sorting to high density areas as was shown to be the case in Figure 2.

The estimates of the agglomeration elasticity is lower when the effective density is used rather than the density simpliciter. In the model where education is not included the reduction is 24% while it is 22% for the model where education is

Table 1: One-step estimation (no individual fixed effects)

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | log wage | | | |
| | (1) | (2) | (3) | (4) |
| Log density | 0.0268*** | | 0.0227*** | |
| | (0.0001) | | (0.0001) | |
| Log effective density | | 0.0203*** | | 0.0178*** |
| | | (0.0001) | | (0.0001) |
| Experience | 0.0506*** | 0.0501*** | 0.0498*** | 0.0495*** |
| | (0.0002) | (0.0002) | (0.0002) | (0.0002) |
| Experience$^2$ | -0.0020*** | -0.0020*** | -0.0020*** | -0.0020*** |
| | (0.00002) | (0.00002) | (0.00002) | (0.00002) |
| Jobtenure | 0.0110*** | 0.0110*** | 0.0114*** | 0.0114*** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| Jobtenure$^2$ | -0.0008*** | -0.0008*** | -0.0007*** | -0.0007*** |
| | (0.00001) | (0.00001) | (0.00001) | (0.00001) |
| Medium occupational skill | 0.2101*** | 0.2099*** | 0.1577*** | 0.1569*** |
| | (0.0005) | (0.0005) | (0.0005) | (0.0005) |
| High occupational skill | 0.3979*** | 0.3977*** | 0.3303*** | 0.3297*** |
| | (0.0011) | (0.0011) | (0.0011) | (0.0011) |
| Secondary Education | | | 0.0498*** | 0.0506*** |
| | | | (0.0005) | (0.0005) |
| University Education | | | 0.2089*** | 0.2098*** |
| | | | (0.0008) | (0.0008) |
| Male | 0.1345*** | 0.1349*** | 0.1352*** | 0.1354*** |
| | (0.0005) | (0.0005) | (0.0005) | (0.0005) |
| Sector fixed effects | + | + | + | + |
| Year fixed effects | + | + | + | + |
| Observations | 7,251,827 | 7,251,827 | 7,251,827 | 7,251,827 |
| $R^2$ | 0.3357 | 0.3383 | 0.4078 | 0.4108 |
| *Note 1:* | | *p<0.1; **p<0.05; ***p<0.01 | | |

included.

Table 2 reports the results of the estimation including individual fixed effects. The estimates of the elasticity of agglomeration are all within the range of 0.0081-0.0090. Including individual fixed effects leads to a sizeable reduction of more than 50% in the estimate of the elasticity of agglomeration across all specifications. This can be seen by comparing the estimates of the elasticity of agglomeration in Table 2, with the estimates from the analogous model (same column number) in Table 1.

The reduction in the estimate of the elasticity of agglomeration when introducing the individual fixed effects is in line with previous studies. Puga and De La Roca (2017) report a drop of 47% for Spain, while Combes et. al. (2010) report a drop of 35% for France and Mion and Naticchioni (2004) report a drop of 67% for Italy. The drop suggests that sorting of workers on unobserved time constant characteristics to high density areas is important.

Table 2: One-step estimation with individual fixed effects

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | log wage | | | |
| | (1) | (2) | (3) | (4) |
| Log density | 0.0090*** | | 0.0081*** | |
| | (0.0002) | | (0.0002) | |
| Log effective density | | 0.0092*** | | 0.0084*** |
| | | (0.0002) | | (0.0002) |
| Experience | 0.0417*** | 0.0417*** | 0.0415*** | 0.0415*** |
| | (0.0004) | (0.0004) | (0.0004) | (0.0004) |
| Experience$^2$ | -0.0019*** | -0.0019*** | -0.0015*** | -0.0015*** |
| | (0.00002) | (0.00002) | (0.00002) | (0.00002) |
| Jobtenure | 0.0099*** | 0.0099*** | 0.0100*** | 0.0100*** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| Jobtenure$^2$ | -0.0006*** | -0.0006*** | -0.0007*** | -0.0007*** |
| | (0.00001) | (0.00001) | (0.00001) | (0.00001) |
| Medium occupational skill | 0.0262*** | 0.0261*** | 0.0213*** | 0.0213*** |
| | (0.0005) | (0.0005) | (0.0005) | (0.0005) |
| High occupational skill | 0.0676*** | 0.0675*** | 0.0640*** | 0.0639*** |
| | (0.0008) | (0.0008) | (0.0008) | (0.0008) |
| Secondary Education | | | 0.1981*** | 0.1980*** |
| | | | (0.0017) | (0.0017) |
| University Education | | | 0.3384*** | 0.3377*** |
| | | | (0.0023) | (0.0023) |
| Individual fixed effects | + | + | + | + |
| Sector fixed effects | + | + | + | + |
| Year fixed effects | + | + | + | + |
| Observations | 7,251,827 | 7,251,827 | 7,251,827 | 7,251,827 |
| $R^2$ | 0.8597 | 0.8598 | 0.8629 | 0.8630 |
| *Note 1:* | | *p<0.1; **p<0.05; ***p<0.01 | | |

The estimates of the elasticity of agglomeration reduce 10% when education is included but are very robust to the use of density versus effective density. This is consistent with the findings of Briant et. al. (2010) who argue that the econometric specification and choice of spatial units is of secondary importance to the inclusion of individual fixed effects.

Table 3: One-step estimation with individual fixed effects and instruments

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | log wage | | | |
| | (1) | (2) | (3) | (4) |
| Log density | 0.0081*** | | | |
| | (0.0002) | | | |
| Log effective density | | 0.0084*** | | |
| | | (0.0002) | | |
| Log density (IV) | | | 0.0029*** | |
| | | | (0.0003) | |
| Log effective density (IV) | | | | 0.0072*** |
| | | | | (0.0005) |
| Decay | | 6.0740*** | | 5.2301*** |
| | | (0.0083) | | (0.0105) |
| Experience | 0.0415*** | 0.0415*** | 0.0416*** | 0.0415*** |
| | (0.0004) | (0.0004) | (0.0004) | (0.0004) |
| $\text{Experience}^2$ | -0.0015*** | -0.0015*** | -0.0015*** | -0.0015*** |
| | (0.00002) | (0.00002) | (0.00002) | (0.00002) |
| Jobtenure | 0.0100*** | 0.0100*** | 0.0099*** | 0.0100*** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| $\text{Jobtenure}^2$ | -0.0007*** | -0.0007*** | -0.0007*** | -0.0007*** |
| | (0.00001) | (0.00001) | (0.00001) | (0.00001) |
| Medium occupational skill | 0.0213*** | 0.0213*** | 0.0215*** | 0.0213*** |
| | (0.0005) | (0.0005) | (0.0005) | (0.0005) |
| High occupational skill | 0.0640*** | 0.0639*** | 0.0639*** | 0.0639*** |
| | (0.0008) | (0.0008) | (0.0008) | (0.0008) |
| Secondary Education | 0.1981*** | 0.1980*** | 0.1992*** | 0.1984*** |
| | (0.0017) | (0.0017) | (0.0017) | (0.0017) |
| University Education | 0.3384*** | 0.3377*** | 0.3398*** | 0.3383*** |
| | (0.0023) | (0.0023) | (0.0023) | (0.0023) |
| Individual fixed effects | + | + | + | + |
| Sector fixed effects | + | + | + | + |
| Year fixed effects | + | + | + | + |
| Observations | 7,251,827 | 7,251,827 | 7,251,827 | 7,251,827 |
| $R^2$ | 0.8629 | 0.8630 | - | - |
| *Note 1:* | | | | *p<0.1; **p<0.05; ***p<0.01 |

Table 3 column (3)-(4) reports the results of the one step estimation method with individual fixed effects using historical instruments. For comparison column (1) and (2) of the table reproduce the analogous results from Table 2 estimated without instruments. Using instruments results in an elasticity of agglomeration of 0.0029 column (3) for density and 0.0072 for effective density column (4). When density is used the estimate drop from 0.0081 to 0.0029, which is a surprisingly large reduction compared with previous studies and with the reduction that occurs when using effective density. Combes et. al. (2010) report a drop from 0.033 to 0.027 of approximately 17%, which is more in line with the reduction on 9% that I find when using the effective density.

Table 3 also reports two estimates for the decay parameter $\hat{\alpha}$. Using instruments leads to a drop in the estimate from 6.1 to 5.2, which is equivalent to an increase in the half-time from 6.8 to 8.3 minutes[1]. Assuming an average speed of 60km/hour this is equivalent to 6.8-8.3 km. of distance. This finding is in line with the literature, usually finding that agglomeration effects tend to reduce significantly between 5 and 10 km from the origin (Graham and Gibbons 2019, p. 9).

## 5.2 Two-step estimation method

This section reports the result of the two-step estimation procedure. This method is defined by equations (10) and (11) and procedes by first estimating $\hat{\sigma}_{ct}$ year-area fixed effects and then regressing these on either density or effective density.

The results of the first step of the two-step estimation method using time-area is given in Table 4. The estimates of the coefficients of observed worker characteristics are numerically very similar to the results obtained when using the one-step method. The coefficient on the experience is reduced from approximately 0.5 to 0.4 when introducing individual fixed effects and remains robust to the introduction of education. The same is the case for the one-step estimation method as see Table 1 and 2. The estimate for the coeffient on jobtenure is reduced from 0.011 to 0.01 when introducing individual fixed effects and are likewise robust to the introduction of education. This behavior of the estimates is again similar to

[1]The halftime is the time difference $\Delta T := T_2 - T_1$ such that $0.5 = \exp(-\hat{\alpha}T_2)/\exp(-\hat{\alpha}T_1)$ this implies the well known rule that $\Delta T = \ln(2)/\hat{\alpha}$. Since I measure time in hours $\Delta T$ is multiplied with 60 to get the results stated in the text.

results from the one-step estimation method (Table 1 and 2).

Experience and jobtenure are as expected both concave with a positive coefficient on the linear terms and a negative coefficient on the squared terms. Considering the model estimated with education and individual fixed effects given in Table 4 column (4) the first year of experience increases wage with on average 4.1% while the 5th year on average increases wage 3.1%. For jobtenure the first year increases wages with 1.0% while the 3rd year of jobtenure increases wage with 0.7%.

Table 4: Two step estimation results

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | log wage | | | |
| | (1) | (2) | (3) | (4) |
| Experience | 0.0502*** | 0.0490 *** | 0.0411 *** | 0.0409 *** |
| | (0.0002) | (0.0002) | (0.0004) | (0.0004) |
| Experience$^2$ | -0.0020 *** | -0.0019 *** | -0.0019 *** | -0.0016 *** |
| | (0.00002) | (0.00002) | (0.00002) | (0.00002) |
| Jobtenure | 0.0110 *** | 0.0113 *** | 0.0098 *** | 0.0099 *** |
| | (0.0001) | (0.0001) | (0.0001) | (0.0001) |
| Jobtenure$^2$ | -0.0007 *** | -0.0007 *** | -0.0006 *** | -0.0006 *** |
| | (0.00001) | (0.00001) | (0.00001) | (0.00001) |
| Medium occupational skill | 0.2058 *** | 0.1510 *** | 0.0257 *** | 0.0212 *** |
| | (0.0005) | (0.0006) | (0.0005) | (0.0005) |
| High occupational skill | 0.3967 *** | 0.3474 *** | 0.0670 *** | 0.0637 *** |
| | (0.0011) | (0.0011) | (0.0008) | (0.0008) |
| Secondary education | | 0.0479 *** | | 0.1995 *** |
| | | (0.0006) | | (0.0017) |
| University education | | 0.2079 *** | | 0.3296 *** |
| | | (0.0008) | | (0.0023) |
| Individual fixed effects | - | - | + | + |
| Sector effects | + | + | + | + |
| Time-area | + | + | + | + |
| Observations | 7,251,827 | 7,251,827 | 7,251,827 | 7,251,827 |
| $R^2$ | 0.3453 | 0.3791 | 0.8610 | 0.8641 |

*Note:* $^{*}$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01
standard errors clustered at individual level.

Without controlling for education and individual fixed effects occupational skill increases wages with 20% for medium skilled and 40% for high skilled. Controlling for education decreases the effect of occupational skill down to respectively 15%

and 35%. When controlling for both education and individual fixed effects this drops down to merely 2% and 6% identified by the approximately 4-5% of the labour force that change occupational skill in each year.

Secondary education and university education increase wages by respectively 5% and 20%. These estimates increase when controlling for individual fixed effects to 20% and 32%, respectively. The group of the labour force identifying these coefficients are workers changing educational level and are therefore individuals who are first active on the labour market but then choose to enter education and reappear in the panel or hold a job while under education. This group of individual workers is as small as 1-2% of the labour force in any year.

### 5.2.1 Elasticity of agglomeration

The estimates of the elasticity of agglomeration using the two-step method are in the range of 0.86% to 4.05% as given in Table 5. Controlling for education reduces the estimates with approximately 10% going from column (1) to (2) or from column (3) to (4). Even with the individual fixed effects, controlling for education reduces the estimate of the elasticity of agglomeration. Including individual fixed effects in the first stage decreases the estimates of the elasticity of agglomeration with approximately 50-60% going from column (1) to (3) or from column (2) to (4). A reduction of this size is in line with the literature (see Melo, Graham, and Noland 2009) and is similar to the reduction seen when using the one-step method.

Using the effective density rather than density simpliciter seems to increase the estimates 25-40% when instruments are not applied. This is apparent when going from row (1) to row (2) in Table 5. However this tendency decrease when instruments are used in which case the increase is approximately 5-15%.

This confirms that for the two-step method the inclusion of education and use of density rather than effective density is of secondary importance compared to the inclusion of fixed effects. This is the same result as obtained for the one-step estimation method and documented in Briant et. al. (2010).

For the preferable model using instruments and controlling for both education as well as individual fixed effects, the one-step and two-step estimates are in the range of 0.072 to 0.0107 depending on whether the effective density or the density simpliciter is used as measure for the economic activity. This implies that doubling

Table 5: Two-step estimation of agglomeration elasticity

| | | *Fixed effects included in first stage estimation* | | | |
|---|---|---|---|---|---|
| Individual fixed effects: | | No | No | Yes | Yes |
| Sector fixed effects: | | Yes | Yes | Yes | Yes |
| Time-area fixed effects: | | Yes | Yes | Yes | Yes |
| Education: | | No | Yes | No | Yes |
| | | *Dependent variable: Year-time fixed effects* | | | |
| Log density | (1) | 0.0304*** | 0.0270*** | 0.0140*** | 0.0125*** |
| | | (0.0004) | (0.0007) | (0.0004) | (0.0007) |
| Log eff-density | (2) | 0.0405*** | 0.0350*** | 0.0225*** | 0.0206*** |
| | | (0.0003) | (0.0028) | (0.0019) | (0.0022) |
| Log density (IV) | (3) | 0.0195*** | 0.0176*** | 0.0097*** | 0.0086*** |
| | | (0.0004) | (0.0004) | (0.0006) | (0.0009) |
| Log eff-density (IV) | (4) | 0.0202*** | 0.0182*** | 0.0116*** | 0.0107*** |
| | | (0.0018) | (0.0011) | (0.0036) | (0.0031) |
| *Note:* | | *p<0.1; **p<0.05; ***p<0.01 | | | |

the employment density increases wage within the range of 0.50 to 0.74%.

## 5.3 The decay parameter

Table 6 shows the estimates $\hat{\alpha}$ for the decay parameter based on the two-step estimation method. All the estimates fall within the range 6.9-8.0, which is equivalent to a half time of 5.2-6.0 minutes. Using again the assumption of an average speed of 60km/hour, this is equivalent to 5.2-6.0 km. of distance, which is in line with the literature and similar to the estimates found using the one-step method.

Columns (1) and (2) in Table 6 reports the estimates of the decay parameter based on first stage estimations not including individual fixed effects. Comparing these estimates to those given in Table 6 column (3) and (4), where individual fixed effects are included in the first stage estimation, it is apparent that the estimate of the decay parameter is reduced when individual fixed effects are included. Using instruments and controlling for education the reduction the estimates for

Table 6: Two-step estimation of deacy parameter

| | | *Fixed effects included in first stage estimation* | | | |
|---|---|---|---|---|---|
| Individual fixed effects: | | No | No | Yes | Yes |
| Sector fixed effects: | | Yes | Yes | Yes | Yes |
| Time-area fixed effects: | | Yes | Yes | Yes | Yes |
| Education: | | No | Yes | No | Yes |
| | | *Dependent variable: Year-time fixed effects* | | | |
| Decay | (1) | 8.00*** | 7.69*** | 7.08*** | 7.38*** |
| | | (0.013) | (0.031) | (0.0023) | (0.0027) |
| Deacy - IV | (2) | 7.84*** | 7.54*** | 6.92*** | 7.08*** |
| | | (0.031) | (0.0028) | (0.049) | (0.035) |
| *Note:* | | *p<0.1; **p<0.05; ***p<0.01 | | | |

the decay paramater changes from 7.54 to 7.08 approximately a reduction of 8%. The estimates of the decay parameter is thus less sensitive to the inclusion of the individual fixed effects than the etsimate of the elasticity of agglomeration.

The effect of controlling for education is less systematic. When individual fixed effects are not included in the first stage estimation the estimate of the decay parameter is reduced when edcution is included (see column (1) and (2) in Table 6). However, controlling for education when fixed effects are included in the first stage estimation, leads oppositely to an increase in the estimate of the decay parameter (see column (3) and (4) in Table 6). In either case the sensitivity of the estimate of the decay parameter is lower than that of the elasticity of agglomeration.

Because the two-step estimation method does not assume that $Cov(\mathbf{x}_{it}, \eta_{ct}) = 0$ I prefer this method to the one-step method. In general, the both estimation methods indicate the importance of controlling for sorting of workers both on observed characteristics such as education and unobserved characteristic in the form of individual fixed effects. My favorite estimate for the elasticity of agglomeration allowing for spill-over effects is 0.0107 as given in Table (5) column (4) row (4). The associated estimate of the decay parameter is $\hat{\alpha} = 7.07$ given in Table 6 column (4) row (2).

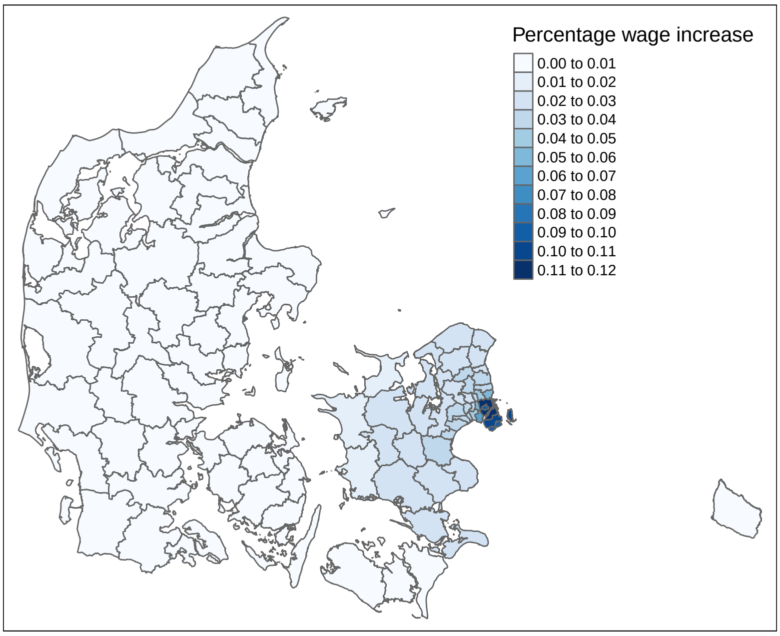



Figure 3: Percentage wage increases from Copenhagen-Frederiksberg population growth (2030)

To illustrate the spill-over effect I consider a 17% population increase in Frederiksberg and Copenhagen as projected for the year 2030. Then applying the method suggested by Graham et. al. (Graham and Gibbons 2019) for quantifying the wider economic benefits I calculate the wage increases using the formula

$$\Delta \log w_c = \hat{\lambda} \cdot \Delta z_c(\hat{\alpha}) = \hat{\lambda} \log \sum_{a \in \mathcal{C}} (\Delta d_a) \cdot \exp(-\hat{\alpha} \cdot T(c,a)), \tag{15}$$

with $\hat{\lambda} = 0.0102$ and $\hat{\alpha} = 7.07$. The resulting wage increases are of a relative low order 0-0.12 percent and the area affacted by the spill-over effect is similarly relatively small as illustrated in Figure 3. While the municipalities surrounding Frederiksberg and Copenhagen are realtively small and therefore in close proximity, the travel times between these municipalities are still relative high due to congestion.

# 6 Conclusion

Thise paper confirms that controlling for worker-specific unobservable heterogeneity to reduce bias in the estimates of the elasticity of agglomeration is of primary importance. Controlling for worker-specific unobservable heterogeneity leads to a in with the reduction in estimates larger than 50% across econometric specifications. Correcting for reverse causality between earnings and agglomeration using instruments are also important to reduce the bias and leads to a reduction of 12% for the elasticity of agglomeration. Controlling for education is also found to be important, reducing estimates for the elasticity of agglomeration by approximately 8%. This is consistent with workers sorting both on observed and unobserved characteristics. I also conclude that for the estimation of the decay parameter, the inclusion of individual fixed effects as well as eduction is less important for the estimation of the decay parameter.

This paper estimates the elasticity of agglomeration to be 0.0107, such that a log point increase in the effective density leads to 0.0107 log point increase in wages. Increasing the effective density of an area by 100%, therefore leads an 0.74% increase in wages. For the same specification the decay parameter is found to be 7.08 which is equivalent to a half-time of $(\log(2)/7.08) \cdot 60 = 5.87$ minutes of transportation time.

The estimate of the elasticity of agglomeration and the decay parameter are important for transport and urban planner. Cost Benefit Analysis is widely used to evaluate the change in net "welfare" arising from transport improvements. Conventional CBA captures only part of the benefits resulting from investments in transportation. The Wider Economic Impacts (WEI) seeks further to incorporate impacts arising from externalities and from forms of imperfect competition. Procedures suggested for the calculation WEI requires the use of the estimate of the elasticity of agglomeration and the decay parameter.

## References

Ahlfeldt, Gabriel M. and Elisabetta Pietrostefani (2019). "The Economic Effects of Density: A Synthesis". In: *Journal of Urban Economics* 111, pp. 93–107.

Ahlfeldt, Gabriel M., Stephen J. Redding, Daniel M. Sturm, and Nikolaus Wolf (2015). "The Economics of Density: Evidence From the Berlin Wall". In: *Econometrica* 83.6, pp. 2127–2189.

Amemiya, Takeshi (1974). "The Nonlinear Two-Stage Least-Squares Estimator". In: *Journal of Econometrics* 2.2, pp. 105–110.

— (1977). "The Maximum Likelihood and the Nonlinear Three-Stage Least Squares Estimator in the General Nonlinear Simultaneous Equation Model". In: *Econometrica* 45.4, p. 955.

— (1983). "Chapter 6 Non-Linear Regression Models". In: *Handbook of Econometrics*. Vol. 1. Elsevier, pp. 333–389.

— (1985). *Advanced Econometrics*. Cambridge, Mass: Harvard University Press.

Arellano, Manuel (2003). *Panel Data Econometrics*. Oxford, UK: Oxford University Press.

Briant, A., P.-P. Combes, and M. Lafourcade (2010). "Dots to Boxes: Do the Size and Shape of Spatial Units Jeopardize Economic Geography Estimations?" In: *Journal of Urban Economics* 67.3, pp. 287–302.

Brueckner, Jan K and Yves Zenou (1999). "Why Is Central Paris Rich and Downtown Detroit Poor? An Amenity-Based Theory". en. In: *European Economic Review*, p. 17.

Ciccone, Antonio (2002). "Agglomeration Effects in Europe". In: *European Economic Review*, pp. 213–227.

Ciccone, Antonio and Robert Hall (1996). "Productivity and the Density of Economic Activity". In: *The American Economic Review* 86.1, pp. 54–70.

Combes, P.-P., G. Duranton, and L. Gobillon (2011). "The Identification of Agglomeration Economies". In: *Journal of Economic Geography* Vol. 11.No. 2, pp. 253–266.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2008). "Spatial Wage Disparities: Sorting Matters!" In: *Journal of Urban Economics* 63.2, pp. 723–742.

Combes, Pierre-Philippe and Laurent Gobillon (2015). "The Empirics of Agglomeration Economies". In: *Handbook of Regional and Urban Economics*. Vol. 5. Elsevier, pp. 247–348.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux (2010). "Estimating Agglomeration Economies with History, Geology, and Worker Effects". In: *Agglomeration Economics*. Ed. by Edward L. Glaeser. University of Chicago Press.

— (2012a). "Sorting and Local Wage and Skill Distributions in France". In: *Regional Science and Urban Economics* 42.6, pp. 913–930.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux (2012b). "The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection". In: *Econometrica* 80.6, pp. 2543–2594.

D'Costa, Sabine and Henry G. Overman (2014). "The Urban Wage Growth Premium: Sorting or Learning?" In: *Regional Science and Urban Economics* 48, pp. 168–179.

De Borger, Bruno, Ismir Mulalic, and Jan Rouwendal (2019). *Productivity Effects of an Exogenous Improvement in Transport Infrastructure: Accessibility and the Great Belt Bridge*. Tech. rep. 065/VIII.

Duranton, Gilles and Diego Puga (2004). "Chapter 48 Micro-Foundations of Urban Agglomeration Economies". In: *Handbook of Regional and Urban Economics*. Vol. 4. Elsevier, pp. 2063–2117.

Enrico, Moretti (2011). "Local Labor Markets". In: *Handbook of Labor Economics*. Vol. 4. Elsevier, pp. 1237–1313.

Fujita, Masahisa and Hideaki Ogawa (1982). "Multiple Equilibria and Structural Transition of Non-Monocentric Urban Configurations." In: *Regional Science and Urban Economics* 12, pp. 161–196.

Gallant, A.Ronald (1977). "Three-Stage Least-Squares Estimation for a System of Simultaneous, Nonlinear, Implicit Equations". In: *Journal of Econometrics* 5.1, pp. 71–88.

Glaeser, Edward L. and David C. Maré (2001). "Cities and Skills". In: *Journal of Labor Economics* 19.2, pp. 316–342.

Graham, Daniel J and Stephen Gibbons (2018). *Quantifying Wider Economic Impacts of Agglomeration for Transport Appraisal: Existing Evidence and Future*

*Directions*. Tech. rep. No 1561. The London School of Economics and Political Science, pp. 1–52.

Graham, Daniel J. and Stephen Gibbons (2019). "Quantifying Wider Economic Impacts of Agglomeration for Transport Appraisal: Existing Evidence and Future Directions". In: *Economics of Transportation* 19, p. 100121.

Graham, Daniel J, Stephen Gibbons, and Ralf Martin (2010). *The Spatial Decay of Agglomeration Economies: Estimates for Use in Transport Appraisal*. Tech. rep. Imperial College London, p. 40.

Graham, Daniel J., Patricia S. Melo, Piyapong Jiwattanakulpaisarn, and Robert B. Noland (2010). "Testing for Causality Between Productivity and Agglomeration Economies". In: *Journal of Regional Science* 50.5, pp. 935–951.

H. Hanson, Gordon (2005). "Market Potential, Increasing Returns and Geographic Concentration". In: *Journal of International Economics* 67.1, pp. 1–24.

Harris, C.D. (1954). "The Market as a Factor in the Localization of Industry in the United States." In: *Annals of the Association of American Geographers* 44, pp. 315–348.

Holman, Nancy, Alan Mace, Antoine Paccoud, and Jayaraj Sundaresan (2015). "Coordinating Density; Working through Conviction, Suspicion and Pragmatism". In: *Progress in Planning* 101, pp. 1–38.

Jr., Robert E. Lucas and Esteban Rossi-Hansberg (2002). "On the Internal Structure of Cities". en. In: *Econometrica* 70.4, pp. 1445–1476.

Kuminoff, Nicolai V., V. Kerry Smith, and Christopher Timmins (2013). "The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets". In: *Journal of Economic Literature* 51.4, pp. 1007–1062.

Melo, Patricia C., Daniel J. Graham, and Robert B. Noland (2009). "A Meta-Analysis of Estimates of Urban Agglomeration Economies". In: *Regional Science and Urban Economics* 39.3, pp. 332–342.

Mion, Giordano (2004). "Spatial Externalities and Empirical Analysis: The Case of Italy". In: *Journal of Urban Economics* 56.1, pp. 97–118.

Newey, Whitney K. (1990). "Efficient Instrumental Variables Estimation of Nonlinear Models". In: *Econometrica* 58.4, p. 809.

— (1993). "16 Efficient Estimation of Models with Conditional Moment Restrictions". en. In: *Handbook of Statistics*. Vol. 11. Elsevier, pp. 419–454.

Newey, Whitney K and Daniel McFadden (1994). “Large Sample Estimation and Hypothesis Testing”. In: *Handbook of Econometrics*. Ed. by Daniel McFadden and R. Engle. Vol. 4. Elsevier, p. 135.

Ottaviano, Gianmarco and Jacques-François Thisse (2004). “Chapter 58 Agglomeration and Economic Geography”. In: *Handbook of Regional and Urban Economics*. Vol. 4. Elsevier, pp. 2563–2608.

Puga, Diego (2010). “The Magnitude and Causes of Agglomeration Economies”. In: *Journal of Regional Science* 50.1, pp. 203–219.

Roback, Jennifer (1982). “Wages, Rents, and the Quality of Life”. In: *Journal of Political Economy* 90.6, pp. 1257–1278.

Roca, Jorge De La and Diego Puga (2017). “Learning by Working in Big Cities”. In: *The Review of Economic Studies* 84.1, pp. 106–142.

Rosenthal, Stuart S. and William C. Strange (2004). “Chapter 49 Evidence on the Nature and Sources of Agglomeration Economies”. In: *Handbook of Regional and Urban Economics*. Vol. 4. Elsevier, pp. 2119–2171.

Scotchmer, Suzanne (2002). “Chapter 29 Local Public Goods and Clubs”. In: *Handbook of Public Economics*. Vol. 4. Elsevier, pp. 1997–2042.

Starrett, David (1978). “Market Allocations of Location Choice in a Model with Free Mobility”. In: *Journal of Economic Theory* 17.1, pp. 21–37.

Storper, M. and A. J. Scott (2008). “Rethinking Human Capital, Creativity and Urban Growth”. In: *Journal of Economic Geography* 9.2, pp. 147–167.

Venables, Anthony J. (2007). “Evaluating Urban Transport Improvements: Cost-Benefit Analysis in the Presence of Agglomeration and Income Taxation”. In: *Journal of Transport Economics and Policy* 41.2, pp. 173–188.

Wooldridge, Jeffrey M. (1996). “Estimating Systems of Equations with Different Instruments for Different Equations”. In: *Journal of Econometrics* 74.2, pp. 387–405.

Wooldridge, Jeffrey M (2010). *Econometric Analysis of Cross Section and Panel Data*. Second. London: The MIT Press.

# A Appendix

The estimation of the decay parameter is performed using a version of the non-linear 2SLS estimator. The estimator is defined as the solution to a quadratic minimization problem loosely stated as

$$\min_{\theta} \rho(\theta)^{\top}\Xi \ \rho(\theta), \tag{16}$$

where $\Xi = Z(Z^{\top}Z)^{-1}Z^{\top}$ with $Z$ being the matrix of instruments and $\rho(\theta)$ being the error function. This estimator was initially proven to be consistent by Amemiya (1974) and is also treated in Amemiya (1983) as a simultaneous equations model. In Amemiya (1977) it is expanded to the 3SLS version which is also covered by Gallant (1977) and Wooldridge (1996). The estimator could either be considered as an M-estimator (extremum estimator) or as a GMM estimator the asymptiotic properties of which are dealt with in Amemiya (1985) as well as Newey and McFadden (1994). I follow Wooldridge (1996) in treating the instruments as stochastic and assume that the $T \times 1$ vector $\rho(\mathbf{y}, \mathbf{W}, \theta)$ has the conditional mean property

$$\mathbb{E}[\rho(\mathbf{y}_i, \mathbf{W}_i, \theta_0)|\mathbf{Z}_i] = \mathbf{0}, \tag{17}$$

where $\theta_0$ with subscript is the true value of the parameter vector and $(\mathbf{y}_i, \mathbf{W}_i, \mathbf{Z}_i)$ are observed data for the $i$'th worker. The distinction between $\mathbf{Z}_i$ and $\mathbf{W}_i$ is there to allow for some variables in $\mathbf{W}_i$ to be endogenous whereas the instruments $\mathbf{Z}_i$ are exogenous. This assumption is stronger than what is necessary in the case of linear instrumental regression, with the conditional zero moment implying that the error is uncorrelated with any function of the instruments.

