I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12765

Endogenous Local Labour Markets, Regional Aggregation and Agglomeration Economies

Jordy Meekes
Wolter H. J. Hassink

NOVEMBER 2019

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12765

Endogenous Local Labour Markets, Regional Aggregation and Agglomeration Economies

Jordy Meekes
The University of Melbourne, Utrecht University, IZA and LCC

Wolter H. J. Hassink
Utrecht University and IZA

NOVEMBER 2019

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Endogenous Local Labour Markets, Regional Aggregation and Agglomeration Economies

This paper examines the role of regional aggregation in measuring agglomeration externalities. Using Dutch administrative data, we define local labour markets (LLMs) based on the worker's commuting outcomes, gender and educational attainment, and show that high-educated workers and male workers are characterised by a relatively large LLM. We find that the effect of employment density on workers' wages increases in the level of regional aggregation, explained by larger agglomeration externalities at a higher spatial scale. We quantify subgroup differentials and find that high-educated workers have agglomeration externalities twice as high as low-educated workers. We show that workers who lose their job in denser LLMs experience positive agglomeration externalities on job matching, with more modest losses in wages and again larger density effects at higher levels of regional aggregation.

JEL Classification: R12, R23, J31, J6

Keywords: urban wage premium, job loss, local labour markets, commuting, agglomeration

Corresponding author:
Jordy Meekes
Melbourne Institute: Applied Economic & Social Research
The University of Melbourne
VIC 3010
Australia
E-mail: jordy.meekes@unimelb.edu.au

1. Introduction

For decades, economists have identified and explained the mechanisms of within-country regional differences in labour market outcomes. Urban economics explains the existence of cities and the clustering of economic activity by agglomeration economies (Proost and Thisse, 2019). Agglomeration economies refer to positive externalities derived from the spatial concentration of economic processes that may lead to better employment prospects and a more productive job match (Glaeser and Maré, 2001; Moretti, 2011). The externalities may benefit workers and firms in various ways, including improved matching of employers to workers and other inputs, sharing of resources and risk, and learning through knowledge generation, diffusion and accumulation (Duranton and Puga, 2004). Importantly, studies that investigate regional differences in economic outcomes need to operationalise geographic space. Research uses the concept of local labour markets (LLMs) to define self-contained regional areas of residence and work activity, but ignores differences among workers in the size of LLMs when measuring agglomeration externalities.

The literature operationalises LLMs by using pre-defined 'exogenous' regional classifications that are identical for all workers or firms in the same location even if they have different attributes.[1] However, the fields of transportation economics and labour economics show that workers' LLM size depends on their geographical location and individual attributes (Farmer and Fotheringham, 2011; Manning and Petrongolo, 2017; Nimczik, 2018). Differences in the size of LLMs are explained by workers' opportunity costs of commuting through financial and time constraints (Small, 2012). The present paper defines LLMs that are endogenous to the worker's commuting outcomes, gender and educational attainment. We show that the LLMs of low-educated workers and of female workers are smaller than those of high-educated workers and male workers, respectively. The main motivation of our paper is to assess the importance of spatial scale for measuring agglomeration externalities on wages, which we do in two ways.

First, we examine the role of the spatial unit size of workers' LLM in the measurement of agglomeration externalities for the entire population of employees. We analyse the effect of agglomeration externalities on the productivity of labour, which is referred to in the literature as the urban wage premium (UWP) and measured by the effect of employment density or population density at the LLM level on individual wages (Glaeser and Maré, 2001; Ahlfeldt and Pietrostefani, 2019). The impact represents the net effect of the positive agglomeration forces based on the matching, sharing and learning mechanisms and the negative dispersion forces of crowding

[1] Recently there has been a surge in research that uses pre-defined classifications to study within-country regional differences as well as differences among subgroups of workers. For example, see the literature on LLMs and trade shocks (Autor et al., 2013, 2015; Helm, 2019), LLMs and polarisation (Autor, 2019), worker and firm sorting (Combes et al., 2012; Eeckhout et al., 2014; Gaubert, 2018) and mismatch (Sahin et al., 2014; Marinescu and Rathelot, 2018).

and spatial frictions. However, in the urban economics literature there is no general consensus on which spatial unit size to use as there is no consensus on the spatial scale at which agglomeration economies are prevalent (Rosenthal and Strange, 2004; Combes and Gobillon, 2015). The literature suggests a theoretical as well as an empirical mechanism is at work.

The theoretical mechanism explains agglomeration economies are prevalent at a high spatial scale even if interactions among workers and firms are highly localised. Specifically, Kerr and Kominers (2015) introduce a cluster-based framework of agglomeration that explains externalities at a high spatial scale through overlapping regions of localised interactions. Alternatively, an empirical mechanism that suggests the Modifiable Areal Unit Problem (MAUP) leads to attenuation bias and is less severe when using larger spatial units (Duranton and Overman, 2005; Briant et al., 2010). The MAUP is caused by scale effects due to arbitrary regional aggregation of small spatial units into large areas as well as by zonation effects due to arbitrary borders of regional areas (Openshaw and Taylor, 1979). We define and use multiple sets of LLMs and placebo LLMs to assess how these mechanisms affect the measurement of agglomeration externalities on wages.

Second, we analyse the role of the spatial unit size in agglomeration externalities on employment and wages to workers who are displaced because of firm bankruptcy. Using exogenous employment separations, we focus on whether agglomeration improves job matching between workers and employers.[2] A priori, the direction of agglomeration externalities in post-displacement job matching is ambiguous. Denser labour markets are characterised by lower search costs that could improve employment prospects, but also by increased search complexity and congestion that hinders matching (Helsley and Strange, 1990; Wheeler, 2001; Bleakley and Lin, 2012). Alternatively, more job opportunities in denser areas may lead to lower mismatch and lower market power of firms over employees, making denser LLMs more competitive. This could allow workers to be more selective in wages and to acquire a larger share of their marginal unit of labour in the matching process (Manning, 2003, 2010; Petrongolo and Pissarides, 2006; Hirsch et al., 2019). Moreover, Dauth et al. (2018) argue that wages in denser cities are higher because of increased assortative matching, that is high-quality workers are matched to high-quality firms. Thus, after a job loss, denser markets may improve workers' employment prospects (a quantity effect) and lead to a more modest wage loss (a heterogeneity effect).

Our empirical analysis is based on rich administrative linked employer-employee data sets from Statistics Netherlands that contain the entire population of Dutch individuals, households and firms over the period 2006 to 2014. We follow the literature by focusing on differences in

[2]The extensive literature on job displacement uses the quasi-experimental empirical design involving job loss due to firm bankruptcy or mass layoffs as an exogenous unemployment shock, to assess displacement effects on employment and wages (Jacobson et al., 1993; Huttunen et al., 2011) as well as heterogeneity in these displacement effects (Ichino et al., 2017; Neffke et al., 2018; Gathmann et al., 2018; Meekes and Hassink, 2019).

workers' LLM through differences in workers' commuting flows from residence to workplace (e.g., see Farmer and Fotheringham (2011); Brezzi et al. (2012); Duranton (2015); Foote et al. (2017)).[3] We use a flow-based cluster algorithm, flowbca, to define multiple sets of LLMs that vary in spatial unit sizes (Meekes and Hassink, 2018).[4] The main input for flowbca is relational data on commuting flows that the algorithm uses to iteratively group two spatial units into one. In each iteration, flowbca selects the 'source' unit from which the maximum directed relative commuting flow starts and aggregates this unit to the 'destination' unit. We start from a set of 398 units and iteratively aggregate units until only 7 distinct units remain. We estimate the density effects using the continuum of regional aggregations ranging from 398 to 7 distinct spatial units. We endogenise workers' LLMs to gender, education level and commuting outcomes, which allows us to use multiple sets of aggregate local labour markets (ALLM), subgroup-specific local labour markets (SLLM) and placebo local labour markets (PLLM). We also use pre-defined regional classifications of the Netherlands to facilitate a comparison with the literature, including 398 municipalities, 40 NUTS 3 areas and 35 public employment services (PES) areas.

We contribute to the literature in two ways. First, we contribute to the literature on the spatial scale of agglomeration externalities on wages using a continuum of regional aggregations. There is abundant evidence that the net effect of the agglomeration mechanisms of matching, sharing and in particular learning are highly localised and sharply attenuate with distance.[5] This evidence is based on identification strategies that estimate the impact of the economic size over a distance horizon holding the spatial unit sizes constant or by using a few administrative regional classifications at different regional aggregation levels. In contrast, we use a novel approach by estimating the agglomeration externalities using a continuum of regional aggregations, allowing LLMs to vary in size and shape. We provide evidence that agglomeration externalities increase in the size of LLMs, where larger LLMs were defined by iteratively aggregating the two units characterised by the highest relative commuting flow out of all bilateral flows into one.

Specifically, we show that the OLS estimate of the UWP equals 2.6 per cent when we use a set of 398 highly disaggregated spatial units, whereas it increases over the level of regional ag-

[3]National government departments also use commuting flows to define self-contained areas of residence and work activity, such as for the US commuting zones and the UK Travel-to-Work-Areas. An alternative approach to model differences in workers' LLM is to focus on job search behaviour of workers or employers using job-to-job flows (e.g., see Nimczik (2018)). We use commuting flows of the entire population of workers instead of job-to-job flows of job movers only, which ensures we define LLMs based on information on all workers.

[4]So far, the spatial economics literature has taken a firm perspective and used distance-based clustering or concentration indices based on densities that are non-directional by nature (e.g., see Duranton and Overman (2005); Murata et al. (2014); Delgado et al. (2016)). We use a flow-based cluster algorithm to examine the spatial scale of workers' LLM, which is directed by nature as the main input is relational data on commuting flows.

[5]See Rosenthal and Strange (2001, 2003, 2008); Rice et al. (2006); Arzaghi and Henderson (2008); Ellison et al. (2010); Andersson et al. (2016); Verstraten et al. (2019); Håkansson and Isacsson (2019).

gregation and peaks at 6.6 per cent when we use a set of 13 aggregated LLMs, other things held constant. Similarly, using the FE estimator the UWP estimate increases in the level of regional aggregation from 0.3 to 1.4 per cent. This pattern is also observed for the different subgroups of workers, and we show that high-educated workers have agglomeration externalities twice as high as low-educated workers. A methodological contribution of our paper is to use placebo LLMs that were defined by aggregating two units with low commuting connectivity, for which the pattern of larger density effects when using larger spatial units is not observed. This is important as the literature suggests a smaller downward border bias when using larger spatial units (Duranton and Overman, 2005; Briant et al., 2010). Together, these findings suggest that the MAUP is not the main cause of higher externalities when using larger spatial units. Instead, our research shows that the magnitude of the UWP more than doubles using larger LLMs, explained by larger agglomeration externalities on wages at a higher spatial scale.

Second, we contribute to the literature on job matching by focusing on the density effects on workers who lose their job following bankruptcy of the firm. The geographical matching-function literature shows that market scale effects lead to higher wages but not to more rapid re-employment (e.g., see Petrongolo and Pissarides (2006)). Moreover, recent work by Dauth et al. (2018) and Hirsch et al. (2019) on the role of matching in the UWP shows that density effects lead to higher wages. Our paper adds to these studies by using a continuum of regional aggregations, again showing that the level of regional aggregation matters and externalities are stronger using larger spatial units. We show that for displaced workers the returns to agglomeration in wages are only significant using relatively large spatial unit sizes to operationalise workers' LLM. The economic size of the effect is comparable to the UWP, as after job loss the loss in hourly wage is about 1.4 percentage points smaller for workers who reside in an LLM that is twice as dense. We find no density effects on employment for workers who have been displaced. Our findings suggest that for displaced workers the positive returns to agglomeration are larger at a higher spatial scale causing smaller wage losses with a comparable probability of re-employment.

2. Background and conceptual setting

2.1. Pre-defined administrative regional classifications

In many countries there are only a few administrative regional classifications available at different levels of regional aggregation. Examples include US Standard Metropolitan Statistical Area (MSA), US commuting zone (CZ), European Nomenclature des Unités Territoriales Statistiques (NUTS) areas and UK Travel-to-Work-Area (TTWA). In the Netherlands, the COROP regional classification, defined in 1971, was set out to identify economically and socially integrated areas. 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 US concept of CZs and the UK concept of TTWAs.

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 40 NUTS 3 areas: each NUTS 3 area consists of a core and hinterland area, while the borders of the 12 provinces are never crossed. The Dutch NUTS 3 areas have an average area size of 842 square kilometres and 70 per cent of the workers live and work in their own area, whereas the US CZs are on average 11,000 square kilometres and 90 per cent of the workers live and work in their own CZ.[6] The UK TTWAs are on average 1,000 square kilometres and about 78 per cent of the employed individuals live and work in their own LLM.[7] We use the 398 Dutch municipalities, 40 NUTS 3 areas and 35 PES areas as reference sets of LLMs, which facilitate a comparison to the sets of aggregate LLMs and subgroup-specific LLMs we define using flowbca.

2.2. Discrete, continuous, and endogenous local labour markets

In research on regional differences in economic outcomes, geographic space is generally operationalised by using a pre-defined regional classification. The advantage of using a pre-defined 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, pre-defined regional classifications represent discrete 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 in the UK is by Manning and Petrongolo (2017), who use a continuous nature of geographic space that allows for overlapping LLMs of two workers who reside in an administratively different but geographically close location. The use of continuous LLMs limits mismeasurement of workers' LLM as they remove arbitrary regional border effects.

We define discrete LLMs endogenous to demographic characteristics and commuting outcomes. By allowing for differences in workers' LLM when they meet different characteristics, we provide an alternative view of overlapping LLMs and complement the study by Manning and Petrongolo (2017). We apply flow-based clustering to define discrete LLMs instead of continuous LLMs for several reasons: discrete areas (i) are easier to visualize and interpret in the context of choices on home and work location, (ii) require fewer assumptions, less detailed data and

[6]The US CZs are defined based on commuting flows across counties. By using the US counties as building blocks for CZs, a lot of structure is already imposed on the shape of CZs. Specifically, the average area size of US counties equals about 3,000 square kilometres (Fowler et al., 2018), whereas the average size of Dutch municipalities that we use as building blocks for LLMs equals 85 square kilometres.

[7]See `http://geoportal.statistics.gov.uk/datasets/travel-to-work-areas-2011-guidance-and-information`

have fewer computational complexities; and (iii) are more easily accessible to and usable for researchers. These reasons may explain why most of the existing literature at the intersection of spatial economics uses discrete non-overlapping spatial units to operationalise geographic space.

2.3. Conceptual setting

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

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

The parameter α represents the baseline returns to agglomeration based on the matching, sharing and learning mechanisms. The role of the level of regional aggregation in the returns to agglomeration is represented by β. The parameter β equals zero if the agglomeration externalities do not depend on the level of regional aggregation. Notably, 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. We hypothesise that β is not equal to zero. Specifically, estimates of the returns to agglomeration could be increasing in the level of regional aggregation if localised interactions are in fact prevalent at a high spatial scale. Our theoretical considerations are as follows.

