The three-dimensional activity index

Joel Emanuel Fuchs

jfuchs@uni-wuppertal.de
University of Wuppertal, Institute of Sociology, Gaußstraße 20, D-42119 Wuppertal (Germany)

Introduction

In sociological data analysis the comparison of data from different countries or institutions occurs all the time. The comparison of absolute values is often problematic due to the different sizes of the observed entities. One solution for comparing different sized entities is the activity index (AI). It enables the normalized international or inter-institutional contrasting of various fields. Although the AI is a long-used instrument, it lacks self-specific instruments to analyse itself. In this paper, we first want to present the AI. After that, we will introduce a new measure called the three-dimensional activity index (3D-AI) motivated by the statistical expected value. In the last part we will show how to use the 3D-AI to centre the basic activity index.

Used Data

For the sake of comprehension, data is used to visualise the new indices. Data from the European Patent Office (EPO) is open, easy to understand and traceable. So, we decided to use the granted patents per field of technology and per country of residence for 2011-2015 available from https://www.epo.org. The data is smoothed by a 3-year binomial filter to visualize the field and country specific trends better.

From the dataset follows, that the basic population consists of 46 countries plus 1 residual category divided into 35 fields over five years. All 8,225 data points will be used for calculation, but to keep the visual analysis clear, only the two countries Germany and the United Kingdom (UK) and the two fields 'Food chemistry' (FC) and 'Semiconductors' (SC) will be represented. This choice is arbitrary; the focus lies on the formulae presented later.

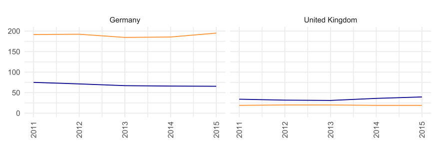

Figure 1. Granted patents.

The activity index

Of course, we can see in figure 1, that the shares of FC and SC of the overall granted patents must be more similar to each other in the UK than in Germany. But 'seeing' or comparing the absolute values is too intangible. Therefore, a relative index is often used, which directly reveals such differences between countries regarding the underlining fields. We will call it the activity index (AI), as denominated by Narin et al. (1987). But it is also known under the revealed technological advantage (Soete & Wyatt, 1983), revealed comparative advantage (Balassa, 1965) or the Balassa index (Ibid.).

Let x_{ijt} be the granted patents of country j regarding the field i in the year t. The AI relates the share of one field of a country $(x_{ijt}/\sum_i x_{ijt})$ to the share of the same field but of all countries $(\sum_j x_{ijt}/\sum_{ij} x_{ijt})$. So, we get

Formula 1. Activity index.

$$AI_{ijt} := AI(x_{ijt}) := \frac{x_{ijt}/\sum_i x_{ijt}}{\sum_j x_{ijt}/\sum_{ij} x_{ijt}}.$$

We calculated the AI for all 8,225 data points, but in figure 2 we will again only show Germany and the UK as well as FC and SC. Because of the different sums used for the AI, it is important to mention which values were calculated and which data points were used.

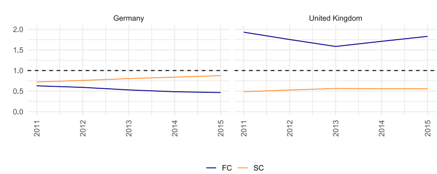

Figure 2. Activity index.

Figure 2 shows the AI corresponding to figure 1. The dashed line represents the average across all combinations of country and field. It is obvious that Germany is closer to the average than the UK. The second observation is that, except of Semiconductors in Germany, all other fields veer away from the average over time.

The temporal activity index

The values of the AI are calculated year by year. This is done, because all 1,645 data points of one year are integrated into the calculation of a single AI value by the composed sums. If we want AI values, that do not depend on the year, we could summarise all values by country and field over all five years. We get an AI that is constant over time so we will denote it the Temporal Activity Index (TAI).

