Technological Forecasting & Social Change 80 (2013) 132–147

Contents lists available at SciVerse ScienceDirect

Technological Forecasting & Social Change

Driving factors during the different stages of broadband diffusion: A non-parametric approach

Vagia Kyriakidou *, Christos Michalakelis, Thomas Sphicopoulos

Department of Informatics and Telecommunications, National and Kapodistrian University of Athens, Panepistimiopolis, 15784, Ilissia, Athens, Greece

ARTICLE INFO

Article history:
Received 29 July 2010
Received in revised form 2 May 2012
Accepted 4 July 2012
Available online 28 July 2012

Keywords:
Socio-economic
Broadband driving factors
Non-parametric analysis
Dimension reduction
SIR

ABSTRACT

A number of socio-economic factors influence broadband diffusion and are assumed to be responsible for the different levels of adoption among countries. Most of the approaches met in literature do not take into account the different stages of the diffusion process and they are parametric. They are based on the construction and study of econometric models, which include all the influential parameters. Regression functions are defined prior to the analysis and they are usually linear in nature. However, this increases the complexity of the system, since the less influential variables are derived during the final stages of the analysis. On the contrary, non-parametric methods can achieve dimension reduction during the early stages, while they do not require a definition of the regression function.

The present work studies the effect of a wide range of social, economic and political factors over the broadband diffusion process, following a non-parametric approach and comparing the results with these of the parametric. Based on criteria from information theory the link function between the level of penetration and the rest variables is derived, providing highly accurate results. The evaluation of the methodology was performed over countries from the wider European area. It is proven that the different stages of broadband diffusion process, defined by the inflection point, are affected by different factors.

1. Introduction

The level of broadband diffusion is considered as an important indicator for a country, related directly to its development, due to the varied capabilities offered by the information and communications technologies [1]. In the area of the European Union (EU) the different levels of broadband penetration among member countries raise entrepreneur and political considerations. The main obstacle for a wider and faster broadband adoption used to be the lack of infrastructures or the limitation of transmission data traffic [2]. EU noticed this necessity and decided to subsidize initiatives regarding the deployment of broadband infrastructures [3], since infrastructures' competition enhances broadband development [4]. A number of studies were published during the last years, aiming to understand the development of broadband adoption by revealing common influential factors [5]. The initial purpose of these researches was the evaluation of the forecasting approach [6–9]. Forecasting methodologies were also used in order to estimate digital divide convergence among countries. As EU declared an "information society for all" through policy framework adopting in June 2005 [3], digital convergence became one of the main issues of relevant studies [10].

In line with the early studies regarding the factors fuelling broadband demand among and within countries, a number of social, economic, demographical and other factors have been extensively studied in literature. These factors are considered as the driving determinants of the broadband diffusion process. Depending on the factors in each examined case, results provide useful information to telecom operators and to policy makers for a more effective strategy deployment. Apart from the identification of the most important determinants, the usage of different statistical analyses and tools could lead to different conclusions as well.

* Corresponding author. Tel.: +30 210 7275184; fax: +30 210 7275214.
E-mail addresses: bkiriak@di.uoa.gr (V. Kyriakidou), michalak@di.uoa.gr (C. Michalakelis), thomas@di.uoa.gr (T. Sphicopoulos).

0040-1625/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.techfore.2012.07.001

In addition, results provided by different studies are affected by the time that the corresponding research took place. Due to the dynamic nature of the broadband diffusion process these results usually reflect only the present situation. Furthermore, results are depending on the geographic areas considered in each study. Thus, an obvious digital gap is identified among OECD and European Union member countries, in terms of broadband penetration rate. Despite the renew Lisbon strategy, stated in 2007 and aiming for 30% broadband penetration in EU until 2010, a digital gap among member states still exists [11,12]. Apart from the level of penetration, differences among countries include, but are not limited to, prices, availability and broadband connection speeds.

The contribution of this work is, firstly, to use a non-parametric regression in order to identify the a priori unknown nature of the relation between the dependent and the independent variable. Secondly is to assume that the influence of socio-economic factors (independent variables) over broadband diffusion (dependent variable) differs during the two main stages of the diffusion process.

Most of the studies regarding the estimation of broadband diffusion are usually based on an econometric analysis, aiming to relate diffusion with a number of external factors which are assumed to affect it. The corresponding econometric models are mainly linear in nature, incorporating either the values of the variables or their logarithms. However, this assumption does not always hold since in many cases the linear relation between the dependent and the independent variables is not the most appropriate to describe their dependence, in terms of estimation accuracy. These cases are better to be described by other functional forms, such as exponential or sinusoid. A very useful tool for identifying the nature of their relation is a non-parametric method. Non-parametric methods do not require the prior knowledge of the econometric model; the regression function is constructed during a later step of the analysis. Another benefit of the non-parametric approach is the identification of the less influential variables, which are excluded from further consideration. The procedure is known as “dimension reduction” and is performed during the early stages of evaluation of the methodology. This is in contrast to the conventional econometric method, where possible variable elimination occurs at a final step, according to the derived results.

In addition to the above, most of the analyses focus on describing the total response of the independent over the dependent variables. For this, they use the available dataset as a whole when evaluating the econometric model. Nevertheless, there is quite often the case that the data must be clustered, in order to reveal the peculiarities of particular segments and allow important information to be derived. The main assumption of this work is that the different stages of the broadband diffusion process are affected by different factors. Therefore, instead of using the whole data to describe the influence of the socio-economic variables over broadband diffusion, the data are appropriately divided into two groups and two separate analyses are consequently performed. Data segmentation is performed according to the inflection point of the diffusion process in each country. The data used in the analysis describe the European area, aiming to build a methodology which consists of two steps.

The first step corresponds to the development of a non-parametric approach aiming to study the driving factors of broadband diffusion in each stage of the process. In addition to this, dimension reduction is performed, based on sliced inverse regression — SIR [13].

During the second step, results of the influential variables are evaluated based on appropriate statistical measures. Hence, the validation of the accuracy of the proposed dimension reduction is provided, without the regression model being specified yet. In order to identify the models that better describe the diffusion process, local polynomial regression (LPR) has been used. Based on criteria from the information theory, such as the Akaike information criterion (AIC) [14] and the Bayesian information criterion (BIC) [15] the best model has been chosen.

The present analysis is based on a large dataset consisting of statistics over a number of European countries. The participating variables were considered over a prolonged period of time. In addition, the study is conducted at the macro level which means that results do not reflect specific countries' behavior but the total response. Moreover, separation of the dataset into two groups is performed based on the inflection point (IP) of the broadband diffusion process in each country. IP provides a crucial turning point in the evolution of diffusion. Before that point, the parameters affecting broadband penetration are expected to be different from those after IP, mainly because of the fact that the maturity level is changing. Besides, it seems that different regression models could be more suitable for describing, in terms of accuracy, the process during its different stages. Estimation of the IP of each country is based on the non-symmetric Gompertz model [16] and the process is described later in this paper.

Finally, in order to boost the selection of non-parametric approach, parametric regression was also conducted in this work, allowing the comparison of results. In the following sections the appropriate theoretical framework is presented, together with the development of the methodology.

2. Background

There are mainly two approaches used regarding the study of the driving factors of broadband adoption. They correspond to either regression analysis based on statistical data or surveys' outputs based on questionnaires [17].

