

Insua, Manuel; Frias, Zoraida; Pérez Martínez, Jorge

**Conference Paper**

## Application of multiple regression analysis to the prices of the spectrum in the IMT band

28th European Regional Conference of the International Telecommunications Society (ITS): "Competition and Regulation in the Information Age", Passau, Germany, 30th July - 2nd August, 2017

**Provided in Cooperation with:**

International Telecommunications Society (ITS)

*Suggested Citation:* Insua, Manuel; Frias, Zoraida; Pérez Martínez, Jorge (2017) : Application of multiple regression analysis to the prices of the spectrum in the IMT band, 28th European Regional Conference of the International Telecommunications Society (ITS): "Competition and Regulation in the Information Age", Passau, Germany, 30th July - 2nd August, 2017, International Telecommunications Society (ITS), Calgary

This Version is available at:

<http://hdl.handle.net/10419/169467>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



# Application of multiple regression analysis to the prices of the spectrum in the IMT band

**Manuel Insua, Zoraida Frias, Jorge Pérez Martínez**

Universidad Politécnica de Madrid  
Dept. Signals, Systems and Radiocommunications  
28040 Madrid, Spain  
m.insua.f@gmail.com  
{zfrias; jperez}@gtic.ssr.upm.es

## Abstract

This paper explores the effects of different variables on the price paid for the spectrum by Mobile Service Providers (MSP) in several European countries. Multiple regression analysis is used to find relations between these variables and the Price/MHz/pop, commonly used in the industry to compare prices across bands and markets. The results show that the most important factor that affects spectrum prices is the frequency band, followed by the country, and, to a lesser extent, the number of competitors and the year of allocation. The analysis is also used to create a regression function capable of predicting future prices. The model has been cross-validated with data from past tenders, with mixed results. The lowest prediction errors are around 10%. Even if accuracy is not high, the model provides a valid range of price estimates in which real prices oscillate. Further research is needed regarding variables not being currently captured in the model, as the balance between supply and demand of spectrum at any one awarding process.

**Keywords** Spectrum management; Spectrum pricing; Spectrum auction; Mobile Networks;

## 1 INTRODUCTION.

### 1.1 MOTIVATION.

Radio spectrum is becoming an increasingly important resource as the number of mobile devices and their ability to download content grow. According to Cisco (Cisco, 2017), mobile data traffic has grown 18-fold over the past five years and will increase sevenfold between 2016 and 2021. The limited nature of radio spectrum makes it a valuable resource that must be shared across a wide variety of uses.

Due to this scarcity, spectrum allocation mechanisms have long followed a valuation method that aims to guarantee efficiency (Coase, 1959). In recent years, Europe has carried out several reforms of the management framework to make available more spectrum for the rollout of new technologies. This has been done fundamentally in two ways. On the one hand, the harmonization of new frequency bands to be allocated to mobile services (European Commission, 2008). On the other, legacy spectrum has been refarmed to enable greater efficiencies in the bands that were being used by old technologies (European Commission, 2009) (European Commission, 2009).

Expecting a spectrum shortage in the near future, the World Communications Conference attributed the 700MHz band to International Mobile Telecommunication (IMT) services in 2012. This band ranges from 694 MHz to 790 MHz, and it is shared with broadcasting services.



Following this decision, several spectrum auctions are expected to take place in Europe over the next few years, since all Member States will need to repurpose the 700 MHz band for mobile by 2020. Indeed, countries like Germany and France have already allocated 700 MHz spectrum, and others, like the United Kingdom, have started the consultation processes with different stakeholders. Other bands, like the 2.3 GHz and 3.4 GHz bands, are also foreseen to be auctioned in the short run.

This study aims to find which factors will influence these bidding processes and a way to model spectrum prices based on previous data from spectrum allocation processes and related market variables at the time the spectrum was allocated.

## 1.2 RELATED WORK.

As mobile data traffic grows and the need to distribute spectrum resources efficiently increases, a number of studies have emerged seeking to find appropriate spectrum pricing methods.