We start from the cluster-based framework of agglomeration by Kerr and Kominers (2015), who show that overlapping regions of firm interaction yield clusters that are larger in spatial scale than the underlying localised interactions between firms. We follow this framework in the context of agglomeration externalities to workers, as these externalities occur at the firm level as well as at the individual level through interactions among firms and workers (Duranton and Puga, 2004). We approximate interactions using workers' commuting flows. For workers, the choice on the work location depends on the interaction benefits such as higher wages and the interaction costs such as higher commuting. Workers and firms interact across spatial units, which leads to a large regional cluster of economic activity in which agents are interconnected. In this cluster the agglomeration economies could be prevalent at a larger spatial scale than only the distance at which a pair of agents has localised interactions. The turning point where the cluster's net agglomeration externalities are at its maximum and will show decreasing returns to scale is when dispersion forces such as crowding outweigh the interaction benefits. Crowding occurs as the number of agents and the cluster area size increases, because workers and firms impose congestion costs on each other through for example labour-market competition and congestion on roads or in office space. Taken together, the net effect of agglomeration forces and dispersion forces depends on the spatial scale.

Notably, workers have different attributes, which could lead to a difference in the spatial scale of LLMs and to a difference in agglomeration externalities. Assuming that a substantial share of agglomeration externalities occurs at the individual level, the returns to agglomeration may differ among subgroups for three reasons. First, agglomeration externalities could differ among subgroups through differences in the capacity to exploit the matching and learning mechanisms (see α in Equation (1)). The vast majority of the literature focuses on this mechanism, by assessing gender and education differentials in the agglomeration benefits for wages (e.g., see Phimister (2005), Gould (2007), Di Addario and Patacchini (2008), Rosenthal and Strange (2008) and Andersson et al. (2016)). Second, subgroups differ in localised interactions through differences in the spatial scale at which agglomeration externalities are prevalent (see β in (1)). Andersson et al. (2016) focus on this mechanism, showing that for high-educated workers the density effects attenuate most with distance. The authors explain this finding as a result of learning being highly localised and disproportionally important for high-educated workers. Third, workers are characterised by an LLM that differs in spatial scale according to their demographic characteristics (Farmer and Fotheringham, 2011) (see regional aggregation in (1)). For example, women work closer to home and have a smaller LLM than men, which the literature explains by a difference in labour supply because of different opportunity costs of commuting given financial and time constraints (Fernandez and Su, 2004; Roberts et al., 2011; Barbanchon et al., 2019).

The conceptual setting guides our empirical analysis and leads to several testable hypotheses. First, subgroups of workers with lower opportunity costs of commuting such as male workers or high-educated workers are characterised by fewer, larger LLMs. Second, agglomeration externalities increase in the level of regional aggregation of clusters (i.e., for larger and fewer distinct LLMs), up to a turning point where net agglomeration externalities decrease as the dispersion forces such as crowding outweigh the agglomeration forces based on the matching, sharing and learning mechanisms. Third, subgroups of workers with larger LLMs benefit from larger agglomeration externalities.

3. Data and flowbca

3.1. Data sets

We used various administrative micro data sets from Statistics Netherlands covering the period of 2006 to 2014 (CBS, 2019). The micro data sets contain data on the entire population 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 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. 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. Workers were defined as displaced workers if they lost their job between six months before the date of bankruptcy, to include the so-called early leavers, and up to twelve months after bankruptcy (Schwerdt, 2011). In addition, to ensure a strong labour market attachment, for displaced workers and their controls a minimum job tenure of three years was imposed.

3.2. 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 data set that contains annual data, we constructed workers' hourly wage of the month of December. Thereby, the hourly wage and commuting distance were constructed based on data about the same job in 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 one 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. This set of variables was used for the descriptive analysis. We used the cross-section of flows in the year 2014, as the number of distinct municipalities decreased in the period 2006 to 2014. We examined the temporal changes in the sets of commuting flows over the period, which were relatively small. For convenience, we used time-invariant LLMs. Aggregate LLMs were defined based on a set of commuting flows across municipalities of all workers together.[8] The subgroup-specific LLMs were defined using separate sets of commuting flows for workers who differ in gender or education.

The second set was used to approximate agglomeration externalities 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.[9] 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. For subgroup-specific LLMs, the employment density and area size may differ between workers if they reside in the same LLM but meet different demographic characteristics.

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.

[8] Unfortunately, the worker's work location is not consistently observed. Specifically, Statistics Netherlands has only data on the number of firm plants, each plant location and the number of employees at each specific plant. Statistics Netherlands imputes the work location by using data on the place of home and location of firm plants, linking employees to the closest firm plant while not exceeding the number of workers at firm plants. Hence, the amount of commuting interaction between municipalities is likely 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. In addition, the commuting flows are not observed of workers who are employed abroad, which represents about 0.5 per cent of the Dutch labour force in 2014 (CBS, 2019).

[9] For the aggregate LLMs and subgroup-specific LLMs, the within-LLM variation in employment density is very limited as the annual growth rate in the number of employed workers is small. Note that for the random placebo LLMs, we used for convenience time-constant values of employment density based on the year 2014.

3.3. Flow-based cluster algorithm

We use flowbca, discussed by Meekes and Hassink (2018), which is an implementation of a flow-based agglomerative hierarchical cluster algorithm that is able to define LLMs by clustering disaggregated spatial units into aggregated spatial units.[10] We 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 spatial units into one depends on the level of interaction. In our analysis, the level of interaction between spatial units is approximated by relative commuting flows from residence to workplace. The main input for the algorithm is a set of commuting flows across 398 municipalities. Alternative sets of aggregate LLMs were constructed at low to high levels of regional aggregation with a number of distinct LLMs between 398 and 7. Subgroup-specific LLMs were defined by separately using commuting flows of subgroups of workers, which include groups of both female workers and male workers varying in three education levels.

Flowbca can be described as follows. LLMs were defined by iteratively aggregating two spatial units into one. In each iteration, the algorithm selects two units that will be aggregated based on an optimisation function. The optimisation function identifies the maximum directed 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. The relative commuting flows are in each iteration computed by taking each absolute commuting flow from source unit to destination unit relative to the source unit's total of absolute outgoing flows. We use a directed flows approach that identifies the maximum single flow from one unit to another, instead of the undirected flows approach that identifies the maximum of the sum of the two flows between two units. The directed flows approach ensures we endogenously define the destination unit as the core of the LLM. We use relative commuting flows instead of absolute flows, as relative flows function as weights that account for the relative importance of a unit. The use of relative flows allows a spatial unit that is relatively small and has few absolute flows to be aggregated to a large spatial unit.[11]

The iterative process is repeated until a stopping criterion is met. The stopping criteria we use include if exactly 7 distinct LLMs have been defined, as well as if an 80 per cent level of self-containment has been achieved. After the algorithm is terminated, the level of self-containment of

[10] Existing pre-defined labour market areas such as the US commuting zones and the UK travel to work areas are also defined using an agglomerative hierarchical cluster algorithm, with commuting flows as main input.

[11] For example, consider three spatial units: A, B and C. Of the residents who live in A, 10 work in B, 15 work in C and 5 work locally in A. Although there are more residents in C, as C has a total population of 100, the commuting flows are more dispersed: 33 work in A, 33 work in B and 34 work locally in C. The same holds for unit B, as B has 30 residents and 10 work in each of the units A, B and C. In the first iteration, the maximum directed relative commuting flow, out of all bilateral flows across A, B and C, is the flow from A to C that equals 50 per cent. In the second iteration, unit A has been aggregated to C and only B and C remain. Of the residents who live in B, 20 work in C and 10 work locally. Of the residents in C, 43 work in B and 87 work locally. The maximum directed relative commuting flow, out of the two directed flows between B and C, is the flow from B to C of 66 per cent.

an LLM is defined as the population weighted local employment rate. The population weighted local employment rate is computed by dividing the total number of workers who live and work 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.

4. Descriptive results

In the descriptive results we show the application of flowbca. 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. 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.[12]

4.1. Endogenous local labour markets

We show how the aggregate LLMs and subgroup-specific LLMs, defined with flowbca, vary in the local employment rate. The algorithm that we used to define LLMs iteratively aggregates a spatial unit to another spatial unit, based on the maximum directed relative commuting flow out of all bilateral flows.[13] 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 1 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 is defined as the relative number of workers who live and work in their LLM. For the aggregate LLMs, the local employment rate varies over the number of distinct LLMs and is higher than 80 per cent for K equal to or lower than 27.

Figure 1 here

Figure 1 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. Observe that the local employment rate of the aggregate LLMs is much higher than that of the 40 NUTS 3 areas and 35 PES areas while holding the number of distinct LLMs constant. This finding can be explained by the fact that the borders of the NUTS 3 areas do

[12]In Appendix E, we explain why we focus on the six subgroups that vary in gender and education. Moreover, in Appendix E we document the changes in commute over the last decades.

[13]See Appendix D for graphs on the relative commuting flow at which two units are aggregated for the sets of aggregate LLMs and subgroup-specific LLMs.

not cross provincial borders, as well as that flowbca allows for more variation across LLMs in the number of employed workers and area size.[14] Specifically, the aggregate LLM local employment equals about 79 per cent for K equal to 40, while both the NUTS 3 and PES regional classification are characterised by a local employment of about 69 per cent. The Dutch pre-defined regional classifications are characterised by a relatively low local employment rate compared to for example the US commuting zones that are characterised by a local employment rate of about 90 per cent (Fowler et al., 2018). Overall, the algorithm that is used to cluster spatial units, flowbca, does relatively well in constructing self-contained regional areas of residence and work activity.

Figure 2 here

Figure 2 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 of the quantile regressions of commuting distance on worker characteristics in Table E.1. Note that the local employment rate is higher than 80 per cent for K equal to or lower than 107, 36, 14, 151, 76 and 26 for the subgroups of low-educated men, average-educated men, high-educated men, low-educated women, average-educated women and high-educated women, respectively.

Figure 3 here

Figure 3 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. Although any threshold is fundamentally arbitrary, the differences in LLMs between subgroups of the population also hold for stopping criteria with other levels of local employment. Figure 3 shows that the number of distinct LLMs is decreasing in the education of workers and is lower for men. In this regard, using a pre-defined regional classification, high-educated workers and male workers are characterised by an LLM that is relatively less self-contained. This observation suggests that pre-defined regional classifications are generally too large for low-educated and female workers, but too small for high-educated and male workers. Significantly, Figure 3 suggests that workers' education is more important for the LLM spatial scale than workers' gender,

[14] See Table B.4 for the minimum, maximum, median and mean of employed workers by regional classification.

as differences in the spatial scale of LLMs are more pronounced between education levels.

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 point out 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 assess the roles of aggregate and subgroup-specific LLMs in the returns to agglomeration. Moreover, we analyse subgroup differentials in agglomeration externalities.

5. Methodology

In this section, we will discuss the main identification challenges that required our particular attention. Following, we provide the empirical models that we use for the estimation of the UWP and the impact of job displacement.

5.1. 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, individual-level endogeneity in employment density and local-level endogeneity in employment density.

The first identification challenge concerns the MAUP (Openshaw and Taylor, 1979; 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 operationalisation of space. The literature on agglomeration economies uses a wide range of regional classifications to operationalise the worker's LLM. The 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 as represented by vacancies relative to unemployment. 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 in a continuous variable, the mismeasurement leads to a parameter estimate attenuated towards zero.

However, the mismeasurement in workers' LLM spatial scale might be non-random. Specifically, there is a worker-specific component in the spatial scale of workers' LLM, as workers who live close but vary in characteristics 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 direction of the bias could be upward or downward, depending on the correlation between the mismeasurement and the true underlying value of the independent variable of interest. Then the mismeasurement could also lead to a sign reversal of the estimated coefficient. Conversely, the literature argues that the bias caused by the MAUP attenuates towards zero and becomes less severe as the level of regional aggregation increases, because with fewer distinct spatial units the arbitrary border effects will be smaller and the incidence that workers do not work in the LLM where they live will be lower (Duranton and Overman, 2005; Briant et al., 2010). We assess the implications of this identification challenge by using aggregate, subgroup-specific and placebo LLMs.

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 might affect the location choice and labour market outcomes (Matano and Naticchioni, 2012; Combes et al., 2012). We limit the potential bias from individual-level endogeneity by exploiting our rich micro data controlling for many factors that affect location, home change and employment decisions. For example, education level is included to control for regional sorting based on skill. 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. Note that in the subgroup-specific analyses of the UWP and job displacement, we use the subgroup-specific LLMs and estimate the model separately for the six subgroups. Effectively we compare subgroups of workers across LLMs, which overcomes problems that make subgroups incomparable such as differences in the demand (thinness) and supply (willingness to commute) of the labour market, as well as education-biased sorting of workers across regional areas.

For the empirical analyses on the returns to agglomeration in wages and employment following job displacement, we apply a quasi-experimental design involving job displacement due to firm bankruptcy. This design is useful to examine the returns to agglomeration in job matching, 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 reduces the number of residential relocations, because in the Dutch context displaced workers relocate less frequently to a different home (Meekes and Hassink, 2019). Thereby, the 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).[15] Con-

[15]The displaced workers are matched to non-displaced workers in the specific month of the job displacement. For the displaced and non-displaced, this month will be referred to as the actual and potential month of job displacement,

sequently, the selection bias into displacement based on observables, for example based on age or industry, is greatly reduced. The identifying restriction rests on whether displaced and non-displaced workers, respectively, have parallel trends in the outcome variables before the month of actual and potential job displacement. Figure C.1 of Appendix C shows 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 externalities, but are explained by a higher productivity of firms. Although, Combes et al. (2012) show that firms in denser areas are more productive because of agglomeration externalities instead of sorting. One strategy to control for this endogeneity issue is to include location-specific fixed effects. However, there are concerns with including spatial fixed effects (Combes and Gobillon, 2015; Ahlfeldt and Pietrostefani, 2019). First, agglomeration effects will be identified based on a small number of workers who move across LLMs, and this mobility across areas is most likely endogenous. Second, there is not enough within-individual variation across locations for all sets of LLMs, as geographic mobility across small spatial units is relatively low. Consequently, we have not included the location fixed effects in the empirical analyses on the UWP. We also refrain from the instrumental variable (IV) estimators that the urban wage premium literature frequently uses to cope with local-level endogeneity, as under the non-classical measurement error IV estimates are biased away from zero (Hyslop and Imbens, 2001; Bingley and Martinello, 2017).

5.2. Urban wage premium empirical model

An empirical model, shown in (2), is specified to estimate the agglomeration externalities, represented by employment density, on wages – also referred to as the urban wage premium. The dependent variable is the natural logarithm of the hourly wage and the model is given as

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

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

respectively. The set of matching variables contains 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.

where subscripts i, r and t denote the worker, regional employment area and year, respectively. Column vector J consists of the logarithmic transformations of the variables employment density and area size. For each regional classification and regional aggregation level, the values of the variables in J are different as the spatial unit sizes of the regional areas r are different. The main parameter of interest is represented by vector δ, which includes the impact of the logarithm of employment density on wages and measures the effect of increasing either the local number of employed workers or the local employment density.[16] 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. In each of the specifications that are shown in (2), (3) and (4), all parameters refer to a different estimate. Note that for the subgroup-specific analyses of the UWP and job displacement, we use the subgroup-specific LLMs and estimate the model separately for the six subgroups. Moreover, we estimate the model separately using various regional classifications. The column vector X represents a set of covariates, including demographic characteristics and job characteristics. Individual-specific fixed effects are referred to by α. Annual dummies are denoted by D. ε refers to the idiosyncratic error term.