Most of the existing studies are associated with the OECD countries, probably due to the availability of data. The main categories of derived statistics are based on social, economic, political, demographical and technological factors [18]. Such kind of data is used in a number of studies in order to measure the digital divide [19] or to rank countries according to their broadband performance [20]. Zupan [21] forecasted the maturity of electronic commerce in Slovenia and predicted a rapid growth in the following years, which is directly related to the corresponding broadband penetration rate. Furthermore, new indexes measuring broadband in terms of performance, efficiency and adoption level are suggested [22,23]. Nevertheless, indexes are also developed for policy usage, mainly taking into account liberalization and competition levels [24]. An extensive study was also performed by

Kum [25], where the effectiveness of governmental intervention in the broadband diffusion process was measured among developed and developing countries. Stanton [26] focused on the demographic factors that affect the broadband diffusion process, while in a relative study [27] three pre-specified factors were analyzed aiming to provide policy suggestions. Park and Yoon [28] showed the impact of both market and technological forces to the broadband diffusion in Korea. The study of Howick and Whalley [29] emphasized on the understanding of adoption drivers between rural and remote areas in Scotland and suggested that policy makers should enhance public view about the necessity of broadband usage in both, households and businesses. Lee and Brown [30] studied the platform competition (i.e. when similar technologies could provide similar services to end-users) and concluded that competition is positively related to fixed-broadband deployment. Modis [31] investigated the future growth of Internet and proved that the forthcoming growth is expected to be low, as most of the developed countries have reached saturation levels.

Studies derived from surveys based on questionnaires face the analysis of broadband diffusion in a similar way. Horrigan [32] performed a national survey in the United States and explained the reasons of slowing broadband adoption, based on a broadband adoption model. The Office of Technology Policy [33] conducted a survey and found that cost and content, along with convenience and confidence, are the most important driving factors accelerating broadband demand. Another survey, which was based on residential data, revealed that income and education tend to become less important in the process of broadband adoption [34]. Hulicki [35] provided an overview of broadband access driving factors, focusing mainly on the technological ones. He suggested that the digital gap is continuously shrinking and that E-services could be an important factor of the process of broadband adoption. Stern et al. [36] analyzed a large data set from the wider Australian area and conducted a research based on tree analysis. They concluded that income should not be considered as a barrier. In addition, they suggested that policy makers and users should pay attention on the content, as it seems to positively affect broadband diffusion. A recent study based on results drawn from a survey across 26 countries worldwide shows that almost 80% of adults regard Internet access as their fundamental right [37]. Finally, in a study conducted for small and medium sized enterprises (SMEs) in Ireland, it is assessed that broadband adoption is a derivative of ISPs' competition and regional market concentration [38].

It is noteworthy that the idea of dividing a diffusion process in two regimes was also introduced by Vakratsas and Kolsarici [39], who proposed a dual-market model for the diffusion of new prescription pharmaceuticals. They used a dynamic probabilistic switch in order to determine a cut-off between different patterns. This approach, although it imposes certain similarities, is not considered suitable to be applied in the case of technological markets, as they mention themselves. The dual-market hypothesis assumes that there is no link between the two regimes, i.e. their adoption processes are disconnected. In the high technology market a single process is assumed to describe the whole diffusion, which is divided into two stages that are not disconnected. These two stages correspond to two discrete segments on the diffusion curve, separated by the inflection point, which is an important event of the diffusion process, since it corresponds to the peak of the adoption rate. After this point, the adoption rate decreases, until it becomes equal to zero and the diffusion cycle comes to its end. Therefore, the inflection point can be considered as a point where penetration of the new technology has reached a significant level of maturity. Thus, it is important to examine if the factors that drive the process affect it in the same way before and after that crucial point of maturity.

3. Methodology

In this section the necessary mathematical concepts that are used in this paper are presented, providing an overview of the two main approaches of regression analysis: parametric and non-parametric. Although the non-parametric regression is adopted in this work, a parametric analysis is briefly presented as well, in order for the differences of both methodologies to be clear.

3.1. Regression analysis

Regression analysis provides useful information regarding the relationship between a dependent variable Y and a group of covariates $x=(x_1, \ldots, x_n)$, $X \in R^p$. In statistical literature there are vast studies regarding regression methodologies [40,41]. However, in the cases where the dataset consists of many variables or long data series, parametric regression could lead to misunderstandings and false conclusions, due to the high complexity of the constructed system. The general form of the considered model is presented in Eq. (1):

$$\text{BroadbandPenetration} = f\begin{pmatrix}\text{social}, & \text{economic}, & \text{political}, \\ \text{demographic}, & \text{technological}\end{pmatrix} + \varepsilon_{it}. \tag{1}$$

The physical meaning of the above Eq. (1) is that there is a response, dependent variable (Broadband Penetration in this case), which is related to a number of independent variables, through a link (or regression) function, f, which can be linear or non-linear in nature.

There are two major approaches which utilize Eq. (1), aiming to describe the system dynamics and identify the underlying relations. They are known as “parametric” and “non-parametric” regression, respectively. In many real-world problems the response variable depends nonlinearly on the explanatory variables (i.e. the link function f of Eq. (1) is non-linear). In some of these cases of non-linear regression functions, the form of the non-linear function is defined prior to the analysis and is parameterized in terms of “basis” functions. For example, polynomial regression is a form of linear regression in which the relationship between the independent and the dependent variables is modeled as an nth order polynomial and the process seeks to determine the coefficients of the polynomial. These types of regression are known as parametric regression, since they are based on models that require the estimation of a finite number of parameters.

However, there are many cases that the functional form is neither known nor can be safely assumed and thus it cannot be parameterized in terms of any basis functions. In these cases a non-parametric approach is employed in order to determine the link function.

Both approaches (parametric and non-parametric) are described below and, as mentioned before, their main difference lies in the prior knowledge or definition of the link function, *f*. More specifically, in the classical parametric approach the regression function between the dependent and the independent variables must be explicitly defined prior to the conduction of the statistical analysis. Such functions are usually described by linear, exponential, logarithmic or any other functional formulations. Linear regression, possibly with transformed variables, has been the method of choice for a long time [42]. It has the advantages that the parameters are easy to compute and easy to interpret. The main drawback is that the assumption of a linear model is a strong one. To overcome this, polynomial regression could be used, but the relationship among the variables may not be approximated well by a polynomial of low degree, and a polynomial of high degree may introduce a problem of overfitting. Transformation of variables, such as the Box–Cox transformation, is an alternative approach but may still result in lack of fit.

Generalized linear models (GLM) have gained popularity, mainly due to their flexibility to incorporate discrete variables and their feasibility. However, these models typically require the specification of the link function, in addition to either the likelihood or the variance specifications.

On the other hand, non-parametric regression overcomes these constraints, since it does not require the prior knowledge of the regression function. It mainly targets into the estimations of the coefficients of the variables and dimension reduction, without relying on the knowledge of the link function, which is constructed in a later step based on a local smoothing procedure.

In the context of this work the non-parametric approach is compared with the parametric for the needs of the analysis. The objective of using a non-parametric technique is to provide an alternative to the parametric approach, where the link functions are predefined (they are usually assumed linear). This simplified assumption makes the parametric analysis a widely used tool. Parametric methods make more assumptions than the non-parametric. If these assumptions are true the parametric methods can produce quite accurate estimates, a fact that partially explains the popularity of the method. However, if the assumptions are not correct, parametric results can be quite inaccurate. The formulation of the parametric models is usually simpler to develop and easier to compute, without the need of special statistical software, while the non-parametric approach includes calculations that require the use of appropriate software and computational effort. In addition, parametric techniques are more flexible than the non-parametric, since the former can describe a greater range of interactions between variables. Finally, non-parametric tests use less information available in the statistics than the parametric ones, where ordinal information is commonly used [43].

