

# Determinants of Patent Protection Regimes: A Self-Organizing Map Approach

Caner Demir, Raif Cergibozan<sup>1</sup>

**Abstract:** This study investigates the determinants of patent protection regimes with the use of the self-organizing map (SOM) method. Unlike any previous attempts in the existing literature, this paper takes into consideration the lack of analytical techniques in the past and tries to demonstrate a potential relationship between the patent protection and its determinants through the employment of a newer, more consistent method. The study consists of two main parts. Firstly, the patent protection strengths of 111 countries have been classified via a SOM-based model and it turns out that three types of clusters can be found around the world; low-, mid- and high-protection. The results also show that the densities of these clusters have dramatically changed in the post-1980 period. In the second part of the study, the determinants of the patent rights are examined for 49 developed and developing countries. After revisiting the older econometric models with recent data, this study also analyses the determinants with the SOM method. The findings suggest that there is a significant relationship between GDP per capita, human capital, R&D, market freedom, political rights and patent protection for about two-thirds of the sample; which implies that the patent policies of these countries are in accordance with the selected economic and social factors.

**Key words:** intellectual property rights, innovation, self-organizing map, artificial neural networks

**JEL Classification:** O34, C45

*Received: 23 December 2017 / Accepted: 16 May 2018 / Sent for Publication: 3 September 2018*

## Introduction

The subject of intellectual property rights (IPR) has been taking up a dominant place in innovation debates and this trend has naturally motivated researchers to investigate IPR-related topics. Once it has been understood that technological progress and correspondingly innovative activities have a significant impact on production possibilities, the sources and determinants of innovation have become an important focal point. There are many observable and unobservable determinants of technological progress. Today's research on technological progress and innovation solely focuses on the observable

---

<sup>1</sup> Kırklareli University, Department of Economics, Kırklareli, Turkey, caner.demir@klu.edu.tr, raif.cergibozan@klu.edu.tr.

determinants such as research and development (R&D) expenditures, number of researchers, number of patents, etc. We may consider the R&D expenditures and number of researchers to be exogenous factors because the authorities can control these factors by implementing any necessary policies. However, the number of patents is a different case. New patents are nothing different from new creative ideas which are considered “innovative”; and innovation cannot be made directly by a political intervention of the authorities. Today, among prominent international institutions such as the World Trade Organization (WTO) and the World Intellectual Property Organization (WIPO), the subject of IPR is perceived as one of the main driving sources of innovation and it is hereby put in the centre of innovation debates.

Although the origins of the idea of intellectual property protection date all the way back to the 15<sup>th</sup> century, the greatest economic impacts of IPR emerged in the post-industrial revolution era and continue to exist until today (Hall, 2007: 569). Nowadays, the mainstream perspective on IPR suggests that the protection of ideas constitute an incentive mechanism for creators and inventors; without this protection mechanism, no one is expected to create new ideas (WIPO, 2011:5).

In the academic literature, IPR protection has been measured via several indices developed by Rapp-Rozek (1990), Seyoum (1996), Sherwood (1997) and Ginarte-Park (1997). However, the Ginarte-Park Index (GPI), which gets updated every five years and has a broader scope, is at the forefront. The value of GPI ranges from zero to five and the index consists of five sub-categories; the coverage, the membership status in international treaties, the duration of protection, the enforcement mechanisms, and the restrictions on patent rights (Park, 2008). The components of each subcategory are laid out in detail in the third section. The GPI is also used to demonstrate the strength of the protection, while the SOM approach allows us to observe the clusters formed by countries with similar IPR protection strengths.

According to the Convention Establishing the World Intellectual Property Organization, which was signed in 1967 and revised in 1979, any member country of the Paris Union or the United Nations is also accepted as a member of the WIPO. As of June 2018, there are 191 member countries in the WIPO.

Already well-known from the theory of economic development, it is irrational to expect that these 191 countries are all in the same developmental path. Moreover, in the member list there are high-protective and low-protective countries as well as there are developed and developing countries. In other words, it is clear that the WIPO members do not constitute a homogenous sample. Here we can make the first proposition of the study which is going to be tested in the empirical analysis section: even though the mainstream perspective on IPR requires a globally harmonized IPR regime, the WIPO members constitute a heterogeneous sample and might be located as separated clusters.

Another objective of the study is to investigate potential determinants of IPR protection. In the existing literature, the only empirical study of the determinants of patent protection so far has been Ginarte and Park (1997). However, due to certain statistical shortcomings which shall be described in the next sections their estimation methods and examined the time periods are inconsistent and inadequate respectively. In our paper, we investigate the potential determinants of IPR with the SOM approach which is a non-parametric method. The other analysis of the study concentrates on some outstand-

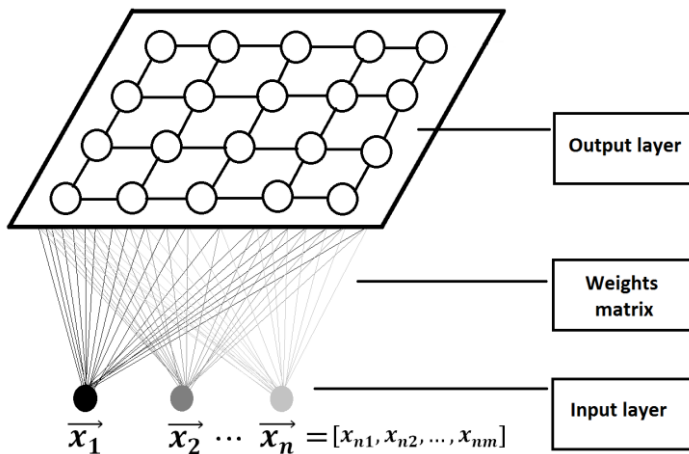
ing potential determinants: market freedom, GDP per capita, human capital, the share of R&D, political rights, and openness. The second proposition of our study is that these factors might determine the strength of IPR protection, and moreover, they may vary across country groups.

First, we shall outline the methodology of the SOM approach, follow with our findings.

## Methodology

SOM also known as Kohonen map is an effective software tool used to examine and analyse existing patterns in high-dimensional data sets. Basically, SOM provides the graphical visualisation of the similarity between the input data. It also offers important advantages to analysts by supplying low-dimensional images of high-dimensional data. SOM is an unsupervised, competitive learning methodology first introduced to the literature on artificial neural networks by Kohonen (1982). Unlike many other analysis techniques, SOM is a non-linear and non-parametric analysis method which is not based on any strict assumptions. The topology of the SOM network is shown in Figure 1.

**Figure 1. The topology of the SOM network**



Source: Mostafa (2010) and authors' own visualisation.