By the iterated law of expectation it follows that

$$\mathbb{E}[\mathbf{Z}_i\mathbf{P}_i\rho(\mathbf{y}_i, \mathbf{W}_i, \theta_0)] = \mathbf{0}, \tag{18}$$

where $\mathbf{P}_i$ is some deterministic projection matrix which appears in order to project out the fixed effects. This is conceptually very similar to demeaning the data in the case where the model only includes individual fixed effects. In this case where the model includes both individual, sector and time fixed effects the routine is slightly more complicated.

To define the relevant error function I consider the model

$$\log w_{it} = \mu_i + \mathbf{x}_{it}^\top \beta + z_{ct}^\top \lambda + \eta_{ct} + \phi_k + \delta_t + \epsilon_{it}, \tag{19}$$

as defined in equation (9) and stack the $T$ observations for individual $i$ to get the matrix form

$$\mathbf{y}_i = \mathbf{X}_i \beta + \mathbf{z}_i(\alpha)\lambda + D_i^\mu \mu + D_i^\phi \phi + D_i^\delta \delta + \epsilon_i \tag{20}$$

where $\mathbf{y}_i := \log \mathbf{w}_i$. From here the fixed effects are collected in a single matrix $D_i = [D_i^\mu \; D_i^\phi \; D_i^\delta]$ and the vector $\psi := (\mu, \phi^\top, \delta^\top)^\top$ such that

$$\mathbf{y}_i = \mathbf{X}_i \beta + \mathbf{z}_i(\alpha)\lambda + D_i \psi + \epsilon_i. \tag{21}$$

The next step is to project out out the fixed effects using $\mathbf{P}_i$ to get the error function evaluated at $\theta_0$

$$\mathbf{P}_i \rho_i(\mathbf{y}_i, \mathbf{W}_i, \theta_0) = \mathbf{P}_i(\mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{z}_i(\alpha)\lambda) = \mathbf{P}_i \epsilon_i, \tag{22}$$

using the definition $\rho_i(\mathbf{y}_i, \mathbf{W}_i, \theta_0) = \mathbf{y}_i - \mathbf{X}_i \beta - \mathbf{z}_i(\alpha)\lambda$ with $\mathbf{W}_i = (\mathbf{X}_i, \mathbf{z}_i(\alpha))$ from which it follows that $\mathbb{E}[\rho(\mathbf{y}_i, \mathbf{W}_i, \theta_0)|\mathbf{Z}_i] = \mathbf{0}$ if $\mathbb{E}[\epsilon_i|\mathbf{Z}_i] = \mathbf{0}$. Given these moment condition it is fairly easy to arrive at the non-linear 2SLS estimator in a general method of moments framework. In the next section this framework is introduced and the references to some results on asymptotic behavior are provided along with some results by Newey 1990 and 1993 on optimal instruments.

## A.1 GMM estimation with optimal instruments

The data used for estimation is assumed to be a random sample size $N$ of iid observations of the vector $w_i$. Specifically the data is assumed to be a panel data and for notational convience assumed to be balanced with $T$ time periods observed for each $i$.

The model under consideration is summarized by the $T \times 1$ generalized residuals vector function $\rho(w_i, \theta)$ where $\theta$ is the parameters of the model. For the true value of the paramters $\theta_0$ the variables $x_i$ are assumed to be exogenous as given in

Assumption (1).

**Assumption 1 (Exogeneity)** *The first assumption is that the* $L \times 1$ *vector* $\rho(w_i, \theta)$ *of generalized errors satisfies the zero conditonal moment condition*

$$\mathbb{E}[\rho(w_i, \theta_0)|x_i] = \mathbf{0} \tag{23}$$

For each observation the exogenous variables are used to create a $L \times T$ matrix $Z(x_i)$ of instruments being functions of the exogenous variables.

$$g(w_i, \theta) = Z(x_i)^\top \rho(w_i, \theta) = \sum_{t=1}^{T} z(x_i)^\top \rho_t(w_i, \theta), \tag{24}$$

where $\rho_t(w_i, \theta)$ is $t$'th component of $\rho(w_i, \theta)$.

**Example 1 (Multivariate non-linear regression)** *Consider non-linear multivariate regression where* $y_i = F(x_i, \theta_0) + \epsilon$ *where* $y_i = (y_{i1}, ..., y_{iT})^\top$, $y_i = (\epsilon_{i1}, ..., \epsilon_{iT})^\top$ *and* $F(x_i, \theta_0) = (f(x_{it}, \theta_0), ..., f(x_{it}, \theta_0))^\top$ *are all* $T \times 1$ *vectors. Then*

$$\rho(w_i, \theta) = \rho(y_i, x_i, \theta) = y_i - F(x_i, \theta) \tag{25}$$

*for the true value* $\theta_0$ *it follows that* $\rho(w_i, \theta_0) = \epsilon$ *such that the assumption* $\mathbb{E}[\rho(w_i, \theta_0)|x_i] = \mathbf{0}$ *is equivalent to assuming* $\mathbb{E}[y_i|x_i] = F(x_i, \theta_0)$*.*

Using Assumption (1) it follows by law of iterated expectation that

$$\mathbb{E}[g(w_i, \theta_0)] = \mathbb{E}[Z(x_i)^\top \rho(w_i, \theta_0)] = \mathbf{0}, \tag{26}$$

in analogue to which the sample moments

$$g_N(\theta) = \frac{1}{N} \sum_{i=1}^{N} Z(x_i)^\top \rho(w_i, \theta), \tag{27}$$

are defined and the GMM estimate $\hat{\theta}$ is found as the minimizer of the quadratic form

$$Q_N(\theta) = g_N(\theta)^\top W_N g_N(\theta), \tag{28}$$

where $W_N \xrightarrow{p} W_0$ positive semidefinite. Intuitively the function $Q_N(\hat{\theta})$ goes toward some function $Q_0(\theta)$ as the sample grows and since $\theta_0$ is the minimizer of the $Q_0(\theta)$ the GMM estimator $\hat{\theta}$ goes towards $\theta_0$ in probability as the sample grows. Theorems establishing consistency are given various places in the econometrics literature see for example Amemiya 1985, p.107, Arellano 2003, p. 184 or Theorem 14.1 in Wooldrige 2010. The same sources also establish the asymptotic normality of the GMM estimator and that for the case of iid observations the asymptotic variance is given as

$$V = (D_0^\top W_0 D_0)^{-1}(D_0^\top W_0 V_0 W_0 D_0)(D_0^\top W_0 D_0)^{-1}, \tag{29}$$

where

$$V_0 := \mathbb{E}[g(w_i, \theta_0) g(w_i, \theta_0)^\top] \tag{30}$$

$$D_0 := \mathbb{E}\left[\frac{\partial g(w_i, \theta_0)}{\partial \theta}\right]. \tag{31}$$

The above framework can generally be used to set up an objective function to define a consistent and asymptotically normally distributed estimator given some choice of a weight matrix $W_N$. The estimate of the covariance matrix are found simply by using $W_N$ and by using the analogous sample moments for $V_0$ and $D_0$ evaluated at $\hat{\theta}$.

**Example 2 (Multivariate non-linear regression continued)** *Consider again the case of non-linear multivariate regression where $y_i = F(x_i, \theta_0) + \epsilon$. The objective function using the $L \times L$ matrix*

$$W_N = \left(\frac{1}{N} Z(x_i)^\top Z(x_i)\right)^{-1}$$

*as weight matrix is then given as*

$$Q_N(\theta) = \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^\top (y_i - F(x_i, \theta))\right)^\top \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^\top Z_i\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N} Z_i^\top (y_i - F(x_i, \theta))\right)$$

*where $Z_i = Z(x_i)$. This is the non-linear 2SLS estimator proven consistent by*

*Amemiya 1974.*

## A.2 Optimal weight matrix

To get an efficient estimator it is necessary to choose the proper weight matrix $W_N$ given the instruments and to select the proper instuments. The optimal weight matrix is given as

$$W_0^* = V_0^{-1}, \tag{32}$$

from which it follows that the variance of the GMM estimator reduces to

$$V = (D_0^\top V_0^{-1} D_0)^{-1}. \tag{33}$$

**Example 3 (Multivariate non-linear regression)** *Consider again the case of non-linear multivariate regression where $y_i = F(x_i, \theta_0) + \epsilon$ and where $\rho(y_i, x_i, \theta) = y_i - F(x_i, \theta)$. It then follows that*

$$V_0 := \mathbb{E}[g(w_i, \theta_0) g(w_i, \theta_0)^\top] = \mathbb{E}[Z_i^\top \epsilon_i \epsilon_i^\top Z_i] = \mathbb{E}[Z_i^\top \Omega Z_i],$$

*where the final identity use an assumption of system homoscedasticity $\Omega := Var(\epsilon_i) = \mathbb{E}[\epsilon_i \epsilon_i^\top]$ which is standard in panel data application because it allows for errors $\epsilon_{it}$ and $\epsilon_{i't'}$ for the same observational unit $i = i'$ to be correlated but not errors across observations $i \neq i'$. The assumption of constancy is convenient because the obvious alternative is to assume that $\Omega(x_i)$ varies with $x_i$ but then one has to estimate a covariance matrix function which is standardly more challenging than estimating the covariance matrix under the assumption of system homoscedasticity. Given an estimate $\hat{\Omega}$ the objective function can be set up as*

$$Q_N(\theta) = \left( \frac{1}{N} \sum_{i=1}^N Z_i^\top (y_i - F(x_i, \theta)) \right)^\top \left( \frac{1}{N} \sum_{i=1}^N Z_i^\top \hat{\Omega} Z_i \right)^{-1} \left( \frac{1}{N} \sum_{i=1}^N Z_i^\top (y_i - F(x_i, \theta)) \right),$$

*which is the non-linear 3SLS estimator as defined in Wooldridge (1996).*

*If one is willing to assume homoscedasticity simpliciter $\Omega = \sigma^2 I_T$ it follows* *that*

$$\mathbb{E}[Z_i^\top \Omega Z_i] = \sigma^2 \mathbb{E}[Z_i^\top Z_i], \tag{34}$$

*in which case the optimal weight matrix is simply*

$$W_N = \left(\frac{1}{N} Z_i^\top Z_i\right)^{-1}$$

*as used in the previous example.*

The previous examples basically show how under the classical assumption of homoscedasticity the instrumental non-linear 2SLS estimator is an efficient GMM estimator. With panel data the assumption of homoscedasticity is however unrealistic hence the assumption of system homoscedasticity and the justification of the non-linear 3SLS estimator. However to achieve the best GMM estimator it is not enough to choose the optimal weighting matrix given the instruments, one must also choose the optimal instruments.

## A.3 Optimal instruments

The optimal instruments are given as

$$Z^*(x_i) := \Omega(x_i)^{-1} B(x_i), \tag{35}$$

where the $T \times p$ matrix $B(x_i)$ and the $T \times T$ matrix $\Omega(x_i)$ are given as

$$B(x_i) := \mathbb{E}\left[\frac{\partial \rho(w_i, \theta_0)}{\partial \theta^\top} | x_i\right] \tag{36}$$

$$\Omega(x_i) := \mathbb{E}[\partial \rho(w_i, \theta_0) \partial \rho(w_i, \theta_0)^\top | x_i], \tag{37}$$

as explained in Newey 1990 and 1993. Consider the general expression of the covariance matrix for the GMM estimator using the optimal weight matrix given as $V = (D_0^\top V_0^{-1} D_0)^{-1}$. The matrix $D_0$ is given as

$$D_0 := \mathbb{E}\left[\frac{\partial g(w_i, \theta_0)}{\partial \theta}\right] = \mathbb{E}\left[Z(x_i)^\top \frac{\partial \rho(w_i, \theta_0)}{\partial \theta^\top}\right] = \mathbb{E}\left[B(x_i)^\top \Omega(x_i)^{-1} \frac{\partial \rho(w_i, \theta_0)}{\partial \theta^\top}\right],$$

using the optimal instruments. Using iterated expectation it follows that

$$D_0 = \mathbb{E}\left[\mathbb{E}\left[B(x_i)^\top \Omega(x_i)^{-1} \frac{\partial \rho(w_i, \theta_0)}{\partial \theta^\top} | x_i\right]\right] = \mathbb{E}[B(x_i)^\top \Omega(x_i)^{-1} B(x_i)]. \qquad (38)$$

For the matrix $V_0$ it follows that

$$V_0 = \mathbb{E}[g(w_i, \theta_0) g(w_i, \theta_0)^\top] = \mathbb{E}[Z(x_i)^\top \rho(w_i, \theta_0) \rho(w_i, \theta_0)^\top Z(x_i)] \qquad (39)$$

which under the assumtion of optimal instruments by use of iterated expectation become

$$V_0 = E[B(x_i)^\top \Omega(x_i)^{-1} B(x_i)] = D_0, \qquad (40)$$

implying that the GMM estimators covariance can be further reduced to

$$V = V_0 = E[B(x_i)^\top \Omega(x_i)^{-1} B(x_i)] \qquad (41)$$

**Example 4 (Multivariate non-linear regression continued)** *Consider again the case of non-linear multivariate regression where $y_i = F(x_i, \theta_0) + \epsilon$. Stack the data vector to express the objective function in matrix notation*

$$Q_N(\theta) = \frac{1}{N} (\mathbf{y} - \mathbf{F}(X, \theta))^\top Z (Z^\top Z)^{-1} Z^\top (\mathbf{y} - \mathbf{F}(X, \theta)), \qquad (42)$$

*and take the first order derivative to get*

$$\frac{\partial Q_N(\theta)}{\partial \theta} = \frac{1}{N} \frac{\partial \mathbf{F}(X, \theta)}{\partial \theta}^\top Z (Z^\top Z)^{-1} Z^\top (\mathbf{y} - \mathbf{F}(X, \theta)), \qquad (43)$$

*it is then apparent that since the function $\partial \mathbf{F}(X, \theta)^\top / \partial \theta$ only depend on exogenoys variables it qualifies as being an instrument since the instruments were defined as any functions of exogenous variables. Using this it follows that I can set $Z = \partial \mathbf{F}(X, \theta) / \partial \theta^\top$ such that the first order conditions reduces to*

$$\frac{\partial Q_N(\theta)}{\partial \theta} = \frac{1}{N} \frac{\partial \mathbf{F}(X, \theta)}{\partial \theta}^\top (\mathbf{y} - \mathbf{F}(X, \theta)). \qquad (44)$$

*Two observations are worth making in this case. First of all $\frac{\partial \mathbf{F}(X, \theta)}{\partial \theta^\top}$ is simply the* *stacking of* $\frac{\partial F(x_i,\theta)}{\partial\theta^\top}$ *and this derivative is the derivative of the generalized error* $\frac{\partial\rho(y_i,x_i,\theta)}{\partial\theta^\top} = \frac{\partial F(x_i,\theta)}{\partial\theta^\top}$ *and further because it only depends on* $x_i$ *the exogenous variables and not* $y_i$ *the endogenous variable it follows that* $\mathbb{E}[\frac{\partial\rho(y_i,x_i,\theta)}{\partial\theta^\top}|x_i] = \frac{\partial F(x_i,\theta)}{\partial\theta^\top}$ *hence the derivative is the optimal instrument up to a constant of proportionality* $\sigma^2$ *under the assumption of homoscedasticity* $\Omega(x_i) = \sigma^2 I_T$. *Secondly it is observed that the first order condition* $\frac{\partial Q_N(\theta)}{\partial\theta} = 0$ *is also first order condition to* $\sum_i (y_i - F(x_i,\theta))^\top (y_i - F(x_i,\theta))$ *the solution to which is the pooled non-linear least squares estimator. Hence pooled non-linear least squares estimator is efficient under homoscedasticity in the sense that it is equivalent to the GMM estimator with optimal instruments and weight matrix.*

*If on the other hand system homoscedaticity is assumed* $\Omega(x_i) = \Omega$ *the optimal instriments are given as*

$$Z^*(x_i) = -\Omega^{-1}\frac{\partial F(x_i,\theta)}{\partial\theta^\top}, \tag{45}$$

*and in this case the objective function becomes*

$$Q_N(\theta) = \frac{1}{N}\left(\sum_{i=1}^N G_i^\top\Omega^{-1}(y_i - F(x_i,\theta))\right)^\top\left(\sum_{i=1}^N G_i^\top\hat{\Omega}^{-1}G_i\right)^{-1}\left(\sum_{i=1}^N G_i^\top\Omega^{-1}(y_i - F(x_i,\theta))\right),$$

*where* $G_i = \frac{\partial F(x_i,\theta)}{\partial\theta^\top}$. *This estimator is the estimator treated in Newey (1990) applied to the problem of non-linear multivariate regression.*

## A.4 Non-linear regression with factors and an endogenous variable

Consider now the non-linear regression problem

$$y_{1i} = F(y_{i2}, x_i, \theta_0) + D_i\alpha_0 + \epsilon_i \tag{46}$$

where $D_i$ is a factor matrix $D_i = [D_{i1}, ..., D_{iK}]$ and $\alpha_0$ is factor parameter. In the simples case $K = 1$ and $D_i$ is a matrix with one column of ones and $N-1$ columns of 0's such that $D_i\alpha_0 = \iota_T\alpha_{0i}$ simply picking out the individual specific fixed effect as in a fixed effects model. It is assumed that

$$\mathbb{E}[\epsilon_i|x_i] = \mathbf{0}, \tag{47}$$

implying that

$$\mathbb{E}[M_{D_i}(y_{1i} - F(y_{i2}, x_i, \theta_0))|x_i] = \mathbf{0} \tag{48}$$

where $M_{D_i}$ is the annihilator matrix for $D_i$ defined as $I - D_i(D_i^\top D_i)D_i^\top$. It follows that it is possible to define

$$g(w_i, \theta) = Z(x_i)^\top M_{D_i}(y_{1i} - F(y_{i2}, x_i, \theta)), \tag{49}$$

with $w_i = (y_i, x_i)$, $y_i = (y_{i1}, y_{i2})$ and $\rho(y_i, x_i, \theta) = M_{D_i}(y_{1i} - F(y_{i2}, x_i, \theta))$ such that the moment condition

$$\mathbb{E}[g(w_i, \theta_0)] = \mathbb{E}[Z(x_i)^\top \rho(y_i, x_i, \theta_0)] = \mathbb{E}[Z(x_i)^\top M_{D_i}(y_{1i} - F(y_{i2}, x_i, \theta_0))] = \mathbf{0}, \tag{50}$$

is satisfied. Under the assumption of system homoscedastiscity

$$\mathbb{E}\left[\rho(y_i, x_i, \theta_0)\rho(y_i, x_i, \theta_0)^\top\right] = \Omega, \tag{51}$$

the optimal instruments are given as

$$Z^*(x_i) = -\Omega^{-1}\mathbb{E}\left[M_{D_i}\frac{\partial F(y_{i2}, x_i, \theta)}{\partial \theta^\top}|x_i\right] \tag{52}$$

# Do female workers benefit less from agglomeration?

Jesper Hybel[*] Ismir Mulalic[†] Malte Borghorst[‡]

**Abstract**

This paper explores the urban wage premium, focusing on gender differences in the benefits from agglomeration. Women still experience greater domestic burden, leading to locally restricted because of and consequently higher commuting costs and therefore may benefit less from agglomeration. Using a panel of the full working population in Denmark for the years 2008-2016, we first demonstrate the existence of an urban wage premium not only for the wage level but also for wage growth. We then show that the portable part of the value of experience is lower for female workers than for male workers. Specifically, we show that the lower return of experience for female workers is rooted in the lower additional gains from experience accumulated in high-density areas and the larger additional benefits from using experience in cities.

**Keywords:** agglomeration economies, learning, wage premium, gender.
**JEL codes:** J16, J31, R23.

[*]Technical University of Denmark, DTU Management, Bygningstorvet, Building 116, DK-2800 Kgs. Lyngby, Denmark, email: jeshyb@dtu.dk. Jesper Hybel is supported by the Innovation Fund Denmark via the Urban Project.

[†]Copenhagen Business School, Department of Economics, Porcelænshaven 16A, DK-2000 Frederiskberg, Denmark, email: imu.eco@cbs.dk. Ismir Mulalic is a Visiting Professor at VU Amsterdam.

[‡]BUILD Department, Aalborg University, A. C. Meyers Vænge 15, 2450 Copenhagen, email: mjb@sbi.aau.dk.

# 1 Introduction

Female workers earn significantly less than male workers. The economic literature offers numerous explanations for the gender wage gap such as differences in education, labor market participation and work experience, and more recently norms and psychological attributes, see e.g. Blau and Kahn 2017. Furthermore, Kleven, Landais, and Søgaard 2018 argue that the gender wage gap is a child care penalty and Borghorst, Mulalic, and Ommeren 2020 find that female workers commute shorter distances and choose jobs farther from agglomerations because they have greater domestic burdens and consequently higher commuting costs. This paper investigates the link between benefits from agglomeration and the gender wage gap, more specifically the role of dynamic learning effects from agglomeration.

Combes et al. 2012 show that firms located in cities enjoy a productivity advantage from density, i.e. agglomeration.[1] Consequently, average wages are higher in areas of dense economic activity, viz. cities (Henderson 2003). The agglomeration economies are usually seen as emerging from improved sharing, matching, or learning in dense labor markets (Duranton and Puga 2004). In cities, these advantages raise productivity and result in an urban wage premium (Puga 2010; Enrico 2011). Besides high-quality worker-firm matches and more productive firms in cities, workers also accumulate human capital faster, i.e. gain more valuable work experience, compared to workers in thinner labor markets (Roca and Puga 2017). This paper seeks to identify and examine the potential differences in the gender-specific human capital accumulation and the urban wage premium.

Following the seminal contribution by Glaeser and Maré 2001, a large body of empirical literature has identified a significant urban wage premium (see e.g., Rosenthal and Strange 2004, Puga 2010), Melo, Graham, and Noland 2009 and Combes and Gobillon 2015). The three main explanations for this urban wage premium offered in the economic literature are static advantages, sorting of workers based on heterogeneous initial ability, and dynamic advantages. The static advantages associated with dense urban areas refer to the benefits only enjoyed while working in cities (Rosenthal and Strange 2004; Puga 2010 and Holmes 2010).

[1]The benefits related to proximity to other economic agents are generally referred to as agglomeration effects and are the subject of an extensive literature in spatial economics (Duranton and Puga 2004, Rosenthal and Strange 2004, Puga 2010, Gaubert 2018).

Sorting is related to the possibility that more productive workers may choose to locate in cities. In the Danish context of this paper, higher educated and wealthier workers are more likely to work in bigger cities and closer to agglomerations (Hybel and Mulalic 2020, Mulalic and Rouwendal 2020, Gutiérrez-i-Puigarnau, Mulalic, and van Ommeren 2016). Combes, Duranton, and Gobillon 2008 suggest that the impact of sorting on the urban wage premium is similar in magnitude as the static advantages. Finally, cities often facilitate dynamic advantages through learning and superior interactions with other economic agents. Workers in cities thus accumulate more high-quality experience. Roca and Puga 2017 argue convincingly that this accumulated human capital remains beneficial even when a worker relocates.

The agglomeration benefits might be different for male and female workers. Hirsch, König, and Möller 2013 find that the gender wage gap in Germany varies between rural and urban areas. Phimister 2005 estimate the static wage premium for the UK and find that the heterogeneous cognitive and social skills across genders is related to the gender-specific urban wage premium. Rosenthal and Strange 2012 find those female businesses profit less from network effects and therefore less from agglomeration due to higher household burden and subsequently higher commuting costs. This paper also aims to identify the nature and causes of the gender-specific learning effect. Denmark is an interesting case because of its equality-friendly labor market. Women's participation rate in Denmark is high and it already in 2007 exceeds 70 %. Female workers constitute about 50 % of the entire workforce.

Using a rich administrative data set for Denmark that follows workers over a decade and across municipalities, we estimate the gender-specific returns of experience acquired and used at different locations. We first construct worker and time-specific measurements of working experience collected at different job places and urban areas. Then we use Mincerian wage regressions to learn about place-specific wages corrected for observed worker characteristics and to identify the initial unobserved worker ability as reflected in the worker (individual) fixed effects.

Finally, utilizing the panel structure of the available data, we explore if this estimated value of knowledge accumulated in cities is different for male and female workers and if it persists after relocating. We find that the portable part of the value of experience for female workers is lower than for males. Responsible are the lower additional gains from experience accumulated in cities, as well as the larger additional benefits from using experience in the top density areas relative to males.

The remainder of the paper is organized as follows. In section 2, we describe the data, provide descriptive statistics, and presents several empirical observations that suggest the gender-specific differences in the urban (growth) wage premium. Section 3 describes and discusses the empirical model and the estimation strategy. Section 4 presents the empirical results and section 5 discusses the gender-specific urban wage premium. Section 6 concludes.

# 2 Data

The data used in the empirical analysis are derived from annual register data from Statistics Denmark for the years 2008–2016. For each year, we have information on the full population of workers including the workers' workplace location at the municipality level, worker hourly wages, worker experience, job tenure and a range of explanatory variables such as age, gender, and education.[2] Experience for each worker is computed as the cumulative sum of the worker's work activity starting from the year 2003. The work activity is measured as the number of days worked during the year in primary employment. We also observe job tenure – the length of time a worker has been with an employer – for all workers.

## 2.1 Selection of sample and descriptive statistics

We focus on a sample of employed workers aged between 17 and 65. We exclude observations for workers who work in the public sector (health, education, and administration) or mining and agriculture. We also exclude immigrants. For these workers, we do not observe education and do not have sufficient information to correctly calculate the work experience. We are then left with 7,246,703 observations (1,155,612 workers); this contains 704,008 male workers (4,441,873 observations) and 451,604 female workers (2,636,868 observations). Table 1 shows the descriptive statistics. The descriptive statistics of the female workers and the male workers

[2] To protect the identity of the companies for which data exist and to provide sufficient confidentiality protection, Statistics Denmark does not provide the exact workplace addresses for companies, but it does provide the municipality code for each establishment.

are almost identical, except for hourly wage which is about 15 % higher for males, see also Figure 9 in the Appendix. Mean hourly wages are 231 and 200 DKK for the male workers and the female workers, respectively.[3] Male workers are slightly older, have slightly longer job-tenure, and have more often full-time jobs. Female workers obtain more frequently tertiary education, while male workers hold more often leading positions.

Table 1: Descriptive statistics

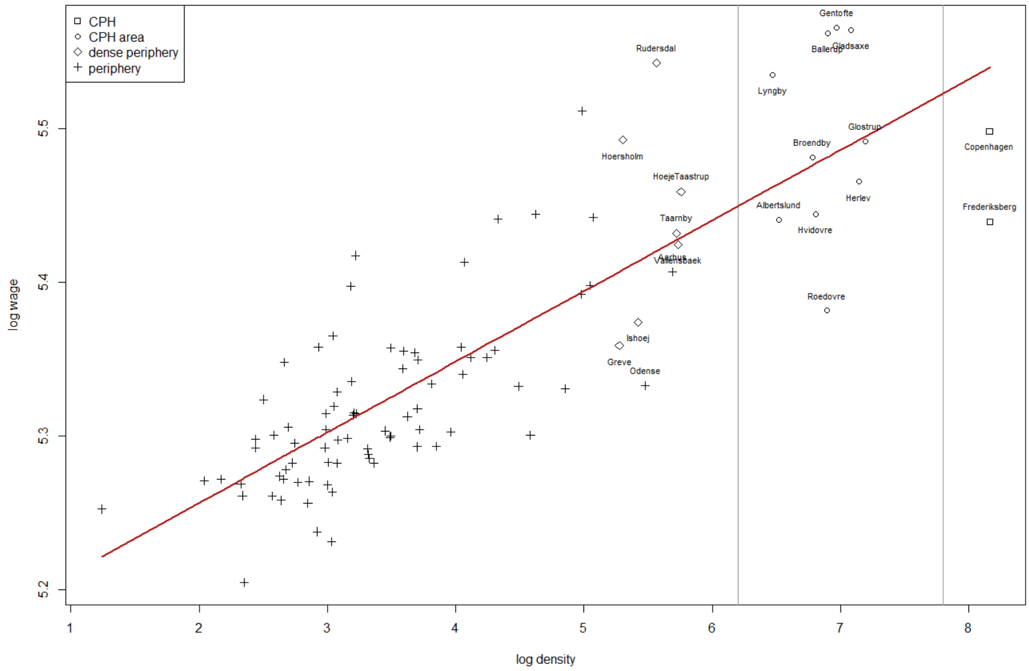
| | Male workers | | Female workers | |
|---|---|---|---|---|
| | mean | std.dev. | mean | std.dev. |
| Hourly wage (DKK) | 230.59 | 79.34 | 200.38 | 67.60 |
| Age | 43.23 | 10.99 | 42.41 | 10.80 |
| Activity (p.a., share) | 0.91 | 0.21 | 0.91 | 0.21 |
| Experience (based on activity) | 10.48 | 3.32 | 10.35 | 3.37 |
| Job-tenure (year) | 4.97 | 4.52 | 4.88 | 4.48 |
| Full-time (share) | 0.92 | 0.28 | 0.82 | 0.39 |
| Education (share) | | | | |
| Primary | 0.24 | | 0.20 | |
| Secondary | 0.52 | | 0.46 | |
| Tertiary | 0.23 | | 0.33 | |
| Occupational skill (share) | | | | |
| Basic skill | 0.65 | | 0.62 | |
| High skill | 0.29 | | 0.36 | |
| Leading position | 0.06 | | 0.02 | |
| Number of workers | 704,008 | | 451,604 | |
| Number of observations | 4,441,873 | | 2,636,868 | |

We group the 98 municipalities covering Denmark into four categories: i) Copenhagen (highest density, the municipalities Copenhagen and Frederiksberg), ii) Copenhagen area (high density around Copenhagen in the metropolitan area), iii) dense periphery (including among other municipalities Aarhus, the second-largest city in Denmark), and iv) periphery (smaller towns and rural areas). It

[3] 1 DKK $\approx$ 0.13 EUR.

appears that there is a significant positive correlation between log hourly wages and log job density, see Figure 1. Moreover, the figure also shows the wage increase for the considered four density groups.

Figure 1: Wage against density anno 2016




The mean worker experience is about 10 years for male workers and female workers. The distributions of experience are also remarkably similar for both genders, see Figure 2. Table 2 suggest however that the female workers use the accumulated experience, conditional on the area where the experience has been accrued, more intensively in high-density areas, compared to the male workers. For example, of the all collected experience in the highest density area, the female workers use 77% of the experience in the same area while male workers use 75%. The same is true for all the other considered areas, i.e. the female workers use the accumulated experience more intensively in high-density areas compared to the male workers. We also find that the share of the female workers increases with the job density, see Figure 3.

