The normalized activity index – an alternative transformation to logarithmic calculus of the activity index

Joel Emanuel Fuchs (jfuchs@uni-wuppertal.de)
Bergische Universität Wuppertal https://orcid.org/0000-0002-9951-3492

Research Article

Keywords: Activity Index (AI), logarithmic AI (LAI), Normalized AI (NAI), European Patent Office (EPO), patent analysis

Posted Date: April 22nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-939228/v2

The normalized activity index – an alternative transformation to logarithmic calculus of the activity index

Joel Emanuel Fuchs[1]

[1] *jfuchs@uni-wuppertal.de*
University of Wuppertal, Gaußstr. 20, 42119 Wuppertal, Germany
ORCiD: 0000-0002-9951-3492

Abstract
The Activity Index (AI) is a well-known index for comparing the contribution of different entities on various fields, for example scientific articles with authorships from different countries structured into various subjects as arts, engineering, economics and so on. This index lacks important properties; the most demonstrative one is its characteristic to be lower but not upper bounded. Further, we will show that the AI is a log-normal distribution and that it is common in literature to transform the AI by the logarithm to a normal distribution. Last, we will present an alternative transformation special for longitudinal data, that transforms the AI to a normal distribution, too, without the negative properties of the logarithm like the loss of data if the logarithm is applied. This newly introduced index called Normalized AI (NAI) will be calculated by expansion the relation of the AI in dividend as in divisor. It will not converge to the logarithm of the AI, but to the logarithm of the AI if z-standardized by each entity-field combination.

Keywords: Activity Index (AI); logarithmic AI (LAI); Normalized AI (NAI); European Patent Office (EPO); patent analysis;

1 Introduction
In sociological data analysis the comparison of data from different countries or institutions occurs all the time. Thereby, the comparison of absolute values is often problematic due to the different size of the observed entities, e.g. a country like Germany, that has eleven times the landmass of Belgium and seven times of its population, will unavoidable be leading in producing output like scientific publications or other (measured in absolute numbers). It is common to use relative data if comparing different sized entities as for example publications per capita, or even relation indexes like the activity index (AI). It enables the normalized international or inter-institutional contrasting of various fields and has the advantage that it is a single-variable transformation unlike the other two which we mentioned. As we will see below, the AI sets the focal variable in relation to sums over subgroups of itself, making it especially interesting for data sets that have only one measuring variable per observation.

Unfortunately, the AI possesses some characteristics that are undesirable. As we will see below, its range is bounded on one side and – as a consequence of that – the AI is log-normal distributed, while many analytic techniques desire a normal distribution, for example Pearson's correlation, z-transformation, or a non-bounded range like some regression techniques. Most common solution is to apply the natural logarithm to receive a normal distribution. This will lead us to a further index called LAI (logarithmic AI).

As mentioned above the LAI has a normal distribution, and so we can apply the z-standardization. If we do this on the macro level, we will state no major difference to non-standardized LAI values, because they already tend to a z-standardization (even though LAI values never will reach a z-standardization without normalizing them). However, on the micro level – defined as the longitudinal data of each entity-field

combination – the z-standardization results into a comparison of self-performance determining, in which years an entity performs over its own mean regarding a focal field.

The second part of this paper will introduce a new index derived from the AI: the normalized Activity Index (NAI). It is a transformation of the Activity Index specifically for use on longitudinal data. Not only will we motivate this new index, but also show some of its characteristics and its similarities to the LAI. In the end it will replace the z-standardization of the LAI on the micro level and additionally to the AI will characterize the performance of an entity in a focal field over the observed years.

We admit that this paper covers a very specific field, the analysis of longitudinal single-variable[1] data sets identifiable over two categorial. We hope, the presented information of AI and LAI extent the potential audience. Nevertheless, the additional analysis on micro level can help to differentiate data furthermore. Finally, how much information can be derived by a single variable and where starts just the fishing for new indices, this question is left to the reader.

2 European Patent Office

To demonstrate the application of the various indexes, we chose a real-life example. Besides five requirements, the choice of the example data is arbitrary. These requirements are: 1. open-access data, 2. scientifically approved, 3. longitudinal, 4. unique observations identifiable by entity and field, and 5. non-negative integer as main variable. We decided to use data from the European Patent Office (EPO), more precisely to use the granted patents per field of technology and per country of residence for 2011-2015 (EPO 2021).

We have 8,225 observations reflecting the number of patents that were granted by the EPO. Each observation is uniquely identifiable by the patents' country, the patents' field, and the year they were granted. The forth variable of our data set reflects, how many patents belong to a certain country/field/year combination and is therefore a non-negative integer variable. Further, 46 individual countries and areas are mentioned in addition to one residual value (called "other countries"). The patents are divided into 35 different fields. The three identifiers country/field/year make 8,225 ($= 47 \cdot 35 \cdot 5$) observations possible, that is the number of observations of our data set. We conclude that we have a comprehensive survey.

Table 1:Key data of the EPO data set.

Observations	Years	Countries	Fields
8,225	5	47	35

[1] As single-variable data set we denote a data set containing a single numeric variable and arbitrary many categorial variables (including year) to uniquely identify each observation.

