July 2019

The correlation between the number of patents and the patent portfolio value of companies – a statistical analysis

Dierk-Oliver Kiehne (InTraCoM GmbH)

Introduction

When talking about patent valuation, literature sources offer a big variety of suggestions how to do it. Typically, those are derived from cost, income or market analogy approaches. Also, indicator-based approaches are found more and more in literature[1], they offer a better possibility to compare patents and even patent portfolios using statistical methods. In combination with the market approach even monetary value ranges can be assigned.
Generally, patents are assumed to be early innovation indicators: they are typically applied long before a product is introduced in the market where the invented technology/procedure or formulation is used, thereby they are an early prediction source for technologies[2][3]. According to the uncertainty about patent quality it may make sense to simply have a look at the quantities of patents that are filed. Many studies regarding innovative abilities simply use the plain application data of patents, to e.g. compare regions or companies. On the other hand, it is also known that not every invention matures into an innovation. It is widely known that only a small share of inventions is finally successful in the market. Furthermore, it can be assumed that those successful products are based on valuable patents.
Within this study it will be investigated how strong values are correlating to application numbers. It uncovers the typical statistical traps.

Description of the methods used

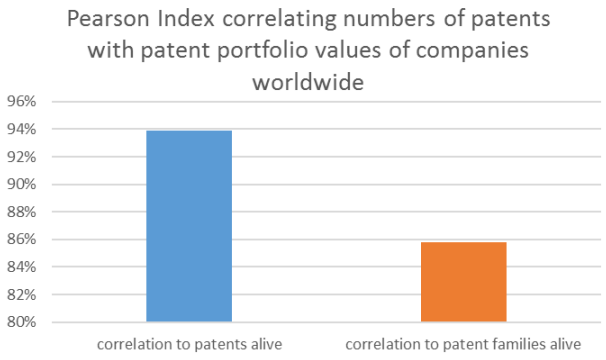
For the statistical analysis datasets of all patents worldwide were considered. In order to have comparable data, all values and data were taken from a specific time 2018-09-30. Apart from the raw patent data, also the patent value data (valued patents) were used[4]. The value data were compared to the alive single patents as well as the alive patent families[5]. All patents were properly assigned to their ultimate owners. In order to be able to also compare regions, all patent owners were assigned to their origin countries. Additionally, revenue data of the single patent owners sourced from Bureau van Dijk were used in order to determine the size of a certain company. The main industrial sector was also determined analysing the patent portfolio of all patent owners by taking the IPC classification of each patent family into account. In order to avoid

[1] Squicciarini, Mariagrazia, Hélène Dernis, and Chiara Criscuolo. "Measuring patent quality." (2013).
[2] Benson, Christopher L., and Christopher L. Magee. "Quantitative determination of technological improvement from patent data." *PloS one* 10.4 (2015): e0121635.
[3] Fankhauser, Manuel, Christian Moser, and Theodor Nyfeler. "Patents as early indicators of technology and investment trends: analyzing the microbiome space as a case study." *Frontiers in bioengineering and biotechnology* 6 (2018): 84.
[4] Patent value data taken from emposis IP-Business Information patent valuation platform. Patent values are assigned using the indicator based market analogy approach.
[5] Simple/DOCDB family definition where all family members refer to the same priorities

statistical superelevation of small portfolios, only companies that contained at least 10 alive patent families were used for the study. Very small companies or those with just one patent might have had a strong bias on the statistical analysis. In all the analysis the statistical representativeness was considered.

