DOI:10.4121/21438021

Data underlying the publication: The limited regional employment benefits of XXL-logistics centres in the Netherlands

Merten Nefs, Jeroen van Haaren & Frank van Oort

Paper published in the Journal of Transport Geography, May 2023
https://doi.org/10.1016/j.jtrangeo.2023.103603

Technical appendix

In this technical appendix to the paper, we explain the sources and treatment of the data. Additionally, we show in detail how the threefold research methodology was applied by introducing the formulae and algorithms that were used. The code files, (sample) data, graphical outputs and a Workflow diagram are included in the repository for the sake of reproducibility.

Data sources, selections and treatment

Sources

As described in the paper, the analysis uses two types of specific data: Lisa company microdata for the years 2000, 2010 and 2017 in the Netherlands; and a dataset of logistics buildings in the Netherlands in 2021. The former is proprietary data, which cannot be shared publicly. The latter is a largely open access dataset available on another repository (Nefs, 2022). In the analysis, however, we use an extended version of this data, containing again employment numbers (from Lisa 2020) that cannot be shared publicly.

Apart from these two datasets, we have used open statistics regarding population from the Dutch statistics office (CBS, 2000, 2010, 2017), and building geometry from OpenStreetMap (OSM, 2021).

Selections

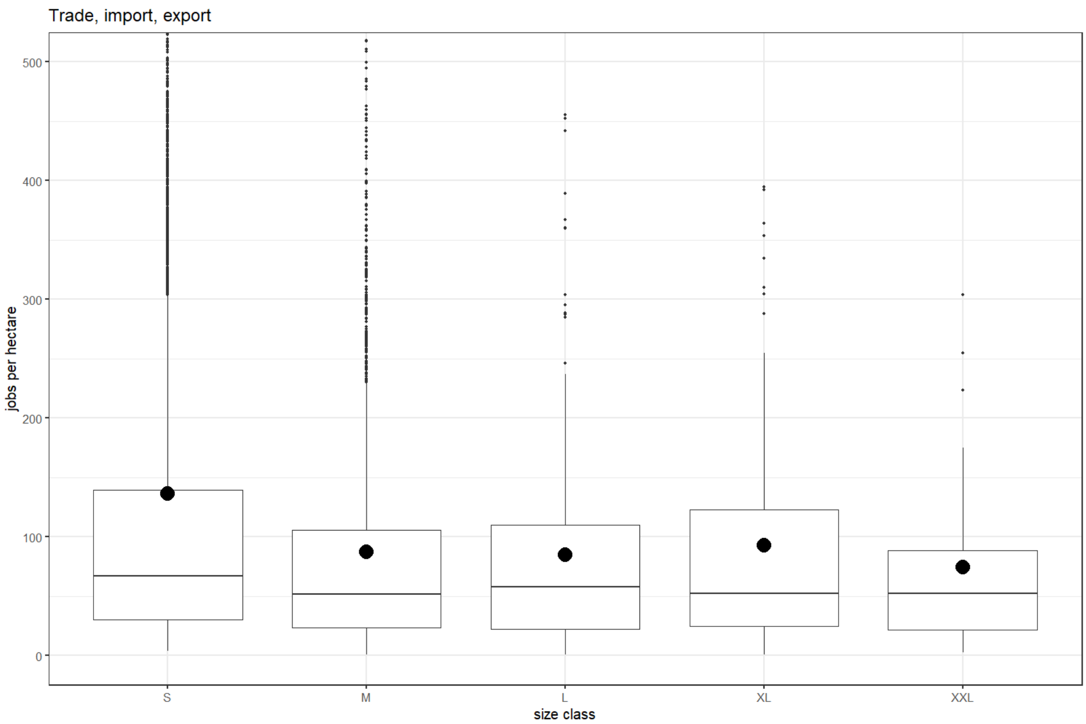
From the Lisa company microdata we have made selections of certain sectors: logistics (a broad definition including trade, import and export, and transport, warehousing and logistics services), agrifood, tech/manufacturing and the remaining (sub)sectors. The exact SBI codes for the selections are documented in the R code in the repository. The SBI code labels, similar to NAICS codes, are listed in the accompanying PDF file (SBI_codelist.pdf).