Each circle in the figure shows neurons – usually pictured as hexagons in the analyses. Since the grid is 4x5, this network has 20 neurons. This particular SOM network consists of a two-layer node: input and output. It is important to note here that each output layer has one coordinate and allows us to easily calculate easily the distances between output layers. They will hence be located in a two-way dimensional map according to the similarity of each output layer weight. Neurons of similar weight locate themselves close to one another, while dissimilar neurons are far from each other. The self-organizing process involves four basic elements: initialization, competition, cooperation, and adaptation. The SOM algorithm comprises the following steps (Vesanto and Alhoniemi, 2000; Sarlin, 2013):

*Step 1. Initialization:* All connections are provided with their initial weights.

*Step 2. Competition:* Using the Euclidean distance, each input vector  $x_j$  is compared with each output vector  $m_i$  and the best match  $m_b$  is defined. The winning node is commonly called the Best Matching Unit (BMU).

$$\|x - m_i\| = \min_i \|x - m_i\| \quad (1)$$

Here, the smallest Euclidean distance might be considered the best match point.

*Step 3. Cooperation and adaptation:* In the input space, the BMU and its topological neighbours are located close to the input vector. The update rule for the prototype vector of unit  $i$  is

$$m_i(t+1) = m_i(t) + \alpha(t)h_{bi}(t)[x(t) - m_i(t)] \quad (2)$$

where  $t$  is time,  $\alpha(t)$  is the adaptation coefficient and  $h_{bi}(t)$  is neighbourhood kernel centred on the winner unit. The neighbourhood function is often calculated using the Gaussian neighbourhood function.

$$h_{bi(j)} = \exp\left(-\frac{\|r_b - r_i\|^2}{2\sigma^2(t)}\right) \quad (3)$$

where  $r_b$  and  $r_i$  are two-dimensional coordinates of the reference vectors,  $m_b$  and  $m_i$ , respectively, and the radius of the neighborhood  $\sigma(t)$  is a monotonically decreasing learning factor at time  $t$ .

*Step 4.* Step 2 is repeated for the sufficient iteration procedure of convergence process.

*Step 5.* Using the SOM-based k-means clustering method, the nodes are gathered into a small number of clusters.

As stated by Mostafa (2010), to be able to see the structure of the clusters, Unified Distance Matrix (U-Matrix) is used in many studies. But it is very difficult to do any classification using U-Matrix due to the lack of the clarity of clusters' visuality. In this paper, to avoid such an issue, the SOM-based k-means clustering method is used and different clusters are thus seen more clearly.

## Empirical Analysis

The empirical part of the study consists of two main analyses. In line with the propositions given in the introduction, we are first going to carry out a SOM analysis to determine the potential patent protection regime clusters around the world. The purpose here is to investigate whether the policies of WIPO ensure a globally harmonized IPR regime; in other words, whether the country-specific patent protection regimes constitute a homogenous sample. For this analysis, a data set covering 111 developed and developing countries within the 1960-2010 period has been used.

The second analysis of the study aims to investigate any potential determinants of the patent protection regime. In order to do so, the study first revisits the Ginarte and Park (1997) by employing the classic econometric models estimated via the ordinary least squares (OLS) and the generalized least squares (GLS) methods. Next, the same inves-

tigation is applied via the SOM method. Due to the lack of data, the second part of the empirical analysis only covers 49 countries for the 1996-2010 period.

Note that for both analyses, the strength of patent protection regime has been measured with the Ginarte-Park Index which was last updated in 2010. The subcategories and components of the GPI are given in detail in Table 1.

**Table 1. The subcategories and components of the Ginarte-Park Index**

1. Coverage	Available	Not Available
Patentability of pharmaceuticals	1/8	0
Patentability of chemicals	1/8	0
Patentability of food	1/8	0
Patentability of surgical products	1/8	0
Patentability of microorganisms	1/8	0
Patentability of utility models	1/8	0
Patentability of software	1/8	0
Patentability of plant and animal varieties	1/8	0
2. Membership in international treaties	Signatory	Not Signatory
Paris convention and revisions	1/5	0
Patent cooperation treaty	1/5	0
Protection of new varieties (UPOV)	1/5	0
Budapest treaty (microorganism deposits)	1/5	0
Trade-related intellectual property rights (TRIPS)	1/5	0
3. Duration of protection	Full	Partial
	1	$0 < f < 1$ *
* where f is the duration of protection as a fraction of 20 years from the date of application or 17 years from the date of grant (for grant-based patent systems).		
4. Enforcement mechanisms	Available	Not Available
Preliminary (pre-trial) injunctions	1/3	0
Contributory infringement	1/3	0
Burden of proof reversal	1/3	0
5. Restrictions on patent rights	Yes	No
Working requirements	1/3	0
Compulsory licensing	1/3	0
Revocation of patents	1/3	0

Source: Park (2008)

Before we investigate the importance of the potential determinants of the IPR protection, let us start with the first of our analyses which tries to classify countries according to their patent protection regimes.

### **A Self-Organizing-Map Based Classification of Patent Protection Regimes**

In this subsection, the patent protection regimes of the examined countries are classified by clustering similar protection strengths via the SOM method. To make such a classification, the IPR protection levels of countries have been used as the inputs in Figure 1, which shows the topology of the SOM network. Since Walter Park, who currently updates the GPI, provides new data only once in every five years, a five-year span dataset, giving the time dimension of the value of 11, has been used.<sup>2</sup> This should be a reasonable choice considering IPR policy implementations do not occur too often. Thus, the first SOM analysis of the study uses each country's IPR strength for each year resulting in 1232 inputs in total.

Figure 2 displays the component matrices which clearly show each year's cluster structure and the U-matrix. Following the example of Sarlin (2011) and Sarlin and Peltonen (2013), to cope with the convergence issue the data used in the analysis has been normalized by considering the variance values. In the SOM analysis, the component matrices correspond to variables used in the building of the U-matrix. In the study it is clear to see that the time dimension of the data constitutes the variables and enables us to observe the variation in time. Here, the dark neurons represent countries with the lowest IPR protection, while white neurons indicate countries with the highest IPR protection. In other words, the lighter the neuron cells are, the stronger the protection is, and vice versa. It could be thus said that component matrices close in shade of colour can be taken for years with similar protection levels.

It is important to note that due to the normalizing process applied to the data, the scale of each component matrix corresponds to different intervals. Therefore, even if the colour arrays of certain components look similar, the intensities of the clusters are different. It is evident from the component matrices that the average values on the scales increase in time, implying that the general global protection strength rises in time as well. It is hence better to evaluate each component matrix using its individual scale.

