Utrecht University

School of Economics

Endogenous local labour markets, regional aggregation and agglomeration economies

Jordy Meekes
Wolter H.J. Hassink

U.S.E. Discussion Paper Series nr: 18-03

Tjalling C. Koopmans Research Institute
Utrecht University School of Economics
Utrecht University

Kriekenpitplein 21-22
3584 EC Utrecht
The Netherlands
telephone +31 30 253 9800
fax +31 30 253 7373
website www.uu.nl/use/research

The Tjalling C. Koopmans Institute is the research institute and research school of Utrecht University School of Economics. It was founded in 2003, and named after Professor Tjalling C. Koopmans, Dutch-born Nobel Prize laureate in economics of 1975.

In the discussion papers series the Koopmans Institute publishes results of ongoing research for early dissemination of research results, and to enhance discussion with colleagues.

Please send any comments and suggestions on the Koopmans institute, or this series to J.M.vanDort@uu.nl

How to reach the authors

Please direct all correspondence to the first author.
Jordy Meekes*#
Wolter H.J. Hassink#~
*University of Melbourne
Melbourne Institute of Applied Economic and Social Research
Grattan Street,
Parkville, Victoria, 3010
Australia
#Utrecht University
Utrecht University School of Economics
Kriekenpitplein 21-22
3584 TC Utrecht
The Netherlands.
~Institute for the Study of Labor (IZA)
Schaumburg-Lippe-Strasse 5-9
53113 Bonn
Germany
E-mail: Jordy.meekes@unimelb.edu.au
W.H.J.Hassink@uu.nl

Utrecht University School of Economics
Tjalling C. Koopmans Research Institute
Discussion Paper Series 18-03

Endogenous local labour markets, regional aggregation and agglomeration economies

Jordy Meekes[ab]
Wolter H.J. Hassink[bc]

[a]Melbourne Institute of Applied Economic and Social Research
University of Melbourne

[b]Utrecht University School of Economics
Utrecht University

[c]Institute for the Study of Labor
Germany

July 2018

Abstract
In this paper, we examine the structure of workers' local labour market (LLM) and its economic consequences. We endogenise workers' LLM to commuting outcomes and worker characteristics. The descriptive results indicate that both male workers and high-educated workers especially are characterised by large LLMs. The empirical results show that the urban wage premium (UWP), explained by the returns to agglomeration in wages, increases by a magnitude of two to three in the level of regional aggregation. We also focus on subgroup differentials in the returns to agglom-eration economies. High-educated workers experience a higher UWP than low-educated workers, but we find no systematic difference between the UWP of men and women when holding the re-gional aggregation level constant. In addition, we examine the returns to agglomeration in wages and employment for workers who experienced job displacement. We show that at a relatively high level of regional aggregation, displaced workers in dense LLMs, compared to displaced workers in more sparse LLMs, experience modest losses in wages and comparable losses in employment.

Keywords: Local labour markets, Urban wage premium, Employment, Commuting, Regional aggregation

JEL classification: R12, R23, J31, J63, J64, J65

Acknowledgements
We wish to thank seminar participants at the 32nd Annual Conference of the European Society for Population Economics, and Utrecht University School of Economics. We are grateful to Statistics Netherlands for giving us access to the administrative data.

1. Introduction

Economists and geographers use the concept of a local labour market (LLM) to describe a self-contained regional area of residence and work activity. LLMs have received much attention from researchers and policy-makers, as they reveal regional differences in population and employment density, employment opportunities, productivity, wages and housing prices (Topel, 1986; Bhaskar et al., 2002; Moretti, 2011). The regional differences across LLMs are explained by agglomeration economies, which refer to benefits derived from spatial concentration of economic processes. The urban economics literature generally makes a distinction between two types of agglomeration economies. The first type is urbanisation economies, which corresponds to externalities from overall economic activity and diversity. The second type is localisation economies, which corresponds to externalities from specialisation of firms located in proximity of each other. The theoretical micro-foundations of agglomeration economies include improved matching of employers to workers and other inputs, sharing of resources and risk, and learning through the generation, diffusion and accumulation of knowledge (Duranton and Puga, 2004). The structure of an LLM differs among subgroups of the population (Farmer and Fotheringham, 2011), which could lead to subgroup differentials in agglomeration economies. Positive spillovers from agglomeration economies may lead to higher wages and better employment prospects for workers.

There is no general consensus, however, on the spatial scale of LLMs at which agglomeration economies take place (Rosenthal and Strange, 2001; Combes and Gobillon, 2015). From a policy perspective, a better understanding of the structure of workers' LLM is important to improve the effectiveness of regional policies that aim to stimulate agglomeration benefits and regional productivity growth. On the one hand, the returns to agglomeration could be decreasing in the spatial scale. For example, the transfer of knowledge might be more prevalent at a low spatial scale if the transmission of knowledge requires face-to-face contact or happens at accidental encounters. Arzaghi and Henderson (2008) show that for advertising firms, knowledge spillovers from density are large but attenuate rapidly with geographical distance. On the other, the returns to agglomeration are increasing in the spatial scale. Sharing of resources and market access for final and intermediate goods could be more prevalent at a high spatial scale, as state-level transportation modes may affect the location patterns of industries that are sensitive to shipping costs (Rosenthal and Strange, 2001). In addition, the matching mechanism could be prevalent at a high spatial scale, as for example workers with a higher education often become employed far away from their home. An approach to examine the role of spatial scale in the returns to agglomeration in wages is introduced and applied by Rosenthal and Strange (2003, 2008), who examine the attenuation of agglomeration benefits by drawing concentric rings at various distances around the worker's Work Public Use Micro Area (PWPUMA). They find that the effect of agglomeration on wages rapidly attenuates with distance.

The main aim of our paper is to examine the importance of regional aggregation for the returns to agglomeration. So far, studies use different types of administrative regional classifications to assess the implications of geographical aggregation (Briant et al., 2010; Lindo, 2015). In many countries, there are generally two to three regional classifications available at different levels of regional aggregation. We introduce a new approach to assess the role of the spatial scale of LLMs in agglomeration economies by using a flow-based cluster algorithm to define LLMs at various levels of regional aggregation. A higher level of regional aggregation and spatial scale is represented by fewer distinct LLMs. Additionally, two economic applications of the returns to agglomeration are analysed. The first application concerns the effect of agglomeration externalities on wages, which is referred to as the urban wage premium (UWP). We examine the differences in the UWP across LLMs by using various levels of regional aggregation, to assess whether the matching, sharing and learning micro-foundations of agglomeration economies are more prevalent at different spatial scales. The second application focuses on the returns to agglomeration through one specific micro-foundation: whether a dense labour market facilitates better job matching between workers and employers. To examine the role of agglomeration economies in job matching, we use an empirical design involving job displacement due to firm bankruptcy (hereafter: job displacement). A priori, the role of agglomeration in job matching is ambiguous. Denser labour markets are characterised by lower search costs that facilitate matching, but also by increased search complexity and congestion that hinders matching (Wheeler, 2001). We focus on the role of spatial scale and its effect of job displacement on two margins of labour adjustment: the wage margin that represents a heterogeneity effect and the employment margin that represents a quantity effect.

Another aim of our paper is to examine whether and how the returns to agglomeration differ by subgroup of workers. So far, the literature on LLMs (Manning and Petrongolo, 2017; Monte et al., 2018) and agglomeration economies (for an overview, see Combes and Gobillon (2015)) has been using pre-defined regional classifications to operate aggregate LLMs at different spatial scales, which could explain the mixed results when trying to answer the question whether agglomeration economies are gender-biased and skill-biased. We will define subgroups-specific LLMs and analyse three different ways in how the returns to agglomeration may differ among subgroups. First, the magnitude of the returns to agglomeration could differ among subgroups through differences in the ability to exploit the positive spillovers from agglomeration economies. The vast majority of the literature focuses on this mechanism, assessing gender differentials and education differentials in the agglomeration benefits for wages (e.g., see Phimister (2005), Di Addario and Patacchini (2008) and Rosenthal and Strange (2008)). Second, the role of regional aggregation in the returns to agglomeration might differ among subgroups. This holds if the prevalence of agglomeration economies depends on the worker's demographic characteristics and the worker's spatial scale of LLMs. An example would be that spillovers from agglomeration improve job matching of only high-educated workers and only at a relatively high spatial scale. Third, the returns to agglomeration could differ among subgroups of workers if there are differences in the structure of workers' LLMs. The differences in the opportunity costs of commuting through financial and time constraints suggest that workers vary in LLM size. Indeed, Farmer and Fotheringham (2011) show that the worker's LLM structure is endogenous to worker characteristics. We assess the structure of workers' LLM and its economic consequences by examining whether the benefits from agglomeration economies and the patterns over the level of regional aggregation differ among subgroups of workers.

Our empirical analysis is based on rich administrative linked employee-employer data sets that cover the period 2006 to 2014. We follow the literature by focusing on differences in workers' LLM through differences in workers' commuting flows from residence to workplace (e.g., see

Farmer and Fotheringham (2011); Brezzi et al. (2012)).[1] We use flowbca, which is a new flow-based cluster algorithm that can be used to define LLMs at various levels of regional aggregation (Meekes and Hassink, 2018).[2] The algorithm is flexible and able to define LLMs for different subgroups of workers. We endogenise workers' LLMs to commuting outcomes, gender and education level. The use of endogenous LLMs prevents the problem that holds for pre-defined regional classifications, which is that these are defined in line with administrative needs instead of economic relevance. To be able to compare with the majority of the studies on agglomeration economies, we apply the concept of endogenous LLMs to the analysis of static externalities of agglomeration.[3] Specifically, we analyse the impact of employment density on wages and employment using separate reduced-form regressions.[4] We estimate the returns to agglomeration using multiple sets of aggregate local labour markets (ALLM) and subgroup-specific local labour markets (SLLM) that vary in the level of regional aggregation. We also use pre-defined regional classifications to facilitate a comparison with the literature, including 398 Dutch municipalities, forty European Nomenclature of Territorial Units for Statistics (NUTS) 3 areas and thirty-five public employment services (PES) areas.

Our analysis provides two sets of novel results. First of all, we show that estimates of the UWP are increasing by a magnitude of two to three in the level of regional aggregation, for example by using forty Dutch NUTS 3 areas instead of the 398 Dutch municipalities. For empirical models in which employment density is operated by using fewer distinct LLMs, estimates of the returns to agglomeration in wages are higher. This pattern holds if we use the aggregate LLMs returned by flowbca (Meekes and Hassink, 2018). Hence, the results suggest that a large share of the benefits from agglomeration including improved matching, sharing and learning takes place at a relatively high spatial scale. In this regard, our results imply that the matching and sharing mechanisms are more important for agglomeration benefits than the learning mechanism, as the former mechanisms are considered to take place at a relatively high spatial scale Rosenthal and Strange (2001). In addition, we show that at a relatively high spatial scale, the loss in hourly wage for displaced workers who reside in relatively dense LLMs, compared to displaced workers who reside in more sparse LLMs, is significantly more modest. The returns to agglomeration in post-displacement wages are increasing by a magnitude of two in the level of regional aggregation. We find no returns to agglomeration in employment for workers who have been displaced. Our findings also suggest that for displaced workers, the positive returns to agglomeration allow them to become more selective in wages with a comparable probability of re-employment. We contribute

[1]Note that alternative ways to model differences in workers' LLM can be based on differences in job search behaviour by workers or employers, such as focusing on job-to-job flows (e.g., see Nimczik (2018)).

[2]So far, the literature has used distance-based clustering or concentration indices based on densities that are non-directional by nature (e.g., see Duranton and Overman (2005); Delgado et al. (2016)). We use a flow-based cluster algorithm to examine the structure of LLMs, which is directed by nature as the main input is directional data of commuting flows.

[3]More recent studies examine the dynamic externalities of agglomeration that may benefit local productivity (e.g., see De La Roca and Puga (2016) and Matano and Naticchioni (2016)).

[4]Note that we focus on urbanisation effects of employment density and area size, not on localisation effects of specialisation. We emphasise urbanisation economies as they play a more important role in the returns to agglomeration, and the use of variables to approximate localisation economies leads to more serious endogeneity concerns (Combes et al., 2008; Groot et al., 2014; Combes and Gobillon, 2015).

to the urban economics literature by focusing explicitly on the matching mechanism of agglomeration economies. Our findings are consistent with the geographical matching-function literature, which shows that market scale effects lead to higher wages but not to more rapid re-employment (Petrongolo and Pissarides, 2006).

Additionally, we show that the returns to agglomeration are more pressured downwards for workers who are characterised by relatively large LLMs, such as male and high-educated workers. We find that the UWP for high-educated workers compared to low-educated workers is about 100 per cent higher, holding the level of regional aggregation and thus the number of distinct LLMs constant. This finding is consistent with Manning (2003), as it suggests that workers who are characterised by thinner and larger LLMs benefit more from denser labour markets, because their wage offer curve is more steep. We find no systematic difference between the UWP of men and women. Importantly, the literature that focuses on subgroup differentials in the returns to agglomeration uses regional classifications that represent relatively large areas that do not differ among subgroups of workers. On the contrary, we show in our descriptive analysis that the structure of workers' LLM differs among subgroups. Specifically, the LLM of low-educated workers and female workers is relatively small. This descriptive finding is consistent with theoretical mechanisms that suggest that workers vary in the opportunity costs of commuting. In this regard, our results suggest that the returns to agglomeration are substantially overestimated for workers who are characterised by relatively small LLMs. Therefore, we contribute to the literature on subgroup differentials in agglomeration effects by examining the role of the endogenous structure of workers' LLM in economic outcomes.

2. Previous Research and Conceptual Setting

2.1. The Returns to Agglomeration

The seminal paper on the UWP is by Ciccone and Hall (1996), who focus on the static agglomeration effects on local productivity. For the U.S., Ciccone and Hall (1996) argue that the average labour productivity of a county increases by 6 per cent if the employment density doubles. Empirical urban economists devoted much attention to deal with individual-level endogeneity and local-level endogeneity, including the issues of skill-biased sorting across LLMs and endogenous local determinants that could bias the estimate of the UWP. Glaeser and Maré (2001) were the first to exploit micro data using individual-specific fixed effects to eliminate the potential of sorting more able workers into larger LLMs. Combes et al. (2008) introduce a two-step procedure to control for correlations between local-time unobservables and individual covariates. More recent studies take care of endogenous sorting across LLMs, using more advanced frameworks based on structural models (Gould, 2007; Baum-Snow and Pavan, 2012). See Rosenthal and Strange (2004) and Combes and Gobillon (2015) for comprehensive overviews.

An important question on the returns to agglomeration is to what extent the matching, sharing and learning mechanisms are relevant (Combes and Gobillon, 2015). We examine the importance of one specific mechanism: whether denser markets lead to better job matching between employers and workers in terms of employment and wage outcomes. Specifically, we assess the returns to agglomeration for workers who have been displaced, focusing on whether the externalities from denser LLMs affect the post-displacement employment probability and post-displacement wages.

We are not the first to examine the role of employment density and the matching mechanism in employment and wage outcomes. However, most studies ignore various selection problems that arise due to correlations among workers' exit rate into unemployment, cause of unemployment and location choice. For example, the intensity of job-to-job search and the probability of labour turnover is relatively low for workers with relatively high wages (Bhaskar et al., 2002) and low commutes (Crane, 1996).

The theoretical mechanisms underlying the role of agglomeration in post-displacement outcomes are ambiguous. A relatively dense market is characterised by low search frictions, more job opportunities and a high distribution of accepted wages. In turn, the higher number of workers and firms may lead to more productive job matching and lower market power of firms over employees (Manning, 2003). In addition, the literature suggests that workers in dense labour markets experience a greater degree of assortative matching, which leads to better job matches (Helsley and Strange, 1990; Wheeler, 2001; Duranton and Puga, 2004; Andersson et al., 2007). Thereby, displaced workers in dense markets may experience a relatively modest loss in post-displacement employment. Alternatively, a relatively high number of job opportunities may lead workers to become more selective in wages (Duranton and Puga, 2004; Berliant et al., 2006; Petrongolo and Pissarides, 2006). In turn, the positive spillovers from denser markets may not improve employment prospects, but instead increase workers' reservation wage leading to a more modest wage loss. It ultimately is an empirical question whether displaced workers who reside in denser markets experience more modest losses in employment and wages.

2.2. Gender Differentials and Education Differentials in the Returns to Agglomeration

The differences in agglomeration economies among subgroups of workers are driven by the theoretical micro-foundations that are internalised by the worker, not by the firm. For this reason, differences in agglomeration benefits among subgroups of workers are likely to be driven by improved job matching and learning of workers instead of improved sharing of risk and resources by firms. The gender differentials in the returns to agglomeration are theoretically ambiguous. The agglomeration benefits could be lower for female workers than for male workers, as women generally work closer to home due to a difference in opportunity costs of commuting through financial and time constraints (White, 1977; Madden, 1981; White, 1986; Hanson and Pratt, 1988; Zax, 1991; Clark et al., 2003; Fernandez and Su, 2004; Roberts et al., 2011). Therefore, on average, women have a smaller LLM. In turn, smaller LLMs are associated with a less productive matching of jobs (Petrongolo and Pissarides, 2006; Di Addario, 2011), leaving women in less productive jobs and lower wages. In addition, women are characterised by a lower labour market attachment and a lower number of working hours, which may reduce the ability to internalise knowledge spillovers. Alternatively, female workers may not experience hindrance from relatively small LLMs, as they have less concentrated labour market opportunities (Madden and Chiu, 1990). Moreover, female workers could experience higher returns to agglomeration, as the benefit of locating in relatively dense areas with more job opportunities and better job matching diminishes the negative effect of having a relatively low willingness to commute and thus a small LLM (Phimister, 2005).

The role of education level in the returns to agglomeration is also theoretically ambiguous. High-educated workers are likely to have higher returns to agglomeration, as larger markets are characterised by lower search costs. Lower search costs enable high-productive firms to conduct more efficient searches for high-productive workers, which generate better job matches and leaves low-educated workers in less productive jobs (Wheeler, 2001). Moreover, high-educated workers might be more able than low-educated workers to exploit knowledge spillovers, because they have the capacity to do so (Rosenthal and Strange, 2008). Alternatively, high-educated workers can have lower returns to agglomeration if larger markets are characterised by a strong amenity advantage for which high-educated workers have a higher marginal willingness to pay.[5] The higher willingness to pay will increase the labour supply of high-educated workers and decrease their wages (Black et al., 2009). Moreover, low-educated workers might be more able than high-educated workers to learn from human capital in close proximity, as they have a greater potential left to benefit (Rosenthal and Strange, 2008). In the literature, the results are mixed. Several studies find that the returns to population density increase with the number of years of education (Wheeler, 2001; Rosenthal and Strange, 2008; Bacolod et al., 2009; Carlsen et al., 2016). Gould (2007) argues that white-collar workers benefit more from working in larger LLMs than blue-collar workers. Other studies find that the returns to population fall with education level (Adamson et al., 2004; Di Addario and Patacchini, 2008; Black et al., 2009; Lee, 2010).

2.3. Regional Aggregation and the Relevance of Endogenous Local Labour Markets

For economic applications in the field of economic geography, the spatial scale of LLMs is operated by using a specific regional classification. So far, most studies operate LLMs by using pre-defined "exogenous" regional classifications. Examples of exogenous regional classifications include the U.S. counties Hoynes (2000), the U.S. Standard Metropolitan Statistical Area (MSA) (Glaeser and Maré, 2001), the U.S. commuting zone (CZ) (Autor et al., 2013), European NUTS areas (Ciccone, 2002), and UK Travel-to-Work-Area (TTWA) (Petrongolo and Pissarides, 2006; Manning, 2009; Faggio et al., 2016). Although many studies in the field of economic geography use various regional classifications, it is surprising that they generally neglect the role of regional aggregation in the returns to agglomeration. A notable exception is the paper by Briant et al. (2010), who focus on the Modifiable Areal Unit Problem (MAUP). They examine the importance of the size and shape of regional areas for estimates of, among others, the UWP. The authors use six regional classifications and argue that specification issues are the most important to limit biases, and the heterogeneity in the shape of LLMs is of higher-order importance. Interestingly, the authors show that heterogeneity in the size of LLMs is more relevant for the returns to agglomeration than the shape of LLMs: the use of small zoning systems instead of large zoning systems lowers the estimate of the UWP by a magnitude of two. Other studies also reveal stronger agglomeration spillovers when a regional classification at a relatively high level of regional aggregation is used to operate LLMs (e.g., see Ellison and Glaeser (1997) and Rosenthal and Strange (2001)).

The advantage of using a pre-defined exogenous regional classification is that within-country differences in economic outcomes can easily be investigated while research outcomes remain comparable across studies and through time. However, the validity of using exogenous regional classifications is questionable for two reasons. First, the exogenous regional classifications are based on

[5]Note that the amenity advantage from denser markets especially holds for European cities (Brueckner et al., 1999). In several U.S. metropolitan areas, disamenities from denser areas, e.g., crime and pollution, may outweigh positive amenities.

old definitions and are potentially outdated. Over the last decades, the labour force composition experienced changes caused by technological, demographic and economic shifts on a global level. These shifts led to a decrease in the demand for low-educated workers and greater wage dispersion (Fernandez, 2001; Goos et al., 2009, 2014), a rise in women's labour force participation (Costa, 2000) and an increase in commuting from home to workplace (Van der Laan and Schalke, 2001; Crane, 2007). Second, the exogenous regional classifications represent non-overlapping areas in the sense that they only vary between areas and not among individuals within areas. A recent paper on the effective size of LLMs is by Manning and Petrongolo (2017), who aim to limit mismeasurement of UK workers' LLM by using a continuous nature of geographic space that allows for overlapping LLMs of two workers who reside in an administratively different but geographically close location. They develop a job search across space framework, in which the worker's size of the LLM depends on the cost of distance. Importantly, Manning and Petrongolo (2017) do not distinguish between different types of workers and model low-skill labour markets that tend to be relatively local, as their data sample contains workers with a relatively low education level.