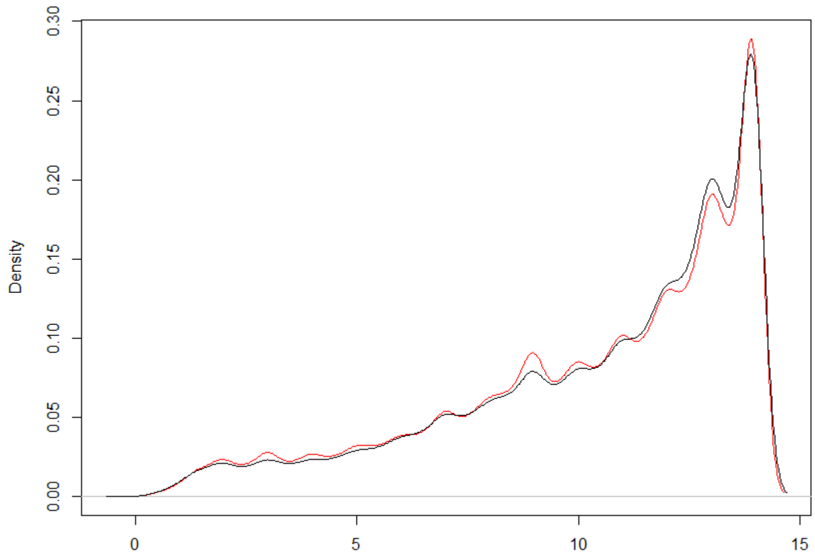
As we have seen, several systematic patterns emerge between the male workers

Table 2: Areas of the used experience in shares, by gender

| Origin | CPH | CPH area | Dense periph. | Periphery | Sum |
|---|---|---|---|---|---|
| Used in | | | Male workers | | |
| CPH | 0.75 | 0.11 | 0.04 | 0.09 | 1.00 |
| CPH area | 0.12 | 0.71 | 0.05 | 0.12 | 1.00 |
| Dense periph. | 0.11 | 0.13 | 0.63 | 0.13 | 1.00 |
| Periphery | 0.01 | 0.02 | 0.01 | 0.96 | 1.00 |
| | | | Female workers | | |
| CPH | 0.77 | 0.10 | 0.04 | 0.08 | 1.00 |
| CPH area | 0.14 | 0.71 | 0.05 | 0.10 | 1.00 |
| Dense periph. | 0.15 | 0.12 | 0.62 | 0.11 | 1.00 |
| periphery | 0.02 | 0.02 | 0.01 | 0.96 | 1.00 |

Notes: CPH (highest density) includes municipalities Copenhagen and Frederiksberg, CPH area (second highest density) includes the rest of the Copenhagen metropolitan area, dense periphery (third highest density) includes among other municipalities Aarhus, the second largest city in Denmark, and periphery (lowest density) includes smaller towns and rural areas.

Figure 2: Distributions of experience for male workers (red) and female workers (black)




Notes: This figure depicts the Gaussian kernel density distribution of the accumulated experience in the year 2016

and the female workers across the density areas in our sample. We summarize them in three facts about the gender specific urban wage premium:

i wages and the share of female workers increase with job density;

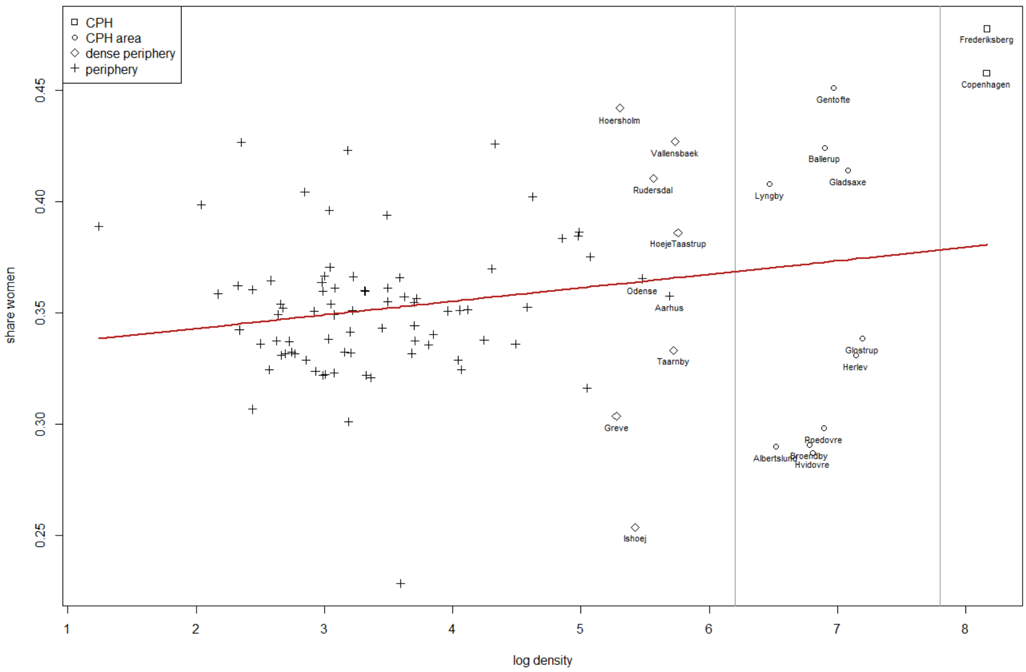
ii distributions of the experience are similar for male and female workers, and

iii female workers use the accumulated experience more intensively in high density areas compared to male workers.

## 3 The econometric model

In this section, we introduce a reduced form wage model that includes the dynamic effects of experience. We introduce the stylized model in subsection 3.1 and describe how the dynamic effects of experience include a portable benefit, that we refer to as a *city premium* and a non-portable benefit, that we refer to as a *city use premium*. We then in subsection 3.2 explain how we specify the learning effects.

Figure 3: Share of the female workers against density



In subsection 3.3 we study the simplified model using linear wage paths. Finally in subsection 3.4 we derive the bias in the estimates of the area fixed effects when the dynamic effects of experience or unobserved worker heterogeneity are ignored.

## 3.1 The wage equation

We use $w_{ait}$ to denote the log wage of worker $i$ in time period $t$ employed in area $a$ and assume that the log wage is given by the equation

$$w_{ait} = \sigma_a + \mu_i + l_{it} + \mathbf{x}_{it}^\top \beta + \epsilon_{it}, \tag{1}$$

where $\sigma_a$ is the unobserved area $a$ fixed effect, $\mu_i$ is the unobserved individual fixed effect, $l_{it}$ are the learning effects to be specified later, $\mathbf{x}_{it}$ is the vector of observable worker characteristics, $\beta$ is a vector of parameters and $\epsilon_{it}$ is unobservable error term.

We use the function $a(i,t)$ to specify the area $a$ in which worker $i$ at time $t$ is employed and let $1[a(i,t) = j]$ be the indicator function that the area of employment is area $j$. The indicators for the areas $j = 1, ..., J$ are collected in the $J \times 1$ vector $\iota_{it} := (1[a(i,t) = 1], ..., 1[a(i,t) = J])^\top$ just as the area fixed effects $\{\sigma_j\}_{j=1}^J$ are collected in the vector $\sigma := (\sigma_1, ..., \sigma_J)^\top$. We assume that workers choice of area of employment $1[a(i,t) = j]$ is uncorrelated with the individual and time specific error terms $\epsilon_{it}$.[4]

The *static* advantages of working in high-density areas are the advantages gained while working there but lost immediately upon being employed elsewhere. A worker changing area of employment from $a = a(i,t)$ to $a' = a(i,t')$ will immediately experience a change in wage due to the difference in the area fixed effects $\sigma_{a'} - \sigma_a$. This change is immediately lost again, should the worker change her area of employment back to the area $a$. The wage equation (1) therefore allows for a static earnings premium of being employed in a high-density area if area fixed effects $\{\sigma_j\}_{j=1}^J$ are positively correlated with the employment density.

We refer to workers with an above-average value of $\mu_i$ as initial high wage earners. The inclusion of the unobserved individual fixed effect $\mu_i$ allows also for sorting, where initial high wage earners are predominantly employed in areas of

[4]See Appendix B in Combes et al. (2008) for a detailed discussion of this assumption in a dynamic framework.

high density. When this is the case, the covariance $Cov(1[a(i,t) = a], \mu_i)$ will be positive for all high density areas $a$. For areas with some employment this is equivalent to $\mathbb{E}[\mu_i | a(i,t) = a] > \mathbb{E}[\mu_i]$, such that the workers of the area $a$ have a higher expected value of $\mu_i$ than the expected value $\mathbb{E}[\mu_i]$ for the population in general. Such sorting effects imply that high-density areas offer certain amenities favored by the high wage earners. The initial high wage earners are therefore willing to pay higher housing prices of high-density areas as suggested by Glaeser and Maré (2001).

Finally the model allows for learning effects $l_{it}$ which capture part of the value of the workers experience distinguished by where the experience is accumulated and where it is used. The value of worker $i$'s experience at time $t$ is given by

$$V(\{e_{ait}\}) = \sum_{a=1}^{J} \phi_{a'a} 1[a(i,t) = a'] e_{ait}, \tag{2}$$

where $e_{ait}$ is the years of experience accumulated in area $a$ at time $t$ by worker $i$. The coefficient $\phi_{aa'}$ measures the value of a year of experience accumulated in area $a$ when used in area $a'$.

To estimate a model that allows for these wage effects of experience we arrange the areas according to their level of employment density into groups $g = \mathcal{G}(a(i,t))$, with a set $g_0$ consisting of areas with low employment density serving as reference group. We then specify the learning effects as

$$l_{it} = \sum_{g \neq g_0} \lambda_g e_{git} + \sum_{g} \delta_g \tilde{e}_{git}, \tag{3}$$

where $\lambda_g e_{git}$ is the value of experience $e_{git}$ accumulated in any area belonging to the group $g$ and $\delta_g \tilde{e}_{git}$ the value of experience accumulated in group $g$ when not used in group $g_0$, achieved by defining $\tilde{e}_{git} := 1[a(i,t) \notin g_0] e_{git}$.[5]

This allows for a *city premium* where experience accumulated in the high-density areas are worth more used anywhere in a case of $\lambda_g > 0$. It also allows for a *city use premium* where experience of different origin $g$ is rewarded higher when used in high density areas $a(i,t) \notin g_0$ in which case $\delta_g > 0$. Importantly the *city premium* $\lambda_g$ is the portable part of experience in comparison to the *city*

[5] The estimated specification also allows for non-linear effects, but we ignore these for now for ease of the presentation.

*use premium* $\delta_g$ which is not portable, hence lost when the worker is employed in areas belonging to the low-density group of areas $g_0$.

## 3.2 The specification of learning effects

In this section, we derive a specification for the learning effects of the form given in equation (3) when there are only two areas: a highly urbanized area $c$ (city) and a less urbanized area $r$ (rural area).

The experience accumulated by worker $i$ at time $t$ by working in the city is denoted by $e_{cit}$ and the experience accumulated by working in the rural area is denoted by $e_{rit}$. The total experience accumulated is simply denoted by $e_{it}$ and equals the sum of the accumulated experiences in the city and in the rural area, i.e. $e_{it} = e_{cit} + e_{rit}$. The wage benefit of experience is distinguished by the area where the experience is used. We use $\phi_{hj}$ as the benefit for an extra year of experience used in area $h$ accumulated in area $j$. The log-wage $w_{rit}$ for a worker currently employed in the rural area is now given by

$$w_{rit} = \sigma_r + \phi_{rr} e_{rit} + \phi_{rc} e_{cit} + u_{it}, \tag{4}$$

where $\sigma_r$ is the static wage effect of the rural area and $u_{it}$ is the unobserved error term. We define $\lambda_{rc} := (\phi_{rc} - \phi_{rr})$ as the measure of the wage premium of the city experience relative to rural experience, when used in the rural area. Adding and subtracting $\phi_{rr} e_{cit}$ to the wage equation for the worker employed in the rural area we get

$$w_{rit} = \sigma_r + \phi_{rr} e_{it} + \lambda_{rc} e_{cit} + u_{it}, \tag{5}$$

that includes only the total experience $e_{it}$ of the worker and the city experience.

We then consider the wage of a worker employed in the city, which is given by

$$w_{cit} = \sigma_c + \phi_{cr} e_{rit} + \phi_{cc} e_{cit} + u_{it}, \tag{6}$$

consisting of a static city specific wage effect $\sigma_c$ and the wage benefits from experience accumulated in the rural area and the city. Adding and subtract the wage benefit $\phi_{rr} e_{rit} + \phi_{rc} e_{cit}$ we get

$$w_{rit} = \sigma_c + \phi_{rr} e_{rit} + \phi_{rc} e_{cit} + (\phi_{cr} - \phi_{rr}) e_{rit} + (\phi_{cc} - \phi_{cr}) e_{cit} + u_{it}, \tag{7}$$

where $\delta_{cr} := (\phi_{cr} - \phi_{rr})$ measures the extra benefit a worker recieves from experience accumulated in the rural area when this experience is used in the city and similarly $\delta_{cc} := (\phi_{cc} - \phi_{cr})$ measures the extra benefit a worker receives from experience accumulated in the city area when used in the city. Importantly a worker changing employment from the rural area to the city, having the stock of expereince $(e_{rit}, e_{cit})$, would immediately receive a wage change $\sum_{a\in\{r,c\}} \delta_{ca} e_{ait} = \delta_{cr} e_{rit} + \delta_{cc} e_{cit}$ in addition to the change $\sigma_c - \sigma_r$ from the static area effects.

Finally we rewrite $\phi_{rr} e_{rit} + \phi_{rc} e_{cit}$ for the worker employed in the city to $\phi_{rr} e_{it} + \lambda_{rc} e_{cit}$ using the same approach as for the worker employed in the rural area and combine equation (5) and (7) to get

$$w_{a(i,t),it} = \iota_{it}^{\top} \sigma + \phi_{rr} e_{it} + \lambda_{rc} e_{cit} + \delta_{cr} \tilde{e}_{rit} + \delta_{cc} \tilde{e}_{cit} + u_{it}, \tag{8}$$

where $\iota_{it} = (1[a(i,t) = r], 1[a(i,t) = c])^{\top}$, $\tilde{e}_{rit} := 1[a(i,t) = c] e_{rit}$, $\tilde{e}_{cit} := 1[a(i,t) = c] e_{cit}$. In this specification the learning effect

$$l_{it} = \lambda_{rc} e_{cit} + \delta_{cr} \tilde{e}_{rit} + \delta_{cc} \tilde{e}_{cit} \tag{9}$$

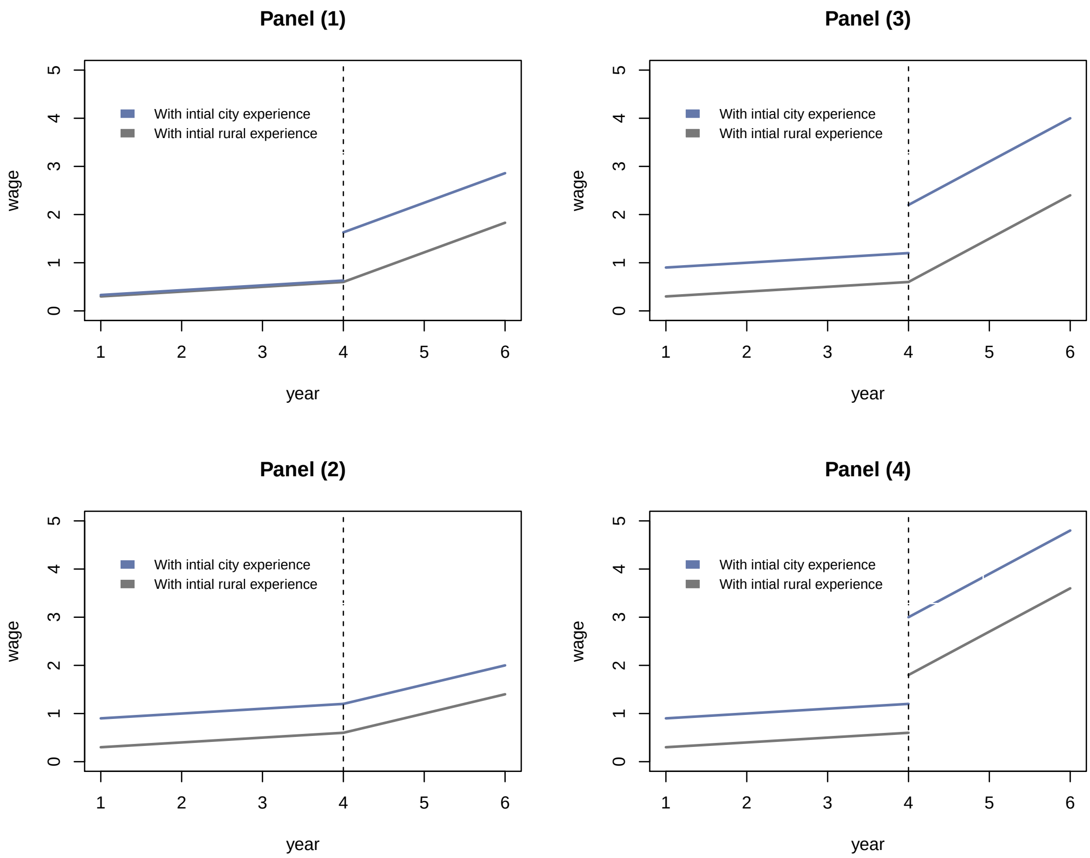
consists of the *city premium* $\lambda_{rc}$ – the wage benefit of the city experience gained irrespective of where it is used – in addition to the *city use premiums* for the rural experience used in the city $\delta_{rc}$ and the city experience used in the city $\delta_{cc}$. In the next section we illustrate the learning effects of this stylized model using linear wage paths.

## 3.3 Wage paths in the stylized model

Figure (4) illustrates the wage paths of four different scenarios using the stylized model given in equation (8). The scenarios differ only in terms of the payoff of experience, see Table 3. In all the scenarios we consider two workers for $T = 6$ periods. Both workers are employed in rural areas for the first 4 years and then change the area of employment to the city for the last two years. We assume that workers accumulate one year of experience with the passing of every period. We also assume that the workers differ in their initial experience. The first worker is assumed to have 2 years of initial rural experience, while the second worker is assumed to have 2 years of initial city experience.

Figure 4 shows that in three scenarios - panel (1), (3), and (4) - there is a jump in at least one of the worker's wage path. At the time of change in job location

Figure 4: Wage curve scenarios




the worker with the initial city experience has also accumulated rural experience and the worker, therefore, has both types of experience. For this worker, the jump is therefore present if either the city use premium of city experience $\delta_{cc}$ or the city use premium of rural experience $\delta_{cr}$ are positive. In comparison, the worker with the initial rural experience has only accumulated rural experience at the time of change in job location and therefore there is only a jump in the wage path of this worker in panel (4) where $\delta_{cr} > 0$.

We also explore the wage growth rate which is determined by the value of experience at the current area of employment. This is trivial because the current area of employment determines what type of experience the worker accumulates. This implies that the wage growth is determined solely by $\phi_{rr} = 0.1$ while the

Table 3: Scenarios

| | Exp. $\phi_{rr}$ | City exp. $\lambda_{cr}$ | Rural exp. used in city $\delta_{cr}$ | City exp. used in city $\delta_{cc}$ | Static effect $\sigma_r = \sigma_c$ |
|---|---|---|---|---|---|
| Panel (1) | 0.1 | 0 | 0 | 0.5 | 0 |
| Panel (2) | 0.1 | 0.3 | 0 | 0 | 0 |
| Panel (3) | 0.1 | 0.3 | 0 | 0.5 | 0 |
| Panel(4) | 0.1 | 0.3 | 0.2 | 0.5 | 0 |

workers are employed in the rural area. When employed in the city the growth rate increase to $\lambda_{cr} + \delta_{cc} + \phi_{rr}$ adding the sum of the city premium and the city use premium of city experience. The higher growth rate of wage during the city employment, due to $\lambda_{cr} > 0$ and $\delta_{cc} > 0$, implies that the value of experience for both workers is above the average while they are employed in the city. The city use premium of rural experience $\delta_{cr}$ cannot affect any wage growth rate because the rural experience is never being accumulated when the worker is employed in the city.

## 3.4 The bias of static wage effects

In this section we expand the simple wage model of equation (8) to include an individual fixed effect $\mu_i$, observed worker characteristics $\mathbf{x}_{it}$ and the learning effects $l_{it}$. The model is therefore equivalent to our full model given in (1). However we still assume that there are only two areas of employment. The wage of the individual worker is therefore given as

$$w_{a(i,t),it} = \iota_{it}^{\top}\sigma + \mathbf{x}_{it}^{\top}\beta + \mu_i + l_{it} + \epsilon_{it} \tag{10}$$

$$l_{it} = \lambda_{rc} e_{cit} + \delta_{cr}\tilde{e}_{rit} + \delta_{cr}\tilde{e}_{cit}$$

where $\mathbf{x}_{it}^{\top}\beta$ includes the wage pay-off of the experience $\phi_{rr}e_{it}$ not distinguished by the origin or the place of use.

We first consider the bias of the city fixed effects estimates if the econometrician fails to control both for the unobserved individual effects $\mu_i$ and the learning effects

$l_{it}$. The workers employed in the city have higher unobserved fixed effects $\mu_i$ than workers in general. This induces the $bias(\hat{\sigma}_c) := \hat{\sigma}_c - \sigma_c$ in a positive direction. The same is the case for the learning effects under the assumption that the combined gain of the city experience relative to the rural areas $\lambda_{cr} + \delta_{cc}$ is positive. Moreover, workers with the city experience are not uniformly distributed across the urban landscape, hence the city workers on average have more city experience. When this experience is valued higher, the city area becomes more productive. Failing to control for this explicitly therefore affects the $bias(\hat{\sigma}_c) := \hat{\sigma}_c - \sigma_c$ positively. See Appendix B for a detailed derivation.

We now consider the bias of the city fixed effect estimates if the econometrician fails to control for the leaning effect but uses the within estimator to control for the unobserved individual fixed effects $\mu_i$. In this case, it is only workers who change their area of employment who identify the area fixed effects. Under the assumption that the city premiums of experience $\lambda_{cr}$, $\delta_{cc}$ and $\delta_{cr}$ are all positive, the workers who change their area of employment from the rural area to the city will - while working in the city - have a higher value of experience than their average value of experience and will, therefore, affect the $bias(\hat{\sigma}_c) := \hat{\sigma}_c - \sigma_c$ positively. The workers who migrate away from the city will, on the other hand, affect the $bias(\hat{\sigma}_c) := \hat{\sigma}_c - \sigma_c$ negatively, unless it is the case that the value of the city experience to a very high degree is not portable. For a detailed derivation see Appendix C.

# 4 Empirical results

This section presents the empirical results. We first estimate a specification ignoring both the learning effect and the unobserved worker heterogeneity. The results of this estimation are given in subsection 4.1. We then in subsection 4.2 include individual fixed effects but still do not control for learning effects. Finally, in subsection 4.3 we also consider the learning effects.

## 4.1 The static wage gains of density

In this section we estimate the static wage gains of density not controlling for the individual fixed effects or the learning effect. The wage equation is therefore given by

$$w_{a(i,t),it} = \iota_{it}^{\top}\sigma + \mathbf{x}_{it}^{\top}\beta + \gamma_s + \eta_t + v_{it}, \tag{11}$$

where $\gamma_s$ are sector fixed effects and $\eta_t$ are year fixed effects.

The results of the estimation are reported in Table 4 column (1). The wage value of the experience is concave. The first year of experience increases wages by 4.5% and the fifth year of the experience by 3.1%.[6] Job tenure is also concave. The first year of employment at a workplace implies a wage increase of 1.1%. The wage increase by the fifth year of employment at the same workplace is reduced to 0.5%. As expected, the wages increase with the level of occupational skill and with the level of education.

Column (2) in Table 4 shows the results of regressing the estimates of the area fixed effects against the log density. We instrument employment density with the population densities for the years 1801 and 1834. The elasticity of wages with respect to employment density is estimated to 0.0226.[7].

The estimate of the elasticity of wage with respect to employment density is likely biased due to the bias in the estimates of the area fixed effects. This bias was found in the two-area example to be positive for the high-density areas, assuming that workers in these areas have higher values of the individual fixed effects $\mu_i$ and that the experience gained and used in the high-density areas has a higher value $(\lambda_{cr} + \delta_{cc}) > 0$ (see equation 17). It was also assumed that workers did not change their areas of employment. This assumption is justified by the fact that, when individual fixed effects are not included, the area fixed effects are also identified by the workers who do not change the employment area. These workers make up the majority of workers in our sample.

[6] These increases as calculated as $(\exp(0.0453 - 0.0016) - 1) \cdot 100$ and as $(\exp(0.0453 \cdot 5 - 0.0016 \cdot 5^2 - (0.0453 \cdot 4 - 0.0016 \cdot 4^2)) - 1) \cdot 100$.

[7] We have also estimated the elasticity of wages with respect to employment density using the area-year fixed effects. It is then 0.0176. We have finally estimated the elasticity of wages with respect to effective density using area-year fixed effects. This elasticity is 0.0182 (see Table 5 in Hybel (2020)). The effective density is a transportation time-weighted employment density. This method has been used by Combes et. al. (2010).

Table 4: Estimation of the static wage gains of density

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | Log wage | Area indicator | Log wage | Area indicator |
| | (1) | (2) | (3) | (4) |
| Log density | | 0.0226*** | | 0.0107*** |
| | | (0.0033) | | (0.0017) |
| Area fixed effects | + | | + | |
| Individual fixed effects | - | | + | |
| Experience | 0.0453*** | | 0.0408*** | |
| | (0.0002) | | (0.0004) | |
| Experience$^2$ | −0.0016*** | | −0.0014*** | |
| | (0.00001) | | (0.00001) | |
| Job tenure | 0.0121*** | | 0.0097*** | |
| | (0.0001) | | (0.0001) | |
| Job tenure$^2$ | −0.0008*** | | −0.0006*** | |
| | (0.00001) | | (0.00001) | |
| Medium skilled occupation | 0.1583*** | | 0.0209*** | |
| | (0.0005) | | (0.0005) | |
| High skilled occupation | 0.3351*** | | 0.0644*** | |
| | (0.0011) | | (0.0008) | |
| Secondary Education | 0.0487*** | | 0.1906*** | |
| | (0.0005) | | (0.0017) | |
| University Education | 0.2040*** | | 0.3202*** | |
| | (0.0008) | | (0.0023) | |
| Male | 0.1412*** | | | |
| | (0.0005) | | | |
| Sector fixed effects | + | | + | |
| Year fixed effects | + | | + | |
| Observations | 7,246,703 | 98 | 7,246,703 | 98 |
| $R^2$ | 0.4163 | - | 0.8630 | - |

*Note:* *p<0.1; **p<0.05; ***p<0.01

## 4.2 The static wage gains of density and the individual fixed effects

In this section, we estimate the static wage gains of density controlling for the individual fixed effects but still not for the learning effects. The wage equation is now given by

$$w_{a(i,t),it} = \iota_{it}^{\top}\sigma + \mu_i + \mathbf{x}_{it}^{\top}\beta + \gamma_s + \eta_t + u_{it}. \tag{12}$$

The results of the estimation are reported in Table 4 column (3). The coefficients for experience and job-tenure are robust to the inclusion of the individual fixed effects. The wage value of the first year of experience is slightly reduced to 4.0% (versus 4.5%) and the fifth year of experience to 2.9% (versus 3.1%). The wage benefit for job tenure is also slightly reduced, with a wage increase for the first year of employment of 0.9% (versus 1.1%), while the wage increase of the fifth year of employment is reduced to 0.4% (versus 0.5%). Also in this specification, the wage increases with the level of occupational skill and the level of education. The estimates are however less robust to the addition of the individual fixed effects and are associated with the significantly higher wage benefit.

Column (4) in Table 4 shows the results of regressing the estimates of the area fixed effects against the log density of the area using historical instruments of population density for the years 1801 and 1834. The elasticity of wage with respect to employment density is reduced with more than 50% (from 0.0226 to 0.0107). This is close to the 47% reduction reported by Puga and De La Roca (2017).[8]

Ignoring the learning effects likely results in a biased elasticity of wage with respect to employment density. When individual fixed effects are included in the estimation equation, only workers who change employment areas identify the area fixed effects. Workers changing their area of employment to the high-density areas bias the estimates of area fixed effects positively if the city premium and the city use premium are positive (see equation 30). This happens because workers who relocate are more likely to have a value of experience greater than their average value of experience while working in the high-density area. On the other hand, workers moving to a low-density area can bias the estimates for the area fixed

[8]Using data for Denmark, Knudsen, Hjorth, and Pilegaard 2019 and De Borger, Mulalic, and Rouwendal 2019 estimate the elasticity of wage with respect to job accessibility of a similar magnitude.

effects either positively or negatively, depending on how portable is the value of the accumulated experience (see equation 31). If the value of experience is extremely portable, these workers will have a value of experience that is lower than their average value of experience, while they work in the high-density area, inducing a negative bias of the high-density area fixed effects.

## 4.3 Dynamic wage benefits of density

In this section we estimate both the static and dynamic wage gains of density. The wage equation we estimate is now

$$w_{a(i,t),it} = \iota_{it}^{\top}\sigma + \mu_i + \mathbf{x}_{it}^{\top}\beta + l_{it} + \gamma_s + \eta_t + \epsilon_{it} \tag{13}$$

$$l_{it} = \sum_{g \neq g_0} \lambda_g e_{git} + \sum_{g \neq g_0} \alpha_g e_{git} e_{it} + \sum_{g} \delta_g \tilde{e}_{git} + \sum_{g} \psi_g \tilde{e}_{git} e_{it},$$

where the learning effects $l_{it}$ include both linear and non-linear effects. Specifically we allow for the city premiums $\lambda_g e_{git} + \alpha_g e_{git} e_{it}$ to depend on the workers total experience $e_{it}$ and similarly for the city use premiums $\delta_g \tilde{e}_{git} + \psi_g \tilde{e}_{git} e_{it}$.

The areas are divided into four different groups $g \in \{1, ..., 4\}$ depending on their level of employment density: three groups with the high employment density collective referred to as *the top* and the final group of comparatively low employment density (the reference group). The group with the highest level of employment density consists of municipalities Copenhagen and Frederiksberg which together make up the center of the Greater Copenhagen area, the largest urban area in Denmark. The group with the second-highest density consists of municipalities located in the proximity of this center and are thus all part of the Greater Copenhagen area. The third group includes dense municipalities in mostly rural areas including second-largest city in Denmark, Aarhus. The low employment density group includes periphery, i.e. smaller towns and rural areas.

The estimation results are reported in Table 5 column (1). The wage gain of the experience accumulated outside the top is 3.4% (versus 4.0%) for the first year and 2.4% (versus 2.9%) for the fifth year when the worker uses the experience outside the top region. This is a reduction compared to similar estimates in the model given in equation 12 without learning effects. The wage gain of experience accumulated in the low-density areas is higher if it is used in the top areas. The coefficient of experience from the areas of lowest density, when used in the top, is 0.0067 and the second-order term is -0.0004. The first year of experience accumulated in the low-density area gives thus a 4.0% wage increase, while the fifth year gives a 2.7% wage increase. The experience gained outside the high-density areas is rewarded higher when used in the high-density areas. How much a worker gains from working in the high-density area depends therefore on the level of the accumulated experience. This result is in line with the results of Puga and De La Roca (2017).

The experience accumulated in the areas of the highest and the second-highest density is rewarded higher independent of where it is used. This is apparent from the coefficients of 0.0147 and 0.0078 for experience gained in the areas of highest density and for the experience gained in the areas of second-highest density, respectively. The first year of experience from the highest density areas is 1.5% more rewarding than the first year of experience from the low-density areas and 0.8% for the areas of the second-highest density. The experience from the areas of the third-highest density is insignificantly different from the experience accumulated in the low-density areas.

The value of the experience accumulated in the high-density areas is only partly portable. Experience gained in the top - the areas of highest, second highest, and third highest density - is rewarded higher when used in the top. This implies for example, that the wage gain from the experience accumulated and used in the areas of the highest density is 5.9% for the first year and 4.0% for the fifth year, while the portable gain is slightly lower at respectively 4.9% for the first year and 3.6% for the fifth year.

Finally, column (2) in Table 5 shows that the estimate of the elasticity of wage with respect to employment density, conditional on the learning effects, is reduced again with about 50%. Notice however that for workers with the 7 years of experience accumulated in the high-density area (city) this elasticity increases more than 4 times (from 0.0052 to 0.0228), see column (3) in Table 5.