Results

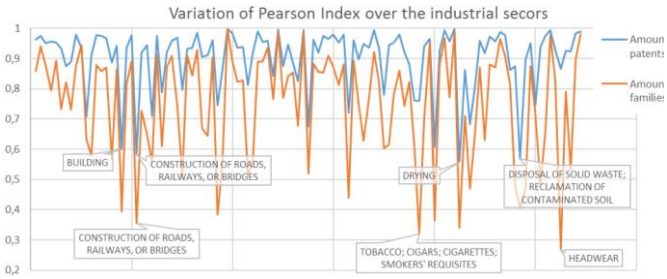
The most typical assumption is that patent values correlate to the patent filings. Patent filings is the most typical way to use patents for any kind of analysis or prognosis. Here often the patent applications are used as a measurement instrument no matter if they ever have a chance to survive the official examination process. For the ownership of patents it must made clear that at least those that were revoked, declined or are ceased are not taken into account. So for this analysis only the alive patents as an accurate comparison value will be used.
For the comparison simple statistical methods like the Pearson Correlation Score and average values are calculated. Even more accurate is the usage of patent families – they are describing an invention in the different jurisdictions:



Graph 1: Pearson correlation between the total amount of alive patents and the patent portfolio values as well as the alive families correlated to the patent portfolio values of all commercial patent owners.

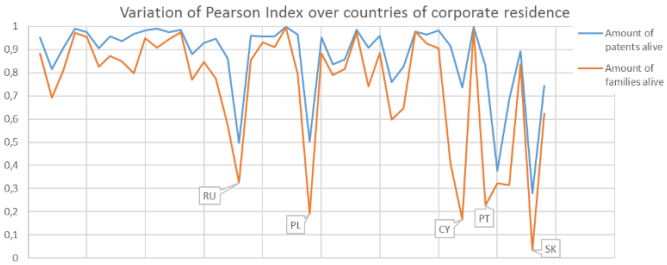
Graph 1 shows at a very first sight a strong correlation between the total amount of (alive) patents (applications and granted patents) of patent owners and their patent portfolio values. The correlation can be seen as significant with 94%. Not that significant are the patent family values in correlation with the portfolio values: Even though patent families describe the amount of inventions and are supposed to represent the innovative abilities better than the amount of all filings (also the same invention in different jurisdictions), the correlation score is just 86%. This value is surprisingly lower but still significant. The problem of this simple statistical analysis is that strong divergences are hidden in the big mass of data. And for most applications those strong deviating companies, are the interesting ones e.g. those where the values are unexpectedly higher than the number of patents expected it. Depending on the main application of this information it may lead to severe statistical misinterpretation: It can be compared to the statistical averaging problem of income where a beggar and a millionaire both have a statistical 500k fortune.
In order to show the problems of this kind of statistical correlation analysis in common, the same analysis was done for different industrial sectors: The whole set was separated into industrial sectors, these were found by analysis of the different patent portfolios of all companies. In order to have a statistical validity, only those sectors with more than 30 companies have been considered.
Graph 2 shows that there are strong variations of the Pearson index when different sectors are compared. The strongest deviations are in sectors like "Disposal of solid waste", "drying", "construction of roads, railways" or "building". According to the former analysis, the variations are even stronger when patent families were considered. The biggest difference and the smallest Pearson index was found in the sector "Headwear" (e.g. helmets, protection wear) – the Pearson index goes down to 27%. In this sector were 42 different companies considered. The biggest group with the smallest statistical significance was the sector "Construction of roads, railways or bridges": here the Pearson coefficient goes down to 35% (patent families) respectively 58% (alive patents). In this group 681 companies were considered.

InTraCoM GmbH | Lengsdorfer Hauptstr. 73 | 53127 Bonn / Tel.: +49 (0) 711/ 797 328-0 | Fax: +49 (0) 711/ 797 328-29 | Internet: www.intracomgroup.de



Graph 2: Pearson correlation coefficient depending on the different industrial sectors.

In a second analysis the country of origin was investigated. All companies were selected by their specific residence. For all these countries the Pearson correlation was calculated. Also here a strong variation is visible. According to expectations, smaller countries (with respect to their total amount of registered companies) were also leading to smaller Pearson correlations. Also here only those countries were taken into account, where at more than 30 companies (with at least 10 alive patent families) were registered.