We define LLMs endogenous to demographic characteristics and commuting outcomes, which allows us to operate workers' LLM at various levels of regional aggregation for different subgroups of workers. By allowing for differences in workers' LLM when they meet different characteristics, we provide an alternative view of overlapping LLMs. We argue that endogenising workers' LLM to worker characteristics is relevant, as theory suggests that the LLM structure is worker-specific. The LLM structure differs among workers, as workers face differences in utility functions and opportunity costs. An example would be that women put a higher value on working closer to home than men do (White, 1977; Roberts et al., 2011). The gender difference in valuation of leisure and commuting can be explained by various reasons (Fernandez and Su, 2004), including women's dual role as a mother and worker, involving responsibilities for the household and children (Madden, 1981; Crane, 2007). Alternatively, more educated workers have thinner labour markets and are therefore characterised by a higher commuting distance (Manning, 2003).[6] Workers in thinner and larger labour markets are likely to benefit more in terms of wages from longer commuting, as the wage offer curve is raised when the skill-level rises (White, 1988). Consequently, high-educated workers face lower opportunity costs of commuting and work further away from home. Both examples suggest that the LLM structure varies in different subgroups of workers. In turn, differences in the LLM structure among workers may lead to differences in the returns to agglomeration.

2.4. Conceptual Setting

The model shown in (1) is specified to display the implications of the level of regional aggregation for estimates of the returns to agglomeration.

$$\text{Returns to Agglomeration} = \alpha + \beta \times \text{Regional Aggregation} \tag{1}$$

[6] In contrast, other literature argues that high-skilled workers have a relatively high value of time and are therefore characterised by shorter commutes (e.g., see Brueckner et al. (2002)). In the context of the Netherlands, we argue that the mechanism of thinness on the labour market outweighs the mechanism of a higher valuation of time. This observation is consistent with the empirical finding that commutes and education are positively correlated.

The parameter α represents the baseline returns to agglomeration. The role of the regional aggregation in the returns to agglomeration is represented by β. The parameter β equals zero if estimates of the returns to agglomeration do not depend on the level of regional aggregation. The vast majority of the literature that examines the returns to agglomeration focuses on the estimation of α and implicitly assumes that β equals zero. Also, the literature examines whether there are gender differentials and education differentials in the level of the returns to agglomeration α. Notably, the literature generally neglects the role of β in the returns to agglomeration. We hypothesise that β is not equal to zero. For example, estimates of the returns to agglomeration could be increasing in the level of regional aggregation, if agglomeration economies are more prevalent at a higher spatial scale. We contribute to the literature by examining the role of regional aggregation in the returns to agglomeration, by examining whether the role of regional aggregation differs among subgroups of workers, and by examining whether subgroups of workers are characterised by a different level of regional aggregation to represent the structure of their LLM.

3. Background, Data and Flowbca

3.1. Background on the Dutch Regional Classifications

In the Netherlands, the COROP regional classification, defined in 1971, was set out to identify economically and socially integrated areas (CBS, 2018). COROP literally stands for the Coordination Commission Regional Research Programme (in Dutch: Coördinatiecommissie Regionaal Onderzoeksprogramma). The COROP classification is equivalent to the European concept of NUTS 3 areas and comparable to the U.S. concept of Commuting Zones and the UK concept of Travel-To-Work-Areas. The COROP areas (hereafter: NUTS 3 areas) were defined based on journey-to-work and place-of-work statistics that reflected the typical commuting outcomes of Dutch employed workers. In total, there are forty NUTS 3 areas: each NUTS 3 area consists of a core and hinterland area, while the provincial borders are never crossed. Besides administrative units such as provinces and municipalities, Statistics Netherlands and the Dutch government use the forty NUTS 3 areas and the thirty-five public employment services (PES) areas for analytical and political purposes. We use the 398 Dutch municipalities, forty NUTS 3 areas and thirty-five PES areas as reference sets of LLMs, which facilitate a comparison to the alternative sets of defined aggregate LLMs and subgroup-specific LLMs.

3.2. Data Sets

We used various administrative micro data sets, retrieved from Statistics Netherlands, covering the period of 2006 to 2014. The micro data sets contain data of individuals, households and firms. The data set Work Location Register (*Gemstplbus*) was used to incorporate data on the geographical employment location of employees at the municipality level. We used a set of 403 distinct Dutch municipalities that existed in 2014. For the sake of convenience, we removed five municipalities that represent the small and isolated Wadden Islands in the northern part of the Netherlands. The work location is observed annually in December. The Population Register (*Gbapersoontab*, *Gbahuishoudensbus*, *Gbaburgerlijkestaatbus*, *Gbaadresgebeurtenisbus*), which is based on municipal and tax office administration, was used to incorporate data on individuals' date of birth, gender, marital status, number of household members and changing home. We removed observations of workers who were aged below 18 or over 65 years. The Address Object Register (*Gbaadresobjectbus*, *Vslgwbtab*) was used to incorporate data on individuals' home address and location at the municipality level. The Job and Wages Register (*Polisbus*), which is based on income statements of employees to the tax office administration, was used to incorporate data on the type of job (full-time or part-time), type of contract (fixed or temporary), economic sector, number of hours worked and gross wage. We removed observations of workers who were employed less than 0.8 full-time equivalent or 128 hours a month, to make the labour market outcomes of workers who differ in especially gender more comparable. Moreover, we removed observations of workers who earned an hourly wage lower than 3 euro. The Main Job Register (*Hfdbaanbus)* was used to select the main job of the worker, which is the job with the highest annual wage. The Bankruptcy Job Endings Register (*Failontslagtab)* was used to incorporate data on the worker, firm and date of workers' job displacement due to firm bankruptcy. The Highest Education Register (*Hoogsteopltab*) was used to incorporate data on workers' highest level of attained education. The highest level of attained education contains three groups, i.e. low, average and high educational attainment. This categorisation is based on the International Standard Classification of Education (ISCED) and corresponds to lower, secondary and tertiary education, respectively.

3.3. Key Variables and Covariates

The key dependent variables include hourly wage and employment. The worker's hourly wage was constructed by taking the natural logarithm of the monthly contractual gross wage relative to the number of contractual hours worked per month. Note that for the urban wage premium annual data set, we constructed workers' hourly wage of the month of December. Thereby, the hourly wage and commuting distance were constructed based on data of the same job of the month of December. The job displacement data set contains monthly data. The worker's employment status was represented by a zero-one indicator variable that equals 1 if the worker is employed, and zero otherwise. The key independent variables can be divided into two sets.

The first set of key independent variables was used to construct the aggregate LLMs and subgroup-specific LLMs, containing a cross-section of commuting flows across municipalities in the year 2014.[7] This set of variables was used for the descriptive analysis. For convenience, we used the cross-section of flows in the year 2014, as the number of distinct municipalities decreased in the period 2006 to 2014. Unfortunately, we were not allowed by Statistics Netherlands to export commuting flows across municipalities between 1 to 9 workers. These flows represent about 3 per cent of the total number of flows and were omitted. Aggregate LLMs were defined based on a set of commuting flows across municipalities of all workers together.[8] The subgroup-specific

[7]We examined the temporal changes in the sets of commuting flows over the period 2006 to 2014, which were relatively small. For the sake of convenience, we use time-invariant LLMs.

[8]Unfortunately, the work location is not entirely consistent as Statistics Netherlands has only data on the number of firm plants, the plant locations and the number of employees at each specific plant. The individual's work location is imputed by Statistics Netherlands using data on the place of work and place of home. Each individual is, based on the home location, linked to the closest firm plant, conditional on not exceeding the number of individuals who were employed at that specific plant. Hence, the amount of commuting interaction between municipalities is likely

LLMs were defined using separate sets of commuting flows for workers who differ in gender and education.

The second set was used to approximate agglomeration spillovers and consists of variables that represent the natural logarithm of employment density and the natural logarithm of area size. This set of variables was used for the empirical analysis. Workers' employment density was constructed by taking the number of employed workers in the LLM relative to the area size in kilometres of the LLM. Various regional classifications were used to represent the worker's LLM, including the Dutch municipalities, NUTS 3 areas, PES areas, aggregate LLMs and subgroup-specific LLMs. For a given worker, each regional classification gives different values of the employment density and area size. For a specific number of distinct aggregate LLMs, the employment density and area size differ between the LLMs, but not between workers who reside in the same LLM.[9] For a specific number of distinct subgroup-specific LLMs, the employment density and area size may differ between workers if they reside in the same LLM but vary in gender or education level.

A set of covariates that was used for the empirical analysis contains zero-one indicator variables that represent female, highest attained education (low, average and high education), Dutch nationality, age (18-25, 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60 and 60-65 years), having children aged 18 or lower, having a partner, number of household members (1, 2, 3-4 and more than 4 members), economic sector of the firm (66 categories), the size of the firm (1-9 employees, 10-49 employees, 50-99 employees, 100-499 employees and more than 499 employees), job tenure (3-6, 6-12, 12-18 and over 18 years) and year of job displacement (2007, 2008, 2009, 2010 and 2011). Note that the variables job tenure and displacement year are only used in the empirical analyses on the returns to agglomeration for workers who have been displaced.

3.4. Flow-Based Cluster Algorithm

We use flowbca, discussed by Meekes and Hassink (2018) and extending earlier work by Duranton (2015), which is a flow-based agglomerative hierarchical cluster algorithm that is able to cluster regional units into meaningful self-contained LLMs. Flowbca is flexible and able to define LLMs for different subgroups of workers at various levels of regional aggregation. From a theoretical point of view, the functional criterion to pair two regional units into one depends on the level of interaction. In our analysis, the level of interaction between regional units is approximated by relative commuting flows from residence to workplace. The main input for the algorithm is a set of commuting flows across municipalities. Alternative sets of aggregate LLMs were constructed at a number of distinct LLMs between 398 and 7. A higher level of regional aggregation leads to fewer distinct LLMs. Subgroup-specific LLMs were defined by separately using commuting flows

to be underestimated, in particular for subgroups who are characterised by relatively large LLMs. Consequently, the variation between subgroups in the size of the LLM is also likely to be underestimated.

[9] For the year 2014, the minimum, maximum, median and average number of employed workers in a NUTS 3 area (40 distinct units) equals 13,960, 753,749, 109,372, and 170,759 workers, respectively. For the year 2014, the minimum, maximum, median and average number of employed workers in a PES area (35 distinct units) equals 52,194, 722,819, 141,689, and 195,153 workers, respectively. For the year 2014, the minimum, maximum, median and average number of employed workers living in an aggregate LLM (35 distinct units) defined by our algorithm equals 9,452, 1,576,821, 77,836 and 194,855 workers, respectively.

of subgroups of workers, which include groups of both female workers and male workers, varying in three education levels.

The algorithm that we used to define LLMs can be described as follows. LLMs were defined by iteratively aggregating two regional units into one. A higher number of iterations implies a higher level of regional aggregation. In each iteration, the algorithm selects two units that will be aggregated based on an optimisation function. The optimisation function identifies the maximum relative commuting flow out of all bilateral commuting flows. The "source" unit from which the largest relative commuting flow starts is aggregated to the "destination" unit. This process is repeated until a stopping criterion is met. Examples of stopping criterions are if a specific number of distinct LLMs has been defined or if a specific average level of self-containment has been met. After the algorithm is terminated, the level of self-containment of an LLM is approximated by its population weighted local employment rate. The population weighted local employment rate is constructed by dividing the total number of workers who work local i.e. those who work and live in their LLM by the total number of employed workers. A higher local employment rate implies a stronger connectivity within the LLM and a weaker connectivity to outside LLMs. See the paper by Meekes and Hassink (2018) for more information about flowbca.

Two limitations of the algorithm require additional attention. First, Dutch municipalities that are relatively large in terms of population, for example Amsterdam and Utrecht, dominate the absolute number of outgoing flows to other municipalities. This leads to a situation where the algorithm aggregates the larger units to smaller destination units. Consequently, the larger regional units will not be defined as the core of an LLM. To overcome this limitation, we used relative flows that function as weights to account for the relative importance of a unit. Second, if the interaction based on commuting flows across regional units is not sparse enough, the algorithm defines several relatively large LLMs and many small isolated LLMs. For a given LLM, this limitation could lead to the situation where multiple municipalities in the LLM are hardly connected to one another. Figure 7 shows that flowbca is able to define meaningful LLMs at various levels of regional aggregation.

4. Descriptive Results

We apply flowbca to define LLMs for various subgroups at different levels of regional aggregation. Commuting flows across municipalities are used as the main input for the algorithm to define LLMs. In the first part of the descriptive results, we document the changes in commute over the last decades and we explain why we focus on the six subgroups that vary in gender and education. In the second part of the descriptive results, we show the application of flowbca. We document to what extent the level of self-containment of a set of LLMs depends on the level of regional aggregation. Moreover, we visualise LLMs for workers who vary in gender or education level.

4.1. Commuting

We provide Table 1 to provide a better understanding of which worker characteristics explain the largest share of variation in workers' commuting distance. Table 1 displays the quantile regressions of commuting distance in kilometres on various worker characteristics. The 0.05, 0.25, 0.50, 0.75 and 0.95 quantile regression are provided in Columns (1) to (5), respectively.

Table 1
Quantile regressions of commuting distance on worker characteristics.

	Commuting distance (km)				
	q05	q25	q50	q75	q95
	(1)	(2)	(3)	(4)	(5)
FEMALE	-0.0614***	-0.3460***	-1.2343***	-2.8691***	-6.2561***
	(0.0066)	(0.0131)	(0.0309)	(0.0569)	(0.2389)
AVERAGE-EDUCATED	0.1264***	0.5218***	1.2518***	2.3795***	4.7847***
	(0.0095)	(0.0183)	(0.0407)	(0.0758)	(0.3389)
HIGH-EDUCATED	0.2982***	1.0884***	3.5444***	7.6499***	13.6635***
	(0.0101)	(0.0166)	(0.0554)	(0.0925)	(0.4143)
25 < *AGE* ≤ 30 years	0.0345***	0.1091***	0.5168***	1.2896***	3.0203***
	(0.0117)	(0.0147)	(0.0316)	(0.0917)	(0.3867)
30 < *AGE* ≤ 35 years	0.0998***	0.2356***	0.8445***	1.8653***	3.6084***
	(0.0095)	(0.0180)	(0.0481)	(0.0692)	(0.3601)
35 < *AGE* ≤ 40 years	0.1556***	0.4728***	1.2471***	2.2744***	4.5206***
	(0.0112)	(0.0240)	(0.0440)	(0.1023)	(0.4297)
40 < *AGE* ≤ 45 years	0.1472***	0.4714***	1.2491***	2.3537***	5.6296***
	(0.0151)	(0.0232)	(0.0623)	(0.0748)	(0.4839)
45 < *AGE* ≤ 50 years	0.1233***	0.3591***	0.9420***	1.7746***	4.6354***
	(0.0146)	(0.0215)	(0.0456)	(0.1055)	(0.4929)
50 < *AGE* ≤ 55 years	0.1078***	0.3371***	0.6551***	1.3828***	4.7094***
	(0.0199)	(0.0280)	(0.0678)	(0.1200)	(0.6200)
55 < *AGE* ≤ 60 years	0.1094***	0.2407***	0.4272***	1.1703***	3.9197***
	(0.0165)	(0.0245)	(0.0480)	(0.1123)	(0.4151)
60 < *AGE* ≤ 65 years	0.0637**	0.0758	0.1959*	0.7882***	5.2194***
	(0.0294)	(0.0512)	(0.1063)	(0.2138)	(0.9823)
DUTCH NATIONALITY	0.0934***	0.0714***	-0.0281	-0.1698***	-1.8535***
	(0.0108)	(0.0191)	(0.0438)	(0.0628)	(0.2567)
NO CHILDREN	-0.0062	0.0871***	0.1331***	0.3740***	1.5561***
	(0.0094)	(0.0163)	(0.0322)	(0.0550)	(0.3455)
PARTNER	0.0500***	0.2213***	0.3664***	0.2964***	-0.8860***
	(0.0083)	(0.0144)	(0.0352)	(0.0706)	(0.2888)
Number of observations	946,043	946,043	946,043	946,043	946,043

Notes: The dependent variable is the commuting distance measured in kilometres. Parameter estimates of the covariates are reported. Bootstrapped standard errors are in parentheses. ***,**,*, correspond to the significance level of 1%, 5%, 10%, respectively. The reference categories of *FEMALE*, *EDUCATED*, *AGE*, *NATIONALITY*, *NO CHILDREN*, *PARTNER*, consist of workers who are male, low-educated, aged between 20 and 25, have a non-Dutch nationality, children and no partner, respectively. The quantile regression analyses include indicator variables for the number of household members (3), firm economic sector (66), firm size (4), the NUTS 3 location of the household (39) and the calendar year (8). The period under observation is from 2006 to 2014. Data set: the administrative data from Statistics Netherlands. Sample: a five per cent random sample.

Table 1 shows that female workers and low-educated workers are characterised by a relatively short commuting distance. Moreover, Table 1 reveals that the estimates for gender and education, compared to other worker characteristics, are relatively economically significant. This observation holds in particular for the regressions of the 75th percentile and above. Also, the differences among the commuting quantiles is highest for gender and education. The difference in commuting outcomes among subgroups of workers suggest that subgroups are characterised by a different LLM structure. We particularly focus on gender- and education subgroups, because these worker characteristics explain the largest share of variation in commuting outcomes.

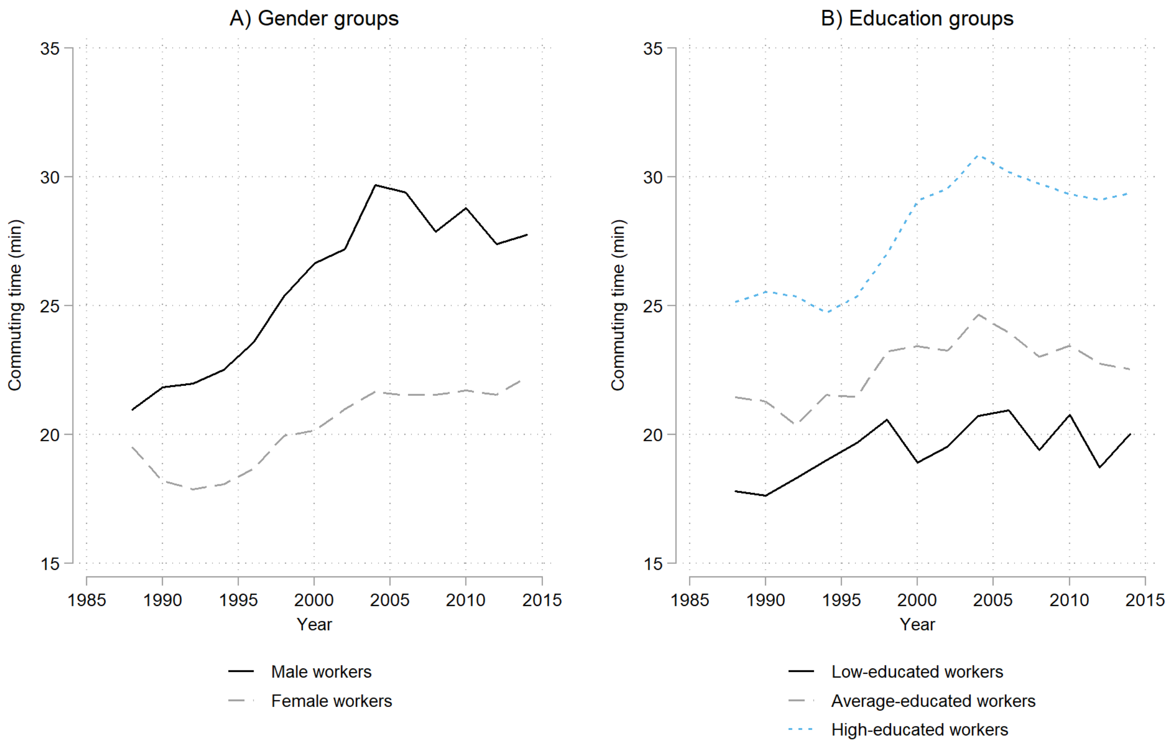


Fig. 1. Changes in the average commuting time of workers by gender and education groups over the period 1988 to 2014. *Notes:* Data set: the SCP labour supply panel. Sample size: 41,275 observations.

Figure 1 is the only figure in this paper that is not based on data retrieved from Statistics Netherlands. We use data from the Dutch SCP labour supply panel (in Dutch: *SCP Arbeidsaanbodpanel*) to observe differences in commuting over the last decades (SCP, 2015). Figure 1 shows that for men and women the average commuting time increased in the period from 1988 to 2014. The increase in commuting time is most severe for high-educated workers. Moreover, Figure 1 shows that workers' commuting time from place of residence to place of work differs among subgroups. Men, compared to women, and high-educated workers, compared to low-educated workers, commute longer. The change in commute over the last decades indicates that regional classifications that have been defined a long time ago, for example the NUTS 3 areas, might be outdated. Moreover, the findings suggest that workers' LLM has become larger over the last decades.