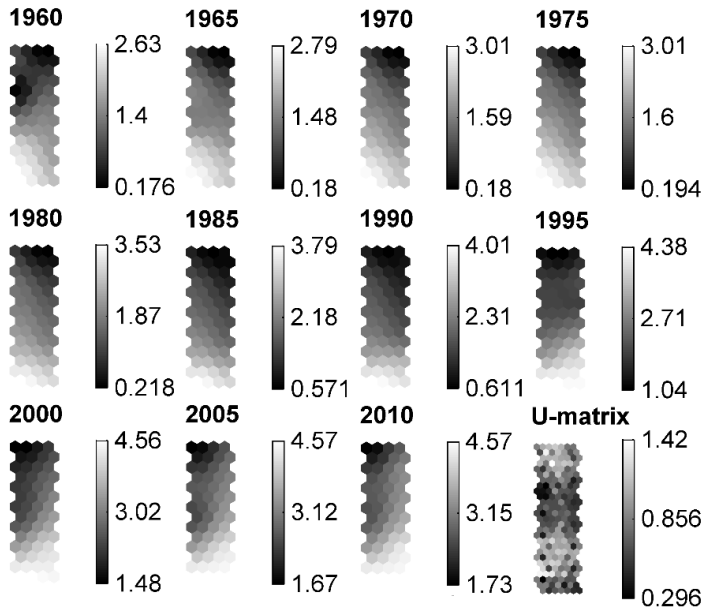
According to the results, there had been only a small increase in the IPR protection until the 1980s. This fact can be observed looking at the mean values on the scales, which were 1.40 in 1960, 1.48 in 1965, 1.59 in 1970, and 1.60 in 1975. As for the post-1980 period, the protection tightens. The upper value of the scale suddenly increases to 3.53 in 1980, which implies a global transformation in patent protection regimes. The mean value for 2010 is 3.15, while in 1980 it is 1.87. Moreover, in 2010 the top and bottom values of the scale are 4.57 and 1.73, while in 1980 they are 3.53 and 0.22. These values

---

<sup>2</sup> The last available update for the data comes from 2010. For that reason, we are unable to extend the data until 2015.

explicitly reveal that in the post-1980 period more protective policies on IPR were implemented around the world. The component matrices also show that the number of dark neurons dramatically decreases in time, which means the size of the low-protective cluster grows smaller. Most of the formally low-protective countries have shifted to the mid-protection path, which can be observed on the neurons of grey colour tones.

**Figure 2. Clustering patterns of patent protection regimes by years**

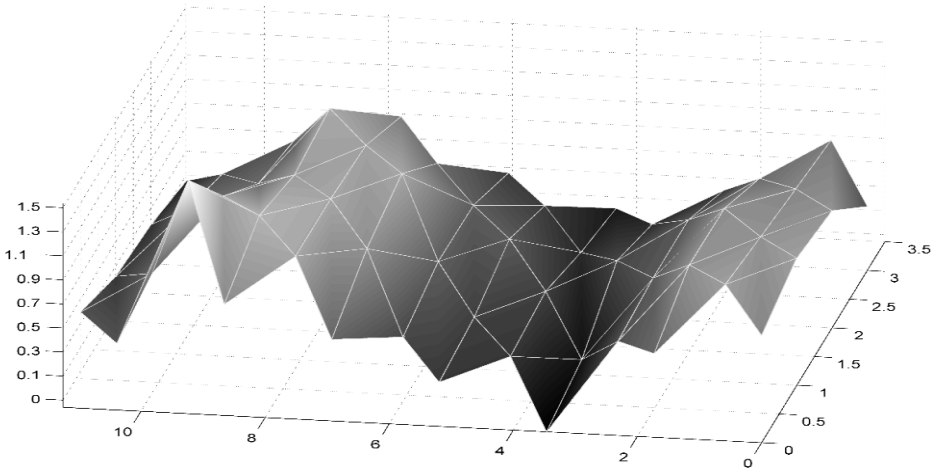


Source: Authors' own calculations. Analyzed and visualized using the SOM toolbox in the MATLAB.

Note: Each year has its own component matrix and the U-Matrix consists of the clustering patterns for each year.

The U-Matrix, which arises from the component matrices, is located in the bottom-right corner of Figure 2 and shows the clusters regarding the IPR protection. To interpret the U-matrix, it should be noted that the values on the vertical scale and colours in the matrix correspond with the distances between the clusters. Therefore, the interpretation of the U-matrix is different from the explanation of the component matrices. Consequently, dark colour implies clustering of countries with similar protection levels. The white and light grey areas are the transition zones (the borders) of the clusters. There are two transition zones in the U-matrix which are located on the top and bottom of the matrix with white and light grey neurons. Here, above the top transition zone, the low-protective cluster is located, while the high-protection cluster is below the bottom transition zone. The intensity of the transition zones can be also seen when the cluster distances and the two-dimensional U-matrix are applied together in Figure 3. From this 3D figure, the numerical value of the peaks and the troughs can be observed. Both Figure 2 and Figure 3 demonstrate that there are three different clusters around the world concerning the IPR protection.

**Figure 3. The distance matrix**

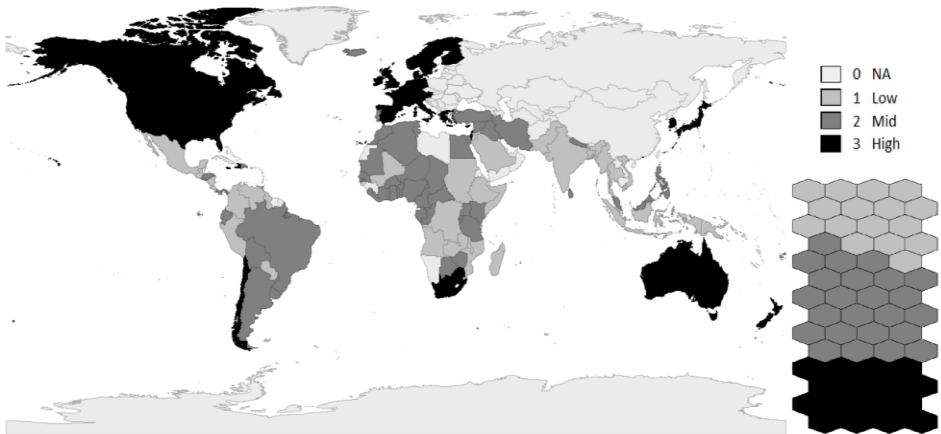


Source: Authors' own calculation. Analysed and visualized using the SOM toolbox in the MATLAB.

Note: Illustrated combining the U-matrix and the distances of countries. The vertical axis shows the level of distance.

These clusters can also be seen in the world map in Figure 4. The results say that out of the 111 sample countries altogether 30 of those are low-protective, 52 are mid-protective, and 29 countries are high-protective.

**Figure 4. IPR protection clusters around the world (1960-2010)**



Source: Authors' own calculations. Created by applying the results obtained from the MATLAB on the GeoDa software.

The results suggest that there still are three different patent protection regimes around the world, despite the fact that there has been a great strengthening pressure on coun-



tries since the post-1980 period. At this point arises the question of which macroeconomic and social indicators determine the patent protection policy. Thus, in the second and the main analysis of the study, the determinants of the patent policy are investigated. Moreover, as we already know from the first analysis of the study, the heterogeneity between countries is also considered when observing the significance of the determinants.

