Appendix: Why cities must drive growth in the EU's single market

Our **gravity model** is constructed using inter-regional EU services and goods trade, from an experimental set of statistics constructed by Siyu Huang and Pantellis Koutrompis at Oxford University.[1] The regions are at the NUTS 2 level. The dependent variables are Eurostat data on regional GDP and population, and Eudistance data on regions' distance from each other.[2] Separate models are constructed for goods and services exports, and for the 2008 and 2018 years. Formally, our model is:

The log of each region's exports = constant + β log of exporting region's GDP + β log of exporting region's population + β log of importing region's GDP + β log of importing region's population + β log distance between exporting and importing region + β exporting country fixed effects + β importing country fixed effects + error.

All coefficients are statistically significant at the 0.01 level, apart from several exporting and importing country fixed effects.

For the data, regression tables and Stata code, please contact the authors.

Our **PCA methodology** condenses our large dataset of regional characteristics and their services and goods exports with many variables into a smaller set of patterns, known as 'components', using machine learning. The PCA has two goals: first, to identify patterns and clusters of regions with similar characteristics and, second, to determine which characteristics correlate with higher trade in goods and services.

We applied PCA to a dataset constructed from Eurostat data and the European Commission's competitiveness index from 2013, and 2019. The competitiveness indices from 2013 and 2019 are 'backward-looking', with observations from 2009/2010 and 2017/2018, aligning with our regional trade data. Apart from the trade data, the merged dataset includes variables for European regions (at the NUTS2 level) across seven categories:

★ **Infrastructure:** Railway connections, accessibility to flights, motorway density.

★ **Institutions:** Perceptions of the rule of law, corruption, and effectiveness of government services.

★ **Connectivity**: The proportion of households with broadband access, proportion of individuals who can order goods online.

★ **Education:** The proportion of the population with tertiary education.

★ **Innovation:** including patent applications, employment in technology and knowledge sectors, scientific publications, creative employment.

★ **Business Sophistication:** for example, employment in professional, scientific, and technical activities as a share of total employment; small and medium sized firms with innovation cooperation as a share of total firms.

1: Siyu Huang and Pantellis Koutrompis, 'European multi-regional input-output data for 2008-2018', 2023.

2: Eurostat, GDP of NUTS 2 regions by year and population of NUTS 2 regions by year; Marcell Kurbucz and Attila Katona, 'eudistance: Distance calculator for the different levels of European NUTS regions', Software Impacts, August 2022.

PCA transforms the original variables into principal components through the following steps: Given a dataset with n observations and p variables, let X be the n x p data matrix.

1. Standardization (if necessary):

$Z = (X - \mu) / \sigma$

where μ is the mean and σ is the standard deviation of each variable.

Standardization of the data is performed 'if necessary'—typically when variables are on different scales. This process ensures that each variable contributes equally to the analysis, preventing those with larger scales from dominating the results.

2. Covariance Matrix:

$C = (1 / (n-1)) * Z^T * Z$

3. Eigenvalue Decomposition:

$C * V = V * \Lambda$

where V is the matrix of eigenvectors (principal components) and Λ is the diagonal matrix of eigenvalues.

4. Principal Components:

$PC = Z * V$

where PC is the matrix of principal component scores.

5. Variables:

X = [Matrix].

After reducing the dimensionality of the data through PCA to identify the principal components, we used correlation analysis to investigate the relationships between these components and regional trade success. By calculating correlation coefficients, we quantitatively assessed how each principal component—representing a combination of regional characteristics—was associated with the volume of goods and services exported. This allows us to identify the underlying patterns and traits that define clusters of regions.

The PCA biplots provide a visual representation of our findings, illustrating the relationships between regions characteristics and export prowess in goods and services. A PCA biplot visually represents both variables (as vectors) and observations (as points) in a reduced-dimensional space, showing the relationships between variables and observations based on their positions and directions relative to each other. These plots help us understand how variables such as GDP, population, human capital, technological advancements, infrastructure, and governance quality correlate with trade volumes.

