Levelling up innovation: Boosting R&D in underperforming regions

TECHNICAL APPENDIX

Data

Business expenditure on R&D (BERD) data comes from Eurostat, and can be broken down by NUTS2 region, and by sector according to 2- or 3-digit NACE industry codes (dependent on sector). Eurostat does not, however, provide a crosstab necessary for this analysis.

To resolve this, we estimate the crosstab by using GVA data from the ONS[i] (which does provide a crosstab), and by applying a technique called the Furness method.

Let X be a matrix where each element X_{ij} represents the total business R&D expenditure in region i in sector j. The Eurostat data tells us the row sums $\sum_j X_{ij}$ and the column sums $\sum_i X_{ij}$, but not the individual elements. Let $X^{(0)}$ be a matrix with the same dimensions of X, but where the individual elements are the GVA for each region and sector.

The Furness method enables estimation of the elements of X iteratively. Let the elements of matrix $X^{(1)}$ and $X^{(2)}$ be defined as follows:

$$X_{ij}^{(1)} = \frac{X_{ij}^{(0)} \sum_j X_{ij}}{\sum_j X_{ij}^{(0)}}$$

$$X_{ij}^{(2)} = \frac{X_{ij}^{(1)} \sum_i X_{ij}}{\sum_i X_{ij}^{(1)}}$$

In simple terms, in the first iteration, we divide all the entries in $X^{(0)}$ by the row sums of $X^{(0)}$, and multiply them by the row sums of X, which means the row sums of $X^{(1)}$ and X are the same. We then repeat this process with the column sums to create $X^{(2)}$. This process converges rapidly, such that after 12 iterations, the row/column sums of matrix $X^{(12)}$ are within 0.1 percent of the row/column sums of X.

Revealed comparative advantage

The sectoral strengths of each region are identified by calculating *revealed comparative advantage* (RCA) from the previously defined matrix X. RCA for region i in sector j is a measure of how much region i invests in sector j compared to the average region. It is defined as:

$$RCA_{ij} = \frac{X_{ij} / \sum_j X_{ij}}{\sum_i X_{ij} / \sum_{ij} X_{ij}}$$

RCA values greater than 1 indicate that the region has more business R&D investment in that sector than the average region. We can then construct a matrix M showing the strengths of each region, where:

$$M_{ij} = \begin{cases} 1, & RCA_{ij} \geq 1 \\ 0, & \text{otherwise} \end{cases}$$

Which is represented as a heatmap in figure 1.

Similarity indices

The regional similarity matrix $\widetilde{M}$ is defined as:

$$\widetilde{M}_{ii'} = \sum_j \frac{M_{ij} M_{i'j}}{u_j d_i}$$

Where $d_i = \sum_j M_{ij}$ and $u_i = \sum_i M_{ij}$. These measures are known as *diversity* (the number of sectoral strengths a particular region has) and *ubiquity* (the number of regions that have a strength in a particular sector). Likewise, a sectoral similarity matrix $\widehat{M}$ can be defined:

$$\widehat{M}_{jj'} = \sum_i \frac{M_{ij} M_{ij'}}{u_j d_i}$$

The similarity indices used to order the regions and sectors in figure 1 are defined as the eigenvector associated with the second largest eigenvalue of $\widetilde{M}$ and $\widehat{M}$ respectively.

For further information on interpreting these similarity matrices, and why the second eigenvector produces a similarity index, see Mealy et al., 2019.[ii]

---

[i] https://www.ons.gov.uk/economy/grossvalueaddedgva/datasets/regionalgvanuts2

[ii] https://advances.sciencemag.org/content/advances/5/1/eaau1705.full.pdf