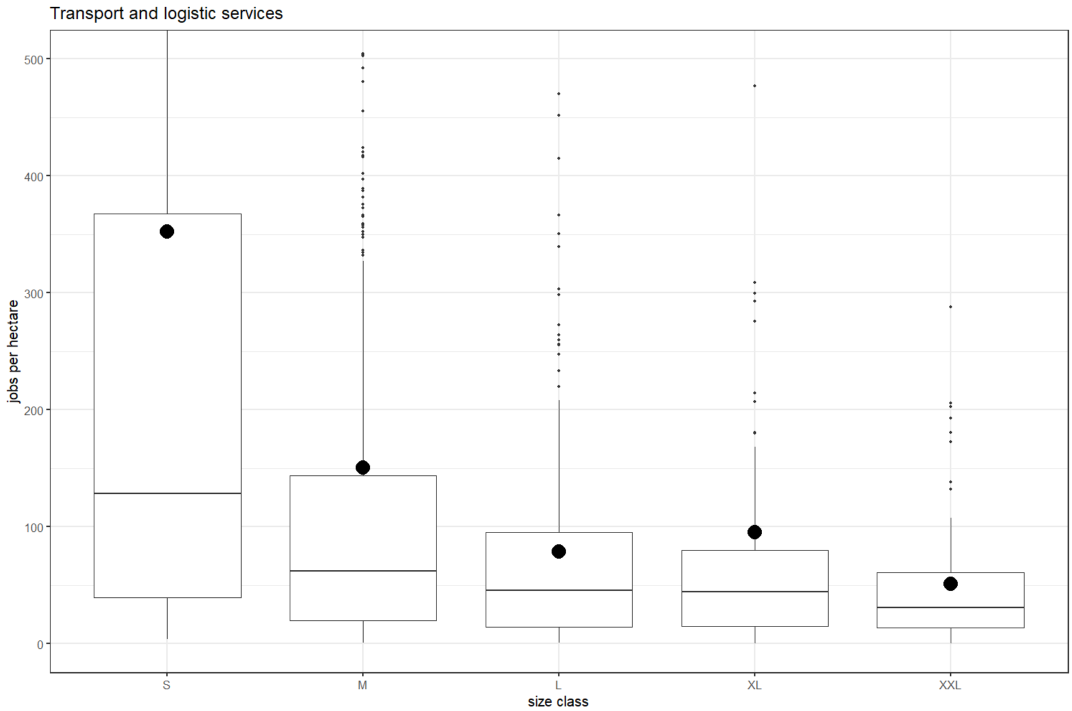
From the logistics building dataset, various selections were made. For the analyses in the paper, the logistics building type of ‘XXL retail’ (large shops / self-service warehouses, located on business estates) were excluded. A specific subset of logistics buildings in the data concerns warehouses that have been known to have a logistics function in 2017, through Lisa data, and still exist in 2021, however without a Lisa entry in 2020. For more information on this category, see the Readme file in (Nefs, 2022). We assume these buildings to be functioning in 2021, due to the overheated real estate market and low vacancy at that time. The jobs related to the buildings may be registered at a headquarters or other warehouse location, explaining the missing entry. Therefore, in analysis 1 (employment density heatmap) we include these buildings, while we exclude them in analysis 3 (shift-share).

Treatment of missing data

There are missing employment data in the logistics buildings data, for logistics buildings without a business registration - and therefore no job registration in the Lisa dataset. We filled these missing data fields by interpolation, using the mean job density per building size class and type of logistics. Job estimation was made by multiplying the job density of the relevant type and size category with the footprint of the building.