Some of these methodologies are based on market mechanisms used in other economic sectors, for example, the model of Modification to the Benchmark Price (Wang, Wang, & Sun, 2010) or the application of the Smith-NERA method (Doyle, 2006). The Smith-NERA method consist in identifying the technical substitution rate between two different inputs, assuming that the amount of production of a company remains constant. In the context of wireless mobile networks, the alternative input that can replace the use of new spectrum is typically considered to be the deployment of more base stations. The Smith-Nera method values radio spectrum as how much it would cost to deploy the additional base stations to provide the same service level.

On the other hand, other methods proposed in the literature of dynamic spectrum allocation systems are based on the physical properties of spectrum (Kerans, Vo, Conder, & Krusevac, 2011). The authors explore a method for assigning value to spectrum based solely on its physical properties for information conveyance, i.e. frequency. This approach arises along with the emergence of systems with dynamic access to spectrum and software defined radio, which need for real-time spectrum pricing.

To sum up, existing literature has focused on both the analysis related to propagation properties of radio spectrum and applying dynamic valuation methods used in other economic sectors. This paper proposes to create a model that combines both approaches through a multiple regression analysis using historic prices.

The approach we propose has been used in literature (Bohlin, Madden, & Morey, 2010) to analyze the factors affecting 3G spectrum prices, although the study focuses only on 86 licenses, all of them belonging to the same band (2100 MHz).

## 1.3 CONTRIBUTION.

The recent allocation of additional spectrum bands for IMT systems allows for a more comprehensive analysis regarding the factors affecting spectrum prices across frequencies and markets. In addition, other factors, such as market competition, GDP, or interference risk, may affect, to a lesser extent, the spectrum value.

This study aims to find a way to quantify the extent to which different factors affect spectrum prices. To this end, we have carried out a multiple regression analysis, which, in addition, allows to create a model intended to predicting spectrum prices based historical information of awarded licenses. Finally, we have tested the model prediction capabilities retrospectively and assessed its performance.



## 2 METHODOLOGY.

### 2.1 DATA COLLECTION.

The data used in the analysis comprises information from 276 spectrum licenses of nine different countries. Each license is reflected in the database as a sample with 20 variables that represent several characteristics of the allocation process and the conditions under which the license is awarded.

In order to collect the price information, we rely on the information provided by the pertinent NRAs (National Regulatory Authorities), which have been in charge of the awarding processes (e.g. Ofcom in the UK or l'Arcep in France). Generally, the documentation published by these NRAs includes information on frequency, bandwidth, channel duplexing, year of commencement of the license, number of competitors who participated in the awarding process, whether the block of frequencies is abstract or specific and whether the license has obligations associated with it. All this information is reflected in the database in the form of variables.

Since it is not always possible to obtain all the information from these sources, we have additionally used Eco Report 03 published by the CEPT to complete the information of channel duplexing, as well as the start year of the license. This report also provides information about the duration of the license and whether it is tradeable. We have also added a variable that indicates whether the licensed block may suffer from interference from other sources. This variable is of boolean type, and it is decided for each case based on adjacent uses.

Regarding the variables that capture socioeconomic aspects that may affect the license, we have relied on different sources. We have used the World Bank and the World Health Organization to obtain information on population and population density. Information on GDP per capita and ARPU in the retail mobile market has been obtained from the Digital Agenda Scoreboard (European Commission, n.d.).

Finally, the Herfindahl-Hirschman Index (HHI), a measure of market concentration, has been obtained from a report published by Altran (Altran, 2015) regarding the telecommunications market (United States Department of Justice, 2015).

Table 1 summarizes the set of variables used in the analysis.

<b>Variable</b>	<b>Definition</b>	<b>Type</b>	<b>Units</b>
<i>PRICE</i>	Price paid per MHz and per inhabitant	Continuous	EUR/MHz/pop
<i>COUNTRY</i>	Country the license is assigned in	Categorical. 9 Options	n/a
<i>FREQ</i>	Frequency of the band the block is in	Continuous	MHz
<i>BAND</i>	Frequency band of the block	Categorical. 6 Options	n/a
<i>BW</i>	Bandwidth of the block	Continuous	MHz
<i>DUPLEX</i>	Whether the uplink and the downlink are separated in frequency	Categorical. 2 Options	n/a
<i>DURAT</i>	Difference between the ending date and the starting date of the license	Continuous	Days
<i>TRADE</i>	Ability to sell the license in the future	Categorical. 2 Options	n/a
<i>YEAR</i>	Year of the starting date of the license	Continuous	Years



