

Use of prediction methods for patent and trademark applications in Spain

Antonio Hidalgo Samuel Gabaly

A B S T R A C T

Patent and trademark offices which run according to principles of new management have an inherent need for dependable forecasting data in planning capacity and service levels. The ability of the Spanish Office of Patents and Trademarks to carry out efficient planning of its resource needs requires the use of methods which allow it to predict the changes in the number of patent and trademark applications at different time horizons. The approach for the prediction of time series of Spanish patents and trademarks applications (1979–2009) was based on the use of different techniques of time series prediction in a short-term horizon. The methods used can be grouped into two specific areas: regression models of trends and time series models. The results of this study show that it is possible to model the series of patents and trademarks applications with different models, especially ARIMA, with satisfactory model adjustment and relatively low error.

1. Introduction

Since economic analysis has taken an interest in studying the problems associated with technological change (in the decade of the 1980's), adequate indicators have been identified which make it possible to provide a coherent explanation for technological activities and their relationship with economic efficiency. Although the earliest studies focused on analyzing the activities of research and development (R&D), at present the focus of analysis has shifted to another type of variable, more oriented toward the gathering of knowledge and capabilities, in which patents provide relevant information [1,2,3]. For this reason, current developments in the demand for patent applications constitute an element of reference, because they make it possible to identify models associated with technological change.

Without a doubt, because patents constitute an excellent indicator of technological change, studying the predictions of patent and trademark applications has recently become a field of interest, in particular on the part of the main patent offices, such as the United States Office of Patents and Trademarks (USPTO), the Japanese Patent Office (JPO), and the European Patent Office (EPO). This interest also results from the need which these organizations have to effectively design their business strategies and plan their resources, with the support of sophisticated models which allow

them to provide information in advance on developments in technological change. The USPTO uses three different types of models for annual prediction of patent applications: the Naive model, the ARIMA model and the econometric model [4]. As for the EPO, it uses two specific prediction models: one linear model, known as trend analysis, based on time series methods, and another model of transference, based on regression methods [5]. Different combinations of these models are also used in a more specific way by various national patent offices, as is the case of the Federal Intellectual Property Institute of Switzerland, for analyzing changes in trademark applications [6].

In the case of the Spanish Office of Patents and Trademarks, its ability to carry out efficient planning of its resource needs requires the use of technological forecasting methods which allow it to predict future changes in the number of patent and trademark applications at different time thresholds, including the short, medium and long-term. With this objective in mind, the Spanish Office of Patents and Trademarks has promoted a research project so as to develop a methodology which predicts the changes in the number of national patent and trademark applications for a time horizon of three years.

This paper summarizes the main results found in this research project, in which the aggregate annual data on national patent and trademark applications between the years of 1979 and 2009 were used. The predictive focus used for the historical series of patents and trademarks consisted of the application of different time series prediction techniques which were compared to obtain a series of predictions for the target variables in a short-term time horizon. In

this research report, we considered a total of four time series to be modeled in the analysis of their change over time and future prospects:

- National patent applications.
- Trademark applications for products.
- Trademark applications for services.
- Total trademark applications (products, services and unclassified).

This paper is structured into four sections. First of all, a review of the literature is performed in terms of the relationship between economic activity and patent production, which serves as a support for the creation of two prediction models. Secondly, the data used in the research are described, and then the empirical models used and results obtained for each of the time series studied are described. The final section presents the most significant conclusions with a prediction of the changes which will occur in the time series of patent and trademark applications.

2. Patents and economic activity

Ever since the literature on economic development began paying attention to the cases of countries which were industrialized after the first industrial revolution, one aspect analyzed has been that involving the role of technology as a factor which promotes or delays the process of catching up with technology leaders [7,8,9]. Within this context, patents have become an indicator of reference, because they are intrinsically related with the processes of innovation and with scientific and technological change [10].

The first studies carried out by Schmookler made it clear that technical inventions have an economic cause and that a country's level of inventive activity is influenced by a set of variables which include the number of workers with technical qualification, the industrial inputs and Gross Domestic Product [11]. Years later, this same author studied the percentage of patents used commercially and identified the different dimensions displayed by innovation when it is actually developed: earnings or losses in economic terms resulting from the innovation, the relationship between output and units of input, the effect which patents have on the income from products which compete with each other, as well as with those which are complementary to the innovation, and the change in Gross Domestic Product caused by the innovations and imitations [12,13].

Continuing along this line of studies, Pavitt [14] found evidence which suggests that an increase in the number of patents in a country is directly related with an increase in spending on R&D. This same hypothesis was verified by Scherer [15], who studied the relationship between spending on R&D and patent production in a sample containing 443 companies and 15,112 American patents of residents in the United States in the period of 1976–1977. Bosworth and Westaway [16] found the existence of a delay of approximately one year between the implementation of R&D expenditure and the patent application produced as a result.

On the basis of these studies, the use of the information contained in the patents has undergone more than significant change, such that, at present, it is almost impossible to find studies on innovation processes or the technological capabilities of any organization (company, university, research center) without using patents as associated indicators. One focus of interest aims toward acknowledging that efforts in R&D have a positive impact on a company's stock of knowledge [17]. Assuming the hypothesis that knowledge gathered depreciates over time, causing the contribution by R&D to decrease in value, these authors used patent applications as one possible indicator of the added value of the knowledge produced over a period of time.

Along these same lines, later studies compared the impact of the value of protecting patents with spending on R&D, and they suggest that patents have a limited but important impact on these expenditures. The method consists of calculating the *rates of subsidies equivalent* to the patents, or in other words, dividing their estimated value by the company's spending on R&D to produce those patents. The rates calculated are equal to the subsidies which the companies would need to maintain R&D at its current level, if there were no patents. Using this methodology, Pakes [18] calculated the rates of equivalent subsidies in the R&D financed by companies in three European countries during the decade of the 1970's, having calculated estimates of 6.8% in France, 5.6% in Germany and 5.7% in the United Kingdom. Schankerman [19], using data on all of the patent applications and renewals in France during the time periods of 1969–1982 and 1969–1987, detected a rate of equivalent subsidies equal to 15.6%. In comparison with other instruments of technological policy such as tax reliefs, these figures seem credible [20].

In a more recent study, based on a 1994 survey amongst R&D directors at American companies, Arora, Ceccagnoli and Cohen [21] used another methodology known as the *patent premium* to calculate the value of protecting patents and its effect on R&D spending. These authors define the patent premium as the difference between the value of innovation before and after having been patented, and therefore its value may be positive or negative. The study reflects the fact that this premium, before the innovation is patented, has a negative value on average, or in other words the foreseen value of an innovation would be reduced by 10%–50% if the innovation were patented, which makes it clear that many innovations are not patented because their inventors believe that protecting the patent would be ineffective, whereas publishing the patent would increase the risk that the innovation might be imitated. On the contrary, when the study focuses on patented innovations, a positive patent premium of between 75% and 125% is detected which increases as a strong recovery of investment in R&D of innovative firms.

Other analyses focus on the companies' interest in using patents as a signal to attract financial investors [22], and therefore patents become a decisive factor in the value of companies. In fact, capital markets use companies' patents as indicators of their inventive activities, and normally they assume that patents bear a correlation with a company's ability to innovative. However, they also associate it with characteristics that are more difficult to observe, such as intellectual capital or the level and productivity of R&D spending [23]. In an empirical study of 4800 American manufacturing companies during the period of 1957–1995, Hall [24] detected that “a greater return on a patent for every million spent on R&D is associated with a 2% increase in the company's market value”.