### **The Determinants of Patent Protection Regimes**

Patent protection regimes or policies are composed of laws and legislations the countries brought to force. From this aspect, it can be assumed that these policies are exogenously and independently implemented by these countries. However, as it has been stated and empirically demonstrated in the previous analysis, the international economic and political environment has a significant impact on the IPR protection strength. Especially in the post-1980 period, the protection standards have dramatically increased in almost each of the countries. The globalization process has transformed both the political and the economic structure of countries and neither IPR nor any other factor could be exempted from its impacts. In this section the possible relationships between patent protection regime and selected factors are hence examined.

The only empirical study on the determinants of IPR up to now belongs to Ginarte and Park (1997). Their study investigates the determinants using the ordinary least squares and generalized least squares methods. To establish an inclusionary study, our study revisits the paper of Ginarte and Park (1997) using the same variables: the same econometric methods are used, and then the SOM approach is applied due to the following issues of the econometric methods. First, the paper of Ginarte and Park (1997) estimates three different samples: the full sample, above the median sample, and below the median sample. However, as it has been empirically proven in our previous analysis, there are three different patent protection regimes around the world. In such a global environment, splitting the sample just based on the median may produce a weak analysis. Secondly, there is a high possibility of multicollinearity within the explanatory factors, and of endogeneity between these factors and the IPR index. These issues could lead to inconsistent and biased results. As a non-linear and non-parametric method, the SOM does not depend on any strict assumptions and is a better choice in this investigation.

The data used in the analysis, their definitions and sources are shown in Table 2. Since Penn World Tables provided by Feenstra et. al (2015) allow for more (balanced) observation of the selected period, GDP per capita and human capital index data are obtained from there. Due to some data constraints with R&D investments, our second analysis only covers the 1996-2010 period; and since the IPR index is calculated only once in five years, the time dimension of the second analysis is only 3. With 49 countries in the data set, our panel data sample has produced 147 observations in total. It should be pointed out that due to such data constraints, the first analysis did not include Ukraine in the assessment, but as the data for Ukraine is available for the 1996–2010 period, we have decided to use it in the second analysis. In accordance with the study of Ginarte and Park (1997), the variables that are investigated in this study as potential determinants of IPR are those of GDP per capita, human capital index, share of R&D investments, openness index, market freedom index, and political rights index.

**Table 2. Variables and data sources**

Variable	Definition	Source
IPR index	The strength of Patent Rights Protection	Walter Park, American University ( <i>freely released on W. Park's official web page</i> )
GDP per capita	Gross Domestic Product per head	Penn World Tables 8.0
Human capital index	Based on years of schooling and returns to education (for further details see Feenstra et. al, 2015)	Penn World Tables 8.0
R&D expenditures (% of GDP)	The share of R&D expenditures	World Development Indicators, World Bank
Openness index (X+M)/GDP)	The share of trade openness	World Development Indicators, World Bank
Market freedom index	Based on freedom from corruption, fiscal freedom, business freedom, labor freedom, monetary freedom, trade freedom, investment freedom, financial freedom	The Heritage Foundation
Political rights	The degree of political freedom. (Note that this is an inverse index varying between 1 and 7, where 1 implies a better situation regarding political rights.	Freedom House

*Note: The variables are selected by consulting the base study of Ginarte and Park (1997). For any further information, please see the websites of the data sources.*

Table 3 shows the panel regression results of our study following the methodology of Ginarte and Park (1997). The first two columns present findings obtained from the OLS estimator, while the other four columns present findings obtained from the GLS estimator. It can be concluded that GDP per capita has proven to be the most significant determinant of IPR for all of the estimated models. The share of R&D investments is assessed as significant by the OLS and GLS estimators for the full sample, while the above the median sample does not estimate it to be so. This is an interesting result which implies that R&D investments in relatively poor countries trigger a stronger IPR protection. This needs to be interpreted considering the necessities of developing countries where an increasing R&D requires better protected legal rights, which makes governments and policymakers implement stronger patent protection regimes. In addition, as the GLS estimation has demonstrated, human capital is one of the significant determinants of IPR for both the full sample and the above the median sample, which reveals that in relatively rich countries higher education standards pave the way for higher intellectual property standards. Such an effect cannot be observed in developing countries as the average education level is lower than the significant threshold level. The rest of the control variables have been estimated as insignificant or with unexpected signs. In this study we propose that despite its useful outcome, the OLS and GLS estimators give inconsistent and biased results because of the multicollinearity and endogeneity issues. Moreover, splitting the samples based on the median of the full sample is not a sufficient way to sort the countries as developed and developing. For this reason, we shall carry out the same models using the findings from the first SOM analysis of this study, dividing the countries into three subsamples.

**Table 3. OLS and GLS regression results**

Dependent variable: IPR						
	Full Sample				Split Sample	
	OLS(1)	OLS(2)	GLS(3)	GLS(4)	GLS(5)	GLS(6)
Constant	-0.599 (0.101)	1.060 (0.396)	-0.571 (0.145)	0.494 (0.425)	0.705 (0.271)	-0.247 (0.749)
GDPPC	0.198 (0.010)	0.154 (0.025)	0.195 (0.015)	0.119 (0.029)	0.085 (0.022)	0.136 (0.044)
RD		0.050 (0.012)		0.052 (0.015)	0.025 (0.016)	0.062 (0.023)
HC		0.069 (0.068)		0.189 (0.099)	0.174 (0.096)	0.183 (0.143)
POLIT		-0.026 (0.023)		-0.001 (0.0260)	0.030 (0.032)	0.013 (0.038)
OPEN		-0.001 (0.017)		0.006 (0.024)	-0.072 (0.020)	0.028 (0.044)
MARKET		-0.306 (0.105)		-0.121 (0.102)	-0.087 (0.061)	0.030 (0.178)
F [p-value]	366.3	78.5	172.0	210.2	48.5	69.0
$\chi^2$ [p-value]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
R <sup>2</sup>	0.72	0.77	0.72	0.76	0.19	0.59
Adj R <sup>2</sup>	0.71	0.76				
No. of obs.	147	147	147	147	74	73

*Note: Standard errors are in parentheses. All variables are in natural logarithms. The full sample consists of 49 countries observed over the three different time periods. The first-half of the split sample (column 5) includes countries above the median sample income and the second half (column 6) of countries below the median income.*

Table 4 presents results of the three different subsamples determined using an additional SOM analysis covering the 1996-2010 period. The findings demonstrate that GDP per capita is a significant determinant of IPR for the high-protection sample, as opposed to the mid- and low-protection samples. Contrary to the previous panel regression results, human capital is estimated as significant only for the low-protective countries. All of the other variables in Table 4 are assessed as insignificant or with an unexpected sign. Similar to the results in Table 3, the high-, mid- and low-protective subsamples suffer from the collinearity and endogeneity issues, which leads to biased and inconsistent coefficients in the regressions.