3.2. Parametric approach

The classical econometric method corresponds to the parametric approach introduced above. In order to evaluate the performance of the proposed methodology, a conventional regression analysis is also conducted in order to be used as a benchmark, in terms of the corresponding results. The analysis is performed by employing a number of models based usually on either linear or logarithmic functions, as Eqs. (2) and (3) show.

$$\text{Broadband Penetration}_{j,t} = \beta_0 + \sum_{i=1}^{n} \beta_i x_{j,t,i} \tag{2}$$

$$\ln\left(\text{Broadband Penetration}_{j,t}\right) = \beta_0 + \sum_{i=1}^{n} \beta_i \ln\left(x_{j,t,i}\right) \tag{3}$$

where i denotes the i-th considered parameter (out of the n), at time t and for each country, j.

The level of statistical significance is determined by the calculated variables' correlation, p-values and t-tests. In addition, the effectiveness of a model is determined by relevant statistical measures, such as R^2, mean squared error, etc.

The above formulations aim to reveal the relationship between the dependent variable, which in the studied case is broadband penetration, and the group of covariates which consist of all the considered factors. Through the analysis, the important factors are divulged, providing useful information towards the explanation of broadband Internet diffusion. Such kinds of studies are conducted seeking to reveal the differences of either among the countries or within the boundaries of a single country. The differences in social, economic, political etc. background could lead to different conclusions regarding broadband proliferation, based on the factors taken under consideration each time.

Finally, the logarithmic transformation of the data is adopted in many cases, in order to reduce the dispersion among statistical data and improve interpretability, as some of them are expressed in percentage (e.g. broadband penetration) and some others in absolute numbers (e.g. population). These approaches are based on the hypothesis that all considered variables contain useful information. However, it is quite possible that the precision and the accuracy of results deteriorate, due to the number of irrelevant regressors included as part of the analysis.

3.3. Non-parametric approach

The main advantage of non-parametric regression and dimension reduction is related to the fact that among a number of considered parameters, the most crucial influential factors are revealed during the early stage of the analysis and thus the irrelevant ones are excluded from further consideration. This reliefs the complexity of the system and allows to study the factors that have an important effect over the process. Moreover, the link function between response and the independent variables is not subjectively predetermined but is derived by adopting a theoretical approach, such as sliced inverse regression. In this way, the bias of explicitly setting the link function is suppressed, as the latter is determined following an appropriate statistical analysis.

Nonparametric regression analysis traces the dependence of a response variable Y on one or several predictors X_i without specifying in advance the function that relates the response to the predictors, which is constructed based on the information carried by the data. It is distinguished from linear regression, in which the function relating Y to the X_is is linear in the parameters

$$Y = b_0^{'} + b_1^{'}X_1 + \ldots + b_k^{'}X_k + \varepsilon$$

and from traditional nonlinear regression, in which the function, f, which relates Y to the X_is, though nonlinearly in terms of its parameters, is explicitly specified:

$$Y = f\left(X_1, \ldots, X_k; b_1^{'}, \ldots, b_k^{'}\right).$$

In traditional regression analysis, the objective is the estimation of the model's parameters, whereas in nonparametric regression the objective is the estimation of the parameters and the form of the regression function, f. In addition, non-parametric analysis seeks to achieve dimension reduction, by excluding some of the examined variables, without losing information.

Typically, non-parametric analysis consists of the following discrete steps:

- Sliced inverse regression (SIR), in order to achieve dimension reduction
- Local polynomial regression (LPR) with a Kernel smoothing, in order to derive information regarding the shape of the regression function f
- Application of appropriate information criteria, such as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), for selecting the most appropriate function, among the pool of the candidate ones, that best describes the data.

The above introduced steps of non-parametric analysis are described in the following paragraphs.

3.3.1. Sliced inverse regression — SIR

SIR is based on the framework proposed in [44] who considered the following model:

$$Y = f\left(b_1^{'}X_1, \ldots, b_k^{'}X_k, \varepsilon\right) \tag{4}$$

where f represents the unknown link function and ε the error, which is assumed to be independent of the X_is. The main advantage of this methodology is its ability to estimate the direction vector of the parameter b' without knowing the shape of f, which is estimated at the next step.

The SIR algorithm is designed to determine the k-dimensional subspace that includes all the necessary information regarding the link function between Y and the covariate X_i. The k-dimensional space spanned by the k vectors is called the effective dimension reduction (e.d.r.) space and any vector in this space is referred to as an e.d.r. direction. The primary goal of this approach is to estimate the e.d.r. directions so that one can plot Y against the e.d.r. variates for the visual exploration of the regression structure. This would allow the effective application of various low-dimensional regression techniques to the reduced space.

A typical algorithm implementing the SIR is described below:

- Let $(y1, x1), \ldots, (yn, xn)$ be the original data set with $(p+1)$ variables and n observations. Table 1 provides an example of such a dataset.
- Sort the data by Y.

Table 1
Sliced inverse regression — SIR.

Y_1	$x_1(=(x_{11}, x_{12,}\ldots\ldots, x_{1p})')$
Y_2	$X_2(=(x_{21}, x_{22,}\ldots\ldots, x_{2p})')$
Y_3	$X_3(=(x_{31}, x_{32,}\ldots\ldots, x_{3\mathrm{p}})')$
Y_4	$X_4(=(x_{41}, x_{42,}\ldots\ldots, x_{4\mathrm{p}})')$
Y_5	………………
…	…………………
Y_n	$x_n(=(x_{n1}, x_{n2}, \ldots \ldots, x_{np})')$

- Divide the data set into H slices as equally as possible. Let n_h be the number of cases in slice h.
- Within each slice, compute the sample mean of x, $\bar{x}_h = n_h^{-1} \sum_{(i) \in \text{slice } h} x_{(i)}$.
- Compute the covariance matrix for the slice means of x, weighted by the slice sizes: $\hat{\Sigma}_n = n^{-1} \sum_{h=1}^{H} n_h (\bar{x}_h - \bar{x})(\bar{x}_h - \bar{x})'$. In the last equation $\bar{x}$ is the mean of the whole sample $\bar{x} = n^{-1} \sum_{i=1}^{n} x_i$.
- Compute the sample covariance for xi's $\hat{\Sigma}_x = n^{-1} \sum_{i=1}^{n} n_h (\bar{x}_i - \bar{x})(\bar{x}_h - \bar{x})'$.

■ Find the SIR directions by conducting the eigenvalue decomposition of $\hat{\Sigma}_n$ with respect to $\hat{\Sigma}_x$: $\hat{\Sigma}_n \hat{\beta}_i = \hat{\lambda}_i \hat{\Sigma}_x \hat{\beta}_i \hat{\lambda}_i \geq \hat{\lambda}_2 \geq \ldots\ldots \geq \hat{\lambda}_p$. The i-th eigenvector $\hat{\beta}_i$ is called the i-th SIR direction. The first few SIR directions can be used for dimension reduction.

■ Project x along the SIR directions; that is, use each SIR direction to form a linear combination of x. $\hat{\beta}'_1 x$ is called the first SIR variate, $\hat{\beta}'_2 x$ the second SIR variate and so on.

■ Plot of Y against the SIR variates. These 2-D or 3-D plots offer a graphical summary useful for revealing the regression structure.