Last of all, though there is little empirical evidence regarding this topic in the literature, patents also seems to constitute a key factor allowing newly created companies to attract venture capital [22,23,25]. Venture capital investors believe that patents with a high value are one of the most important factors when reaching investment decisions, especially in the early stages. Within this context, newly created companies have real incentives to file for patents, which could explain why, in two empirical studies carried out in the American semi-conductor industry [26] and in the American manufacturing sector [27], newly created companies displayed a greater propensity toward patenting than did other companies.

3. Evolution of patent and trademark applications in Spain

The analysis of historical time series of national patent and trademark applications in Spain focuses on the period of 1979–2009, and the developments therein are partially

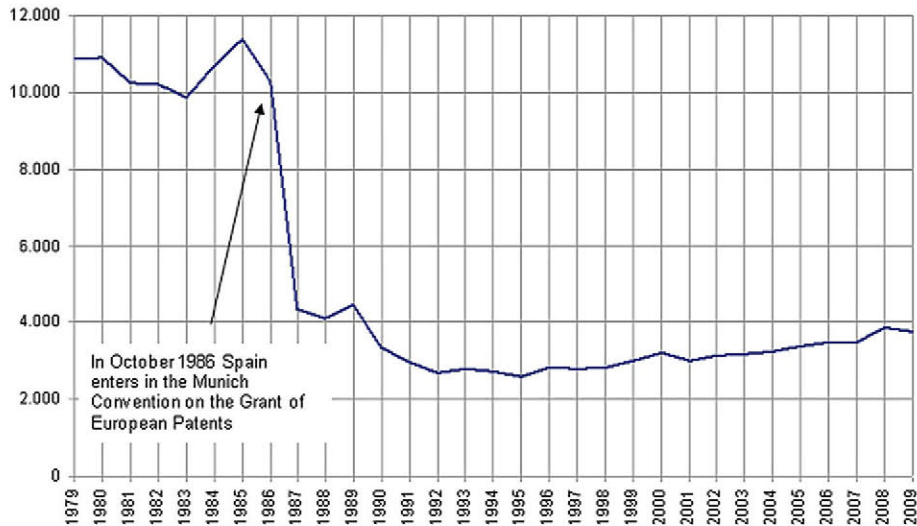


Fig. 1. Historical evolution of the number of national patent applications.

conditioned by the different milestones and regulatory changes that took place over time. Fig. 1 shows the changes in the number of national patent applications at the Spanish Office of Patents and Trademarks. The main characteristics which are observed in the series can be summarized as follows:

- A drastic drop in the number of national patent applications in the year of 1986 as a result of the change in Spanish legislation, because the Munich Convention on European Patents took effect in Spain. This change brought with it the shift from an old patent system to a more modern, stricter system, which would then filter innovations.
- This trend toward a decrease continued until the year of 1989, when there was a specific increase in patent requests which might be related with Spain's entry into the Patent Cooperation Treaty (PCT).
- As of the year 1995, a rising trend in the number of patent applications could be seen, reaching 1988 levels in the year of 2008 (approximately 4000 applications), thereby confirming a trend toward a recovery in the number of applications,

though far from the nearly 10,000 patent applications reached in 1986.

The changes in the historical series of national trademark applications is shown in Fig. 2, in which trademark applications for products and services are distinguished, in accordance with the 1957 Classification of Nice. It must be mentioned that the series of total trademark applications includes those for products and those for services, plus a small number of unclassified trademark applications.

Examining the series of national trademark applications makes it possible to see three specific peaks in the years of 1989, 2000 and 2006, the last of these as a result of the fact that the electronic trademark application took effect in 2005, making the procedure easier and less expensive. As of the year of 2007, the trend has been toward a decrease in all of the classes analyzed, which is probably due to the negative effects of the economic crisis in Spain. It is also important to emphasize how the number of trademark applications for services surpassed the number of trademark applications for products for the first time in the year 2000.

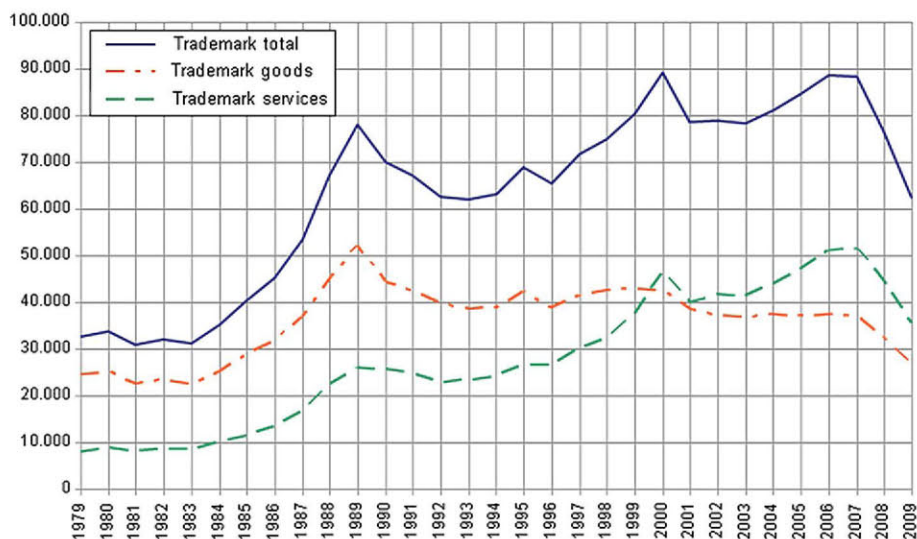


Fig. 2. Historical evolution of the number of national trademark applications.

Table 1
Significant milestones in patents and trademarks in Spain.

Patents	
1986	New Law on Patents (Law 11/1986)
1986	Entry into the Munich Convention (European Patent)
1989	Entry into the Patent Cooperation Treaty (PCT)
1990	Implementation of the report on the State of the Art in the agro-food sector
1991	Implementation of the report on the State of the Art in all sectors
1992	End of the transitional period for the patentability of product patents
1992	Regulation (EEC) 1768/92, supplementary protection certificate for medicinal products
1996	Instruction 2/1995 of the Spanish Office of Patents and Trademarks regarding TRIPs
1996	Regulation (EEC) 1610/96, supplementary protection certificate for phytosanitary products
2000	Prior voluntary examination in agro-food sector put in place
2001	Prior voluntary examination in all sectors put in place
2002	EU Biotechnology Directive
2008	Written opinion in patents put in place
2009	Electronic patent application
Trademarks	
1993	Regulation (EEC) 40/94, on the EU trademark
2001	Law on Trademarks (Law 17/2001)
2003	Law of Industrial Design (Law 20/2003)
2005	Electronic trademark application
2007	Protocol concerning the Madrid Arrangement regarding the International Registration of Trademarks
2009	Reduction of fees on EU trademarks

Although this analysis has focused on the most significant milestones in the time period considered, it is possible that in the respective developments of the series of patents and trademarks applications that other milestones may have had an influence that are more difficult to explain. Because of this, Table 1 is provided to show all of the milestones of a regulatory nature that occurred during the time period analyzed.

4. Prediction models and results

There are three different types of prediction models. According to Meade [28], selecting the most appropriate prediction method is based on two specific statements:

- The characteristics of the time series of data are an important factor when determining the relative performance of the various prediction methods.
- The most complex and sophisticated statistical methods do not necessarily lead to the most accurate results.