### 4.3.1 Earnings Profiles

In this section, we illustrate the results from Table 5 using earning profiles. We assume that workers initially have zero years of experience and accumulate one year of experience each year, for every year of the period under consideration. We

Table 5: Estimation of the dynamic and static wage gains of density

| | *Dependent variable:* | | |
|---|---|---|---|
| | Log wage | Area indicator | Area indicator + City premium (7 years of experience) |
| | (1) | (2) | (3) |
| Log density | | 0.0052*** | 0.0228*** |
| | | (0.0017) | (0.0021) |
| Experience | 0.0344*** | | |
| | (0.0004) | | |
| Experience$^2$ | −0.0012*** | | |
| | (0.00002) | | |
| Experience (highest density) | 0.0147*** | | |
| | (0.0010) | | |
| Experience (second highest density) | 0.0078*** | | |
| | (0.0009) | | |
| Experience (third highest density) | 0.0003 | | |
| | (0.0013) | | |
| Experience (highest density) × experience | −0.0003*** | | |
| | (0.0001) | | |
| Experience (second highest density) × experience | −0.0001 | | |
| | (0.0001) | | |
| Experience (third highest density) × experience | 0.0001 | | |
| | (0.0001) | | |
| Experience (lowest density) used in top | 0.0067*** | | |
| | (0.0007) | | |
| Experience (highest density) used in top | 0.0107*** | | |
| | (0.0011) | | |
| Experience (second highest density) used in top | 0.0079*** | | |
| | (0.0010) | | |
| Experience (third highest density) used in top | 0.0094*** | | |
| | (0.0013) | | |
| Experience (lowest density) × experience used in top | −0.0004*** | | |
| | (0.0001) | | |
| Experience (highest density) × experience used in top | −0.0008*** | | |
| | (0.0001) | | |
| Experience (second highest density) × experience used in top | −0.0005*** | | |
| | (0.0001) | | |
| Experience (third highest density) × experience used in top | −0.0006*** | | |
| | (0.0001) | | |
| Job tenure | 0.0097*** | | |
| | (0.0001) | | |
| Job tenure$^2$ | −0.0006*** | | |
| | (0.00001) | | |
| Medium skilled occupation | 0.0211*** | | |
| | (0.0005) | | |
| High skilled occupation | 0.0641*** | | |
| | (0.0008) | | |
| Secondary Education | 0.1939*** | | |
| | (0.0017) | | |
| University Education | 0.3156*** | | |
| | (0.0023) | | |
| Observations | 7,246,703 | 98 | 98 |
| $R^2$ | 0.8635 | 0.2590 | 0.7031 |

*Note:* *p<0.1; **p<0.05; ***p<0.01

also assume that workers change the employment area after five years.

Figure 5: Migration to high density area

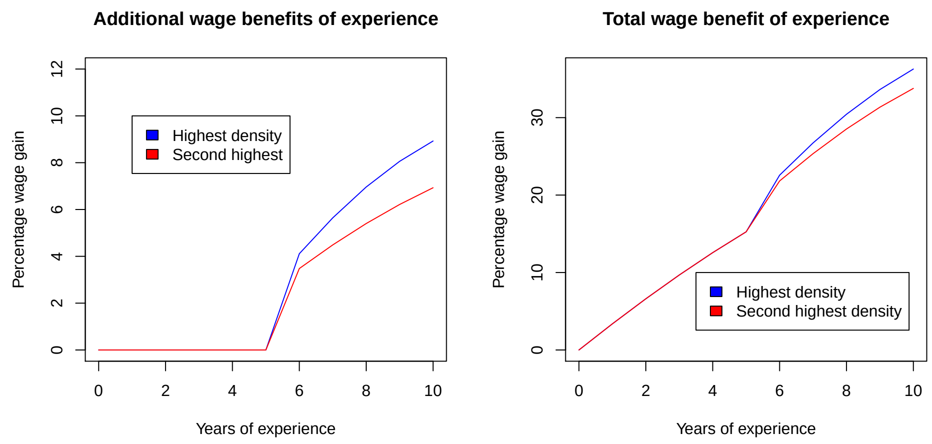



Figure 5 illustrates the case of a worker who relocates from the low-density area to either the highest density area (blue wage-path) or the second-highest density area (red wage-path). The left panel shows the wage benefit of the learning effects as defined in equation (13). These wage benefits do not include benefits of the experience accumulated and used in the low-density areas. Because the areas of the low employment density are used as the reference, these benefits are part of the wage benefits earned by the experience accumulated and used anywhere. For the first five years, the worker accumulates and uses experience in an area with low employment density. The wage benefits of the learning effects are therefore zero. However, as illustrated in the right panel of Figure 5, the wage still increases due to the benefit of experience accumulated and used in a low-density area.[9]

When a worker relocates, the additional wage benefits of experience increases to approximately 4%. This increase consists of two parts. The first part is due to the additional wage benefit of using the already acquired experience in a new employment area. The size of this part does not depend on whether the worker migrates to an area of highest employment density or second-highest employment density. The size does, however, depend on how much experience the worker has

[9]The wage benefit of the experience accumulated and used in the low-density area is determined by the coefficients of the experience and experience squared in Table 5.

accumulated before the relocation. The second part shows one additional year of experience accumulated from year 5 to year 6, in the new area of employment. It depends on (i) the wage benefit of experience accumulated in the new employment area, and (ii) the wage benefit of the experience accumulated in the new area of employment when used in the top density area. These two components combined are larger for the areas of the highest employment density compared to the second-highest employment density area (see Table 5). This difference explains why the jump from year 5 to year 6 is slightly larger when a worker migrates to the top density area, rather than to the area of the second-highest employment density. Furthermore, it also explains why the growth of the additional benefits is larger for the high-density areas. After 5 years of working in the highest density area, the additional wage benefit reaches a level of approximately 6-8%.[10]

Figure 6 illustrates the reverse case, i.e. when a worker relocates from a high density to a low-density area. The left panel of Figure 6 shows how the additional wage benefit increases, the longer the worker is employed in the high-density area. After the first year, the gain of the experience accumulated and used in one of the areas of the highest density relative to the experience accumulated and used in an area of the low density is only 1.5%, while after only 5 years it reaches a level of about 10%. Due to the additional wage benefits of experience accumulated and used in the high-density areas, the additional wage benefits decrease when the worker leaves the city.

The right panel of Figure 6 pictures how a reduction in the additional wage benefit of the accumulated experience results in a decrease in the total benefit of the accumulated experience. Having accumulated as much as five years of experience in the high-density area, workers who migrate to the low-density areas thus have to accept a decrease in the total wage benefits of experience. Hence, the additional wage benefits of experience are not completely portable and are potentially functioning as a strong disincentive for workers with many years of experience to leave the high-density areas, v.i.z. cities.

[10]It is easy to see from Figure 5, that during the period of employment in the area with high employment density, the additional wage benefits are above average. Using equation (30) This implies that workers who migrate to an area of the highest or the second-highest density will bias the estimates of the area fixed effects positively.

Figure 6: Migration from high density area


# 5 Gender and the dynamic wage gains of density

In this section we focus on the gender-specific dynamic gains of density. We estimate the full wage model as specified in equation (13) separately for females and males workers. The results are given in columns (1) and (2) in Table 6.

We first focus on the rewards of experience gained outside the top density areas. Here we find that the wage gains of experience are lower for women than for men when the experience is used outside the top density areas. For example, male worker's wages increase by 3.6% in the first year, compared to a lower 3.3% wage increase for the female workers. For both genders, the top density areas offer opportunities to increase the returns to experience accumulated outside the top density areas. However, the additional increase in wages when using the experience in the top density areas is slightly larger for the female workers. For the first year of experience, male workers get an additional 0.5% wage increase, when using the experience in the top area, while female workers get a higher 0.7% additional wage increase. However, despite this additional gain, male workers still get a higher wage increase for the first year of the experience accumulated in the low-density areas and used in the top density areas, 4.2%, compared to the female workers who get 4.0%. The use of the experience in top density areas by female workers reduces the gender wage gap but does not close it.

We now focus on the rewards of the experience gained in the top density areas

Table 6: Estimation of the dynamic wage gains of density for male and female

| | *Dependent variable:* | |
|---|---|---|
| | Log wage (male) | Log wage (female) |
| | (1) | (2) |
| Experience | 0.0368*** | 0.0332*** |
| | (0.0005) | (0.0007) |
| Experience$^2$ | −0.0012*** | −0.0012*** |
| | (0.00002) | (0.00002) |
| Experience (highest density) | 0.0179*** | 0.0110*** |
| | (0.0014) | (0.0015) |
| Experience (second highest density) | 0.0086*** | 0.0072*** |
| | (0.0012) | (0.0015) |
| Experience (third highest density) | 0.0014 | −0.0015 |
| | (0.0017) | (0.0021) |
| Experience (highest density) | −0.0004*** | −0.0004*** |
| × experience | (0.0001) | (0.0001) |
| Experience (second highest density) | −0.0002* | 0.00003 |
| × experience | (0.0001) | (0.0001) |
| Experience (third highest density) | 0.00005 | 0.0001 |
| × experience | (0.0001) | (0.0002) |
| Experience (lowest density) | 0.0057*** | 0.0078*** |
| used in top | (0.0009) | (0.0011) |
| Experience (highest density) | 0.0107*** | 0.0099*** |
| used in top | (0.0014) | (0.0015) |
| Experience (second highest density) | 0.0058*** | 0.0107*** |
| used in top | (0.0012) | (0.0015) |
| Experience (third highest density) | 0.0082*** | 0.0118*** |
| used in top | (0.0017) | (0.0022) |
| Experience (lowest density) | −0.0004*** | −0.0004*** |
| × experience used in top | (0.0001) | (0.0001) |
| Experience (highest density) | −0.0009*** | −0.0006*** |
| × experience used in top | (0.0001) | (0.0001) |
| Experience (second highest density) | −0.0004*** | −0.0007*** |
| × experience used in top | (0.0001) | (0.0001) |
| Experience (third highest density) | −0.0006*** | −0.0006*** |
| × experience used in top | (0.0001) | (0.0002) |
| Job tenure | 0.0099*** | 0.0091*** |
| | (0.0001) | (0.0001) |
| Job tenure$^2$ | −0.0006*** | −0.0005*** |
| | (0.00001) | (0.00001) |
| Medium skilled occupation | 0.0221*** | 0.0202*** |
| | (0.0007) | (0.0006) |
| High skilled occupation | 0.0649*** | 0.0639*** |
| | (0.0010) | (0.0014) |
| Secondary Education | 0.2370*** | 0.1101*** |
| | (0.0022) | (0.0024) |
| University Education | 0.3793*** | 0.2022*** |
| | (0.0031) | (0.0032) |
| Observations | 4,441,873 | 2,636,868 |
| $R^2$ | 0.8511 | 0.8735 |

*Note:* *p<0.1; **p<0.05; ***p<0.01

when used outside the top density areas. We find, that for both male and female workers the experience accumulated in the highest and the second-highest density areas is rewarded higher than the experience gained elsewhere when used outside the top density areas. This additional gain from the accumulated experience is lower for the female workers than for the male workers.[11] This implies that the first year of experience accumulated in the highest density area awards male worker with an additional 1.8% wage increase relative to the experience accumulated in the low-density areas. Female workers receive the lower 1.1% additional wage increase. The comparative percentages for the experience accumulated in the areas of second-highest density are 0.9% and 0.7% for the male workers and female workers respectively. The gender gap in the additional reward for the experience accumulated in high-density areas is, therefore, more pronounced in cities.

The experience accumulated in the high-density areas for both genders is rewarded extra when used in these areas. Moreover, only for the experience gained in the area of the highest density, the additional wage increase of using the experience in the top density areas is higher for the male workers than for the female workers. The coefficient of the experience for the highest density areas, when used in the top, is 0.0107 for the male workers and 0.0099 for the female workers (see Table 6). This is equivalent to 1.1% and 1.0% increases in the wages for the male workers and the female workers, respectively. For the experience accumulated in the areas of the second-highest density when used in the top, the female workers get a larger 1.1% wage gain while the male workers get a lower 0.6% wage gain. Hence, the first year of experience when used in the top density areas awards the male workers by 2.8% additional wage increase compared to the experience accumulated in the low-density areas and used in the low-density areas. For female workers, this additional wage increase is only 2.0%. The wage growth rate of working in the areas of highest density thus appears larger than the wage growth rate of working outside the top density areas, with a higher growth rate for the male workers than for the female workers. However, for the experience accumulated in the second-highest areas and used in the top density areas, both genders get almost the same benefit, i.e. 1.36% and 1.40% for the male and the female workers, respectively. For the

[11] For the male workers, the coefficient of the experience accumulated in the highest density area is 0.0179 compared to 0.0110 for the female workers. It is 0.0086 for the areas of the second-highest density for male workers, while only 0.0072 for the female workers (see Table 6).

experience accumulated in the third-highest area and used in the top, the male workers get additionally 0.8%, while the female workers get 1.1%.[12]. High-density areas thus offer male and female workers similar opportunities, i.e. the faster wage growth.

Finally, we compare the share of the additional value of portable experience for male and female workers, accumulated in the high-density areas. For the areas with the highest density, the portable share for the male workers is 0.63 while for the female workers it is 0.53.[13] For the areas of the second-highest density, the portable share for the male workers is 0.60 while for the female workers it is 0.40. Finally for the areas of the third-highest density the portable share is 0 for both genders.[14] For the female workers, the portable part of the value of experience is thus lower than for the male workers, partly due to the lower additional gains from the experience accumulated in high-density areas, but also due to the larger additional benefits from using the accumulated experience in the top density areas relative to the male workers. Therefore, the female workers continually have to work in the high-density areas to have similar benefits as male workers.

## 5.1 Gender and earnings profiles

In this subsection, we illustrate the results shows in Table 6 using earning profiles. We do this because using only percentages for the value of the first year of experience, understates the economic significance of the results since the difference becomes larger as workers accumulate more experience.

[12]This is computed using only the coefficients of experience accumulated in the areas of third highest density when used in the top density areas because the coefficients on the experience accumulated in the areas of the third-highest density are statistically insignificant.

[13]For male workers, the share is calculated using the coefficient of experience accumulated in the highest density areas and the coefficient for experience accumulated in the highest density areas when used in the top, as given in Table 6. Hence the portable share for the male workers is $0.063 = 0.0179/(0.0179 + 0.0107)$, ignoring the second-order effects from the non-linear terms. The calculation of the other shares is performed similarly.

[14]We assume that the coefficients on experience accumulated in the areas of third highest density are 0 because they are not statistically significant (see Table 6).

### 5.1.1 Migration to a low-density area

Figure 7 illustrates the additional wage gains from the experience accumulated in a high-density area for the male and the female workers. These gains are the learning effects defined in equation 13. We assume again that all workers initially have zero years of experience and each year accumulate one year of experience from the area of current employment. For the first 5 years, the workers are assumed to work in the high-density area, then they migrate to one of the low-density areas.

Figure 7: Accumulated additional gains from experience when migrating from the city


The left panel in Figure 7 shows the additional wage gains from the experience accumulated in the high-density area. The gains are larger for male workers. While employed in the high-density area, the additional wage gains from the experience include the additional wage benefit of experience accumulated in the area of the highest density, as well as the additional wage benefit of the experience accumulated in the area of the highest density when used in the top. After 5 years of work in the high-density area, the benefit from an accumulated additional wage increase for the male workers is approximately 12%, compared to only 8% for the female workers.

Both genders experience a drop in the additional wage gains when relocating. This happens because they lose the additional wage benefit of the experience accumulated and used in the high-density area. The size of this reduction is approximately 4%. Furthermore, relative to the additional wage increase, at the time of migration, the reduction is larger for the female workers, who loose 4% out of the 8% gained, while the male workers lose 4% out of the 12% gained. The smaller relative loss for the male workers reflects how their additional wage returns are more portable than the additional wage return for the female workers. After migration, the additional wage returns of experience slightly decrease as the workers accumulate further experience.

The right panel of Figure 7 shows the outcome of relocation from the area of the second-highest density. Now the additional wage gains from the experience are larger for the female workers than for the male workers. After 5 years of work in the high-density area, the female workers have reached an additional wage increase of approximately 7.5% while the male workers have reached approximately 6.5%. When migrating the additional wage gains of experience are reduced to approximately 4%. Therefore, the reduction is the largest for female workers.

Figure 8 illustrates the total wage gains of the experience. Here the gains of experience earned and used anywhere are added to the additional gains illustrated in Figure 7. The left panel in Figure 8 illustrates again migration from a high to a low-density area. For both genders, the wages increase faster for the first 5 years of employment in the high-density area compared to the latter 5 years of employment in the low-density area. The reduction in the wage growth rate is due to workers no longer receive the benefits of using the accumulated experience in the top, as well as due to accumulating low-density experience, rather than high-density experience[15]. In conclusion, the wage increases faster for the male workers than for the female workers both for the first 5 years of employment in the high-density area as well as for the latter 5 years of employment in the low-density area.

[15]There is also a third effect due to the non-linearity of the learning effects. This effect is because the worker's experience accumulated in the high-density area becomes less worth for every year of experience accumulated in the low-density area. This effect is very little as shown in Figure 7 and is therefore ignored.

Figure 8: Total accumulated gains from experience when migrating from the city


# 6 Conclusions

The gender pay gap is a long-standing inequality and its origins are multidimensional. Empirically, we know very little about the impact of the agglomeration economies on this inequality. This article seeks to identify the gender-specific urban wage premium. Using register data for workers in Denmark, we first prove the existence of an urban premium for wage levels and a city size premium on wage growth. The estimated effects imply individual-level compensating differentials for agglomeration economies as predicted by urban economic models that allow for productivity advantages emerging from improved sharing, matching, or learning in dense labor markets (Roca and Puga 2017; Duranton and Puga 2004). We also identify three empirical facts about the gender-specific urban wage premium: i) wages and the share of female workers increase with job density, ii) distributions of the work experience are similar for both genders, and iii) female workers use the accumulated experience more intensively in cities. Finally, our empirical findings suggest that the value of the portable part of the accumulated work experience for the female workers is below that of the male workers. This explains, at least partially, the puzzle of simultaneously positive correlations between, wages and the share of female workers on the one hand, and between job density and significantly higher mean hourly wages for the male workers on the other.

Policymakers and academics who are interested in the gender wage gap, agglomeration economies, and urbanization may be interested in our results. We emphasize that our results do not say anything explicitly about the gender pay gap because it might reflect broader inequalities in society, but do indicate that female workers gain less from agglomeration. It is plausible that our results do not hold for other countries with different labor market structure. It would be interesting to apply the methodology introduced to countries with a larger number of cities that vary in size to examine the underlying mechanisms in more detail.

## References

Blau, Francine D. and Lawrence M. Kahn (2017). "The Gender Wage Gap: Extent, Trends, and Explanations". In: *Journal of Economic Literature* 55.3, pp. 789–865.

Borghorst, M., I. J. Mulalic, and J. van Ommeren (2020). *Commuting and the Gender Pay Gap.* Technical Report. Manuscript.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon (2008). "Spatial Wage Disparities: Sorting Matters!" In: *Journal of Urban Economics* 63.2, pp. 723–742.

Combes, Pierre-Philippe and Laurent Gobillon (2015). "The Empirics of Agglomeration Economies". In: *Handbook of Regional and Urban Economics.* Vol. 5. Elsevier, pp. 247–348.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux (2010). "Estimating Agglomeration Economies with History, Geology, and Worker Effects". In: *Agglomeration Economics.* Ed. by Edward L. Glaeser. University of Chicago Press.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux (2012). "The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection". In: *Econometrica* 80.6, pp. 2543–2594.

De Borger, Bruno, Ismir Mulalic, and Jan Rouwendal (2019). *Productivity Effects of an Exogenous Improvement in Transport Infrastructure: Accessibility and the Great Belt Bridge.* Tech. rep. 065/VIII.

Duranton, Gilles and Diego Puga (2004). “Chapter 48 Micro-Foundations of Urban Agglomeration Economies”. In: *Handbook of Regional and Urban Economics*. Vol. 4. Elsevier, pp. 2063–2117.

Enrico, Moretti (2011). “Local Labor Markets”. In: *Handbook of Labor Economics*. Vol. 4. Elsevier, pp. 1237–1313.

Gaubert, Cecile (2018). “Firm Sorting and Agglomeration”. In: *American Economic Review* 108.11, pp. 3117–53.

Glaeser, Edward L. and David C. Maré (2001). “Cities and Skills”. In: *Journal of Labor Economics* 19.2, pp. 316–342.

Gutiérrez-i-Puigarnau, Eva, Ismir Mulalic, and Jos N. van Ommeren (2016). “Do Rich Households Live Farther Away from Their Workplaces?” In: *Journal of Economic Geography* 16.1, pp. 177–201.

Henderson, J.Vernon (2003). “Marshall’s Scale Economies”. In: *Journal of Urban Economics* 53.1, pp. 1–28.

Hirsch, Boris, Marion König, and Joachim Möller (2013). “Is There a Gap in the Gap? Regional Differences in the Gender Pay Gap”. In: *Scottish Journal of Political Economy* 60.4, pp. 412–439.

Holmes, Thomas J. (2010). “Structural, Experimentalist and Descriptive Approaches to Empirical Work in Regional Economics”. In: *Journal of Regional Science* 50.1, pp. 5–22.

Hybel, Jesper and Ismir Mulalic (2020). *Transportation and Quality of Life*. Working Paper. DTU.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard (2018). *Children and Gender Inequality: Evidence from Denmark. Tech. Rep. National Bureau of Eco- Nomic Research.* Technical Report. National Bureau of Economic Research.

Knudsen, E.S., K. Hjorth, and N. Pilegaard (2019). *Wages and Accessibility – Evidence from Denmark*. Tech. rep. DTU.

Melo, Patricia C., Daniel J. Graham, and Robert B. Noland (2009). “A Meta-Analysis of Estimates of Urban Agglomeration Economies”. In: *Regional Science and Urban Economics* 39.3, pp. 332–342.

Mulalic, Ismir and Jan Rouwendal (2020). “Does Improving Public Transport Decrease Car Ownership? Evidence from a Residential Sorting Model for the

Copenhagen Metropolitan Area". In: *Regional Science and Urban Economics* 83, p. 103543.

Phimister, Euan (2005). "Urban Effects on Participation and Wages: Are There Gender Differences?" In: *Journal of Urban Economics* 58.3, pp. 513–536.

Puga, Diego (2010). "The Magnitude and Causes of Agglomeration Economies". In: *Journal of Regional Science* 50.1, pp. 203–219.

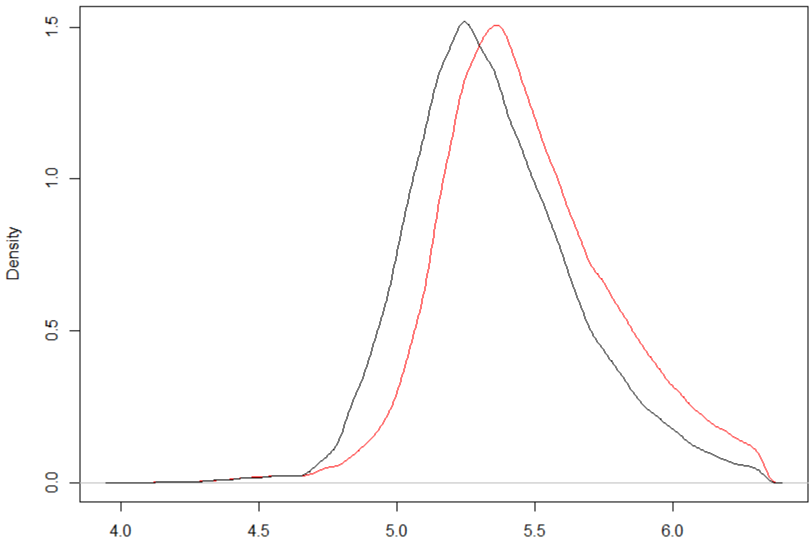
Roca, Jorge De La and Diego Puga (2017). "Learning by Working in Big Cities". In: *The Review of Economic Studies* 84.1, pp. 106–142.

Rosenthal, Stuart S. and William C. Strange (2004). "Chapter 49 Evidence on the Nature and Sources of Agglomeration Economies". In: *Handbook of Regional and Urban Economics*. Vol. 4. Elsevier, pp. 2119–2171.

— (2012). "Female Entrepreneurship, Agglomeration, and a New Spatial Mismatch". In: *Review of Economics and Statistics* 94.3, pp. 764–788.

# Appendix A: Data

Figure 9: Distributions of the log hourly wages for male workers (red) and female workers (black)




Notes: This figure depicts the Gaussian kernel density distribution of the hourly wages by gender for the period 2008 until 2016

# Appendix B: Ignoring individual fixed effects and learning effects

In the case where both individual fixed effects and learning effects are ignored the econometrician uses the estimation equation

$$w_{a(i,t),it} = \iota_{it}^{\top}\sigma + \mathbf{x}_{it}^{\top}\beta + v_{it}, \tag{14}$$

implying that the estimates $(\hat{\sigma}_c, \hat{\sigma}_r)$ of the static area effects can be written as

$$\begin{pmatrix} \hat{\sigma}_c \\ \hat{\sigma}_r \end{pmatrix} = \begin{pmatrix} \sigma_c \\ \sigma_r \end{pmatrix} + (\mathbf{D}^{\top}\mathbf{D})^{-1}\mathbf{D}^{\top}\mathbf{X}(\beta - \hat{\beta}) + (\mathbf{D}^{\top}\mathbf{D})^{-1}\mathbf{D}^{\top}\mathbf{v}, \tag{15}$$

where $\mathbf{D}^{\top} = [\mathbf{D}_1^{\top}, ..., \mathbf{D}_N^{\top}]$, $\mathbf{D}_i^{\top} = [\iota_{i1}, ..., \iota_{iT}]$ and with $\mathbf{X}$ and $\mathbf{v}$ being defined analogously to $\mathbf{D}^{\top}$ as stacked matrices of $\mathbf{x}_{it}^{\top}$ and $v_{it}$ respectively. The error term $v_{it} = l_{it} + \mu_i + \epsilon_{it}$ contains both the left out learning effects and the left out unobserved fixed effects. To derive an approximate formula of the bias we assume that $\mathbf{D}$ is orthogonal to $\mathbf{X}$ such that the only source of bias stems from $(\mathbf{D}^{\top}\mathbf{D})^{-1}\mathbf{D}^{\top}\mathbf{v}$. The matrix $\mathbf{D}^{\top}\mathbf{D}$ is diagonal with diagonal terms $\sum_i T(i,a)$ where $T(i,a) := \sum_{t=1}^{T} 1[a(i,t) = a]$ is the duration of time individual $i$ is employed in area $a$. In this case the bias can therefore be written as

$$bias(\hat{\sigma}_a) = \frac{\sum_{i=1}^{N} T(i,a)\mu_i}{\sum_{i=1}^{N} T(i,a)} + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} 1[a(i,t) = a]l_{it}}{\sum_{i=1}^{N} T(i,a)}, \tag{16}$$

so a sum of duration weighted averages of the unobserved individual characteristic $\mu_i$ and duration weighted averages of learning effects $l_{it}$ (see the following subsection for a more detailed derivation).

In the case where no worker changes area of employment, where $\mu_i = \mu$ for all workers in the city and where $\mu_i = 0$ for workers in the rural area, it follows that

$$\frac{\sum_{i=1}^{N} T(i,a)\mu_i}{\sum_{i=1}^{N} T(i,a)} = \mu.$$

Because no worker changes area of employment, the city workers have no rural experience such that the rural experience used in the city $\tilde{e}_{rit} = 0$. No migration also implies that their city experience $e_{cit}$ is equal to their city experience used in the city $\tilde{e}_{cit} := 1[a(i,t) = c]e_{cit}$. It therefore follows that the learning effects for a city worker is given as

$$l_{it} = (\lambda_{cr} + \delta_{cc})e_{cit},$$

in which case the total bias becomes

$$\mu + (\lambda_{cr} + \delta_{cc})\frac{\sum_{i=1}^{N}\sum_{t=1}^{T} 1[a(i,t) = c]e_{cit}}{\sum_{i=1}^{N} T(i,a)} = \mu + (\lambda_{cr} + \delta_{cc})\left(\frac{T+1}{2}\right), \tag{17}$$

where $(T+1)/2$ is the average number of years of city experience of a worker employed in the city.

## 6.1 Technical note

To derive the bias we first consider estimation of the panel data model with static area specific effects $\sigma = (\sigma_1, ..., \sigma_J)^\top$ defined by

$$w_{a(it),it} = \mathbf{z}_{it}^\top\theta + u_{it} = \iota_{it}^\top\sigma + \mathbf{x}_{it}^\top\beta + u_{it} \tag{18}$$

where $\mathbf{z}_{it}^\top := (\iota_{it}^\top \; \mathbf{x}_{it}^\top)$, $\theta := (\sigma^\top, \beta^\top)^\top$ and $\iota_{it}^\top := (1[a(i,t) = a_1], ..., 1[a(i,t) = a_J])$ is the vector of dummy variables for the location of employment of individual $i$ at time $t$. Stacking the model in the time index to get

$$\mathbf{w}_i = \mathbf{Z}_i\theta + \mathbf{u}_i = \mathbf{D}_i\sigma + \mathbf{X}_i^\top\beta + \mathbf{u}_i, \tag{19}$$

and defining the associated pooled OLS estimator $\hat{\theta}_N$ as the solution to the normal equations

$$\left(\sum_i \mathbf{Z}_i^\top\mathbf{Z}_i\right)\hat{\theta}_N = \left(\sum_i \mathbf{Z}_i^\top\mathbf{w}_i\right), \tag{20}$$

it follows by matrix partition that

$$\begin{pmatrix} \sum_i \mathbf{D}_i^\top\mathbf{D}_i & \sum_i \mathbf{D}_i^\top\mathbf{X}_i \\ \sum_i \mathbf{X}_i^\top\mathbf{D}_i & \sum_i \mathbf{X}_i^\top\mathbf{X}_i \end{pmatrix}\begin{pmatrix} \hat{\sigma}_N \\ \hat{\beta}_N \end{pmatrix} = \begin{pmatrix} \sum_i \mathbf{D}_i^\top\mathbf{w}_i \\ \sum_i \mathbf{X}_i^\top\mathbf{w}_i \end{pmatrix}, \tag{21}$$

using the equations associated with $\hat{\sigma}_N$ it follows that

$$\left(\sum_i \mathbf{D}_i^\top\mathbf{D}_i\right)\hat{\sigma}_N + \left(\sum_i \mathbf{D}_i^\top\mathbf{X}_i\right)\hat{\beta}_N = \sum_i \mathbf{D}_i^\top\mathbf{w}_i \tag{22}$$

where we then substitute $\mathbf{w}_i$ with the model equation to get

$$\hat{\sigma}_N = \sigma + \left(\sum_i \mathbf{D}_i^\top \mathbf{D}_i\right)^{-1} \left(\sum_i \mathbf{D}_i^\top \mathbf{X}_i\right)(\beta - \hat{\beta}_N) + \left(\sum_i \mathbf{D}_i^\top \mathbf{D}_i\right)^{-1} \left(\sum_i \mathbf{D}_i^\top \mathbf{u}_i\right), \tag{23}$$

multiplying and dividing with $N$ and taken probability limits under the assumption that $\mathbb{E}[\mathbf{D}_i^\top \mathbf{X}_i] = \mathbf{0}$ it follows that

$$plim\ \hat{\sigma}_N = \sigma + plim\ \left(\frac{1}{N}\sum_i \mathbf{D}_i^\top \mathbf{D}_i\right)^{-1} \left(\frac{1}{N}\sum_i \mathbf{D}_i^\top \mathbf{u}_i\right). \tag{24}$$

The error term $u_{it} = \mu_i + l_{it} + \epsilon_{it}$ when the true model includes individual fixed effects and learning effects as in model (1) implying that

$$plim\ \hat{\sigma}_{N,a} = \sigma_a + plim\ \frac{\sum_{i=1}^N T(i,a)\mu_i}{\sum_{i=1}^N T(i,a)} + plim\ \frac{\sum_{i=1}^N \sum_{t=1}^T 1[a(i,t) = a] l_{it}}{\sum_{i=1}^N T(i,a)}, \tag{25}$$

under the assumption that $\mathbb{E}[\mathbf{D}_i^\top \epsilon_i] = \mathbf{0}$ where $T(i,a) := \sum_{t=1}^T 1[a(i,t) = a]$.

In the more general case where the assumption that $\mathbb{E}[\mathbf{D}_i^\top \mathbf{X}_i] = \mathbf{0}$ is not imposed the expression for the probability limit includes further the term

$$-\left(\frac{1}{N}\sum_i \mathbf{D}_i^\top \mathbf{D}_i\right)^{-1} \left(\frac{1}{N}\sum_i \mathbf{D}_i^\top \mathbf{X}_i\right)(plim\ \hat{\beta}_N - \beta). \tag{26}$$

# Appendix C: Ignoring learning effects

In the second case the econometrician estimates the equation

$$w_{a(i,t),it} = \iota_{it}^\top \sigma + \mu_i + \mathbf{x}_{it}^\top \beta + u_{it}, \tag{27}$$

controlling for the unobserved individual fixed effects while still ignoring the learning effects. In this case the city fixed effects are only identified by the workers who change their area of employment. This can be seen by using dot notation $\dot{z}_{it} := z_{it} - (1/T)\sum_t z_{it}$ and writing the wage equation time demeaned

$$\dot{w}_{a(i,t),it} = \sum_a \sigma_a(\iota_{ait} - \bar{\iota}_{ai}) + \dot{\mathbf{x}}_{it}^\top \beta + \dot{u}_{it}, \tag{28}$$

where $\iota_{ait} := 1[a(i,t) = a]$ such that for any worker not changing area of employment $(\iota_{ait} - \bar{\iota}_{ai}) = 0$. Because the time demeand learning effects are contained in the error term $\dot{u}_{it} = \dot{l}_{it} - \dot{\epsilon}_{it}$ the bias will in general depend on the time demeand value of experience $\dot{l}_{it}$. Specifically it can be shown that the probability limit of the estimate of the city fixed effect is given as

$$plim\ \hat{\sigma}_c = \sigma_c + plim\ \left(\frac{1}{N}\sum_i T(i,c)\left(1 - \frac{T(i,c)}{T}\right)\right)^{-1}\left(\frac{1}{N}\sum_i \mathbf{d}_i^\top \dot{\mathbf{l}}_i\right), \qquad (29)$$

the derivation of which is given in the following subsection.