3.3.2. Local polynomial regression — LPR

LPR is a generalization of local mean smoothing. Instead of fitting a local mean, one fits a local pth-order polynomial. Calculations for local polynomial regression are naturally more complex than those for local means, but local polynomial smoothing methods have better statistical properties. The basic idea is to partially approximate the dependent variable Y_i, using a specific bandwidth and based on a formulation as the following:

$$Y_i = a + f_1\left(x_{i,1}\right) + f_2\left(x_{i,2}\right), \ldots. + f_k\left(x_{i,k}\right) + \varepsilon_i. \tag{5}$$

Kernel-based methods are most popular non-parametric estimators, since they can uncover structural features in the data which a parametric approach might not reveal [45], and consists of applying a function, known as the kernel, to each data point in the time-series. The approach is nonparametric, in the sense that it does not assume a fixed structure associated with the data. Kernel smoothing belongs to the class of weighted moving averages, which means that all the points in the time-series are weighted using as weights the results of the computation of the Kernel function. Every Kernel function has the following properties:

- All kernel values are positive or zero.
- The kernel functions are normally symmetric.
- Kernel function values decrease to zero from a central (maximum) value.

Among the most widely used Kernel functions are the normal (Gaussian) and the Epanechnicov, which are applied in the evaluation of the proposed methodology.

The importance of using smoothing techniques is that statistical analyses often aim to generalize small, individual datasets to large, general populations. Therefore, it is often desirable to “smooth” the data, changing a rigid histogram into a curve. Nonparametric kernel smoothing does exactly this by applying a kernel function to the dataset. Plotting of the regression (smoothed) line of the density function will allow for the inspection of how the density estimation transformed the shape of the original data. As a next stage, the regression line is used to determine a pool of candidate functions that could fit the data. The selection of the most appropriate function is based on an information criterion, such as AIC and BIC (Eqs. (6) and (7), respectively) [15,46], applied over the candidate functions.

$$\mathrm{AIC} = 2k + n[\ln(RSS/n)] \tag{6}$$

$$\mathrm{BIC} = \ln\left(\sigma_\varepsilon^2\right) + \frac{k}{n}\ln(n) \tag{7}$$

The model that is finally chosen to describe the process is the one that has the lowest value in each of the above criteria. The performance of Kernel smoothing is evaluated based on appropriate statistical measures, such as the mean integrated squared error (MISE) or the asymptotic mean integrated squared error (AMISE).

4. Steps of the analysis

In accordance to the concepts presented in the preceding sections, the following paragraphs describe the steps of the analysis performed. Thus, the problem faced in this study is the influence of social, economic and other parameters over the diffusion of broadband connections, described by data from the wider European area.

The discrete parts of the analysis are the following:

- Determination of the relevant social, economic and other parameters and data collection for each considered country, from Eurostat's database.
- Estimation of the IP of each country is based on the non-symmetric Gompertz model according to the following procedure: based on the existing diffusion data, for each country, the Gompertz model [16] was trained, using non-linear squares (NLS). This procedure concluded with the estimation of the diffusion parameters, i.e. the diffusion rate and the saturation level, for

each country. Following this, the inflection point is calculated at the 37.79% of the estimated saturation level, as derived from the formulation of the model.

- Division of the data into two groups, based on the estimated inflection point of the diffusion process, for each country, according to the procedure described in the previous step. Diffusion data that have lower values than the inflection point are collected into the first group, while the rest are collected into the second. In addition, there is a third group which consists of the whole dataset which is also analyzed, mainly for comparison reasons.
- Application of the sliced inverse regression (SIR) algorithm for all three groups. This will result into the identification of the most influential parameters and the exclusion of the less relevant ones, leading to dimension reduction with no loss of information.
- Application of a Kernel smoothing technique, in order to reveal the behavior of the dependent against the independent variables. This will serve as a driver for the construction of the pool of the candidate link functions.
- Application of the AIC and the BIC over all of the candidate functions, in order to conclude to the link function that best describes the modeled system, in terms of appropriate statistical measures.

All the above steps are applied and discussed in the following subsections.

4.1. Selection of parameters and corresponding data

During this step of the analysis the socio-economic factors assumed to affect broadband diffusion are identified and data were extracted from Eurostat's database [47]. These statistics correspond to the national total of social, economic, demographic and political factors available. They were also cross-checked with subscription data published by the OECD [48]. Descriptive statistics of the datasets used are presented in Appendix A.

The selection of the influential parameters to be included in the model was based on the existing results of the relevant literature and some conceptual arguments. Economic factors, such as *GDP* per capita and *Price*, are considered in the analysis, as they have been proven among others in [25,33,34], to promote the adoption of new technologies. *GDP* per capita is expressed on a logarithmic basis in order to eliminate deviation among values. Income inequality (*Ineq*) is also considered in order to better describe the impact derived from economic factors. *Ineq* can describe the effect of income on broadband diffusion better than *GDP* per capita as it reflects the distribution of income and not just an average depending solely on population. Governmental intervention seems to influence broadband penetration as well [5,18,25]. The impact of public practices can be reflected by E-government online availability (*E_gov*), information technology expenditure (*IT_exp*) and communications expenditure (*Com_exp*). Although *IT_exp* and *Com_exp* are directly related with the above intervention, they are also referred to technological factors. One more technological factor, i.e. using PCs connected to the Internet (*Us_PC*), is considered in the analysis, which is also linked to economic issues. According to other studies [26,27], demographic factors including *Age* and population density (*Dens*) seem to influence broadband diffusion. In addition, education level, such as *School* and tertiary graduates in science and technology (*S&T*) are related with the development of broadband penetration [32]. Furthermore, the effective use of new technologies requires the acquisition of certain skills, i.e. Internet skills (*Int_sk)* and computer skills (*PC_sk*) [49]. Finally, the technological maturity of a country, as described by the corresponding factors of human resources in science and technology (*HR*) and E-commerce (*E_com*), reveals its relevance with new technologies [21,28,50].

The broadband penetration, expressed in percentages, describes the number of dedicated, high-speed connections (with a capacity equal to or higher than 144 kbits/s) per 100 inhabitants. Various technologies are covered; ADSL, cable modem as well as other types of access lines.

Table 2 illustrates the factors that have been considered in the context of the present study, with a short description of them.

Table 2
Parameters and short description.

Parameters	Description
GDP	Gross domestic product per capita (in €)
HR	Human resources in science and technology as a share of labor force
E_gov	E-government online availability — percentage of online availability of 20 basic public services
S&T	Tertiary graduates in science and technology per 1000 of population aged 20–29 years
Us_PC	Persons employed using computers connected to the Internet
Ineq	Inequality of income distribution — income quintile share ratio
Age	Proportion of population aged 25–49
School	School expectancy — expected years of education over a lifetime
E_com	Individuals using the Internet for ordering goods or services — percentage of individuals aged 16 to 74
IT_exp	Information technology expenditure as a percentage of GDP
PC_sk	Individuals' level of computer skills — percentage of the total number of individuals aged 16 to 74
Int_sk	Individuals' level of Internet skills — percentage of the total number of individuals aged 16 to 74
Price	Price of telecommunications — local calls (10 min)
Com_exp	Communications expenditure as a percentage of GDP
Dens	Population density

It should be stated that the country-level panel data correspond to the total, average response of each country and they do not take into explicit account the possible country specific heterogeneity. This would indeed be an interesting future extension of the proposed approach, provided that the appropriate data are available.

Following that, a correlation analysis was performed over the data in order to conclude the factors that should be considered in the econometric analysis. Multicollinearity was tested by performing bivariate correlations and specifically two-tailed Pearson analysis. The threshold was set to a value of about 75% correlation and variables showing a correlation above that value were excluded from further consideration. Despite the fact that they could carry information regarding the diffusion process, they present a high level of correlation with other independent variables and had to be excluded, in order to avoid redundancy and biasness. More specifically, these variables are human resources in science and technology (*HR*), individuals using the Internet for ordering goods or services (*E_com*) and information technology expenditures (*IT_exp*).