In accordance with this last claim, the predictive focus used for the historical series of patent and trademark applications in Spain has consisted of the use of different techniques for predicting time series which have been compared to obtain a series of predictions for the target variables in a short-term timeframe. The methods used can be grouped into two specific sets: regression models of trends and advanced models of time series.

4.1. Regression models of trends

The first approach, based on series of original data on patent and trademark applications, has been to calculate estimates of regression models built on the basis of the extrapolation of trends observed in the annual series of data [29]. In this stage, linear and quadratic regression models were created with different time periods.

The representation of a model with these characteristics was performed in response to the following explanation: if y_t is used to represent the applications filed in a specific year t (whether patents or trademarks, as appropriate), and e_t represents the unexplained portion of the model, or term of error, which is assumed to be distributed independently at each point in time with a mean value of 0, a linear trend model would have the form $y_t = a + bt + e_t$, and a quadratic trend model would follow the equation $y_t = a + bt + ct^2 + e_t$ [30].

In accordance with the guidelines of the European Patent Office, in this research we considered three time horizons for the linear models: long-term (from 1979 through 2009), medium-term (from 1994 through 2009) and short-term (from 2002 through 2009). The quadratic model was created on the basis of a series of medium-term data, or in other words from 1994 through 2009. The criteria followed to select the most appropriate model, was based on each model's goodness of fit with the data.

Fig. 3 shows the projections of time series created on the basis of the linear and quadratic models for national patent applications. It can be seen that the behavior of the short and medium-term linear models, as well as the quadratic model, is very similar, showing a slightly rising trend. These three models have a good fit with the data in the series, with values that fluctuate between 0.89 and 0.94. As for the long-term linear model, it shows a trend very distant from the forecasts of the other models, which is due to the fact that, because it spans a longer period of data, it includes a set of sudden changes in trend.

The analysis of the time series projections created on the basis of the linear and quadratic models for national trademark applications for products is represented in Fig. 4. It shows that, although the long-term linear trend is toward an increase, albeit with a fit which is not

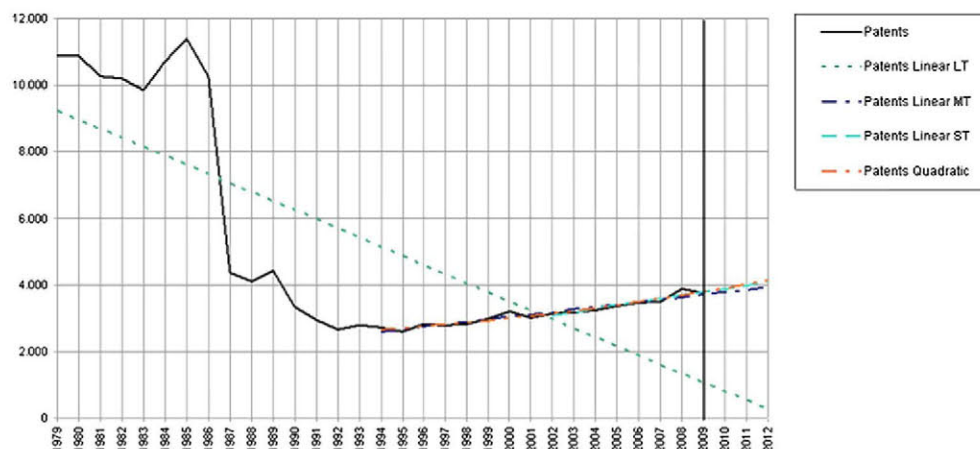


Fig. 3. Regression models of trends applied to national patent applications.

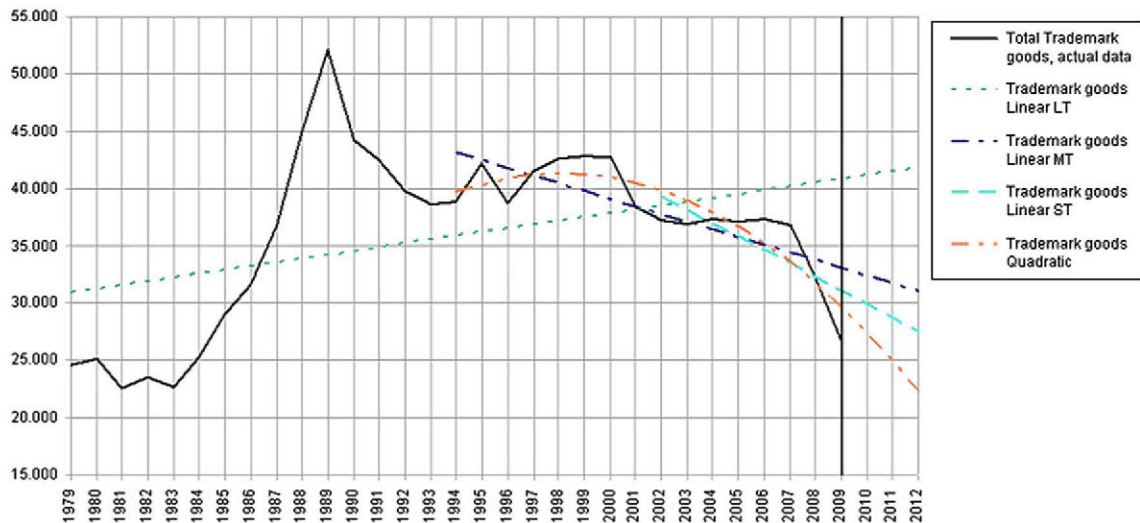


Fig. 4. Regression models of trends applied to national trademark applications for products.

very significant ($R^2 = 0.15$), in the medium and short-term the linear and quadratic trends are toward a decrease, indicating a clear downturn in trademark applications for products as of the year 2000, which is accentuated as of the year 2008, probably due to the effects of the economic crisis in Spain. Of the models evaluated in Fig. 4, the quadratic model (medium-term) is the one which best seems to fit the data from the series of trademarks for products, with an R^2 of 0.79.

The linear and quadratic models for the time series of national trademark applications for services are shown in Fig. 5. Therein one can verify that, in both the long and medium-term, the linear and quadratic trends have positive slopes, though the quadratic models changes in trend as of 2006, becoming a decreasing trend, as occurred in the time series of trademark requests for products. On the contrary, the linear model in the short-term, which spans the period from 2002 to 2009, shows a practically flat trend with a very low fit ($R^2 = 0.01$).

Last of all, the analysis of these regression models applied to the time series of total national trademarks (products, services and unclassified) shows a rising trend in the behavior of the linear models in the medium and long-term, though the last of these is the one which shows the best fit ($R^2 = 0.74$). Both the short-term

linear model and the quadratic model trend toward a decrease and have less significant fits, in particular the linear model in the short-term ($R^2 = 0.12$) (Fig. 6).

Summarized in Table 2 are the estimated regression parameters of the trend models analyzed: linear in the long, medium and short-term, and quadratic (in the medium-term). As can be seen, the models which have the best fit are:

- For patent requests, the quadratic ($R^2 = 0.94$), medium-term linear ($R^2 = 0.93$) and short-term linear ($R^2 = 0.90$) models.
- For trademark applications for products, the quadratic model ($R^2 = 0.79$).
- For trademark applications for services, the long-term linear model ($R^2 = 0.91$) and the quadratic model ($R^2 = 0.81$).
- For total trademark applications, the long-term linear model ($R^2 = 0.74$).