<i>SPECIF</i>	Whether the block is abstract or specific prior to the bidding	Categorical. 2 Options	n/a
<i>OBLIG</i>	Whether the license include obligations of any kind or not	Categorical. 2 Options	n/a
<i>GDP</i>	Current GDP per capita. Eurostat code: nama_gdp_c	Continuous	Euros
<i>ARPU</i>	Average Revenue Per User in the Retail Mobile Market	Continuous	Euros
<i>POP</i>	Number of people living in the country of the license	Continuous	People
<i>POPDEN</i>	Average number of people living in each squared kilometer in the country of the license	Continuous	People/Km <sup>2</sup>
<i>HHI</i>	Herfindahl-Hirschman Index. Commonly accepted measure of market concentration	Continuous	Number
<i>COMPET</i>	Number of MNOs that obtained any spectrum in the same band during the same auction	Continuous	Number
<i>INTERF</i>	Whether the frequency block is interfered or not	Categorical. 2 Options	n/a

TABLE 1 SUMMARY OF THE ANALYZED VARIABLES

## 2.2 REGRESSION ANALYSIS.

In this article, we have used regression analysis for two different purposes. Firstly, as a measure for the level of correlation between two variables using simple regressions. Secondly, to create prediction models using multiple regressions.

In the first place, a simple regression analysis allows to assess the extent to which a dependent variable, as the price of spectrum in EUR/MHz/pop, can be explained by a single predictor or control variable. In addition, we obtain measures of statistical significance for each case, that is, the probability of one variable affecting the other. In addition, a set of values and statistic parameters help us identify the quality of the model obtained, as explained next.

For the subsequent analysis, the concept of *residuals* needs to be defined. In statistics, the residuals are measures of deviation between observations and theoretical values. Specifically, a residual ( $\varepsilon_i$ ) is the difference between the actual value of the dependent variable ( $y_i$ ) and the predicted value for it ( $\hat{y}_i$ ) (Kutner, Nachtsheim, & Neter, 2004). This is:

$$\varepsilon_i = y_i - \hat{y}_i$$

Through regression analysis, we obtain the estimates of the coefficients of the model, which correspond to the terms  $\beta$  of the function described by a regression model, presented below (Kutner, Nachtsheim, & Neter, 2004).

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_N x_N$$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 \text{ for simple regression}$$

For each variable included in the model, a coefficient  $\hat{\beta}_n$  is obtained which refers to the weighting factor of the variable in the dependent variable ( $y$ ). In addition to these coefficients,  $\hat{\beta}_0$  refers to the constant term of the function. For each of these



coefficients  $\widehat{\beta}_n$ , three different statistics are calculated: the standard error, the t-value and  $Pr(>|t|)$ .

The standard error ( $s_{\widehat{\beta}_n}$ ) represent the average residuals normalized to the deviation of the control variables, as it is noted in the equation below. In other words, is a measure of the extent to which the estimates of the coefficients ( $\widehat{\beta}_n$ ) affect the average of the dependent variable ( $\widehat{y}$ ). This variable also allows to calculate a confidence interval around the regression function. To create a reliable model, it is desirable to have the minimum standard error in relation to the coefficient they refer to. The standard error is then calculated as in (Steel & Torrie, 1960):

$$s_{\widehat{\beta}_n} = \sqrt{\frac{\frac{1}{I-2} \sum_{i=1}^I \widehat{\varepsilon}_i^2}{\sum_{i=1}^I (x_i - \bar{x})^2}}$$

$$\text{where } \bar{x} = \frac{1}{I} \sum_{i=1}^I x_i$$

The t-value is a measure of how many times the standard deviation deviates from our estimate of zero. The further from zero, the greater the probability of being right when rejecting the null hypothesis, that is, stating that there is a relation between the explanatory variables and the dependent variable. The t-value can be calculated as follows (Box, Hunter, & Hunter):

$$t_n = \frac{\widehat{\beta}_n - \beta_n}{s_{\widehat{\beta}_n}}$$

Finally,  $Pr(>|t|)$  corresponds to the probability of observing a value equal to or greater than  $|t|$ . According to that, we want the value to be close to zero because if  $Pr(>|t|)$  is very low, the probability of finding an equal or greater t-value is very unlikely. Therefore, a low  $Pr(>|t|)$  indicates that we can discard the null hypothesis.