The dataset, that the analysis was based on, consists of cross-sectional, time-series data, over 26 European countries. More specifically, semiannual data were used, starting from the second semester of year 2001 up to December 2009, for all countries. Before the analysis a first-order autocorrelation test was conducted for the whole dataset and results indicated a strong autocorrelation of 0.820. For this reason and in order to remedy autocorrelation effects, the second-order autocorrelation test was conducted. The corresponding value was 0.636, which was considered borderline acceptable. Therefore, in addition to this and in order to remove any possible autocorrelation effects, the regression was performed after a dataset transformation, based on the generalized differences [51]. Thus, the final dataset consists of annual transformed data on which autocorrelation is very low (0.237).

4.2. Diffusion process stages and data separation

The next assumption was to consider the influential effect of the factors that result to a diffusion process which does not remain constant throughout its whole cycle but changes during its different stages. In this context, certain points of time, which correspond to crucial turning points of the diffusion process, have a significant importance in the analysis; one of these being the inflection point. This is the time when the market potential (the adoptions per time unit) has reached its maximum value and from that point on is expected to decline. A common approach in similar studies is to consider the available data and factors as a total, resulting into identifying their influence over the diffusion process as a whole. However, this is not appropriate for identifying the effects of the factors during the different stages of the diffusion, especially before and after the inflection point. The point of inflection is quite important and it is related to different policies and managerial decisions that should be considered. For example, Netherlands reached IP in 2004 (Q4) when the penetration rate was 14.7%, while the respective rate for Norway was 15.7% in 2005 (Q4). On the contrary, it is estimated that Greece and Latvia reached their IP in 2008 (Q2) and 2007 (Q2) respectively when the corresponding penetration rates were 9% and 9.2%.

Thus, in order to identify the driving factors of broadband diffusion during these two discrete stages of the process, data were split into two parts, corresponding to the stage before and after the inflection point, respectively. Two analyses were consequently performed, based on these two datasets and for comparison reasons the whole dataset was also analyzed.

4.3. SIR and dimension reduction

This step corresponds to the application of the sliced inverse regression (SIR) methodology for estimating of the model's parameters and performing dimension reduction.

Application of the SIR algorithm described in the previous sections resulted into the calculation of the system's corresponding eigenvalues and eigenvectors. None of the eigenvalues were equal to zero, although only the first ones turned out to be the most significant.

After estimating the SIR variates, the t-ratios for each of the considered factors are calculated [52]. The SIR variates are suitably transformed according to the estimated eigenvectors and dimension reduction is performed based on the first, most important, eigenvalues, which are presented in Table 3.

4.4. Local polynomial regression and information criteria

After applying the SIR, in order to identify the most important factors by performing dimension reduction, local polynomial regression with a kernel smoothing is conducted, seeking to identify the form of the link function. As bandwidth selection bears

Table 3
Eigenvalues for the three datasets.

	Eigenvalues	
	1st	2nd
Whole	0.754	0.273
After IP	0.757	0.419
Before IP	0.713	0.196

danger of over or under smoothing results, different bandwidths were tested in order to perform Gaussian and Epanechnikov Kernel smoothing analyses. The identification of the regression function is based on the last step of the previously introduced SIR algorithm and it corresponds to the plotting of Y against the SIR variates, seeking for a graphical summary, useful for revealing the regression structure.

According to the regression results, certain families of functions can be determined by the visual inspection of the smoothed line. It is often the case, that different functions can describe the diffusion process more accurately in its various stages. Thus, it would not be appropriate to arbitrarily select a function to describe the whole process, since this would probably lead to inaccurate results. In order to conclude to the most accurate candidate function, the values of the expressions of AIC and BIC are calculated. The analysis is conducted for the first two eigenvalues, although the results of the larger eigenvalues for the three datasets (whole, after IP and before IP) are only presented. This is because both the AIC and the BIC calculations based on the second best eigenvalues assessed the less significant outputs.

5. Results

This section is devoted to the presentation of the results, which were derived following the application of the methodology described above. According to these results, the exponential model provides a better fit in almost all of the evaluated cases. Therefore it provides a better description of the examined system, than the linear function that would be constructed following a conventional econometric analysis.

Corresponding results are presented in Table 4, for the three datasets (the whole dataset — whole, after the inflection point — after IP and before the inflection point — before IP). The first line of each variable contains the value of the parameter and the second (in the parenthesis) the value of the t-ratio. Numbers in bold indicate that the corresponding parameter turned out to be statistically important, based on the calculation of the t-ratio (at the 95% level of confidence).

The highlighted factors that are above the critical value of t-ratio (1.96) at the 95% level of confidence are statistically significant. Depending on the dataset they are four or six. However, the determination of the significance of factors based solely on the t-ratios could lead to the omission of influential factors, due to their statistical insignificance. Therefore, in order to accurately identify the significant determinants, the set of considered factors are sorted based on absolute values of t-ratios and this ranking allows the consideration of twelve candidate models. For each of these models application of sliced inverse regression led to the estimates and the link function $g_k(k=1, \ldots 12)$ was then obtained by applying the local polynomial regression. The k-th model is made up of the factors with the k higher t-ratios. Eventually the model with the lowest AIC_C is chosen as the ultimate solution [52,53].

As mentioned earlier, the construction of the pool of candidate link functions is based on the provided regression line. It is apparent that a link function, based on the plots of the SIR variants (the linear combination of the parameter values multiplied by

Table 4
SIR results for the three datasets.

	Whole	After IP	Before IP
GDP	−0.275	−0.298	−0.229
	(−1.242)	(−0.885)	(−0.650)
E_gov	**0.022*****	**0.015*****	**0.026*****
	(8.699)	**(3.794)**	**(5.170)**
S&T	−0.019	−0.019	**−0.042***
	(−1.637)	(−1.088)	**(−2.372)**
Us_PC	**0.052*****	**0.076*****	**0.058*****
	(5.577)	**(5.913)**	**(3.525)**
Ineq	**0.108***	0.048	0.130
	(2.339)	(0.610)	(1.804)
Age	−0.036	−0.054	−0.075
	(−1.004)	(−1.190)	(−0.975)
School	−0.032	−0.041	−0.138
	(−0.645)	(−0.595)	(−1.384)
PC_sk	0.017	−0.021	0.004
	(1.199)	(−0.727)	(0.211)
Int_sk	0.006	−0.017	**0.042****
	(0.723)	(−1.116)	**(2.947)**
Price	−0.118	−0.118	−1.003
	(−0.293)	(−0.187)	(−1.543)
Com_exp	−0.161	**0.320***	**−0.532***
	(−1.497)	**(2.148)**	**(−2.587)**
Dens	**0.001*****	**0.001****	**0.002****
	(3.688)	**(2.793)**	**(2.963)**

t-Statistics are in parentheses.
* $p<0.05$.
** $p<0.01$.
*** $p<0.001$.

their corresponding weights), could be arbitrarily considered, which would sometimes indeed coincide with the link function selected based on Kernel smoothing and local polynomial regression. However, the arbitrary selection of the link function could impose objectivity and bias and should be avoided. Despite that in the present case study the selected function may be easily identified by visual inspection of the plots of SIR variants, this is not always the case.

Using the first, most significant, eigenvalue, the plots of the actual broadband penetration against the SIR variants are created. Application of local polynomial regression resulted in a smoothed line which implies two families of link functions (among the families of simple functions) that could describe broadband diffusion better: the linear and the exponential.