Graph 3: Correlation coefficient in comparison to the different company's' origin

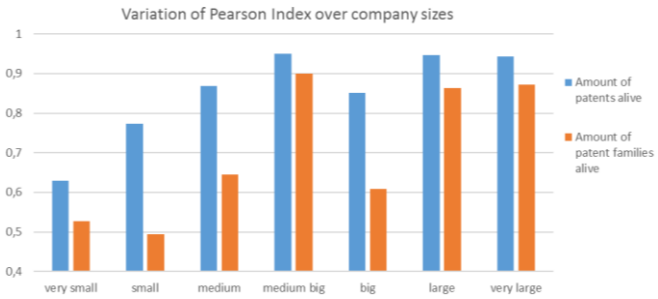
Again, the pattern, where smaller statistical groups (in this context countries with less registered companies having at least 10 alive patent families) tend to have smaller correlations, can be seen here. Very surprising that the correlation coefficient goes down to even 3% (alive patent families) respectively to 28% (alive patents) in Slowak Republic. 35 different companies have been considered here, a comparably small group. But e.g. the Russian companies have a similar small correlation coefficient: 32% (alive families) and 50% (alive patents) where 603 companies have been participating in the analysis, a comparably big group. The other obvious deviations were in Poland (332 companies), Cyprus (85 companies) and Portugal (58 companies). Again, the analysis shows big differences in certain countries so that also here a general high correlation expectation between value and total number of patents /patent families leads to wrong conclusions.

A third segmentation was done, using the revenue sizes of the different companies.

Very small companies having less than 1 m€ revenue in their last available balance sheet, small up to 5m€, medium up to 10m€, medium big up to 100m€, big up to 1bn€ large up to 10 bn€ and very large more than 10 bn€ of revenue according their last balance sheet.

The results are even more surprising than the others before: There's no clear, linear dependency between size (and expected total amount of patents) and correlation coefficient. Surprisingly the big sized companies (with 4,364 companies in that group) show a smaller correlation between alive patents and the total patent portfolio value (85%), even more extreme the low correlation between the alive patent families and the value (61%), even though in the two neighbour-groups the correlation is bigger, where 15,954 medium-big and 969 large companies were considered. For the group of small (5,541 companies) and very small companies (4,865) the correlation is also much smaller. The lowest correlation between amount of alive patents and the portfolio value is seen at the very small companies (Pearson coefficient 63%) and the small companies with 78%. For the group of small companies also the Pearson correlation varies between the number of patents (78%) and the number of patent families (49%), which can be explained by bigger patent families for this specific size group. Generally, a big variety of the correlation among the company size groups is seeable, where no clear pattern seems to be available.

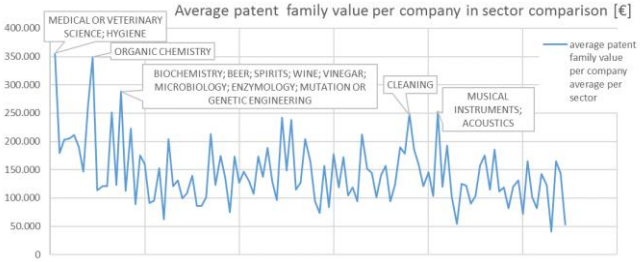
InTraCoM GmbH | Lengsdorfer Hauptstr. 73 | 53127 Bonn / Tel.: +49 (0) 711/ 797 328-0 | Fax: +49 (0) 711/ 797 328-29 | Internet: www.intracomgroup.de



Graph 4: variation of the correlation coefficient depending on the company sizes.