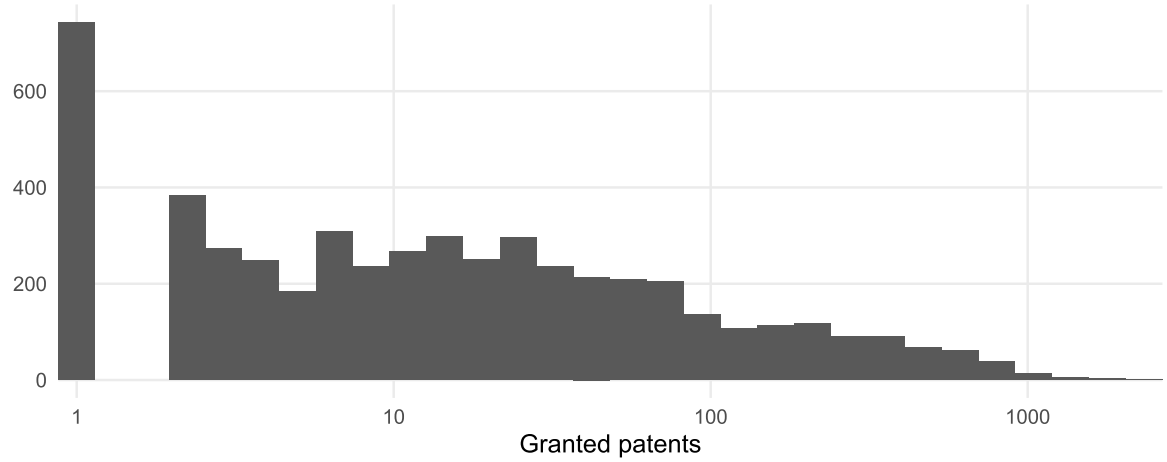

Figure 1: Histogram of number of patents granted per country, field and year (zero excluded; logarithmic scale on x axis).

Let us first focus on absolute numbers of our data set. 36.8% of the observations are zero (not shown in Figure 1 due to the logarithmic scale), 30.1% are between 1 and 10 (see also Figure 1), 24.1% are between 11 and 100, and 9.0% are over 100. Table 2 shows the five countries with the most granted patents in our time period.

Table 2: Five highest values of granted patents of our data set aggregated by country.

Country	United States	Germany	Japan	France	Switzerland
Patents	72,293	67,518	58,339	24,673	13,629

Already between these five countries there are huge differences in the amount of granted patents, so we get a first feeling for the difficulties of analysing absolute data of our set. We have some few big players (United States, Germany, and Japan) dominating our observations, and making it difficult to find country specific characteristics. In the next example we show why their domination disguises the characteristics of other countries.

Let us have a look at the two countries Germany and the United Kingdom (UK), that was granted 10,200 patents between 2011 and 2015, and the two fields ‘Food chemistry’ (FC) and ‘Semiconductors’ (SC). Figure 2 visualizes the observations of these two countries and fields over the years 2011-2015. As expected, Germany is dominating both fields in absolute numbers of granted patents. Table 3 complements the graphical information with the mean over the years.

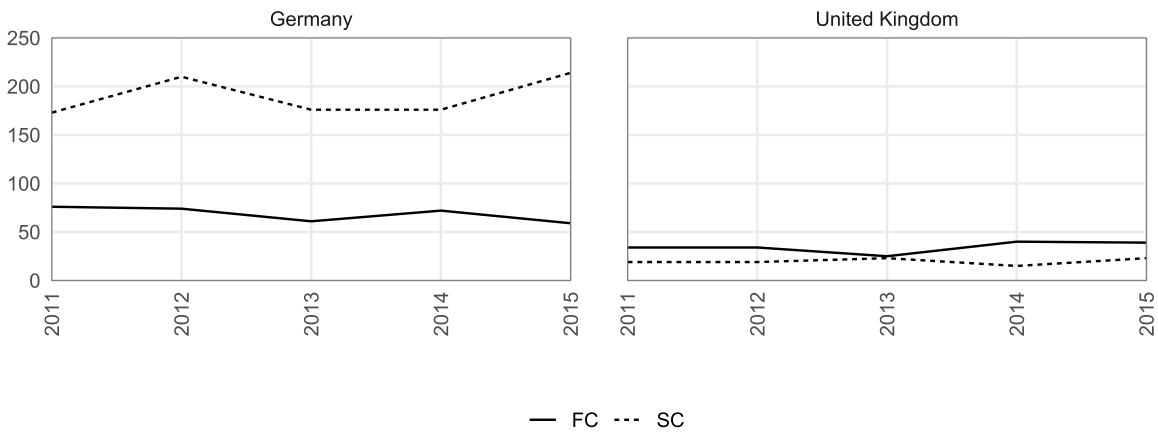

Figure 2: Granted patents.

Table 3:Arithmetic Mean of granted patents of the years 2011-2015.

	Germany	UK
Food chemistry	68.4	34.4
Semiconductors	189.8	19.8

Even though Germany is clearly dominating, the UK reflects a contrary setup regarding the two fields FC and SC. While in Germany patents of SC are dominating over the other field, in the UK patents of FC are dominating the granted patents of SC. Following questions come in mind: what is the norm, dominating FC or SC? Does FC dominate SC in the UK to the same degree as SC dominates FC in Germany? Or is the gap between SC and FC in the UK negligible? These questions can be answered by the AI. Additionally, it evens the differences in size between the countries, what enables a direct comparison of AI values from different countries.

3 The Activity Index

In literature the AI is known by several names. 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 or RTA (Soette & Wyatt 1983), revealed comparative advantage or RCA (Balassa 1965), or the Balassa index (Ibid.).

Let *D* be a comprehensive data set with a summable variable *v*, i.e. the values of *v* are non-negative integers, and each observation of *D* is uniquely identifiable by the tuple of two categorial variables *I* and *J,* e.g. *I* for a set of countries and *J* for a set of fields, and an observation exists to each tuple. We denote by v_{ij} the value of variable *v* identified by $i \in I$ and $j \in J$. Than we define the AI as

Equation 1: Activity Index (AI)

$$AI_{ij} := AI(v_{ij}) := \frac{v_{ij} / \sum_{j \in J} v_{ij}}{\sum_{i \in I} v_{ij} / \sum_{i \in I, j \in J} v_{ij}}.$$

The AI often operates on single-variable data sets like our EPO data, that allows to use the expressions AI_{ij} and $AI(v_{ij})$ equally. The AI is a relation of shares, as which we want to reformulate it to gain a deeper understanding of what exactly the AI is

measuring. A share can be meant as due or fraction of a value, but we also use the word share as synonym for a number $x \in [0,1]$. It follows a verbal version of the AI.