Figure 2 shows the density plots of the gender shares (Fig. 2A) and education shares (Fig. 2B) across 398 municipalities. The shares are separately given for employed individuals in their home municipality and work municipality. Figure 2A provides us with several insights. First, there are on average more men than women in the sample. This observation can be explained by the fact that there are more men employed than women. Second, for both men and women, the distribution of workers is much wider than the distribution of residents. A wider distribution suggests higher concentration ratios in specific municipalities. Male and female workers are relatively concentrated in specific municipalities, but male and female residents are more evenly distributed across municipalities. This observation suggests that there exists substantial regional mismatch between the home location and employment location of both male and female workers.

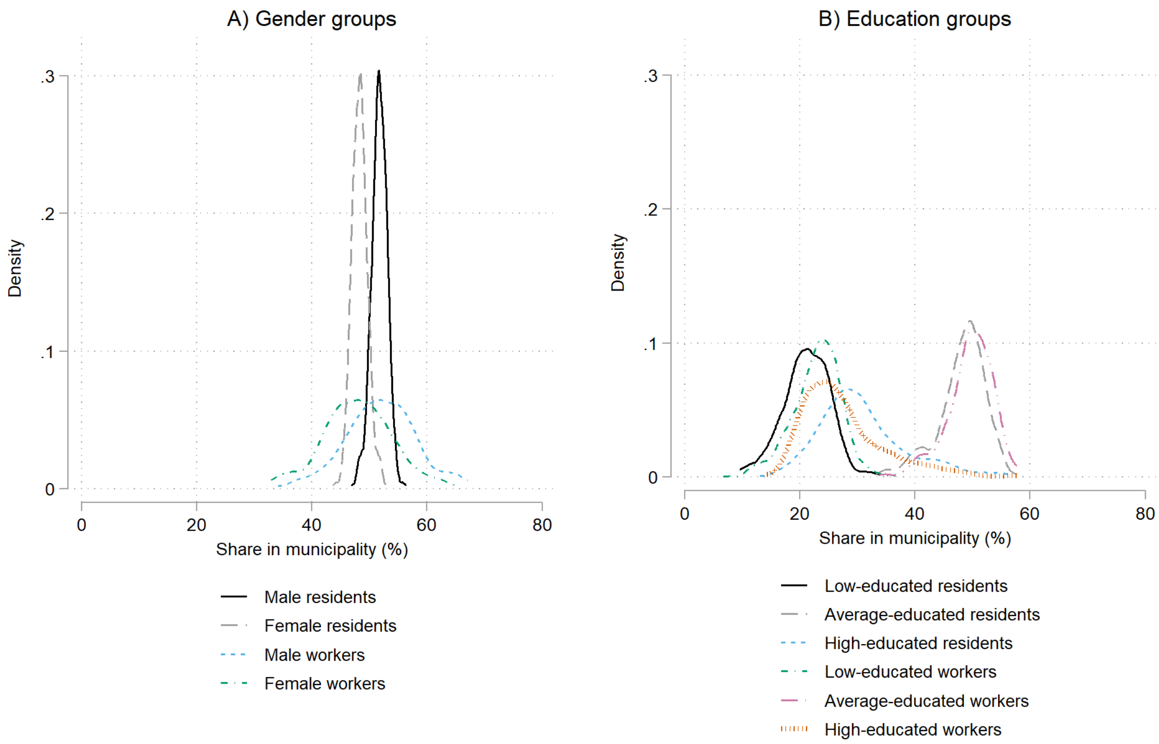


Fig. 2. Distribution plot of gender and education shares across municipalities. *Notes:* The gender and education shares are constructed by taking the subgroup-specific fraction, separately for residents and workers, in the municipality. The sample contains fractions for 398 distinct municipalities. Data set: the administrative data from Statistics Netherlands.

The distribution of high-educated workers is relatively wide (see Fig. 2B), which implies that high-educated workers are more concentrated in specific municipalities than low-educated workers. Moreover, Figure 2B reveals that the distributions do not differ between residents and workers who belong to the identical education group. Hence, there is not much education-biased regional mismatch between home and employment locations. However, the differences in the concentration ratios between education categories suggest that there is substantial education-biased sorting across municipalities.

Overall, in this subsection, we have shown that workers' gender and education explain the largest share of variation in commuting distance. Moreover, we have shown that workers' commuting time has been increasing over the last decades. This finding underscores the relevance of defining LLMs with more recent data on commuting flows. In addition, the results suggest that there is substantial regional mismatch between workers' residence and work location for both women and men. Also, the results indicate substantial education-biased sorting of workers across regional areas. Our descriptive results motivate the use of subgroup-specific LLMs for workers who differ in gender and education, as these demographic characteristics are the most important for commuting outcomes.

4.2. Endogenous Local Labour Markets

In this subsection, we show how the aggregate LLMs and subgroup-specific LLMs, defined with flowbca, vary in the local employment rate. The local employment rate refers to the relative number of workers who live and work in their LLM. Moreover, we visualise LLMs of men and women separated by three education levels.

Figure 3 shows the maximum relative commuting flow in each iteration of flowbca. The algorithm that we used to define LLMs iteratively aggregates a regional unit to another regional unit, based on the maximum relative commuting flow out of all bilateral flows. The starting set of units contains 398 distinct municipalities. After each iteration of the algorithm, the number of distinct LLMs (K) decreases by one. Figure 3 shows that the relative commuting flow at which units are aggregated is decreasing in the number of iterations. This observation holds as with fewer distinct LLMs there is more connectivity within a given LLM and less connectivity to outside LLMs. However, observe that the relative commuting flow at which units are aggregated is not uniformly decreasing in the number of iterations. This observation can be explained by the following example. Consider three regional units: A, B and C. Unit C has a relative flow of about 25 per cent to unit A and also to unit B. However, unit A is aggregated to unit B as the relative flow from A to B, which is the maximum of all relative flows, equals 30 per cent. After A has been aggregated to unit B, unit C will be aggregated to the combination of A and B, as C has a relative flow of 50 per cent to the new LLM that consists of A and B together.

Figure 4 shows the maximum relative commuting flow at which regional units were aggregated to construct the subgroup-specific LLMs for each of the six subgroups. Two observations are in place. First, when aggregating from about 10 to 100 distinct LLMs, women are characterised by a lower relative commuting flow than men. This observation suggests that women work closer to home than men. For a higher number of distinct LLMs, this distinction is less obvious. Second, high-educated workers have generally higher values of the relative commuting flows at which regional units are aggregated. This observation suggests that high-educated workers, compared to low-educated workers, work more often outside their LLM. Figure 4 suggests that the extent to which a regional classification reflects workers' LLM strongly depends on the worker's gender and education.

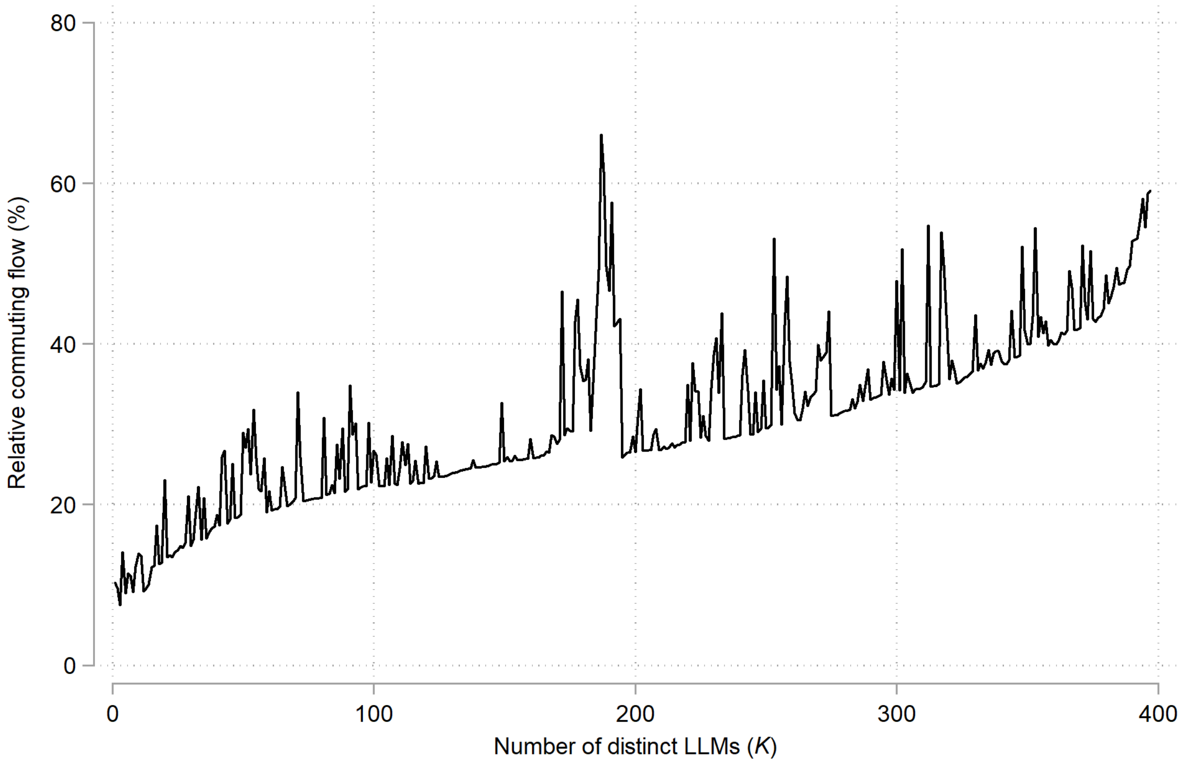


Fig. 3. Relative commuting flow at which two units are aggregated. *Notes:* In each iteration, starting from a set of 398 distinct municipalities, the cluster algorithm selects the regional unit with the highest relative flow and aggregates the source unit to the receiving destination unit. The relative commuting flows are computed by taking each absolute commuting flow from source unit to destination unit relative to the source unit's total of absolute outgoing flows. In total 7,291,815 commuting flows were used. Data set: the administrative data from Statistics Netherlands.

Figure 5 shows the population weighted average local employment expressed as a percentage, based on the aggregate LLMs, NUTS 3 areas and PES areas. The local employment rate represents the rate at which workers live and work in the same LLM. For the aggregate LLMs, the local employment rate varies over the number of distinct LLMs. Figure 5 shows that local employment decreases in the number of distinct LLMs. This is not surprising, as after two units are aggregated the workers who commute between the two aggregated units will work locally. It is important to note that the local employment rate of the aggregate LLMs is much higher, with an identical number of distinct LLMs, than that of the forty NUTS 3 areas and thirty-five PES areas. This observation suggests that the algorithm that is used to cluster regional units does relatively well in constructing self-contained regional areas of residence and work activity.

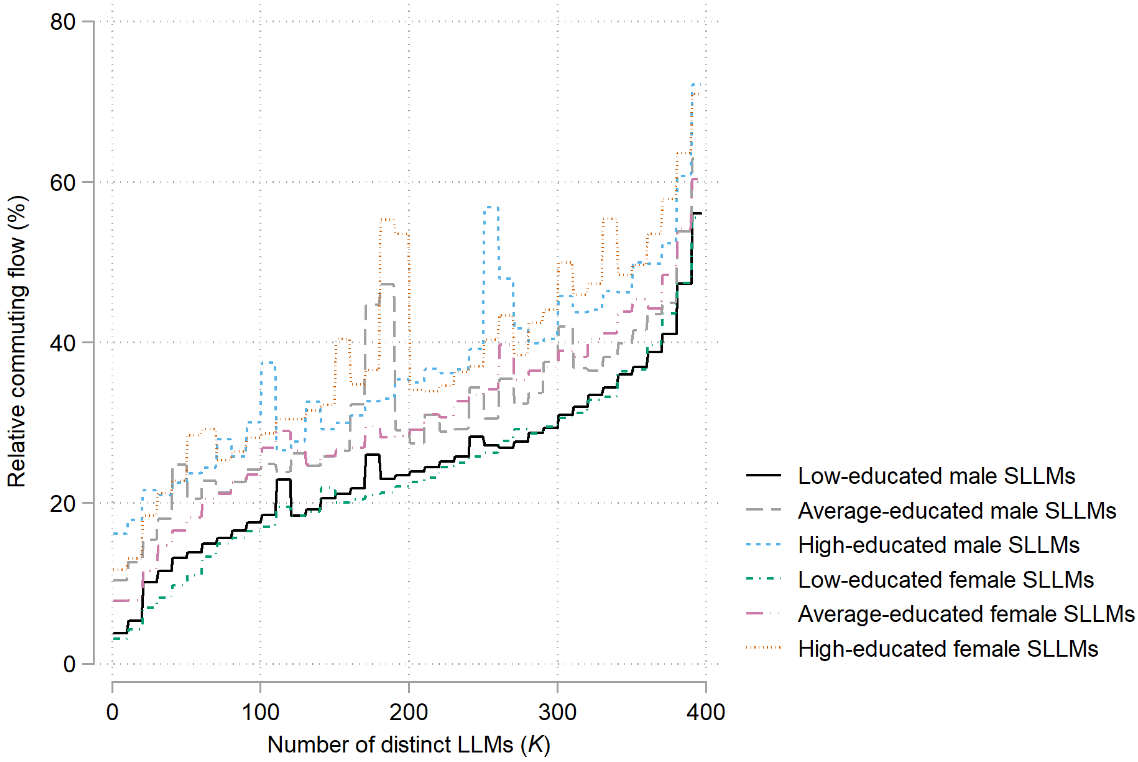


Fig. 4. Subgroup-specific relative commuting flow at which two units are aggregated. *Notes:* The median of the relative commuting flows, in increments of ten, is given to smooth out the lines and to provide visible patterns. The values of the relative commuting flow, in each iteration, are available upon request. See Figure 3 for additional notes.

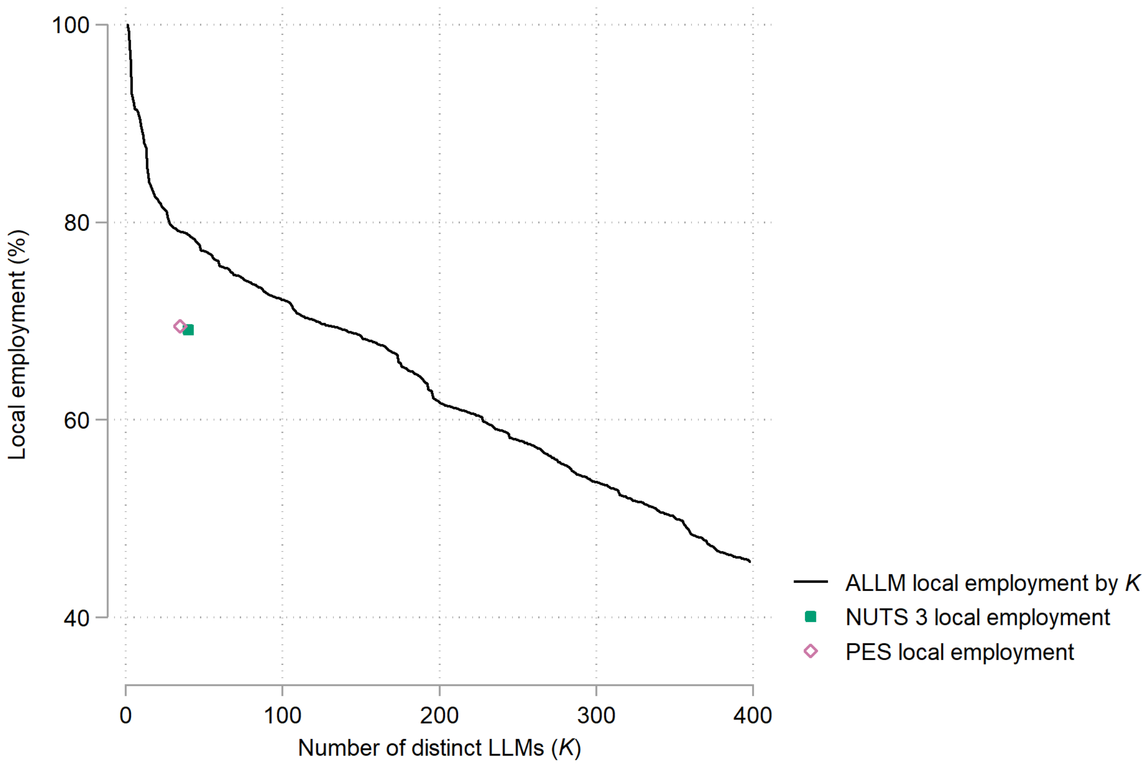


Fig. 5. Local employment by regional classification. *Notes:* The local employment rate is constructed by taking the number of workers who live and work in their LLM relative to the total number of workers. See Figure 3 for additional notes.

Figure 6 reveals the extent to which the local employment varies over the number of distinct subgroup-specific LLMs. Both male workers and high-educated workers are characterised by lower local employment compared to female and low-educated workers, respectively. This observation suggests that male and high-educated workers are characterised by a relatively high commuting distance and a large LLM, which is consistent with the results in Table 1.

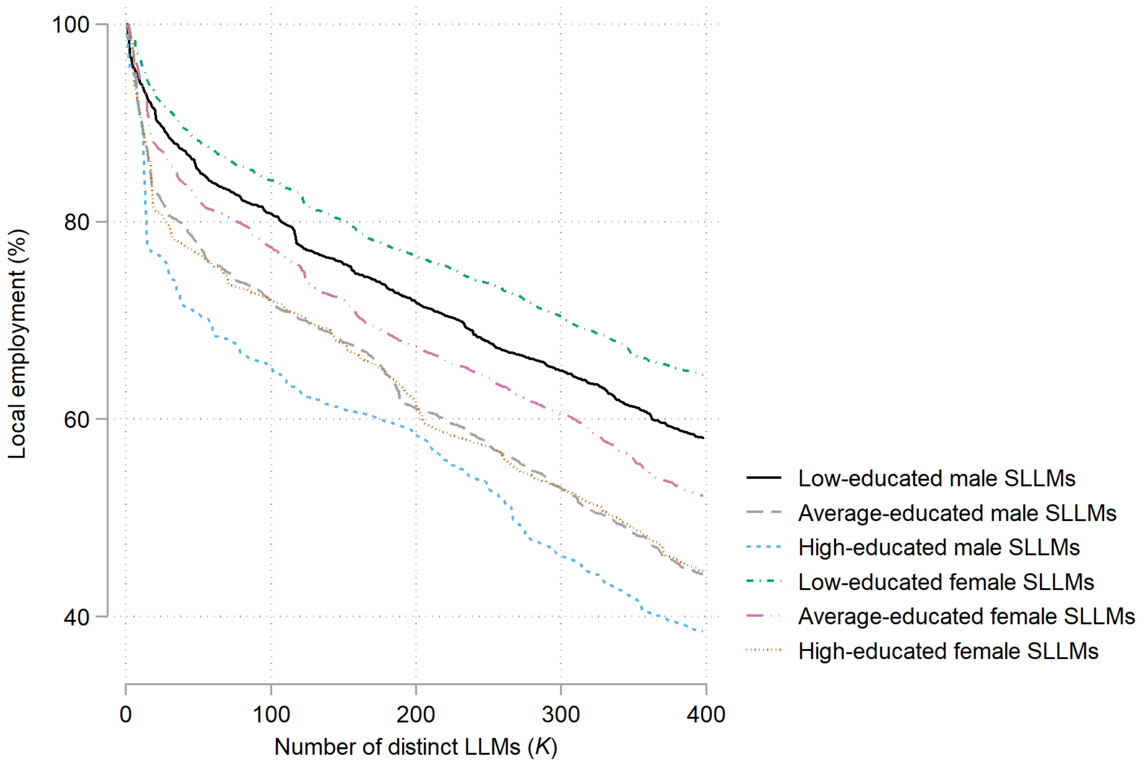
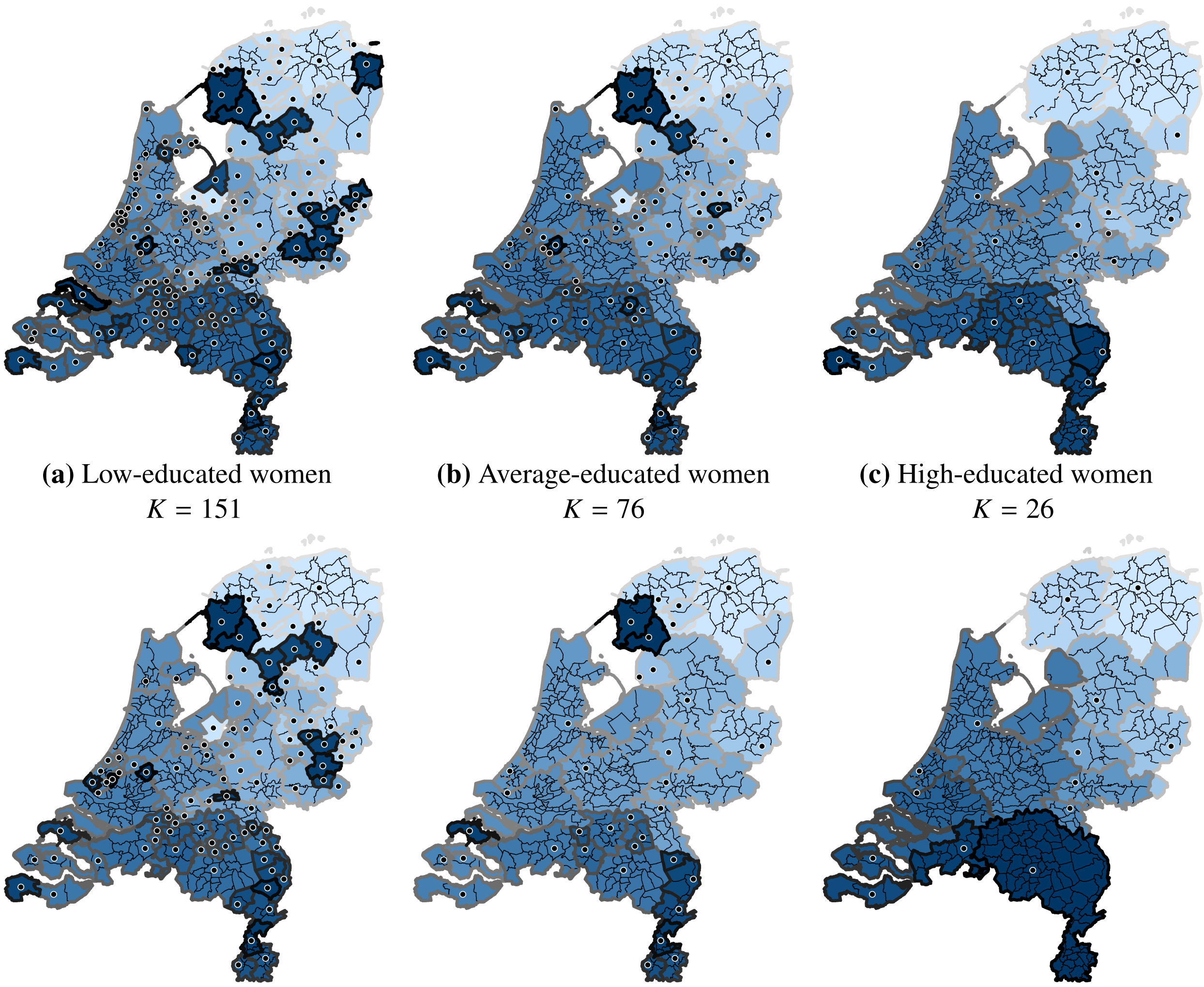


Fig. 6. Subgroup-specific local employment by regional classification. *Notes:* Local employment by subgroups and regional classifications. See Figure 5 for additional notes.