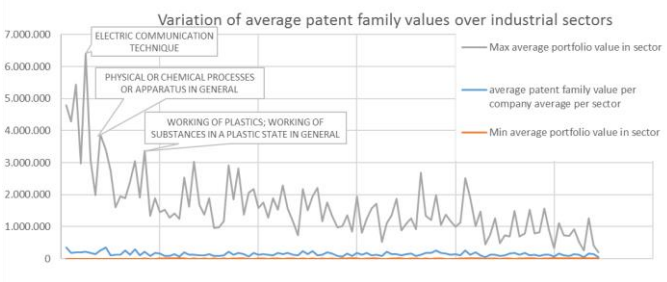
For all outlined cases the correlation coefficient leads to strong variations depending on the different grouping that was used.
However, it is still not clear what the reason for the strong differences in the correlation coefficient is. Generally, this can be explained by very poor values compared to the total amount of patents and compared to the average or it may be due to very high values. Of course, this is of central importance: the high value patents indicate that here may be innovative products (in future).
That is why in a second set of analysis the average patent family values were calculated – for the same segmented groups as used before.

In the first analysis the different sectors were compared again. Here the averages of all average patent family values per sector were considered.



Graph 5: "average of averages" - variation of the average patent family value: sector averages in comparison. The highest sector average is found in "Medical or Veterinary Science", the lowest sector averages in "yarns; mechanical finishing of yarns or ropes; warping or breaming"

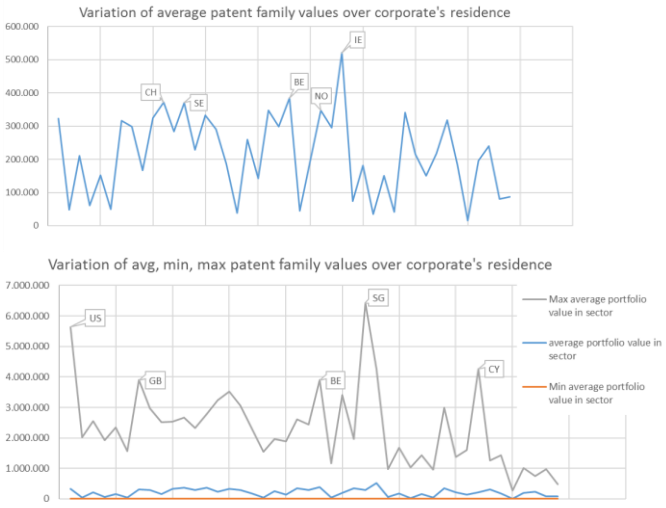
The graph shows extremely strong deviations even though here average values of company averages in the different sectors are shown. This represents the strong heterogeneity of patent portfolio values for single companies. As shown before the different sector sizes are from 31 to 7,327 companies.



Graph 6: the maximum (upper line), averages (same line as in Graph 5) and minimum average values per company shows also the strong variation inside sectors.

When looking at the variations of average values the strong heterogeneity becomes even more visible. The highest average value for a certain company was measured in the sector Electric Communication Technique.

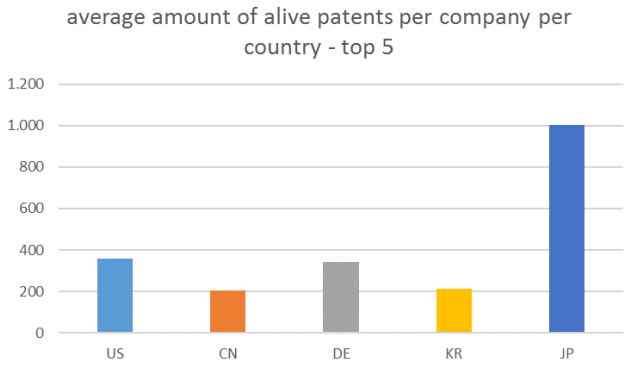
A similar picture comes with the comparison of the different countries, where the companies are registered. The companies based in the countries Singapore, USA, Cyprus have the highest average patent family values – in country average. Having a look at the highest averages per company Ireland, Belgium and Switzerland have the highest values in their companies. Very surprising.



Graph 7: upper graph: the average values over the companies' origin (country of residence) and the max, min and average values in comparison, lower graph.