Equation 2: verbal version of AI.

$$AI_{ij} := AI(v_{ij}) := \frac{\text{The share of } v_{ij} \text{ on all accumulated obs. of } i.}{\text{The share of all accumulated obs. of } j \text{ on all accumulated obs..}}$$

That means, the AI compares the share of field *j* of a country *i* with the average share of field *j* (see Figure 3).[2] We can conclude some properties of the AI directly from this formula. If $AI_{ij} > 1$, then the share of field *j* in country *i* is higher than the average share of field *j*. Contrary, if $AI_{ij} < 1$, then the share of field *j* in country *i* is lower than the average share of field *j.*

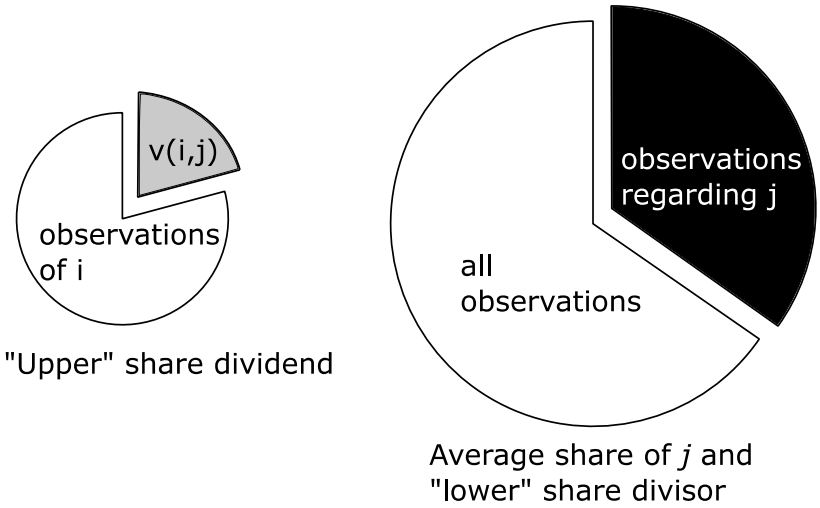

Figure 3: Visualization of the shares of the AI with AI < 1.

The AI has three main advantages. First, contrary to other relations like patents per capita the AI does not need further variables to be calculated. Second, it is a relation index; the size of the countries does not matter any longer. It only matters what shares the fields have within each country. Third, the AI can be calculated for each country-field combination without making any further choices like setting a reference country. Indeed, the reference for each country is the same, that is the average share of *j.*

We computed the AI for all of our 8,225 observations by year as seen in Figure 4; we recognize a log-normal distribution at once. Unfortunately, we lose 350 observations by division through zero. This happens for example at North Macedonia, that did not have one granted patent in all years and fields.

[2] If we talk of the average share of field *j*, we mean the overall share of *j* in the data set, i.e. $\frac{\sum_{i \in I} v_{ij}}{\sum_{i \in I, j \in J} v_{ij}}$, not to mistaken for the arithmetic mean of the share of field *j*, that is $\frac{1}{\#I} \sum_{i \in I} \frac{v_{ij}}{\sum_{j \in J} v_{ij}}$ where #*I* is here and everywhere else the number of categories in *I* (e.g. number of countries).

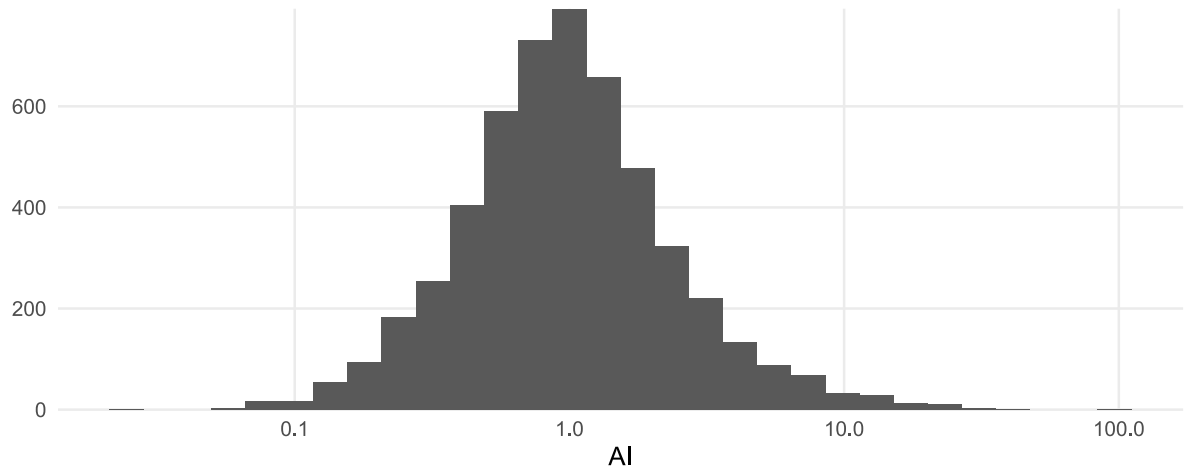

Figure 4: Activity Index for granted patents in 2011-2015 (zero excluded; logarithmic scale on x axis).

Figure 5 shows the AI for the countries Germany and UK regarding both fields FC and SC. The figure is complemented by Table 4 showing the average AI values of these country-field combinations. We derive some expectable and some new information.