In addition to these three parameters referred to characteristics of the coefficients ( $\widehat{\beta}_n$ ), there are other useful statistic values referred to the model as a whole. These statistics are: the residual standard error, the value of and of adjusted  $R^2$  for the number of variables, the value of F-statistic, and the p-value.

The residual standard error is a measure of how the regression is adjusted to the samples used and, therefore, the quality of the model. It refers to the mean deviation between the samples of the dependent variable and the prediction obtained by the regression model. Its value is expressed as (James, Witten, Hastie, & Tibshirani, 2014):

$$RSE \text{ (residual standard error)} = \sqrt{\frac{1}{I-2} \sum_{i=1}^I \varepsilon_i^2}$$

The statistic  $R^2$ , also known as coefficient of determination, allows us to measure how well the model adjusts to real data. Specifically, it is a measure of the proportion of the variance in the data that is able to explain the model created. There are several ways to calculate this coefficient, but the one that will be used in this work is the following (R Project):



$$R^2 = 1 - \frac{\sum_{i=1}^I \varepsilon_i^2}{\sum_{i=1}^I (y_i - \bar{y})^2}$$

In addition to the value of  $R^2$ , we provide the value of  $R^2$  adjusted to the number of variables. This value is interesting because when adding explanatory variables to a model, the value of  $R^2$  always increases. The value of  $R^2$  adjusted allows to correct the value of the statistic taking into account the number of variables. This allows comparing  $R^2$  from different models with unequal number of variables. To obtain  $R^2$  the following expression is used (R Project):

$$R_{adj}^2 = 1 - \frac{(1 - R^2) (I - 1)}{I - N - 1}$$

Where  $I$  is the number of samples and  $N$  is the number of explanatory variables.

*F-statistic* is a measure of how good the regression is when compared to a random prediction. This is expressed in relation to one. If the value is less than one, the prediction is worse than a random alternative. If it is greater than one, the prediction is better, and the higher the better the model predicts and thus, the null hypothesis can be discarded. It is calculated as the ratio between the root of the residual sum of squares (RSS) of the model created and the RSS of a model generated using the Fisher-Sendecho distribution as a null hypothesis.

Finally, we can assess the model reliability through the p-value. This parameter is the equivalent to  $Pr(>|t|)$  obtained for each coefficient  $\hat{\beta}_n$ , but for the whole model. The smaller p-value, the better the model can explain the breakdown of the samples provided into individual components.

Finally, an iterative process has been carried out to improve the model performance, in which variables were added or eliminated depending on whether the predictions were improved. To assess such an improvement, the statistics described above were used.

## 2.3 VALIDATION.

To validate the model developed, a cross-validation methodology is used. This allows us to assess the performance of the model when using data that the model has not managed previously. To this end, we have split the data into training and test sets so as to evaluate the errors delivered by a model built on the training set when predicting the values of the test set (Devijver & Kittler, 1982).

In the particular case of spectrum, cross-validation needs to take into account the time dimension, as the samples follow a longitudinal sequence. Therefore, we have applied this methodology bearing in mind that the training set needs to contain samples temporarily prior to the ones that will be used in the tests. This way of proceeding allows for a better representation of real spectrum pricing processes, in which the information of previous allocations is available, but not that of the future ones.

# 3 RESULTS.

## 3.1 RELATION BETWEEN VARIABLES AND PRICE/MHZ/POP.

As explained in section 2.2, a simple regression analysis has been performed between the spectrum prices (as of EUR/MHz/pop) and each of the explanatory variables, before carrying out the multiple regression analysis.



The results of the analysis are shown below in table 2. It is important to note that the samples regarding 2100 MHz spectrum have been left out, except for the samples of the auction held by Germany in 2016 around 2000 MHz. This is because spectrum prices in this band are well above those of other bands, indeed more than four times bigger than the rest due to the dot-com bubble. This price imbalance undermines the analysis since these few samples greatly worsen the correlation results and cause the regression models to deviate too much to fit these samples.