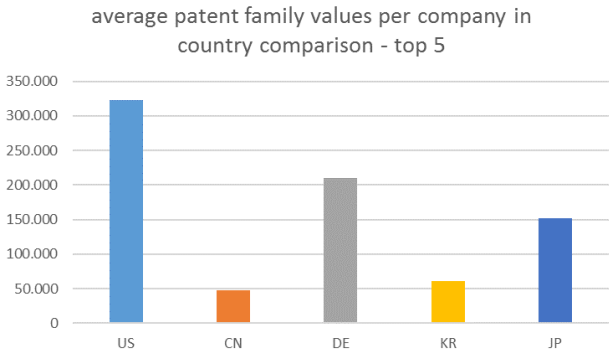
InTraCoM GmbH | Lengsdorfer Hauptstr. 73 | 53127 Bonn / Tel.: +49 (0) 711/ 797 328-0 | Fax: +49 (0) 711/ 797 328-29 | Internet: www.intracomgroup.de

In the beginning if the study it was shown that the highest Pearson correlation was found in the correlation between value and amount of alive patents (vs. alive families), that is why the amount of patents and average values for the top countries is analysed: When looking at the 5 top regions where most patents are filed (patenting companies with their origin in the specific countries), the huge difference in quality versus quantity is impressively visible. The Japanese companies are holding most alive patents in average (1,005). The other regions are comparably similar distributed: the average amount of alive patents is between 213 (CN) and 359 (US).



Graph 8: Average amount of patents per company for the top 5 filing countries (companies with their origin in these countries). Japanese companies hold 3 times more patents than the others.

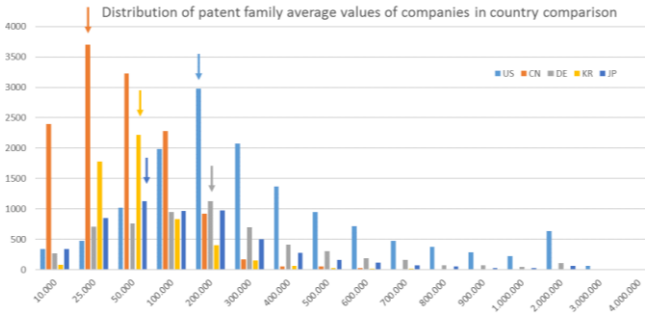
However, when comparing the average values for these regions, the picture is way more differentiated.



Graph 9: Average patent family value per company for the top 5 filing countries (companies with their origin in these countries). US located companies have in average the double average values than the Japanese ones even though they have 3 times more patents. The lowest average values were found for the Chinese companies in the top 5 comparison.

Japan is now only the third, US is leading the ranking in the comparison. But what is more is the huge differences in average values between the first (US: 322,000) and the last in the ranking (CN: 47,000) of these top 5 filing countries.

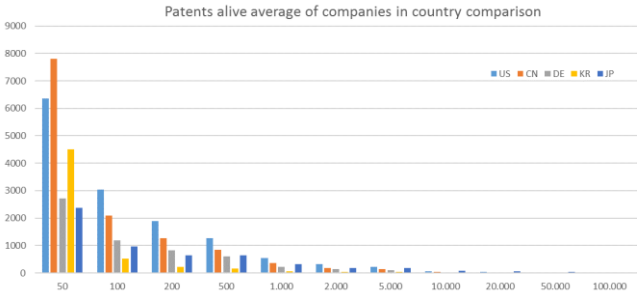
For making this even more transparent and understandable, instead of averages histograms of the value distributions were built.



Graph 10: Average value distribution histogram of the top 5 patent filing countries (patent filing companies with their origin in these countries)

For all shown countries a Rayleigh distribution of values is obvious, but their vertex is completely different what was already expected when looking at the averages before: Chinese companies have the vertex of their distribution between 10,000 and 25,000 € average patent family value. So most companies (app. 3,700) are in this average patent family value range. In contrast to that most US companies (app. 3,000) have an average value range of 100,000 to 200,000€. The scaling is not linear that means that the distance between these two countries is even bigger than this graph suggests. And the distribution offers a second obvious peak for the US companies: 639 companies have an average patent family value of 1-2 m€.