The lowest AIC_C values emerged for the first seven and eight parameters for the after-IP and before-IP datasets, respectively, with the exponential link function. Thus, it can be assumed that all these variables are important for broadband penetration and despite the statistical insignificance are worth to be included in the analysis. Results illustrated in the figures below (Figs. 1–4) refer to the models with the lowest AIC_C. More specifically, for the after-IP dataset, the parameters considered, in order of importance, are *Us_PC*, *E_gov*, *Dens*, *Com_exp*, *Age*, *Int_sk* and *S&T*. The corresponding parameters for the before-IP dataset are *E_gov*, *Us_PC*, *Dens*, *Int_sk*, *Com_exp*, *S&T*, *Ineq* and *Price*.

In order to conclude the link function that better describes the data, the AIC and BIC criteria for the linear and the exponential functions are calculated and results are presented in Table 5.

It seems that the exponential model has the lowest values, in both criteria (AIC and BIC). Therefore, it is considered as the most appropriate function to describe the broadband diffusion process in the majority of the cases.

According to the obtained results it is obvious that the estimated influential factors vary, depending on the maturity level of broadband penetration in both the order of importance and in the sign. There are three factors that must be certainly regarded as facilitators in broadband diffusion. More precisely, these are E-government online availability (*E_gov*), employees using computers connected to the Internet (*Us_PC*) and population density (*Dens*). Despite the fact that income inequality (*Ineq*) seems to be a driving factor for the whole dataset, further analysis over the before and after the inflection point datasets, revealed that *Ineq* turned out to be an influential parameter only for the first stage of broadband diffusion i.e. before-IP.

The results corresponding to the data after the inflection point showed that *Us_PC*, *E_gov*, *Dens* and *Com_exp* are the dominant factors in the diffusion of broadband services. All these parameters have a positive relationship with broadband penetration.

Us_PC seems to enable further use of broadband services. In fact, all parties involved in an economy, such as government, businesses and individuals, could incorporate the opportunities derived from broadband use and enhance their comparative advantages by exploring the benefits offered by new technologies. In addition, governmental practices should have a direct application to the use of new technologies and for this reason E-government online availability must be constantly updated. Population density (*Dens*) turned out to be significant in the process of further broadband diffusion. In urban areas both the easier development of infrastructures promoting broadband access to inhabitants and the network effects are the main reasons for a

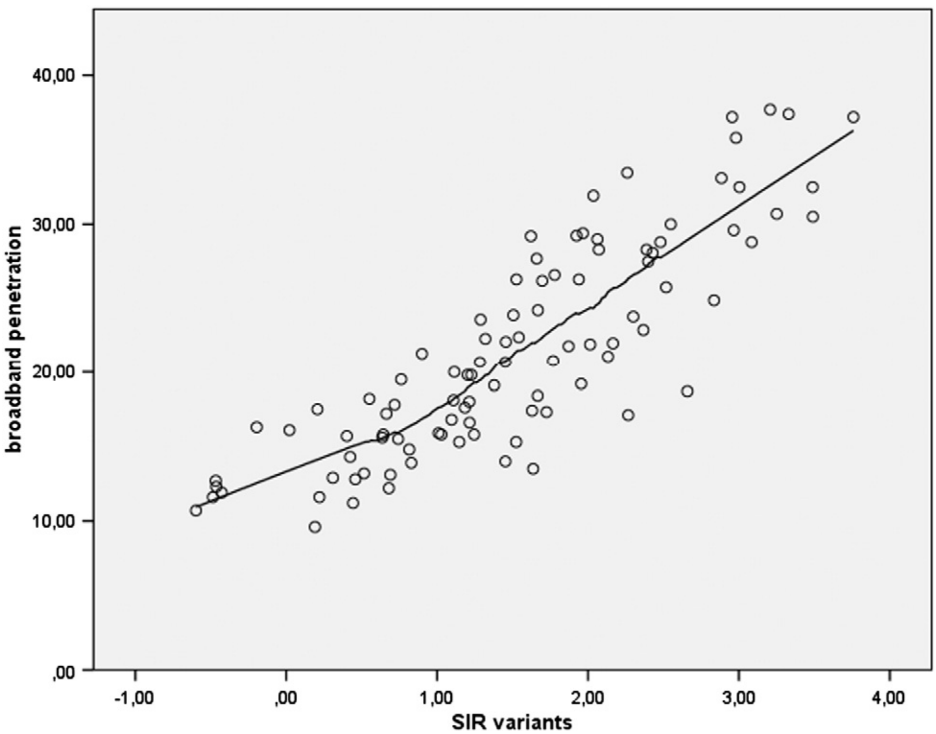

Fig. 1. Gaussian Kernel smoothing with bandwidth 0.8 — after IP.

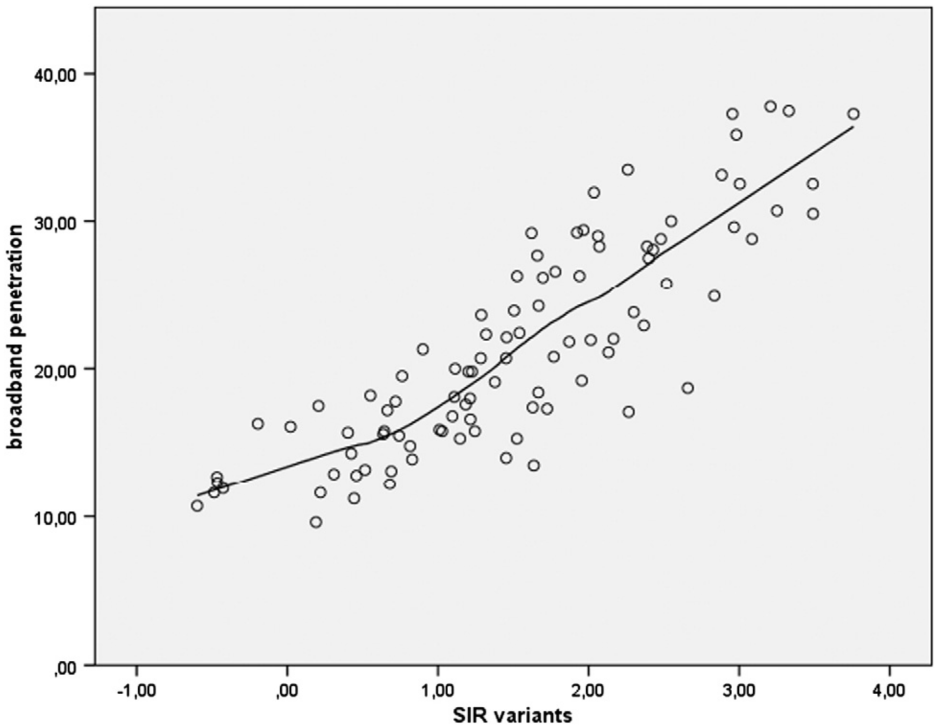

Fig. 2. Epanechnikov Kernel smoothing with bandwidth 0.8 — after IP.

higher level of broadband diffusion. Moreover, *Com_exp* seems to enhance broadband penetration as it referred to investments made for the development of the communications sector which is in turn directly linked to the broadband diffusion. Such kind of expenditures could be the deployment of networks or the subsidization of telecommunications services.