Consider the first the case where migration is from the rural area to the city. For the first $m$ periods the workers are employed in the rural area, accumulating one year of experience each year. Their rural experience is therefore given as

$$e_{rit} = \begin{cases} t & \text{if } t \leq m \\ m & \text{if } t > m \end{cases},$$

and their city experience is given as

$$e_{cit} = \begin{cases} 0 & \text{if } t \leq m \\ t - m & \text{if } t > m \end{cases}.$$

Since the workers have no city experience prior to migration the city experience used in the city $\tilde{e}_{cit}$ is equal to the city experience $e_{cit}$ such that the learning effects become

$$l_{it} = (\lambda_{cr} + \delta_{cc})e_{cit} + \delta_{cr}\tilde{e}_{rit}.$$

The sum of the time demeaned learning effects while working in the city $\mathbf{d}_i^\top \dot{\mathbf{l}}_i$ can therefore be written as the sum of the two components of experience

$$(\lambda_{cr} + \delta_{cc})\sum_{t=1}^{T} 1[a(i,t) = c]\dot{e}_{cit} + \delta_{cr}\sum_{t=1}^{T} 1[a(i,t) = c]\dot{\tilde{e}}_{rit}.$$

The first component $\sum_{t=1}^{T} 1[a(i,t) = c]\dot{e}_{cit}$ is equal to $\sum_{t=1}^{T} 1[a(i,t) = c]e_{cit} - ((T-m)/T)\sum_{t=1}^{T} 1[a(i,t) = c]e_{cit}$. Isolating the factor $\sum_{t=1}^{T} 1[a(i,t) = c]e_{cit} = (T-m)(T-m+1)/2$ it follows that

$$\sum_{t=1}^{T} 1[a(i,t) = c]\dot{e}_{cit} = \frac{(T-m+1)}{2}(T-m)\left(1 - \frac{T-m}{T}\right).$$

For the rural experience used in the city the component $\sum_{t=1}^{T} 1[a(i,t) = c]\dot{\tilde{e}}_{rit} = \sum_{t=1}^{T} 1[a(i,t) = c]\tilde{e}_{rit} - ((T-m)/T)\sum_{t=1}^{T} 1[a(i,t) = c]\tilde{e}_{rit}$. Now isolating the factor $\sum_{t=1}^{T} 1[a(i,t) = c]\tilde{e}_{rit} = m(T-m)$ and use that in all the $(T-m)$ periods where rural experience is used in the city, it is $m$ years of experience determined as the number of periods of previous employment in the rural area. We therefore find that

$$\sum_{t=1}^{T} 1[a(i,t) = c]\dot{e}_{rit} = m(T-m)\left(1 - \frac{T-m}{T}\right).$$

Dividing this with the factor $T(i,c)\left(1 - \frac{T(i,c)}{T}\right) = (T-m)\left(1 - \frac{T-m}{T}\right)$ it follows that the bias is given as

$$bias(\hat{\sigma}_c) = (\lambda_{cr} + \delta_{cc})\frac{(T-m+1)}{2} + \delta_{cr}m, \tag{30}$$

showing that the workers who migrate to the city will have a value of experience greater than their average value of experience under the assumption that $(\lambda_{cr}+\delta_{cc})$ and $\delta_{cr}$ are both positive.

Next we consider the case where workers are employed for the first $m$ periods in the city and for the latter $T-m$ periods in the rural area. Because the workers have no rural experience while working in the city the learning effects while employed in the city can be written as

$$\dot{l}_{it} = \lambda_{cr}\dot{e}_{cit} + \delta_{cc}\dot{\tilde{e}}_{cit}.$$

It follows that

$$\sum_{t=1}^{T} 1[a(i,t) = c]\dot{l}_{it} = \lambda_{cr}\sum_{t=1}^{T} 1[a(i,t) = c]\dot{e}_{cit} + \delta_{cc}\sum_{t=1}^{T} 1[a(i,t) = c]\dot{\tilde{e}}_{cit}.$$

Dividing both sums $\sum_{t=1}^{T} 1[a(i,t) = c]\dot{\tilde{e}}_{cit} = \frac{m(m+1)}{2}(1-m/T)$ and $\sum_{t=1}^{T} 1[a(i,t) = c]\dot{e}_{cit} = \frac{m(m+1)}{2}(1-m/T) - \frac{m^2(T-m)}{T}$ with the factor $T(i,c)\left(1 - \frac{T(i,c)}{T}\right) = m(1-m/T)$ it follows that the total bias is

$$bias(\hat{\sigma}_c) = (\lambda_{cr} + \delta_{cc})\frac{m+1}{2} - \lambda_{cr}m, \tag{31}$$

which is negative if and only if $\theta > 1/2 + 1/2m$ where $\theta := \lambda_{cr}/(\lambda_{cr} + \delta_{cc})$ is the share of value of experience that is portable. If the value of experience is highly portable it is possible that workers while working in the city have a lower value of experience than their average value of experience. The portability of the value of experience implies that when workers change area of employment they do not experience a large drop in the value of their experience. On the other hand if experience is not portable then workers who initially work in the city experience a fast growth in wages and may reaching a peak level. They then migrate to the rural area and experience a large drop in their wages never again reaching the peak level. This implies that these workers, while in the city, have a value of experience above their average value of experience and hence they contribute positively to the bias of the city fixed effect.

## 6.2 Technical note

For the case where the estimation is done with individual fixed effects the derivations of the bias are very similar. We start by defining then the symmetric, idempotent, rank $T-1$ demeaning matrix $Q = I_T - \iota_T(\iota_T^\top \iota_T)^{-1}\iota_t^\top$ and premultiplying the stacked version of the model with $Q$ to get

$$Q\mathbf{w}_i = Q\mathbf{Z}_i\theta + Q\mathbf{u}_i = Q\mathbf{D}_i\sigma + Q\mathbf{X}_i\theta + Q\mathbf{u}_i, \tag{32}$$

using dot notation $\dot{\mathbf{A}} = Q\mathbf{A}$ the model can then be written as

$$\dot{\mathbf{w}}_i = \dot{\mathbf{Z}}_i\theta + \dot{\mathbf{u}}_i = \dot{\mathbf{D}}_i\sigma + \dot{\mathbf{X}}_i\theta + \dot{\mathbf{u}}_i. \tag{33}$$

Using derivations similar to the case without individual fixed effects it can be shown that

$$plim\ \hat{\sigma}_N = \sigma + plim\ \left(\frac{1}{N}\sum_i \mathbf{D}_i^\top \dot{\mathbf{D}}_i\right)^{-1}\left(\frac{1}{N}\sum_i \mathbf{D}_i^\top \dot{\mathbf{u}}_i\right), \tag{34}$$

under the assumption that $\mathbb{E}[\dot{\mathbf{D}}_i^\top \mathbf{X}_i] = \mathbb{E}[\mathbf{D}_i^\top \dot{\mathbf{X}}_i] = \mathbf{0}$. The matrix $\sum_i \mathbf{D}_i^\top \dot{\mathbf{D}}_i$ is not invertable if all areas are included, however this is easily fixed by dropping a row of $\mathbf{D}_i$ using one area as a reference level.

Before simplifying the expression for the bias further we note that in the more general case where the assumption that $\mathbb{E}[\dot{\mathbf{D}}_i^\top \mathbf{X}_i] = \mathbf{0}$ is not imposed the expression for the probability limit includes further the term

$$-\left(\frac{1}{N}\sum_i \mathbf{D}_i^\top \dot{\mathbf{D}}_i\right)^{-1}\left(\frac{1}{N}\sum_i \mathbf{D}_i^\top \dot{\mathbf{X}}_i\right)(plim\ \hat{\beta}_N - \beta). \tag{35}$$

Simplification of the bias under the assumption $\mathbb{E}[\mathbf{D}_i^\top \dot{\mathbf{X}}_i] = \mathbf{0}$ is still more challenging than for the case without individual fixed effects because the matrix $\mathbf{D}_i^\top \dot{\mathbf{D}}_i$ is not diagonal. However in the two area case, the matrix reduces to a scalar because a row corresponding to one area is removed. In this case the probability limit of the area fixed effect for the area not used as reference level is given as

$$plim\ \hat{\sigma}_{N,a} = \sigma_a + plim\ \left( \frac{1}{N} \sum_i \mathbf{d}_i^\top \dot{\mathbf{d}}_i \right)^{-1} \left( \frac{1}{N} \sum_i \mathbf{d}_i^\top \dot{\mathbf{u}}_i \right), \tag{36}$$

where $\mathbf{d}_i = (1[a(i,1) = a], ..., 1[a(i,T) = a])^\top$. The scalar $\mathbf{d}_i^\top \dot{\mathbf{d}}_i$ can also be written as

$$\mathbf{d}_i^\top \dot{\mathbf{d}}_i = \sum_{t=1}^{T} 1[a(i,t) = a] \left( 1[a(i,t) = a] - (1/T) \sum_{t=1}^{T} 1[a(i,t) = a]) \right) \tag{37}$$

$$= T(i,a) \left( 1 - \frac{T(i,a)}{T} \right), \tag{38}$$

from which it follows that if we consider a worker who works in the city for the first $m$ periods and the changes employment location to the rural area for the last $T - m$ periods then

$$\mathbf{d}_i^\top \dot{\mathbf{d}}_i = m \left( 1 - \frac{m}{T} \right), \tag{39}$$

when the rural area is used as reference. Analogously if the worker is employed for the first $m$ periods in the rural area and then changes job location to the city for the last $T - m$ periods then

$$\mathbf{d}_i^\top \dot{\mathbf{d}}_i = (T - m) \left( 1 - \frac{(T - m)}{T} \right). \tag{40}$$

still assuming the rural area is used as reference.

# Transportation and quality of life

## Evidence from Denmark

Jesper Hybel* Ismir Mulalic†

May 14, 2020

**Abstract**

This paper investigates the importance of transportation for the quality of life. We first estimate a quality of life index that measures a representative household's willingness-to-pay for local amenities. We find that the quality of life is high in large cities. Wages and rents are also substantially higher in the urban areas that are dense. Our empirical results suggest that the quality of public transport system is in particular important for the quality of life.

**Keywords:** quality of life, rent gradients, wage gradients, commuting costs, amenities, transportation.

**JEL codes:** H4, J3, O52, R1, R4.

---

*Technical University of Denmark, DTU Management, Bygningstorvet, Building 116, DK-2800 Kgs. Lyngby, Denmark, email: jeshyb@dtu.dk. Jesper Hybel is supported by the Innovation Fund Denmark via the Urban Project.

†Copenhagen Business School, Department of Economics, Porcelnshaven 16A, DK-2000 Frederiskberg, Denmark, email: imu.eco@cbs.dk. Ismir Mulalic is a Visiting Professor at VU Amsterdam.

# 1 Introduction

This paper studies the importance of transportation for the quality of life (QOL). Countries around the world devote significant share of public funds to transport infrastructure investments and maintenance. For the highest income countries the investment in transport infrastructure has stabilized around 1% of the GDP and is expected to raise over the coming decades (OECD and ITF, 2013).[1] Moreover, households devote about 20% of their expenditures to transportation (see e.g. Berri et al., 2014 and Couture et al., 2018). The average commuter spent about 1 hour per day on commuting in 2016 (OECD, Statista 2019). It is therefore important to recognize the importance of transportation for the quality of life.

Roback (1982) and Rosen (1979) pioneered estimation of the quality of life (QOL) index for urban areas by adjusting the city wages for local cost-of-living.[2] They show that, in cities, higher nominal wage levels may compensate for both higher housing costs and disamenities.[3] This implies that (homogenous) households accept lower real wages or suffer higher housing costs in order to live in a place with preferable amenities measured by the the QOL, viz. a measure of neighbourhood quality. Beeson and Eberts (1989) and Gabriel and Rosenthal (1996) compare local wages to rents to measure QOL. This methodology implicitly includes the value of all – observed and unobserved – local urban amenities. Albouy (2008) estimates more plausible QOL index by adjusting the quality of life indices for taxes, non-housing costs, and non-labor income, and shows that these measures are positively correlated with popular "liveability" rankings. While hedonic methods are usually applied to estimate value of specific amenities (e.g. traffic noise (Theebe, 2004), air quality (Chay and Greenstone, 2005), crime (Pope, 2008; Gautier et al., 2009) and proximity to water (Rouwendal et al., 2017)), the one-dimensional QOL index offers an economically intuitive measure of "liveabil-

[1] According to OECD and ITF (2013), advanced economies will need to improve the quality of ageing transport networks and emerging economies will need more transport infrastructure to support economic growth.

[2] See Albouy and Lue (2015) for an exhaustive review of literature on the estimation of the QOL.

[3] Hoch (1972) has also argued that the city wage premium is a result of higher urban productivity and low urban quality of life (e.g. noise, pollution, congestion and crime).

ity" that provides the value households place on all local amenities.[4] Moreover, Albouy (2008) argues convincingly that QOL indices from the empirical hedonic literature in practice offer counter-intuitive results, e.g. by producing odd rankings of cities and city rankings that negatively correlates with city size (Burnell and Galster, 1992).

Although the property value hedonics is the workhorse model for valuation of urban amenities, these methods are often biased by the housing sorting.[5] Structural approaches account for the household residential sorting and relate household sorting to local urban amenities including the provision of local public goods.[6] This may be important when studying the importance of transportation for the QOL, because the provision of public transport and transport infrastructure has some of the characteristics of a local public good and is likely associated with Tiebout sorting (see e.g. Epple and Sieg (1999)). For example, the density of railroad stations and bus stops is related to the population density and usually shows substantial differences over space. The structural models are however computationally-intensive and do not offer a clear measure of the QOL but instead provide the value (heterogeneous) households place on considered local urban amenities. This paper focuses on the one-dimensional QOL indices.

Transportation infrastructure impacts the spatial organisation of economic activity between urban areas and household sorting (Redding and Turner, 2015). It also facilitates interaction within cities. It enables workers to combine living in high quality residential areas with working at the most productive places (Redding and Rossi-Hansberg, 2017). Ahlfeldt et al. (2015) demonstrate that because of the presence of clustering benefits, better transportation possibilities, that reduce the

[4]Rosen (1974) showed that the first derivative of the hedonic price function with respect to the individual attribute equals the marginal willingness to pay (wtp) for this attribute. Economists have relied on Rosens hedonic model of market equilibrium to measure the wtp for specific amenities. See Palmquist (2006) for a review of empirical hedonic literature.

[5]See Kuminoff et al. (2010) for a more detailed discussion of omitted variable and simultaneity problems in the empirical hedonic literature. Bajari and Benkard (2005) generalize Rosen's model in some aspects (allows for oligopoly, discrete characteristics, and unobserved characteristics) but is more restrictive in other aspects, e.g. it assumes a specific functional form of the utility function.

[6]See Kuminoff et al. (2013) for an overview of the literature on residential sorting models. The methodology employed in the residential sorting models was developed by Berry (1994) and Berry et al. (1995). Bayer et al. (2007) pioneered the application of this approach to housing market analysis. Bayer and Timmins (2005, 2007) discuss the equilibrium properties of residential sorting models.

burden of commuting, result in more specialisation. Heblich et al. (2018) confirm this by showing how emergence of rail mass transport in 19th century London implied a substantial increase in the specialization of the inner city in production. Glaeser et al. (2001) provide empirical evidence on the growing importance of consumer amenities, which are often clustered in central cities. Mulalic and Rouwendal (2020) show that the extension of the metro network in Copenhagen results in a substantial increase in the interest among the highly educated for living in areas close to metro stations, which affects the demographic composition of neighbourhoods.[7] Moreover, Baum-Snow (2007) shows that the construction of highways has contributed to the suburbanization of households. Transport infrastructure is therefore related to the attractiveness of urban areas and consequently also to the QOL.

Transportation is derived demand as individuals often consume the service not because they benefit from consumption directly, but because they partake in other consumption or activities elsewhere (see e.g. Small and Verhoef (2007)). Transportation allows households to buy consumption goods and activities, get to work and enjoy leisure.[8] Households therefore face in general trade-off between, on one hand productivity and consumption advantages (high-paying jobs and high quality local urban amenities), and on the other hand higher costs of living and disamenities (high housing costs, congestions and pollution), when they decide where to live. It is therefore important also to recognize the importance of commuting costs for the QOL.

This paper follows Albouy and Lue (2015) and estimates a transport adjusted QOL index for the 98 urban areas - municipalities - covering Denmark. We first compare housing and commuting costs to local wages to estimate a representative (mobile) households willingness-to-pay (wtp) for local amenities, viz. the QOL

[7]Urban economic theory predicts that workers with high wages have different commuting patterns than those with low wages. In a standard monocentric city model workers with higher wages will have longer commuting distances (Alonso, 1964; Muth, 1969; Brueckner, 1987). Assuming that the workers commuting costs include time costs that positively depend on wages, the relationship between wage on commuting distance is ambiguous (Henderson, 1977; Fujita, 1989, pp.31.). The relationship between income and commuting distance depends also on the spatial distribution of residential amenities (Brueckner et al., 1999). Gutirrez-i Puigarnau et al. (2016) show that for Denmark, conditional on the workplace location, the income elasticity of distance is negative.

[8]Travel may also have direct consumption value (Couture et al., 2018). This value is however negligible, so we ignore it in this study.

index. We consider household *taste heterogeneity* as well as *commuting costs*. More precisely, we estimate local wages by place of work to reduce potential biases from unobserved skills, correct for local taxes, and add commuting costs to housing expenditures. The average Dane spends about 55 minutes on transport per day (DTU, 2013) and the average household's expenditure devoted to transport is about 20% of the total household budget (Berri et al., 2014).[9] We therefore also analyse the importance of transportation for the QOL in Denmark. We regress the estimated QOL indices on the observed amenities to infer how much quality of life is associated to urban amenities, and in particular to transportation.

We find that local wages and rents vary considerably between municipalities, and are substantially higher in the urban areas that are dense. We find also that worker heterogeneity is important when estimating wage differentials, i.e. correction for worker heterogeneity reduces the percentage wage gap between areas with the lowest and the highest wages by about 50%. The estimated QOL index ranks the Greater Copenhagen Area and other large cities in Denmark highest. This is plausible because these high density urban areas are considered as highly attractive. Moreover, we find strong positive relationship between the QOL indices and the population growth. The QOL indices are also positively associated with the local urban amenities related to transportation demonstrating, in particular, the importance of public transport for the quality of life. Our empirical findings suggest that 10 extra departures by public transport per hectare per day are associated with DKK 15 (about 2 Euros).[10]

The remainder of the paper is organized as follows. Section 2 describes the theoretical model that guides our empirical methodology. Section 3 presents the data, provides descriptive statistics and discusses our empirical strategy. Empirical findings are presented and discussed in Section 4, emphasising the associations between transportation and the QOL. Section 5 concludes.

[9]About 90% of the average household's expenditure devoted to transport are allocated to private transport including purchases of cars and car usage costs (Berri et al., 2014).

[10]1 DKK $\approx$ 0.13 Euro.

# 2 Theoretical framework

This section describes the theoretical framework that we use. We first introduce the basic model in subsection 2.1. In subsection 2.2 we show how to operationalize this model.

## 2.1 The model

We follow Albouy and Lue (2015) and extend the Rosen (1979) model by including commuting costs. Households are assumed to be homogeneous, perfectly mobile and fully informed about the municipality characteristics. This implies that households have full information on housing prices, wages, commuting costs and amenities. We further simplify by assuming zero moving costs. This implies a spatial equilibrium in which utility levels are equalized across municipalities.

Households consume housing $y$ at municipality specific price $p_j$, a traded good $x$ with the price normalized to one, as well as leasure time $l$ and commuting time $f$. Each municipality provides access to a vector of amenities $\mathbf{Z}$ aggregated into a single index $Q = Q(\mathbf{Z})$.[11] The preferences of households are represented by the quasi-concave utility function $U(x, y, l, h, f, Q)$ that is increasing in $x, y, l, Q$ and decreasing in commuting time $f$ and work hours $h$.

Households choose combination $(j, k)$ of a municipality of residence $j$ and a municipality where they work $k$. Residence locations $(j)$ differ in local prices $p_j$ and local amenities $Q_j$, while workplace locations $k$ differ in local wages $w_k$ and monetary commuting costs $cf_{jk}$, where $c \geqslant 0$ is the monetary cost per unit of time spent on commuting. They also choose consumption levels of $x$, $y$ and labour supply $h$, and pay local taxes $\tau$. The resulting household budget constraint is then $x + p_j y \leq w_k h - \tau(w_k h) - cf_{jk}$. Households are also constrained with respect to the time available which is standardized to 1 and used on commuting $f$, working $h$ and leisure $l$, so $h + l + f_{jk} \leq 1$. Assuming the spatial equilibrium, the net

[11]Amenities in municipalities that are physically close to municipality $j$ may have a direct impact on the utility of households with residence or/and job in that municipality $(j)$, e.g. restaurants, parks or recreational facilities. van Duijn and Rouwendal (2015) develop a model in which this is explicitly taken into account. In our model this is captured by municipality specific indices.

expenditure for a household with the utility $\overline{u}$ can be expressed as:

$$E(p_j, w_k, f_{jk}; Q_j, \overline{u}) := \min_{x,y,h,l} \{x + p_j y - w_k h + \tau(w_k h) + c f_{jk} \tag{1}$$

$$: l + f_{jk} + h \leq 1, U(x, y, l, f_{jk}; Q_j) \geq \overline{u}\},$$

where $\overline{u}$ is the equilibrium level of utility. This expenditure function is increasing in the local prices $p_j$ and the time of commute $f_{jk}$ and decreasing in local wages $w_k$ and local amenities $Q_j$, i.e. assuming that eq.(1) is differentiable, $\frac{\partial E}{\partial p} \geq 0$, $\frac{\partial E}{\partial f} \geq 0$, $\frac{\partial E}{\partial w} \leq 0$ and $\frac{\partial E}{\partial Q} \leq 0$. Moreover, in equilibrium households chose combinations $(j, k)$ providing the same level of utility $\overline{u}$, so all households are equally satisfied. For households with homogeneous preferences, free mobility and perfect information, the expenditure incurred at equilibrium utility $\bar{u}$ must be the same for all locations $j$. Formally this can be written:

$$E(p_j, w_k, f_{jk}; Q_j, \overline{u}) = 0 \tag{2}$$

In order to learn about differences in local prices and local wages we implicitly differentiate eq.(2) with respect to $j$ and $k$ (by varying the municipality of residence or municipality of work):

$$\frac{\partial E}{\partial p} dp_j + \frac{\partial E}{\partial f} df_j + \frac{\partial E}{\partial Q} dQ_j = 0 \tag{3}$$

$$\frac{\partial E}{\partial w} dw_k + \frac{\partial E}{\partial f} df_k = 0. \tag{4}$$

Eq.(3) represents the housing price gradient and shows that households are compensated for higher housing prices by lower commutes or higher level of amenities. Eq.(4), the wage gradient, shows that wages increase with commutes, or in other words, that workers are compensated for longer commutes with higher wages.[12]

Finally, we combine eq.(3) and eq.(4) and derive a household's willingness to

[12]This is a standard result in monopsony models, see e.g. Manning (2003a,b). There is also evidence that in Denmark, the country of our study, employees who face longer commutes receive a small wage increase. Mulalic et al. (2014) shows that for Denmark, 1 km increase in commuting distance induces a wage increase of about 0.15 % that corresponds to approximately half of the commuting costs.

pay ($wtp$) for change in the QOL ($dQ_j$):

$$-\frac{\partial E}{\partial Q}dQ_j = \frac{\partial E}{\partial p}dp_j + \frac{\partial E}{\partial f}df_{jk} + \frac{\partial E}{\partial w}dw_k \tag{5}$$

where $df_{jk} := df_k + df_j$ is the total difference in commuting time. Applying the envelope theorem and evaluating the derivatives at the national average we can rewrite eq.(5) to:

$$-\frac{\partial E}{\partial Q}dQ_j = \bar{y}\bar{p}_j + [c + (1-\tau')\bar{w} - \alpha]df_{jk} - (1-\tau')\bar{h}dw_k, \tag{6}$$

where $\alpha := (\partial U/\partial f)/(\partial U/\partial x)$ is the the "leisure-value" of commuting. Note here that $\frac{\partial E}{\partial Q}dQ_j$ is the marginal willingness-to-pay (wtp) for QOL ($Q_j$). Moreover, this expression relates urban benefits (amenities and employment opportunities) $\frac{\partial E}{\partial Q}dQ_j + (1-\tau')\bar{h}dw_k$ and urban costs $\bar{y}\bar{p}_j + [c + (1-\tau')\bar{w} - \alpha]df_{jk}$. For example, households pay higher urban costs to get access to higher level of urban amenities and better employment opportunities, or receive higher wages as compensation for high housing price or low level of amenities. It also allows quantification of unobserved QOL ($Q_j$) as a weighted sum of local costs of living $p_j$, local wages $w_k$ and commuting costs $f_{jk}$.

## 2.2 Model operationalization

In order to operationalize the model and construct the QOL index, we first express differentials in terms of log-differentials ($\hat{z} := (z-\bar{z})/\bar{z}$, $z = p, w, f$) and divide eq.(6) with the national average of income $\bar{m}$:

$$-\frac{\partial E}{\partial Q}\frac{dQ_j}{\bar{m}} = s_y\hat{p}_j + \left[s_c + s_w\frac{\bar{f}}{\bar{h}}\right]\hat{f}_{jk} - s_w\hat{w}_k, \tag{7}$$

where $s_y := \bar{y}\bar{p}/\bar{m}$ is the income share for housing, $s_c := c\bar{f}/\bar{m}$ is share of income spent on commuting, and $s_w := (1-\tau')\bar{h}\bar{w}/\bar{m}$ is income share from labour. This model ignores household heterogeneity, so the shares apply only to a representative household. We furthermore assume that the marginal commuting time is valued as work time such that $\alpha = 0$. Finally we multiply with the share of residents in municipality $j$ working in municipality $k$ ($\pi(k|j)$) and sum over workplaces in order to get:

$$\hat{Q}_j = s_y \hat{p}_j + \left[ s_c + s_w \frac{\bar{f}}{\bar{h}} \right] \hat{f}_j - s_w \hat{w}_j, \tag{8}$$

where $\hat{f}_j := \sum_k \hat{f}_{jk} \pi(k|j)$ and $\hat{Q}_j := -\frac{\partial E}{\partial Q} \frac{dQ_j}{\bar{m}}$. The left hand side is the marginal willingness-to-pay for local amenities as a fraction of household income.

Finally, the shares $s_y$, $s_w$ and $s_c$ are based on the official statistics from Statistics Denmark.[13] The income share from labour (household disposable income as a fraction of total expenditures) is 83 %. This implies that about one fifth of income comes from other sources than labour. The expenditure on housing as a share of total income is 32 % and the share of income spent on commuting is 14.0 %. Moreover, the ratio of time spent commuting (one hour in average) to time spent working (about 8 hours a day) is approximately 12.4 % of a working day.

# 3 Data

The data we use to estimate local housing prices and local wages are derived from administrative registers for all Danish households for the year 2010. We observe about 2 M households. The households in our sample are distributed over 98 municipalities in which they choose to live and to work. The average area of a municipality is 432.59 $km^2$ and the average population density is 130 people per $km^2$. The geographical size of municipalities decreases with population density. The municipalities are therefore smaller in the Greater Copenhagen Area (GCA).[14] We discuss the data and the estimation of local housing prices and local wages in the following two subsections. In the last subsection we show how we estimate commuting costs by combining Danish register data on commuting flows with the data on travel times, mode choice and trip frequencies from the Danish National

[13] See `https://www.statbank.dk`. As income $\bar{m}$ we use the total consumption as defined in Table `FU09`, which for the year 2010 is approximately Euro 41.000 (DKK 305.000). We also calculate the disposable income $(1-\tau)\bar{h}w$ as a share of consumption. To calculate $s_y$ and $s_c$ we use Table `FU02`. The average number of work hours is set at $\bar{h} = 7.4$ which is the official number of work hours for a full time employee. Based on the assumption that each worker travels to and from work every work day the average number of hours spent on transport is $\bar{f} = 0.91$ calculated as $\bar{f} = \sum_{jk} \pi_{jk}(f_{jk} + f_{kj})$.

[14] The GCA is part of the Danish island Zealand. Copenhagen (the capital city of Denmark) is its centre. It is the political, administrative and educational core region of Denmark.

Transportation model.[15]

## 3.1 Housing prices

The housing price index $\hat{p}_j$ is constructed using a dataset of all the real estate transactions for the year 2010. The data set includes transaction prices and the structural attributes from the Building and Dwelling Register (BBR), such as age of building, size (sqm) and number of rooms. We restrict our sample to so-called arm's length sales where the buyer is a private individual. The final sample includes 13,087 realized real estate transactions. Table 1 shows the descriptive statistics. The mean realised price is DKK 1.8 M. The average house is 57 years old (was constructed in 1953), has four rooms and 123 sqm. About one third of the traded units were single-family houses. More importantly, there is a high degree of variation in almost every quality attribute. This is very useful for the identification of the housing price indices.

Table 1: Descriptive statistics for the real estate transactions for year 2010

| | mean | std. dev. | min | max |
|---|---|---|---|---|
| Price (1000 DKK) | 1,822.01 | 1,014.23 | 190.0 | 5,900.00 |
| Space (sqm) | 123.55 | 43.31 | 1.00 | 680.00 |
| Age | 56.95 | 37.05 | 0.00 | 409.00 |
| Number of rooms | 4.26 | 1.44 | 1.00 | 16.00 |
| Number of toilets | 1.52 | 0.58 | 1.00 | 6.00 |
| Single-family house (share) | 0.23 | 0.42 | 0.00 | 1.00 |

*Notes*: Number of observations is 13,087. 1 DKK $\approx$ 0.13 EUR.

Standardized house price has been compiled from a hedonic model with municipality fixed effect. The log of the sales price is regressed on housing characteristics $\mathbf{X}_m$ and a municipality indicator $\mu_{j(m)}$ with $j(m)$ being the municipality where the house $m$ is located. The regression equation is given as

$$\log p_m = \mathbf{X}_m^\top \beta + \mu_{j(m)} + \epsilon_m,$$

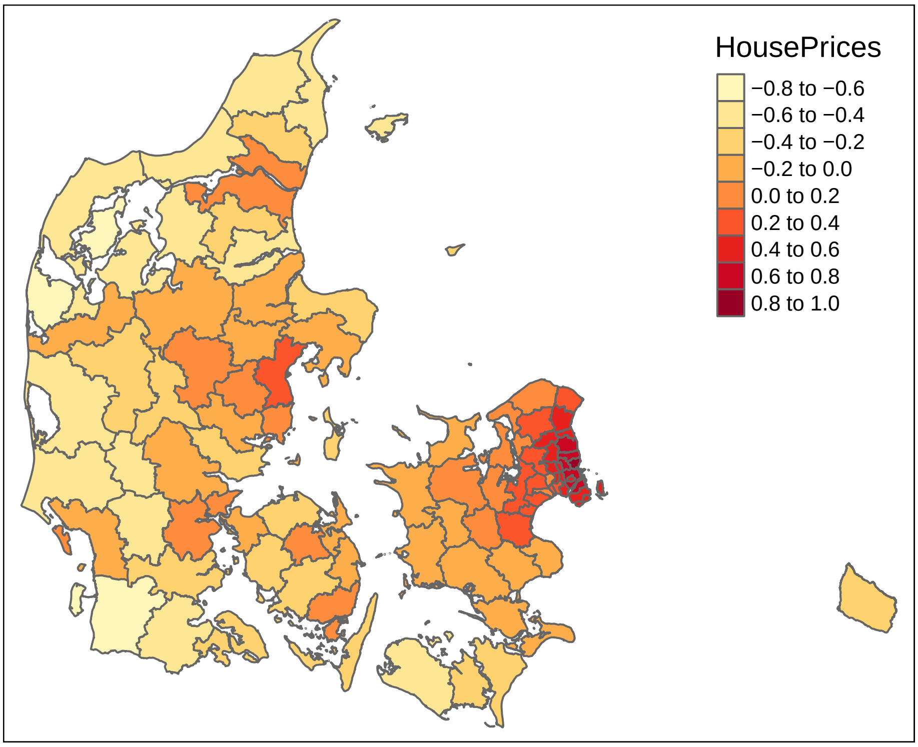
and the estimates $\hat{\mu}_j$ are used as the housing price index. Figure 1 shows the

[15]Recall here that in addition to the micro data sets and the transport model data, we also use the aggregate data from Statistics Denmark that allows us to calculate $s_y$, $s_c$ and $s_w$.

resulting housing price index across municipalities and Table A1 in the Appendix A reports the estimated coefficients.