<i>Variable</i>	<i>Residual std. Error</i>	<i>Adjusted R<sup>2</sup></i>	<i>F-statistic</i>	<i>P-value</i>	<i>Samples</i>
<i>COUNTRY</i>	0.1989	0.164	7.0	3.23E-08	243
<i>Log10(FREQ)</i>	0.1295	0.646	441.8	2.20E-16	243
<i>BAND</i>	0.1159	0.716	123.2	2.20E-16	243
<i>BW</i>	0.2176	-0.001	0.9	3.49E-01	243
<i>DUPLEX</i>	0.2120	0.050	13.9	2.47E-04	243
<i>DURAT</i>	0.2127	0.035	9.7	2.09E-03	242
<i>TRADE</i>	0.2317	0.001	1.2	2.85E-01	207
<i>YEAR</i>	0.2084	0.074	20.2	1.06E-05	242
<i>SPECIF</i>	0.2195	0.002	1.5	2.29E-01	231
<i>OBLIG</i>	0.2156	0.021	6.0	1.53E-02	229
<i>GDP</i>	0.2163	0.006	2.3	1.34E-01	206
<i>ARPU</i>	0.2294	-0.002	0.6	4.23E-01	179
<i>POP</i>	0.2151	0.023	6.6	1.10E-02	243
<i>POPDEN</i>	0.2179	-0.003	0.2	6.78E-01	243
<i>HHI</i>	0.2201	-0.003	0.3	5.97E-01	228
<i>COMPET</i>	0.2103	0.065	17.9	3.38E-05	243
<i>INTERF</i>	0.2192	0.004	2.0	1.62E-01	231

TABLE 2 SIMPLE REGRESSION ANALYSIS RESULTS

As can be seen in table 2, the variables that affect prices the most are those that refer to the frequency and the country of the license, followed by the number of competitors in the awarding process and the start year of the license.

### 3.2 PREDICTION MODEL.

The results obtained from simple regression analysis provides valuable information to create a model that relays on several of these variables. Until a final model was reached, we performed several iterations, varying the number of control variables.

The first iteration, not presented here, included the six variables that we considered that explain price variability best (*Log10(FREQ)*, *BAND*, *COUNTRY*, *YEAR*, *COMPET*, *DUPLEX*). As the variables *FREQ* and *BAND* contain similar information, only *BAND* has been used in the iterative process. In addition, the variables *DURAT*, *SPECIF*, *OBLIG* and *INTERF* were added in the process because they allow the model to fit better to the prices of the licenses.

The estimates of the optimal coefficients for the regression function are shown below in Table 3 and Table 4. Along with the estimates, the statistical values explained in Section 2.2 are presented.

	<b>Estimate</b>	<b>Std. Error</b>	<b>T value</b>	<b>Pr (&gt; t )</b>
$\beta_0$	-4.08E+01	8.38E+00	-4.868	2.23E-06
<i>BAND = 800MHz</i>	3.69E-01	3.90E-02	9.468	< 2e-16



<i>BAND = 900MHz</i>	1.07E-02	3.78E-02	0.282	7.78E-01
<i>BAND = 1800MHz</i>	-1.25E-01	3.32E-02	-3.777	2.07E-04
<i>BAND = 2100MHz</i>	-2.29E-01	5.43E-02	-4.208	3.84E-05
<i>BAND = 2600MHz</i>	-2.30E-01	3.67E-02	-6.268	2.07E-09
<i>COUNTRY = ES</i>	1.07E-01	2.65E-02	4.024	8.02E-05
<i>COUNTRY = FR</i>	2.56E-01	4.63E-02	5.534	9.37E-08
<i>COUNTRY = IE</i>	-7.64E-03	4.89E-02	-0.156	8.76E-01
<i>COUNTRY = IT</i>	8.89E-02	2.42E-02	3.665	0.000314
<i>COUNTRY = NL</i>	-2.29E-02	6.25E-02	-0.366	7.15E-01
<i>COUNTRY = PT</i>	-2.67E-01	4.48E-02	-5.949	1.13E-08
<i>COUNTRY = SE</i>	-2.43E-02	5.07E-02	-0.48	0.631903
<i>COUNTRY = UK</i>	-8.42E-02	6.36E-02	-1.324	0.187059
<i>YEAR</i>	2.06E-02	4.17E-03	4.949	1.54E-06
<i>COMPET</i>	4.16E-02	1.32E-02	3.154	0.001849
<i>DUPLEX = YES</i>	3.17E-02	1.71E-02	1.847	0.066144
<i>DURAT</i>	-8.49E-05	1.41E-05	-6.022	7.71E-09
<i>SPECIF = Specific</i>	9.40E-02	3.35E-02	2.809	0.005442
<i>OBLIG = YES</i>	-1.17E-01	3.09E-02	-3.794	1.94E-04
<i>INTERF = YES</i>	-1.51E-01	3.02E-02	-5.009	1.17E-06