4.2. Advanced time series models

With the goal of making more accurate predictions of values in the time series studied, a set of more advanced models of time

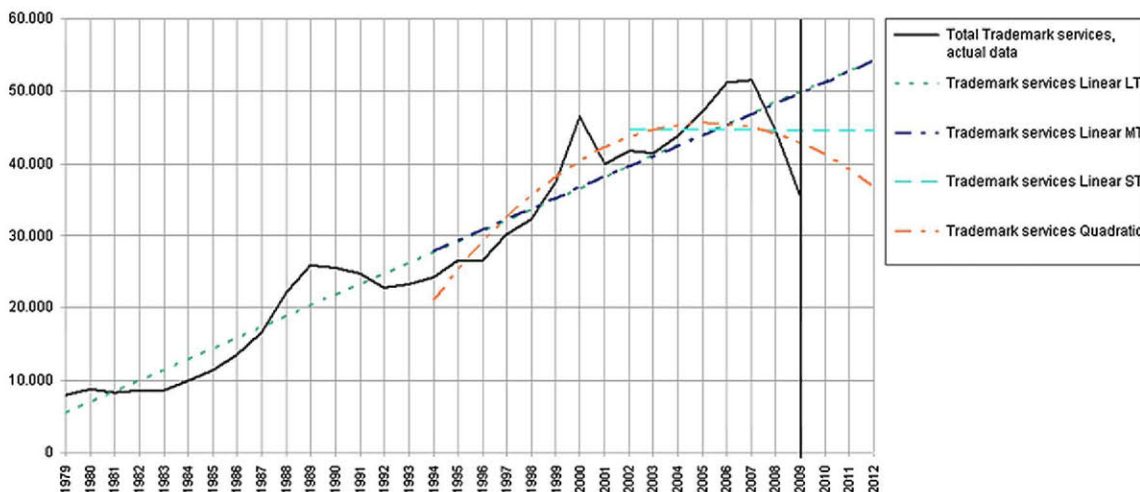


Fig. 5. Regression models of trends applied to for national trademark applications for services.

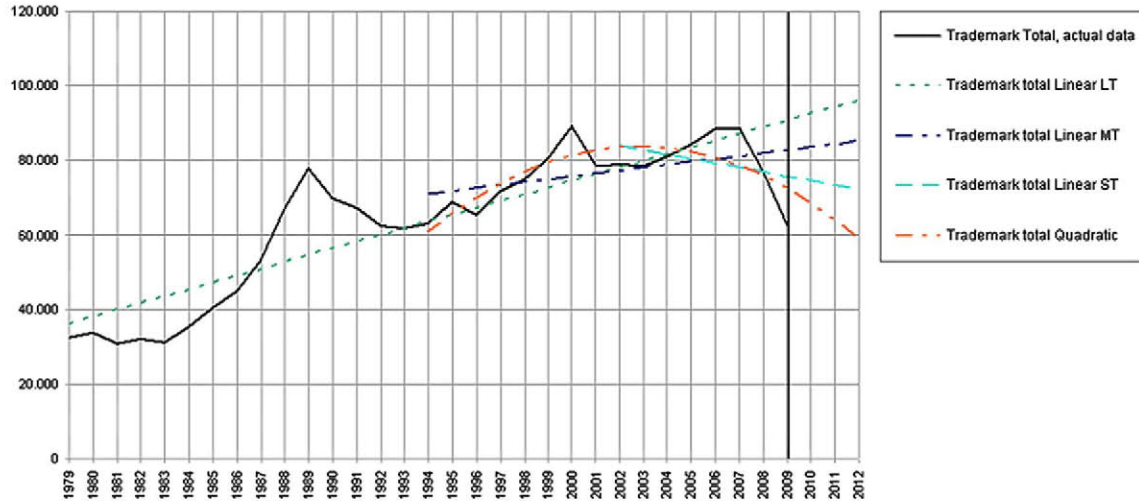


Fig. 6. Regression models of trends applied to national trademark applications (total).

series with a better fit have been used [31]. The time series models used are as follows:

- Exponential smoothing model (Holt type).
- Auto-regressive model of order 1 (AR1).
- ARIMA model with estimate of coefficients.

4.2.1. Exponential smoothing model

In this stage of the analysis, an exponential smoothing model of the Holt trend type was used [32] to model and make three-year predictions of the time series of applications for patents, product trademarks, services trademarks and total trademarks. The use of this model is based on the fact that it is appropriate for series in which there is no seasonality and because it allows for the detection of possible linear trends in the data. The Holt type model used is more general than other models of exponential smoothing, such as Brown's model or the simple models, though the time needed to calculate its algorithms may be somewhat greater for time series, especially for long series. At the same time, it is very similar to an ARIMA model with zero orders of auto-regression, two orders of differentiation and two orders of mobile average, or in other words, an ARIMA model (0,2,2).

In Holt's exponential smoothing model used, the smoothed values of immediately preceding periods were used and were

adjusted by the value F_{t+1} for a percentage of the level in the present period X_t . In this model, the prediction of the period $t+1$ is equal to the prediction for the preceding periods plus α times the difference between the current value and its prediction. Therefore, it can be expressed by way of the following equation:

$$F_{t+1} = (F_t + T_t) + \alpha(X_t - F_t - T_t) \quad (1)$$

in which F_{t+1} is the smoothed value for the period $t+1$, X_t is in the real value of the series in the period t , T_t is the trend estimate, and α is the constant or parameter of smoothing for the level. As for the trend T_t , it is estimated using the equation $T_t = \beta(F_t - F_{t-1})X_t + (1 - \beta)T_{t-1}$, in which the difference of the last two smoothed values is calculated. β represents the constant or smoothing parameters of the trend.

Table 3 shows the values of the parameters of the models calculated. It is made clear that in nearly all of the time series, the level parameter α holds most of the weight, with a unitary value, thereby ruling out the existence of intrinsic linear trends in the analyzed series. Only in case of the series of trademark applications for products is a certain trend component found, because the value of the trend parameter β is somewhat higher (0,2), though with a relatively low statistical t (1,23).

Table 2
Summary of trend models and estimated regression parameters.

Time series	Model type	Model summary					Parameter estimated		
		R^2	F	gl1	gl2	Sig.	Constant	b	c
Patents	Linear LT	56.9%	38.283	1	29	0.000	9501.9	-271.81	
	Linear MT	92.5%	172.795	1	14	0.000	2531.0	74.42	
	Linear ST	89.5%	50.947	1	6	0.000	2994.6	101.06	
	Quadratic	94.0%	101.398	2	13	0.000	2647.7	35.50	2.29
Trademark of products	Linear LT	15.5%	5.328	1	29	0.028	30584.9	331.07	
	Linear MT	57.7%	19.097	1	14	0.001	43791.7	-671.64	
	Linear ST	55.5%	7.474	1	6	0.034	40459.2	-1172.85	
	Quadratic	78.8%	24.097	2	13	0.000	38743.3	1011.15	-98.99
Trademark of services	Linear LT	90.7%	281.703	1	29	0.000	4064.8	1478.68	
	Linear MT	61.8%	22.629	1	14	0.000	26466.5	1452.60	
	Linear ST	0.0%	0.000	1	6	0.985	44712.7	-17.96	
	Quadratic	80.6%	27.034	2	13	0.000	16486.0	4779.43	-195.70
Trademark (total)	Linear LT	73.9%	82.138	1	29	0.000	34644.0	1810.75	
	Linear MT	18.1%	3.098	1	14	0.100	70255.4	783.27	
	Linear ST	12.0%	0.816	1	6	0.401	85195.5	-1189.55	
	Quadratic	61.1%	10.230	2	13	0.002	55235.6	5789.86	-294.51

Table 3
Parameters of the Holt type exponential smoothing models.