Not surprisingly we find that the housing prices are higher in the GCA and in the north of this area that is considered as highly attractive, and in other larger cities in Denmark, e.g. in Aarhus (the second largest city in Denmark). Low price houses are spread throughout most of western and southern Denmark.

Figure 1: Housing price index $p_j$




## 3.2 Local wages

We use a micro data set for the full population of workers to construct the wage index $\hat{w}_k$. The dataset is derived from annual register data from Statistics Denmark for the year 2010 and includes information on workers residence and workplace (both at the municipal level), hourly wages, and a range of explanatory variables for each worker: educational level, age, gender, full-time versus part-time, and the sector of employment. We select workers who had been employed for at least one year. Our sample then includes 1,209,928 observations (workers). Table 2 reports the descriptive statistics for workers.

Table 2: Descriptive statistics for workers

| | mean | std.dev. | min | max |
|---|---|---|---|---|
| Hourly wage (DKK/hour) | 215.99 | 91.35 | 85.58 | 1345.27 |
| Age | 43.57 | 10.58 | 16.00 | 93.00 |
| Male (share) | 0.55 | 0.50 | 0.00 | 1.00 |
| Primary education (share) | 0.15 | 0.35 | 0.00 | 1.00 |
| Upper secondary education (share) | 0.04 | 0.19 | 0.00 | 1.00 |
| Vocational education and training (share) | 0.40 | 0.49 | 0.00 | 1.00 |
| Qualifying educational programmes (share) | 0.02 | 0.15 | 0.00 | 1.00 |
| Short cycle higher education (share) | 0.06 | 0.24 | 0.00 | 1.00 |
| Vocational bachelors educations (share) | 0.20 | 0.40 | 0.00 | 1.00 |
| Bachelors programmes (share) | 0.01 | 0.12 | 0.00 | 1.00 |
| Masters programmes (share) | 0.11 | 0.31 | 0.00 | 1.00 |
| PhD programmes (share) | 0.01 | 0.09 | 0.00 | 1.00 |

*Notes*: Number of observations is 1,209,928. 1 DKK $\approx$ 0.13 EUR.

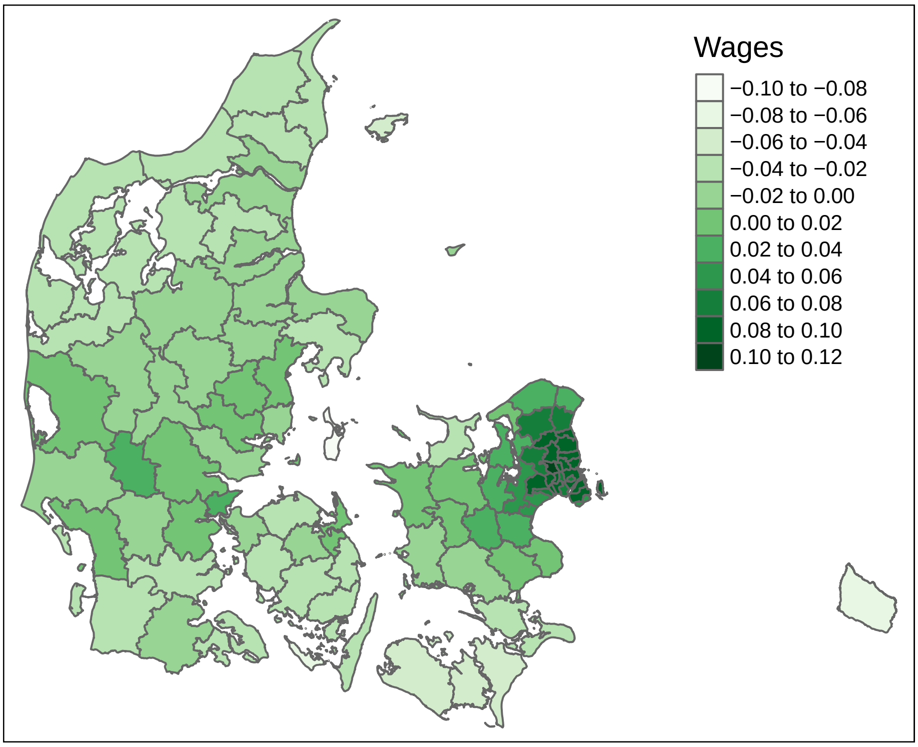
We regress the log of wages on the work place indicators $\mu_k$ as well as controls for the observed worker attributes $\mathbf{X}_i$:

$$\log w_i = \mathbf{X}_i^\top \beta + \mu_{k(i)} + \epsilon_i \tag{9}$$

where $k(i)$ is the place of work municipality of individual $i$. More importantly, we first estimate $\hat{\mu}_k$ for the place of work, and then we use the estimated $\hat{\mu}_k$ to calculate the wage differentials $\hat{w}_j = \sum_k \hat{\mu}_k \pi(k|j)$ for workers with residence in municipality $j$, where $\pi(k|j)$ is the share of residents in municipality $j$ working in municipality $k$. In other words, we average $\hat{\mu}_k$ according to the proportion of workers $\hat{\pi}_{jk}$ living in municipality $j$ and working in municipality $k$. Note here that many workers work in a different municipality from their residence municipality.

We find that the wage differentials $\hat{w}_j$ are substantially higher in the GCA and other large cities in Denmark (Aarhus, Odense and Aalborg) as illustrated in Figure 2. We also find that heterogeneity of workers is important when estimating wage differentials $\hat{w}_j$. For example, before correction for worker heterogeneity, the percentage wage gap between the municipality with the lowest and the municipality with the highest wages is about 50 %. This gap reduces significantly when correcting for the observed heterogeneity. Table A2 in Appendix A reports the estimation results of the Mincerian wage regression. Moreover, local wages and housing prices are positively correlated suggesting that households in Denmark are at least partly compensated for the higher housing costs by higher urban wages.

Figure 2: Wage index by place of work $w_k$




## 3.3 Commuting costs

The commuting time index $\hat{f}_{jk}$ is based on a data on travel times, mode choice and trip frequencies between 907 traffic zones from the Danish National Transportation model for the year 2010 designed for detailed traffic modelling (Rich et al., 2010). These travel times are derived using the complete road network structure including all minor roads and one-way restrictions and include congestion delays and transition times for public transport. The computations of the travel times within the traffic zones include also trips not crossing the zone borders, so the diagonal elements of the travel time O-D matrix are different from zeros (positive).

We combine the data on travel times with the register data on commuting flows between municipalities to compute the commuting time index. To see how, let the set $M$ be the set of municipalities covering Denmark. The workers choose combination of a municipality of residence and a municipality where they work $(j, k) \in M \times M$. From the Danish National Transport Model we have data on the travel times $f(z_g, z_h, l)$ and the number of trips $n(z_g, z_h, l)$ from zone $z_g$ to zone $z_h$ using the transport mode $v$, which can be either public transport or car. To aggregate the travel time data from the level of transport zones to the level of municipalities we first estimate the expected travel time using the number of trips as weights. Specifically we define the travel time from municipality $j$ to municipality $k$ as:

$$f_{jk} := \widehat{\mathbb{E}}[f(z_g, z_j, v)|z_g \in j, z_h \in k] :=$$
$$\sum_{z_g \in j} \sum_{z_h \in k} \sum_{v} f(z_g, z_h, v) \frac{n(z_g, z_h, v)}{\sum_{z_g \in j} \sum_{z_g \in k} \sum_{l} n(z_g, z_h, v)},$$

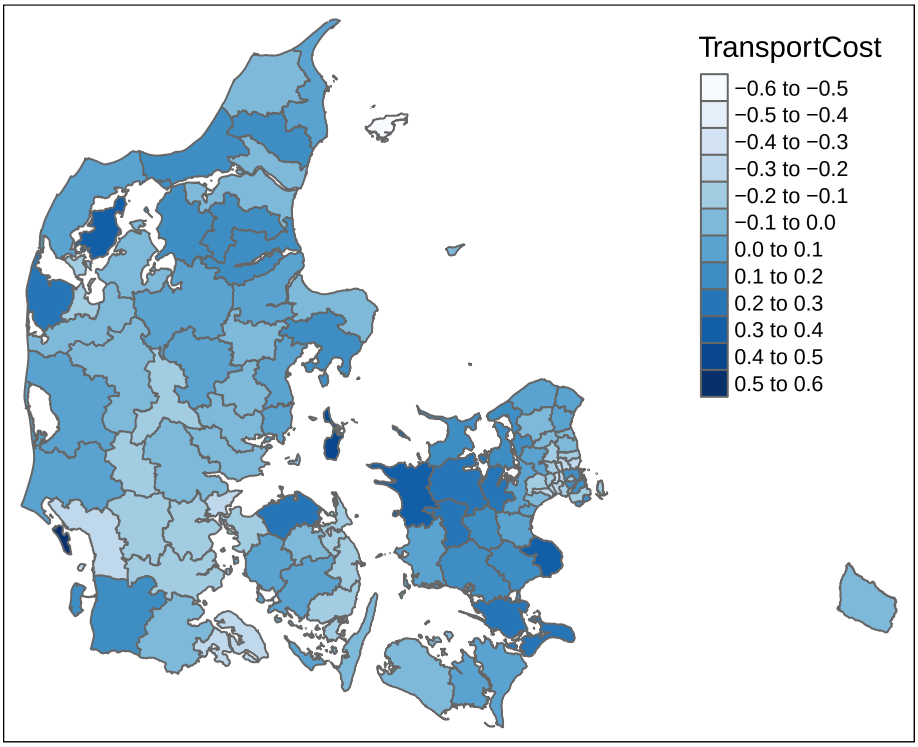
and then compute the commuting time differentials $\hat{f}_j = (f_j - \bar{f})/\bar{f}$ for a specific municipality as:

$$f_j := \sum_{k \in M} \pi_{jk}(f_{jk} + f_{kj}), \qquad \bar{f} := \sum_{jk} \pi_{jk}(f_{jk} + f_{kj}) \tag{10}$$

where the commuting times differentials are averaged in proportion to the number of workers living in municipality $j$ and working in municipality $k$. We assume that each worker travels to and from work every day.

Figure 3 shows the results. The commuting time index is lower in large cities and in particular in the GCA. This is consistent with standard urban models that predict increasing commuting costs with the distance from the CBD (Alonso, 1964; Muth, 1969). Moreover, in the GCA commuting times are higher in the CBD and lover in the suburban areas. The similar patterns were also observed for other larger cities by e.g. Bayer et al. (2007).

Figure 3: Commuting time index $\hat{f}_j$



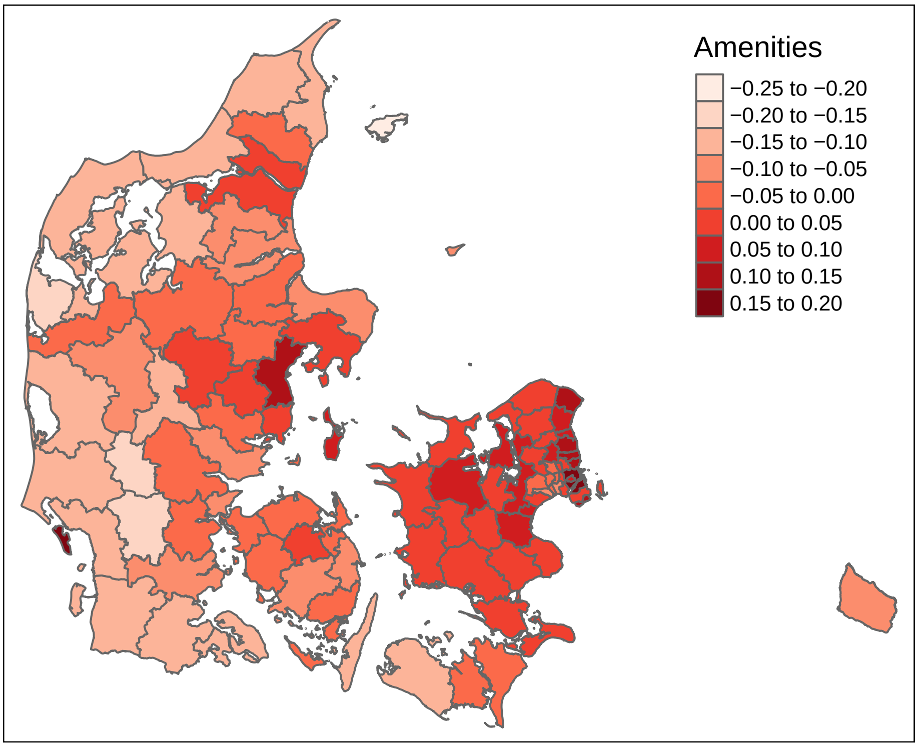

# 4 Empirical Results

In this section, we turn to the empirical results. In the first subsection, we presents information on the QOL index ($\hat{Q}_j$). The following subsection discusses the specific role of transportation for the QOL in Denmark.

## 4.1 Quality of life index

We combine information on the estimated housing prices ($\hat{p}_j$), local wages ($\hat{w}_j$), and commuting differentials ($\hat{f}_j$) to estimate the average local willingness-to-pay for amenities (quality of life (QOL) index) from eq.(8). Our estimation results suggest as expected that the marginal willingness-to-pay for local amenities ($\hat{Q}_j$) is higher in the GCA and other larger cities in Denmark (Aarhus, Odense and Aalborg), see Figure 4.

Figure 4: QOL index ($\hat{Q}_j$)




**Notes:** This QOL index represents the marginal willingness-to-pay for local amenities $\hat{Q}_j$.

Table 3 reports the top five and the bottom five municipalities based on the QOL.[16] The highest QOL index in Denmark is in the Copenhagen Municipality, that is the core of the GCA and well-known for its high-quality restaurants, large number of museums and other cultural amenities, high quality public transport, shopping opportunities, and "the best place to work". This municipality is also characterised with high housing costs and high wages. This is also the case for the third (Dragr), the fourth (Rudersdal) and the fifth municipality (Lyngby-Taarbk) on the ranking list, all located in the GCA. One notable exception is the second ranked municipality. Fan is a smaller (56 sqkm) island in the south-west Denmark (in the North Sea) with a population of about 3000. It is connected with the main

[16] Table A3 in Appendix A reports ranking of all municipalities in Denmark based on the QOL.

land with a ferry service.[17] For this municipality wages are low and commuting costs are high, but the households are compensated with relatively low housing costs and high level of amenities, most likely beautiful nature and clean air.[18] The lowest quality of life is found in municipalities further away from the big cities located on Jutland (Hjrring, Lemvig, Vejen and Billund) and island Ls. In these municipalities both wages and housing costs are low. We find in general a strong positive correlation between $\hat{w}_j$ and $\hat{p}_j$ (correlation coefficient is 0.76).

Table 3: Top- and bottom-five municipalities in Denmark based on the QOL

| Rank | Municipality | $\hat{Q}_j$ | $\hat{w}_j$ | $\hat{p}_j$ | $\hat{f}_j$ |
|---|---|---|---|---|---|
| | | Top five | | | |
| 1 | Kbenhavn | 0.16 | 0.08 | 0.64 | 0.11 |
| 2 | Fan | 0.16 | -0.03 | 0.01 | 0.54 |
| 3 | Dragr | 0.15 | 0.06 | 0.49 | 0.16 |
| 4 | Rudersdal | 0.14 | 0.10 | 0.75 | -0.07 |
| 5 | Lyngby-Taarbk | 0.13 | 0.09 | 0.74 | -0.12 |
| | | Bottom five | | | |
| 94 | Hjrring | -0.14 | -0.03 | -0.50 | -0.05 |
| 95 | Lemvig | -0.16 | -0.02 | -0.78 | 0.29 |
| 96 | Vejen | -0.16 | -0.02 | -0.48 | -0.11 |
| 97 | Billund | -0.19 | 0.03 | -0.42 | -0.14 |
| 98 | Ls | -0.23 | -0.06 | -0.46 | -0.53 |

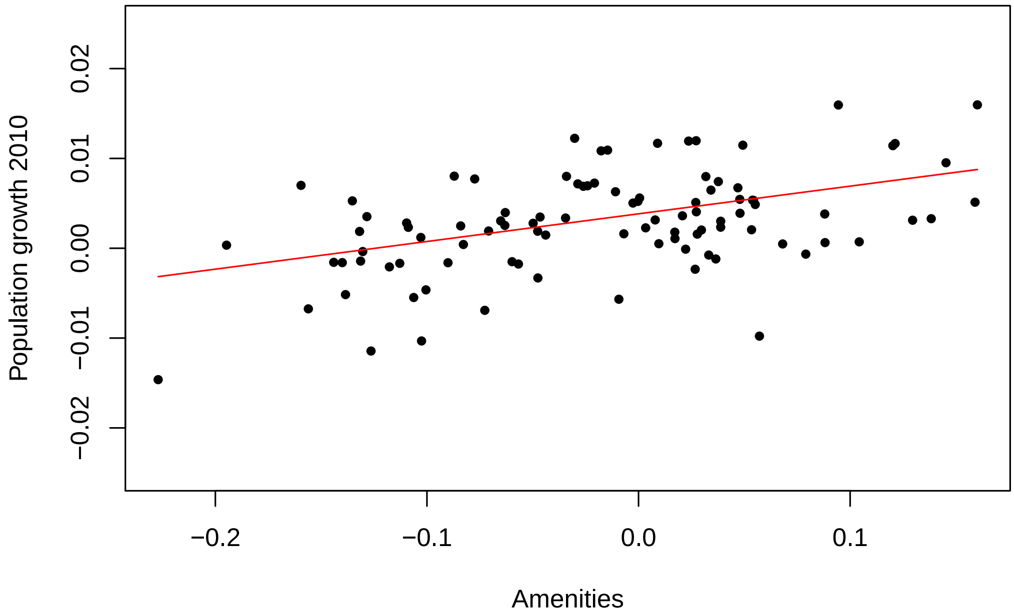
We also find that population has grown faster in high-amenity areas, see Figure 5. It suggests that households in Denmark are attracted by high amenity areas, i.e. cities. This was also observed for the U.S. by Glaeser et al. (2001) who show empirically that high amenity cities have grown faster than low amenity cities. They also find that in the U.S. urban rents have raised faster than urban wages, suggesting that the demand for living in cities has risen also because of increasing demand for urban amenities. Although we do not estimate changes in urban rents and urban wages, it is well-known that also in Denmark urban rents have raised faster than urban wages and that the evolving urbanisation process is likely caused, not only by the raising incomes, but also by an increase in the demand for urban

[17] The ferry ride takes 12 minutes.

[18] The whole island's western shore is a long beach. About 30,000 tourists visit this island each summer.

amenities (Gutirrez-i Puigarnau et al., 2016).[19]

Figure 5: Population growth and QOL in Denmark




## 4.2 Transportation and the QOL in Denmark

The QOL index captures per definition the net value of all amenities within a municipality. Some of these amenities are positively evaluated by at least some household types, such as parks, monuments and public transport, and some are not appreciated, such as pollution and congestion. There exists also amenities that are not observable by researchers, such as e.g. nice neighbourhood atmosphere. Many amenities are related to transport. We use a multivariate regression of the estimated QOL index ($\hat{Q}_j$) on a vector of observed municipality-level amenities

[19]Note here that rearranging eq.(8) gives $\hat{p}_j = \frac{s_w}{s_y}\hat{w}_j - \frac{\left(s_w(\bar{f}/\bar{h})+s_c\right)}{s_y}\hat{f}_j + \hat{Q}_j = A_j + \hat{Q}_j$, where $A_j$ denotes the compensation for housing costs in terms of wages corrected for commuting costs. Figure A1 in Appendix A shows the relationship between housing price index and local wages corrected for commuting costs.

to explore the relationship between QOL and transport related amenities. The considered amenity variables are summarized in Table A4 in Appendix A.

We are in particular interested in the impact of transport on the QOL. We use two variables – *number of departures with public transport per sq km* and *distance to the nearest highway ramp* – as proxies for different forms of transport infrastructure. Share of workers commuting to or from an municipality are also related to transportation. We also use four additional amenity variables to proxy for other relevant QOL aspects. The considered amenities variables are endogenous at different levels to the local population and are likely related to households residential sorting. Therefore the derived and discussed monetary values of these amenities are only illustrative and should be taken with caution.[20]

Table 4 shows the estimation results. The important element of the QOL index in Denmark is the demographic composition of neighbourhoods. The regression shows significant and large coefficients of (1) *share of population with higher education*, and (2) *share of pupils in private schools.* It is often argued in the urban economics literature that the attractiveness of living in a particular area is partly determined by the demographic composition of that neighbourhood. The importance of this factor for location choice within the San Francisco Bay area was documented by Bayer and Timmins (2007) and for the GCA by Mulalic and Rouwendal (2020). The strongly significant coefficient related to private schools comes as a surprise. All households in Denmark have a universal access to primary schools and only minor share of pupils attend the private schools.[21] However, this positive correlation is likely related to school quality. Private schools allow for more time for each student, which results in better schooling. Moreover, the supply of private schools in cities is also related to higher educated parents, who are conscious of their children receiving high-quality schooling, and with the provision of public goods. This is also confirmed by the positive significant correlation between the service level (the municipality service expenses) and the QOL index despite the

[20] Notice here that the regression residuals result mostly from unobserved amenities and measurement error, but likely also from mis-specification. Consequently, the estimated regression models are not fulfilling requirements for an orthogonal error term.

[21] In Denmark, every child is guaranteed a place in the tuition-free public schools in proximity to its residence. About 80 % of all pupils in primary and lower secondary schools attended the tuition-free public schools, 15 % attended the private schools, and 5 % attended the other (special) schools. Some parents choose private schools because they are smaller, or because they have a particular educational approach, e.g. for religious reasons.

Table 4: Urban amenities and the QLI

| | Dependent variable: $\hat{Q}$ | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| No. of publ. transp. departures per sq km | 0.049*** | | 0.164** |
| | (0.016) | | (0.082) |
| Log distance to the nearest highway ramp | -0.016*** | | -0.003 |
| | (0.006) | | (0.004) |
| Service level (municipality service expenses) | | 0.293* | 0.264* |
| | | (0.150) | (0.149) |
| Share of population with higher education | | 0.006*** | 0.006*** |
| | | (0.001) | (0.001) |
| Share of pupils in private schools | | 0.002*** | 0.002*** |
| | | (0.001) | (0.001) |
| Population density | | -0.002 | -0.062** |
| | | (0.005) | (0.031) |
| Share of workers commuting from munic. | | 0.003*** | 0.003*** |
| | | (0.001) | (0.001) |
| Share of workers commuting to munic. | | -0.002*** | -0.002*** |
| | | (0.001) | (0.001) |
| Constant | -0.001 | -0.530*** | -0.493*** |
| | (0.013) | (0.146) | (0.146) |
| $R^2$ | 0.189 | 0.671 | 0.686 |
| Adjusted $R^2$ | 0.172 | 0.649 | 0.658 |
| Number of obs. | 98 | 98 | 98 |

**Notes:** High education includes bachelor, long-cycle higher education and PhD-degree; *p<0.1; **p<0.05; ***p<0.01; standard errors are in parentheses.

fact that local taxes are controlled for. The service level is usually higher in cities, where the concentration of economic activity is higher.

More importantly for this study, the second set of amenities that we show in the regressions illustrates the role of transport infrastructure. The main transport amenities explain about 17 % of the variation in $\hat{Q}_j$ (see model (1)) and all amenities together about 66 % (see model (3)). We find a negative relationship between *distance to the nearest highway ramp* and $\hat{Q}_j$. Our empirical results suggest that 1 % reduction in the distance to the nearest highway ramp is related to 0.2 % increase in the marginal willingness-to-pay for local quality of life $\hat{Q}_j$. This corresponds to about EUR 1,000 per kilometre. However, this relationship diminish significantly when we include the full set of amenities. *The number of departures with public transport per sq km* is strongly associated with $\hat{Q}_j$. Additional 100 departures per sqkm per day, or 10 departures per hectare per day, is associated with about EUR 2. This strongly suggests that the provision and quality of public transport is an important amenity. Car ownership and use are relatively expensive in Denmark, car ownership is low (0.81 cars per household in Denmark), and the share of multiple car households is low (8.2 % of households in Denmark), even though the share of households with two workers is high.[22] Many workers therefore have to use public transport and presumably, accessibility to this facility is important. Moreover, we find a significant negative relationship between population density and QOL. One interpretation of this fact is that, conditional on other amenities, population density variables is a proxy for disamenities such as congestion, noise and pollution. Finally, our empirical results suggest also that the specialized employment areas with many jobs are associated with the lower QOL, i.e. the coefficient associated with *the share of workers commuting to municipality* is negative. On the other hand, municipalities with a larger share of workers commuting from municipality are more attractive. This suggest that households in Denmark prefer to separate workplace locations from residence locations in order to enjoy urban amenities in their neighbourhoods and benefit from production benefits from concentration (agglomeration). The presence of these facts also

[22]In Denmark, the purchase-tax of a car is 105 % for the value of the car below about EUR 10.500 and 180 % of the value of the car above. In addition there is an annual ownership tax depending on the characteristics of the car. Mulalic and Rouwendal (2015) show that the mean annual total expenditure associated with ownership and use of a new car in Denmark is about EUR 11,000.

suggest that better transportation possibilities, that reduce the burden of commuting, results in more specialisation also identified for London (Heblich et al., 2018). So, on the one hand, production benefits from agglomeration and demand for urban amenities at the residence location force workers to accept commutes. Better transport possibilities on the other hand ease commuting.

In summary, the empirical analyses have shown that the transport infrastructure, and in particular public transport, are important for the QOL in Denmark.

# 5 Conclusion

This paper estimates the Quality of Life (QOL) index that measures the value a representative household places on the local amenities. The estimated QOL index produces plausible ranking of the 98 municipalities covering Denmark. It is high for the capital and other larger cities, while it is low in rural municipalities. We also find strong positive relationship between the QOL index and the population growth suggesting that the urbanisation process is likely related to the increasing demand for urban amenities.

The importance of transportation for the quality of life is confirmed by our empirical results. We find that proximity to the nearest highway ramp and the quality of public transport are positively related to the QOL indices. Our empirical findings also suggest that households prefer to separate the workplace locations from the residence locations that induces commuting.

Over the past years, policymakers and urban planners have expressed concerns regarding the urbanization. Households tend however to move to the areas which best satisfies theirs preferences for urban amenities (e.g. public goods and nature), or in other words, to the areas that offer higher QOL. Our empirical findings suggest that changes in the transport infrastructure have important implications for the attractiveness of the residential and work locations, and finally for the QOL.

One of the main objectives of the regional policy in many countries is to give the local authorities (e.g. municipalities) the same financial basis through the equalisation schemes. For example, the purpose of the equalisation scheme in Denmark is to even out the differences in the economic situation in the municipalities due to differences in the tax base and the demographic composition. This equalisation scheme is based on the so-called net equalisation method, i.e. municipality's estimated structural surplus or deficit per inhabitant. For the individual municipalities the net payments or the net receipts can be substantial. However, the calculation of the structural surplus per inhabitant ignores the differences in the housing costs and the value of amenities. Our findings can be useful for improving the equalisation schemes.

# References


Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., and Wolf, N. (2015). The economics of density: Evidence from the berlin wall. *Econometrica*, 83(6):2127–2189.

Albouy, D. (2008). Are big cities bad places to live? estimating quality of life across metropolitan areas. Working Paper 14472, National Bureau of Economic Research.

Albouy, D. and Lue, B. (2015). Driving to opportunity: Local rents, wages, commuting, and sub-metropolitan quality of life. *Journal of Urban Economics*, 89:74–92.

Alonso, W. (1964). *Location and Land Use. Toward a General Theory of Land Rent*. Harward University Press.

Bajari, P. and Benkard, C. L. (2005). Demand estimation with heterogeneous consumers and unobserved product characteristics: A hedonic approach. *Journal of Political Economy*, 113(6):1239–1276.

Baum-Snow, N. (2007). Did Highways Cause Suburbanization?*. *The Quarterly Journal of Economics*, 122(2):775–805.

Bayer, P., Ferreira, F., and McMillan, R. (2007). A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, 115(4):588–638.

Bayer, P. and Timmins, C. (2005). On the equilibrium properties of locational sorting models. *Journal of Urban Economics*, 57(3):462–477.

Bayer, P. and Timmins, C. (2007). Estimating equilibrium models of sorting across locations. *Economic Journal*, 117(518):353–374.

Beeson, P. E. and Eberts, R. W. (1989). Identifying productivity and amenity effects in interurban wage differentials. *The Review of Economics and Statistics*, 71(3):443–452.

Berri, A., Lyk-Jensen, S. V., Mulalic, I., and Zachariadis, T. (2014). Household transport consumption inequalities and redistributive effects oftaxes: A repeated cross-sectional evaluation for france, denmark and cyprus. *Transport Policy*, 36:206–216.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics*, 25(2):242–262.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Brueckner, J. (1987). The structure of urban equilibria: A unified treatment of the muth-mills model. In Mills, E. S., editor, *Handbook of Regional and Urban Economics*, volume 2, chapter 20, pages 821–845. Elsevier, 1 edition.

Brueckner, J., Thisse, J., and Zenou, Y. (1999). Why is central paris rich and downtown detroit poor?: An amenity-based theory. *European Economic Review*, 43(1):91–107.

Burnell, J. D. and Galster, G. (1992). Quality-of-life measurements and urban size: An empirical note. *Urban Studies*, 29(5):727–735.

Chay, K. and Greenstone, M. (2005). Does air quality matter? evidence from the housing market. *Journal of Political Economy*, 113(2):376–424.

Couture, V., Duranton, G., and Turner, M. A. (2018). Speed. *The Review of Economics and Statistics*, 100:725–739.

DTU (2013). Tu-rapport om danskernes transportvaner.

Epple, D. and Sieg, H. (1999). Estimating equilibrium models of local jurisdictions. *Journal of Political Economy*, 107(4):645–681.

Fujita, M. (1989). *Urban Economic Theory*. Cambridge University Press.

Gabriel, S. A. and Rosenthal, S. S. (1996). Commutes, neighborhood effects, and earnings: An analysis of racial discrimination and compensating differentials. *Journal of Urban Economics*, 40(1):61–83.

Gautier, P. A., Siegmann, A., and Van Vuuren, A. (2009). Terrorism and attitudes towards minorities: The effect of the theo van gogh murder on house prices in amsterdam. *Journal of Urban Economics*, 65(2):113 – 126.

Glaeser, E. L., Kolko, J., and Saiz, A. (2001). Consumer city. *Journal of Economic Geography*, 1(1):27–50.

Gutirrez-i Puigarnau, E., Mulalic, I., and van Ommeren, J. (2016). Do rich households live farther away from their workplaces? *Journal of Economic Geography*, 16(1):177–201.

Heblich, S., Redding, S. J., and Sturm, D. M. (2018). The Making of the Modern Metropolis: Evidence from London. CEP Discussion Papers dp1573, Centre for Economic Performance, LSE.

Henderson, J. (1977). *Economic Theory and the Cities*. Academic Pres.

Hoch, I. (1972). Income and city size. *Urban Studies*, 9(3):299–328.

Kuminoff, N. V., Parmeter, C. F., and Pope, J. C. (2010). Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities? *Journal of Environmental Economics and Management*, 60(3):145–160.