As we have already commented on the results from the econometric models and identified the insufficiency of these estimators, we can now move on with the analysis of the factors affecting the patent protection regime using the SOM method.

In this part of the analysis we examine whether the self-organized pattern of the potential determinants shows any signs of similarity to the patterns of the actual IPR patterns. In case it does, it is possible to state that the factors are significant determinants of IPR protection.

**Table 4. GLS regression results (subsamples identified with the SOM method)**

Dependent variable: IPR			
Cluster	High	Mid	Low
Model	GLS(1)	GLS(2)	GLS(3)
Constant	1.195 (0.281)	0.846 (0.578)	-0.248 (1.951)
GDPpc	0.099 (0.025)	0.054 (0.040)	0.033 (0.115)
RD	0.026 (0.018)	0.018 (0.018)	0.067 (0.046)
HC	0.036 (0.095)	0.129 (0.126)	0.513 (0.298)
POLIT	-0.001 (0.024)	0.056 (0.041)	0.049 (0.065)
OPEN	-0.027 (0.015)	0.031 (0.034)	0.042 (0.102)
MARKET	-0.188 (0.065)	-0.049 (0.162)	0.156 (0.399)
$\chi^2$ [p-value]	44.2 [0.00]	78.5 [0.00]	19.9 [0.00]
R <sup>2</sup>	0.23	0.30	0.41
No. of obs.	69	48	30
No. of groups	23	16	10

*Note: Standard errors are in parentheses. All variables are in natural logarithms*

SOM is not a regression model. However, to make this statement easier to understand, let us define our case as a linear regression model:

$$y = f(x_1, x_2, \dots, x_k) + \varepsilon$$

where  $\varepsilon$  stands for the random, unknown variables.

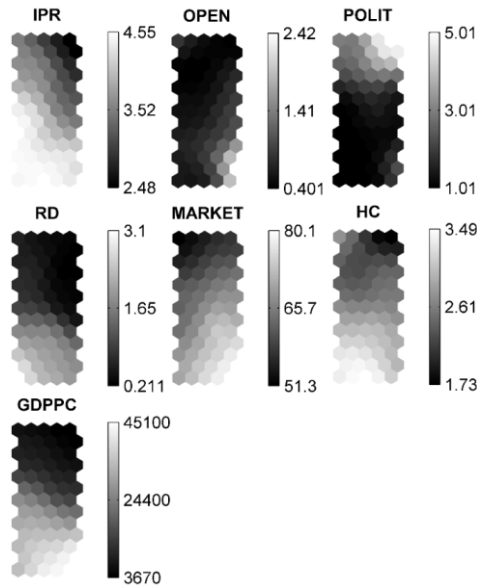
Here, assuming the expected value of  $\varepsilon$  is zero [ $E(\varepsilon) = 0$ ], it could be suggested that  $y$  is determined by the  $x$  variables. Accordingly, if we follow the same assumption, the IPR model can be constructed as follows:

$$IPR = f(GDPpc, RD, HC, OPEN, POLIT, MARKET) + \varepsilon$$

Figure 5 displays the clustering patterns of each variable. Note that the second analysis covers the 1996-2010 period with five-year span data of 49 countries. Thus, the inputs are each year’s values of those variables. In other words, the component matrix for a variable consists of the variable’s 1996-2000, 2001-2005 and 2006-2010 values. Similarities in component matrices could be interpreted as a correlation between the variables. For example, we can see with the IPR index on Figure 5 that the high-protective countries are clustered on the bottom left corner of the component matrix. Almost the same clustering patterns can be observed with human capital. This is the first sign of a significant relationship between IPR and human capital for the full sample. As evident from the component matrices, three different colour transitions exist for the component matrices of these variables: white, grey, and black. As for the variables of R&D, market freedom, GDP per capita and political rights, the same colour transitions can also be

observed along with the same relationship. Note that the political rights index is an inverse index, and its clustering patterns hence appear to be the exact opposite of the other related variables. Table 6 and 7 show the detailed correlation coefficients calculated using different methods.

**Figure 5. The clustering patterns and potential determinants of patent protection regime (full sample)**



Source: Author's own calculations. Analysed and visualized using the SOM toolbox in the MATLAB.

The basic evidence for the determinants of IPR is visualized in Figure 5 for the full sample. However, as defined in the introduction as one of the two main propositions of the study, the determinants of IPR may vary across different country groups. Indeed, the basic sample statistics for high-, mid- and low-protective countries indicate remarkable differences between country samples. In Figure 6, 7 and 8 the same component matrices are created for high-, mid- and low-protective countries, respectively. It is important to mention here that the size of the component matrices depends on the number of inputs. Since the number of countries in the low-protection group is 10, in the mid-protection group 16, and in the high-protection group 23, the size of the component matrices for each country group is different.

**Table 5. Sample statistics for different clusters**

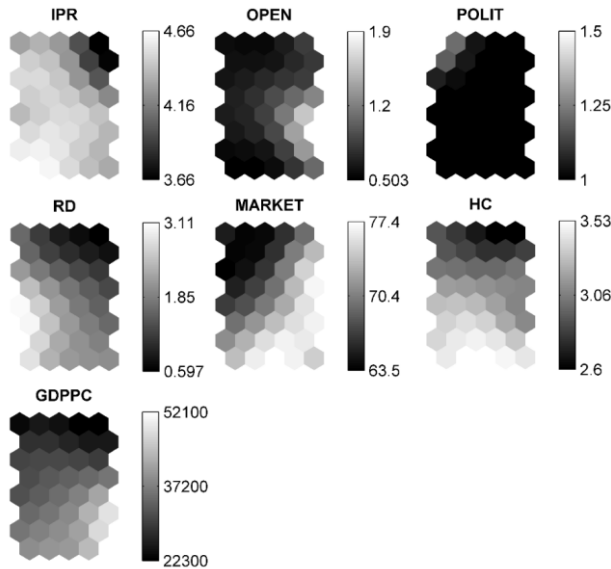
Cluster	Freq(%)	IPR	GDPPC	HC	RD	POLIT	MARKET	OPEN
High	47.69	4.41	37197.52	3.24	2.20	1.19	72.54	0.94
Mid	29.23	3.69	17302.59	2.56	0.51	1.26	66.42	0.72
Low	30.77	3.04	6202.94	2.19	0.38	3.63	57.41	0.51

Note: Authors' own calculations.

In Figure 6, we can see that the component matrices for R&D, GDP per capita and human capital have almost the same clustering patterns as IPR, which implies that in high-protective countries there is a relationship between these factors and patent protection regime. Similar outcome can be also seen in Figure 7 with mid-protective and Figure 8 with low-protective countries.