In the PCA biplot, variables are represented as arrows and individual regions as points. Arrows pointing in similar directions indicate a positive correlation between variables, showing they share similar trends or behaviours within the data. Conversely, arrows pointing in nearly opposite directions suggest a strong negative correlation. The length of each arrow reflects the strength of its variable's contribution to the principal component, with longer arrows indicating a greater influence.

Each point on the biplot represents a region, and the distance between any two points illustrates their similarity or dissimilarity based on traits such as institutional quality, infrastructure, and innovation capacity. Small angles between two arrows signify a positive correlation, indicating that as one variable increases, the other tends to increase as well. Large angles, especially those near 180 degrees, indicate a negative correlation, where one variable increases as the other decreases. For example, if the GDP arrow and ICT innovation capacity arrow are close to parallel, this suggests that regions with higher GDP are likely to have greater innovation capacity.

Chart A illustrates a biplot for the 2008 goods trade. In this visual, the 'total_x' variable, highlighted in light green and positioned in the top right quadrant, represents the exports of goods. The alignment of this variable with others in the same direction indicates positive correlations, showing that regions with higher goods exports also tend to have higher regional GDP, larger populations, and more patenting activities. The arrows for these variables are tightly clustered and point in the same direction, reinforcing their strong interrelationships. The second panel of the chart details the NUTS2 codes for the regions forming these clusters. Notable among them are outliers such as Paris (FR10), Stuttgart (DE21), and Noord-Brabant (NL41), which are significant due to their high performance on these metrics. This quadrant also includes a broad array of populous, old industrial regions in Western Europe, highlighting a pattern where traditional industrial strength correlates with higher goods exports.

Chart A: PCA clustering of regional characteristics and exports of goods in 2008

Chart B, representing the 2018 data, shows a dramatic change in the clustering of variables. The variables for regional GDP, population, and knowledge workers are now orthogonal to the goods exports ('total-x') variable, indicating no significant correlation in the new data. This suggests a significant shift in manufacturing and export hubs from populous, affluent regions in Western Europe to more peripheral, less populous, and economically less developed regions, particularly in Central and Eastern European countries.