Kuminoff, N. V., Smith, V. K., and Timmins, C. (2013). The new economics of equilibrium sorting and policy evaluation using housing markets. *Journal of Economic Literature*, 51(4):1007–1062.

Manning, A. (2003a). *Monopsony in motion: Imperfect competition in labor markets*. Princeton University Press.

Manning, A. (2003b). The real thin theory: monopsony in modern labour markets. *Labour economics*, 10(2):105–131.

Mulalic, I., Ommeren, J. N. V., and Pilegaard, N. (2014). Wages and Commuting: Quasinatural Experiments' Evidence from Firms that Relocate. *Economic Journal*, 124(579):1086–1105.

Mulalic, I. and Rouwendal, J. (2015). The impact of fixed and variable cost on automobile demand: Evidence from denmark. *Economics of Transportation*, 4(4):227 – 240.

Mulalic, I. and Rouwendal, J. (2020). Does improving public transport decrease car ownership? evidence from a residential sorting model for the copenhagen metropolitan area. *Regional Science and Urban Economics*, 83:103543.

Muth, R. F. (1969). *Cities and Housing*. University of Chicago Press.

OECD and ITF (2013). ITF transport outlook 2013: Funding transport. Technical report, OECD.

Palmquist, R. B. (2006). Property value models. In Mler, K. G. and Vincent, J. R., editors, *Handbook of Environmental Economics*, volume 2, chapter 16, pages 763–819. Elsevier, 1 edition.

Pope, J. C. (2008). Fear of crime and housing prices: Household reactions to sex offender registries. *Journal of Urban Economics*, 64(3):601 – 614.

Redding, S. J. and Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, 9(1):21–58.

Redding, S. J. and Turner, M. A. (2015). Chapter 20 - transportation costs and the spatial organization of economic activity. In Duranton, G., Henderson, J. V., and Strange, W. C., editors, *Handbook of Regional and Urban Economics*, volume 5 of *Handbook of Regional and Urban Economics*, pages 1339–1398. Elsevier.

Rich, J., Nielsen, O., Brems, C., and Hansen, C. (2010). Overall design of the danish national transport model. Technical report, Annual Transport Conference at Aalborg University.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy*, 90(6):1257–1278.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy*, 82(1):34–55.

Rosen, S. (1979). Wages-based indexes of urban quality of life. In Mieszkowski, P. and Straszheim, M., editors, *Current Issues in Urban Economics*, chapter 3, pages 74–103. John Hopkins University, Baltimore.

Rouwendal, J., Levkovich, O., and van Marwijk, R. (2017). Estimating the value of proximity to water, when ceteris really is paribus. *Real Estate Economics*, 45(4):829–860.

Small, K. A. and Verhoef, E. T. (2007). *The Economics of Urban Transportation*. Routledge.

Theebe, M. A. J. (2004). Planes, trains, and automobiles: The impact of traffic noise on house prices. *The Journal of Real Estate Finance and Economics*, 28(2):209–234.

van Duijn, M. and Rouwendal, J. (2015). Sorting based on Urban Heritage and Income: Evidence from the Amsterdam Metropolitan Area. Tinbergen Institute Discussion Papers 15-030/VIII, Tinbergen Institute.

# A Appendix

Table A1: Hedonic price equation with municipality fixed effect, OLS

| | log(price) |
|---|---|
| Space (sqm) | 0.007*** |
| | (0.001) |
| Space squared | -0.00001*** |
| | (>0.00001) |
| Age, years | -0.065*** |
| | (0.005) |
| Age squared, years | 0.00002*** |
| | (>0.00001) |
| Number of rooms | 0.004 |
| | (0.004) |
| Dummy indicating 2 toilets | 0.125*** |
| | (0.008) |
| Dummy indicating 3 toilets | 0.196*** |
| | (0.020) |
| Dummy indicating 4 toilets | 0.039 |
| | (0.107) |
| Dummy indicating 5 toilets | 0.218* |
| | (0.129) |
| Dummy indicating 6 toilets | 0.147 |
| | (0.258) |
| Dummy indicating single-family house | 0.267*** |
| | (0.016) |
| Municipality fixed effect | *yes* |
| Constant | 73.956*** |
| | (4.628) |
| Adjusted $R^2$ | 0.589 |
| Observations | 13,087 |

*Notes*: $^*$p<0.1; $^{**}$p<0.05; $^{***}$p<0.01.

Table A2: Mincerian wage regression with municipality fixed effect, OLS

| | log wage |
|---|---|
| Age | $0.043^{***}$ |
| | (0.0002) |
| Age squared | $-0.0004^{***}$ |
| | (>0.00001) |
| Dummy indicating male | $0.175^{***}$ |
| | (0.0005) |
| Dummy indicating primary education | $-0.144^{***}$ |
| | (0.001) |
| Dummy indicating upper secondary education | $0.020^{***}$ |
| | (0.002) |
| Dummy indicating vocational education and training | $-0.054^{***}$ |
| | (0.001) |
| Dummy indicating short cycle higher education | $0.045^{***}$ |
| | (0.002) |
| Dummy indicating vocational bachelors educations | $0.104^{***}$ |
| | (0.001) |
| Dummy indicating bachelors programmes | $0.126^{***}$ |
| | (0.002) |
| Dummy indicating masters programmes | $0.301^{***}$ |
| | (0.001) |
| Dummy indicating PhD programmes | $0.333^{***}$ |
| | (0.003) |
| Work place municipality fixed effect | *yes* |
| Constant | $4.173^{***}$ |
| | (0.005) |
| Adjusted $R^2$ | 0.329 |
| Observations | 1,209,928 |

*Notes*: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table A3: QOL ($\hat{Q}_j$), housing prices ($\hat{p}_j$), local wages ($\hat{w}_j$), and commuting differentials ($\hat{f}_j$) across municipalities in Denmark

| Rank | Municipality | $\hat{Q}_j$ | $\hat{w}_j$ | $\hat{p}_j$ | $\hat{f}_j$ |
|---|---|---|---|---|---|
| 1 | Kbenhavn | 0.16 | 0.08 | 0.64 | 0.11 |
| 2 | Fan | 0.16 | -0.03 | 0.01 | 0.54 |
| 3 | Dragr | 0.15 | 0.06 | 0.49 | 0.16 |
| 4 | Rudersdal | 0.14 | 0.10 | 0.75 | -0.07 |
| 5 | Lyngby-Taarbk | 0.13 | 0.09 | 0.74 | -0.12 |
| 6 | Aarhus | 0.12 | 0.01 | 0.38 | 0.05 |
| 7 | Gentofte | 0.12 | 0.10 | 0.82 | -0.23 |
| 8 | Helsingr | 0.10 | 0.03 | 0.40 | 0.03 |
| 9 | Frederiksberg | 0.09 | 0.08 | 0.59 | -0.12 |
| 10 | Hrsholm | 0.09 | 0.09 | 0.58 | -0.09 |
| 11 | Roskilde | 0.09 | 0.04 | 0.36 | 0.05 |
| 12 | Fredensborg | 0.08 | 0.07 | 0.41 | 0.02 |
| 13 | Solrd | 0.07 | 0.05 | 0.33 | 0.02 |
| 14 | Sams | 0.06 | -0.09 | -0.39 | 0.42 |
| 15 | Kge | 0.06 | 0.03 | 0.20 | 0.08 |
| 16 | Holbk | 0.05 | 0.01 | 0.02 | 0.22 |
| 17 | Fures | 0.05 | 0.09 | 0.49 | -0.13 |
| 18 | Frederikssund | 0.05 | 0.03 | 0.17 | 0.12 |
| 19 | Gladsaxe | 0.05 | 0.09 | 0.62 | -0.27 |
| 20 | Odder | 0.05 | -0.02 | 0.09 | 0.02 |
| 21 | Lejre | 0.05 | 0.03 | 0.06 | 0.22 |
| 22 | Allerd | 0.05 | 0.09 | 0.41 | -0.03 |
| 23 | Halsns | 0.04 | 0.02 | 0.05 | 0.15 |
| 24 | Gribskov | 0.04 | 0.02 | 0.14 | 0.05 |
| 25 | Silkeborg | 0.04 | -0.01 | 0.06 | 0.06 |
| 26 | Vordingborg | 0.04 | -0.03 | -0.13 | 0.23 |
| 27 | Sor | 0.03 | 0.01 | -0.04 | 0.20 |
| 28 | Kalundborg | 0.03 | 0.01 | -0.13 | 0.35 |
| 29 | Hillerd | 0.03 | 0.06 | 0.29 | -0.02 |
| 30 | Hvidovre | 0.03 | 0.08 | 0.43 | -0.16 |
| 31 | Stevns | 0.03 | 0.01 | -0.14 | 0.31 |
| 32 | Nstved | 0.03 | -0.00 | -0.05 | 0.17 |

| Rank | Municipality | $\hat{Q}_j$ | $\hat{w}_j$ | $\hat{p}_j$ | $\hat{f}_j$ |
|---|---|---|---|---|---|
| 33 | Egedal | 0.03 | 0.07 | 0.21 | 0.08 |
| 34 | Trnby | 0.03 | 0.09 | 0.42 | -0.14 |
| 35 | Herlev | 0.03 | 0.09 | 0.47 | -0.20 |
| 36 | Skanderborg | 0.02 | 0.00 | 0.10 | -0.03 |
| 37 | Rdovre | 0.02 | 0.07 | 0.46 | -0.25 |
| 38 | Odense | 0.02 | -0.00 | 0.13 | -0.10 |
| 39 | Greve | 0.02 | 0.06 | 0.27 | -0.07 |
| 40 | Faxe | 0.02 | 0.01 | -0.08 | 0.20 |
| 41 | Odsherred | 0.01 | -0.02 | -0.18 | 0.18 |
| 42 | Ringsted | 0.01 | 0.02 | 0.00 | 0.11 |
| 43 | Syddjurs | 0.01 | -0.02 | -0.13 | 0.13 |
| 44 | Slagelse | 0.00 | -0.00 | -0.07 | 0.09 |
| 45 | Aalborg | 0.00 | -0.01 | 0.02 | -0.05 |
| 46 | Ballerup | -0.00 | 0.10 | 0.41 | -0.18 |
| 47 | Nordfyns | -0.00 | -0.03 | -0.31 | 0.29 |
| 48 | Svendborg | -0.01 | -0.02 | 0.04 | -0.16 |
| 49 | r | -0.01 | -0.07 | -0.28 | 0.08 |
| 50 | Randers | -0.01 | -0.01 | -0.09 | 0.03 |
| 51 | Favrskov | -0.01 | -0.00 | -0.05 | -0.01 |
| 52 | Glostrup | -0.02 | 0.09 | 0.46 | -0.36 |
| 53 | Vejle | -0.02 | 0.01 | -0.05 | -0.01 |
| 54 | Hje-Taastrup | -0.02 | 0.08 | 0.22 | -0.11 |
| 55 | Viborg | -0.03 | -0.01 | -0.16 | 0.06 |
| 56 | Kolding | -0.03 | 0.01 | 0.01 | -0.11 |
| 57 | Horsens | -0.03 | 0.00 | -0.03 | -0.08 |
| 58 | Middelfart | -0.03 | -0.01 | -0.00 | -0.15 |
| 59 | Brnderslev | -0.03 | -0.03 | -0.31 | 0.16 |
| 60 | Vallensbk | -0.04 | 0.08 | 0.27 | -0.22 |
| 61 | Albertslund | -0.04 | 0.08 | 0.19 | -0.15 |
| 62 | Kerteminde | -0.05 | 0.01 | -0.03 | -0.12 |
| 63 | Guldborgsund | -0.05 | -0.04 | -0.30 | 0.04 |
| 64 | Assens | -0.05 | -0.02 | -0.25 | 0.05 |
| 65 | Holstebro | -0.05 | -0.02 | -0.16 | -0.06 |
| 66 | Ishj | -0.06 | 0.07 | 0.16 | -0.21 |
| 67 | Brndby | -0.06 | 0.09 | 0.22 | -0.23 |

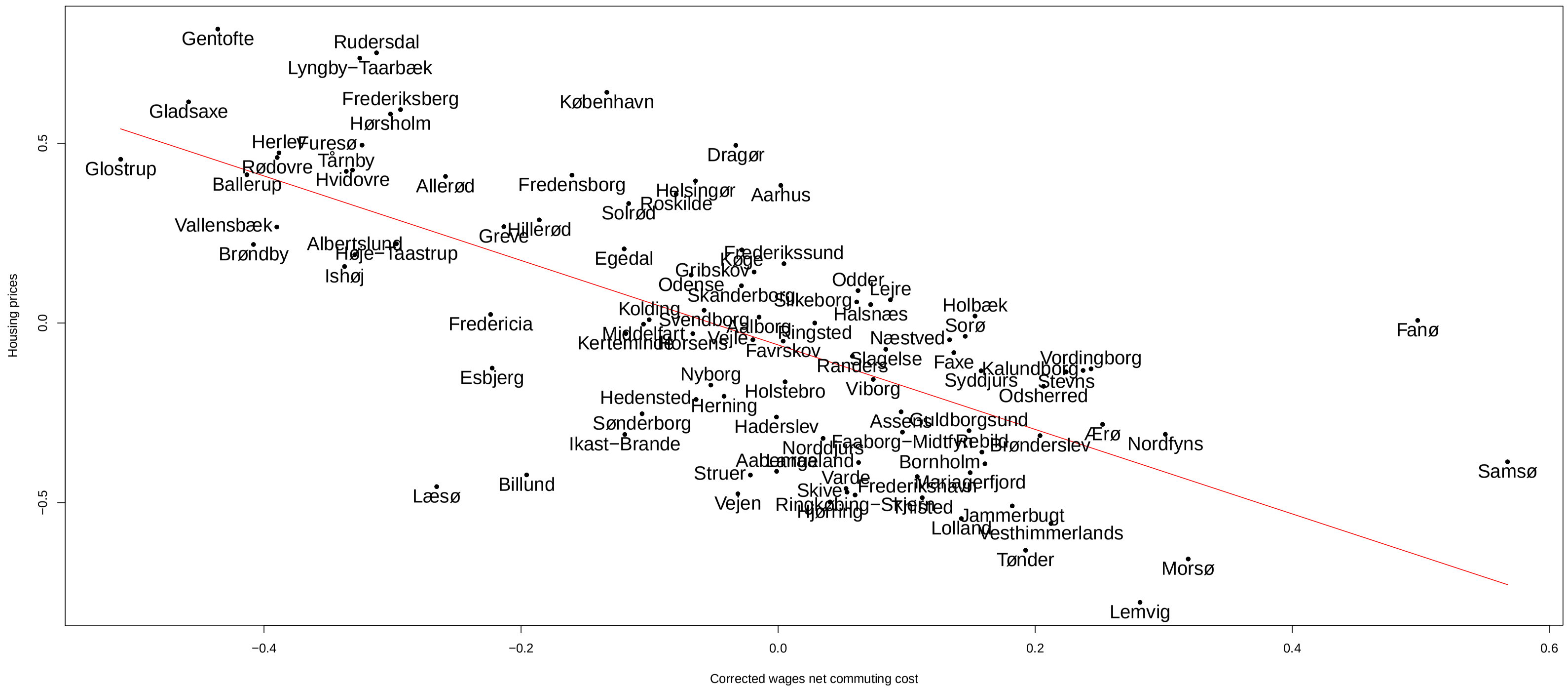
| Rank | Municipality | $\hat{Q}_j$ | $\hat{w}_j$ | $\hat{p}_j$ | $\hat{f}_j$ |
|---|---|---|---|---|---|
| 68 | Fredericia | -0.06 | 0.03 | 0.02 | -0.20 |
| 69 | Rebild | -0.06 | -0.02 | -0.36 | 0.13 |
| 70 | Faaborg-Midtfyn | -0.07 | -0.03 | -0.30 | 0.02 |
| 71 | Nyborg | -0.07 | -0.02 | -0.17 | -0.14 |
| 72 | Bornholm | -0.07 | -0.07 | -0.39 | -0.04 |
| 73 | Herning | -0.08 | -0.01 | -0.20 | -0.09 |
| 74 | Haderslev | -0.08 | -0.03 | -0.26 | -0.10 |
| 75 | Mariagerfjord | -0.08 | -0.02 | -0.42 | 0.13 |
| 76 | Hedensted | -0.09 | -0.00 | -0.21 | -0.10 |
| 77 | Norddjurs | -0.09 | -0.02 | -0.32 | -0.02 |
| 78 | Frederikshavn | -0.10 | -0.02 | -0.43 | 0.06 |
| 79 | Langeland | -0.10 | -0.03 | -0.39 | -0.01 |
| 80 | Jammerbugt | -0.10 | -0.03 | -0.51 | 0.13 |
| 81 | Mors | -0.11 | -0.03 | -0.66 | 0.30 |
| 82 | Vesthimmerlands | -0.11 | -0.03 | -0.56 | 0.18 |
| 83 | Esbjerg | -0.11 | 0.02 | -0.13 | -0.23 |
| 84 | Snderborg | -0.11 | -0.03 | -0.25 | -0.25 |
| 85 | Thisted | -0.12 | -0.03 | -0.49 | 0.06 |
| 86 | Lolland | -0.13 | -0.06 | -0.54 | -0.01 |
| 87 | Varde | -0.13 | -0.01 | -0.46 | 0.03 |
| 88 | Aabenraa | -0.13 | -0.01 | -0.41 | -0.03 |
| 89 | Skive | -0.13 | -0.03 | -0.47 | -0.03 |
| 90 | Ringkbing-Skjern | -0.13 | 0.00 | -0.48 | 0.09 |
| 91 | Ikast-Brande | -0.14 | -0.00 | -0.31 | -0.17 |
| 92 | Tnder | -0.14 | -0.03 | -0.63 | 0.14 |
| 93 | Struer | -0.14 | -0.03 | -0.42 | -0.14 |
| 94 | Hjrring | -0.14 | -0.03 | -0.50 | -0.05 |
| 95 | Lemvig | -0.16 | -0.02 | -0.78 | 0.29 |
| 96 | Vejen | -0.16 | -0.02 | -0.48 | -0.11 |
| 97 | Billund | -0.19 | 0.03 | -0.42 | -0.14 |
| 98 | Ls | -0.23 | -0.06 | -0.46 | -0.53 |

Table A4: Descriptive statistic for amenity variables

| | mean | std.dev | min | max |
|---|---|---|---|---|
| No. of publ. transp. departures per sqkm | 178.27 | 498.65 | 0.09 | 3952 |
| Log distance to the nearest highway ramp (km) | 1.63 | 1.41 | -1.83 | 4.95 |
| Service level (municipality service expenses index) | 1.00 | 0.04 | 0.93 | 1.14 |
| Share of population with higher education (%) | 23.38 | 8.44 | 13.70 | 51.20 |
| Share of pupils in private schools (%) | 13.84 | 6.12 | 0.00 | 29.40 |
| Population density (people per sqkm) | 557.52 | 1,362.04 | 17.00 | 11,028.00 |
| Share of workers commuting from munic. (%) | 44.58 | 20.05 | 5.20 | 83.50 |
| Share of workers commuting to munic. (%) | 39.11 | 20.56 | 3.30 | 87.10 |

**Notes:** Number of observations is 98. High education includes bachelor, long-cycle higher education and PhD-degree.

Figure A1: Housing price index and local wages corrected for commuting costs



# B Appendix

Consider the expenditure minimization problem defined as

$$\begin{aligned}\min_{x,y,h,l,f} \quad & x + p_j y + cf + \tau(w_k h + I) - w_k h - I \\ s.t. \quad & U(x,y,h,l,f,Q_j) - u \geq 0 \\ & T - h - l - f \geq 0 \\ & f - f_{jk} \geq 0,\end{aligned} \tag{11}$$

unlike in the text we allow the agent to also minimize over transport time $f$ however since we assume $\partial U/\partial f < 0$ in optimum $f = f_{jk}$ so the minimization problem could be reformulated simply inserting $f_{jk}$ for $f$ as in the text. Another deviation from the text is that we include the endowment of non-labour income $I$ which is also seen to be insubstantial, hence we have ignored it in the text.

The associated Lagrangian function for the minimization problem is given as

$$\begin{aligned}\mathcal{L}(x,y,h,l,f,\gamma,\mu,\delta) = {} & x + p_j y + cf + \tau(w_k h + I) - w_k h - I \\ & - \gamma(U(x,y,h,l,f,Q_j) - u) \\ & - \mu(T - h - l - f) \\ & - \delta(f - f_{jk}),\end{aligned} \tag{12}$$

from which by KKT-conditions it follows by envelope theorem that

$$\frac{\partial \mathcal{L}}{\partial x} = 1 - \gamma \frac{\partial U}{\partial x} = 0 \tag{13}$$

$$\frac{\partial \mathcal{L}}{\partial y} = p_j - \gamma \frac{\partial U}{\partial y} = 0 \tag{14}$$

$$\frac{\partial \mathcal{L}}{\partial h} = -(1-\tau')w_k - \gamma \frac{\partial U}{\partial h} + \mu = 0 \tag{15}$$

$$\frac{\partial \mathcal{L}}{\partial l} = -\gamma \frac{\partial U}{\partial l} + \mu = 0 \tag{16}$$

$$\frac{\partial \mathcal{L}}{\partial f} = c - \gamma \frac{\partial U}{\partial f} + \mu - \delta = 0. \tag{17}$$

Assuming that $\partial U/\partial x > 0$ it follows that $\gamma > 0$ such that the first constraint is binding. Assuming that $\partial U/\partial l > 0$ it follows that $\mu > 0$ and hence the second constraint is binding. Assuming further that $c \geq 0$ and $\partial U/\partial f < 0$ it follows that $\delta > 0$ and the third constraint is binding such that $f = f_{jk}$.

Furthermore it follows that

$$\frac{\partial \mathcal{L}}{\partial p_j} = y \tag{18}$$

$$\frac{\partial \mathcal{L}}{\partial Q_j} = -\gamma \frac{\partial U}{\partial Q_j} \tag{19}$$

$$\frac{\partial \mathcal{L}}{\partial f_{jk}} = \delta \tag{20}$$

$$\frac{\partial \mathcal{L}}{\partial w_k} = -(1-\tau')h. \tag{21}$$

From equation (20), (17), (15) and (13) it follows that

$$\frac{\partial \mathcal{L}}{\partial f_{jk}} = \delta = (1-\tau')w_k + c - \alpha, \tag{22}$$

where $\alpha := \frac{\frac{\partial U}{\partial f} - \frac{\partial U}{\partial h}}{\partial U/\partial x}$. Assuming that $\alpha = 0$ when $\partial U/\partial f = 0$ then requires the assumption stated in the text that $\partial U/\partial h < 0$.

The expenditure function defined by the minimization problem satisfy that

$$E(p_j, w_k, f_{jk}, Q_j, u) = 0, \tag{23}$$

such that

$$\frac{\partial E}{\partial p_j} dp_j + \frac{\partial E}{\partial f_{jk}} df_j + \frac{\partial E}{\partial Q_j} dQ_j = 0 \tag{24}$$

$$\frac{\partial E}{\partial w_k} dw_k + \frac{\partial E}{\partial f_{jk}} df_k = 0, \tag{25}$$

which combined implies that

$$-\frac{\partial E}{\partial Q_j}dQ_j = \frac{\partial E}{\partial p_j}dp_j + \frac{\partial E}{\partial f_{jk}}df_{jk} + \frac{\partial E}{\partial w_k}dw_k, \tag{26}$$

with $df_{jk} := df_j + df_k$.

Evaluating the derivatives at the national average and applying the Envelope theorem $\partial E/\partial z = \partial \mathcal{L}/\partial z$ it follows that

$$-\frac{\partial E}{\partial Q_j}dQ_j = \bar{y}dp_j + [(1-\tau')\bar{w} + c]df_{jk} - (1-\tau')\bar{h}dw_k, \tag{27}$$

under the assumption that $\alpha = 0$. Defining differentials as $\hat{z} = (z - \bar{z})/\bar{z}$ it then follows that

$$-\frac{\partial E}{\partial Q_j}dQ_j = \bar{p}\bar{y}\hat{p}_j + \left[(1-\tau')\bar{w}\bar{h}\frac{\bar{f}}{\bar{h}} + c\bar{f}\right]\hat{f}_{jk} - (1-\tau')\bar{h}\hat{w}_k. \tag{28}$$

Let $\bar{m}$ be some standardizing constant and define

$$s_y := \frac{\bar{p}\bar{y}}{\bar{m}} \quad s_c := \frac{c\bar{f}}{\bar{m}} \quad s_w := \frac{(1-\tau')\bar{w}\bar{h}}{\bar{m}}, \tag{29}$$

in out application we let $\bar{m} = (1-s)(1-\tau')[\bar{w}\bar{h} + I]$ such that $\bar{m}$ is the spending of consumption with $s$ being the share of disposable income $(1-\tau')[\bar{w}\bar{h}+I]$ saved. Dividing the previous equation with $\bar{m}$ it follows that

$$-\frac{\frac{\partial E}{\partial Q_j}}{\bar{m}}dQ_j = s_y\hat{p}_j + \left[s_w\frac{\bar{f}}{\bar{h}} + s_c\right]\hat{f}_{jk} - s_w\hat{w}_k. \tag{30}$$

Finally by multiplying this equation with $\pi(k|j)$ summing over $k$ and defining

$$\hat{f}_j := \sum_k \hat{f}_{jk}\pi(k|j) \quad \hat{w}_j := \sum_k \hat{w}_k\pi(k|j), \tag{31}$$

it follows that

$$-\frac{\frac{\partial E}{\partial Q_j}}{\bar{m}}dQ_j = s_y\hat{p}_j + \left[s_w\frac{\bar{f}}{\bar{h}} + s_c\right]\hat{f}_j - s_w\hat{w}_j. \tag{32}$$

# A model for estimation of the demand for on-street parking

Albert Gragera* Edith Madsen† Ismir Mulalic‡
Jesper Hybel Pedersen§

May 28, 2020

**Abstract**

This paper presents a stylized econometric model for the demand for on-street parking with focus on the estimation of the elasticity of demand with respect to the full cost of parking. The full cost of parking consists of a parking fee and the cost of searching for a vacant parking space (cruising). The cost of cruising is usually unobserved. Ignoring this issue implies a downward bias of the elasticity of demand with respect to the total cost of parking since the cost of cruising depends on the number of cars parked. We demonstrate that, even when the cost of cruising is unobserved, the demand elasticity can be identified by extending the econometric model to include the spatial interaction between the parking facilities. We illustrate the model with on-street parking data from Copenhagen and find a 55% larger parking demand elasticity to total cost than to parking fees, suggesting a significant cruising bias that is likely to be present in usually reported figures in the literature

**Keywords**: on-street parking, demand estimation, cruising-for-parking.
**JEL classification**: C51, R41, L91.

*Technical University of Denmark, DTU Management, Bygningstorvet 116, 2800 Kgs. Lyngby, Denmark, email: albg@dtu.dk.

†Copenhagen Business School, Porcelanshaven 16A, 2000 Frederiksberg, Denmark, email: ema.eco@cbs.dk .

‡Copenhagen Business School, Porcelanshaven 16A, 2000 Frederiksberg, Denmark, email: imu.eco@cbs.dk. Ismir Mulalic is a Visiting Professor at VU Amsterdam.

§Technical University of Denmark, DTU Management, Bygningstorvet 116, 2800 Kgs. Lyngby, Denmark, email: jeshyb@dtu.dk. Jesper Hybel is supported by the Innovation Fund Denmark via the Urban Project.

# 1 Introduction

Cities around the world use parking policies to regulate the demand for on-street parking and, to some extent, also the level of urban congestion. It is therefore of relevance to estimate the sensitivity of the demand for on-street parking to cost. The full cost of parking (the generalised cost of parking) consists of a parking fee and in addition the cost of searching for a vacant parking space (cruising). The cost of cruising is typically unobserved,[1] but ignoring it biases the estimate of the demand elasticity because the cost of searching for a vacant parking space depends on the number of cars parked, i.e. the demand for parking (Inci et al., 2017; Zakharenko, 2016). This paper proposes a solution to this problem[2]. We formulate an econometric model with both parking fees and cruising for parking as arguments for the demand elasticity for parking. We show how this demand elasticity can be identified, even in situations where cruising for parking is unobserved, when the model is extended to include spatial interaction between the parking facilities.

The economic literature has shown a growing interest for regulatory parking policies and provides a comprehensive treatment of parking pricing (Inci, 2015). Verhoef et al. (1995) analyse different parking policies as a substitute to road pricing and find that the use of parking fees is superior to physical restrictions on parking space supply. Fosgerau and De Palma (2013) show that workplace parking charging schemes can be used as a substitute for the time-varying toll to reduce urban congestion. Moreover, it is typically argued, that parking should be priced at its opportunity cost, just like any other commodity. Arnott et al. (2005) identify a potential triple dividend from optimal parking pricing: reduced cruising for parking, reduced congestion (travel time savings), and the use of parking revenues to lower other taxes (reduced deadweight loss caused by tax distortions). However, in real life parking facilities are often underpriced (Small and Verhoef, 2007). This underpricing leads to cruising for parking which is a pure loss from a social welfare perspective (Shoup, 2005; Calthrop and Proost, 2006). Arnott and Inci (2006)

[1] The exceptions are van Ommeren et al. (2012) and Inci et al. (2017). van Ommeren et al. (2012) examines cruising for parking, but in this study information on parking fees is not available. Inci et al. (2017) show that the mean cruising time can be computed by using parking data about arrival rates and vacancy rates.

[2] This paper is based on a previous working paper by Madsen et. al. (2013) but uses an updated dataset allowing us to achieve robust results.

argue that parking pricing (especially hourly parking fees) has also the downside that it can increase congestion by implying shorter parking durations and thus increase traffic congestion by increasing parking turnover. Arnott et al. (2015) examine the optimal level of curbside parking capacity when both urban transport and curbside parking are underpriced and consider the situation where there is garage parking as an alternative to the curbside.

Our paper adds new insights to the empirical literature that attempts to estimate the price elasticity of parking (see e.g. Feeney 1989; Concas and Nayak 2012; Lehner and Peer 2019). This small – but growing – literature suggests that the price elasticity of parking demand varies depending on many factors (local context, time of the day, trip purpose, income, competing transport options, etc.) and laying in the range between $-0.6$ and $-0.1$; depending on the specification of parking demand (occupancy, dwell time and volume) as shown in Lehner and Peer (2019). Moreover, several studies estimate the price elasticity of demand for parking ignoring the cost of cruising (see e.g. Kelly and Clinch 2009; Hensher and King 2001). However, there is a rather surprising absence of accurate empirical estimates of the effect of the cost of parking on the demand for parking. This effect is important as it is required for rigorous welfare analysis of a parking policy.

In this paper, we propose a new stylized econometric model to identify the elasticity of parking demand to total parking cost, using the usually available data collected by cities (parking occupancy rates and parking fees). We illustrate this model using parking data available for Copenhagen. We show that the effect of the parking fee is always less than the effect of the cost of parking in absolute value. We also show that the effect of the cost of parking can be identified, even if the cost of cruising is unobserved, by extending the econometric model to include the spatial interaction between the parking facilities (streets). Our empirical findings suggest that a significant cruising bias is likely to be present in the parking price elasticity measures in the literature (when interpreted as elasticities to the total parking cost).

The next section introduces an econometric model for the demand for on-street parking; section 3 presents the empirical illustration, and section 4 concludes.

# 2 An econometric model of the demand for parking

In this section, we specify an econometric model for the demand for on-street parking. First, in section 2.1, we describe a very simple model without spatial interactions. Then, in section 2.2, we consider an extension of the model that takes the spatial interaction into account.

For both models, the demand for on-street parking is described in terms of the occupancy rate, i.e. the number of parked cars relative to the number of legal parking lots. The supply of parking lots is assumed to be constant and thus, the occupancy rate reflects the demand for on-street parking. There is no modelling of external factors affecting the demand for parking by e.g. affecting the overall traffic demand or number of cars. In this way, the model proposes a partial description without interaction with other sectors. We also simplify by ignoring the effect on the demand for on-street parking of other parking alternatives (e.g. private parking houses). We suggest that this effect is small and thus of little importance, see section 3.1.