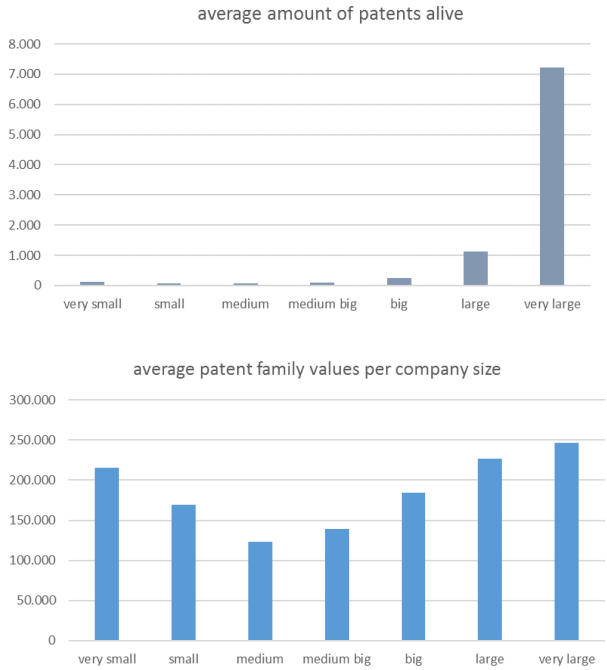
InTraCoM GmbH | Lengsdorfer Hauptstr. 73 | 53127 Bonn / Tel.: +49 (0) 711/ 797 328-0 | Fax: +49 (0) 711/ 797 328-29 | Internet: www.intracomgroup.de



Graph 11: How many different companies are holding how many patents: on the x-axis the amount of patents, the y –axis the amount of companies.

In contrast to that, the pattern for the total amount of filings is similar for all the compared countries. It looks like an exponential distribution (Graph 11). It shows that most companies have a smaller patent portfolio. The distribution thereby does not show obvious significant statistical abnormalities. It also shows that the simple amount of filings does not bring any benefit in terms of differentiation.

Finally, the company sizes regarding their average patent family value and the average amount of filings have been analysed. The main determinants of this analysis is how efficient a patent department works: Is every invention filed – which may lead to a lower average value – or is the filing of a patent the result of internal selection process, like a stage-gate-process.

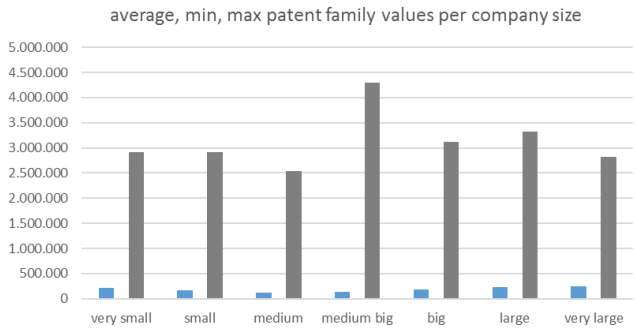


Graph 12: Upper graph shows the average amount of filed alive patents per company size- the biggest companies also hold most of the patents. Lower graph: Big companies also have the best patents? Yes, in average they do but surprisingly also the very small ones seem to have excellent patents. Are here the tomorrows' innovators?

In total the biggest companies also have in average the best average patent family values. Of course, also the assignee size influences the value of a patent due to factors like market access or the ability to push technologies to standards. But on the other hand, it can also be assumed that there are many patents filed that are not primary part of a R&D strategy, maturing in lower patent values.
Surprisingly also the smallest companies are supposed to have very strong patents: so, the small and very small companies represent an anti-trend to the thesis above. Here also the most interesting candidates for investing are expected.
Way less surprising is the average amount of filed patents. Expectedly the biggest companies have most patents filed.