Figure 7 visualises the LLMs of male and female workers separated by the three educational groups. The stopping criterion of the algorithm was set equal to a local employment rate of 80 per cent. That is, if 80 per cent of the workers live and work in their LLM, the algorithm is terminated. The stopping criterion of local employment equal to 80 per cent is arbitrary. Importantly, the differences in LLMs between subgroups of the population also hold for stopping criteria with other levels of local employment. Figure 7 shows that the number of distinct LLMs is decreasing in the education of workers and is lower for men. In this regard, high-educated workers and male workers are characterised by an LLM that is relatively less self-contained. This suggests that exogenous regional classifications are generally too large for low-educated and female workers, but too small for high-educated and male workers. Significantly, Figure 7 suggests that workers' education is more important for the LLM structure than workers' gender, as differences in the structure of LLMs are more pronounced between education levels.



(a) Low-educated women
K = 151

(b) Average-educated women
K = 76

(c) High-educated women
K = 26

(d) Low-educated men
K = 107

(e) Average-educated men
K = 36

(f) High-educated men
K = 14

Fig. 7. Subgroup-specific local labour markets. *Notes:* The stopping criterion of the cluster algorithm is set to a minimum local employment rate of 80 per cent. The number of distinct LLMs is represented by K. The LLMs and its cores (the black dots with a white circle) are returned by flowbca. Each distinct LLM is surrounded by a thick border and highlighted by a colour. Data set: the administrative data from Statistics Netherlands.

Overall, our findings are relevant for research that focuses on quantifying regional differences in economic outcomes, as they suggest that the mismeasurement in workers' LLM strongly depends on the characteristics of the data sample. For example, the magnitude of mismeasurement in workers' LLM is very different for a data sample of women compared to a sample of men. The descriptive results in this subsection show that the extent to which a regional classification reflects a worker's LLM strongly depends on the worker's geographical location, gender and education. For this reason, we focus on the roles of aggregate and subgroup-specific LLMs in the returns to agglomeration in wages and employment. Moreover, we focus on gender- and education differentials in the agglomeration externalities.

5. Methodology

In this section, we provide the empirical models that we use for the two economic applications discussed in the empirical analyses. The two economic applications we consider include the estimation of the UWP and the impact of job displacement. After introducing the empirical models, we will discuss the main identification challenges that required our particular attention.

5.1. Urban Wage Premium

An empirical model, shown in (2), is specified to estimate the returns to agglomeration. The dependent variable is the natural logarithm of the hourly wage and the model is given as

$$w_{irt} = \delta J_{rt} + \beta' X_{irt} + \alpha_i + D_t + \varepsilon_{irt} \tag{2}$$

$i \in 1, 2, ..., N$; $r \in 1, 2, ..., R$; $t \in 2006, 2007, ..., 2014$

where subscripts i, r and t denote the worker, regional employment area and year, respectively. The main parameter of interest is referred to by δ, which measures the agglomeration benefits for wages by increasing either the local number of employed workers or the local employment density. Equation (2) presents a generic empirical model, which is estimated for both the OLS estimator (without the individual-specific fixed effects term α) and the FE estimator.[10] The model is separately estimated using various regional classifications. For each regional classification and regional aggregation level, the values of the variable employment density J in regional area r are different, as the area of the worker's LLM is different. The vector X represents a set of covariates, including demographic characteristics, job characteristics and the area size of the worker's LLM. We include the area size to capture the agglomeration benefits for wages by increasing the spatial extent of a regional area. Individual-specific fixed effects are referred to by α. Annual dummies are denoted by D. ε refers to the idiosyncratic error term.

5.2. Job Displacement

A generic empirical model is specified to estimate both the displacement effect on employment and the natural logarithm of hourly wage, respectively.[11] The empirical model is given as

$$Y_{irt} = \delta(DISPLACED_i \times POST_{it}) + \rho POST_{it} + \beta' X_{it} + \alpha_i + N_r + D_t + \varepsilon_{irt} \tag{3}$$

$i \in 1, 2, ..., N$; $r \in 1, 2, ..., R$; $t \in 1, 2, ..., 108$

[10] Please note that in each of the specifications that are shown in (2), (3) and (4), all parameters refer to a different estimate.

[11] See the paper by Meekes and Hassink (2017) for more information on our quasi-experimental empirical design involving job displacement due to firm bankruptcy.

where subscripts i, r and t denote the worker, regional home area and month, respectively.[12] Note that workers are distinguished by their home location instead of employment location, to prevent the problem where we would not observe a worker's geographical employment location during an unemployment spell. The displacement effects on the outcome variables are represented by parameter δ of the two-way (double) interaction term between the scalar indicator variables *DISPLACED* and *POST*. The time-constant variable *DISPLACED* equals one for workers who have been displaced, and zero otherwise. Note that the main effect of *DISPLACED* is taken care of by including individual-specific fixed effects. The time-varying indicator variable *POST* equals one for the post-displacement period of thirty-six months. The base and omitted reference categories of *DISPLACED* and *POST* are the non-displaced and the period prior to displacement, respectively. The worker's covariates, including demographic characteristics and job characteristics, are represented by vector X. The parameters of the covariates are referred to by vector β. Individual-specific fixed effects are represented by α. N_r represents indicators for the geographical home location at the NUTS 3, PES, aggregate LLM or subgroup-specific LLM level. The aggregate LLMs and subgroup-specific LLMs are returned by flowbca. Calendar month indicators are denoted by D. ε refers to the idiosyncratic error term.

We added various interaction terms to assess the role of agglomeration economies, which is represented by employment density J, in the displacement effects on employment and hourly wage. The empirical model in (3) complements the model in (2) by adding various three-way (triple) and two-way interaction terms among employment density J, *DISPLACED* and *POST*. Moreover, we included interaction terms among a vector of worker characteristics X, *DISPLACED* and *POST*. The vector X includes time-varying variables (area size and other characteristics) as well as time-invariant variables (female, education and other characteristics of the terminated job). The empirical model is

$$\begin{aligned} Y_{irt} = {} & (\theta J_{irt}) \times DISPLACED_i \times POST_{it} + (\iota J_{irt}) \times DISPLACED_i + (\nu J_{irt}) \times POST_{it} \\ & + (\kappa' X_{irt}) \times DISPLACED_i \times POST_{it} + (\gamma' X_{irt}) \times DISPLACED_i + (\eta' X_{irt}) \times POST_{it} \\ & + \delta DISPLACED_i \times POST_{it} + \rho POST_{it} + \mu J_{irt} + \beta' X_{irt} + \alpha_i + N_r + D_t + \varepsilon_{irt} \end{aligned} \tag{4}$$

where the main parameter of interest is represented by θ. The parameter θ measures the role of employment density in the displacement effects on the dependent variable.

5.3. *Identification Challenges*

In our study on the returns to agglomeration in wages and employment, three identification challenges required particular attention. The challenges include the MAUP[13], individual-level endogeneity in employment density and local-level endogeneity in employment density.

[12] Note that we use annual data for the analysis of the urban wage premium and monthly data for the analysis of job displacement. The time period under observation t for the job displacement data sample ranges from 1 to 108, which refers to January 2006 and December 2014, respectively. Note also that vector X contains a different set of covariates in the urban wage premium data sample and job displacement data sample, respectively.

[13] See Appendix A for a detailed discussion of the implications of the MAUP.

The first identification challenge concerns the MAUP (Fotheringham and Wong, 1991; Burger et al., 2008; Briant et al., 2010). The MAUP relates to the issue that results and conclusions of empirical analyses are sensitive to the operation of the size, structure and location of regional units. In the literature on agglomeration economies, a wide range of regional classifications are used to operate the worker's LLM. The type of regional classification that is used is important, as it affects the values of variables that approximate the degree of agglomeration represented by the employed relative to the area size or the degree of tightness represented by the vacancies relative to the unemployed. The worker's employment density is the mean of the true size, given that the classification represents the LLM of a "typical" worker. Under a random (classical) measurement error, the mismeasurement leads to a parameter estimate attenuated towards zero. However, the mismeasurement in workers' LLM structure might be non-random. Specifically, there could be a worker-specific component in the structure of workers' LLM, as workers who live in the same neighbourhood are not likely to have identical LLMs. For example, low-educated workers are likely to have a smaller LLM than the mean of the true size, whereas high-educated workers are likely to have a larger LLM. Under a non-classical measurement error, the degree of attenuation is lower and the mismeasurement could lead to a sign reversal of the estimated coefficient. We assess the implications of this identification challenge by using aggregate and subgroup-specific LLMs to operate the worker's LLM, which allows workers' LLM to depend on their commuting outcomes, gender and education level.

The second challenge concerns the endogeneity in employment density at the individual level, which is caused by non-random location choices of workers. For example, unobserved characteristics like ability or reasons not related to agglomeration economies might affect the location choice and labour market outcomes (Matano and Naticchioni, 2012; Combes et al., 2012; Combes and Gobillon, 2015). We limit the potential bias from individual-level endogeneity by controlling for many factors that affect location, home change and employment decisions. For example, education level and changes in age are included to control for regional sorting based on skill and age. Moreover, we included individual-specific fixed effects to control for other potential confounding effects of time-constant variables such as abilities and knowledge other than education. In addition, for the empirical analyses on the returns to agglomeration in wages and employment following job displacement, we used a quasi-experimental design involving job displacement. This design is useful to examine the returns to agglomeration, as job displacement results in a non-culpable and unforeseen negative employment shock. By using this design, we remove potential confounding effects on post-unemployment outcomes caused by heterogeneity in the hazard rate into unemployment, signalling value, advance notification and severance pay. Moreover, the use of job displacement ensures a low number of residential relocations, because workers, following a negative employment shock, are less able or willing to relocate home (Meekes and Hassink, 2017). Thereby, our quasi-experimental design limits the problem of sorting across regional areas based on job or wage offers (Mion and Naticchioni, 2009). We compare the labour market outcomes of displaced workers with the outcomes of a control group that consists of comparable but non-displaced workers. We applied coarsened exact matching that makes displaced workers and non-displaced workers balanced in observables (Iacus et al., 2011).[14] Consequently, the selection

[14]The displaced workers are matched to non-displaced workers in the specific month of the job displacement. For

bias into displacement based on observables is greatly reduced. The identifying restriction rests on whether displaced and non-displaced workers have parallel trends in the outcome variables prior to the month of actual (for the matched displaced) and potential (for the matched non-displaced) job displacement. In Figure D.1 in Appendix D we show that our design satisfies this restriction.

The third challenge concerns endogeneity at the local level, which is caused by aggregate missing variables. Location choices of firms and workers can be affected by local productivity and local wage levels, or by differences in production and consumption amenities. For example, the more productive firms may self-select into denser LLMs. In this situation, wage premiums cannot be attributed to positive agglomeration spillovers, but are explained by a higher productivity of firms. A way to control for this endogeneity issue is to include location-specific fixed effects. Unfortunately, there is not sufficient within-individual variation in locations for all sets of aggregate LLMs. Hence, to allow for a comparison between the results of the OLS and FE estimator, we have not included the location-specific fixed effects in the empirical analyses on the UWP. See Combes and Gobillon (2015) for a discussion on several other reasons why including location-specific fixed effects does not work out well. Importantly, Combes and Gobillon (2015) argue that the issue of local-level endogeneity is less relevant than individual-level endogeneity, as the estimated effect of employment density changes much more when individual-specific fixed effects are included instead of location-specific fixed effects. As a robustness check, we apply a two-step procedure, in the spirit of Combes et al. (2008), to correct for location-specific differences in amenities, productivity and housing prices over time. Combes et al. (2008) introduce the two-step procedure to limit correlations between local-time unobservables and individual covariates. Thereby, the identification challenge involving endogenous sorting of worker quality across regional areas based on wage expectations would be addressed, and home change to a given LLM when expecting a high wage is no longer a source of bias. However, endogeneity would still be an issue if the location choice of a worker depends on the exact wage offered and obtained. See Appendix B for the application of the two-step procedure.

6. Empirical Results

6.1. Urban Wage Premium

We examine the agglomeration effects on wages (see Eq. (2)). Figure 8 shows the results of the regressions of the natural logarithm of hourly wages on employment density, demographic characteristics and job characteristics. Figure 8A and Figure 8B display the results of the OLS and FE regressions, respectively. The estimates are provided for various sets of aggregate LLMs, in which employment density and area size varies by the number of distinct LLMs (K). A lower number of distinct LLMs implies a higher level of regional aggregation. The estimates of the UWP based on

the displaced and non-displaced, this month will be referred to as the actual and potential month of job displacement, respectively. The default set of matching variables we used in the process of matching consists of the following variables: indicator variables for gender, age (21-30; 30-35; 35-40; 40-45; 45-50 and 50-59 years), children aged 18 or lower, partner, Dutch nationality, LLM-specific geographical home location, type of job (full-time or part-time), type of contract (fixed or temporary), job tenure (3-6; 6-12; 12-18 and over 18 years), firm size (10-49; 50-99; 100-499 and 500 or more employed workers), economic sector of the firm (twenty-one ISIC sectors), calendar month and calendar year.

the NUTS 3 classification and PES classification, which contain 40 and 35 distinct areas, respectively, are also provided. These estimates do not depend on the number of distinct regional units, but allow for a point of comparison. Note that when K is equal to 398, the regional classification that is used to operate employment density is identical to the set of Dutch municipalities.

Fig. 8. Aggregate LLM UWP by empirical specification (Eq. (2)). *Notes:* The dependent variable is the natural logarithm of hourly wage. Each estimate of the natural logarithm of employment density on hourly wage represents a different regression. In each regression, the variables employment density and area size are operated based on a different regional classification. The values at which K is used to operate the aggregate LLMs include 398, 350 to 50 in increments of fifty, 45 to 25 in increments of five, and 22 to 7 in increments of three. K equals 40 and 35 for the NUTS 3 classification and PES classification, respectively. The 95% confidence intervals are constructed using clustered standard errors by LLM. All regression analyses include indicator variables for the worker's gender, education category (2), age group (8), having the Dutch nationality, having a child, having a partner, economic sector of the firm (66), size of the firm (4), number of household members (3) and calendar year (8). The number of estimated parameters for each covariate is provided in parentheses. All regressions include a variable that represents the natural logarithm of the area size of the worker's LLM. The parameter estimates are not reported. Data set: the administrative data from Statistics Netherlands. The period under observation is from 2006 to 2014. The number of individual-year observations equals 18,882,294.

Figure 8A and Figure 8B show that the estimate of the UWP directly decreases in the number of distinct LLMs. For the entire interval of K, the OLS estimates of the UWP ranges between 2.6 and

6.7 per cent. More urbanised LLMs are characterised by a substantial UWP: if the employment density doubles, the increase in wages is about 2.6 to 6.7 per cent. This finding is consistent with those reported by Groot et al. (2014), who also use Dutch data, and find a UWP of 2.1 and 4 per cent using municipalities and NUTS 3 areas to operate LLMs, respectively. Groot et al. (2014) find higher estimates if they use the instrumental variables estimator. However, under the non-classical measurement error, IV estimates can be biased upward (Hyslop and Imbens, 2001). Wheeler (2001) finds a UWP of 2.7 per cent by using the logarithm of population density as the independent variable of interest at the U.S. MSA level. The UWP estimate is generally higher in studies that use a dummy variable to differ between urban and rural areas. For example, the studies by Glaeser and Maré (2001) and Yankow (2006) find that American urban workers earn about 25 or 19 per cent more than American rural workers, respectively. Using UK data, D'Costa and Overman (2014) find a UWP of 8.4 per cent.

Following the literature, we also estimate the UWP controlling for individual-specific fixed effects. Individual-specific fixed effects limit the potential of individual-level endogeneity, which is driven by sorting more able workers into larger LLMs. Our FE estimates of the UWP range from 0.3 to 1.4 per cent. The reduction in the UWP by introducing individual-specific fixed effects is consistent with the literature. After including individual-specific fixed effects, Glaeser and Maré (2001), Yankow (2006) and D'Costa and Overman (2014) find a UWP of 10.9, 5 and 2.3 per cent, respectively. Observe that the OLS and FE estimates of the UWP that are based on the forty and thirty-five distinct aggregate LLMs are higher but not significantly different than the NUTS 3 and PES estimates, respectively.

The difference between the OLS and FE estimates in Figure 8 suggests that the role of time-constant unobserved heterogeneity in the UWP is substantial. By introducing individual-specific fixed effects, the potential of endogeneity based on omitted variables is more limited. However, an alternative explanation is that the individual-specific fixed effects amplify the attenuation bias (Griliches, 1977; Freeman, 1984; Chowdhury and Nickell, 1985; Griliches and Hausman, 1986), which shifts the line of the aggregate LLM estimates downwards. Hence, we consider the range of 0.3 to 6.7 per cent as the lower and upper boundary of the UWP in the Netherlands, respectively.

Figure 9 shows the UWP for various subgroups, in order to better understand the gender differentials and education differentials in the returns to agglomeration. The subgroup-specific LLMs are used to operate the worker's LLM.[15] In Figure 9 and Figure 10, graphs A-F consist of six different subgroups. Subgroups A-C and D-F, represent male and female workers, respectively. Subgroups A and D, B and E, and C and F, represent low-educated, average-educated and high-educated workers, respectively. The UWP estimates, using the subgroup-specific LLMs, NUTS 3 and PES regional classifications, reveal that the UWP increases in the attained education level. We find that the UWP is comparable for male and female workers. More importantly, for all subgroups the UWP increases in the level of regional aggregation. This finding suggests that a large share of the returns to agglomeration takes place at a relatively high spatial scale.

Figure 10 shows the FE estimates of the UWP for the aforementioned six subgroups.[16] Note

[15] See Appendix C for the regression analyses using the aggregate LLMs. The estimates of the UWP using the subgroup-specific LLMs are comparable to the estimates of the UWP using the aggregate LLMs, but the difference between the estimate of the UWP among subgroups is slightly larger using the aggregate LLMs.