## 2.1 A simple model

First, let the demand for parking in street $i$ at period $t$ in terms of the occupancy rate, $O_{it}$, (the number of cars parked divided by the number of parking spaces) be given by

$$O_{it} = \alpha_i + \beta c_{it} + \varepsilon_{it} \quad (1)$$

$$c_{it} = p_{it} + S(O_{it}) \quad (2)$$

where $c_{it}$ is the total cost of parking in street $i$ at period $t$, $\alpha_i$ is a street-specific fixed effect, and $\varepsilon_{it}$ is an idiosyncratic error term. The cost $c_{it}$ consists of a direct cost $p_{it}$ (a parking fee) and an indirect cost, $S(O_{it})$, that reflects the searching costs (cruising) and depends on the occupancy rate $O_{it}$. In line with the literature we assume that the searching cost function $S(\cdot)$ is increasing in the occupancy rate, see e.g. Anderson and De Palma (2004). Altogether equations (1)-(2) express that an increase in the parking fee reduces $O_{it}$ and thus increases the number of vacant parking spaces; this in turn implies a lower cruising time and by that a lower cost of searching. The specification highlights the fact that the cost of searching, and by that the cost $c_{it}$, is an endogenous variable in the parking demand equation.

In our dataset, we do not have any information on searching in terms of time and costs and therefore we will specify the functional relationship between the searching costs and the occupancy rate in order to arrive at a reduced form equation for $O_{it}$ (see below). It is important to note that if we did have information on searching then the total cost of parking $c_{it}$ could be calculated and a valid instrument for $c_{it}$ would be the parking fee $p_{it}$. Consequently, the parameter $\beta$ could be estimated by IV estimation.

The street-specific fixed effects capture all time-invariant differences in the demand for parking between streets such as the distance to the location of shopping and leisure activities and the number of residential parking permits (residents pay an annual fee and in return gain the right to park on-street in a specific area). Very importantly, the inclusion of street-specific fixed effects controls for endogeneity of the average parking fee level in a street. It is typically the case that the fees are higher in the city center where the demand is also high and vice versa in the areas further away from the city center. The street-specific fixed effects allow for this type of endogeneity but exclude the case where a change in the parking fee over time is a response to a change in demand. We find that this assumption is reasonable in most empirical applications to on-street parking. Typically, these adjustments are a result of some political decisions rather than demand reactions.[3]

In order to obtain a reduced form equation for the parking demand in terms of the occupancy rate $O_{it}$ we need to specify how the searching costs depend on the occupancy rate. We assume that the costs of searching are linear in the occupancy rate:

$$S\left(O_{it}\right) = a + bO_{it} \quad \textit{where} \;\; b > 0 \tag{3}$$

Using eq. (3) it is straightforward to show that the reduced form equation implied by equations (1)-(2) is

$$O_{it} = \tilde{\alpha}_i + \tilde{\beta} p_{it} + \tilde{\varepsilon}_{it} \tag{4}$$

where $\tilde{\alpha}_i = \left(\alpha_i + a\beta\right) / \left(1 - b\beta\right)$, $\tilde{\beta} = \beta / \left(1 - b\beta\right)$ and $\tilde{\varepsilon}_{it} = \varepsilon_{it} / \left(1 - b\beta\right)$. For

[3]This is reasonably to be the case for our illustrative example from the city of Copenhagen, see section 3.1.

$\beta < 0$ then $\tilde{\beta} \in ]\beta, 0]$ since $b > 0$ such that the parameter corresponding to $p_{it}$ in the reduced form equation is less than $\beta$ in absolute value. The parameter describes the total effect of increasing the parking fee. The direct effect is that it will decrease the demand for parking and the indirect effect is that this in turn will decrease the searching cost which will increase the demand for parking. The larger the value of $b$ the smaller the absolute value of the total effect. From this reduced form equation it is not possible to identify the parameter $\beta$ in the demand equation and the parameters $a$ and $b$ in the searching cost function separately. However, if the costs of searching are piecewise linear in the occupancy rate then all parameters are identified if there are streets where the occupancy rate is below a threshold value of the occupancy rate where the cost of searching is zero (see Appendix A).

Obviously, the assumption about the searching cost being linear in the occupancy rate is strong and a more realistic assumption would be that the marginal cost of searching is increasing in the occupancy rate. This could for example be modelled as $S(O) = c/(1-O)$ where $c > 0$ as done in Anderson and De Palma (2004). However, this will lead to a more complicated reduced form equation for the occupancy rate which is not useful in empirical work.

## 2.2 Spatial interaction between the parking facilities

The framework in section 2.1 assumes that the demand for parking in a specific street is independent of the cost of parking in all other streets. This assumption is obviously not likely to hold in practice since the demand for parking in a specific street expectedly will also depend on the cost of parking in neighboring streets. We now extend the model to allow for this. More formally, we assume that the demand for parking in street $i$ depends on both the cost of parking in street $i$ and on the cost of parking in neighboring streets $j \neq i$. As before, the cost of parking consists of a parking fee and a searching cost which is increasing in the occupancy rate. The demand for parking in street $i$ at time $t$ is now given by:

$$O_{it} = \alpha_i + \beta c_{it} + \gamma \sum_{j \neq i} w_{ij} c_{jt} + \varepsilon_{it} \quad (5)$$

$$c_{jt} = p_{jt} + S(O_{jt}) \quad (6)$$

The parameter $\gamma$ corresponding to the term $\sum_{j\neq i} w_{ij}c_{jt}$ in eq. (5) describes how the demand for parking in a specific street is affected by the costs of parking in neighbouring streets. The spatial weights $w_{ij}$ for $j \neq i$ are pre-specified and each weight defines the exact neighbouring effect of a specific street. We use the following geographically derived weights:

$$w_{ij} = \exp\left(-\theta d_{ij}\right) \tag{7}$$

where $d_{ij}$ is the the shortest route distance between streets $i$ and $j$, and $\theta > 0$ is a specified constant. The weights are exponentially decreasing in the distance and approaches zero as the distance increases. We use the minimax normalization of the weights (a common scaling of all weights) and note that this normalization preserves the symmetry such that $w_{ij} = w_{ji}$. For a more extensive discussion of spatial weights, see e.g. Upton et al. (1985) and Anselin (2013).

The model defined by equations (5)-(6) allows for substitution between the demand for parking in different streets as given by the spatial weights and the model parameters. The model implies the following own and cross elasticities with respect to the total parking cost:

$$e_{ii} \equiv \frac{\partial O_{it}}{\partial c_{it}} / \frac{O_{it}}{c_{it}} = \beta \frac{c_{it}}{O_{it}} \tag{8}$$

$$e_{ij} \equiv \frac{\partial O_{it}}{\partial c_{jt}} / \frac{O_{it}}{c_{jt}} = \gamma w_{ij} \frac{c_{jt}}{O_{it}} \tag{9}$$

Intuitively, we would expect $\gamma > 0$ such that all other streets are substitutes for parking in one particular street. Everything else equal, the closer two streets are located to each other the higher the substitution effect is, i.e $e_{ij} > e_{ik}$ for $d_{ij} < d_{ik}$ since $w_{ij} > w_{ik}$. It is important to note that the difference in substitution effect between two different streets is determined by the parameter $\theta$ which is pre-specified and not estimated. In this study, the parameter $\theta$ is set at 0.5. This implies that spatial weights are close to zero ($< 0.1$) for streets more than one kilometre away. The need to specify the spatial structure a priori is a limitation in all spatial models, see Gibbons and Overman (2012) for a discussion of this.

As our dataset does not contain information on searching time or searching cost, equations (5)-(6) cannot by used directly in estimation. Instead our approach is to impose assumptions on the relationship between the searching cost and the occupancy rate and use that to reach a reduced form equation that can be estimated. As eq. (3) in section 2.1 we assume that the costs of searching are linear in the occupancy rate, i.e. $S(O) = a + bO$. Using this, equations (5)-(6) can be written as (in matrix notation):

$$O_{nt} = \tilde{\alpha}_n + \tilde{\beta} p_{nt} + \tilde{\gamma} W_n p_{nt} + \lambda W_n O_{nt} + \tilde{\varepsilon}_{nt} \tag{10}$$

where the $n$-vector $\tilde{\alpha}_n$ have elements $(\alpha_i + a\beta + a\gamma \sum_{j \neq i} w_{ij}) / (1 - b\beta)$, parameters are defined as $\tilde{\beta} = \beta/(1 - b\beta)$, $\tilde{\gamma} = \gamma/(1 - b\beta)$ and $\lambda = b\tilde{\gamma}$, the weight matrix $W_n$ has elements $w_{ij}$ and zeros in the diagonal, and the error term $\tilde{\varepsilon}_{nt}$ is i.i.d. $N(0, \tilde{\sigma}^2 I_n)$ with $\tilde{\sigma}^2 = \sigma^2/(1 - b\beta)^2$ across $t = 1, ..., T$. This is the standard Spatial Durbin Model (SDM) with fixed effects $\tilde{\alpha}_n$, exogenous regressors $p_{nt}$ and $W_n p_{nt}$ and the spatially lagged endogenous regressor $W_n O_{nt}$, see e.g. LeSage and Pace (2009). Like in the simple framework of section 2.1 the parameters of main interest, $\beta$ and $\gamma$ in eq. (5), do not appear as parameters in the SDM model and as before we have that when $\beta < 0, \gamma > 0$ and $b > 0$ then $\tilde{\beta} \in ]\beta, 0]$ and $\tilde{\gamma} \in [0, \gamma[$. Therefore estimates of $\tilde{\beta}$ and $\tilde{\gamma}$ will underestimate the marginal effects of increasing parking costs $\beta$ and $\gamma$. However, the structural parameters $\beta, \gamma$ and $b$ in the demand for parking eq. (5) can be obtained as functions of the parameters $\tilde{\beta}, \tilde{\gamma}$ and $\lambda$, as follows:

$$\begin{aligned} b &= \frac{\lambda}{\tilde{\gamma}} \\ \beta &= \frac{\tilde{\beta}}{1 + \lambda/(\tilde{\beta}\tilde{\gamma})} \\ \gamma &= \tilde{\gamma} - \lambda\beta \end{aligned}$$

Estimation of eq. (10) is performed by maximum likelihood as described in Lee (2004). In addition, Lee (2004, 2007) investigate the sources of identification and various reasons for failure to identify the model parameters in different versions of spatial autoregressive (SAR) models. It is shown that in case the exogenous regressors (in our case $p_{nt}$ and $W_n p_{nt}$) and the spatially lagged regressor are collinear the source of identification will be coming from the covariance structure of the error terms. This in turn implies that the covariance structure of the error term in eq.

(10) must be correctly specified. In our case, we assume that the elements in the error term are independent across $i, t$ with constant variance. Obviously, an identification that relies on variation in exogenous variables is more appealing since assumptions imposed on the error term such as constant variance are somewhat arbitrary. The problem is discussed in a recent paper by Gibbons and Overman (2012) and is similar to the identification problem in models where the outcome variable depends on some expected value of the outcome variable, the reflection problem, see Manski (1993).

Finally, Lee and Yu (2010) show that the estimation of a spatial model with unit-specific fixed effects is straight forward. It is done by using results from standard panel data models, i.e. maximization of the conditional likelihood function gives consistent estimators of the model parameters where the conditioning is done with respect to unit-specific averages of the dependent variable as sufficient statistics for the unit-specific effects.

# 3 Empirical illustration

This section of the paper presents an illustration of the application of the econometric model. We use parking data from the city of Copenhagen. With this, it is in principle possible to test the model and estimate demand elasticity of parking with respect to its full cost.

Section 3.1 describes the parking market and parking policies in the city of Copenhagen. It also includes a discussion of a number of key assumptions that underlie the identification of the model and the interpretation of its parameters. The data set provided by the city of Copenhagen for the analysis is described in section 3.2 and estimation results are discussed in section 3.3. We discuss our findings on the parking price elasticity, relate our result to the estimates provided by the existing literature and conclude the section by discussing the results obtained from the estimation of a standard spatial model with street-specific fixed effects.

## 3.1 Parking in the city of Copenhagen

About two-thirds of the parking spaces in the city of Copenhagen are on-street and hence this is the dominating way of parking. The city of Copenhagen has, like many other larger cities, a long history of paid parking (both for publicly provided as well as privately provided parking places). In 1990 the city of Copenhagen initiated a new system for payments for parking, where the inner city was divided into different zones and then successively expanded until the regulation covered the whole city central area in 2007.[4] The purpose of the system was to reduce the traffic and the number of parked cars in the city, especially commuting in cars to workplaces in central Copenhagen. In the zonal system, all on-street parking is charged a fee depending on the duration of the parking, time of the day and the location of the zone. The zones closest to the historical city center are more expensive. Many other European cities use similar systems where payment for on-street parking varies across zones and time-intervals.[5] At present, the zonal system covers three zones: red is the city center with few residents and many shops, restaurants and offices, green and blue have more residents, see figure 1 (a).

## 3.2 Data

The survey data used in the empirical analysis is provided by the city of Copenhagen. The sample covers the years 2008-2016 with semi-annual counts (in April and September, starting with September 2008). Each count includes three daily measurements (at 12:00, 17:00 and 22:00). For all streets in central Copenhagen we know the number of legal parking spaces as well as the number of occupied spaces, for each of the three daily counts. We do not have information about cruising costs or cruising time. Furthermore, we do not have information about alternative parking (e.g. private parking houses and workplace parking).[6] Figure 1 shows both the map of parking regulated zones and the occupancy rates measured

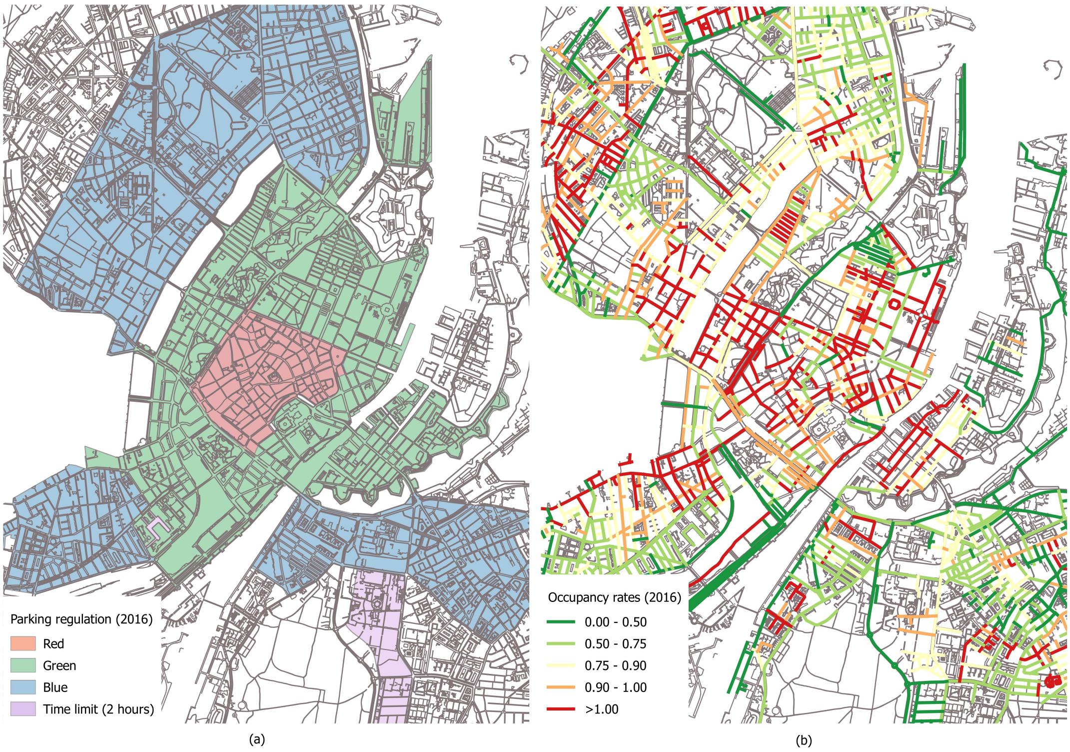
[4] The regulatory area did not change until the introduction of a new (Yellow) parking zone in 2017, besides changes in the pricing scheme.

[5] Special rules apply for residents in a parking zone such that they can purchase parking permits that grant them unrestricted parking close to their home address (when available). The price of a residential parking permit is about €90 per year per car. The parking permit is connected to a specific car and there is no limit to the number of residence parking permits available.

[6] If available, such information could be entered into the model as additional parking lots.

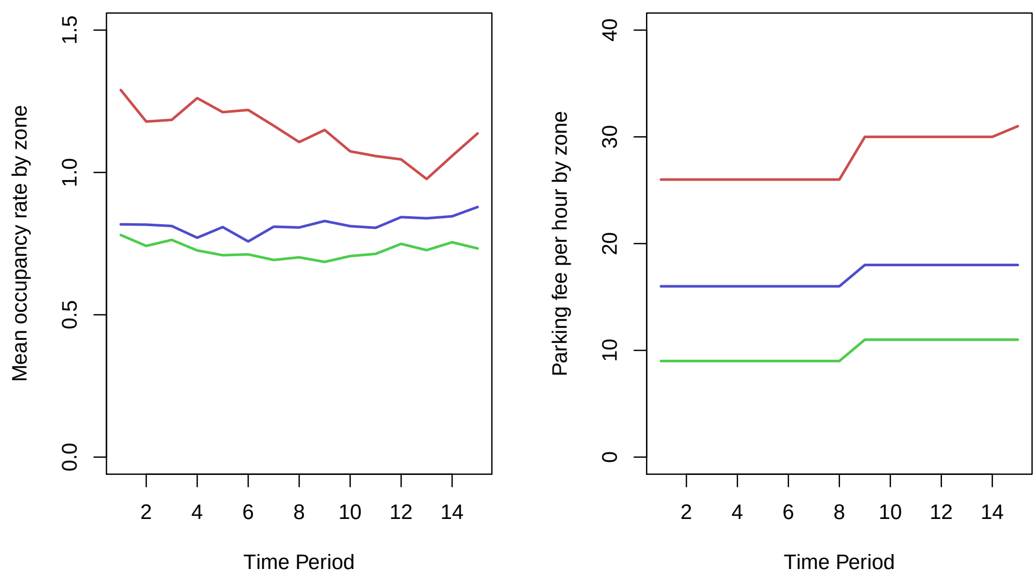
at noon in April 2016.

Figure 1: Map of Copenhagen's parking regulation (a) and occupancy rates at 12:00h (b) in April 2016




In the empirical analysis, we have reduced the dataset in several ways. First, the three different time counts represent different traffic situations. For example, in the Danish National Travel Survey, we see many shoppers and short term parkers at noon while residents are more dominating after work hours. For the following empirical analysis, we choose to use the figures from the noon count (12:00 am). Second, the dataset provides the number of occupied spaces as well as the number of legal parking spaces for each street. With this information we can calculate the occupancy-rate for each street. Note that the occupancy rate can be above 100%. This is possible since the number of legal parking lots is rarely physically marked and thus it is possible to deviate from the estimated number depending on the size of the cars and the density of the standard size of parked cars. Because of this we accept an occupancy rate above 100% in our dataset but choose to censor the occupancy rates above 130%.[7] Third, as long as measures are taken at street level, we also drop the observation for those crossing different regulatory zones as this prevents us from assigning them a single parking fee or discern the occupancy rates at each subsegment. This mainly affects the few major urban avenues connecting different city districts.

Figure 2: Average nominal parking fees (DKK/h) and occupancy rates for red, blue and green zones




Notes: period represents the semi-annual counts (in April and September) starting with September 2008.

Figure 2 shows that the mean occupancy rate for the red zone (central Copenhagen) has consistently stayed above 100% which indicates that there is generally no excess supply of parking places in the zone, i.e. empty spots will generally be filled immediately and thus cruising for parking is present. For the green and blue zones as well as for the outer zone we also find very high occupancy rates indicating little or no excess supply and potentially cruising for parking. We also note that the occupancy rates are highest in the red zone at the 12.00 am (noon)

[7]This rule of censoring occupancy rates above 130% is based on the technical analysis of the parking capacity in the City of Copenhagen.

count. The temporal pattern of occupancy rates does also seem to hint a potential substitution effect between contiguous zones coinciding with pricing changes.

The parking fees for the zones are shown in figure 2. The parking fee for the red zone (the city center) is almost three times as high as for the blue zone. Outside the three zones (the outer city) there are generally no fees for parking. We also see that the nominal prices have been changed only once for the years 2008-2016. This obviously represents a limitation for the econometric analysis.

## 3.3 Empirical results

We now describe the empirical results. We first present our results on the parking price elasticity based on the simple demand model. Next, we discuss the results obtained from the estimation of the standard spatial model.

Table 1 reports the estimation results. All estimated equations include street- and time-specific effects. The street-specific effects control for all the time-invariant systematic differences in the demand for parking at the street level,[8] while the time-specific effects account for unobserved parking demand shocks over time that affects all streets (e.g. business cycles). Column [1] shows the estimates for the simple demand reduced-form equation based on eq. (4). Since the supply of on-street parking has been constant in the period of the observation, we interpret the effect of the parking fee on the occupancy rate as a demand effect. As we expected, an increase in the parking fee decreases the demand for on-street parking. The parameter associated with the parking fee ($\widetilde{\beta}$) in the simple model is estimated to $-0.029$, see column [1]. The parameter estimate is tight and indicates a plausible effect. The 95% confidence interval is estimated to be from $-0.034$ to $-0.023$.

The estimation result allows us to derive the parking fee elasticity.[9] Notice here that the parking fee elasticity is different from the elasticity of demand with respect to total cost of parking, since the total cost of parking consists of a parking fee and the cost of cruising. The parking fee elasticity in the red zone (the historical city center) at the sample average of the occupancy rate in the red zone (1.2) and the parking fee of 31 DKK/hour is $-0.75$, i.e. raising the parking fee in the red

[8]E.g. street attributes (e.g. one-way traffic), number of residential units, the distance to the location of shopping and leisure activities, the number of residential parking permits, the supply of public transport, etc.

[9]The parking fee elasticity is defined as $\varepsilon_{O,p} = \frac{\partial O}{\partial p}\frac{p}{O} = \widetilde{\beta}\frac{p}{O}$.

Table 1: Models for on-street parking in terms of the occupancy rate

| | [1] Standard model Eq. (4) | [2] Spatial Durbin Model (SDM) Eq. (10) |
|---|---|---|
| Parking fee, ($\tilde{\beta}$) | -0.029*** | -0.045*** |
| | (0.003) | (0.005) |
| $W \cdot p$, ($\tilde{\gamma}$) | | 0.0002*** |
| | | (0.0001) |
| $W \cdot O$, ($\lambda$) | | 0.002*** |
| | | (0.0005) |
| Street fixed effect | yes | yes |
| Period fixed effect | yes | yes |
| $R^2$ | 0.010 | |
| Log likelihood | | 3,365 |
| Number of observations | 10,935 | 10,935 |

Note: dependent variable is the occupancy rate $O$ (share); censoring $O = 1.30$; parking fee is measured in DKK/hour; $W$ is the spatial weights matrix; *** indicates that estimates are significantly different from zero at the 0.01 level; in the SDM $\theta = 0.05$; standard errors are in parentheses.

zone by 1% reduces demand for on-street parking in the historical city center by 0.75%. Taking into account previously reported confidence interval, the parking fee elasticity lies in the range $[-0.88, -0.59]$. These finding are consistent with the ones reported by the literature (see e.g. Lehner and Peer (2019)).

However, these are underestimates of the parking demand elasticity because, as shown in section 2.1, the parameter corresponding to the parking fee in the reduced form equation ($\widetilde{\beta}$) is less than the parameter corresponding to the total cost of parking ($\beta$) in absolute value. The *crusing bias* is caused by the fact that while the cost of cruising is usually unobserved, ignoring it bias the estimation of the price elasticity of demand because of the dependence of the costs of cruising on the number of cars parked. Our findings indicate, that due to the cruising bias, the parking demand elasticity (the car drivers' response to an increase in the total cost of parking) is most likely larger than proposed in the literature.

Column [2] in table 1 reports the estimated coefficients for the Spatial Durbin Model (SDM) based on eq. (10).[10] All the estimated coefficients are significant and

[10]Estimation of the spatial model in eq. (10) is in principle straight forward, see section 2.2.

have the expected signs. The $\tilde{\beta}$ is estimated to be $-0.045$; which, for the red zone, translates into a parking fee elasticity of $-1.16$, a 55% larger (in absolute terms) than the elasticity computed from the standard demand model. The coefficients associated with the parking fees in the neighbouring streets ($\tilde{\gamma}$) and with the occupancy rates in the neighbouring streets ($\lambda$) are both significant and positive. This suggests that increasing parking fees and occupancy rates in the neighbouring streets of a street, raise the demand for parking in that street.

We can also use these estimates ($\tilde{\beta}$, $\tilde{\gamma}$ and $\lambda$) to recover the structural parameters ($\beta$, $\gamma$ and $b$) and consequently to learn about the impact of the total cost of parking on the parking demand. We find that coefficient associated with the total cost of parking ($\beta$) is significantly negative and equal to $-0.070$ (std.err is 0.012), $b$ is 8.024 (std.err is 2.545), and $\gamma$ is 0.0004 (std.err is 0.0001). The total parking cost elasticity for the red zone is then $-1.81$, suggesting that a 1% increase in the parking costs will reduce the occupancy rate in the red zone by 1.8%. This elasticity is essential for a rigorous welfare analysis and it should not be mixed up with the parking fee elasticity, as currently used in the literature due to the lack of better estimates. Our results imply that the introduction of paid parking (or changing parking fees) likely have larger welfare effects than previously suggested. Finally, higher parking fees do reduce cruising costs ($b = 8.025$). Consequently, the true impact of an increase in the parking fee on the demand for parking would be underestimated if the cruising costs are ignored.

# 4 Conclusion

This paper deals with estimation of the elasticity of the demand with respect to the full cost of parking for on-street parking. We take into account the data availability, i.e. (city) transport authorities collect parking data that includes the occupancy rates and sporadically and if relevant the parking fees. This paper proposes a new methodological framework to clarify the identification of the effect of the cost of parking, consisting of the costs of searching for parking (cruising) and a parking fee, on the demand when the cost of searching is unobserved. We illustrate the model using on-street data from the city of Copenhagen for the years 2008-2016. Our illustrations suggest that the parking demand elasticity is most likely larger than the one proposed in the literature.

Our findings have a number of implications. First it demonstrates that parking fees can potentially be a useful policy instrument to organize the parking market and to reduce the external costs of traffic such as congestion (cruising), air pollution, and other relevant local environmental externalities. It also demonstrates that, in line with the literature (see Arnott et al. 1991), a spatially differentiated parking fee is necessary to induce the optimal parking pattern. Second, the proposed empirical methodology can be useful for the estimation of other similar reduced form demand equation describing the demand with the constrained capacity. In particular the reduced form demand equation resulting from a bottleneck model is a good example (see e.g. Arnott et al. 1993). Finally, the proposed methodology makes it possible to make a straightforward extension of the demand model to include spatial interactions. In this way many of the identification problems in applied spatial economics can be avoided.

**Acknowledgement 1** *The authors thank Mogens Fosgerau, Jos Van Ommeren, Eren Inci and Ninette Pilegaard for their useful suggestions on an earlier draft. We are grateful to the city of Copenhagen (Åse Boss Henrichsen, Søren Lindgreen and Anders Møller) for providing the data. Seminar participants at the Kuhmo Nectar Conference on Transportation Economics 2012, LATSIS, and the Technical University of Denmark also provided helpful comments. Research support from the Danish Council for Strategic Research is acknowledged. All remaining errors are the authors' alone.*

# References

Anderson, S. and De Palma, A. (2004). The economics of pricing parking. *Journal of Urban Economics*, 55:1–20.

Anselin, L. (2013). *Spatial econometrics: methods and models*, volume 4. Springer Science & Business Media.

Arnott, R., de Palma, A., and Lindsey, R. (1991). A temporal and spatial equilibrium analysis of commuter parking. *Journal of Public Economics*, 45:301–335.

Arnott, R., De Palma, A., and Lindsey, R. (1993). A structural model of peak-period congestion: A traffic bottleneck with elastic demand. *The American Economic Review*, pages 161–179.

Arnott, R. and Inci, E. (2006). An integrated model of downtown parking and traffic congestion. *Journal of Urban Economics*, 60:418–442.

Arnott, R., Inci, E., and Rowse, J. (2015). Downtown curbside parking capacity. *Journal of Urban Economics*, 86(C):83–97.

Arnott, R., Rave, T., Schöb, R., et al. (2005). Alleviating urban traffic congestion. *MIT Press Books*, 1.

Calthrop, E. and Proost, S. (2006). Regulating on-street parking. *Regional Science and Urban Economics*, 36(1):29–48.

Concas, S. and Nayak, N. (2012). A meta-analysis of parking pricing elasticity. Technical report.

Feeney, B. P. (1989). A review of the impact of parking policy measures on travel demand. *Transportation planning and technology*, 13(4):229–244.

Fosgerau, M. and De Palma, A. (2013). The dynamics of urban traffic congestion and the price of parking. *Journal of Public Economics*, 105:106–115.

Gibbons, S. and Overman, H. G. (2012). Mostly pointless spatial econometrics? *Journal of Regional Science*, 52(2):172–191.

Hensher, D. A. and King, J. (2001). Parking demand and responsiveness to supply, pricing and location in the sydney central business district. *Transportation Research Part A: Policy and Practice*, 35(3):177–196.

Inci, E. (2015). A review of the economics of parking. *Economics of Transportation*, 4:50–63.

Inci, E., van Ommeren, J., and Kobus, M. (2017). The external cruising costs of parking. *Journal of Economic Geography*, 47:333–355.

Kelly, J. and Clinch, J. (2009). Temporal variance of revealed preference on-street parking price elasticity. *Transport Policy*, 16(4):193–199.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, 72(6):1899–1925.

Lee, L.-f. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140(2):333–374.

Lee, L.-f. and Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 154(2):165–185.

Lehner, S. and Peer, S. (2019). The price elasticity of parking: A meta-analysis. *Transportation Research Part A: Policy and Practice*, 121:177–191.

LeSage, J. and Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Madsen, E., Mulalic, I., and Pilegaard, N. (2013). A model for estimation of the demand for on-street parking. MPRA Working Paper 52301.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542.

Shoup, D. (2005). *The High Cost of Free Parking*. Planners Press, Chicago.

Small, K. A. and Verhoef, E. T. (2007). *The Economics of Urban Transportation*. Routledge, London.

Upton, G., Fingleton, B., et al. (1985). *Spatial data analysis by example. Volume 1: Point pattern and quantitative data*. John Wiley & Sons Ltd.

van Ommeren, J., Wentink, D., and Rietveld, P. (2012). Empirical evidence on cruising for parking. *Transportation Research Part A: Policy and Practice*, 46:123–130.

Verhoef, E., Nijkamp, P., and Rietveld, P. (1995). The economics of regulatory parking policies: the (im)possibilities of parking policies in traffic regulation. *Transportation Research Part A*, 29:141–156.

Zakharenko, R. (2016). Time dimension of parking economics. *Transportation Research Part B: Methodological*, 91:211–228.

# 5 Appendices

## 5.1 Appendix A

Assume now that the costs of searching are piecewise linear in the occupancy rate

$$S(O_{it}) = \begin{array}{cc} 0 & \text{if } O_{it} < \Theta \\ a + bO_{it} & \text{if } O_{it} \geq \Theta \end{array} \tag{A.1}$$

It means that the cost of searching is zero when the occupancy rate is less than a threshold value $\Theta$ (e.g. $\Theta = 70\%$) and linear and increasing for values above $\Theta$. This might be a more realistic assumption than having the costs being linear in the occupancy rate since if the occupancy rate is low then there will be empty parking spaces and the cost of searching is zero. The threshold value at $\Theta$ reflects that if the occupancy rate is above this level then it is more likely that all spaces are occupied which implies cruising. Note also that, as emphasised by Arnott & Inci (2006), given perfect information about parking spaces and optimal pricing of parking, cruising time is (close to) zero.

The reduced form for the eq. (A.1) is now given by

$$O_i = \begin{array}{cc} \alpha_i + \beta p_i & \text{if } p_i \geq (\Theta - \alpha_i)/\beta \\ \frac{\alpha_i + a\beta}{1-b\beta} + \frac{\beta}{1-b\beta}p_i & \text{if } p_i < (\Theta - \alpha_i)/\beta \end{array} \tag{A.2}$$

In this case all parameters are identified if there are streets where the occupancy rate is less than $\Theta$. This identification strategy utilises the fact that the expression is non-linear in the exogenous variable. The difficulty is related to the correct censoring of the occupancy rate. The threshold value $\Theta$ should be selected at the level at which the cost of searching turns to zero.