Formula 2: Temporal activity index.

$$TAI_{ijt} := TAI(x_{ijt}) := \frac{\sum_t x_{ijt} / \sum_{it} x_{ijt}}{\sum_{jt} x_{ijt} / \sum_{ijt} x_{ijt}}.$$

Arithmetic mean

Let us calculate the arithmetic mean (AM), but not for all AIs, just for the AIs of a single country and a single field. Therefore, y shall be the number of years. Then we get $AM(AI_{ijt}) = \frac{1}{y}\sum_t AI(x_{ijt})$. This describes the AM of the AI values. We could also calculate the AI values of the arithmetic means of each part of the AI, so these would be $AM(x_{ijt}) = \frac{1}{y}\sum_t x_{ijt}$, $AM(\sum_i x_{ijt}) = \frac{1}{y}\sum_{it} x_{ijt}$, $AM(\sum_j x_{ijt}) = \frac{1}{y}\sum_{jt} x_{ijt}$ $AM(\sum_{ij} x_{ijt}) = \frac{1}{y}\sum_{ijt} x_{ijt}$. Using these AMs for calculating the AI, we will get the TAI as aforementioned. We conclude that the TAI is very similar to the AM.

Table 1. Temporal activity index.

Country	FC	SC
Germany	0.535	0.796
United Kingdom	1.759	0.535

The three-dimensional activity index

What does this have to do with the 3D-AI? The 3D-AI is the fraction of the classic AI and the TAI, as

Formula 3. Three-dimensional activity index.

$$3D\text{-}AI(x_{ijt}) := \frac{AI(x_{ijt})}{TAI(x_{ijt})} = \frac{x_{ijt} \cdot \sum_{ij} x_{ijt} \cdot \sum_{it} x_{ijt} \cdot \sum_{jt} x_{ijt}}{\sum_{ijt} x_{ijt} \cdot \sum_t x_{ijt} \cdot \sum_j x_{ijt} \cdot \sum_i x_{ijt}}.$$

So, the 3D-AI centres the AI by its average over the years represented by the TAI. We therefore improve the AI by disadvantaging other parts of it, as we can see in figure 3.

The first two indices of the AI can be interchanged, i.e. $ai(x_{ijt}) = ai(x_{jit})$. We denote the newest value the 3D-AI, because all three indices can be interchanged, i.e. $ai(x_{ijt}) = ai(x_{tij}) = ai(x_{itj})$ and so on.

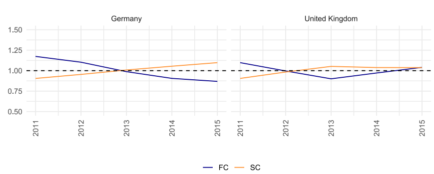

Figure 3. Three-dimensional activity index.

So, what are the advantages and disadvantages? Because we only divide the lines from figure 2 by a constant, the centred 3D-AI values are very similar, as we would expect from centred values. But now we can directly see in which year which country regarding which field performs like its average. For example, UK's FC line intersects with the dashed line about the year 2012 and 2014/2015. Before 2012, UK's FC performed above its own average, same after 2014. This is new information, which the classic AI and the absolute values could not express. The 3D-AI does not replace the classic AI, because it has some disadvantages. Due to the centralisation we cannot measure which field performs above average and which below. There are some more disadvantages, which would go beyond the scope of this paper.

In the future, there is a lot more to do. The TAI can be a better mean for the AI, but for a complete centring we also need an (empirical) variance. Indeed, if we want to analyse data by the activity index, we should think about reconstructing all empirical instruments, not only the mean and the variance, but also differentiation, correlation and so on. Perhaps we would benefit from a whole new toolbox designed especially for inter-institutional analyses.

Acknowledgments

We thank Jennifer Madine for proofreading this paper. Further we thank the European Patent Office for their free data providing service. All graphics and calculations were made by the help of R, RStudio and the huge package service around R, especially magrittr, data.table and ggplot2.

References

Balassa, B. (1965). Trade liberalization and "revealed" comparative advantage. *The Manchester School of Economic and Social Studies,* 32, 99-123.

Narin, F., Carpenter, M. P. & Woolf, P. (1987). Technological assessments based on patents and patent citations. In Grupp, H. (Ed.), *Problems of measuring technological change*, Cologne, 107-119.

Soete, L. G. & Wyatt, S. M. E. (1983). The use of foreign patenting as an internationally comparable science and technology output indicator. *Scientometrics,* 5, 31-54.