[16] See Table C.3 for the coefficients and standard errors of the UWP based on FE estimates for the forty NUTS 3

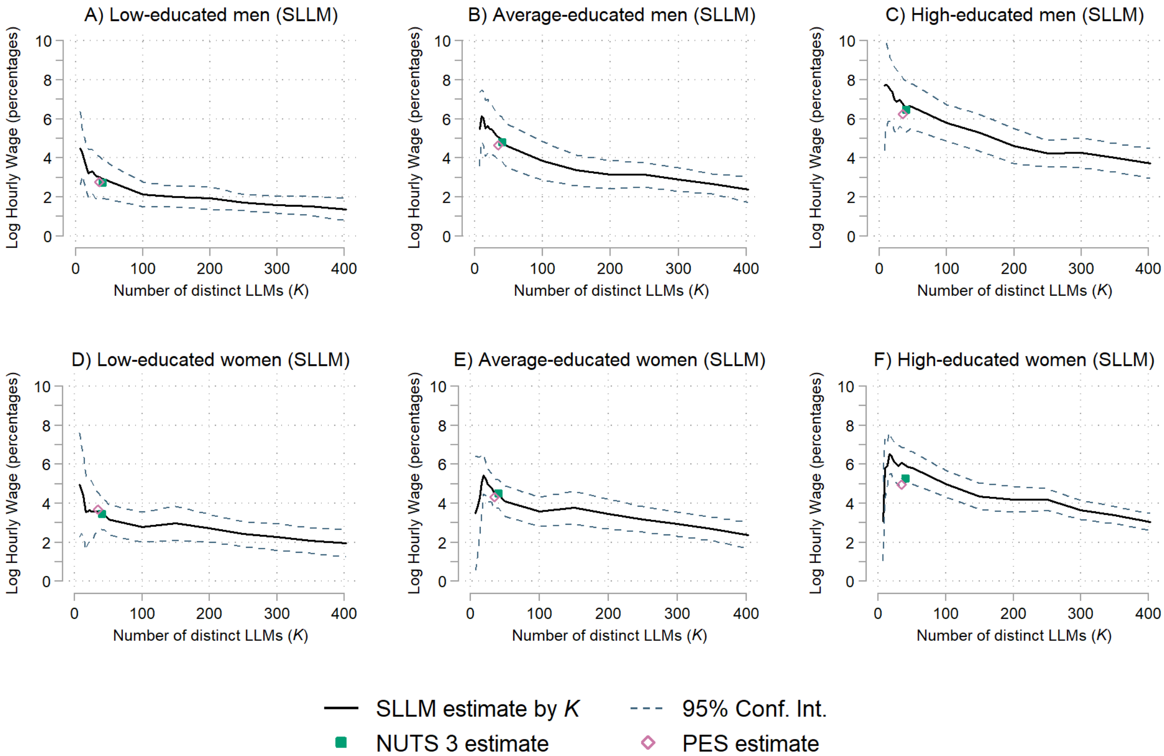


Fig. 9. Subgroup-specific LLM UWP based on OLS estimates (Eq. (2)). *Notes:* The employment density and area size of the subgroup-specific LLMs vary in gender and education level. The number of individual-year observations for the subgroups in Figures 9A-9F equals 2,296,052; 5,400,850; 4,479,115; 864,968; 2,643,962; 3,197,347, respectively. See Figure 8 for additional notes.

on the y-axes that the FE estimates of the UWP are much smaller than the OLS estimates. Consistent with Figure 9, Figure 10 also shows that the UWP is increasing in the level of regional aggregation and workers' education level. Interestingly, we find weak evidence that the UWP for low-educated and high-educated female workers is underestimated when a pre-defined exogenous regional classification is used, as estimates of the UWP based on the NUTS 3 and PES classification are smaller than estimates based on the subgroup-specific LLMs holding the regional aggregation level constant. This finding could be explained by the main input of these regional classifications, which include journey-to-work and place-of-work statistics that reflected the typical commuting outcomes of employed individuals that were predominantly male workers. Figure 10 suggests that the finding that men enjoy a larger UWP than women depends on the level of aggregation. This could explain the mixed evidence in the literature on gender- and education differentials in the returns to agglomeration.

We emphasise several findings. First, we show that at a higher level of aggregation, i.e. at a lower number of distinct LLMs, estimates of the UWP are higher. Moreover, we show that the

areas and forty subgroup-specific LLMs, respectively.

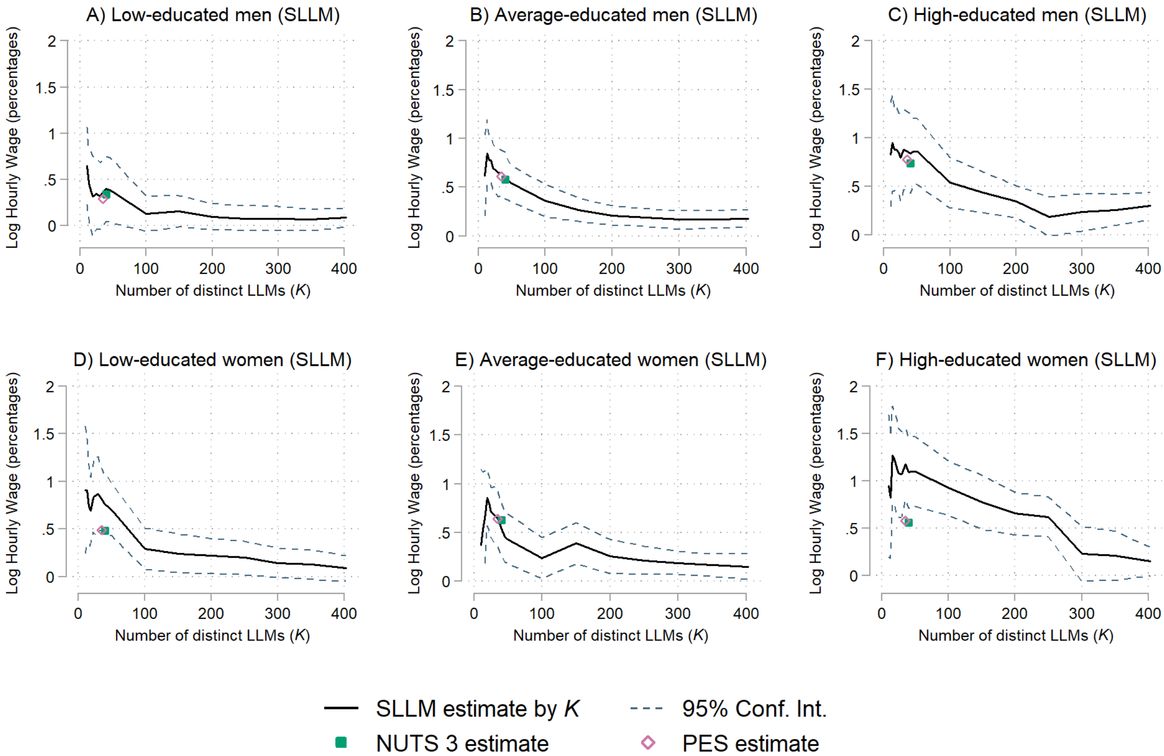


Fig. 10. Subgroup-specific LLM UWP based on FE estimates (Eq. (2)). *Notes:* See Figure 8 and Figure 9 for additional notes.

role of spatial scale in the UWP is similar across the subgroups. It is significant that the descriptive results point out that the measurement error in a worker's LLM depends on the demographic characteristics, since the LLM size and structure depends on the worker's gender and education level. Together, these findings suggest that the MAUP is not the main driver behind the effect of regional aggregation on the returns to agglomeration. Instead, it seems that the micro-foundations of agglomeration economies are more prevalent at a relatively high spatial scale. Then, consistent with the paper of Briant et al. (2010), we find that the empirical specification to estimate the UWP is more important than the regional classification that affects the structure and size of workers' LLM. However, an alternative explanation for differences in estimates of the UWP after introducing individual-specific fixed effects is that the FE estimator amplifies the attenuation bias (Griliches, 1977; Freeman, 1984; Chowdhury and Nickell, 1985; Griliches and Hausman, 1986). Finally, the UWP is increasing in the level of education. We find no gender differential in the UWP. However, our descriptive results suggest that female workers are characterised by smaller LLMs than male workers. In this regard, the empirical analysis suggests that the UWP is overestimated for female workers if a pre-defined regional classification is used to operate LLMs, as pre-defined classifications consist of relatively large LLMs.

6.2. Job Displacement

We examine to what extent displaced workers' loss of employment and wages depend on the employment density of the LLMs where workers are located. Table 2, Table 3 and Table 4 present the displacement effects on employment and wages for low-educated workers, average-educated workers and high-educated workers, respectively (see Eq. (3)). Columns (1) and (2), and (3) and (4), show the displacement effects on employment and hourly wages, respectively. For the variables displacement status (*DISPLACED*) and post-displacement period (*POST*), the omitted categories are the non-displaced workers and the pre-displacement period, respectively.

Tables 2 to 4 show that displaced workers, compared with non-displaced workers, are 20 to 30 percentage points less employed over the post-displacement period of thirty-six months. The negative displacement effect on hourly wage ranges between 3 and 8 per cent. These findings are consistent with those reported in the job displacement literature (e.g., see Schwerdt, 2011; Ichino et al., 2017). Interestingly, we show that male workers experience a more modest loss in employment than female workers. Moreover, low-educated workers experience a relatively large loss in employment. Also, high-educated workers experience relatively modest losses in both employment and wages. Using the three-way interaction models, we examine the role of local employment density in the displacement effects on employment and wages.

Table 2
Displacement effects on employment and hourly wage for low-educated workers (Eq. (3)).

	Low-educated men		Low-educated women	
	Employment (=1)	Hourly wage (log)	Employment (=1)	Hourly wage (log)
	(1)	(2)	(3)	(4)
DISPLACED × *POST*	-0.2317^{***}	-0.0715^{***}	-0.3048^{***}	-0.0309^{**}
	(0.0090)	(0.0056)	(0.0206)	(0.0126)
Number of parameters	133	130	119	119
Number of individuals	4,725	4,725	997	997
Number of observations	259,875	224,453	54,835	46,287

Notes: Each column gives the dependent variable. Parameter estimates of the two-way interaction term are reported. Clustered standard errors by aggregate LLM are in parentheses. $^{***},^{**},^{*}$, correspond to the significance level of 1%, 5%, 10%, respectively. The reference categories of *DISPLACED* and *POST* consist of the non-displaced workers and pre-displacement period, respectively. The regression analyses include individual-specific fixed effects, aggregate home LLM fixed effects and indicator variables for *POST*, age (3), children aged 18 or lower, partner, the number of household members (3), and calendar month (107). Parameter estimates of the covariates are not reported. Data set: the administrative data from Statistics Netherlands. The period under observation is from January 2006 to December 2014, in which displaced and non-displaced workers are observed for 18 months prior until 36 months after the actual and potential month of job displacement, respectively.

Table 3
Displacement effects on employment and hourly wage for average-educated workers (Eq. (3)).

	Average-educated men		Average-educated women	
	Employment (=1)	Hourly wage (log)	Employment (=1)	Hourly wage (log)
	(1)	(2)	(3)	(4)
DISPLACED × *POST*	−0.2003***	−0.0638***	−0.2754***	−0.0753***
	(0.0054)	(0.0040)	(0.0106)	(0.0078)
Number of parameters	148	148	131	130
Number of individuals	10,995	10,995	3,176	3,176
Number of observations	604,725	543,286	174,680	153,170

Notes: See Table 2 for additional notes.

Table 4
Displacement effects on employment and hourly wage for high-educated workers (Eq. (3)).

	High-educated men		High-educated women	
	Employment (=1)	Hourly wage (log)	Employment (=1)	Hourly wage (log)
	(1)	(2)	(3)	(4)
DISPLACED × *POST*	−0.2170***	−0.0358***	−0.2649***	−0.0475***
	(0.0113)	(0.0082)	(0.0158)	(0.0117)
Number of parameters	133	133	128	131
Number of individuals	2,599	2,599	1,500	1,500
Number of observations	142,945	131,785	82,500	74,854

Notes: See Table 2 for additional notes.

Figure 11 reveals to what extent the displacement effects on employment and wages depend on the employment density of LLMs where workers are located (see Eq. (4)). The subgroup-specific LLMs are used to operate the worker's LLM. Note that when K is equal to 398, the regional classification that is used to operate employment density is the set of Dutch municipalities. Figure 11A shows an insignificant three-way interaction effect of employment density on the post-displacement employment probability. For the models in which employment density is operated using the NUTS 3 or PES regional classification, we find that doubling the employment density of a worker's geographical home location increases the loss in employment by about 1 percentage point. Figure 11B shows a positive and significant displacement effect of employment density on hourly wage at a relatively high spatial scale, which include a number of distinct LLMs equal to or lower than 16. Specifically, if the employment density in the geographical home location of displaced workers doubles, the post-displacement loss in wages is about 1 to 2 percentage points lower. The PES estimate of employment density on post-displacement wages is weakly significant and equals 1.4 percentage points. The results suggest that workers who become displaced in dense LLMs, compared to workers in more sparse LLMs, experience a modest loss in wages and an intermediate loss in employment.

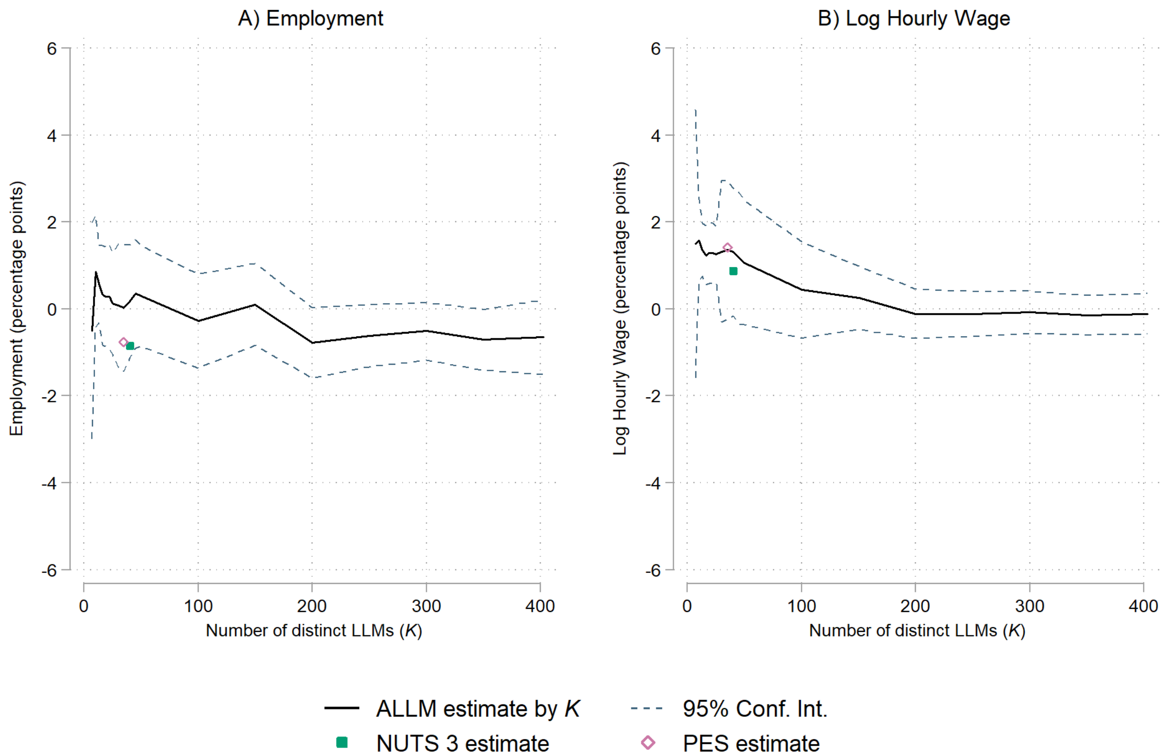


Fig. 11. Aggregate LLM displacement effects on employment and wages (Eq. (4)). *Notes:* Figure 11A and Figure 11B represent regressions of employment and the natural logarithm of hourly wage, respectively. Parameter estimates of the three-way interaction term, among *DISPLACED*, *POST* and *EMPLOYMENT DENSITY*, are reported. The 95% confidence intervals are constructed using clustered standard errors by LLM. In each regression, the natural logarithms of employment density and area size are operated based on a different regional classification. The values at which K is used to operate the aggregate LLMs include 398, 350 to 50 in increments of fifty, 45 to 25 in increments of five, and 22 to 7 in increments of three. The parameter estimates of the main and two-way interaction terms of the aforementioned independent variables are not reported. The regressions include three-way interaction terms among *DISPLACED*, *POST*, and each one of the following covariates. The regressions include a variable that represents the area size of the worker's LLM home location and various zero-one indicator variables for gender, age (3), Dutch nationality, job tenure (3), manufacturing, children aged 18 or lower, partner, number of household members (3) and year of job displacement (4). The estimates of the main, two-way interaction and three-way interaction terms of the covariates are not reported. In addition, the regressions include individual-specific fixed effects, calendar-month fixed effects (107) and LLM-specific home location fixed effects (K-1). The main effects of the LLM-specific home location fixed effects and calendar-month fixed effects are not reported. The period under observation is from January 2006 to December 2014. The number of individual-month observations equals 1,319,560 and 1,173,835 for the model in which employment and hourly wage is the dependent variable, respectively. See Table 2 for additional notes.

Figure 12 and Figure 13 reveal the subgroup-specific roles of employment density in the displacement effects on employment and hourly wages, respectively.[17] Figure 12 shows that high-educated female workers experience a significant negative effect of employment density, operated by subgroup-specific LLMs, on post-displacement employment. Also, using the NUTS 3 areas or the PES areas to operate employment density, the loss in employment is two to four percentage points lower for low-educated female workers if they reside in a geographical home location that is twice as large in terms of density. Moreover, the loss in employment is about five percentage points higher for high-educated female workers in a twice as dense location.

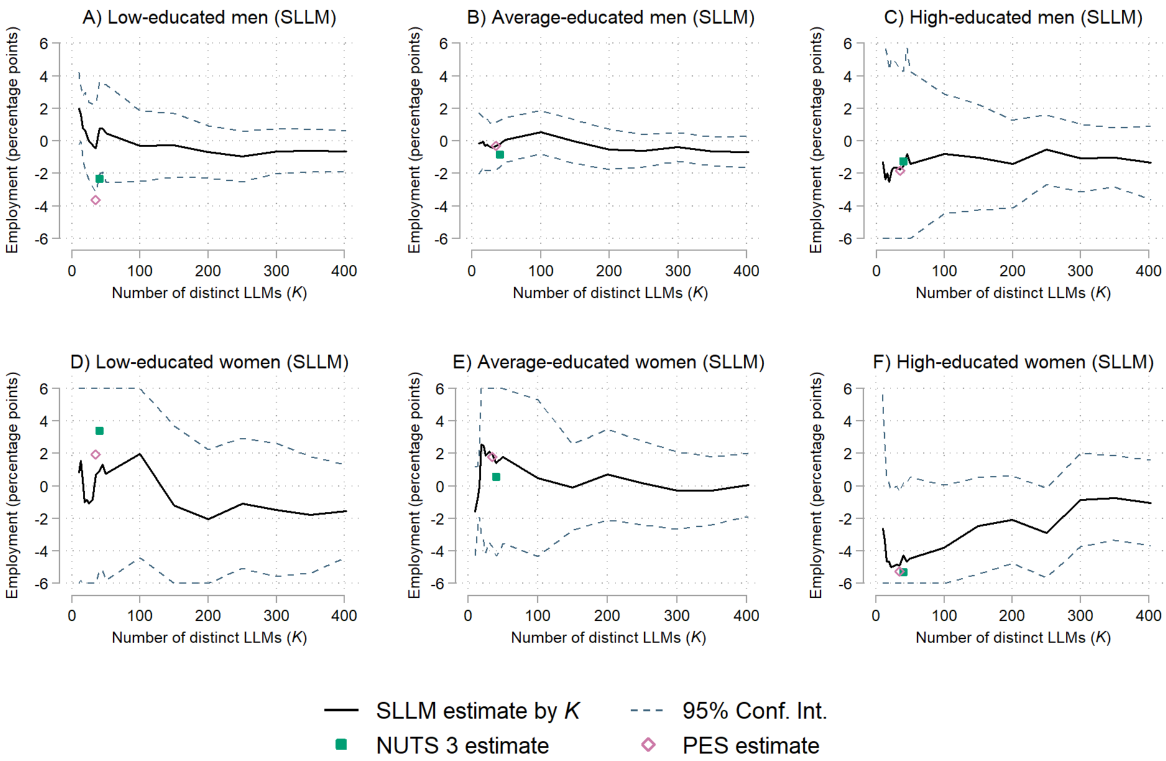


Fig. 12. Subgroup-specific LLM displacement effects on employment (Eq. (4)). *Notes:* Figure 12 represents regressions of employment. Parameter estimates of the three-way interaction term, among *DISPLACED*, *POST* and *EMPLOYMENT DENSITY*, are reported. The number of individual-month observations for the subgroups in graphs A-F equals 269,060; 612,535; 161,975; 58,905; 172,535; 80,355, respectively. See Figure 11 for additional notes.

Figure 13 shows a significant effect of employment density, operated by subgroup-specific LLMs, on post-displacement wages for high-educated men and low-educated women. Both subgroups experience more modest losses in hourly wages if they reside in denser LLMs. Using the

[17]Note that in Figure 12 and Figure 13, the 95 per cent conference intervals are in some cases set at a limit of minus six and plus six percentage points to keep the scales of the vertical axes identical. See Table D.5 for the coefficients and standard errors of the subgroup-specific LLM displacement effects for the forty NUTS 3 areas and forty subgroup-specific LLMs, respectively.

NUTS 3 or PES areas to operate LLMs, we find a significantly lower loss in wages for high-educated female workers. The results suggest that displacement in a denser LLM would lead to a more modest loss in hourly wage. Note, however, that the empirical evidence is relatively weak as the standard errors are relatively high.