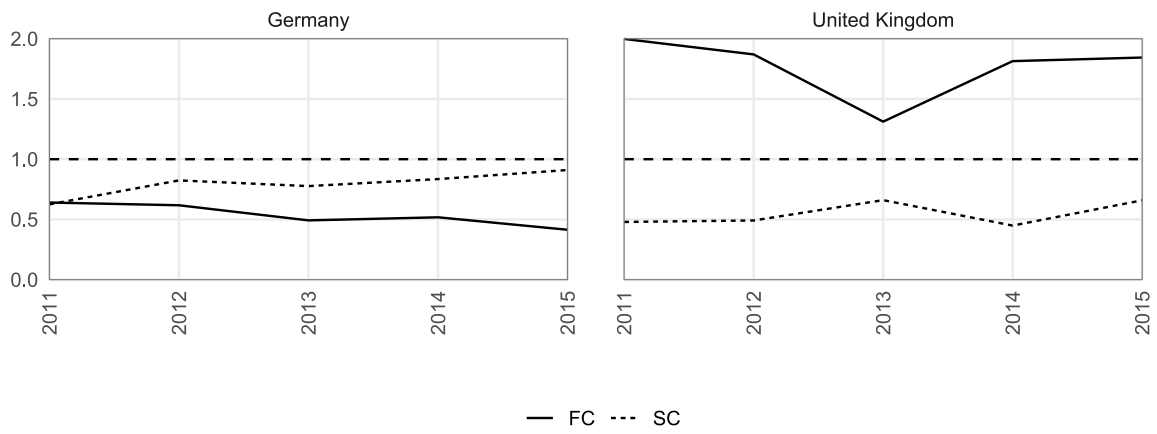

Figure 5: AI of granted patents.

The AI of Germany regarding our two fields FC and SC performs like expected. Even though both AI lines are below 1 (meaning the share of FC respectively SC on all German patents is less then the average), they are correlated to Germany's absolute number of granted patents (Spearman's correlation coefficients for absolute numbers and their corresponding AI are 1 for FC and 0.8 for SC). Having a look at the UK, we derive new information. Table 3 indicated, that both fields FC and SC are treated nearly equally by the UK. But considering Table 4, the share of granted patents in FC is 1.77 times the average share of FC, so the UK places emphasis on this field. On the other hand, the share of granted patents in SC is only 0.55 times the average share of SC, meaning that the international community puts more effort on this field than the UK.

Table 4: Arithmetic mean of the AI of granted patents 2011-2015.

	Germany	UK
Food chemistry	0.536	1.767
Semiconductors	0.794	0.548

The AI is not an index indicating in which field a country is superior, especially with a view to Germany. In absolute numbers it outnumbers the UK on both fields FC and SC. Regarding the AI, UK has a way higher AI in FC than the "big" Germany – but this does not mean, that the UK is superior on the field FC. It only means the UK is more engaged in the field FC with its possible resources than Germany. How can we explain this phenomenon, having high absolute numbers but very low AI values? There are many reasons, and probably the most important one is, that the AI is a relation of shares (Equation 2). Considering the dividend all shares of a country sum up to 1, making it impossible to just have high shares on every field. Indeed, the shares of one country spread from low to high resulting in lower and higher AI values. Therefore, even if a country is superior on all fields, it will still posses lower AI values. (The unrealistic exceptional case would be $AI \equiv 1$.)

We want to briefly examine some properties of the AI. We already talked about its domain, a non-negative integer variable uniquely discriminable by two categorial variables. The range of the AI is clearly $[0, +\infty)$, because a share is located in [0,1], so a relation of shares (Equation 2) is located in $\frac{[0,1]}{[0,1]} = [0, +\infty)$.[3] More interesting is the distribution of the AI, that as a transformation of a variable or of a set of observations is always finite and, in this sense, cannot possess a continuous distribution. Nevertheless, we see in Figure 4 that AI values tend to be log-normal distributed with an expected value of 1. This is not an exception but a rule, although it would go beyond the scope of this paper to proof this.[4]

4 The logarithmic AI

Having log-normal distributed data, suggests to apply the natural logarithm leading to the logarithmic AI (LAI). In literature the LAI is also known by the term revealed patent advantage or RPA (Grupp 1994). It is defined as

Equation 3: The logarithmic Activity Index (LAI).

$$LAI_{ij} := \ln\left(AI_{ij}\right).$$

The range of the LAI is $(-\infty, +\infty)$, that is the logarithm of the range of the AI. Its expected value tends to 0 ($= \ln(E[AI])$) if the number of observations tend to infinity[5].

---

[3] Here we use interval calculation. The Division of two intervals is defined by: $\frac{[a,b]}{[c,d]} := \left[\frac{a}{d}, \frac{b}{c}\right]$. Approximating the interval [0,1] by $\lim_{n\to\infty}\left[\frac{1}{n}, 1\right]$, ee get $\frac{[0,1]}{\lim_{n\to\infty}\left[\frac{1}{n},1\right]} = \lim_{n\to\infty}\left[\frac{0}{1}, \frac{1}{1/n}\right] = \lim_{n\to\infty}[0, n] = [0, \infty)$.

[4] One way to show this is interval calculation again. We divide the interval $[0,1]$ into *n* equal sized subintervals $\left[\frac{i}{n}, \frac{i+1}{n}\right]$ for $0 \le i < n$. We presume, that every subinterval has the same probability to contain a share. If we have two shares $x, y \in [0,1]$, we get a certain probability for $\left[\frac{i}{j+1}, \frac{i+1}{j}\right]$ to contain the share $\frac{x}{y}$, denoted by $P\left(\frac{x}{y}; \left[\frac{i}{j+1}, \frac{i+1}{j}\right]\right)$. For an arbitrary interval $[n_1, n_2] \subset [0,1]$ we have $P\left(\frac{x}{y}; [n_1, n_2]\right) = \sum_{0\le i,j<n} P\left(\frac{x}{y}; [n_1, n_2] \cap \left[\frac{i}{j+1}, \frac{i+1}{j}\right]\right)$. If $[n_1, n_2] \to \frac{x}{y}$ and $n \to \infty$, we get the log-normal distribution for *P*. If the AI values are independent and identically distributed and the number of observations tend to infinity, the AI approximates a log-normal distribution.