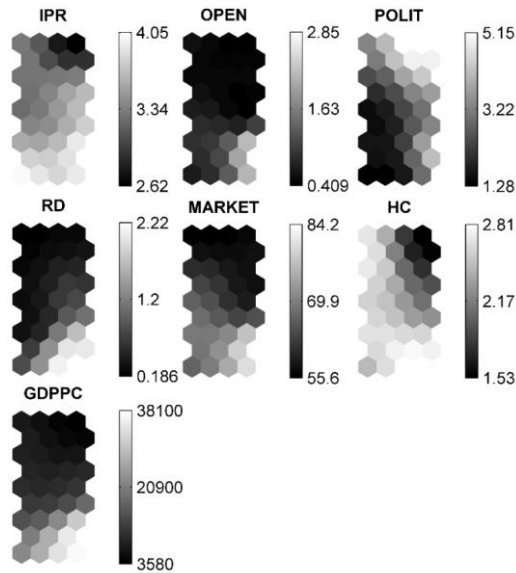
Figure 9 displays label matrices created using the clustering patterns of actual IPR values and predicted IPR values. It is important to mention again here that the study assumes there is not any other explanatory factor of IPR [ $E(\varepsilon) = 0$ ]. Thus, depending on this assumption, we expect that the clustering patterns produced by the potential determinants should be same with the clustering patterns produced by the actual IPR index. Naturally, this is not completely realistic and the match-up ratio may be less than 1. The sizes of the component matrices in Figure 9 are not the same because of the same above-mentioned reason. The patterns are very similar but there are, of course, some differences between the matrices. The success rate of this analysis is 69%, and this rate exceeds two thirds of the full sample. The country list of the second analysis and countries successfully matched up with the actual cluster are given in Table A2.

**Figure 6. The clustering patterns and potential determinants of patent protection regime (high sample)**



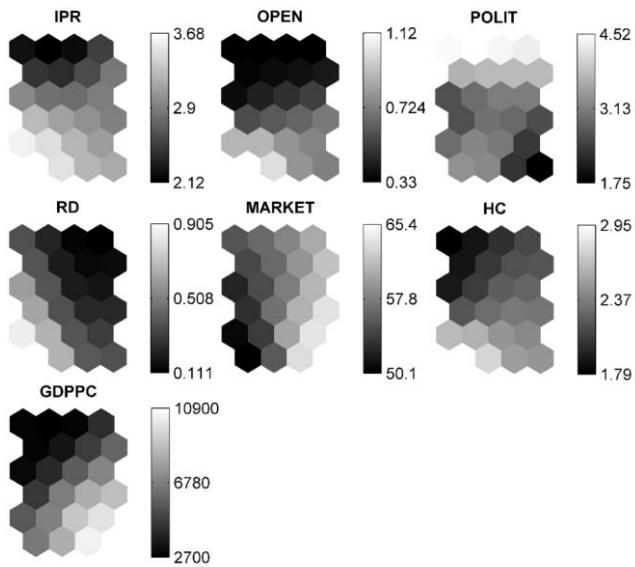
*Source: Authors' own calculations. Analysed and visualized using the SOM toolbox in the MATLAB.*

**Figure 7. The clustering patterns and potential determinants of patent protection regime (mid sample)**



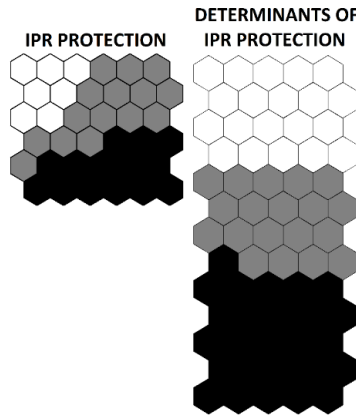
*Source: Authors' own calculations. Analysed and visualized using the SOM toolbox in the MATLAB.*

**Figure 8. The clustering patterns and potential determinants of patent protection regime (low sample)**



*Source: Authors' own calculations. Analysed and visualized using the SOM toolbox in the MATLAB.*

**Figure 9. The clusters for the IPR protection and its determinants**



*Source: Authors' own calculations. Analysed and visualized using the SOM toolbox in the MATLAB.*

The relationship between a patent protection regime and its determinants is given above in the component matrices; but let us now examine the strength of these relationships in detail. For the full sample of 49 countries, the importance of the determinants is sorted using different calculation methods: the structuring index, relative importance index, cluster description index, and Spearman's rank correlation index. In the last column of Table 6, using all of these calculation methods, we have calculated the overall index. Since these measures reveal the main findings of the study, it is advisable to first give some information about their background.

In previous studies, structuring index (SI), relative importance (RI), and cluster description (CD) methods have been used to determine the significance level of each variable in the emergence of different clusters according to SOM outputs<sup>3</sup>. However, these methods can sometimes give conflicting results. Therefore, in this study, the overall index is constructed in order to include each calculation method with equal weights. This way, by reducing the effect of extreme results, more consistent results should be achieved. The SI index was originally developed and used by Park et al. (2005) and Tison et al. (2004, 2005). This study is adapted to determine the factors that affect the formation of different IPR groups. If one of the variables used in the model has a low SI value, it indicates that the effect of that variable on the cluster of the SOM map is also low; whereas variables with a high SI value explain a significant portion of the variance between the cluster groups. The SI value of the variable  $i$  can be calculated as follows:

---

<sup>3</sup> For further information on the calculation of indices, see Ki et al (2016).



$$SI_i = \sum_{j=1}^S \sum_{k=1}^{j-1} \frac{|w_{ij} - w_{ik}|}{\|r_j - r_k\|}$$

where the nominator and denominator respectively show the weight and topological differences between  $j$  and  $k$  map units, while  $S$  represents the total number of map units. Another index used to determine the significance level of variables in the emergence of SOM-based clusters is the relative importance (RI) index. Based on the distance matrix, each variable in the RI index is pictured in a pie chart, proportional to the sum of the variables. In addition, the sum of these effects is standardized to be 100. In other words, the importance of a variables in the model responds to the size it has in the pie chart. Consequently, it is expected that  $i$  variable has a high RI value in order to have a high effect on the SOM structure.

Vesanto (2002) revealed the CD index which, unlike other indices, expresses variance in each cluster. Thanks to the CD index, the internal properties of each cluster can be displayed. It is calculated in the following equation:

$$CC_i = \sum_{l=1}^C S_{li}^D = \sum_{l=1}^C \frac{(C-1)S_{li}^C}{\sum_{m=1, m \neq l}^C S_{mi}^C} \quad \text{where } S_{li}^C = \frac{\sigma_{li}}{\sigma_i}$$

where  $\sigma_{li}$  and  $\sigma_i$ , respectively, indicate the standard deviations of the variable in cluster  $l$  and the whole data set, while  $C$  shows the total number of clusters. A high CD value calculated for a variable means that the importance of that variable is high in more clusters. For all indices the value is high, which implies that the variable is important.