		Estimate	ET	t	Sig.
Patents	Alfa (level)	1.000	.199	5.026	.000
	Beta (Trend)	.001	.068	.010	.992
Trademark of products	Alfa (level)	1.000	.195	5.128	.000
	Beta (Trend)	0.200	.162	1.233	.228
Trademark of services	Alfa (level)	1.000	.234	4.271	.000
	Beta (Trend)	.001	.056	.019	.985
Trademark (total)	Alfa (level)	1.000	.233	4.287	.000
	Beta (Trend)	.001	.159	.004	.997

4.2.2. AR(1) and ARIMA models with automatic estimate of coefficients

Generic ARIMA models combine three different types of processes in one single model: auto-regressive (AR), differentiated (I) and mobile average (MA) processes. The type of formulation used to represent an ARIMA model is as follows: ARIMA (p,d,q), in which p is the order of auto-regression, d is the degree of differentiation, and q is the mobile average order used. In the AR1 model, $p = 1$ and $q = d = 0$.

The following equation expresses an ARIMA model of order (p,d,q):

$$Y_t^{(d)} = C + \underbrace{\phi_1 \cdot Y_{t-1}^{(d)} + \dots + \phi_p \cdot Y_{t-p}^{(d)}}_{\text{Auto-regressive portion}} + \underbrace{\theta_1 \cdot \epsilon_{t-1}^{(d)} + \dots + \theta_q \cdot \epsilon_{t-q}^{(d)}}_{\text{Mobile average portion}} + \epsilon_t^{(d)} \quad (2)$$

The difference existing between a mobile average type process and another which is auto-regressive is slight but very important. Each value in a mobile average type series is a weighted average of the most recent random disturbances, or noise, whereas each value in an auto-regressive series represents a weighted average of prior values, so the effect of a disturbance which occurs in an auto-regressive type process is absorbed with the passage of time. However, in a mobile

average type process, a disturbance has an effect on the whole system for a certain number of periods (order of the series) but then later ceases to affect the series in a sudden manner. Once the parameters which explain an ARIMA model have been established, one can continue building the model for the series under study, in accordance with the methodology of Box and Jenkins [33].

In the process for modeling a time series of the ARIMA type, three steps can be distinguished: identification, estimation and diagnosis. In the identification stage of AR1 and ARIMA models with automatic estimation of coefficients, correlograms have been calculated using the auto-correlation (AC) and partial auto-correlation (PAC) functions that are shown in Fig. 7.

In all of the correlograms evaluated in the four series analyzed, one can see how the auto-correlation function displays a rapid decrease tending toward 0 in few delays, whereas the partial auto-correlation function only displays one significant peak in the first delay. These results indicate that, in all of the analyzed series of patent and trademark applications, one may possibly encounter an ARIMA model (1,0,0) or an ARIMA model (1,1,0). Moreover, these results from the correlograms can also be used to verify that the series are seasonal. In other words, they have the same average and variation throughout the entire process, which has also been verified by way of the *Augmented Dickey-Fuller (ADF)* unit root tests carried out. Nevertheless, to ensure the result, natural logarithms of the series were taken and differentiated in the process for estimation of the final ARIMA model, which made it possible to convert the series into annual growth rates.

Moreover, in the ARIMA model identification state, manual identification was used by statistical experts, along with an algorithm for automatic identification of ARIMA models based on a programmed dynamic analysis of the different values in the comparative statistics of the models and the fit of the functions of the residuals so as to ensure that they are composed of white noise. In all cases, the ARIMA model selected by both the human experts and by the automatic algorithms for selection of ARIMA models was the ARIMA model (1,1,0) of the series in natural logarithms.

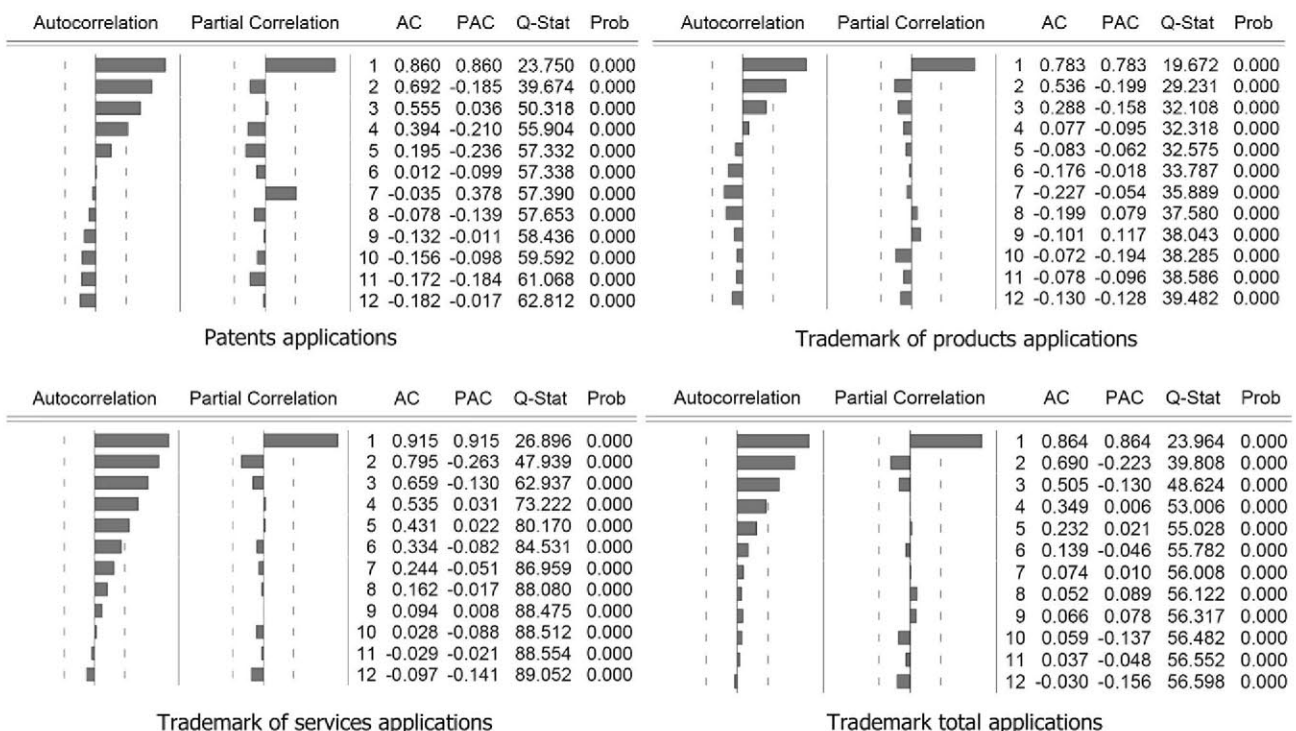


Fig. 7. Correlograms for the series of patent and trademarks applications.

Table 4
Parameters of models AR1 and ARIMA (1,1,0).