[5] If we demand the observations tending to infinity, we presume a uniformly extension of the categorial variables country, field and year. Also, we demand that only finitely many observations belong to each value of country respectively field respectively year.

See Figure 6 for a histogram of the LAI of granted patents (please observe the similarity to Figure 4, except this time the x-axis is normal scaled).

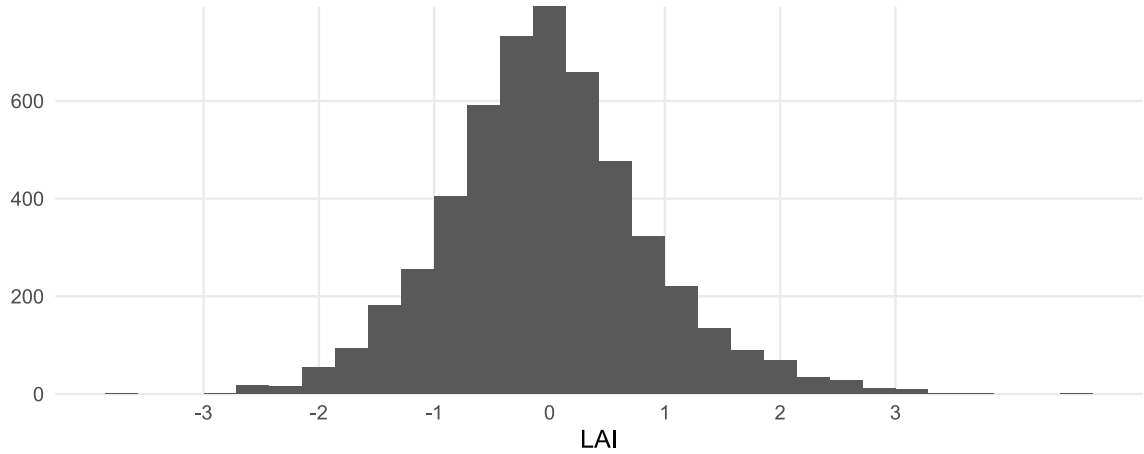

Figure 6: Logarithmic Activity Index for granted patents in 2011-2015.

Figure 6 shows 5,197 observations, that are 2,678 less than AI values and 3,028 less than observations in the data set. The reason, there are 2,678 AI zero values rejected by the logarithm. Overall, we have a loss of data of 37%.

We do not only have a loss in data, but a loss in interpretation, too. The AI offers an immediate way of interpretation, e.g. an AI value of 1.5 means, the share of the focal field is 1.5 times larger in the concerning country than the average share. This immediacy is lost for the interpretation of the LAI, enabling only relative statements like above the average (*LAI > 0*) or below the average (*LAI < 0*).

Interestingly, the LAI seems to tend to a z-standardization. Indeed, we have an expected value of *-0.02* and a deviation of *0.89* regarding our data set. If LAI would be a z-standardization with an expected value of *0* and a deviation of *1*, the related AI distribution would have an expected value of $1.65 = \sqrt{e}$ and a deviation of $2.16 = \sqrt{e^2 - e}$, but we already mentioned that the AI tends to an expected value of *1.* Therefore, the LAI does not tend to a z-standardization and will most times have a standard deviation *<1*.

Figure 7 shows the LAI of Germany and the UK for the fields FC and SC. In this figure, we added the AI values for these four country-field combinations, but we shifted the AI values one unit downwards to draw a better comparison. We see, there are no substantive changes in the curves meaning that we cannot derive information from the LAI that we did not already derive from the AI. We conclude that the LAI is an index with certain qualities (unbounded, nearly z-standardized) which can be important for further analysis, but it is no independent index in the sense that it delivers information which are new to the information of the AI.

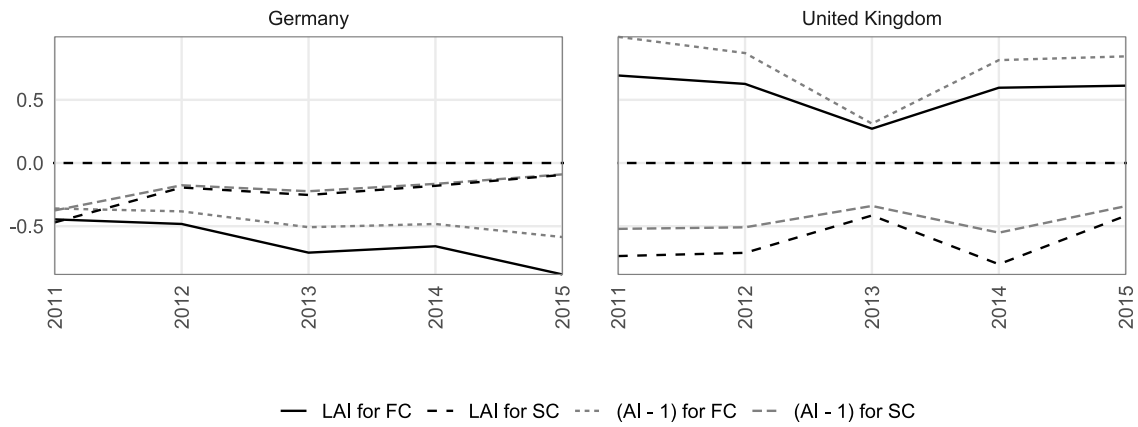

Figure 7: Comparison of AI (shifted down by 1 to scale to LAI) and LAI of granted patents.