5.3. Job displacement empirical model

A generic empirical model is specified to estimate the displacement effects on employment and the natural logarithm of hourly wage. The empirical model is given as

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

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

where subscripts i, r and t denote the worker, regional home area and month, respectively.[17] Note that workers are distinguished by their geographical home location instead of employment location, to prevent the problem where we would not observe a worker's 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

[16] Note that including employment size in vector J instead of employment density gives identical estimates, conditional on including the area size in the model (see Combes and Gobillon (2015) for a discussion on the empirics of agglomeration economies). We control for the area size to isolate the effect of local employment density. In the spirit of Combes et al. (2008), we also apply the two-step procedure. See Appendix A for the application of this procedure.

[17] 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. Displaced and non-displaced workers are followed for eighteen months before until thirty-six months after job displacement. Vector X contains a different set of covariates in the urban wage premium data sample and job displacement data sample.

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, and zero for the month of job displacement and the pre-displacement period of eighteen months. The base and omitted reference categories of *DISPLACED* and *POST* are the non-displaced and the period before 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 externalities in the displacement effects on employment and hourly wage. The empirical model in (4) complements the model in (3) by adding various three-way (triple) and two-way interaction terms among vector J, *DISPLACED* and *POST*. The vector J includes the variables employment density and area size. Moreover, we included interaction terms among a vector of worker characteristics X, *DISPLACED* and *POST*. The vector X includes time-varying variables as well as time-invariant variables (female, education and other characteristics of the terminated job). The empirical model is

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

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

6. Empirical results

6.1. Agglomeration effects on wages: The urban wage premium

We examine the urban wage premium by estimating the effect of employment density on wages (see Eq. (2)). Figure 4 shows the results of the regressions of the natural logarithm of hourly wage on employment density, demographic characteristics and job characteristics. Figures 4A and 4B display the results of the OLS and FE regressions, respectively. The UWP 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 larger spatial units and a higher level of regional aggregation. The UWP estimates based on 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 spatial units, but allow for a point of comparison. Note that when K is equal to 398, the set of Dutch municipalities is used to operationalise LLMs.

Figure 4 shows that the UWP estimates directly decrease in the number of distinct LLMs. Over the interval of K, the OLS estimates of the UWP ranges between 2.6 and 6.6 per cent (see Figure 4A). More urbanised LLMs are characterised by a substantial UWP: if the employment density doubles, the increase in wages is about 2.6 to 6.6 per cent. This finding is consistent with those reported in the literature, as in the comprehensive summary of the quantitative literature on the effects of density by Ahlfeldt and Pietrostefani (2019) the mean and median density elasticity of wages equals 4 per cent. Our UWP estimates are also in line 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 operationalise 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 are amplified and biased upward (Hyslop and Imbens, 2001; Bingley and Martinello, 2017). De La Roca and Puga (2017) use Spanish data and find a UWP of 4.6 per cent. 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. D'Costa and Overman (2014) use UK data and find a UWP of 8.4 per cent.

We also estimate the UWP controlling for individual-specific fixed effects (see Figure 4B). Our FE estimates of the UWP range from 0.3 to 1.4 per cent. Complementing the study by Briant et al. (2010) that argues the estimator is most important for the estimation of the UWP, we show that the regional aggregation level to operationalise workers' LLM is almost as important as using the OLS estimator or FE estimator. Observe in Figure 4 that the OLS and FE estimates of the UWP that are based on the 40 and 35 distinct LLMs are higher but not significantly different from the NUTS 3 and PES estimates, respectively. These findings suggest that using pre-defined regional classifications allows for an accurate estimation of the UWP, as arbitrary border effects seem less relevant. However, Figure 4 reveals that regional aggregation effects are very important, as estimates of the UWP more than double when using larger LLMs.

The reduction in the UWP by introducing individual-specific fixed effects is consistent with the literature. After including individual-specific fixed effects, De La Roca and Puga (2017), Glaeser and Maré (2001), Yankow (2006) and D'Costa and Overman (2014) find a UWP of 2.4, 10.9, 5.0 and 2.3 per cent, respectively. Our estimate of the UWP is low compared to other countries, which could be explained by relatively high regional fragmentation of economic activities and policies in the Netherlands (OECD, 2016). The difference between the OLS and FE estimates in Figure 4 suggests that the role of time-constant unobserved heterogeneity in the UWP is substantial. The literature argues that by introducing individual-specific fixed effects, the potential of endogeneity caused by sorting of more able workers into larger LLMs is more limited (Glaeser and Maré, 2001; Combes et al., 2008). However, De La Roca and Puga (2017) argue that including fixed effects indeed provides an accurate estimate of the static agglomeration externalities, but causes a reduction in the estimate of the UWP as it ignores dynamic agglomeration benefits such as improved learning in cities that benefits wages over a long-term period. They argue that about half of the benefits of working in dense areas are static and the other half are dynamic. Notably, an alternative explanation for differences in estimates of the UWP after introducing individual-specific fixed effects is that the FE estimator amplifies the measurement bias (Griliches, 1977; Griliches and Hausman, 1986), which may shift the line of the aggregate LLM estimates downwards. We will assess this below by applying two placebo checks.

Figure 4 here

Comparing the aggregate LLM estimates with placebo LLM estimates, we assess whether agglomeration externalities are larger or the MAUP is more severe at higher regional aggregation levels. We used aggregate LLMs to operationalise workers' LLM, which are characterised by a strong connectivity in terms of commuting within each LLM and a weak connectivity to outside LLMs. The strong connectivity is caused by the decision criterion to group two spatial units into one LLM according to the highest relative commuting flow. Conversely, we define placebo LLMs that are characterised by a weak connectivity within each LLM and a strong connectivity to outside LLMs. Specifically, we compare the UWP results for aggregate LLMs with placebo LLMs, where placebo LLMs are defined in two ways: (i) aggregating the two units that have the lowest non-zero relative commuting flow to assess the importance of arbitrary regional aggregation and (ii) aggregating a random pair of units to assess the importance of arbitrary borders. We apply two placebo checks according to the two MAUP concerns (Openshaw and Taylor, 1979).

For the first placebo check we define placebo LLMs based on the decision criterion to group two spatial units according to the lowest non-zero relative commuting flow. The lowest non-zero relative commuting flow decision criterion ensures grouping two spatial units with a weak but at least some connectivity. This placebo check addresses the scale effect: variation in results because of arbitrary aggregation of spatial units into larger LLMs. The literature argues that the downward border bias becomes smaller when using fewer, larger spatial units, as there are fewer borders and the incidence of mismeasurement from workers living but working in another LLM is lower (Duranton and Overman, 2005; Briant et al., 2010). If this holds, we will observe that the estimates of the UWP also increase in the level of regional aggregation when using placebo LLMs to operationalise workers' LLM.

The placebo LLM UWP estimates in Figure 4 show that the returns to agglomeration are not increasing over the entire distribution of regional aggregation of placebo LLMs. There are two novel findings: (i) the placebo LLMs show that the scale effect of the MAUP causes a downward bias for normal to low levels of regional aggregation (where $K > 10$). Importantly, Figure 4 shows that the MAUP can be as important as introducing individual-specific fixed effects for estimates of the UWP. (ii) the placebo LLMs show that the scale effect of the MAUP causes an upward bias at very high levels of regional aggregation ($K \leq 10$). This finding indicates a non-random measurement bias in employment density at high levels of regional aggregation, causing an upward bias instead of an attenuation bias towards zero. Overall, the upward bias caused by the MAUP is highest when using ten or fewer spatial units to operationalise LLMs.

For the second placebo check we randomise the starting set of commuting flows across all spatial units and define 100 different sets of placebo LLMs for 13 different levels of regional aggregation (see Table 1). Each of the 1,300 sets of placebo LLMs is characterised by alternative combinations of aggregating spatial units into LLMs, as for each iteration the starting set of commuting flows across spatial units is differently randomised. This placebo check addresses the zonation effect: variation in results due to arbitrary borders when using alternative sets of LLMs with different combinations of spatial units holding the number of distinct LLMs, K, constant. This placebo check addresses the scale effect as well: if the zonation effect of the MAUP is more prevalent at specific levels of regional aggregation, the share of placebo LLMs that gives a higher UWP estimate than the aggregate LLM UWP estimate will depend on K.

Table 1 here

Table 1 illustrates for an interval of the number of distinct LLMs, K, the percentage of placebo LLM UWP estimates that are higher than the corresponding aggregate LLM UWP estimate. The corresponding aggregate LLM UWP refers to the estimate provided in Figure 4, based on an identical K at which the 100 placebo LLM UWP estimates are estimated. Consistent with the first placebo check, Table 1 shows that the upward bias caused by the MAUP becomes more prevalent if the number of distinct LLMs decreases. Specifically, using ten or fewer spatial units to operationalise LLMs, more than 20 per cent of the OLS placebo LLM estimates is higher than the corresponding OLS aggregate LLM estimate, again indicating a non-random measurement bias in employment density causing an upward bias. An alternative way to interpret this is that at high levels of regional aggregation, specifically for $K \leq 10$, the aggregate LLM UWP estimates are not significantly different from zero.

Moreover, Table 1 shows that compared with the OLS estimator, the upward bias caused by the MAUP when using ten or fewer spatial units to operationalise LLMs is more prevalent with the FE estimator. In contrast, the upward bias is less severe at low levels of regional aggregation with the FE estimator. This finding suggests that the FE estimator amplifies the upward bias in the estimation of agglomeration economies at high levels of regional aggregation with few distinct LLMs. Thus at very high levels of regional aggregation the MAUP causes an upward bias and is a concern for the estimation of agglomeration externalities. Based on both placebo checks, considering the MAUP does not cause an upward bias in the UWP over the interval $13 \leq K < 400$, the increase in the UWP when using more aggregated LLMs is explained by capturing larger agglomeration externalities at a higher spatial scale.

6.2. Subgroup-specific differences in the urban wage premium

Figures 5 and 6, for respectively the OLS and FE estimator, show the UWP for subgroups in order to better understand the gender differentials and education differentials in the returns to agglomeration. The subgroup-specific LLMs are used to operationalise the worker's LLM.[18] In Figures 5 and 6, 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 orange dashed line represents the subgroup-specific LLM estimate for the number of distinct LLMs (K) at which the subgroup-specific local employment rate equals 80 per cent (see Figures 2 and 3). As shown in Figure 3, this holds for K equal to 107, 36, 14, 151, 76 and 26 for subgroups A-F, respectively.

The estimates in Figure 5, based on the subgroup-specific LLMs, NUTS 3 and PES regional classifications, reveal that the UWP increases in the attained education level. Moreover, we find that the UWP is comparable for male and female workers when holding the number of distinct LLMs constant. Also, Figure 5 reveals that 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. Importantly, the descriptive results indicate that the size of a worker's LLM depends on the demographic characteristics. The orange dashed line takes this into account by providing the subgroup-specific LLM estimate for the subgroup-specific number of distinct LLMs at which the subgroup-specific local employment rate equals 80 per cent. For average-educated and high-educated workers, it seems that the UWP is gender-biased as men enjoy a higher UWP than women, although note this difference is statistically insignificant.

[18] See Appendix B for the regression analyses using the aggregate LLMs. The UWP estimates based on the subgroup-specific LLMs are comparable to the UWP estimates based on the aggregate LLMs.

Figure 5 here

Figure 6 shows the FE estimates of the UWP for the aforementioned six subgroups.[19] Note on the *y*-axes that the FE estimates of the UWP are much smaller than the OLS estimates. Consistent with Figure 5, Figure 6 also shows that the UWP is increasing in the level of regional aggregation and workers' education level. The orange dashed line reveals that the UWP for low-educated and high-educated workers is overestimated and underestimated, respectively, when a pre-defined regional classification is used. For example, observe that for low-educated workers estimates of the UWP based on the NUTS 3 and PES classification are higher than the subgroup-specific LLM estimates at which the subgroup-specific local employment rate equals 80 per cent. 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. Importantly, using LLMs with a higher local employment rate, such as the US commuting zones with 90 per cent local employment, the likelihood of overestimating agglomeration benefits for low-educated workers is much higher. Figure 6 reveals that the finding that men enjoy a larger UWP than women depends on the level of aggregation and level of education. This observation could explain the mixed evidence in the literature on gender- and education differentials in the returns to agglomeration.

Figure 6 here

6.3. Returns to agglomeration in post-displacement employment and wages

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 presents the effects of job loss on employment and wages in columns (1) and (2), 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. Table 2 shows that displaced workers, compared with non-displaced workers, are about 23 percentage points less employed over the post-displacement period of thirty-six months. The negative displacement effect on hourly wage ranges between 6 and 7 per cent. These findings are consistent with those reported in the job displacement literature (e.g., see Schwerdt, 2011; Ichino et al., 2017).

[19] See Table B.3 for the coefficients and standard errors of the UWP based on FE estimates for the 40 NUTS 3 areas and 40 subgroup-specific LLMs, respectively.

Table 2 here

Figure 7 illustrates the role of local employment density at the aggregate LLM level in the displacement effects on employment and wages, based on the three-way interaction models (see Eq. (4)). The subgroup-specific LLMs are used to operationalise the worker's LLM.[20] When K equals 398, the regional classification that is used to operationalise workers' LLM is the set of Dutch municipalities. Figure 7A shows an insignificant three-way interaction effect of employment density on the post-displacement employment probability. Figure 7B 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 25. Specifically, if the employment density in the geographical home location of displaced workers doubles, the post-displacement loss in wages is about 1.3 to 1.5 percentage points lower. The PES estimate of employment density on post-displacement wages is weakly significant and equals 1.4 percentage points. Overall, we find agglomeration matching benefits for wages that support the literature (e.g., Petrongolo and Pissarides (2006); Dauth et al. (2018); Hirsch et al. (2019)), using a continuum of regional aggregations again showing that agglomeration externalities are stronger using larger spatial units.

Figure 7 here

6.4. Discussion of agglomeration economies results

We emphasise several findings based on our empirical analyses. Consistent with the literature, we show that the LLMs of low-educated workers and of female workers are smaller than those of high-educated workers and male workers, respectively (e.g., see Farmer and Fotheringham (2011); Nimczik (2018)). Importantly, this finding points out that the mismeasurement in workers' LLM when using a pre-defined regional classification depends on the worker's characteristics. For example, using pre-defined classifications that do not differ among subgroups of workers, the LLM is likely to be too large for low-educated female workers whereas it is too small for high-educated male workers. Given that pre-defined regional classifications such as the US commuting zones often contain large spatial units, the observation on the LLM being too large is most relevant.

Using a continuum of regional aggregations, we show that estimates of the UWP more than double using LLMs consisting of larger spatial units. Specifically, using the OLS estimator the UWP estimate increases from 2.6 per cent for 398 spatial units to 6.6 per cent for 13 aggregated spatial units, whereas using the FE estimator the UWP estimate increases from 0.3 to 1.4 per cent.

[20]We find no clear evidence on subgroup differentials in the role of agglomeration economies in displacement effects on employment and wages (see Appendix F).