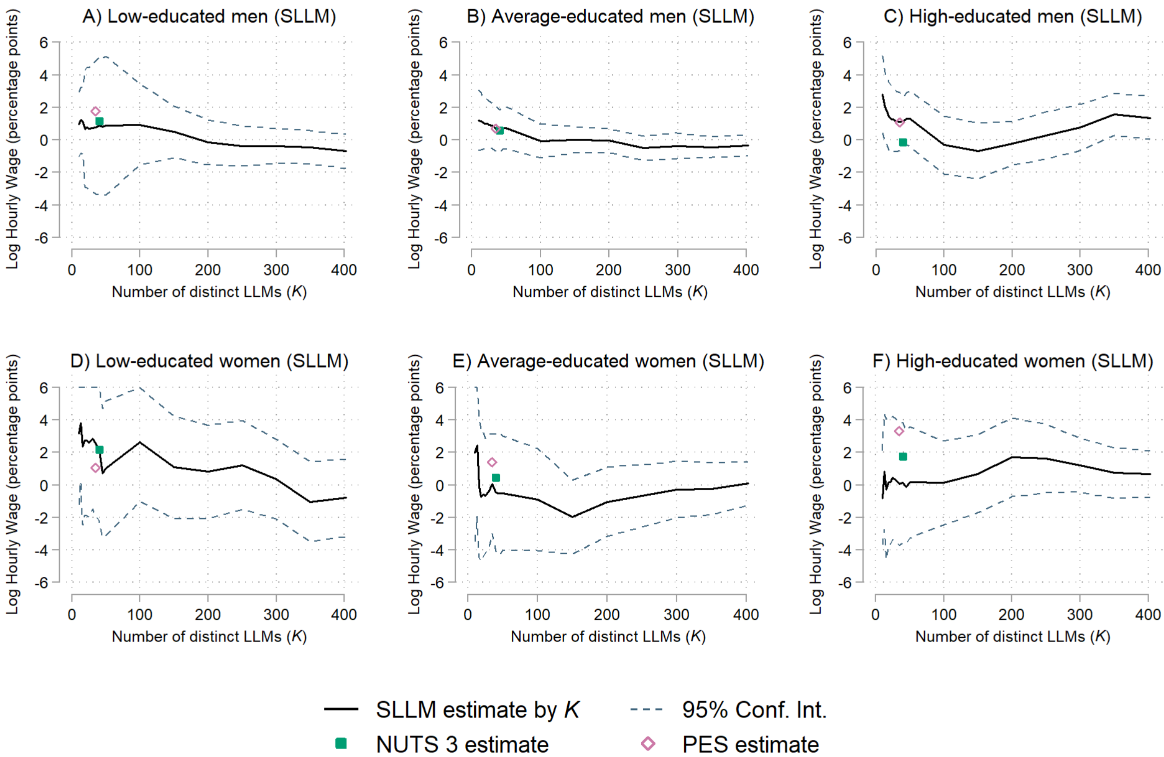


Fig. 13. Subgroup-specific LLM displacement effects on wages (Eq. (4)). *Notes:* Figure 13 represents regressions of the natural logarithm of hourly wages. Parameter estimates of the three-way interaction term, among *DISPLACED*, *POST* and *EMPLOYMENT DENSITY*, are reported. The number of individual-month observations for the subgroups in Figures 13A-13F equals 232,150; 550,028; 149,369; 49,727; 151,393; 72,919, respectively. See Figure 11 for additional notes.

In general, the results suggest that the loss in employment is more modest for low-educated and average-educated workers in more dense labour markets and more pronounced for high-educated workers in more dense labour markets. This could be explained by a more pronounced job search complexity and congestion for high-educated workers in dense labour markets. We find no clear subgroup differentials in the role of employment density in the post-displacement effects on hourly wages for workers who have been displaced.

7. Conclusion

In this paper, we examine the role of regional aggregation in the returns to agglomeration. A higher level of regional aggregation refers to a regional classification with a higher spatial scale and fewer distinct LLMs. We focus on two economic applications of the returns to agglomeration, including the effect of agglomeration externalities on wages, which is referred to as the UWP, and the returns to agglomeration in wages and employment for workers who have been displaced. We apply a unique flow-based cluster algorithm, using commuting flows across municipalities as the main input, to define alternative sets of aggregate and subgroup-specific LLMs. The subgroup-specific LLMs are defined to examine subgroup differentials in the returns to agglomeration. Our conclusions are twofold.

First, we conclude that the returns to agglomeration strongly depend on the spatial scale of LLMs. We show that estimates of the UWP are increasing by a magnitude of two to three in the level of regional aggregation. This finding suggests that a large share of the positive externalities from agglomeration economies, based on the matching, sharing, and learning micro-foundations, takes place at a relatively high spatial scale. Hence, we argue that improved matching and sharing are more important for agglomeration benefits than improved learning, as agglomeration benefits for matching and sharing take place at a relatively high spatial scale Rosenthal and Strange (2001, 2008). In addition, we examine whether agglomeration externalities affect job matching of displaced workers. We show that being located in denser LLMs leads to more modest post-displacement wage losses. Specifically, if a displaced worker is located in an LLM that is a hundred per cent denser, the loss in wage is about 2 percentage points lower. Note that this observation only holds at a relatively high spatial scale. We do not find positive returns to agglomeration in post-displacement employment at any spatial scale. Thus, we argue that the matching mechanism indeed operates at a relatively high spatial scale and leads to heterogeneity effects in job matching through wage differentials, but not to quantity effects in job matching through employment differentials. In line with the matching-function literature (e.g., Petrongolo and Pissarides (2006)), we find that agglomeration spillovers from a denser labour market allow workers to be more selective in wages with a similar re-employment probability.

Second, we conclude that the returns to agglomeration are more attenuated for subgroups of workers who are characterised by large LLMs, such as male workers and high-educated workers. Subgroup-specific local labour markets are relevant, as theory suggests that workers differ in their opportunity costs of commuting through time and financial constraints. We focus on workers who differ in gender and education levels, as we show that these demographic characteristics are the most important drivers behind differences in workers' commuting outcomes. For all subgroups, the returns to agglomeration in the UWP increase equally in the level of regional aggregation. This finding is very important as it suggests that the modifiable areal unit problem is not the main driver behind the role of regional aggregation level in agglomeration benefits, because our descriptive analysis shows that the mismeasurement in workers' LLM structure and size depends on worker's gender and education level. Holding the level of regional aggregation and thus the number of distinct LLMs constant among the different subgroups, we show that the UWP is education-biased and not gender-biased. Compared to low-educated workers, high-educated workers experience a UWP that is about 100 per cent higher. The use of aggregate LLMs or subgroup-specific LLMs provides comparable results of subgroup differentials in agglomeration benefits.

Importantly, our descriptive results show that the structure of workers' LLM is endogenous to the worker's gender, education and geographical location. Female and especially low-educated workers are characterised by small and more distinct LLMs. Given that most studies use regional classifications that represent large areas, we argue that the returns to agglomeration are generally overestimated for workers who are characterised by small LLMs. In this regard, the returns to agglomeration are also gender-biased, as women are characterised by LLMs with a relatively low level of regional aggregation that correspond to smaller returns to agglomeration. We do not find convincing empirical evidence of gender differentials or education differentials in the returns to agglomeration in post-displacement employment and wages. Consequently, our results suggest that the importance of the matching mechanism for subgroup differentials in agglomeration benefits is reasonably small. Overall, the presence of gender- and education differentials in the returns to agglomeration in the UWP is relevant from a societal perspective, as it highlights a trade-off between efficiency and societal wage equality.

Our research gives a better understanding of the structure of workers' LLM and its economic consequences, which is very relevant from a policy perspective. Examples are labour market policies which aim to increase the rate and quality of the job match between worker and employer, or which aim to limit the impact of negative employment shocks (Blumenberg, 2004; Moretti, 2011; Crépon and Van den Berg, 2016). Specifically, we show that denser labour markets lead to more modest losses in wages, but not to variation in the losses in employment. Moreover, our findings are relevant for place-based policies targeted at specific regions or subgroups of the population (Glaeser and Gottlieb, 2008; Neumark and Simpson, 2015). Place-based policies targeted at workers who are characterised by a relatively small LLM, including female workers and low-educated workers, should be focused on smaller geographical locations than policies directed at other subgroups of workers. In this regard, the so-called ripple effect, i.e. the diffusion of the policy impact over the areas next to the targeted location, might be smaller for low-educated workers than for high-educated workers. The role of the spatial scale in the efficiency of policies targeted at different subgroups of the population is a potential area for future research.

All in all, we show how a researcher is able to define meaningful clusters using flowbca, which is done in the context of economic geography by defining aggregate and subgroup-specific LLMs. A key question is whether the use of aggregate and subgroup-specific regional classifications to operate geographic space is important. A main contribution of our paper is that we show that this is indeed the case, as the regional classification and in particular the level of regional aggregation strongly affects estimates of the agglomeration benefits for wages. The implications of the spatial scale in empirical analyses may be even more important for research on larger regional areas such as the U.S. or Europe, as for these areas there are regional classifications available at a much higher level of regional aggregation such as MSAs. This paper, which deals with the importance of (subgroup-specific) regional aggregation for the empirical analysis of agglomeration economies, could aid with a broader body of research that uses regional classifications to estimate regional differences in economic outcomes.

References

Adamson, D. W., Clark, D. E., Partridge, M. D., 2004. Do urban agglomeration effects and household amenities have a skill bias? Journal of Regional Science 44 (2), 201–224.

Andersson, F., Burgess, S., Lane, J. I., 2007. Cities, matching and the productivity gains of agglomeration. Journal of Urban Economics 61 (1), 112–128.

Arzaghi, M., Henderson, J. V., Oct. 2008. Networking off Madison Avenue. Review of Economic Studies 75 (4), 1011–1038.

Autor, D. H., Dorn, D., Hanson, G. H., 2013. The geography of trade and technology shocks in the United States. American Economic Review: Papers and Proceedings 103 (3), 220–225.

Bacolod, M., Blum, B. S., Strange, W. C., 2009. Skills in the city. Journal of Urban Economics 65 (2), 136–153.

Baum-Snow, N., Pavan, R., 2012. Understanding the city size wage gap. Review of Economic Studies 79 (1), 88–127.

Berliant, M., Reed, R. R., Wang, P., 2006. Knowledge exchange, matching, and agglomeration. Journal of Urban Economics 60 (1), 69–95.

Bhaskar, V., Manning, A., To, T., 2002. Oligopsony and monopsonistic competition in labor markets. Journal of Economic Perspectives 16 (2), 155–174.

Black, D., Kolesnikova, N., Taylor, L., 2009. Earnings functions when wages and prices vary by location. Journal of Labor Economics 27 (1), 21–47.

Blumenberg, E., 2004. En-gendering effective planning: Spatial mismatch, low-income women, and transportation policy. Journal of the American Planning Association 70 (3), 269–281.

Brezzi, M., Piacentini, M., Rosina, K., Sanchez-Serra, D., 2012. Redefining urban areas in OECD countries. In: Redefining “Urban”. Organisation for Economic Co-operation and Development, pp. 19–58.

Briant, A., Combes, P. P., Lafourcade, M., 2010. Dots to boxes: Do the size and shape of spatial units jeopardize economic geography estimations? Journal of Urban Economics 67 (3), 287–302.

Brueckner, J. K., Thisse, J.-F., Zenou, Y., 1999. Why is central Paris rich and downtown Detroit poor? An amenity-based theory. European Economic Review 43 (1), 91–107.

Brueckner, J. K., Thisse, J.-F., Zenou, Y., 2002. Local labor markets, job matching, and urban location. International Economic Review 43 (1), 155–171.

Burger, M. J., Van Oort, F. G., Van der Knaap, G. A., 2008. A treatise on the geographical scale of agglomeration externalities and the Modifiable Areal Unit Problem. ERIM Report Series Research in Management ERS-2008-076-ORG.

Carlsen, F., Rattsø, J., Stokke, H. E., 2016. Education, experience, and urban wage premium. Regional Science and Urban Economics 60, 39–49.

CBS, 2018. CBS statline, Centraal Bureau voor de Statistiek, Den Haag/Heerlen.

Chowdhury, G., Nickell, S., 1985. Hourly earnings in the United States: Another look at unionization, schooling, sickness, and unemployment using PSID data. Journal of Labor Economics 3 (1), 38–69.

Ciccone, A., 2002. Agglomeration effects in Europe. European Economic Review 46 (2), 213–227.

Ciccone, A., Hall, R. E., 1996. Productivity and the density of economic activity. American Economic Review 86 (1), 54–70.

Clark, W. A. V., Huang, Y., Withers, S., 2003. Does commuting distance matter? Commuting tolerance and residential change. Regional Science and Urban Economics 33 (2), 199–221.

Combes, P.-P., Duranton, G., Gobillon, L., 2008. Spatial wage disparities: Sorting matters! Journal of Urban Economics 63 (2), 723–742.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., Roux, S., 2012. The productivity advantages of large cities: Distinguishing agglomeration from firm selection. Econometrica 80 (6), 2543–2594.

Combes, P.-P., Gobillon, L., 2015. Chapter 5 - The empirics of agglomeration economies. In: Duranton, G., Henderson, J. V., Strange, W. C. (Eds.), Handbook of Regional and Urban Economics. Vol. 5. Elsevier, pp. 247–348.

Costa, D. L., 2000. From mill town to board room: The rise of women’s paid labor. Journal of Economic Perspectives 14 (4), 101–122.

Crane, R., 1996. The influence of uncertain job location on urban form and the journey to work. Journal of Urban Economics 39 (3), 342–356.

Crane, R., 2007. Is there a quiet revolution in women's travel? Revisiting the gender gap in commuting. Journal of the American Planning Association 73 (3), 298–316.

Crépon, B., Van den Berg, G. J., 2016. Active labor market policies. Annual Review of Economics 8 (1), 521–546.

D'Costa, S., Overman, H. G., 2014. The urban wage growth premium: Sorting or learning? Regional Science and Urban Economics 48, 168–179.

De La Roca, J., Puga, D., 2016. Learning by working in big cities. Review of Economic Studies, 1–37.

Delgado, M., Porter, M. E., Stern, S., 2016. Defining clusters of related industries. Journal of Economic Geography 16 (1), 1–38.

Di Addario, S., 2011. Job search in thick markets. Journal of Urban Economics 69 (3), 303–318.

Di Addario, S., Patacchini, E., 2008. Wages and the city. Evidence from Italy. Labour Economics 15 (5), 1040–1061.

Duranton, G., 2015. Delineating metropolitan areas: Measuring spatial labour market networks through commuting patterns. In: Watanabe, T., Uesugi, I., Ono, A. (Eds.), The Economics of Interfirm Networks. No. 4 in Advances in Japanese Business and Economics. Springer Japan, pp. 107–133.

Duranton, G., Overman, H. G., 2005. Testing for localization using micro-geographic data. Review of Economic Studies 72 (4), 1077–1106.

Duranton, G., Puga, D., 2004. Chapter 48 - Micro-foundations of urban agglomeration economies. In: Henderson, J. V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics. Vol. 4. Elsevier, pp. 2063–2117.

Ellison, G., Glaeser, E. L., 1997. Geographic concentration in U.S. manufacturing industries: A dartboard approach. Journal of Political Economy 105 (5), 889–927.

Faggio, G., Silva, O., Strange, W. C., 2016. Heterogeneous agglomeration. Review of Economics and Statistics 99 (1), 80–94.

Farmer, C. J. Q., Fotheringham, A. S., 2011. Network-based functional regions. Environment and Planning A 43 (11), 2723–2741.

Fernandez, R. M., 2001. Skill-biased technological change and wage inequality: Evidence from a plant retooling. American Journal of Sociology 107 (2), 273–320.

Fernandez, R. M., Su, C., 2004. Space in the study of labor markets. Annual Review of Sociology 30 (1), 545–569.

Fotheringham, A. S., Wong, D. W. S., 1991. The modifiable areal unit problem in multivariate statistical analysis. Environment and Planning A 23 (7), 1025–1044.

Freeman, R. B., 1984. Longitudinal analyses of the effects of trade unions. Journal of Labor Economics 2 (1), 1–26.

Glaeser, E. L., Gottlieb, J. D., 2008. The economics of place-making policies. Brookings Papers on Economic Activity 39 (1 (Spring)), 155–253.

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

Goos, M., Manning, A., Salomons, A., 2009. Job polarization in Europe. American Economic Review: Papers and Proceedings 99 (2), 58–63.

Goos, M., Manning, A., Salomons, A., 2014. Explaining job polarization: Routine-biased technological change and offshoring. American Economic Review 104 (8), 2509–26.

Gould, E. D., 2007. Cities, workers, and wages: A structural analysis of the urban wage premium. Review of Economic Studies 74 (2), 477–506.

Griliches, Z., 1977. Estimating the returns to schooling: Some econometric problems. Econometrica 45 (1), 1–22.

Griliches, Z., Hausman, J. A., 1986. Errors in variables in panel data. Journal of Econometrics 31 (1), 93–118.

Groot, S. P., De Groot, H. L., Smit, M. J., 2014. Regional wage differences in the Netherlands: Micro evidence on agglomeration externalities. Journal of Regional Science 54 (3), 503–523.

Hanson, S., Pratt, G., 1988. Spatial dimensions of the gender division of labor in a local labor market. Urban Geography 9 (2), 180–202.

Helsley, R. W., Strange, W. C., 1990. Matching and agglomeration economies in a system of cities. Regional Science and Urban Economics 20 (2), 189–212.

Hoynes, H. W., 2000. Local labor markets and welfare spells: Do demand conditions matter? Review of Economics and Statistics 82 (3), 351–368.

Hyslop, R., Imbens, G. W., 2001. Bias from classical and other forms of measurement error. Journal of Business & Economic Statistics 19 (4), 475–481.

Iacus, S. M., King, G., Porro, G., 2011. Multivariate matching methods that are monotonic imbalance bounding.

Journal of the American Statistical Association 106 (493), 345–361.
Ichino, A., Schwerdt, G., Winter-Ebmer, R., Zweimüller, J., 2017. Too old to work, too young to retire? Journal of the Economics of Ageing 9, 14–29.
Lee, S., 2010. Ability sorting and consumer city. Journal of Urban Economics 68 (1), 20–33.
Lindo, J. M., 2015. Aggregation and the estimated effects of economic conditions on health. Journal of Health Economics 40, 83–96.
Madden, J. F., 1981. Why women work closer to home. Urban Studies 18 (2), 181–194.
Madden, J. F., Chiu, L. C., 1990. The wage effects of residential location and commuting constraints on employed married women. Urban Studies 27 (3), 353–369.
Manning, A., 2003. The real thin theory: Monopsony in modern labour markets. Labour Economics 10 (2), 105–131.
Manning, A., 2009. The plant size-place effect: Agglomeration and monopsony in labour markets. Journal of Economic Geography 10 (5), 717–744.
Manning, A., Petrongolo, B., 2017. How local are labor markets? Evidence from a spatial job search model. American Economic Review 107 (10), 2877–2907.
Matano, A., Naticchioni, P., 2012. Wage distribution and the spatial sorting of workers. Journal of Economic Geography 12 (2), 379–408.
Matano, A., Naticchioni, P., 2016. What drives the urban wage premium? Evidence along the wage distribution. Journal of Regional Science 56 (2), 191–209.
Meekes, J., Hassink, W. H. J., 2017. The role of the housing market in workers' resilience to job displacement after firm bankruptcy. IZA Discussion Paper 10894.
Meekes, J., Hassink, W. H. J., 2018. flowbca: A flow-based cluster algorithm in Stata. Stata Journal, in press.
Mion, G., Naticchioni, P., 2009. The spatial sorting and matching of skills and firms. Canadian Journal of Economics/Revue Canadienne d'Économique 42 (1), 28–55.
Monte, F., Redding, S., Rossi-Hansberg, E., 2018. Commuting, migration, and local employment elasticities. American Economic Review, in press.
Moretti, E., 2011. Chapter 14 - Local labor markets. In: Card, D., Ashenfelter, O. (Eds.), Handbook of Labor Economics. Vol. 4. Elsevier, pp. 1237–1313.
Neumark, D., Simpson, H., 2015. Chapter 18 - Place-based policies. In: Duranton, G., Henderson, J. V., Strange, W. C. (Eds.), Handbook of Regional and Urban Economics. Vol. 5. Elsevier, pp. 1197–1287.
Nimczik, J. S., 2018. Job mobility networks and endogenous labor markets. Mimeo.
Petrongolo, B., Pissarides, C., 2006. Scale effects in markets with search. The Economic Journal 116 (508), 21–44.
Phimister, E., 2005. Urban effects on participation and wages: Are there gender differences? Journal of Urban Economics 58 (3), 513–536.
Roberts, J., Hodgson, R., Dolan, P., 2011. It's driving her mad: Gender differences in the effects of commuting on psychological health. Journal of Health Economics 30 (5), 1064–1076.
Rosenthal, S. S., Strange, W. C., 2001. The determinants of agglomeration. Journal of Urban Economics 50 (2), 191–229.
Rosenthal, S. S., Strange, W. C., 2003. Geography, industrial organization, and agglomeration. Review of Economics and Statistics 85 (2), 377–393.
Rosenthal, S. S., Strange, W. C., 2004. Chapter 49 - Evidence on the nature and sources of agglomeration economies. In: Henderson, J. V., Thisse, J.-F. (Eds.), Handbook of Regional and Urban Economics. Vol. 4. Elsevier, pp. 2119–2171.
Rosenthal, S. S., Strange, W. C., 2008. The attenuation of human capital spillovers. Journal of Urban Economics 64 (2), 373–389.
Schwerdt, G., 2011. Labor turnover before plant closure: "Leaving the sinking ship" vs. "Captain throwing ballast overboard". Labour Economics 18 (1), 93–101.
SCP, 2015. SCP arbeidsaanbodpanel 1985–2012. DANS, Sociaal en Cultureel Planbureau, Den Haag.
Topel, R. H., 1986. Local labor markets. Journal of Political Economy 94 (3), S111–S143.
Van der Laan, L., Schalke, R., 2001. Reality versus policy: The delineation and testing of local labour market and spatial policy areas. European Planning Studies 9 (2), 201–221.
Wheeler, C. H., 2001. Search, sorting, and urban agglomeration. Journal of Labor Economics 19 (4), 879–899.

White, M. J., 1977. A model of residential location choice and commuting by men and women workers. Journal of Regional Science 17 (1), 41–52.

White, M. J., 1986. Sex differences in urban commuting patterns. American Economic Review: Papers and Proceedings 76 (2), 368–372.

White, M. J., 1988. Location choice and commuting behavior in cities with decentralized employment. Journal of Urban Economics 24 (2), 129–152.

Yankow, J. J., 2006. Why do cities pay more? An empirical examination of some competing theories of the urban wage premium. Journal of Urban Economics 60 (2), 139–161.

Zax, J. S., 1991. Compensation for commutes in labor and housing markets. Journal of Urban Economics 30 (2), 192–207.