While the z-standardization of the LAI values on the macro level will not reveal further information of our data, we sustained observed the micro level even though only as a showcase for four country-field combinations, what raise the question about a z-standardization on the micro level. To state precise formulas, we want to expand the notation of our observation by a year index. Let v_{ijt} be the granted patents of country *i* in field *j* and year *t*, and *T* the set of all years. The AI changes slightly to $AI_{ijt} := AI(v_{ijt}) := \frac{v_{ijt}/\sum_{j\in J} v_{ijt}}{\sum_{i\in I} v_{ijt}/\sum_{i\in I, j\in J} v_{ijt}}$, while the LAI changes to $LAI_{ijt} := LAI_{ijt}(AI_{ijt}) := \ln(AI_{ijt})$. Finally, we get the country-field specific z-transformation by

Equation 4: z-standardisation by each country-field combination.

$$z(LAI_{ijt}) := \frac{LAI_{ijt} - \frac{1}{\#T}\sum_{t\in T} LAI_{ijt}}{\sqrt{\frac{1}{\#T}\sum_{t\in T}\left(LAI_{ijt} - \frac{1}{\#T}\sum_{t\in T} LAI_{ijt}\right)^2}}.$$

Starting from here, we will mean this z-standardized LAI by each country-field combination, if we mention the z-standardization. The z-standardization marks an important point. The calculation of AI and LAI can be done (and originally was intended) with non-longitudinal data. The z-standardization like the NAI depend on longitudinal data. Figure 8 shows the z-standardization of Germany and the UK in the fields FC and SC.

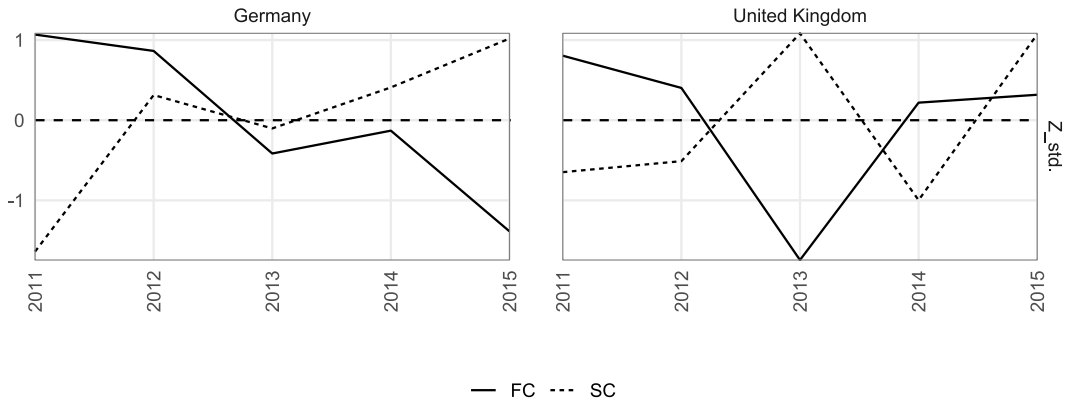

Figure 8: Granted patents as over years z-standardised LAI.

We want to highlight the difference between the various indexes for analysing the data and to do so, we will focus on Germany and the field FC. Analysing absolute numbers, we see that Germany's FC values drop from 76 to 59 (mean 68.4, sd. 7.8, see also Figure 2). Conjecturable, Germany's FC values tend to a constant value between 59 and 76. Otherwise, analysing AI values, we see that they are falling from 0.64 to 0.41 (Figure 5), meaning the FC share on granted patents in Germany is less then the average share and loosing over time. Does this mean, Germany's FC value is not tending to a constant number but has a dropping trend? Further, analysing z-standardized values, we can see that Germany performs in the years 2011 and 2012 over its own average (Figure 8). We clearly can distinguish a Germany before 2012 (year included) and after 2012 (year excluded), supporting the thesis of a dropping trend.

5 A new index: the three-dimensional AI

The z-standardization has three certain disadvantages. First, we stated a loss of data of 37%. Second, ad hoc there is no reason why the z-standardizations of two different country-field combinations should be comparable. Third, it is the third transformation in a row of the original data (from AI to LAI to standardization). Therefore, we proclaim a new index called the Normalized Activity Index (NAI). We will derive it from the AI as a natural development on longitudinal data, although it will be a transformation of the original observations with less loss of data compared to the z-standardization. In addition to that, the single NAI values will be highly comparable to each other. Last but not least we will proclaim, that the NAI reveals the same predications as the z-standardization making the last one overdue for interpretation.

Observing the micro level, it is obvious to consider the mean of each country-field combination as we have done before. For the absolute values of the data set the arithmetic mean is $\overline{v_{ij}} = \frac{1}{\#T}\sum_{t\in T} v_{ijt}$. Especially for the AI there are at least two ways to calculate its average. The first one is the arithmetic mean of the AI, i.e. $\overline{AI_{ij}} = \frac{1}{\#T}\sum_{t\in T} AI_{ijt}$. The second one is the AI of the arithmetic means, in formula

Equation 5: The AI of arithmetic means (AIAM).

$$AIAM_{ij} := AIAM(v_{ijt}) := \frac{\overline{v_{ij}}/\sum_{j\in J}\overline{v_{ij}}}{\sum_{i\in I}\overline{v_{ij}}/\sum_{i\in I,j\in J}\overline{v_{ij}}} = \frac{\sum_{t\in T} v_{ijt}/\sum_{j\in J,t\in T} v_{ijt}}{\sum_{i\in I,t\in T} v_{ijt}/\sum_{i\in I,j\in J,t\in T} v_{ijt}}.$$

The AIAM exists as long as at least one observation of country *i* and at least one observation of field *j* are positive. Hence, we have 8,050 AIAM values for our 8,225 observations. The AIAM seems to be more robust than the AI and – more important – delivers to every AI value a corresponding AIAM value.

Table 5: Comparison of Arithmetic mean of AI and AIAM, both calculated for the years 2011-2015.