This pattern is not observed when using placebo LLMs to operationalise workers' LLM, as we show that the MAUP causes an upward bias in the UWP at high levels of regional aggregation and a downward bias at low levels of regional aggregation. In fact, the upward bias is severe at very high levels of regional aggregation with ten or fewer distinct LLMs, which makes us unable to infer whether at this spatial scale agglomeration externalities decrease because of stronger dispersion forces or because of the MAUP. The question then arises whether aggregation effects are more important for empirical analyses on larger areas such as the US, Australia or Europe, as for these areas there are regional classifications available at a higher level of regional aggregation such as the US commuting zones (Foote et al., 2017; Fowler et al., 2018).

Following, we find that the UWP is education-biased but not gender-biased. Compared to low-educated workers, high-educated workers experience a UWP that is about 100 per cent higher. We find no gender differential in the UWP when holding the number of distinct LLMs constant. Importantly, our descriptive results point out that female workers and low-educated workers are characterised by smaller LLMs than male workers and high-educated workers, respectively. In this regard, we argue that if a pre-defined regional classification with average spatial unit sizes is used to operationalise workers' LLM, the UWP is likely to be overestimated for low-educated and female workers and underestimated for high-educated and male workers. Also, we show that subgroups who differ in characteristics face a similar agglomeration spillover curve over the level of regional aggregation. Together, these findings suggest that the MAUP is not the main driver behind the effect of regional aggregation on the returns to agglomeration when using at least 13 distinct spatial units to operationalise LLMs. Instead, it seems that the agglomeration economies are prevalent at a relatively high spatial scale.

Finally, the results on the returns to agglomeration in post-displacement outcomes suggest that workers who lose their job in dense LLMs, compared to workers who lose their job in more sparse LLMs, experience a modest loss in wages and a comparable loss in employment. These results corroborate the literature on wage benefits from matching in denser markets (e.g., Petrongolo and Pissarides (2006); Dauth et al. (2018); Hirsch et al. (2019)). Specifically, we show that a displaced worker who is located in an LLM that is a 100 per cent denser, the loss in wage is about 1.4 percentage points lower. Again, we find larger density effects on wages at higher regional aggregation levels. We do not find positive returns to agglomeration in post-displacement employment. Thus we argue that the matching mechanism of agglomeration economies is prevalent 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.

7. Conclusion

This paper assesses the role of spatial scale in measuring agglomeration externalities. We analyse the effect of employment density on wages, which is referred to as the urban wage premium, as well as the returns to agglomeration in wages and employment for workers who lost their job following firm bankruptcy.

The purpose of our paper is to examine whether the way to operationalise geographic space is important for the estimation of agglomeration externalities. The good news for existing research is that arbitrary borders of regional areas seem less important, as we yield similar results when using different sets of LLMs with the same number of distinct spatial units. However, the premise of our paper is that aggregation effects matter – using a continuum of regional aggregations we show that the agglomeration externalities on wages more than double using larger LLMs, explained by larger agglomeration externalities at a higher spatial scale. The present 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.

Our research provides new avenues for future research and gives a deeper understanding of the spatial scale of workers' LLM and of agglomeration externalities, which from a policy perspective is relevant for multiple socio-economic reasons. First, 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 such as female workers and low-educated workers, compared to policies directed at other subgroups of workers, may be more effective if they are specific, local and decentralised. The role of the spatial scale in the efficiency of policies targeted at subgroups of the population is a potential area for future research. Second, our research suggests that positive agglomeration externalities, based on the localised matching, sharing and learning mechanisms, are prevalent at a high spatial scale. This suggests urban and regional policies to increase agglomeration benefits and regional productivity growth should tend to be generic and centralised, such as city-region cooperation and geographical upscaling of economic activities. Third, we find that a dense LLM provides economic value as it leads to smaller wage losses after job loss, but not to variation in the losses in employment. This finding is relevant for labour market policies that aim to increase the matching quality of worker to employer or limit wage inequality following negative employment shocks (Moretti, 2011; Crépon and Van den Berg, 2016).

Acknowledgements

We wish to thank seminar participants at the 9th European Meeting of the Urban Economics Association, 2018 SMU Conference on Urban and Regional Economics, 13th Meeting of the Urban Economics Association, the Microeconomic Studies Function of the Federal Reserve Bank of New York, The University of Melbourne, the 32nd Annual Conference of the European Society for Population Economics and Utrecht University School of Economics. We also thank Bram van Besouw, Maarten Goos, Henri de Groot, Ross Hickey, Eric Mengus, Jos van Ommeren, Frank van Oort, Roberto Pinheiro, Jørn Rattsø, Vincent Schippers and Coen Teulings for their insightful comments. We are grateful to Statistics Netherlands for giving us access to the administrative data. This article is based on a chapter in the dissertation of Meekes (2019).

References

Ahlfeldt, G. M. and E. Pietrostefani (2019). The economic effects of density: A synthesis. *Journal of Urban Economics 111*, 93–107.

Andersson, M., J. Klaesson, and J. P. Larsson (2016). How local are spatial density externalities? Neighbourhood effects in agglomeration economies. *Regional Studies 50*(6), 1082–1095.

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

Autor, D. H. (2019). Work of the past, work of the future. *AEA Papers and Proceedings 109*, 1–32.

Autor, D. H., D. Dorn, and G. H. Hanson (2013). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review 103*(6), 2121–2168.

Autor, D. H., D. Dorn, and G. H. Hanson (2015). Untangling trade and technology: Evidence from local labour markets. *The Economic Journal 125*(584), 621–646.

Barbanchon, T. L., R. Rathelot, and A. Roulet (2019). Gender differences in job search: Trading off commute against wage. *Mimeo*.

Bingley, P. and A. Martinello (2017). Measurement error in income and schooling and the bias of linear estimators. *Journal of Labor Economics 35*(4), 1117–1148.

Bleakley, H. and J. Lin (2012). Thick-market effects and churning in the labor market: Evidence from US cities. *Journal of Urban Economics 72*(2), 87–103.

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

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

Burger, M. J., F. G. Van Oort, and G. A. Van der Knaap (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.

CBS (2019). Statistics Netherlands, https://www.cbs.nl/en-gb.

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

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

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

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

Dauth, W., S. Findeisen, E. Moretti, and J. Suedekum (2018). Matching in Cities. NBER Working Paper 25227.

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

De La Roca, J. and D. Puga (2017). Learning by working in big cities. *Review of Economic Studies 84*(1), 106–142.

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

Di Addario, S. and E. Patacchini (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 T. Watanabe, I. Uesugi, and A. Ono (Eds.), *The Economics of Interfirm Networks*, Number 4 in Advances in Japanese Business and Economics, pp. 107–133. Springer Japan.

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

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

Eeckhout, J., R. Pinheiro, and K. Schmidheiny (2014). Spatial Sorting. *Journal of Political Economy 122*(3), 554–620.

Ellison, G., E. L. Glaeser, and W. R. Kerr (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review 100*(3), 1195–1213.

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

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

Foote, A., M. J. Kutzbach, and L. Vilhuber (2017). Recalculating - How uncertainty in local labor market definitions affects empirical findings. CES Working Paper 17-49, Center for Economic Studies, U.S. Census Bureau.

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

Fowler, C. S., L. Jensen, and D. Rhubart (2018). Assessing U.S. labor market delineations for containment, economic core, and wage correlation. `https://doi.org/10.17605/OSF.IO/T4HPU`.

Gathmann, C., I. Helm, and U. Schönberg (2018). Spillover effects of mass layoffs. *Journal of the European Economic Association, In Press.*

Gaubert, C. (2018). Firm sorting and agglomeration. *American Economic Review 108*(11), 3117–3153.

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

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

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

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

Helm, I. (2019). National industry trade shocks, local labor markets, and agglomeration spillovers. Mimeo.

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

Hirsch, B., E. J. Jahn, A. Manning, and M. Oberfichtner (2019). The urban wage premium in imperfect labour markets. CEP Discussion Paper 1608.

Håkansson, J. and G. Isacsson (2019). The spatial extent of agglomeration economies across the wage earnings distribution. *Journal of Regional Science 59*(2), 281–301.

Huttunen, K., J. Møen, and K. G. Salvanes (2011). How destructive is creative destruction? Effects of job loss on job mobility, withdrawal and income. *Journal of the European Economic Association 9*(5), 840–870.

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

Iacus, S. M., G. King, and G. Porro (2011). Multivariate matching methods that are monotonic imbalance bounding. *Journal of the American Statistical Association 106*(493), 345–361.

Ichino, A., G. Schwerdt, R. Winter-Ebmer, and J. Zweimüller (2017). Too old to work, too young to retire? *Journal of the Economics of Ageing 9*, 14–29.

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993). Earnings losses of displaced workers. *American Economic Review 83*(4), 685–709.

Kerr, W. R. and S. D. Kominers (2015). Agglomerative forces and cluster shapes. *The Review of Economics and Statistics 97*(4), 877–899.

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

Manning, A. (2010). The plant size-place effect: Agglomeration and monopsony in labour markets. *Journal of Economic Geography 10*(5), 717–744.

Manning, A. and B. Petrongolo (2017). How local are labor markets? Evidence from a spatial job search model. *American Economic Review 107*(10), 2877–2907.

Marinescu, I. and R. Rathelot (2018). Mismatch unemployment and the geography of job search. *American Economic Journal: Macroeconomics 10*(3), 42–70.

Matano, A. and P. Naticchioni (2012). Wage distribution and the spatial sorting of workers. *Journal of Economic Geography 12*(2), 379–408.

Meekes, J. (2019). *Local labour markets, job displacement and agglomeration economies*. Ph. D. thesis, Utrecht University, Utrecht.

Meekes, J. and W. H. J. Hassink (2018). flowbca: A flow-based cluster algorithm in Stata. *Stata Journal 18*(3), 564–584.

Meekes, J. and W. H. J. Hassink (2019). The role of the housing market in workers' resilience to job displacement after firm bankruptcy. *Journal of Urban Economics 109*, 41–65.

Mion, G. and P. Naticchioni (2009). The spatial sorting and matching of skills and firms. *Canadian Journal of Economics/Revue Canadienne d'Économique 42*(1), 28–55.

Moretti, E. (2011). Chapter 14 - Local labor markets. In D. Card and O. Ashenfelter (Eds.), *Handbook of Labor*

Economics, Volume 4, pp. 1237–1313. Elsevier.

Murata, Y., R. Nakajima, R. Okamoto, and R. Tamura (2014). Localized knowledge spillovers and patent citations: A distance-based approach. *The Review of Economics and Statistics 96*(5), 967–985.

Neffke, F. M. H., A. Otto, and C. Hidalgo (2018). The mobility of displaced workers: How the local industry mix affects job search. *Journal of Urban Economics 108*, 124–140.

Neumark, D. and H. Simpson (2015). Chapter 18 - Place-based policies. In G. Duranton, J. V. Henderson, and W. C. Strange (Eds.), *Handbook of Regional and Urban Economics*, Volume 5, pp. 1197–1287. Elsevier.

Nimczik, J. S. (2018). Job mobility networks and endogenous labor markets. *Mimeo*.

OECD (2016). The Metropolitan Region of Rotterdam-The Hague, Netherlands. OECD Territorial Reviews.

Openshaw, S. and P. Taylor (1979). A million or so correlation coefficients, three experiments on the modifiable areal unit problem. In N. Wrigley (Ed.), *Statistical applications in the spatial science*, pp. 127–144. London: Pion.

Petrongolo, B. and C. Pissarides (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.

Proost, S. V. and J.-F. Thisse (2019). What can be learned from spatial economics? *Journal of Economic Literature 57*(3), 575–643.

Rice, P., A. J. Venables, and E. Patacchini (2006). Spatial determinants of productivity: Analysis for the regions of Great Britain. *Regional Science and Urban Economics 36*(6), 727–752.

Roberts, J., R. Hodgson, and P. Dolan (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. and W. C. Strange (2001). The determinants of agglomeration. *Journal of Urban Economics 50*(2), 191–229.

Rosenthal, S. S. and W. C. Strange (2003). Geography, industrial organization, and agglomeration. *Review of Economics and Statistics 85*(2), 377–393.

Rosenthal, S. S. and W. C. Strange (2004). Chapter 49 - Evidence on the nature and sources of agglomeration economies. In J. V. Henderson and J.-F. Thisse (Eds.), *Handbook of Regional and Urban Economics*, Volume 4, pp. 2119–2171. Elsevier.

Rosenthal, S. S. and W. C. Strange (2008). The attenuation of human capital spillovers. *Journal of Urban Economics 64*(2), 373–389.

Sahin, A., J. Song, G. Topa, and G. L. Violante (2014). Mismatch unemployment. *American Economic Review 104*(11), 3529–64.

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 labour supply panel 1985–2012. DANS, `https://doi.org/10.17026/dans-x8q-46p7`.

Small, K. A. (2012). Valuation of travel time. *Economics of Transportation 1*(1), 2–14.

Verstraten, P., G. Verweij, and P. J. Zwaneveld (2019). Complexities in the spatial scope of agglomeration economies. *Journal of Regional Science 59*(1), 29–55.

Wheeler, C. H. (2001). Search, sorting, and urban agglomeration. *Journal of Labor Economics 19*(4), 879–899.

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.

List of Tables

Table 1

Percentage of UWP estimates that are higher with placebo LLMs than with aggregate LLMs (Eq. (2)).

	OLS	FE
	(1)	(2)
K=7	44%	53%
K=10	35%	42%
K=13	11%	5%
K=16	19%	12%
K=19	10%	1%
K=22	13%	4%
K=25	9%	1%
K=30	6%	0%
K=35	6%	1%
K=40	1%	0%
K=45	2%	0%
K=50	1%	0%
K=100	0%	3%
Number of observations	18,882,294	18,882,294

Notes: The dependent variable is the natural logarithm of hourly wage. Each column gives the estimator. K represents the number of distinct LLMs used to operationalise workers' LLM. For each K, 100 different sets of placebo LLMs were randomly defined. The percentage that is provided represents the share of the 100 different placebo LLM UWP estimates that are higher than the corresponding aggregate LLM UWP estimate. The aggregate LLMs are identical to the sets of aggregate LLMs used in Figure 4. The number of regressions with placebo LLMs on which this table is based is equal to 2,600. See Figure 4 for additional notes.

Table 2

Displacement effects on employment and hourly wage (Eq. (3)).

	Employment (=1)	Hourly wage (log)
	(1)	(2)
DISPLACED × *POST*	-0.2279^{***}	-0.0625^{***}
	(0.0038)	(0.0027)
Number of parameters	150	150
Number of individuals	23,992	23,992
Number of observations	1,319,560	1,173,835

Notes: Columns (1) and (2) give the effects on employment and wages, respectively. The parameter estimates of the interaction term between *DISPLACED* and *POST* are reported. Standard errors are clustered by individual and provided in parentheses. The reference categories of *DISPLACED* and *POST* consist of the non-displaced workers and pre-displacement period, respectively. *** corresponds to the significance level of 1%. The regression analyses include individual-specific fixed effects, aggregate home LLM fixed effects (34) 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. Monthly data are used and the period under observation is from January 2006 to December 2014. Workers are observed for 18 months before until 36 months after the month of job displacement. The month of job displacement refers to the actual and potential month of job loss for the displaced workers and non-displaced workers, respectively.