Models	Series		Estimate	ET	t	Sig.
AR1	Patents	Constant	5715.89	2354.440	2.428	.022
		AR (Delay 1)	.93	.065	14.320	.000
	Trademark of products	Constant	32389.56	5515.786	5.872	.000
		AR (Delay 1)	.90	.078	11.488	.000
	Trademark of services	Constant	25010.26	12397.029	2.017	.054
		AR (Delay 1)	.96	.046	20.699	.000
	Trademark (total)	Constant	56797.92	14367.790	3.956	.001
		AR (Delay 1)	.94	.057	16.482	.000
ARIMA(1,1,0)	Patents Logn	Constant	-.003	.015	-.209	.837
		AR (Delay 1)	.088	.075	1.162	.257
		Difference	1			
	Trademark of products Logn	Constant	.015	.039	.395	.697
		AR (Delay 1)	.654	.170	3.850	.001
		Difference	1			
	Trademark of services Logn	Constant	.024	.059	.401	.692
		AR (Delay 1)	.737	.174	4.226	.000
		Difference	1			
	Trademark (total) Logn	Constant	.014	.042	.341	.737
		AR (Delay 1)	.617	.223	2.771	.012
		Difference	1			

For the estimation and diagnosis of the AR1 and ARIMA models, modeling techniques with iterative calculation algorithms were used. The values of these models' parameters are shown in Table 4.

4.2.3. Results

For the time series of patent applications shows that, as of the year 1992, the three models show a similar fit, with slight divergences in the time horizon as of the year 2009 (Fig. 8). In this case, the ARIMA series (1,1,0) of the natural logarithm of patent applications displays nearly flat future change; the exponential smoothing model predicts a slightly decreasing change, and the AR1 model predicts a slight increase until 2010, surpassing 4000 patent applications in this year.

Fig. 9 shows that the three time series models predict the changes in the time series of trademark applications for products quite well, though as of the year 2010 the model's predictions tend to differ. According to the AR1 model, nearly linear change would occur in the year 2010, then with a slight increase, whereas the exponential smoothing and ARIMA models indicate a likely decline in the number of trademark applications for products. If we stick with the change in the models calculated, the most probable final trend is that which is located midway between the average of the results from all three models studied.

In the case of the predictions for the series of trademark applications for services (Fig. 10), it is verified that all three models result in different scenarios for the change in number of applications, with the exponential smoothing model being the one which offers the most optimistic forecast, in accordance with which there will be approximately 40,000 applications in the year of 2012. The AR1 model shows a more linear prediction, whereas the ARIMA models find decreasing change which is then attenuated as of the year 2011. Last of all, as regards the modeling of the change in the series of total trademark applications, the three prediction models also display certain divergences in the prediction horizon which follow the patterns seen in the series of trademark applications for services (Fig. 11).

Tables 5, 6 and 7 show the statistics for the verification of diagnosis and the main parameters for estimating the exponential smoothing models, AR1 and ARIMA. As can be seen in the tables, the ARIMA model (1,1,0) has a high degree of fit (R^2) with greater values than the remaining models and all with very high and significant values, in addition to displaying very low average quadratic errors, though the exponential smoothing and AR1 models also possess adjustment parameters and errors which could be considered quite good. As for the Ljung-Box statistics from the three types of models calculated, they indicate to us that the

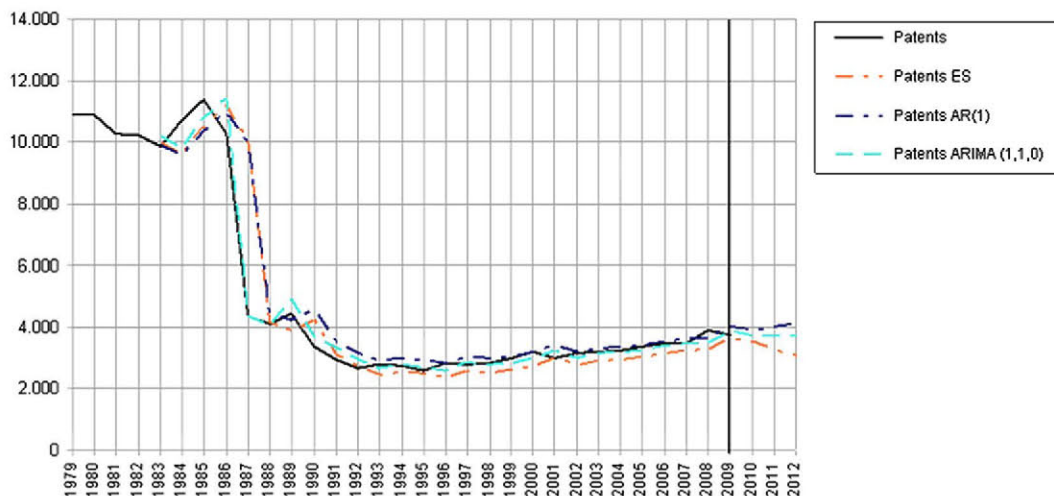


Fig. 8. Models of time series applied to national patent applications.

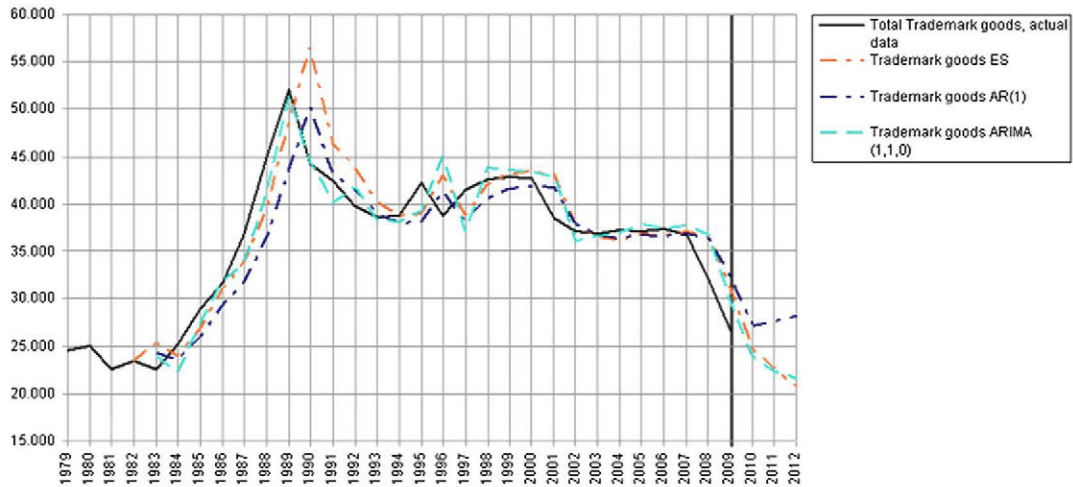


Fig. 9. Models of time series applied to national trademark applications for products.

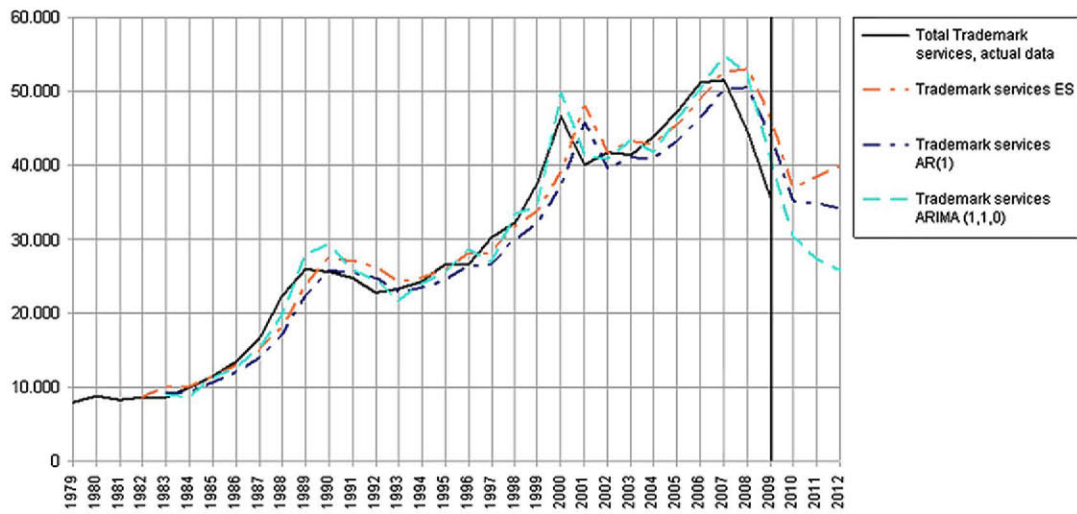


Fig. 10. Models of time series applied national trademark applications for services.