This occurs in 2814 out of the 19526 buildings included in the analysis (14%). Although it slightly increases the inaccuracy of the data, we found it important to fill in the data gaps, since many ‘no-data’ occurrences concern XXL warehouses that can generate large numbers of jobs. Leaving these buildings out of the analysis would render it irrelevant for the discussed topic. The boxplots below demonstrate the decreasing of mean employment per square meter in the two logistics types that were involved in the interpolation in the various building size classes: Trade and Logistics Services.





Methods of the 3 analyses

In this section we introduce the used formulae and algorithms of the three parts of the methodology in the paper. The code files are available in the repository.

1. Direct employment density scores

To produce the employment density heatmap, the data were treated and combined in R (see code files in the repository). Resultant employment point files – required to make a heatmap – were loaded into the Kernel Density Estimation (KDE) Heatmap algorithm (available as a QGIS plugin). This part is made available both as a QGIS graphic modeler file and an equivalent Python script in the repository. The used decay function is Quartic, with a radius of 2 km. For more information on this function, see https://www.geodose.com/2017/11/qgis-heatmap-using-kernel-density.html

The table shown in this part of the analysis, comparing various NUTS3 (corop) regions, highlights (*) three regions with DC-favouring policies. The three regions are consistent high-rankers in the Logistics Hotspot Ranking (see list in the repository). The code that produces the table is included as well.

Files

Analysis1_Heatmap_prep.R
Analysis1_corop_data_selection.R
Analysis1_Table1_regions.R
Analysis1_sectorgrowth_corridor.R
Analysis1_KDE_Heatmap_QGIS.py
Analysis1_KDE_Heatmap_QGIS.model3

2. Ellisison-Glaeser-Kerr co-agglomeration index

The EG coagglomeration index "takes a simple form when applied to industry pairs (as opposed to larger groups). The index for the coagglomeration of industries i and j is

$$\gamma_{ij}^{c} = \frac{\sum_{m=1}^{M}(s_{mi} - x_m)(s_{mj} - x_m)}{1 - \sum_{m=1}^{M} x_m^2}$$

where m indexes geographic areas. smi is the share of industry i's employment contained in area m. xm measures the aggregate size of area m, which we model as the mean employment share in the region across manufacturing industries. The Mathematical Appendix demonstrates that this index can be regarded as a measure of the strength of agglomerative forces in a particular model of firm location." (Ellison et al., 2010)

The application of the EGI is done in Strata (see code in repository). Co-agglomeration analysis run on the NUTS3 regional level produced similar results as the municipal level.

Files

Analysis2_Coagglomeration_Stata.do
Analysis2_coag_graphs.R

3. Shift-share analysis

This part of the analysis is performed according to the formulae below, and applied in the R code available in the repository. The code includes the plotting of four bar charts pertaining to the different sectors analysed.

$RS_{it} = G_{it}^{obs} - G_{it}^{exp}$

$G_{it}^{exp} = G_{nat,t}^{Obs} + G_{ind,t}^{Obs}$

Where:

$RS_{it} = Regionalshare, competitive\ effect\ of region\ i\ at\ time\ t$

$G_{it}^{obs} = Observed\ growth\ rate\ in\ region\ i\ at\ time\ t$

$G_{it}^{exp} = Expected\ growth\ ratein\ region\ i\ at\ time\ t$

$G_{nat,t}^{Obs} = Observed\ growth\ ratein\ the\ national\ economy\ at\ time\ t$