List of Figures

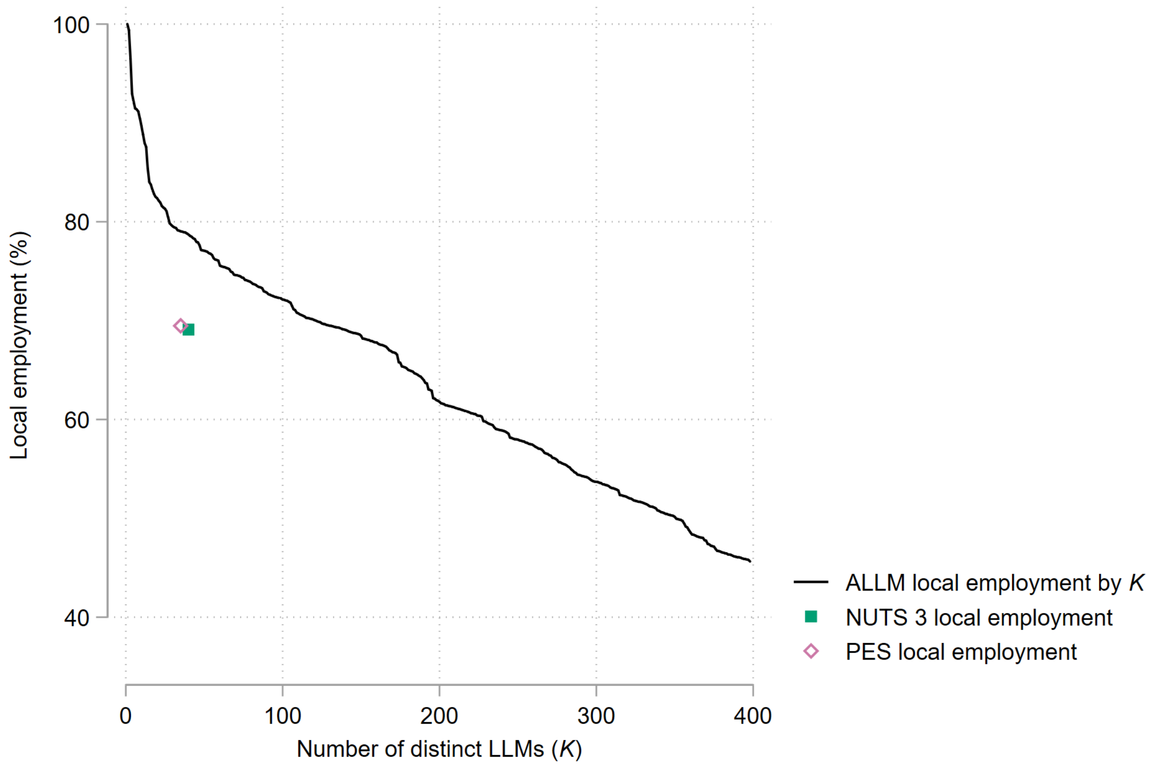


Fig 1 Local employment by regional classification. *Notes:* The local employment rate is constructed by taking the number of workers who live and work in their LLM relative to the total number of workers. The number of distinct LLMs (K) decreases by one after each iteration of the algorithm. In each iteration, starting from a set of 398 distinct municipalities, the cluster algorithm selects the spatial 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.

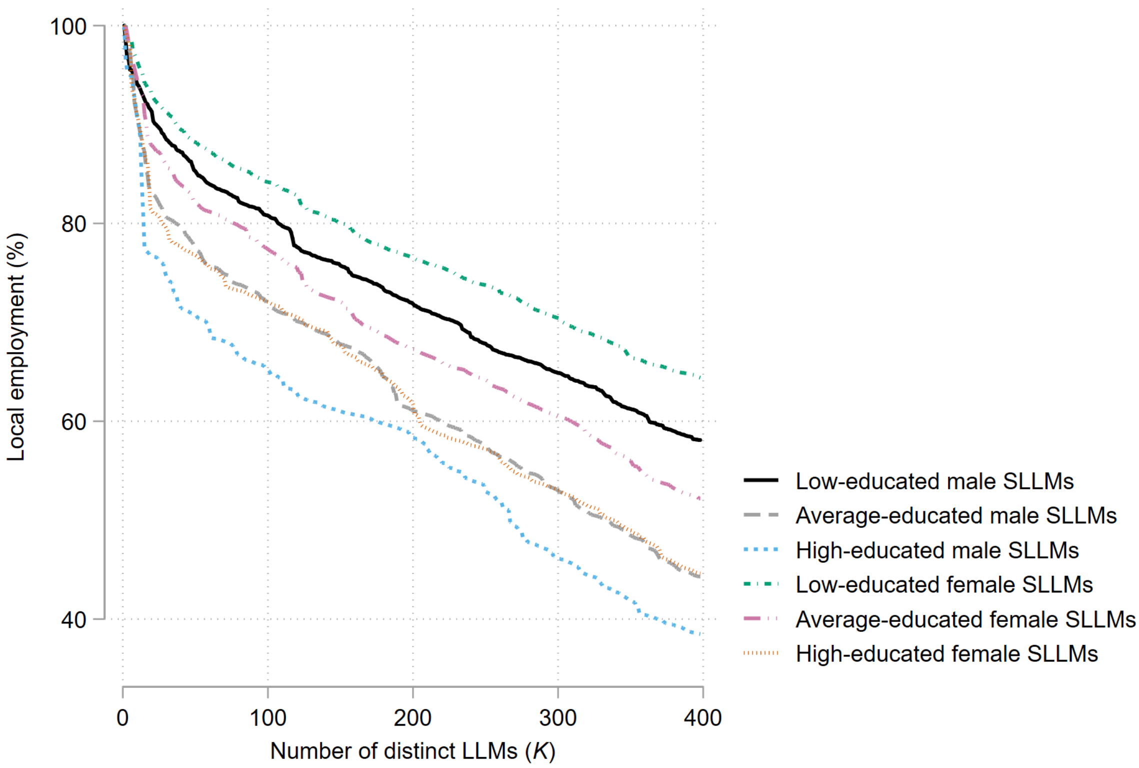
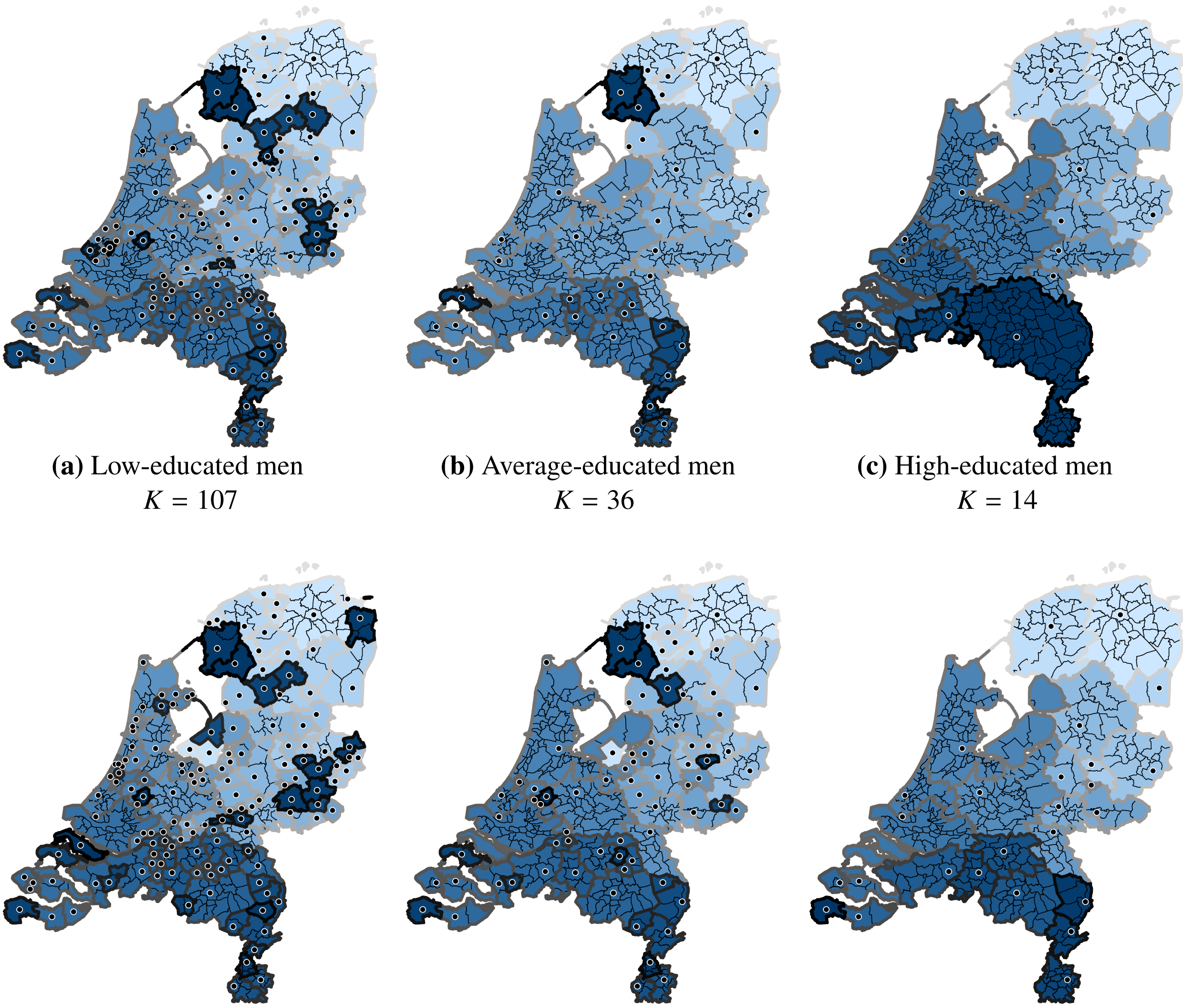


Fig 2 Subgroup-specific local employment by regional classification. *Notes:* Local employment by subgroups and regional classifications. See Figure 1 for additional notes.



(a) Low-educated men
K = 107

(b) Average-educated men
K = 36

(c) High-educated men
K = 14

(d) Low-educated women
K = 151

(e) Average-educated women
K = 76

(f) High-educated women
K = 26

Fig 3 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 different shade of blue.

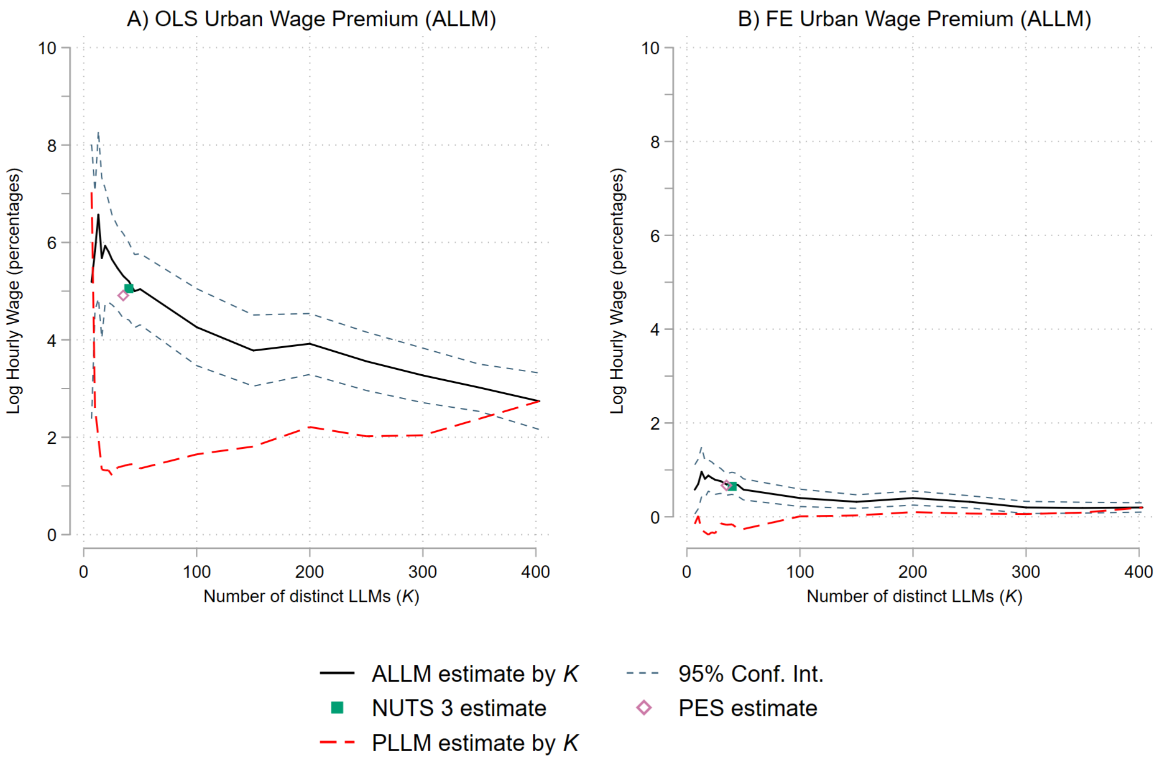


Fig 4 Aggregate LLM UWP and placebo LLM UWP by empirical specification (Eq. (2)). *Notes:* The dependent variable is the natural logarithm of hourly wage. Each estimate of the effect of the natural logarithm of employment density on hourly wage represents a different regression. In each regression, the variables employment density and area size are operationalised based on a different regional classification. K represents the number of distinct LLMs used to operationalise workers' LLM. The values of K 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 aggregate LLMs (ALLMs) and placebo LLMs (PLLMs) are defined by iteratively aggregating the two spatial units characterised by, respectively, the highest and the lowest non-zero directed relative commuting flow. 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 the indicator variables 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. The period under observation is from 2006 to 2014. The number of individual-year observations equals 18,882,294.