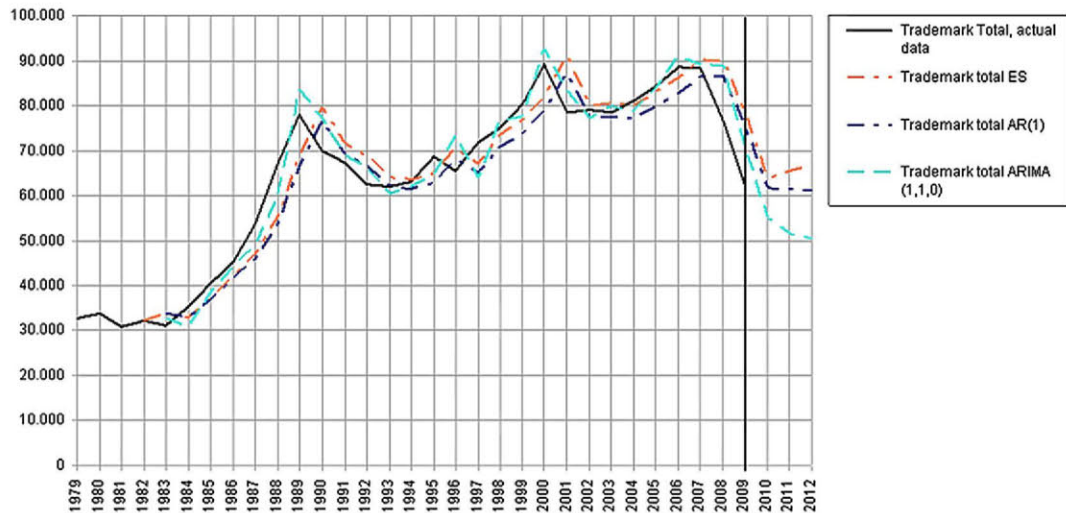


Fig. 11. Models of time series applied to national trademark applications (total).

Table 5
Statistics of the exponential isolation model.

Model	Model fit statistics								Ljung-Box Q(18)		
	R ² stationary	R ²	RMSE	MAPE	MAE	MaxAPE	MaxAE	BIC normalized	Statistics	GL	Sig
Patents	40.4%	82.9%	1205.7	12.931	546.8	129.962	5666.4	14.428	3.924	16	.999
Trademark of products	23.5%	73.0%	3654.7	6.588	2462.2	27.883	12335.9	16.646	16.680	16	.407
Trademark of services	17.6%	91.8%	3869.1	8.129	2526.1	29.819	10574.0	16.760	13.767	16	.616
Trademark (total)	12.9%	84.4%	6823.2	7.640	5037.9	26.036	16188.2	17.894	13.939	16	.603

Table 6
Statistics of model AR1.

Model	Model fit statistics								Ljung-Box Q(18)		
	R ² stationary	R ²	RMSE	MAPE	MAE	MaxAPE	MaxAE	BIC normalized	Statistics	GL	Sig
Patents	74.3%	82.9%	1478.0	13.284	658.9	127.599	5563.3	14.835	7.759	17	.971
Trademark of products	68.2%	68.2%	3966.8	8.031	2819.6	38.033	8924.6	16.809	13.785	17	.682
Trademark of services	86.0%	91.9%	5049.6	15.857	3402.9	189.806	16380.3	17.292	5.558	17	.996
Trademark (total)	77.2%	84.5%	8257.8	10.557	6130.4	76.957	24700.9	18.276	9.295	17	.931

Table 7
Statistics of model ARIMA (1,1,0).

Model	Model fit statistics								Ljung-Box Q(18)		
	R ² stationary	R ²	RMSE	MAPE	MAE	MaxAPE	MaxAE	BIC normalized	Statistics	GL	Sig
Patents (Ln)	88.0%	98.1%	393.9	5.272	248.3	11.733	1176.0	12.440	24.600	17	.104
Trademark of products (Ln)	46.7%	85.0%	2672.9	5.257	1886.8	15.971	6191.5	13.148	9.984	17	.904
Trademark of services (Ln)	54.3%	96.8%	2731.9	6.727	2019.7	17.290	7701.2	16.192	16.543	17	.486
Trademark total (Ln)	49.3%	92.7%	5727.9	5.871	3834.2	15.858	12164.3	18.283	7.919	17	.968

Table 8
Verification of the models' goodness of fit.

Model	Linear LT	Linear MT	Linear ST	Quadratic	Exponential isolation	AR(1)	ARIMA
Patents	56.9%	92.5%	89.5%	94.0%	82.9%	82.9%	98.1%
Trademark of products	15.5%	57.7%	55.5%	78.8%	73.0%	68.2%	85.0%
Trademark of services	90.7%	61.8%	0.0%	80.6%	91.8%	91.9%	96.8%
Trademark (total)	73.9%	18.1%	12.0%	61.1%	84.4%	84.5%	92.7%

models have been correctly specified for the time series analyzed, because they have significance values of more than 0.05 in all cases, which has also been corroborated in the correlograms of residuals, which demonstrated that these were not correlated. Last of all, the statistics on errors (RMSE, MAPE, MAE, MaxAPE and MaxAE) and normalized BICs are also lower in the ARIMA model than in the other models (with the exception of the BIC statistic in the case of the series of total trademarks, by a small difference), which

confirms that the model which is best adapted to the values is the ARIMA model (1,1,0) on the series in natural logarithms.

5. Conclusions

The ability of the Spanish Office of Patents and Trademarks to carry out efficient planning of its resource needs requires the use of methods which allow it to predict the changes in the number of patent and trademark applications at different time horizons. At present, there are different methods for predicting the behavior of patent and trademark applications for a short-term time horizon. Used in this study were different models for analyzing time series and regression series to estimate forecasted changes in the series of national patent and trademark applications (for products, for services and total) in the next three years.

From the methods used in this study, it can be concluded that the advanced models of time series possess a very good fit for the real values of the series in comparison with regression models of trends (Table 8), and therefore it is to be expected that the future values for the analyzed series will be close to the values predicted by the ARIMA, AR1 and exponential smoothing models, and within the confidence interval produced for the prediction. It is important to emphasize, as well, that the results of this study make it clear that it is feasible to model the series of national patent and trademark applications with different models of time series, above all

Table 9
Values of the prediction of changes in the time series of patent and trademark applications in accordance with the ARIMA model.