**TABLE 3 ESTIMATES OF THE COEFFICIENTS OF THE REGRESSION MODEL**

<b>RESIDUAL STD. ERROR</b>	<b>ADJUSTED R<sup>2</sup></b>	<b>F-STATISTIC</b>	<b>P-VALUE</b>	<b>SAMPLES</b>
0.0796	0.8666	75.05	< 2.2e-16	229

**TABLE 4 STATISTICS OF THE MODEL**

As summarized in Table 4 the model reaches a  $R^2$  value of 0.87. This value is not so high as to make highly accurate predictions, but it allows making predictions with errors below 10%.

### 3.3 VALIDATION RESULTS.

As described in section 2.3, price predictions are cross-validated taking into account when the samples of the model were generated, as they belong to different time instances. The tests that have been performed are the following ones:

#### 3.3.1 LATEST AWARDS OF THE 800MHz AND 2600MHz BANDS.

These bands are commonly used for 4G. This scenario intends to assess how good the prediction would have been at that moment. For this, the Italian auction of 2013 is used as the test set because it was the last 4G auction held out of the ones collected for this study. As a training set, it would be desirable to use only the previous awards of all the bands studied, with the exception of the 2100MHz band, due to the inflated prices due the dot com bubble. Nonetheless, it has been necessary to include it to make the model work. Since the regression function uses the country as a predictor, at least one Italian sample must be included in the training set. As the only previous licenses of Italian spectrum are those of the band of 2100MHz we must include those and therefore all other samples of this band. In summary, the training set and the test set are as follows:

- Training set: Licenses prior to the test set, i.e. all licenses of the 800MHz, 900MHz, 1800MHz, 2100MHz and 2600MHz bands with a date prior to 2013.
- Test set: Italian 800MHz and 2600MHz bands licenses.



The results of this test are the following:

<i>COUNTRY</i>	<i>FREQ</i>	<i>Real Value Average EUR/MHz/pop</i>	<i>Prediction Average EUR/MHz/pop</i>	<i>Residual Average</i>
IT	800	0.8080	1.4980	-0.6899
IT	2600	0.0589	1.2272	-1.1683

**TABLE 5 CROSS-VALIDATION RESULTS FOR THE LAST 4G LICENSE**

In this case, the accuracy of the predictions is very low. This is mainly because we are forced to include 3G licenses in the training set, so the model tries to fit to the high prices of these licenses, causing very large errors. This test does not serve to evaluate the reliability of the model since we can not train it without samples of 3G licenses and therefore avoid this error. In contrast, this test serves to illustrate why it is a good idea to exclude from the creation of models the licenses affected by the dot-com bubble.

### 3.3.2 REFORMED SPECTRUM.

In recent years, the bands of 900MHz and 1800MHz have been allocated again mostly due to refarming processes. This test intends to know how the model would have worked in the last awarding process of one of these bands. In this case, the test set consists of the Swedish licenses of the 1800 MHz band. Given that in this case we do have samples of licenses from Sweden other than for the 2100 MHz band, we can exclude this band from the training set. Taking all this into account, the two sets are as follows:

- Training set: 800MHz, 900MHz, 1800MHz and 2600MHz licenses with date before 2013.
- Test Set: Swedish 1800MHz band licenses.

The results of this test are presented next:

<i>COUNTRY</i>	<i>FREQ</i>	<i>Real value EUR/MHz/pop</i>	<i>Prediction EUR/MHz/pop</i>	<i>Residual</i>
SE	1800	0.2503	0.4570	-0.2067
SE	1800	0.2142	0.4570	-0.2428

**TABLE 6 CROSS-VALIDATION RESULTS FOR THE LAST REFORMING LICENSE**

The first observation we can make is that the withdrawal of the 2100MHz band from the training set provides lower predictions and thus better matches to the actual prices of the 1800 MHz spectrum. We can also see that the predictions in this case deviate much less than in the previous test (about 0.2 EUR/MHz/pop). However, the accuracy of the model, although it provides a rough estimate of the price, does not allow predicting the valuation of the spectrum accurately.