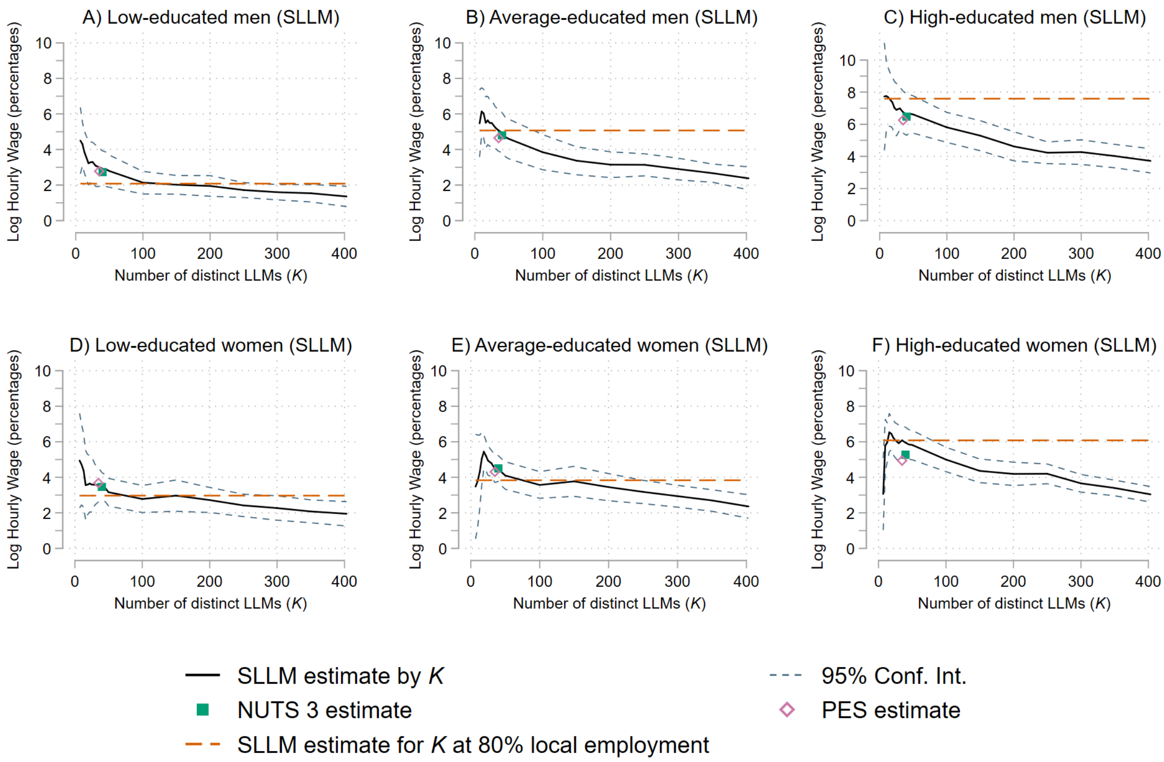


Fig 5 Subgroup-specific LLM UWP based on OLS estimates (Eq. (2)). *Notes:* The orange dashed line shows the subgroup-specific LLM estimate for the number of distinct LLMs (K) at which the subgroup-specific local employment rate equals 80 per cent. This holds for K equal to 107, 36, 14, 151, 76 and 26 for the subgroups of low-educated men, average-educated men, high-educated men, low-educated women, average-educated women, high-educated women, respectively. 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 5A-5F equals 2,296,052; 5,400,850; 4,479,115; 864,968; 2,643,962; 3,197,347, respectively. See Figure 4 for additional notes.

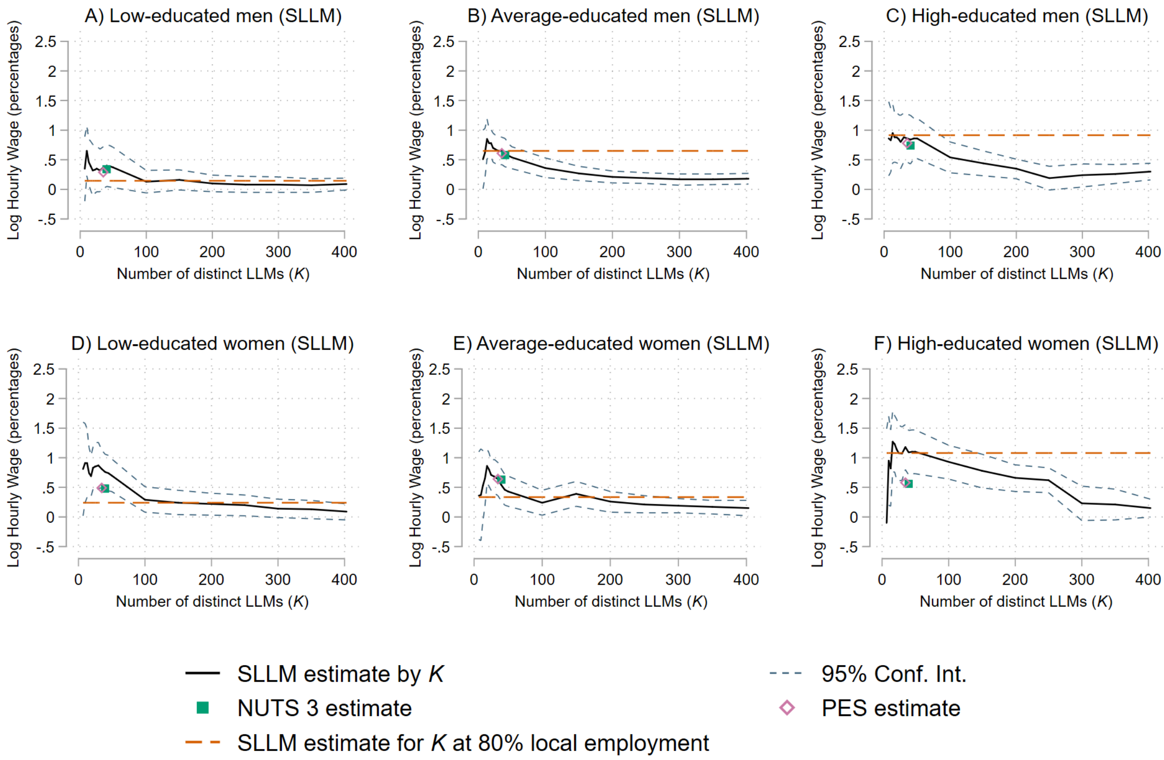


Fig 6 Subgroup-specific LLM UWP based on FE estimates (Eq. (2)). *Notes:* See Figures 4 and 5 for additional notes.

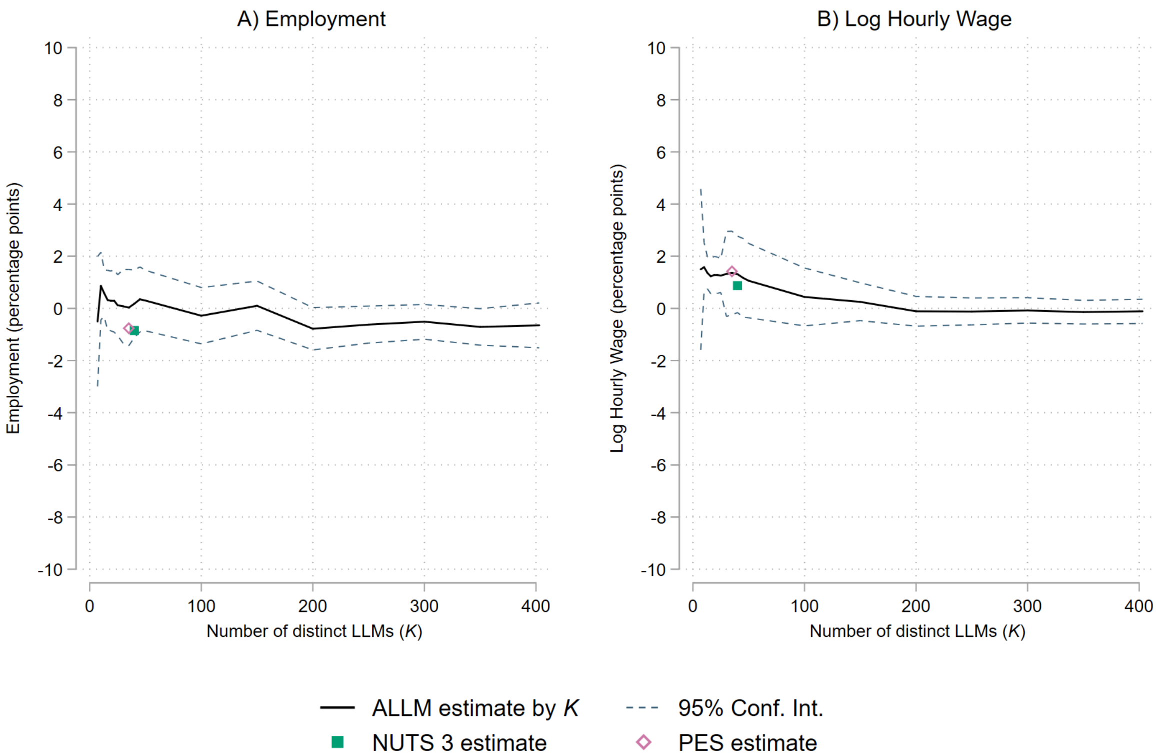


Fig 7 Aggregate LLM displacement effects on employment and wages (Eq. (4)). *Notes:* Figures 7A and 7B represent regressions of employment and the natural logarithm of hourly wage, respectively. Parameter estimates of the three-way interaction term, among *DISPLACED*, *POST* and *EMPLOYMENT DENSITY*, 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 operationalised based on a different regional classification. The values at which K is used to operationalise 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.

Appendices:

Appendix A Urban wage premium: 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), which is a more robust way to compute standard errors. Figure A.1 shows the estimates of the UWP using the aggregate LLMs to operationalise workers' LLM.

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

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

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

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

The results of the two-step approach are provided in Figure A.1 and Figure A.2. Compared to the direct approach of estimating the UWP (see Fig. 4), 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. Figure A.1 shows an effect of log employment density on wages of about 3 to 4 per cent for relatively large areas. 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. For Spain, De La Roca and Puga (2017) find and effect of log city size on wages of 4.5 per cent. For Germany, Dauth et al. (2018) find an effect of log population on wages of 3.7 per cent, and Hirsch et al. (2019) find an effect of log population density on wages between 3.2 and 3.6 per cent.

Figure A.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 operationalise workers' LLM leads to larger differences in the UWP between education levels than when using the subgroup-specific LLMs to operationalise 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.

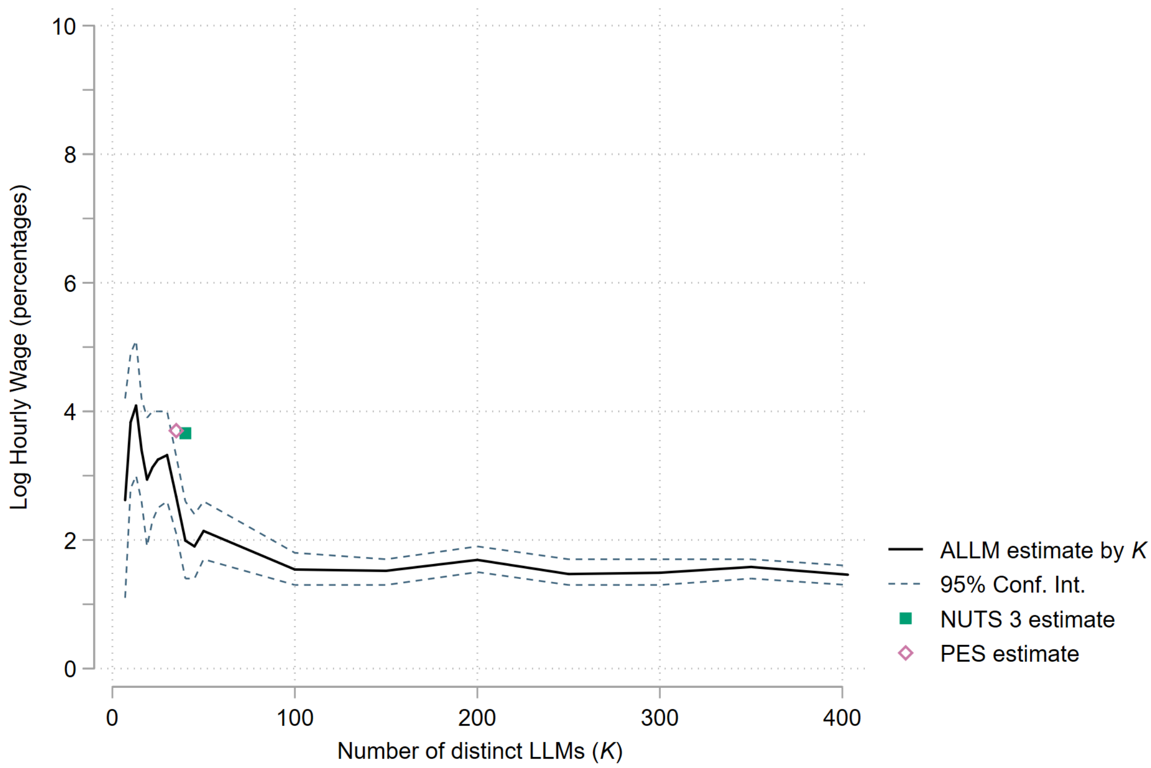


Fig A.1 Aggregate LLM UWP based on the OLS two-step procedure (Eq. (A.2)). *Notes:* Estimates of the second stage are provided. See Figure 4 for additional notes.

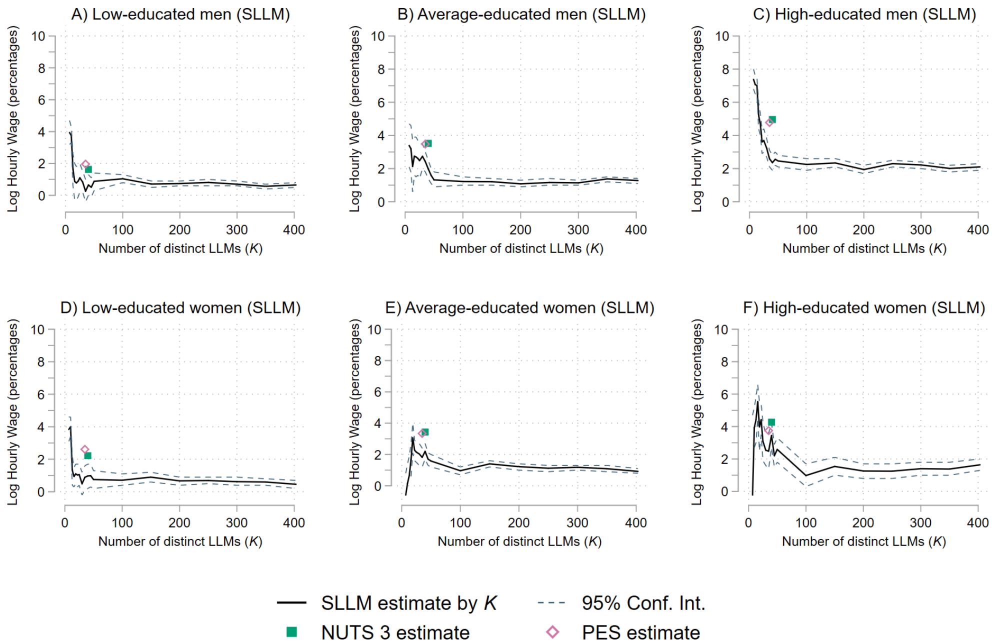


Fig A.2 Subgroup-specific LLM UWP by subgroups based on the OLS two-step procedure (Eq. (A.2)). *Notes:* Estimates of the second stage are provided. See Figures 4 and 5 for additional notes.