Appendix A The MAUP in Workers' Local Labour Market

In this appendix, we provide a detailed discussion about the implications of a measurement error in the size of workers' LLM, also known as the Modifiable Areal Unit Problem (MAUP).

A general model on the UWP is given as

$$Y_{irt} = \beta_0 + \beta_1 J^*_{rt} + \varepsilon_{irt} \tag{A.1}$$

where subscripts i, r and t denote the worker, regional unit and year, respectively. The job density is referred to by J^* at regional unit r. Unfortunately, the job density is mismeasured, for example as the size of r is mismeasured. Hence, we do not observe J^* but only J, where J_{rt}=J^*_{rt}+u_{irt}. Parameter u represents the measurement error. The model we estimate, given J^*_{rt}=J_{rt}-u_{irt}, equals

$$\begin{aligned} Y_{irt} &= \beta_0 + \beta_1 J^*_{rt} + \varepsilon_{irt} = \beta_0 + \beta_1 (J_{rt} - u_{irt}) + \varepsilon_{irt} \\ &= \beta_0 + \beta_1 J_{rt} + (\varepsilon_{irt} - \beta_1 u_{irt}) \\ &= \beta_0 + \beta_1 J_{rt} + v_{irt} \end{aligned} \tag{A.2}$$

where $v_{irt} = \varepsilon_{irt} - \beta_1 u_{irt}$. The OLS estimator for β_1 is given as

$$\begin{aligned} \hat{\beta}_{1_{OLS}} &= \frac{\sum_{i=1}^{N}(J_{rt}, Y_{irt})}{\sum_{i=1}^{N} J_{rt}^2} \\ &= \frac{\sum_{i=1}^{N}(J^*_{rt}, u_{irt})(\beta_0 + \beta_1 J^*_{rt} + \varepsilon_{irt})}{\sum_{i=1}^{N}(J^*_{rt} + u_{irt})(J^*_{rt} + u_{irt})} \\ &= \frac{\beta_1 cov(J^*_{rt}, J^*_{rt}) + cov(J^*_{rt}, \varepsilon_{irt}) + \beta_1 cov(u_{irt}, J^*_{rt}) + cov(u_{irt}, \varepsilon_{irt})}{cov(J^*_{rt}, J^*_{rt}) + 2cov(J^*_{rt}, u_{irt}) + cov(u_{irt}, u_{irt})} \end{aligned} \tag{A.3}$$

And the probability limit of $\hat{\beta}_{1_{OLS}}$, given $N \to \infty$, equals

$$\begin{aligned} plim(\hat{\beta}_{1_{OLS}}) &= \beta\left(\frac{\sigma^2_{J^*} + \sigma_{J^*u}}{\sigma^2_{J^*} + \sigma^2_u + 2\sigma_{J^*u}}\right) + \frac{\sigma_{J^*\varepsilon} + \sigma_{u\varepsilon}}{\sigma^2_{J^*} + \sigma^2_u + 2\sigma_{J^*u}} \\ &= \beta\left(1 - \frac{\sigma^2_u + \sigma_{J^*u}}{\sigma^2_{J^*} + \sigma^2_u + 2\sigma_{J^*u}}\right) + \frac{\sigma_{J^*\varepsilon} + \sigma_{u\varepsilon}}{\sigma^2_{J^*} + \sigma^2_u + 2\sigma_{J^*u}} \end{aligned} \tag{A.4}$$

The classical measurement error model holds if $\sigma_{J^*u} = \sigma_{J^*\varepsilon} = \sigma_{u\varepsilon} = 0$, which implies that the covariance between the true employment density and the measurement error term equals zero ($\sigma_{J^*u} = 0$), the covariance between the true employment density and the equation error term equals zero ($\sigma_{J^*\varepsilon} = 0$), and the covariance between the measurement error and the equation error equals zero ($\sigma_{u\varepsilon} = 0$). Under the classical measurement error, the probability limit equals $plim(\hat{\beta}_{1_{OLS}}) =$

$\beta\frac{\sigma^2_{J^*}}{\sigma^2_{J^*}+\sigma^2_u} = \beta\tau$, where $\tau = \frac{\sigma^2_{J^*}}{\sigma^2_{J^*}+\sigma^2_u}$ and τ represents an attenuation bias as it is between zero and one. The bias in the estimate of the coefficient is $plim(\hat{\beta}_{1_{OLS}} - \beta) = \beta\tau - \beta = -\beta(1-\tau) = -\beta\left(1 - \frac{\sigma^2_{J^*}}{\sigma^2_{J^*}+\sigma^2_u}\right) = -\beta\frac{\sigma^2_u}{\sigma^2_{J^*}+\sigma^2_u}$.

Note that for papers that operate employment density as a zero-one indicator variable, e.g., D'Costa and Overman (2014), $\sigma_{J^*u} = 0$ does not hold. Consequently, the pool of workers who live in a peripheral area cannot under-report and the pool of workers who live in the urban area cannot over-report. This will amplify the attenuation bias, as the measurement error is negative (positive) if the dummy equals one (zero).

For various reasons it seems highly unlikely that the measurement error in employment density is classical. There exist potential confounding factors (e.g. unobserved ability) that direct high ability workers in denser LLMs (Combes et al., 2012). Moreover, there might be aggregate missing variables that influence local outcomes and local characteristics. For example, better individual outcomes attract more firms and workers in some locations, which in turn affect local characteristics. Generally, the issue of aggregate missing variables is considered to be less important (Combes and Gobillon, 2015). However, both mechanisms lead to the situation where employment density is correlated to the equation error, i.e. $\sigma_{J^*\varepsilon} \neq 0$. Consequently, the estimates of J are not likely to be consistent as the variable employment density is endogenous for the aforementioned reasons. As a solution to the endogeneity of job density, the literature tends to include more observables and individual-specific fixed effects to correct for unobserved heterogeneity. For the FE estimator, the key identification assumptions are that job changes across LLMs are random and workers' ability is time-constant.

Unfortunately, including more observables or individual-specific fixed effects amplifies the attenuation bias (Griliches, 1977; Freeman, 1984; Chowdhury and Nickell, 1985; Griliches and Hausman, 1986). To explain the role of individual-specific fixed effects in the attenuation bias, we consider the following model

$$Y_{irt} = \beta_0 + \beta_1 J^*_{rt} + a_i + \varepsilon_{irt} \tag{A.5}$$

We take the first difference to eliminate the individual-specific fixed effect a_i. Note that the same logic holds for the within estimator.

$$Y_{ir,t} - Y_{ir,t-1} = \beta_1(J^*_{r,t} - J^*_{r,t-1}) + \varepsilon_{ir,t} - \varepsilon_{ir,t-1} \tag{A.6}$$

Again, we do not observe $J^*_{r,t}$, but we observe the mismeasured variable $J_{r,t}$. We observe $J_{r,t} = J^*_{r,t} + u_{ir,t}$, which we implement in (A.6) to get

$$\begin{aligned} \Delta Y_{ir,t} = Y_{ir,t} - Y_{ir,t-1} &= \beta_1(J_{r,t} - J_{r,t-1} - u_{ir,t} + u_{ir,t-1}) + \varepsilon_{ir,t} - \varepsilon_{ir,t-1} \\ &= \beta_1(J_{r,t} - J_{r,t-1}) + \varepsilon_{ir,t} - \varepsilon_{ir,t-1} + \beta_1(u_{ir,t-1} - u_{ir,t}) \\ &= \beta_1 \Delta J_{r,t} + v_{ir,t} \end{aligned} \tag{A.7}$$

where $v_{ir,t} = \Delta\varepsilon_{ir,t} + \beta_1(u_{ir,t-1} - u_{ir,t})$. The probability limit of $\hat{\beta}_{1_{FD}}$, given $N \rightarrow \infty$, equals

$$plim(\hat{\beta}_{1_{FD}}) = \frac{\beta\sigma^2_{\Delta J^*}}{\sigma^2_{\Delta J^*} + \sigma^2_{\Delta u}} \tag{A.8}$$

To compute the probability limit of $\hat{\beta}_{1_{FE}}$, we derive the variation in the changes of the true employment density J, i.e. $\sigma^2_{\Delta J^*}$, and the variation in the changes of the measurement error u, i.e. $\sigma^2_{\Delta u}$. This gives

$$\begin{aligned} \sigma^2_{\Delta J^*_{r,t}} &= var(J^*_{r,t}) + 2cov(J^*_{r,t}, J^*_{r,t-1}) + var(J^*_{r,t-1}) = \sigma^2_{J^*_{r,t}}(1-\rho) \\ \sigma^2_{\Delta u} &= var(u_{ir,t}) + 2cov(u_{ir,t}, u_{ir,t-1}) + var(u_{ir,t-1}) = \sigma^2_{u_{ir,t}}(1-r) \end{aligned} \tag{A.9}$$

where $var(J^*_{r,t-1}) = var(u_{ir,t-1}) = 0$ holds under the assumption that $J^*_{r,t-1}$ and $u_{ir,t-1}$ are stationary. The parameter ρ and r represent the first order autocorrelation in J^* and u, respectively. The probability limit equals

$$\begin{aligned} plim(\hat{\beta}_{1_{FE}}) &= \frac{\beta\sigma^2_{J^*}(1-\rho)}{\sigma^2_{J^*}(1-\rho) + \sigma^2_u(1-r)} \\ &= \frac{\beta}{1 + \dfrac{\sigma^2_{u_{ir,t}}(1-r)}{\sigma^2_{J^*_{r,t}}(1-\rho)}} \end{aligned} \tag{A.10}$$

This equation shows that the attenuation bias will be amplified if r goes to zero (i.e. the measurement error is uncorrelated over time) or ρ goes to one (i.e. the value of the employment density J^* is strongly correlated over time). Note that if the measurement error is time-constant (i.e. $u_{i,t} = u_i$), introducing fixed effects is beneficial as it completely eliminates the attenuation bias. If this is not the case, for example due to changes in the demographic composition of cities or the location of the worker or due to temporal heterogeneity in the level of the data, the attenuation bias is amplified by introducing fixed effects. Effectively, the attenuation bias will increase if the serial correlation in the true values of employment density exceeds the serial correlation in the measurement error. An alternative way to put this is that by introducing individual-specific fixed effects to correct for time-constant sorting across areas, the variation in the measurement error σ^2_u is increased while the variation in the true employment density $\sigma^2_{J^*}$ is reduced. This holds for two reasons.

The first being that the relative number of workers with a measurement error in the regional level which is used to operate the LLM is higher with panel data. The higher the number of individual-year observations in the sample, the higher the number of observations that the regional levels are potentially mismeasured. In the context of employment density in a specific regional area, this problem is severe, since the variable is relatively time-constant while there is a measurement error in every period. The second reason is that the agglomeration effects are identified based on a small number of workers who change home across LLMs. Consequently, by using the FE estimator, the coefficient is identified by using a lower number of correct observations. Therefore, the measurement error and sorting of the few households that relocate across LLMs to a different home attenuate the FE coefficients towards zero. Indeed, the literature on agglomeration economies finds lower returns to urbanization in wages when using the FE estimator (Glaeser and Maré, 2001). This finding is typically interpreted as empirical evidence that suggests that the FE estimator limits the potential of time-constant unobserved heterogeneity. An alternative explanation is that the individual-specific fixed effects amplify the attenuation bias.

Appendix B Two-Step Estimation Procedure

In this appendix, we provide the estimates of the UWP using the two-step procedure in the spirit of Combes et al. (2008). Using the two-step procedure, we limit the potential of sorting of worker quality across LLMs. Moreover, the model controls for differences in amenities and housing prices across LLMs. Figure B.1 shows the estimates of the UWP using the aggregate LLMs to operate workers' LLM.

The first step involves the regression of individual wages on worker covariates and LLM-year FE, expressed as

$$w_{irt} = \sum_{r=1}^{R} \sum_{t=2006}^{2014} [\delta_{rt}(N_r + D_t)] + \beta' X_{it} + \alpha_i + \varepsilon_{irt} \tag{B.1}$$

The second step involves the regression of the estimated LLM-year fixed effects on employment density and the annual dummies.

$$\hat{\delta}_{rt} = \beta_1 J_{rt} + D_t + \varepsilon_{rt} \tag{B.2}$$

The results of the two-step approach are provided in Figure B.1 and Figure B.2. Compared to the direct approach of estimating the UWP (see Fig. 8), the estimates using the two-step approach are lower. This observation suggests that the direct approach leads to an overestimation of the UWP. However, the pattern of the UWP over the number of distinct LLMs is comparable: with fewer distinct LLMs the estimate of the UWP is higher. Combes et al. (2008), using French data and a similar empirical specification, find an estimate of the UWP between 3 and 4 per cent, which is consistent with our findings.

Figure B.2 shows the estimates using the two-step approach and subgroup-specific LLMs. Several observations are in place. First, the returns to agglomeration are increasing in the education level and are higher for men. Second, using the NUTS 3 areas to operate workers' LLM leads to larger differences in the UWP between education levels than when using the subgroup-specific LLMs to operate workers' LLM. Finally, compared to the use of the direct approach, the estimates of the UWP are lower if the two-step approach is used.

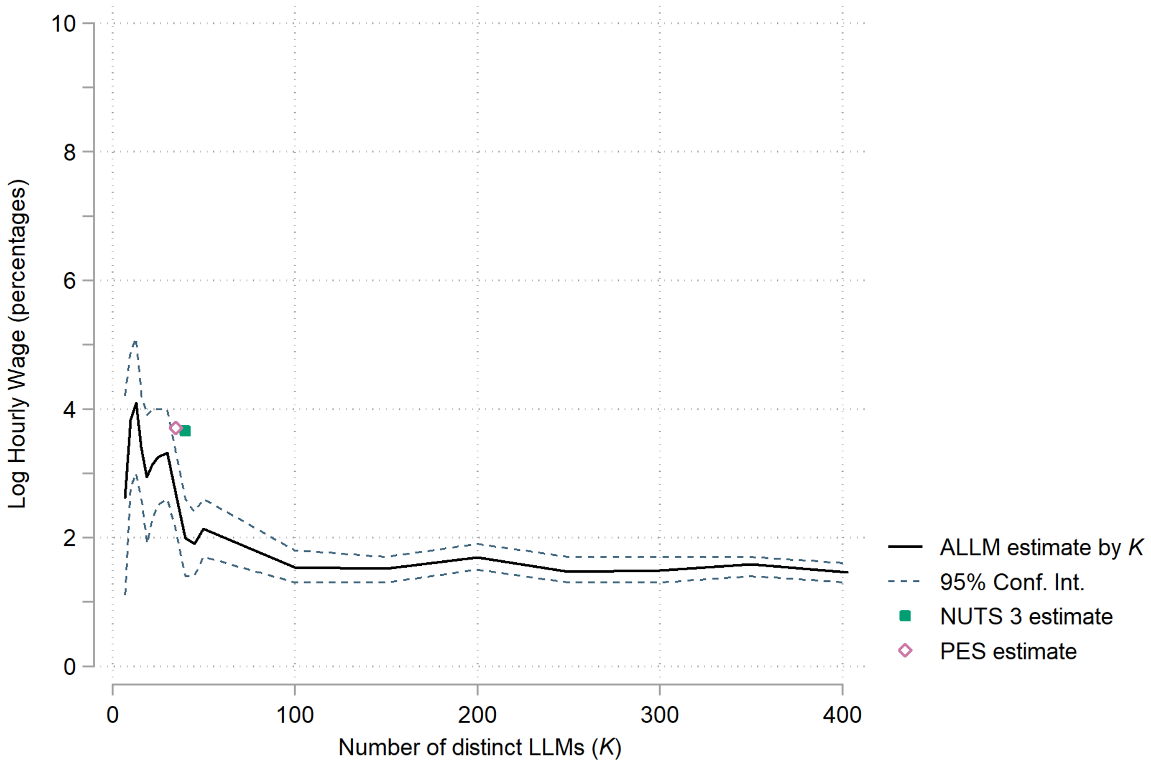


Fig. B.1. Aggregate LLM UWP based on the OLS two-step procedure (Eq. (B.2)). *Notes:* Estimates of the second stage are provided. See Figure 8 for additional notes.

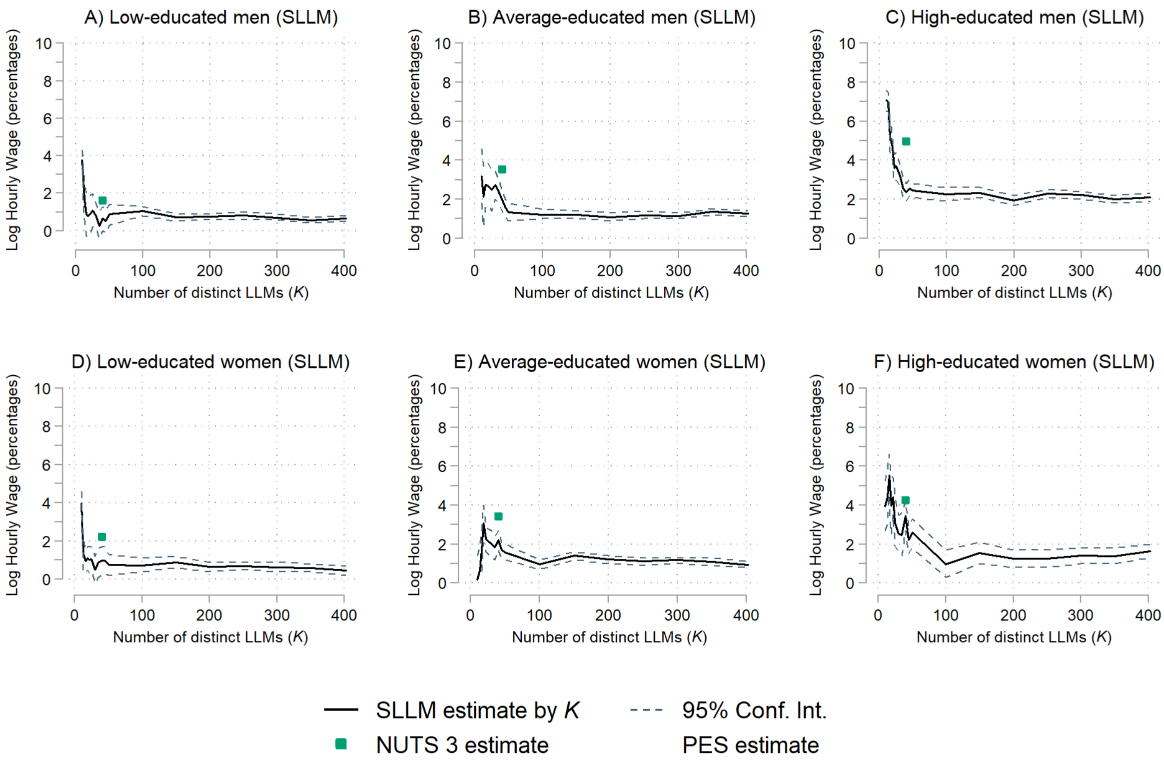


Fig. B.2. Subgroup-specific LLM UWP by subgroups based on the OLS two-step procedure (Eq. (B.2)). *Notes:* Estimates of the second stage are provided. See Figure 8 and Figure 9 for additional notes.

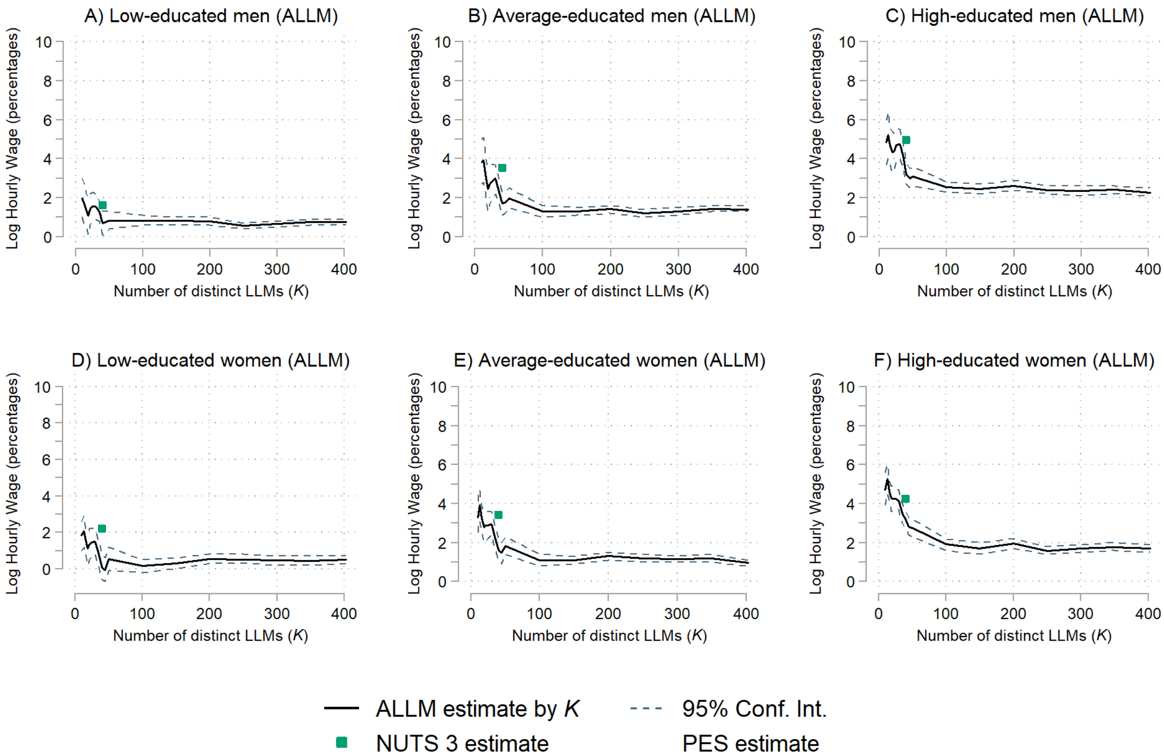


Fig. B.3. Aggregate LLM UWP by subgroups based on the OLS two-step procedure (Eq. (B.2)). *Notes:* Estimates of the second stage are provided. See Figure 8 and Figure 9 for additional notes.