Chart B: PCA clustering of regional characteristics and exports of goods in 2018

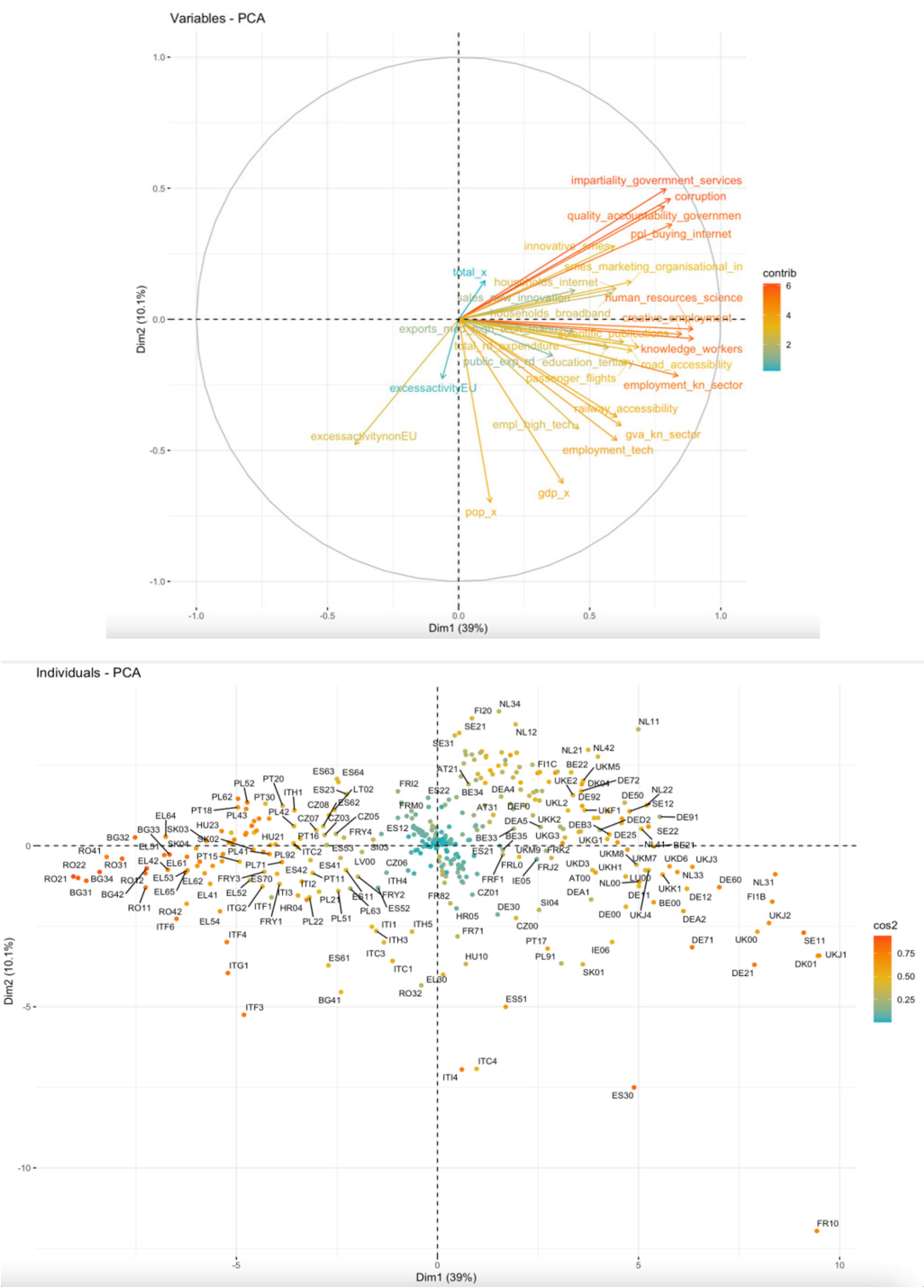
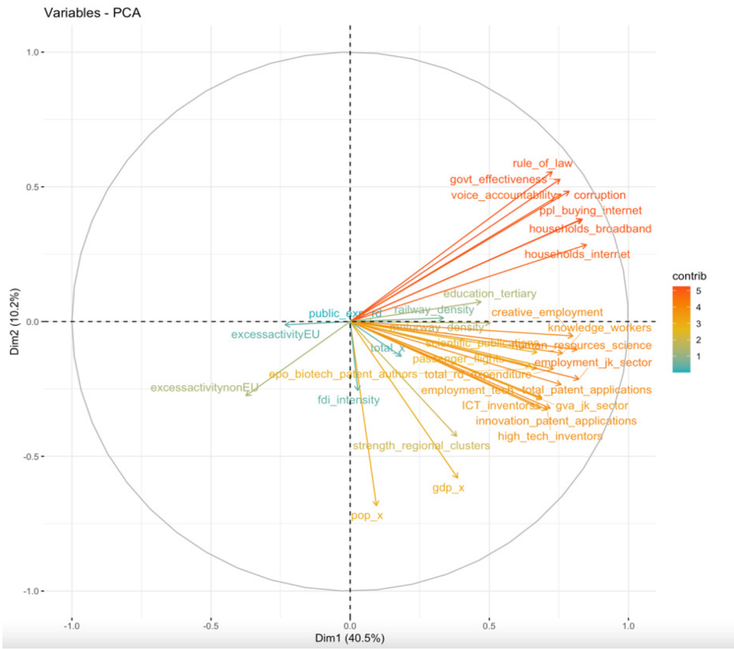


Chart C displays the 2008 services trade biplot, where the 'total_x' variable – representing services exports – is located in the low right quadrant. It is notably aligned with key economic indicators such as regional GDP, population, patent applications, and employment sectors including finance, technology, and science. This alignment suggests that regions excelling in these areas tended to have higher services exports. It's important to note that in PCA biplots, the rotation of variable arrows around the origin does not alter their interpretive value since PCA focuses on the relative directions and distances between arrows, indicating correlation strengths and directions regardless of their quadrant.