### 3.3.3 FIRST AUCTIONS OF 700 MHz.

The French and German auctions of the band of 700MHz have been the first two auctions made in this band, being, therefore, of special interest. In this scenario, we intend to assess whether using all historical samples available before both auctions, we could have predicted the price of these licenses. In this case, the test set consists only of samples belonging to this band (700MHz) and training set consists of all other samples prior to these bids except for those belonging to the 2100 MHz band. The sets are as follows:

- Training set: Licenses of the bands of 800MHz, 900MHz, 1800MHz and 2600MHz with date before 2015.



- Test set: 700 MHz band licenses from Germany and France.

The results are presented next:

COUNTRY	FREQ	Real Value Average EUR/MHz/pop	Prediction Average EUR/MHz/pop	Residual Average
FR	700	0.7244	0.8162	-0.0918
DE	700	0.2051	0.5602	-0.3552

**TABLE 7 CROSS-VALIDATION RESULTS FOR THE 700MHz LICENSES**

The results of this test are the most important since it is the test that most closely resembles the situations that we will be having in the near future. They may appear contradictory but they are encouraging for future research, as explained next.

Firstly, the price predictions of the French licenses are accurate, only deviating 0.1 EUR/MHz/pop from actual values. In the case of the German auction, the error is bigger. Looking at the real values, we can see that in Germany the prices of this band are much lower than those reached in France. This relationship does not correspond to what happens in other bands, in which the valuations in EUR/MHz/pop of both countries used to be very similar and in no case has historically diverged more than 0.1 EUR/MHz/pop.

However, while in France the auction only included blocks of the 700MHz band, the German auction includes blocks from a wide variety of bands including the 900MHz band. This is one of the potential explanations for this, since the frequency bands below 1GHz have very similar propagation characteristics and thus we might consider the 900MHz spectrum as a substitute product for 700MHz. Therefore, in this context of increased supply, prices can be unsurprisingly lower and consequently justify the prediction error of the model for the German case.

In addition, we can consider there are factors yielding to increased demand in sub-1GHz spectrum in France. This is because in this country there is an operator (Free Mobile) that does not have spectrum in this range of frequencies. As suggested in (Frias, González-Valderrama, & Pérez Martínez, 2017), operators not having this type of spectrum in its existing portfolio perceive greater value from sub-1GHz than those that already rely on low frequency spectrum, since the former would benefit from increased savings in future rollout costs. This fact can be the cause of increased demand in France and consequently of higher prices.

Taking into account these differences in both markets, the price differences and errors are logic, since the prediction model does not take into account the supply and demand that occurs during a bidding process.

\*\*\*

Despite the varying accuracy of the results, as time progresses and a larger number of samples are available, the models may make better predictions. In any case, the prediction models with better results present errors close to 10%. This means that predictions made by linear regression models can give a price range, but not a highly accurate prediction.

## 4 CONCLUSIONS AND FUTURE WORK.

This paper explores the effects of different variables over the price paid for spectrum licenses by mobile service providers (MSP) in several European countries. Regression analysis is used to quantify the extent to which each variable explains spectrum prices. The results suggest that the



two main factors affecting spectrum price are the frequency ( $R^2 = 0.647$ ) and the country ( $R^2 = 0.164$ ) awarding the license, while other factors, such as the number of competitors participating in the awarding process or the start year of the license influence spectrum prices to a lesser extent. Finally, sociodemographic factors such as GDP (Gross Domestic Product) per capita or ARPU appear not to affect at all ( $R^2 \sim 0$ ), contrary to what it might have been expected.

The methodology employed also allows us to create a regression model to obtain predictions for spectrum licenses. This model has been tested in several past awarding processes with dissimilar results. Although the predictions improve as we progress over time, since more samples are added to the model, prediction errors are never below 10%. Even if accuracy is below what would have been desirable, the model provides a valid range of price estimates in which the real prices oscillate.