Appendix C Urban Wage Premium: Summary Statistics and Robustness Checks

Table C.1
Summary statistics for hourly wage and commuting distance.

	Hourly wage (log)	Commuting distance (km)
Mean	2.7685	18.5271
St. Dev.	0.4715	25.2737
Variance	0.2223	638.7611
Skewness	0.1903	3.2560
Kurtosis	4.0985	17.3093
1th percentile	1.5706	0.5493
5th percentile	2.0531	1.2541
25th percentile	2.4652	3.7796
50th percentile	2.7496	9.9365
75th percentile	3.0591	22.0567
95th percentile	3.5525	65.8312
99th percentile	3.9886	132.9291
Number of observations	18,893,075	18,893,075

Notes: The urban wage premium data sample.

Table C.2
Individual summary statistics.

	Mean	St. Dev.
Employment (=1)	1	0
Hourly wage (log)	2.7685	0.4715
Hourly wage (€)	17.9216	12.4329
Commuting distance (km)	18.5271	25.2737
Age (in years)	36.2138	11.0393
Female (=1)	0.3552	0.4786
Low-educated (=1)	0.1674	0.3733
Average-educated (=1)	0.4261	0.4945
High-educated (=1)	0.4065	0.4912
Dutch (=1)	0.8995	0.3006
Partner (=1)	0.3901	0.4878
No child (=1)	0.6404	0.4799
Fixed contract (=1)	0.7052	0.4559
Full-time job (=1)	0.7930	0.4051
Manufacturing sector (=1)	0.1913	0.3933
Number of observations	18,893,075	18,893,075

Notes: The urban wage premium data sample.

Table C.3
Coefficients and standard errors of subgroup-specific LLM UWP based on FE estimates (Fig. 10, Eq. (2)).

	Hourly wage (log)	
Subgroup:	NUTS 3 (K = 40)	SLLM (K = 40)
Low-educated men	0.0034	0.0040
	(0.0012)	(0.0017)
Average-educated men	0.0058	0.0062
	(0.0010)	(0.0012)
High-educated men	0.0074	0.0084
	(0.0012)	(0.0020)
Low-educated women	0.0048	0.0076
	(0.0015)	(0.0015)
Average-educated women	0.0063	0.0054
	(0.0014)	(0.0013)
High-educated women	0.0056	0.0109
	(0.0015)	(0.0018)

Notes: Each estimate represents a different regression. The coefficients and standard errors are provided for the regressions in which the employment density and area size are operated based on the forty NUTS 3 areas and forty subgroup-specific LLMs, respectively. See Figure 10 for additional notes.

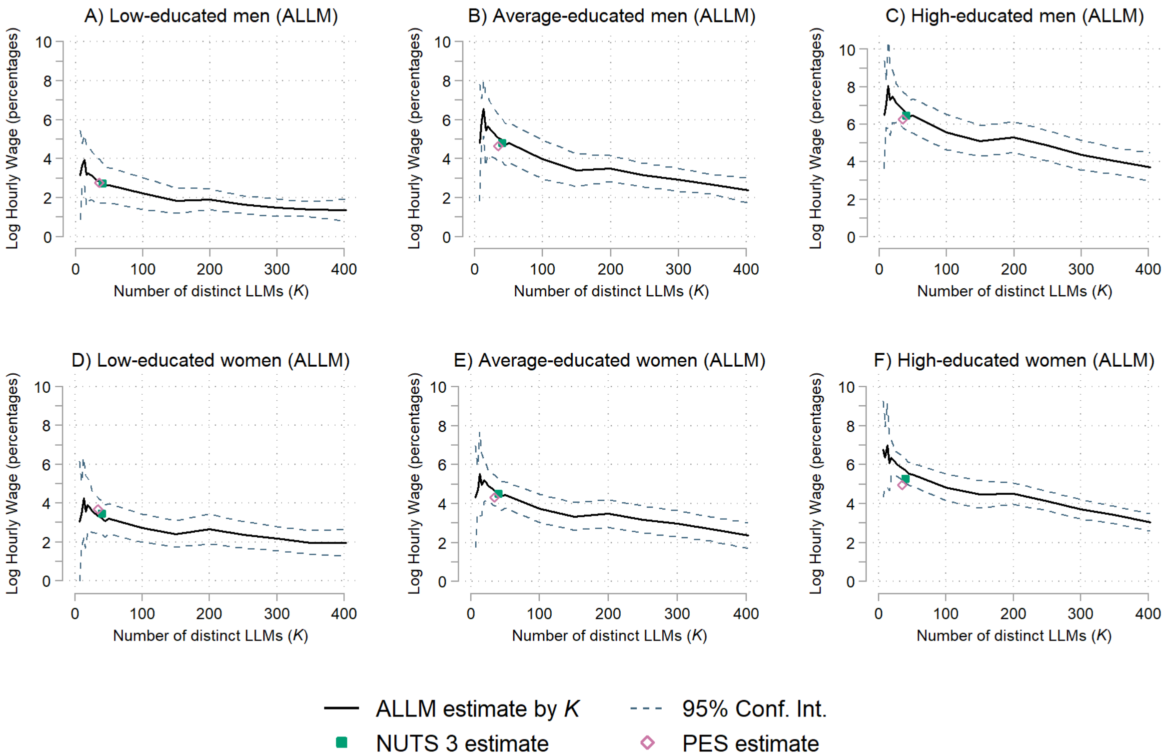


Fig. C.1. Aggregate LLM UWP based on OLS estimates (Eq. (2)). *Notes:* See Figure 9 for additional notes.

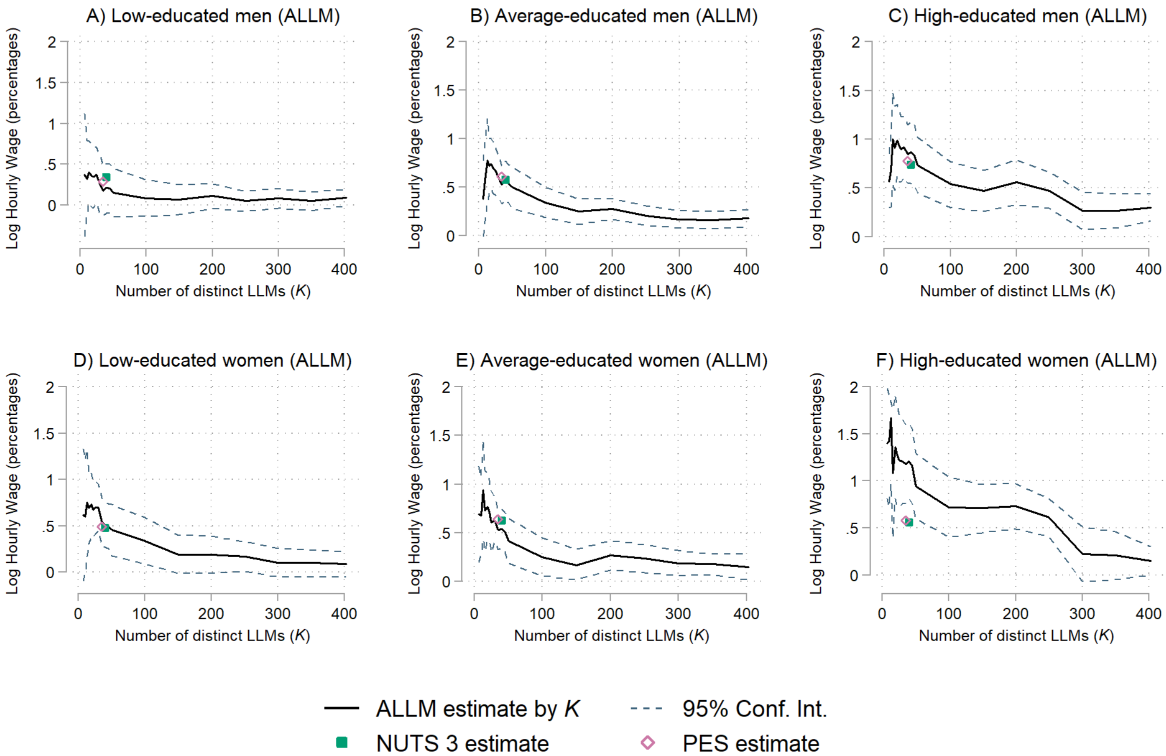


Fig. C.2. Aggregate LLM UWP based on FE estimates (Eq. (2)). *Notes:* See Figure 10 for additional notes.

Appendix D Job Displacement: Summary Statistics and Robustness Checks

Table D.1
The within change in hourly wage and commuting distance.

	Hourly wage (log)		Commuting distance (km)	
	Displaced	Non-displaced	Displaced	Non-displaced
Mean	-0.0187	0.0490	3.4568	0.5570
St. Dev.	0.3401	0.2113	32.7783	15.7735
Variance	0.1157	0.0446	1074.4167	248.8017
Skewness	-0.3814	3.5438	0.8532	0.3645
Kurtosis	29.9120	106.1860	13.6756	46.0938
1th percentile	-1.0382	-0.5346	-98.9588	-50.3882
5th percentile	-0.4812	-0.1939	-38.6010	-9.3345
25th percentile	-0.1286	-0.0012	-2.8020	0
50th percentile	0.0062	0.0386	0	0
75th percentile	0.1100	0.1031	9.8634	0
95th percentile	0.3712	0.2813	51.4306	14.5955
99th percentile	0.8198	0.6151	127.8004	60.6672
Number of observations	7,248	13,591	6,916	13,390

Notes: The job displacement data set. The individual summary statistics are based on the within change, measured by the difference in the values of each variable between the eighteenth month after job displacement and the month of job displacement.

Table D.2
Individual summary statistics using the non-matched job displacement data sample.

	Non-displaced		Displaced		
	Mean	St. Dev.	Mean	St. Dev	t-statistic
Employment (=1)	1	0	1	0	
Hourly wage (log)	2.8711	0.3903	2.7861	0.4181	32.84***
Hourly wage (€)	19.1870	11.6554	18.6162	50.9130	7.24***
Commuting distance (km)	15.5553	21.9180	17.8662	25.3218	-15.88***
Home change (=1)	0.0059	0.0764	0.0052	0.0718	1.37
Annual household income (€)	44,402	22,164	41,926	22,433	16.80***
Age (in years)	40.6143	9.2440	42.0801	9.1878	-23.90***
Female (=1)	0.4683	0.4990	0.2944	0.4558	52.52***
Low-educated (=1)	0.1723	0.3777	0.3097	0.4624	-54.78***
Average-educated (=1)	0.4153	0.4928	0.5368	0.4987	-37.16***
High-educated (=1)	0.4123	0.4923	0.1535	0.3605	79.29***
Dutch (=1)	0.9107	0.2852	0.9023	0.2969	4.44***
Partner (=1)	0.5376	0.4986	0.5598	0.4964	-6.71***
No child (=1)	0.5282	0.4992	0.5543	0.4971	-7.88***
Household members (#)	2.9257	1.3365	2.9222	1.3101	0.39
Fixed contract (=1)	0.9291	0.2566	0.9068	0.2907	13.12***
Full-time job (=1)	0.5916	0.4915	0.7096	0.4539	-36.21***
Tenure in the job (in months)	118.3416	80.9808	126.3223	86.3185	-14.85***
Manufacturing sector (=1)	0.2093	0.4068	0.4640	0.4987	-94.34***
Number of individuals (#)	10,587,265		22,765		

Notes: The individual summary statistics, provided for the month of actual or potential displacement, are based on the sample prior to matching. The time period under observation is from July 2007 to December 2011. Sample means with standard deviations are provided for the treatment group and control group. The t-statistic is provided to assess whether the mean and standard deviation of each variable for the groups of displaced and non-displaced workers are statistically different from each other. ***,**,*, correspond to the significance level of 1%, 5%, 10%, respectively. By construction, all displaced and non-displaced were employed in the month of actual or potential displacement.

Table D.3
Individual summary statistics using the matched job displacement data sample.

	Non-displaced		Displaced		
	Mean	St. Dev.	Mean	St. Dev	t-statistic
Employment (=1)	1	0	1	0	
Hourly wage (log)	2.8369	0.3786	2.8353	0.4151	0.31
Hourly wage (€)	18.4704	9.2714	19.3620	49.0370	-2.16**
Commuting distance (km)	14.9149	20.5429	17.4778	24.2814	-8.90***
Home change (=1)	0.0060	0.0771	0.0050	0.0707	0.99
Annual household income (€)	45,001	22,597	44,164	21,943	2.87***
Age (in years)	41.1290	9.9092	41.7133	9.5521	-4.59***
Female (=1)	0.2298	0.4207	0.2304	0.4211	-0.10
Low-educated (=1)	0.2330	0.4228	0.2557	0.4363	-4.06***
Average-educated (=1)	0.5821	0.4932	0.5749	0.4944	1.12
High-educated (=1)	0.1849	0.3883	0.1694	0.3752	3.10***
Dutch (=1)	0.9685	0.1747	0.9617	0.1919	2.86***
Partner (=1)	0.5759	0.4942	0.5851	0.4927	-1.44
No child (=1)	0.5548	0.4970	0.5519	0.4973	0.45
Household members (#)	3.0299	1.3294	3.0004	1.3189	1.71*
Fixed contract (=1)	0.9667	0.1794	0.9637	0.1872	1.29
Full-time job (=1)	0.7958	0.4031	0.7873	0.4092	1.60
Tenure in the job (in months)	124.8017	88.2240	129.1366	89.7851	-3.75***
Manufacturing sector (=1)	0.4919	0.5000	0.5078	0.5000	-2.45**
Number of individuals (#)	14,876		9,767		

Notes: The individual summary statistics, provided for the month of actual or potential displacement, are based on the sample after matching. The time period under observation is from July 2007 to December 2011. Sample means with standard deviations are provided for the treatment group and control group. The t-statistic is provided to assess whether the mean and standard deviation of each variable for the groups of displaced and non-displaced workers are statistically different from each other. ***,**,*, correspond to the significance level of 1%, 5%, 10%, respectively. By construction, all displaced and non-displaced were employed in the month of actual or potential displacement.

Table D.4
Firm summary statistics using the job displacement data sample.

	Firms			
	Bankrupt firms		Non-bankrupt firms	
	Mean	St. Dev.	Mean	St. Dev.
Firm size:				
1-9 employees (=1)	0	0	0	0
10-49 employees (=1)	0.5881	0.4922	0.7110	0.4534
50-99 employees (=1)	0.1289	0.3351	0.1097	0.3125
100-499 employees (=1)	0.1820	0.3859	0.1067	0.3087
500 or more employees (=1)	0.1010	0.3013	0.0727	0.2596
Firm sector:				
Agriculture, forestry and fishing (=1)	0.0041	0.0638	0.0100	0.0995
Mining and quarrying (=1)	0	0	0	0
Manufacturing (=1)	0.3224	0.4674	0.2540	0.4354
Electricity, gas, steam and air conditioning supply (=1)	0	0	0	0
Water supply; sewerage, waste management and remediation activities (=1)	0.0002	0.0127	0.0007	0.0258
Construction (=1)	0.1988	0.3991	0.1880	0.3908
Wholesale and retail trade; repair of motor vehicles and motorcycles (=1)	0.2112	0.4082	0.2037	0.4028
Transportation and storage (=1)	0.0312	0.1740	0.0503	0.2187
Accommodation and food service activities (=1)	0.0051	0.0714	0.0123	0.1104
Information and communication (=1)	0.0258	0.1585	0.0430	0.2029
Financial and insurance activities (=1)	0.0412	0.1987	0.0360	0.1863
Real estate activities (=1)	0.0014	0.0369	0.0043	0.0657
Professional, scientific and technical activities (=1)	0.0719	0.2584	0.0927	0.2900
Administrative and support service activities (=1)	0.0316	0.1748	0.0537	0.2254
Public administration and defence; compulsory social security (=1)	0	0	0	0
Education (=1)	0.0074	0.0855	0.0060	0.0772
Human health and social work activities (=1)	0.0431	0.2031	0.0353	0.1847
Arts, entertainment and recreation (=1)	0.0022	0.0465	0.0047	0.0682
Other service activities (=1)	0.0026	0.0506	0.0053	0.0728
Activities of households as employers; undifferentiated goods- and services-producing activities of households for own use (=1)	0	0	0	0
Activities of extraterritorial organisations and bodies (=1)	0	0	0	0
Number of firms (#)	3,000		12,487	

Notes: Means and standard deviations are provided at the firm level in the month of job displacement. The time period under observation is from July 2007 to December 2011. Bankrupts firms consist of all distinct firms of which an entity is declared bankrupt and a worker is displaced in the month of actual displacement. Non-bankrupt firms consist of all distinct firms where matched non-displaced workers work in the month of potential displacement.

Table D.5
Coefficients and standard errors of subgroup-specific LLM displacement effects (Fig. 12 and Fig. 13, Eq. (4)).

	Employment (=1)		Hourly wage (log)	
Subgroup:	NUTS 3 (K = 40)	SLLM (K = 40)	NUTS 3 (K = 40)	SLLM (K = 40)
Low-educated men	-0.0234	0.0077	0.0115	0.0088
	(0.0126)	(0.0140)	(0.0136)	(0.0209)
Average-educated men	-0.0085	-0.0017	0.0068	0.0053
	(0.0066)	(0.0071)	(0.0055)	(0.0065)
High-educated men	-0.0126	-0.0154	-0.0015	0.0117
	(0.0243)	(0.0284)	(0.0065)	(0.0073)
Low-educated women	0.0340	0.0096	0.0215	0.0211
	(0.0218)	(0.0300)	(0.0165)	(0.0211)
Average-educated women	0.0056	0.0141	0.0045	-0.0046
	(0.0139)	(0.0282)	(0.0141)	(0.0178)
High-educated women	-0.0528	-0.0425	0.0173	0.0013
	(0.0211)	(0.0218)	(0.0170)	(0.0181)

Notes: Each estimate represents a different regression. The coefficients and standard errors are provided for the regressions in which the employment density and area size are operated based on the forty NUTS 3 areas and forty subgroup-specific LLMs, respectively. See Figure 12 and Figure 13 for additional notes.

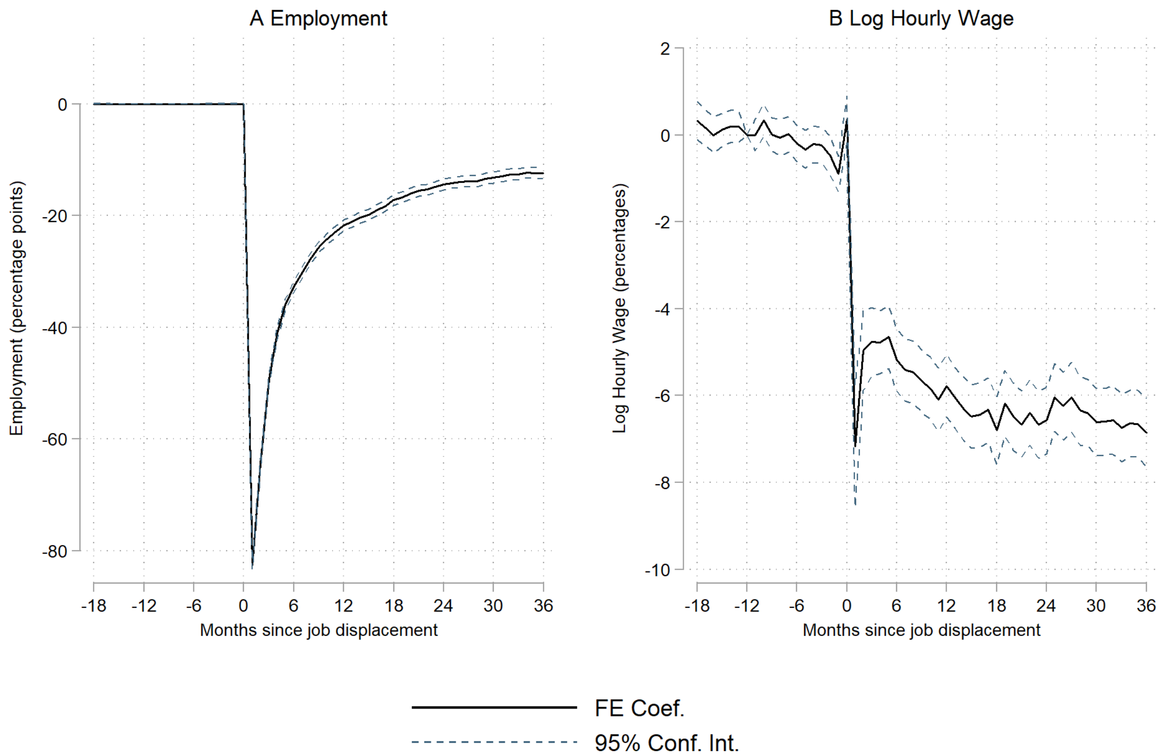


Fig. D.1. Time-dependent displacement effects on employment (A) and log hourly wage (B). *Notes:* Reference group is the group of non-displaced workers. Reference month is the twelfth month prior to job displacement. The 95% confidence intervals are computed using clustered standard errors by individual. All four fixed effects regression models include 260 parameters of which there are 54 two-way interaction terms. See Table 2 for additional notes and statistics.