When having a look at the maximum values the trend is completely different to the averages: The highest total average values are found at the medium big companies: 4.25m€. For all the other groups the maximum average patent values

seems more or less equal and ranges between 2.5 and 3.3 m€. However, this was also the biggest group in the study.



Graph 13: Minimum, average and maximum average values. Minimum values are according to the scale hard to identify in the graph but all range between 0.5 and 1,500 €. Poor values but still these refer to alive patents.

Summary

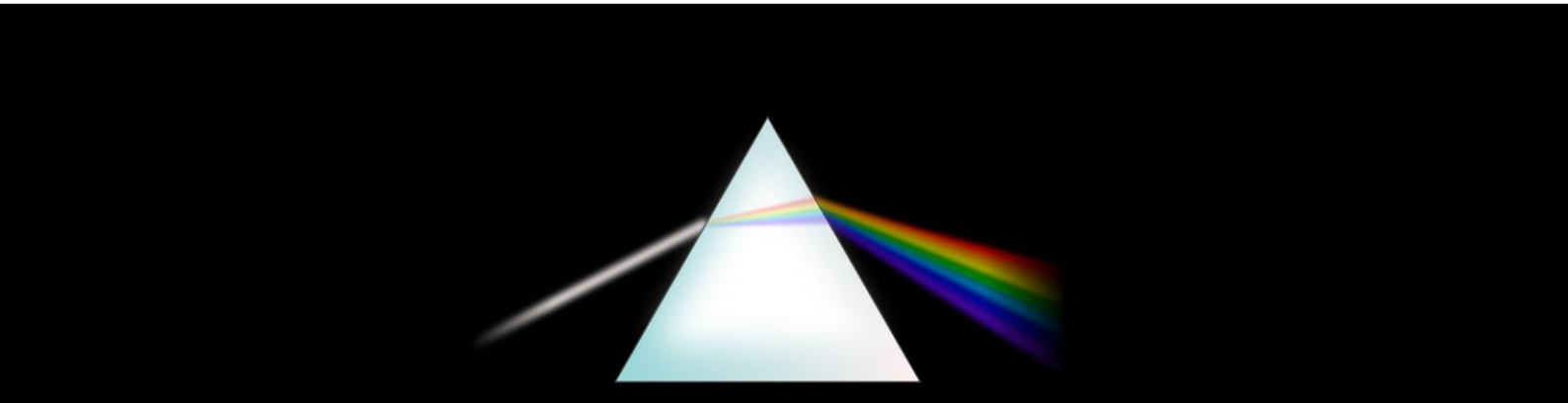
Statistics may be misleading: the bigger the amount of trials, the more likely the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performed. This phenomenon is known as the law of large numbers. Translated for patents and their values this means that if the complete world of patents and their values are correlated, then the statistical correlation is high. When doing a segmentation, e.g. by comparing groups of industrial sectors, countries or company sizes, the correlation becomes more and more less significant. Very impressive is this when it comes to the comparison of the lowest and the highest average patent family values that ranges from 0.5€ to 4.3 m€ - these are 7 orders of magnitude, talking about average values, no absolute patent family values.

This study has shown that no matter what kind of grouping was taken, the number of patents and the average values come to completely different pictures and rankings. Also, the Pearson correlation coefficient varies in these different groups which indicates that the correlation is not constant at all.

For all cases where differentiation is the objective, e.g. by finding the most promising investment candidates or the most innovative sectors, the counting of patents would have led to the wrong result lists. Again, quantity is not equal to quality.

Contact:

Dr. Dierk-Oliver Kiehne
InTraCoM GmbH
Lengsdorfer Hauptstr. 73
53127 Bonn
www.intracomgroup.com
inf@intracomgroup.de
+49 711 7973280



Picture source : Suidroot https://commons.wikimedia.org/wiki/File:Prism-rainbow.svg, this picture was modified