	Arithmetic mean of AI		AIAM	
	Germany	UK	Germany	UK
Food chemistry	0.536	1.767	0.530	1.766
Semiconductors	0.794	0.548	0.790	0.545

Table 5 shows that the arithmetic mean of AI and the AIAM deliver quiet similar results. Indeed, the root mean square error (RMSE), i.e. $\sqrt{\frac{1}{\#I \cdot \#J} \sum_{i \in I, j \in J} \left(\overline{AI_{ij}} - AIAM_{ij}\right)^2}$, is 0.25 and we claim that this number tends to zero, if the number of observations tend to infinity.

Being the average gives rise to normalize AI through the AIAM, probably by division. Why normalize by division and not by subtraction? First, the AI is a relation meaning among others it is a non-negative numeric. Division by the AIAM would keep this property, subtraction would not. Second, even if it is not clear on the first sight, we normalized the z-standardization by dividing the AI, too. The dividend of the z-standardization is $\ln\left(AI_{ijt} \,/\, \sqrt[\#T]{\prod_{t \in T} AI_{ijt}}\right)$ (see also Equation 4). Third, by division we will gain symmetry of the indices *i*, *j* and *t*, i.e. $NAI_{ijt} = NAI_{jti} = NAI_{tij} = \cdots$ (see also Equation 6).

So, the normalization of the AI through division by the AIAM will be called the Normalized AI (NAI) and is defined as

Equation 6:The Normalized Activity Index (NAI)

$$NAI_{ijt} \coloneqq NAI\left(v_{ijt}\right) \coloneqq \frac{AI\left(v_{ijt}\right)}{AIAM\left(v_{ijt}\right)} = \frac{v_{ijt} \cdot \sum_{i \in I, j \in J} v_{ijt} \cdot \sum_{i \in I, t \in T} v_{ijt} \cdot \sum_{j \in J, t \in T} v_{ijt}}{\sum_{i \in I, j \in J, t \in T} v_{ijt} \cdot \sum_{t \in T} v_{ijt} \cdot \sum_{j \in J} v_{ijt} \cdot \sum_{i \in I} v_{ijt}}.$$

Calculating the NAI for our data set we get 6,480 observations. As mentioned before, we get for every AI value a corresponding AIAM value, but AIAM is equal zero in 1,395 cases. Overall, we have a loss of data of 21%, that is 16 percent points better than the LAI. Astonishing is the distribution of the NAI values. The NAI has like AI and AIAM a range of $[0, +\infty)$ and so it is log-normal distributed as we can see in Figure 9. But focussing on the range $[0,2]$ as we see in Figure 10, where we used a non-transformed x-axis, we see a nearly normal distribution of the NAI values. We claim that for a raising number of observations the distribution of NAI values tend to a normal distribution.

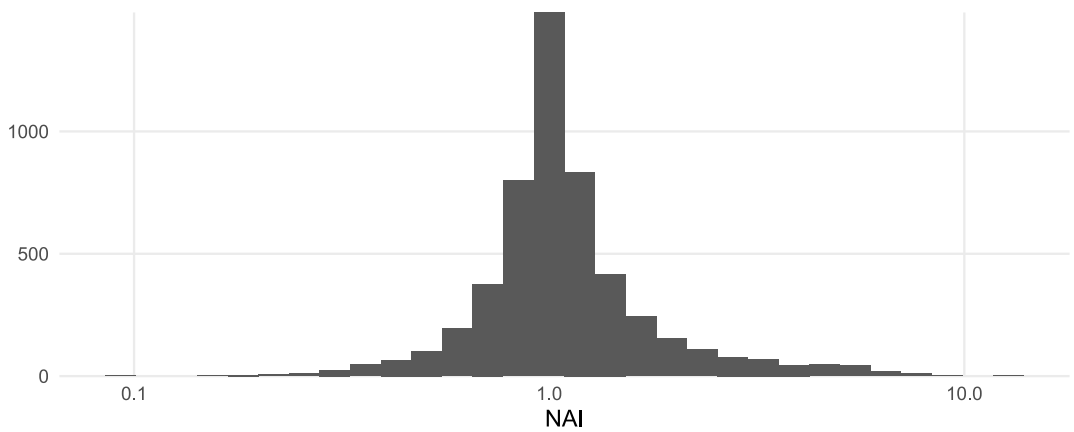

Figure 9: Normalized Activity Index for granted patents in 2011-2015 (zero excluded; logarithmic scale on x-axis).

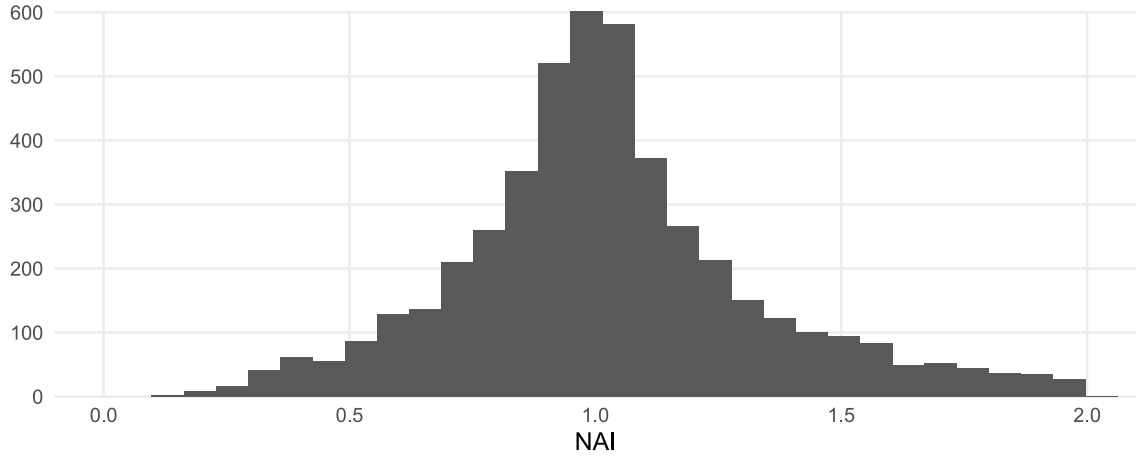

Figure 10: Normalized Activity Index for granted patents in 2011-2015 (values above 2 excluded; normal scale on x-axis).