		2010	2011	2012
Patents	Forecast	3.740	3.738	3.737
	ICS 95%	4.290	4.570	4.792
	ICI 95%	3.246	3.027	2.869
Trademark of products	Forecast	23.778	22.303	21.649
	ICS 95%	27.579	29.565	32.535
	ICI 95%	20.391	16.492	13.796
Trademark of services	Forecast	30.270	27.322	25.757
	ICS 95%	35.825	38.028	42.039
	ICI 95%	25.400	19.090	14.803
Trademark (total)	Forecast	55.071	51.715	50.426
	ICS 95%	65.595	71.663	80.244
	ICI 95%	45.900	36.352	29.983

ARIMA, with satisfactory results in terms of the fit of models and relatively low error levels.

Using the results found in the application of the ARIMA model as a reference, the prediction of the trend in each time series is carried out using upper and lower limits according to the confidence level of 95%, this range being the interval in which it is most likely that the real future value of the series analyzed in the three-year time horizon will fall. Table 9 reflects the numerical values of the trends in the predictions by the ARIMA model for the different time series studied and their confidence intervals in the period of 2010–2012, in which ICS represents the highest confidence interval and ICI the lowest confidence interval.

By observing these data, it can be concluded that the patent applications display a nearly flat trend prediction, though at the upper and lower limits of the confidence interval, they reflect slight decreases or increases in the number of patent applications. As for the trademark applications for products, they display a slightly decreasing trend, something which also occurs in the series of trademarks for services and total trademarks. It is of interest to provide a reminder at this point that the predictions of future changes in trademarks within the exponential smoothing and AR1 models evaluated (Figs. 9, 10 and 11) showed a slight future trend toward increasing or remaining steady, according to the models, and that the confidence levels of the ARIMA model take this possibility into account. It must be underlined that, in the prediction by the ARIMA model for the series of total trademark applications, the sum of the ARIMA predictions form the series of applications for trademarks of products and for trademarks of services is slightly lower than the value of the prediction for the series of total trademarks, which is due mainly to the fact that this series includes a small percentage of unclassified trademark applications.

Acknowledgments

The paper proceeds from an original study based on a recent research financed by the Spanish Patent and Trademark Office and carried out during 2010. The authors wish to thank Gerardo Penas and Esther Arias for their support in the development of research.

References

- [1] Freeman C. The economics of industrial innovation. Cambridge: MIT Press; 1982.
- [2] Pavitt K. Uses and abuses of patent statistic. In: van Raan AFJ, editor. Handbook of quantitative studies of science and technology. Amsterdam: Elsevier; 1988. p. 509–36.
- [3] Dosi G. Sources, procedures and microeconomic effects of innovation. Journal of Economic Literature 1988;XXVI:1120–71.
- [4] Adams K, Kim D, Joutz FL, Trost RP, Mastrogianis G. Modeling and forecasting U.S. patent application filings. Journal of Policy Modelling 1997;19(5): 491–535.
- [5] Hingley P, Nicolas M. Methods for forecasting numbers of patent applications at the European patent office. World Patent Information 2004;26(3):191–204.
- [6] Bock C, Günter M, Haftka TK, Mitter M, Hüsler J. Forecast of trademark applications in Switzerland. World Patent Information 2004;26(4):275–82.
- [7] Gerschenkron A. Economic backwardness in historical perspective. Cambridge: The Helpman Press; 1962.
- [8] Rosenberg N. Perspectives on technology. Cambridge: Cambridge University Press; 1976.
- [9] Hobday M. Innovation in East Asia: the challenge to Japan. New York: Guilford Press; 1995.
- [10] Hidalgo A, Moleró J. Technology and growth in Spain (1950–1960): an evidence of Schumpeterian pattern of innovation based on patents. World Patent Information 2009;31(3):199–205.
- [11] Schmookler J. The level of inventive activity. Review of Economics and Statistics 1954;36(2):183–90.
- [12] Schmookler J. Invention and economy growth. Cambridge: Harvard University Press; 1966.
- [13] Hidalgo A. Analysis of the commercial use of Spanish inventions protected by patents between 1996 and 2006. Journal of Intellectual Property Rights 2009; 14(1):63–9.
- [14] Pavitt K. R&D, patenting and innovative activity: a statistical exploration. Research Policy 1982;11(1):33–51.
- [15] Scherer FM. The propensity to patent. International Journal of Industrial Organization 1983;1(1):107–28.
- [16] Bosworth D, Westaway T. The influence of demand and supply side pressures on the quantity and quality of inventive activity. Applied Economics 1984;16: 131–46.
- [17] Hall B, Griliches Z, Hausman J. Patents and R&D: is there a lag? International Economic Review 1986;27(2):265–83.
- [18] Pakes A. Patents as options: some estimates of the value of holding European patent stocks. Econometrica 1986;54(4):755–84.
- [19] Schankerman M. How valuable is patent protection? Estimates by technology field. RAND Journal of Economics 1998;29(1):77–107.
- [20] Gambardella A, Harhoff D, Verspagen B. The value of European patents. European Management Review 2008;5(2):69–84.
- [21] Arora A, Ceccagnoli M, Cohen W. R&D and the patent premium. In: NBER Working Paper, 9431; 2003.
- [22] Lemley M. Reconciling patents in the age of venture capital. Journal of Small and Emerging Business Law 2000;4:137–48.
- [23] Long C. Patent signals. University of Chicago Law Review 2002;69:625–79.
- [24] Hall B. Innovation and market value. In: Barrell R, Mason G, O'Mahoney M, editors. Productivity, innovation and economic performance. Cambridge: Cambridge University Press; 2000.
- [25] Kamiyama S, Sheehan J, Martínez C. Valuation and exploitation of intellectual property. In: STI Working Paper. OECD; 2006.
- [26] Hall B, Ziedonis R. The patent paradox revisited: an empirical study of patenting in the U.S. Semiconductor industry 1979–1995. RAND Journal of Economics 2001;32(1):129–51.
- [27] Mansfield E. Patents and innovation: an empirical study. Management Science 1986;32(2):173–81.
- [28] Meade N. Evidence for the selection of forecasting methods. Journal of Forecasting 2000;19:515–35.
- [29] Draper NR, Smith H. Applied regression analysis. USA: Wiley; 1981.
- [30] Maddala GS. Introduction to econometrics. USA: Wiley; 2001.
- [31] Pena D, Tiao GC, Tsay RSA. Course in time series analysis. USA: John Wiley and Sons; 2001.
- [32] Holt CC. Forecasting seasonal and trends by exponentially weighted moving averages. Memorandum N 52. Office of Naval Research; 1957.
- [33] Box G, Jenkins G. Time series analysis: forecasting and control. San Francisco: Holden-Day Inc.; 1970.



Antonio Hidalgo is Professor of Technology Strategy, Director of the Research Group of Innovation, IPR and Technology Policy, and Director of the Master in Economics and Innovation Management at the Universidad Politécnica de Madrid, Spain. He received an MBA and a PhD in Industrial Engineering from the Universidad Politécnica de Madrid. He acts as an expert to the European Commission as technology consultant in different European projects. His works have appeared in different international journals as R&D Management; Production, Planning and Control; Intelligent Manufacturing; Technology Transfer; Product Development; Entrepreneurship and Innovation Management; Asian Journal of Technology Innovation; and Innovation and Learning.



Samuel Gabaly is Director of Data Analysis at Gabaly Diseño SL, a company specialized in advanced data analysis, consulting, research and data mining. He is an economist and statistics expert with ICT and data analysis background and received an MBA from the Universidad Complutense de Madrid Business School, being awarded with the Student of the Year award by the Princess Cristina of Spain. He has designed and collaborated in many advanced studies and research projects for private and public sector clients, including European Commission, Government of Spain, Deloitte, Altran and McKinsey in areas like innovation, R&D, ICT, transport and other sectors.