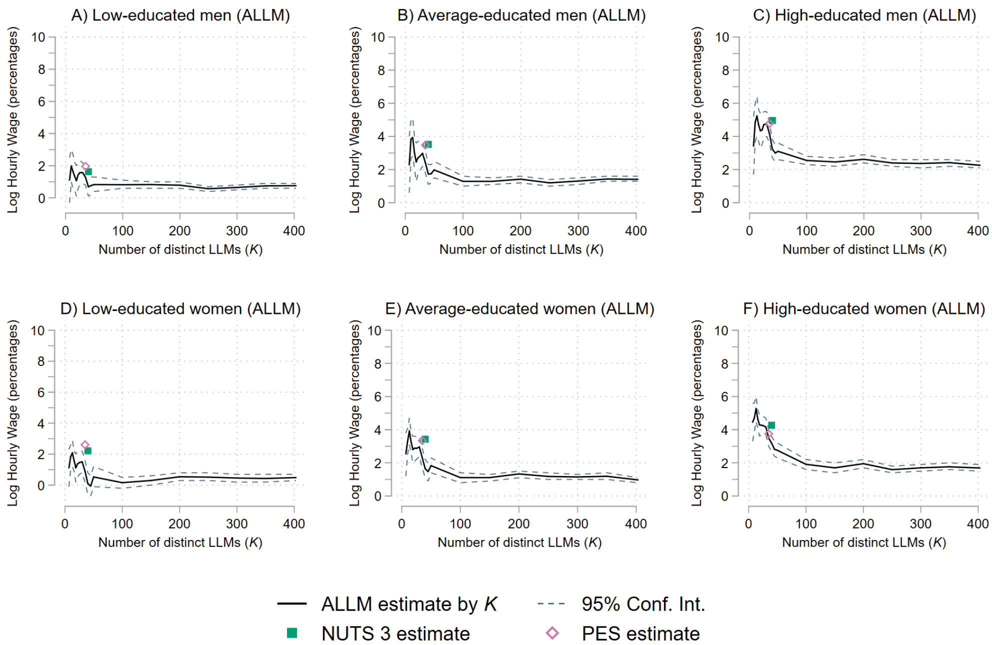


Fig A.3 Aggregate LLM UWP by subgroups based on the OLS two-step procedure (Eq. (A.2)). *Notes:* Estimates of the second stage are provided. See Figures 4 and 5 for additional notes.

Appendix B Urban wage premium: Summary statistics and robustness checks

Table B.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 B.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 B.3
Coefficients and standard errors of subgroup-specific LLM UWP based on FE estimates (Fig. 6, 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 operationalised based on the 40 NUTS 3 areas and 40 subgroup-specific LLMs, respectively. See Figure 6 for additional notes.

Table B.4
Statistics on the number of employed workers by regional classification.

	Minimum	Maximum	Median	Mean
NUTS 3 (K = 40)	13,960	753,749	109,372	170,759
PES (K = 35)	52,194	722,819	141,689	195,153
ALLM (K = 35)	9,452	1,576,821	77,836	194,855

Notes: The urban wage premium data sample. For the year 2014, summary statistics are on the number of employed workers are provided by the NUTS 3 area, PES area (35 distinct units) and aggregate LLM (35 distinct units).

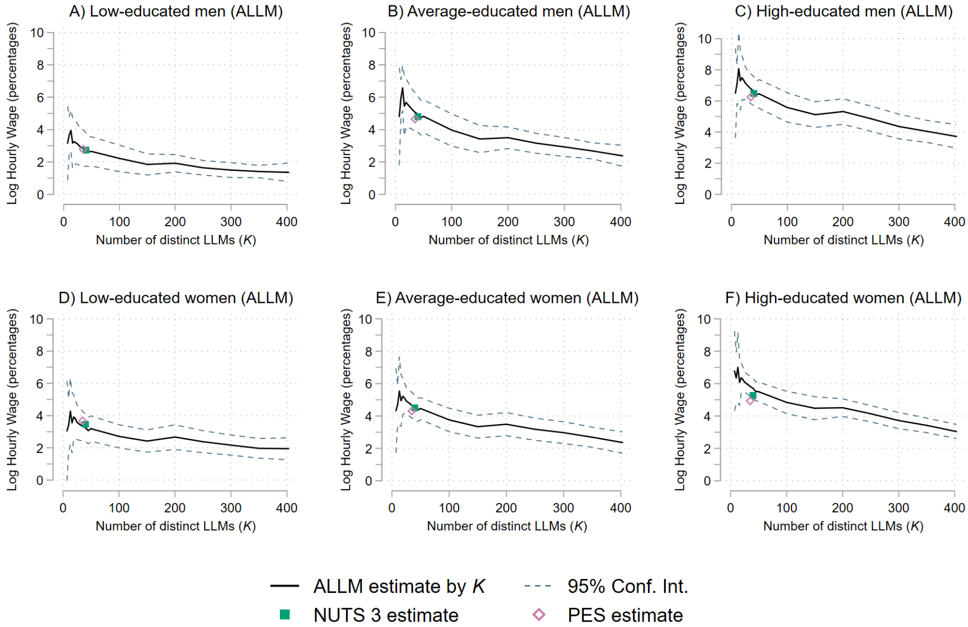


Fig B.1 Aggregate LLM UWP based on OLS estimates (Eq. (2)). *Notes:* See Fig. 5 for additional notes.

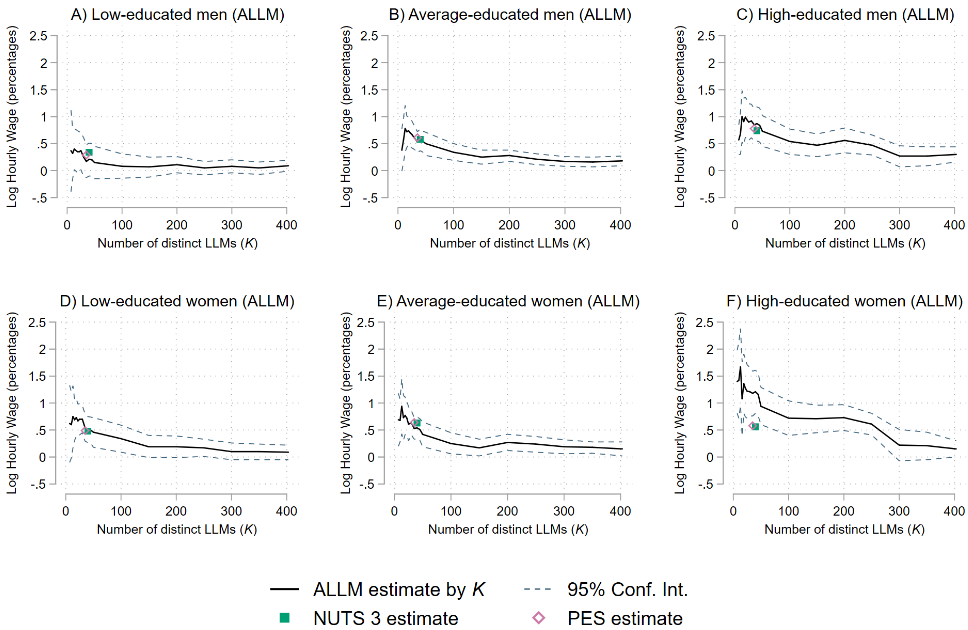


Fig B.2 Aggregate LLM UWP based on FE estimates (Eq. (2)). *Notes:* See Fig. 6 for additional notes.

Appendix C Job displacement: Summary statistics and robustness checks

Table C.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. The within change is measured by the difference in the values between the eighteenth month after job displacement and the month of job displacement.

Table C.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 are provided for the sample before CEM is applied. The statistics are provided based on observations in the period July 2007 to December 2011, for the month of potential and actual displacement of the non-displaced and displaced, respectively. Sample means with standard deviations are provided, and the t-statistic shows whether the values for the displaced workers and non-displaced workers are statistically different from each other. *** corresponds to the significance level of 1%. Note that workers are included conditional on being employed in the month of actual or potential displacement.

Table C.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 are provided for the sample after CEM is applied. The statistics are provided based on observations in the period July 2007 to December 2011, for the month of potential and actual displacement of the non-displaced and displaced, respectively. Sample means with standard deviations are provided, and the t-statistic shows whether the values for the displaced workers and non-displaced workers are statistically different from each other. ***,**,*, correspond to the significance level of 1%, 5%, 10%, respectively. Note that workers are included conditional on being employed in the month of actual or potential displacement.

Table C.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 based on observations in the month of job loss over the period July 2007 to December 2011. The group of bankrupts firms gives information on all distinct firms of which an entity is declared bankrupt. The group of non-bankrupt firms gives information on firms where the matched non-displaced workers are employed.

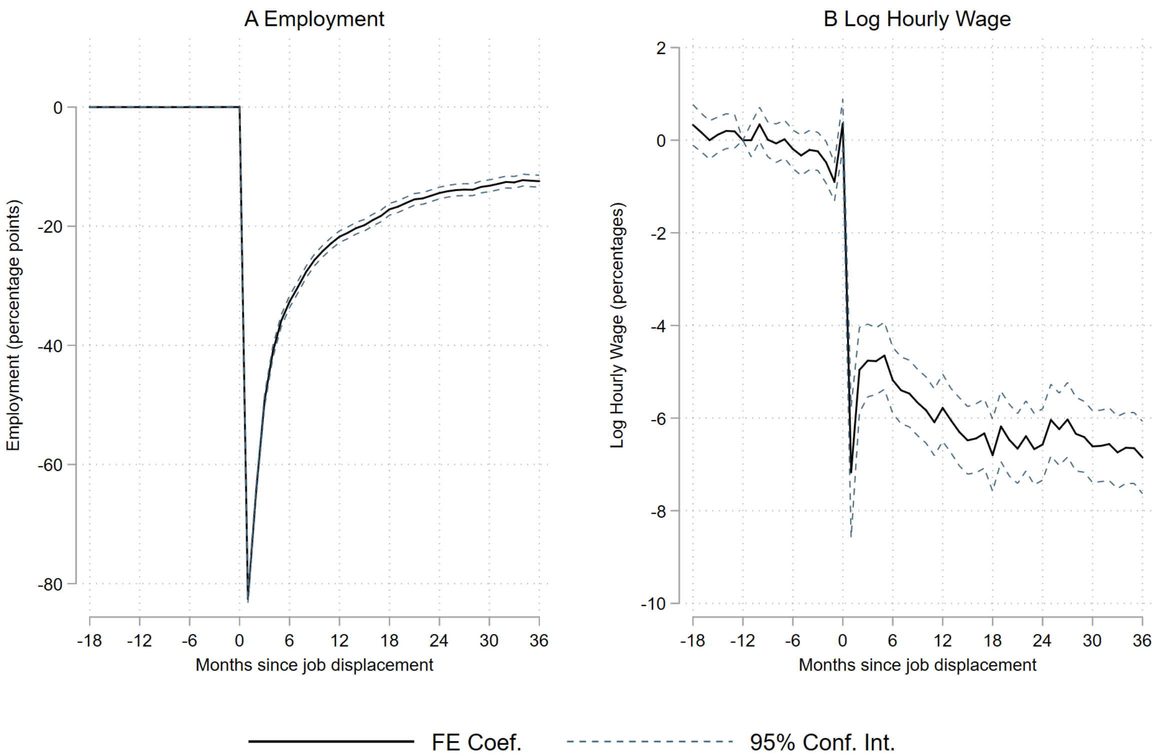


Fig C.1 Time-dependent displacement effects on employment (A) and wages (B). *Notes:* The reference group consists of the non-displaced workers and the reference month is the twelfth month before job displacement. The 95% confidence intervals are computed using clustered standard errors by individual. The two fixed effects regression models include 260 parameters including 54 two-way interaction terms. See Table 2 for additional notes and statistics.

Appendix D Flowbca

Figure D.1 shows the maximum relative commuting flow in each iteration of flowbca. Observe 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.

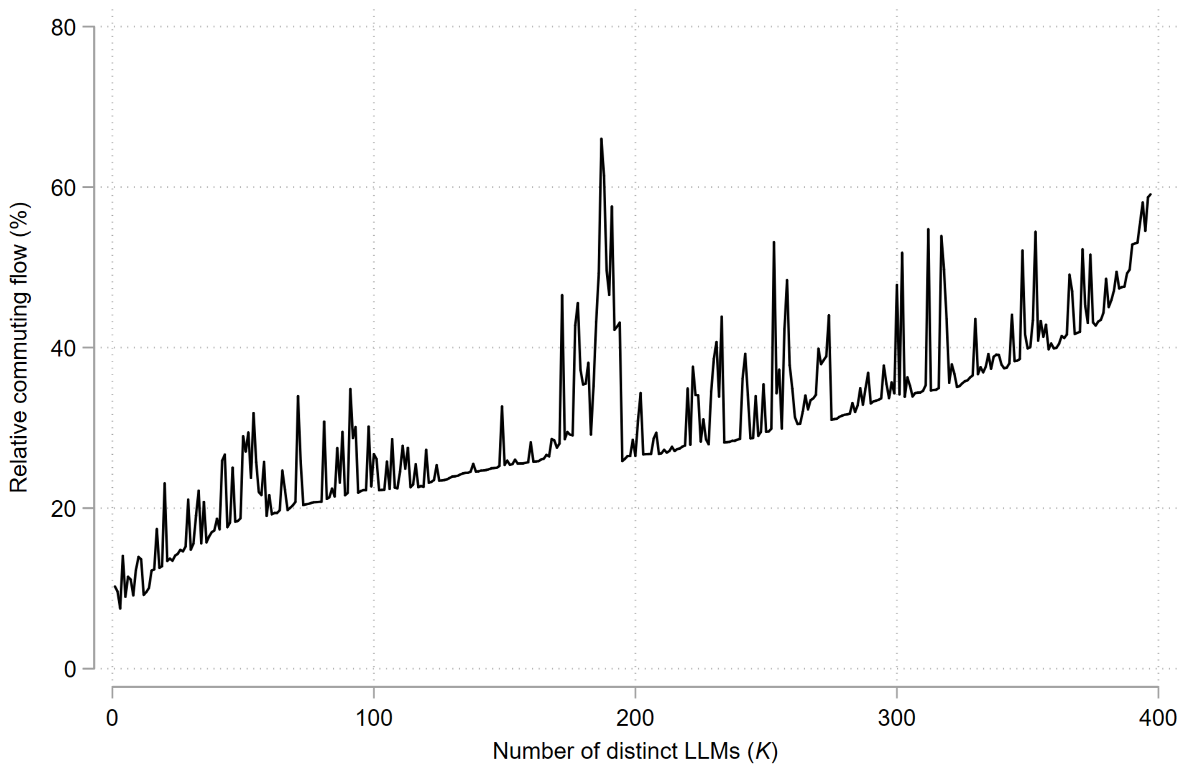


Fig D.1 Relative commuting flow at which two units are aggregated. *Notes:* See Figure 1 for additional notes.

Figure D.2 shows the maximum relative commuting flow at which spatial 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 spatial units are aggregated. This observation suggests that high-educated workers, compared to low-educated workers, work more often outside their LLM. Figure D.2 suggests that the extent to which a regional classification reflects workers' LLM strongly depends on the worker's gender and education.