**Table 6. The importance of the determinants of IPR (full sample)**

Rank	SI	Val.	RI	Val.	CD	Val.	SRC	Val.	Overall	Val.
1	MARK	825.46	HC	25.07	OPEN	6.84	GDPPC	0.81***	HC	88.52
2	GDPPC	772.29	MAR	23.25	RD	5.41	RD	0.78***	GDPPC	87.64
3	HC	724.89	GDPPC	19.66	MAR	5.08	HC	0.73***	RD	80.43
4	RD	686.60	RD	13.86	GDPPC	4.84	POL	-0.67***	MAR	80.31
5	POL	654.27	POL	10.48	HC	4.81	MAR	0.51***	POL	68.33
6	OPEN	597.83	OPEN	7.65	POL	4.39	OPEN	0.19**	OPEN	51.72

*Note: Authors' own calculations.*

According to the results in Table 6 showing the importance of the determinants for the full sample, the primary determinants of IPR are human capital, GDP per capita, R&D investments, and market freedom, as has been previously shown in Figure 5. The importance of the determinants is also estimated for each cluster. Since the SI, RI and CD calculations require the overall sample including the sub-samples (clusters), the importance for each cluster is calculated via the Spearman's rank correlation test. The results given in Table 7 reveal that the top four determinants of patent protection regime in high-protective countries are GDP per capita, R&D investments, human capital and political rights. The ranking of the determinants is different for mid- and low-protective countries. In mid-protective countries the primary determinants are GDP per capita, R&D investments, market freedom, and openness, while in low-protective countries it is R&D investments, human capital, GDP per capita and openness.

**Table 7. Spearman’s rank correlations for cluster groups**

Rank	SRC (High)	Values	SRC (Mid)	Values	SRC (Low)	Values
1	GDPPC	0.368***	GDPPC	0.792***	RD	0.601***
2	RD	0.359***	RD	0.506***	HC	0.573***
3	HC	0.233**	MARKET	0.374**	GDPPC	0.492***
4	POLIT	-0.203*	OPEN	0.331**	OPEN	0.377*
5	OPEN	-0.154	HC	0.316**	POLIT	-0.342*
6	MARKET	0.017	POLIT	-0.257*	MARKET	-0.116

*Note: Authors’ own calculations.*

It is not surprising that GDP per capita is the main determinant of a patent protection regime for each cluster. As countries develop and improve their structural reforms, they try to ameliorate their law system as well. Also, there is a significant correlation between human capital and the patent protection regime for each cluster. Rising awareness of the society results in the need for a stronger protection of rights. Inventors, innovators, thinkers and artists all need to be protected by stronger laws. The correlation between the patent protection regime and political rights is estimated as significant for each cluster, as expected. Since strengthening of the patent protection regime (or the strength of the IPR policy) is nothing short of a strengthening of the law system, there likely is a relationship between these two law subjects. We have also found that the share of R&D investment is a significant determinant of IPR for each cluster, which means that increasing research and technology investments require higher protection standards. This is a major determinant, especially in low-protective countries. The results imply that in low-protective countries policymakers respond to these demands by increasing the effort in R&D sector. It is not so for high- and mid-protective countries; the share of RD investment still is, however, the second most significant determinant of patent protection regime there. It can be thus concluded that R&D is the main driver of a stronger patent protection regime for each cluster.

**Conclusion**

In this study, we have investigated the determinants of IPR. IPR include laws such as patents, copyrights, trademarks, etc. that aim to protect the benefits of creators of new ideas. Despite the fact that the definition of IPR, compared to patents, is broader in scope, it is expected that the main economic impact of IPR – or the most probable interaction of IPR – might be seen with patents. For that reason, patents are subjected to a close examination and various institutions and researchers have established several statistical indicators. There have been many studies investigating the impacts of IPR or patents on economic growth and innovation performances of countries. However, there is only very little evidence for the determinants of patent protection. In our study, the paper of Ginarte and Park (1997), a recognized authority on the subject, has been revisited via a more up-to-date statistical method: SOM. Despite the authors’ efforts to compose a comprehensive IPR index and find the main determinants of IPR, there are some issues with the estimation method Ginarte and Park (1997) used in their paper. Since the OLS and the GLS methods are not resistant to collinearity and endogeneity, the analysis needs to be updated by a more robust method. For that purpose, we have decided to use a SOM as a non-parametric technique.

The study consists of two different analyses. In the first analysis, using the advantages of the SOM, we classify the countries based on the strength of their patent protection. 111 countries from the 1960-2010 period have been used for this and it turns out that there are three different patent protection regimes around the world. Furthermore, in the post-1980 period, which has witnessed the constitution of the globalization process, there is a dramatic increase in the IPR protection for all countries. This could be considered as an outcome of global policies proposed by the WTO and the WIPO. Establishing a globally harmonized patent protection regime and a standardization process for all countries are the main goals for these institutions. The purpose of this study is not the examination of the impacts of IPR, but many other studies have empirically revealed that countries from different development stages are affected by the same policies differently. Especially in the case of some developing countries, there is evidence suggesting that tightening of the protection policy reduces innovation and growth performance of the country.

The second analysis in this study investigates the determinants of patent protection regime. Since the study simulates Ginarte and Park's (1997) paper in a sense, we have used the same variables as potential determinants and revisited the econometric OLS and GLS models from the original paper. Due to data constraints, the second analysis only contains 49 countries for the 1996-2010 period. However, due to the issues with the estimation methods, we have obtained very little significant evidence. Then, we subject the potential determinants and IPR variables to a SOM analysis. Considering the regime clusters, we have made the same analysis for the high-, mid- and low-protective countries as well as the full sample. The results show that in the case of the full sample, human capital, income level, the share of the R&D sector, market freedom, and political rights can be considered to be significant determinants of patent protection regime. As for the sub-samples, the rankings of these determinants start to differ for the high-, mid- and low-protective countries. Although in high-protective countries the correlation coefficients are relatively low, there is a significant relationship between income level, the share of R&D sector, human capital, political rights, and patent protection regime. In mid-protective countries, income level, the share of R&D sector, market freedom, openness rate, human capital and political rights are the most significant determinants of patent protection regime, respectively. Whereas in low-protective countries the most significant determinants are the share of R&D sector, human capital, income level, openness rate and political rights.

In conclusion, based on all of these findings, two main outcomes can be emphasized. Firstly, there are three different patent protection regimes around the world which the methods used in the study assessed as significant. Policy makers and researchers should certainly consider these distinctions between countries in the future. Secondly, patent protection regimes of countries majorly depend on their income level, R&D investment, and human capital. However, the importance of the determinants varies across countries, depending on their stage of development.