This error might be due to the difficulties experienced when gathering proxies for some variables, that clearly affect spectrum prices, such as interference susceptibility or coverage obligations. In both cases, it is difficult to find a clear metric that enables comparing said factors across spectrum bands, and, particularly, across countries. In other cases, variables external to the analysis can explain the errors, as in the case of 3G licenses, which were clearly overrated due to the dot-com bubble.

All in all, regression analysis can offer regulators and practitioners a better insight about the factors affecting spectrum the most, which is particularly useful for better-informed pricing towards coming spectrum auctions. Future work comprises an in-depth analysis of these error sources as well as a comparative analysis with machine-learning algorithms.

## 5 REFERENCES.

- Altran. (2015). *TEM Report - Value Migration in the Telecom & Media Sector in Europe: a European tragedy?*
- Bohlin, E., Madden, G., & Morey, A. (2010). *An Econometric Analysis Of 3G Auction Spectrum Valuations*. Florence: European University Institute. Retrieved from <http://cadmus.eui.eu/handle/1814/14237>
- Box, G., Hunter, W., & Hunter, J. S. (n.d.). In *Statistics for Experimenters* (pp. 66–67). John Wiley & Sons.
- Cisco. (2017). *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper*. Retrieved March 26, 2017, from <http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html>
- Coase, R. H. (1959). The Federal Communications Commission. *The Journal of Law & Economics*, 2, 1-40. Retrieved from <http://www.jstor.org/stable/724927>
- Devijver, P., & Kittler, J. (1982). *Pattern recognition : a statistical approach*. Englewood Cliffs. Retrieved from <http://library.wur.nl/WebQuery/clc/183980>



- Doyle, C. (2006). The economics of pricing radio spectrum. *Warwick Business School*. Retrieved 03 20, 2017, from <http://userpage.fu-berlin.de/~jmueller/its/conf/berlin04/Papers/Doyle.pdf>
- European Commission. (n.d.). *Digital Single Market Scoreboard*. Retrieved from <https://ec.europa.eu/digital-single-market/en/policies/scoreboard>
- European Commission. (2008). *European Commission (2008a). Commission Decision of 13 June 2008 on the harmonisation of the 2500–2690 MHz frequency band for terrestrial systems capable of providing electronic communications services in the Community (2008/477/EC), L 163/37.*
- European Commission. (2009). *European Commission (2009a). Commission Decision of 21 May 2008 on the harmonisation of the 3400–3800 MHz frequency band for terrestrial systems capable of providing pan-European electronic communications services in the Community (2009/766/EC), L 274/32.*
- European Commission. (2009). *European Commission (2009b). Commission decision of 6 May 2010 on harmonised technical conditions of use in the 790–862 MHz frequency band for terrestrial systems capable of providing electronic communications services in the European Union (2010/267/EU),.*
- Frias, Z., González-Valderrama, C., & Pérez Martínez, J. (2017). Assessment of spectrum value: The case of a second digital dividend. *Telecommunications Policy*. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0308596116302786>
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). In *An Introduction to Statistical Learning*. Springer.
- Kerans, A., Vo, D., Conder, P., & Krusevac, S. (2011). Pricing of spectrum based on physical criteria. In *New Frontiers in Dynamic Spectrum Access Networks (DySPAN), 2011 IEEE Symposium on* (pp. 223–230). IEEE. Retrieved 3 19, 2017, from <http://ieeexplore.ieee.org/document/5936210/>
- Kutner, M. H., Nachtsheim, C. J., & Neter, J. (2004). In *Applied Linear Regression Models* (4 ed., p. 25). McGraw-Hill/Irwin.
- R Project. (n.d.). *R Documentation - Summarizing Linear Model Fits*. Retrieved from <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/summary.lm.html>
- Steel, R., & Torrie, J. H. (1960). In *Principles and Procedures of Statistics with Special Reference to the Biological Sciences* (p. 288). McGraw Hill.
- United States Department of Justice. (2015). Herfindahl–Hirschman Index.
- Wang, H., Wang, Q., & Sun, J. (2010). Spectrum pricing based on the method of modification to the benchmark price. In *2010 2nd International Conference on Information Science and Engineering (ICISE)* (pp. 2379–2382). IEEE. Retrieved 03 19, 2017, from <http://ieeexplore.ieee.org/abstract/document/5689007/>