Figure 11 shows the NAI values of Germany and the UK for the fields FC and SC in comparison to the z-standardization. Again, the values of NAI are shifted one unit downwards for better comparability. According to there amount, the curves differ a lot, but disregarding scaling, the curves seem to be very equally. Indeed, the RMSE of all z-standardized values and all shifted NAI values is just 0.77. We claim again that the RMSE tends to zero if the number of observations raises.

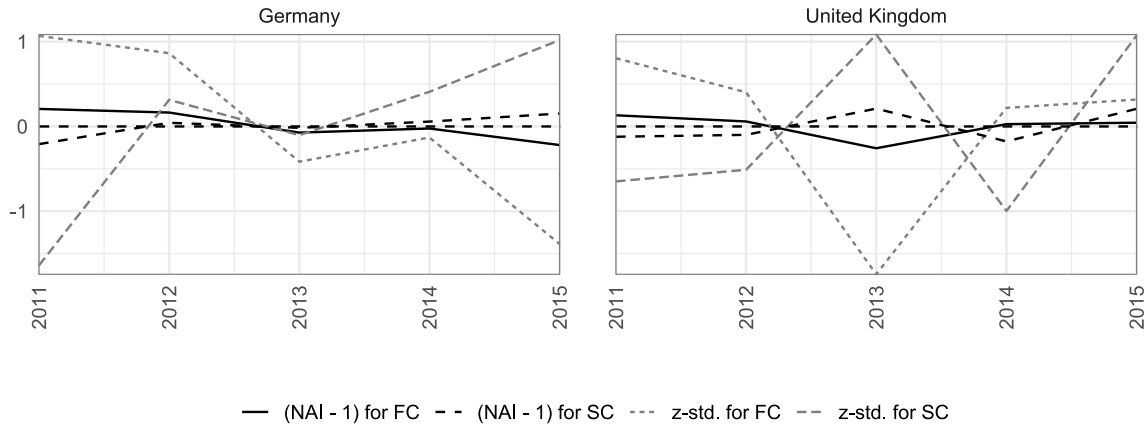

Figure 11: Granted patents as over years z-standardized LAI and as NAI (shifted down by 1 to scale to z-std.).

We conclude, that the NAI approximates the z-standardization, so the interpretation of the NAI results in the same statements. Hence, we can derive from the NAI in which years a field in a country performs over or under its own mean performance.

6 Conclusion

By the NAI we get a well-motivated new index, that is a natural expansion of the AI. It helps to identify performance over and under a country-field specific mean. It approximates the z-standardization and therefore has nearly a normal distribution. Thereby it omits some of the disadvantages of the z-standardized LAI – it has lower loss of data, is easier to calculate and easier to interpret. We advise the reader, to

dismiss the triple for interpreting data of AI, LAI and z-standardization and instead use the new triple AI, LAI and NAI, as the AI is a size-adjusted index to compare the field activity of countries; LAI is the normal distributed transformation of the AI that is useful for further analyse techniques as regression; NAI is the normalized AI and while LAI is useful for analysing data on the macro level, the NAI is strongest in analysing certain country-field combinations to highlight characteristic points in longitudinal data.

For the future we claimed several statements that should be proven. The NAI approximates the z-standardization and therefore is approximately normal distributed, and the AIAM and the arithmetic mean of the AI tend to the same value. These statements can be proven analytically or by a Monte-Carlo experiment. Further, there are more interpretation levels than we presented. The NAI cannot only be presented for a certain country-field combination, but for example also for a certain year-field combination, because the NAI is symmetrical in its indices. And last but not least, if the division of the AI by the AIAM corresponds to the normalization by the expected value, is there a technique or operation corresponding to the normalization by the standard deviation, too, to draw the NAI even nearer to a distribution with expected value of 1 and standard deviation of 1?

Statements and Declarations

There are no competing interests. The choice for EPO data is based on the open-accessibility and scientifically approved standard of this database – otherwise this choice was arbitrary. Except of figure 3, all figures were made with the statistical software R (4.1.0; R Core Team, 2021) and the graphical package ggplot2 (3.3.4; Wickham, 2016). Figure 3 was made with Inkscape (Inkscape Project, 2020).

The content of this paper was first presented as a poster at the proceedings of the 18th International Conference on Scientometrics & Informetrics (ISSI2021; Fuchs, 2021) and was published as preprint at Research Square (www.researchsquare.com; Fuchs, 2021a).

References

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

EPO (2021) European patents granted by field of technology. https://www.epo.org. Accessed 25.03.2021.

Fuchs, J.E. (2021). The three-dimensional activity index. *18th International Conference on Scientometrics & Informetrics ISSI2021. Proceedings*, 1474-1475.

Fuchs, J.E. (2021a). The three-dimensional activity index – an alternative transformation to logarithmic calculus of the activity index, preprint version 1, *Research Square (https://www.researchsquare.com/article/rs-939228/v1)*, DOI: 10.21203/rs.3.rs-939228/v1.

Grupp, H. (1994): The measurement of technical performance of innovations by technometrics and its impact on established technology indicators. *Research Policy* 23, S. 175-193.

Inkscape Project. (2020). Inkscape. Retrieved from https://inkscape.org.

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.

R Core Team (2021). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/.

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.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag,* New York.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- grantedpatents20112020pertechnologyfieldpercountryofresidenceen.xlsx