**Disclosure statement:** No potential conflict of interest was reported by the authors.

## References

- FEENSTRA, Robert C.; INKLAAR, Robert; TIMMER, Marcel P. The next generation of the Penn World Table. *American Economic Review*, 2015, 105.10: 3150-82.
- GINARTE, Juan C.; PARK, Walter G. Determinants of patent rights: A cross-national study. *Research policy*, 1997, 26.3: 283-301., DOI: [10.1016/S0048-7333\(97\)00022-X](https://doi.org/10.1016/S0048-7333(97)00022-X)
- HALL, Bronwyn H. Patents and patent policy. *Oxford Review of Economic Policy*, 2007, 23.4: 568-587., DOI: [doi.org/10.1093/oxrep/grm037](https://doi.org/10.1093/oxrep/grm037)
- KI, Seo Jin; LEE, Seung Won; KIM, Joon Ha. Developing alternative regression models for describing water quality using a self-organizing map. *Desalination and Water Treatment*, 2016, 57.43: 20146-20158., DOI: [10.1080/19443994.2015.1112981](https://doi.org/10.1080/19443994.2015.1112981)
- KOHONEN, Teuvo. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 1982, 43.1: 59-69., DOI: [10.1007/BF00337288](https://doi.org/10.1007/BF00337288)
- MOSTAFA, Mohamed M. Clustering the ecological footprint of nations using Kohonen's self-organizing maps. *Expert Systems with Applications*, 2010, 37.4: 2747-2755., DOI: [10.1016/j.eswa.2009.09.016](https://doi.org/10.1016/j.eswa.2009.09.016)
- PARK, Y. S., GEVREY, M., LEK, S., GIRAUDEL, J.L. (2005). Evaluation of relevant species in communities: development of structuring indices for the classification of communities using a self-organizing map. In: *Modelling Community Structure in Freshwater Ecosystems*. Springer, Berlin, Heidelberg, 2005. p. 369-380, DOI: [10.1007/3-540-26894-4\\_31](https://doi.org/10.1007/3-540-26894-4_31)
- PARK, Walter G. International patent protection: 1960–2005. *Research policy*, 2008, 37.4: 761-766., DOI: [10.1016/j.respol.2008.01.006](https://doi.org/10.1016/j.respol.2008.01.006)
- RAPP, Richard T.; ROZEK, Richard P. Benefits and costs of intellectual property protection in developing countries. *J. World Trade*, 1990, 24: 75-102.
- SARLIN, Peter. Sovereign debt monitor: A visual self-organizing maps approach. In: *Computational Intelligence for Financial Engineering and Economics (CIFER)*, 2011 IEEE Symposium on. IEEE, 2011. p. 1-8.. IEEE, DOI: [10.1109/CIFER.2011.5953556](https://doi.org/10.1109/CIFER.2011.5953556)
- SARLIN, Peter; PELTONEN, Tuomas A. Mapping the state of financial stability. *Journal of International Financial Markets, Institutions and Money*, 2013, 26: 46-76., DOI: [10.1016/j.intfin.2013.05.002](https://doi.org/10.1016/j.intfin.2013.05.002)
- SEYOUM, Belay. The impact of intellectual property rights on foreign direct investment. *The Columbia Journal of World Business*, 1996, 31.1: 50-59., DOI: [10.1016/S0022-5428\(96\)90006-X](https://doi.org/10.1016/S0022-5428(96)90006-X)
- SHERWOOD, Robert M. Intellectual property systems and investment stimulation: The rating of systems in eighteen developing countries. *Idea*, 1996, 37: 261.
- TISON, J., et al. Use of unsupervised neural networks for ecoregional zoning of hydrosystems through diatom communities: case study of Adour-Garonne watershed (France). *Archiv für Hydrobiologie*, 2004, 159.3: 409-422., DOI: [10.1127/0003-9136/2004/0159-0409](https://doi.org/10.1127/0003-9136/2004/0159-0409)

TISON, J., et al. Typology of diatom communities and the influence of hydro-ecoregions: a study on the French hydrosystem scale. *Water Research*, 2005, 39.14: 3177-3188., DOI: [10.1016/j.watres.2005.05.029](https://doi.org/10.1016/j.watres.2005.05.029)

VESANTO, Juha; ALHONIEMI, Esa. Clustering of the self-organizing map. *IEEE Transactions on neural networks*, 2000, 11.3: 586-600., DOI: [10.1109/72.846731](https://doi.org/10.1109/72.846731)

VESANTO, Juha, et al. Data exploration process based on the self-organizing map. Helsinki University of Technology, 2002.

WIPO. (2011). "What is Intellectual Property?", WIPO Publication No: 450 E.

**Appendix****Table A1. Country list for the classification analysis and the regime clusters**

Low	Mid		High
Angola	Argentina	Niger	Australia
Bangladesh	Burundi	Nigeria	Austria
Colombia	Benin	Nepal	Belgium
Costa Rica	Burkina Faso	Panama	Canada
Ethiopia	Bolivia	Philippines	Switzerland
Guatemala	Brazil	Portugal	Chile
Guyana	Botswana	Rwanda	Cyprus
Indonesia	Cent. Afr.	Senegal	Germany
India	Ivory Coast	Singapore	Denmark
Jordan	Cameroon	Sierra Leone	Spain
Madagascar	Congo	El Salvador	Finland
Mexico	Dominican Republic	Syria	France
Mali	Algeria	Chad	United Kingdom
Burma	Ecuador	Togo	Greece
Mozambique	Egypt	Trinidad & Tobago	Hong Kong
Malawi	Fiji	Tunisia	Haiti
Nicaragua	Gabon	Turkey	Ireland
Pakistan	Ghana	Tanzania	Israel
Peru	Grenada	Uganda	Italy
Papua New Guinea	Honduras	Uruguay	Jamaica
Paraguay	Iran	Zimbabwe	Japan
Saudi Arabia	Iraq		Korea
Sudan	Iceland		Luxembourg.
Somalia	Kenya		Netherlands
Swaziland	Liberia		Norway
Thailand	Sri Lanka		New Zealand
Venezuela	Morocco		Sweden
Vietnam	Malta		USA
Democratic Congo	Mauritania		South Africa
Zambia	Mauritius		
	Malaysia		

*Note: The clusters depend on the findings from the analysis in the section 3.1.*

**Table A2. Country list for the determination analysis and the regime clusters**

Low	Mid	High
Burkina Faso*	Argentina*	Australia*
Bolivia*	Brazil	Austria*
Costa Rica	Colombia	Belgium*
Egypt*	Cyprus*	Canada*
Indonesia*	Ecuador	Denmark*
Pakistan*	Iceland	Finland*
Sri Lanka*	India	France*
Thailand*	Israel	Germany*
Uruguay*	Mexico*	Greece
Zambia*	Morocco	Ireland*
	New Zealand	Italy
	Panama*	Japan*
	South Africa*	Korea*
	Trinidad and Tobago*	Luxembourg*
	Turkey	Netherlands*
	Ukraine	Norway*
		Portugal
		Singapore*
		Spain
		Sweden*
		Switzerland*
		United Kingdom*
		USA*

*Note: The findings depend on the analysis in section 3.2. The asterisk sign (\*) indicates countries where the predicted cluster and the actual regime cluster overlap.*