Chart C: PCA clustering of regional characteristics and exports of high-knowledge services in 2008



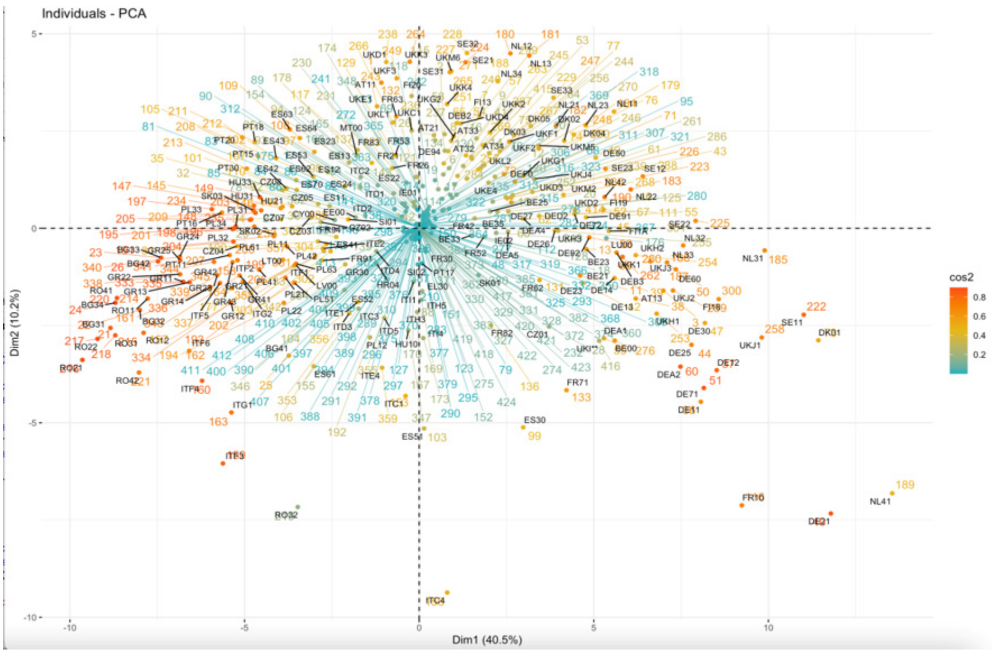


Chart D, depicting the 2018 services trade, shows the 'total_x' arrow now positioned in the upper right quadrant, indicating a continuation and strengthening of its alignment with population, GDP, and sectors like technology and science. This shift suggests an increasing integration of these variables with services trade, pointing to deeper economic interdependencies. The arrows are closer together and point more uniformly in the same direction compared to 2008, highlighting a stronger consensus among these economic drivers in influencing services exports. In the second panel of Chart D you see that many EU capitals are clustered in the quadrant around services exports. The clustering of many EU capitals in this quadrant underscores their significant role in the services sector, reflecting their economic growth and specialization. A key for the NUTS2 identifiers of observations that are located there can be found here:

IE06 – Dublin

CZ00 - Czechia (country)

HR05 - Zagreb

PT17 – Lisbon

PL91 – Warsaw

SK01 - Bratislava

FR71 – Rhône-Alpes

ES30 – Madrid

EL30 – Athens

FR10 - Paris

ITC4 – Lombardy

DEA1 - Düsseldorf

DE00 – Germany (country)

NL00 – Netherlands (country)

LU00 – Luxembourg (country)

UKK1 – London

BE01 – Brussels

For the full dataset, all the variables in detail, and R code, please contact the authors

Chart D: PCA clustering of regional characteristics and exports of high-knowledge services in 2018