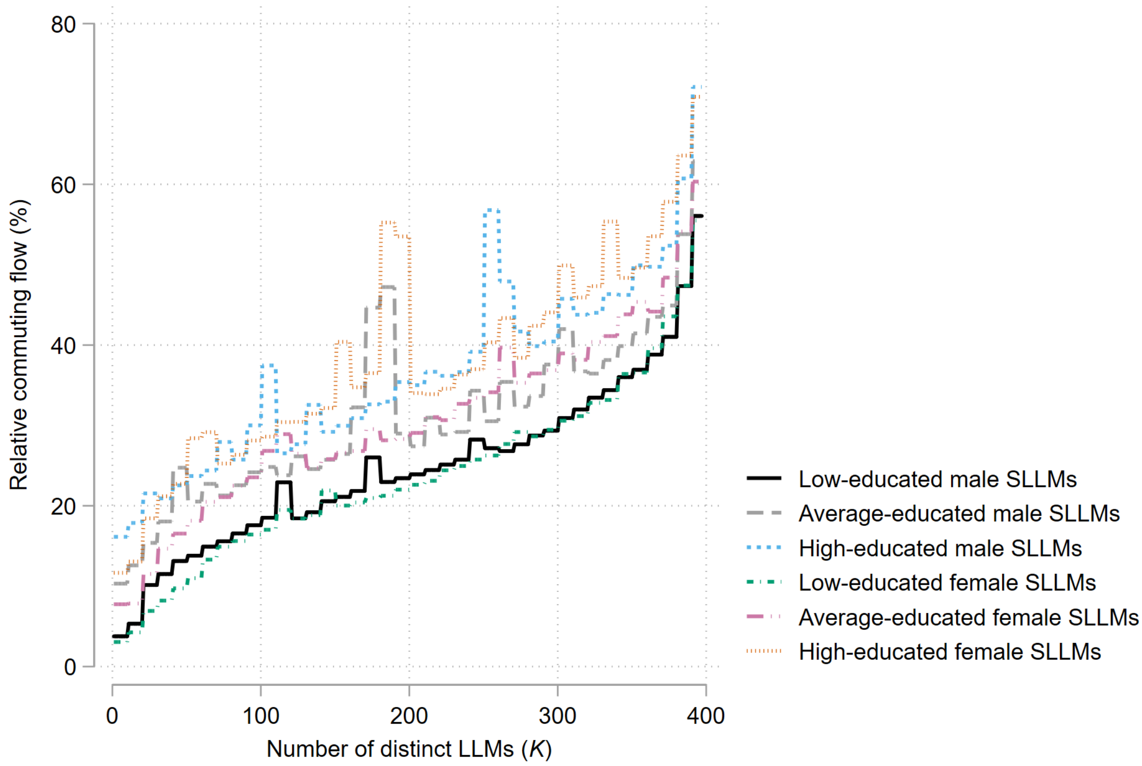


Fig D.2 Subgroup-specific relative commuting flow at which two units are aggregated. *Notes:* The median of the relative commuting flows, in increments of ten, is given to smooth out the lines and to provide visible patterns. The values of the relative commuting flow, in each iteration, are available upon request. See Figure 1 for additional notes.

Appendix E Commuting

Table E.1 provides an understanding of which worker characteristics explain the largest share of variation in workers' commuting distance. Table E.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 E.1 shows that female workers and low-educated workers are characterised by a relatively short commuting distance. Moreover, Table E.1 reveals that the estimates for gender and education, relative to other worker characteristics, are 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 spatial scale. We particularly focus on gender- and education subgroups, because these demographic characteristics explain the largest share of variation in commuting outcomes.

Figure E.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 E.1 shows that for men and women the average commuting time increased in the period from 1988 to 2014. The same pattern has been observed in other countries such as the US (Crane, 2007). The increase in commuting time is most severe for high-educated workers. Moreover, Figure E.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.

Table E.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. Sample: a five per cent random sample.

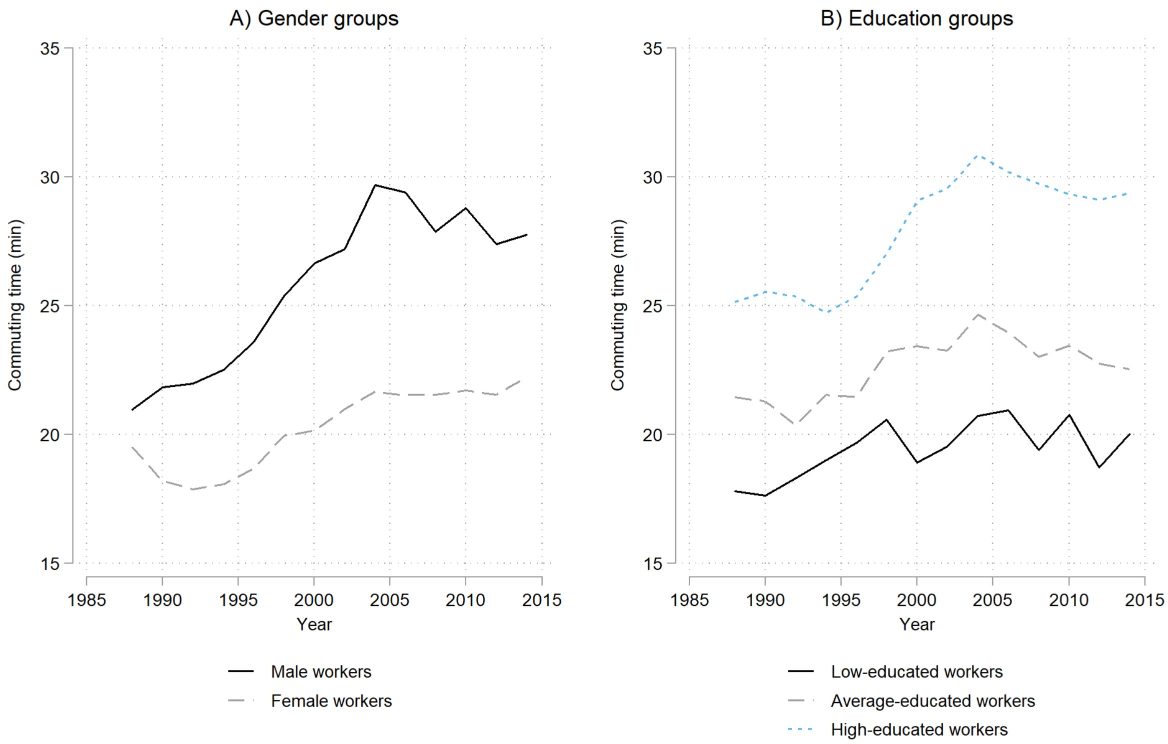


Fig E.1 Changes in the average commuting time of workers by gender and education groups over the period 1988 to 2014. *Notes:* Data set: the SCP labour supply panel. Sample size: 41,275 observations.

Figure E.2 shows the density plots of the gender shares (Fig. E.2A) and education shares (Fig. E.2B) across 398 municipalities. The shares are separately given for employed individuals in their home municipality and work municipality. Figure E.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.

The distribution of high-educated workers is relatively wide (see Fig. E.2B), which implies that high-educated workers are more concentrated in specific municipalities than low-educated workers. Moreover, Figure E.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-

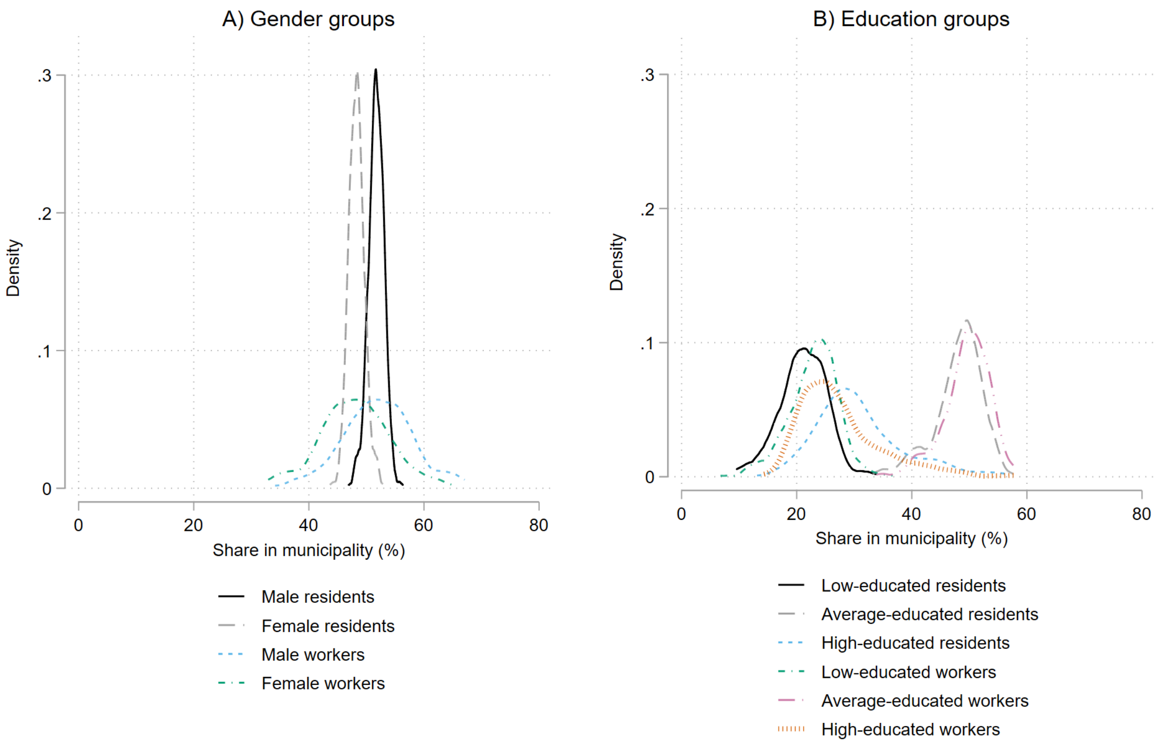


Fig E.2 Distribution plot of gender and education shares across municipalities. *Notes:* The gender and education shares are constructed by taking the subgroup-specific fraction, separately for residents and workers, in the municipality. The sample contains fractions for 398 distinct municipalities.

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 Dutch workers' commuting time has been increasing over the last decades, which is consistent with increasing commuting in other countries such as the US (Crane, 2007). 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 according to differences in gender and education.

Appendix F Subgroup-specific differences in the displacement effects

Figures F.1 and F.2 reveal the subgroup differentials in the importance of employment density for the displacement effects on employment and hourly wage, respectively. Note that in Figures F.1 and F.2, the estimates and 95 per cent confidence intervals are in some cases set at a limit of minus ten and plus ten percentage points to keep the scales of the vertical axes identical. This was especially necessary for classifications with fewer than 13 distinct LLMs, when the MAUP is most prevalent. See Table F.1 for the coefficients and standard errors of the subgroup-specific LLM displacement effects for the 40 NUTS 3 areas and 40 subgroup-specific LLMs, respectively.

Fig F.1 Subgroup-specific LLM displacement effects on employment (Eq. (4)). *Notes:* Figure F.1 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 7 for additional notes.

Figure F.1 shows that high-educated female workers experience a significant negative effect of employment density, operationalised by subgroup-specific LLMs, on post-displacement employment. Specifically, for high-educated female workers the loss in employment is about five percentage points higher in a twice as dense location. Also, we find that for low-educated female workers, using the NUTS 3 areas or the PES areas, the loss in employment is two to four percentage points lower if they reside in a geographical home location that is twice as large in terms of density.

Figure F.2 shows a significant effect of employment density, operationalised by subgroup-specific LLMs, on post-displacement wages for high-educated men and low-educated women. Both subgroups experience more modest losses in hourly wage if they reside in denser LLMs. Using the NUTS 3 or PES areas to operationalise 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 on subgroup differentials is relatively weak as the standard errors are relatively high.

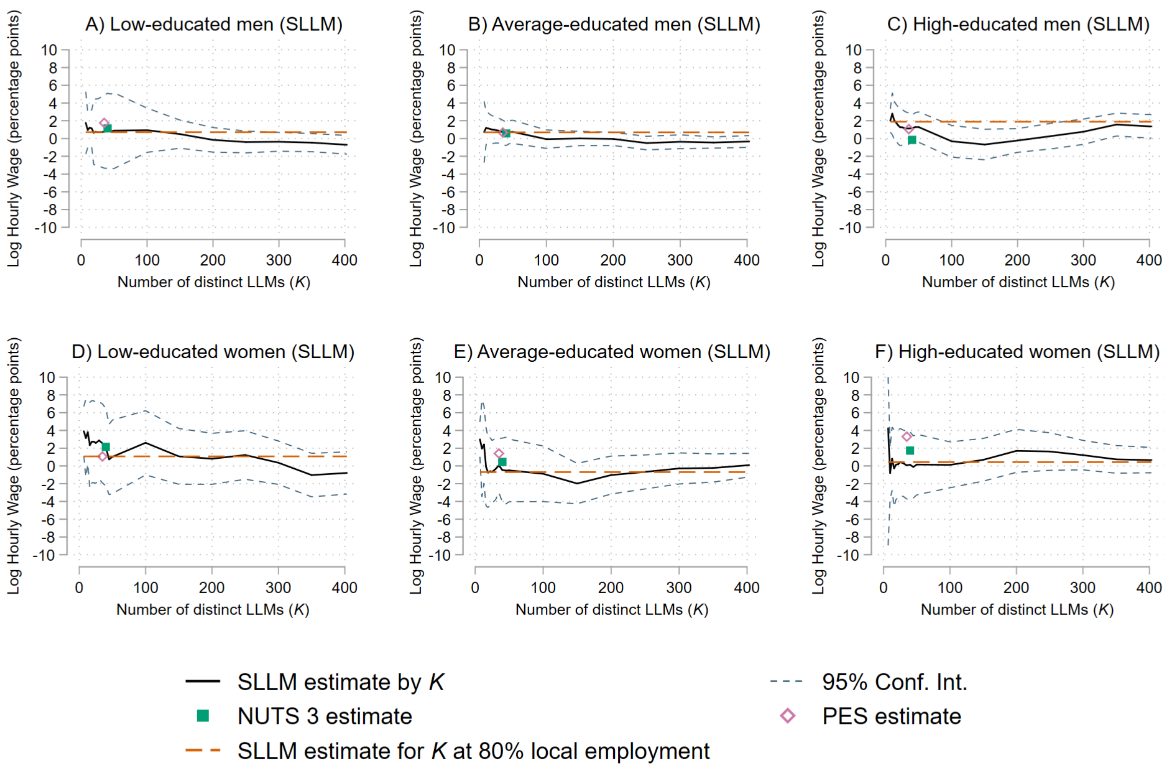


Fig F.2 Subgroup-specific LLM displacement effects on wages (Eq. (4)). *Notes:* Figure F.2 represents regressions of the natural logarithm of hourly wage. Parameter estimates of the three-way interaction term, among *DISPLACED*, *POST* and *EMPLOYMENT DENSITY*, are reported. The number of individual-month observations for the subgroups in graphs A-F equals 232,150; 550,028; 149,369; 49,727; 151,393; 72,919, respectively. See Figure 7 for additional notes.

In general, the results suggest that in more dense labour markets the loss in employment is more modest for low-educated and average-educated workers and more pronounced for high-educated workers. This observation can be explained by increased 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 effects on hourly wage for workers who have been displaced.

Table F.1

Coefficients and standard errors of subgroup-specific LLM displacement effects (Fig. F.1 and Fig. F.2, 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 operationalised based on the 40 NUTS 3 areas and 40 subgroup-specific LLMs, respectively. See Figure F.1 and Figure F.2 for additional notes.