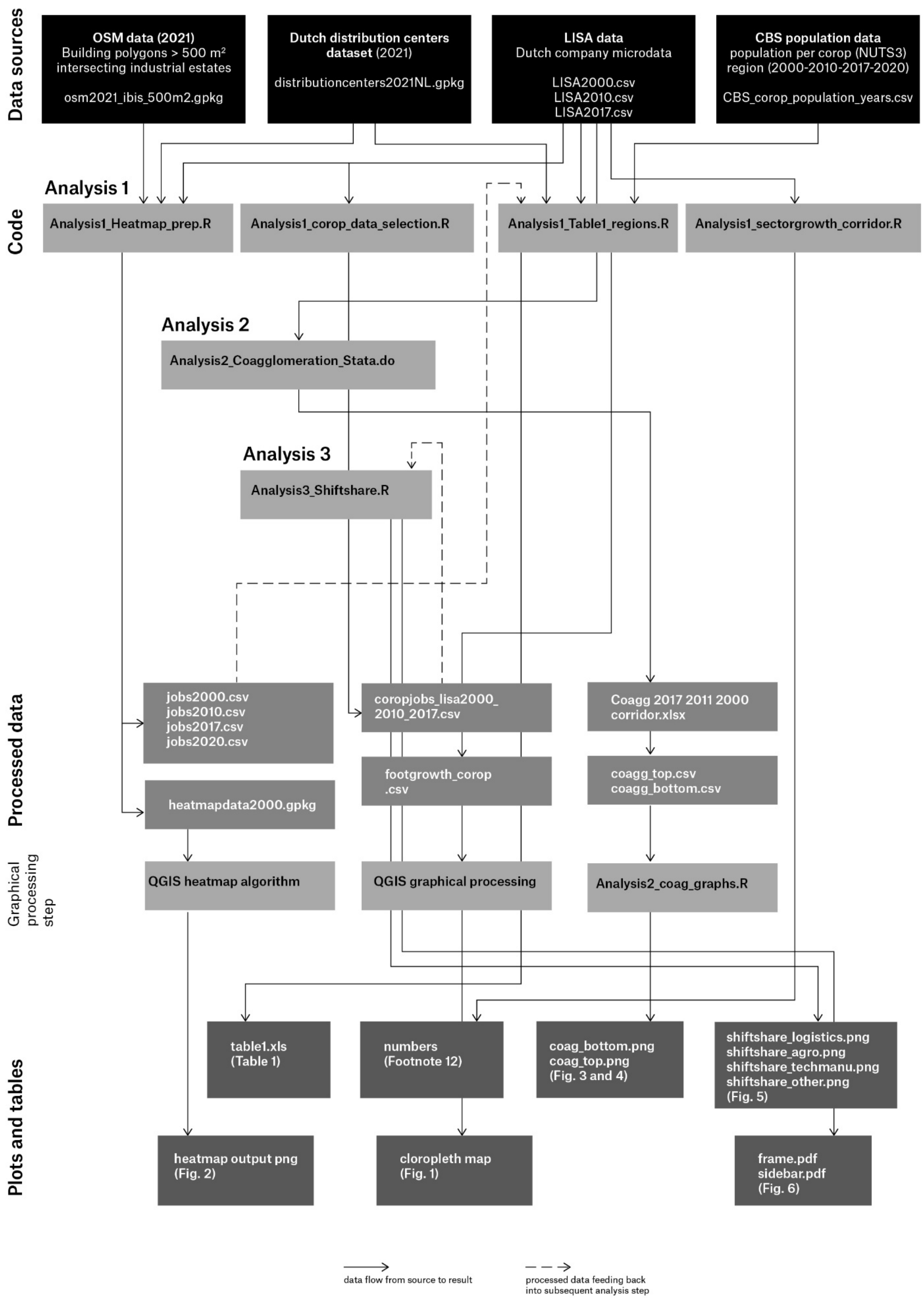
$G_{ind,t}^{Obs} = Observed\ growth\ ratesin\ the\ underlying\ industry\ mix\ at\ time\ t$

The code of this part includes the evaluation diagram.

Files

Analysis3_Shiftshare.R

Workflow diagram



References

Ellison, G., Glaeser, E. L., & Kerr, W. R. (2010). What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, *100*(3), 1195–1213. https://doi.org/10.1257/aer.100.3.1195

Nefs, M. (2022). *Dutch Distribution Centres 2021 Geodata*. 4TU.ResearchData. https://doi.org/10.4121/19361018.v1