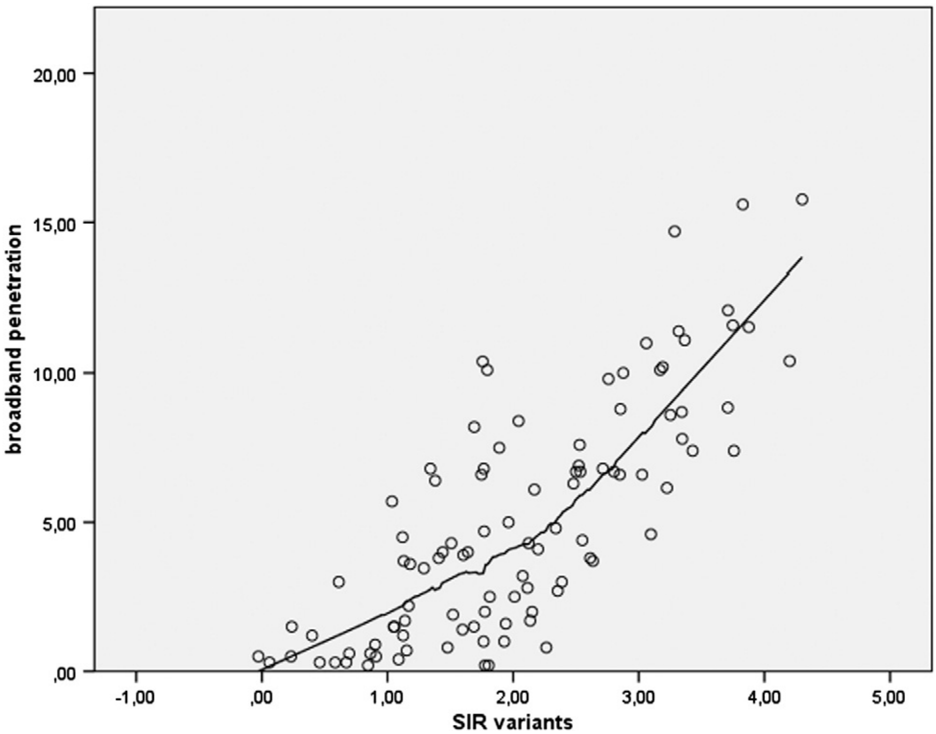

Fig. 3. Gaussian Kernel smoothing with bandwidth 0.8 — before IP.

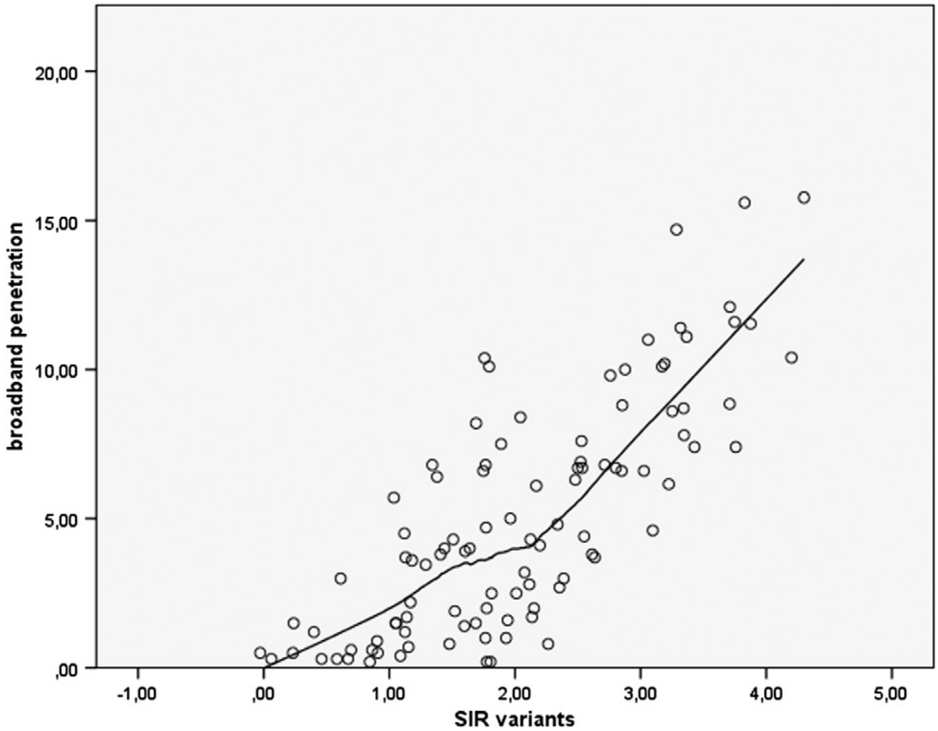

Fig. 4. Epanechnikov Kernel smoothing with bandwidth 0.8 — before IP.

It should be also noted that the coefficient signs carry important information regarding the considered determinants. For example, the variable *Age*, which according to the above analysis seems to influence the process of further diffusion, is negatively related with broadband penetration. It is assumed that the age group 24–45 years belongs to the most potential users of broadband services. However, as the broadband penetration level among that group of age increases, the pool of potential broadband users is running out. Additionally, Internet skills (*Int_sk*) showed a negative sign as well. This may imply that after the inflection point, these skills are obsolete or taken for granted, indicating a certain level of technological maturity in a society. Finally, the sign of the considered proportion of graduates in science and technology (*S&T*) remains always negative. However, the negative influence on broadband penetration after-IP was limited according to the value of the coefficient.

The before the inflection point results, indicate that there are five facilitators regarding broadband diffusion and three barriers. These factors are *E_gov*, *Us_PC*, *Dens*, *Int_sk* and *Ineq* as facilitators and *Com_exp*, *S&T* and *Price* as barriers. Apart from the three common factors (*E_gov*, *Us_PC* and *Dens*) *Int_sk* are positively related with the broadband diffusion in the dataset before the inflection point. It seems that, as the broadband penetration level exceeds the inflection point, skills related to the diffusion of broadband services have reached a desirable or required level as well. For this reason, government interventions should focus on the enhancement of digital literacy such as internet skills. Thus, policy makers can give incentives to further use of new technologies, through the educational system which is mainly determined by them. According to the results, income inequality facilitates the adoption of broadband services in the first stage of diffusion. Not surprisingly, *GDP* per capita cannot be assumed as a driving factor as it reflects the average income of the population. On the contrary, *Ineq* refers to income distribution and concentration which means that broadband services could be prohibitive during the first stage of broadband development.

Moreover, the rapid growth of technology caused an increase in graduates in science and technology. However, in the years that correspond to the dataset before-IP (approximately from 2001 to 2004) graduates cannot exceed market needs and this is the reason why *S&T* has a negative effect on the diffusion process. Moreover, *Com_exp* has a negative influence on broadband

Table 5
Results of AIC and BIC criteria.

	Non-parametric results			
	After IP		Before IP	
	Linear	Exponential	Linear	Exponential
AIC	267.012	263.975	184.433	180.113
BIC	2.989	2.957	1.994	1.950

penetration process as well. In other words, few investments during the early stages of broadband development are an obstacle for rapid diffusion. In accordance with this, in the after-IP dataset, *Com_exp* has a positive influence in the diffusion of broadband services. Finally, *Price* seems to be an obstacle for the early adoption of broadband services. This is mainly due to the higher prices of such kind of services in the first years of launch, before the development of both the technology and competition led to lower cost.

In order to verify the results provided by the non-parametric analysis and enhance the selection of exponential approach, Table 6 presents a comparison between actual and estimated broadband penetration, for some of the considered countries.

As observed, the exponential function can provide more accurate estimates of the actual values, as compared with the estimated values provided by the linear model. However, in some cases the linear approach fits better to the estimation of broadband penetration i.e. in the case of the Czech Republic.

5.1. Comparison with parametric approach

As mentioned above, a parametric analysis is also conducted in order to compare the results with those provided by the non-parametric method, as Table 7 shows.

According to the results, it is important to note that the order of the significant determinants was slightly different between the two approaches. The importance lies into the fact that the most significant determinants define and initiate appropriate policies, seeking to enhance the diffusion of broadband services. In addition, there were marginal changes to some signs e.g. *Price* for the after-IP dataset.

The comparison was based on appropriate statistical measures such as the coefficient of multiple determinations (R^2), the adjusted coefficient of multiple determinations (Ra^2), the residual sum of squares (RSS) and the standard error of the estimates (s.e.). Therefore, the statistical measures referring to both parametric and non-parametric approaches are presented in Table 8.

Comparison of the above results shows that all the statistical measures associated with the non-parametric approach are better, in all cases, than those of the parametric. However, it must be noted that the parametric method provided quite acceptable outputs despite the fact that the non-parametric approach seems to fit the data more accurately.

6. Conclusions and future research

The work presented in this paper aims to identify the functional formulation that describes the relationship between the diffusion of broadband technology and its influential factors. It fills the corresponding gap in literature, where linear econometric models are constructed and used without taking into account the different stages of the diffusion process.

The approach adopted in this work relied on non-parametric techniques. Based on the assumption that the link function between the dependent variable (broadband diffusion) and the independent ones is unknown, non-parametric regression resulted into the construction of the function, performing on the same time dimension reduction.

Based on the assumption that the different stages of diffusion are influenced by different factors, the data used for the analysis were split into two segments, before and after the inflection point of the diffusion process, for each considered country. The analysis revealed that, as initially assumed, the factors that influence diffusion are partially different, during the different stages of the process.

Table 6
Comparison between actual and estimated broadband penetration.

	Actual BB	Estimated BB	
		Linear	Exponential
After IP			
Austria	21.8	23.61	22.79
Czech Republic	15.8	16.11	16.16
Netherlands	22.4	23.61	22.79
Norway	25.7	27.57	27.31
Slovenia	19.1	20.60	19.85
Spain	16.8	18.88	18.34
Sweden	32.5	30.53	31.29
Before IP			
Austria	11.6	10.07	10.73
Czech Republic	4.3	3.57	3.34
Netherlands	6.3	7.21	6.43
Norway	11.5	10.76	12.15
Slovenia	7.8	8.69	8.38
Spain	4.3	5.05	4.36
Sweden	12.1	10.74	12.10

Table 7
Results of parametric approach.

	Whole	After IP	Before IP
GDP	−2.191	−2.951	−0.105
	(−1.014)	(−1.274)	(−0.070)
E_gov	**0.195*****	**0.073***	**0.078*****
	(7.618)	**(2.586)**	**(4.178)**
S&T	−0.169	−0.180	−0.115
	(−1.461)	(−1.429)	(−1.728)
Us_PC	**0.445*****	**0.559*****	**0.164****
	(4.824)	**(6.326)**	**(2.654)**
Ineq	0.630	0.500	0.446
	(1.396)	(0.912)	(1.659)
Age	**−0.903***	−0.521	−0.364
	(−2.549)	(−1.650)	(−1.256)
School	−0.814	−0.282	−0.594
	(−1.651)	(−0.591)	(−1.591)
PC_sk	0.188	−0.147	0.051
	(1.306)	(−0.720)	(0.664)
Int_sk	−0.026	−0.130	**0.141***
	(−0.281)	(−1.238)	**(2.616)**
Price	−0.048	0.444	−3.503
	(−0.012)	(0.102)	(−1.442)
Com_exp	−0.377	1.924	**−1.543***
	(−0.358)	(1.876)	**(−2.005)**
Dens	**0.015*****	**0.011****	**0.009****
	(3.582)	**(2.852)**	**(2.974)**

t-Statistics are in parentheses.
* $p<0.05$.
** $p<0.01$.
*** $p<0.001$.

According to the results, it is proved that governmental actions such as E-government online availability and communication expenditures can enhance the process of broadband diffusion. The level of *Com_exp* could exert either positive or negative influence depending on the level of cumulative communications investments made in the past. Moreover, the proportion of persons employed using computers connected to the Internet (*Us_PC*) is a driving factor for the proliferation of broadband diffusion for all three groups of considered datasets. It seems that penetration in the private sector is directly and positively related to the broadband diffusion over households. Additionally, population density (*Dens*) can be considered as a supporting parameter in the process, as urbanization facilitates the adoption of broadband services.

In the dataset describing the process before the inflection point, Internet skills (*Int_sk*) are added to the influential factors with a positive sign regarding broadband adoption. Thus, policies targeting to motivate people gaining such kind of skills could accelerate broadband diffusion. The results indicate that governmental intervention regardless of the level of penetration rate can promote broadband diffusion. It is also concluded that the most suitable regression model can be different based on the maturity level of broadband diffusion.

The present analysis can be further extended by taking into consideration a larger dataset of influential factors. They could include, among others, weather, races, and technological skills to be considered in future research. In this case, dimension reduction is necessary, as it can ensure the limitation of complexity due to the fewer conclusion parameters.

The understanding of the barriers as much as the driving factors in the process of broadband demand enhancement is very useful for policy interventions, not only at EU but also at the national level. Through this analysis the most and less important socio-economic factors are revealed. This study could reveal regional results if it was carried out for each considered country independently, taking into account the heterogeneity of each country. This could be achieved by performing a multilevel analysis, by means of a hierarchical model or structural equation modeling.

Table 8
Comparison between parametric and non-parametric results.

	After IP		Before IP	
	Parametric	Non-parametric	Parametric	Non-parametric
R^2	0.7331	0.7363	0.6211	0.6335
Ra^2	0.7118	0.7335	0.6009	0.6298
RSS	1313	1297	566	548
S.e.	3.863	3.715	2.455	2.364

Furthermore, a stricter correlation analysis based on multicollinearity should be performed, in order to exclude the variables which add in the complexity of the model. In this study the notion of network effects is not considered, though broadband subscription may positively correlate with previous ones, e.g. mobile or dial-up subscription [54].

Appendix A. Data descriptive statistics

Table 9
Descriptive statistics for data after the broadband inflection point.

	N	Minimum	Maximum	Mean	Std. deviation
BB	96	9.60	37.70	21.44	7.19
GDP		9.25	10.78	10.09	0.31
E_gov		30.00	100.00	67.20	17.25
S&T		6.40	29.10	13.59	4.40
Us_PC		19.00	68.00	41.01	11.67
Ineq		3.30	7.80	4.51	1.08
Age		32.50	40.60	35.97	1.75
School		15.90	20.80	17.95	1.28
PC_sk		8.00	23.00	13.57	3.26
Int_sk		14.00	52.00	32.68	9.94
Price		0.16	0.76	0.37	0.13
Com_exp		1.00	4.20	2.27	0.75
Dens		2.90	488.30	130.43	123.62

Table 10
Descriptive statistics for data before the broadband inflection point.

	N	Minimum	Maximum	Mean	Std. deviation
BB	100	0.20	15.77	5.00	3.88
GDP		8.78	10.59	9.76	0.45
E_gov		3.00	77.50	37.57	19.16
S&T		4.80	24.50	11.33	4.50
Us_PC		5.00	56.00	28.44	12.46
Ineq		3.00	7.90	4.74	1.27
Age		33.60	40.00	36.47	1.28
School		14.90	20.40	17.36	1.21
PC_sk		1.00	23.00	11.47	4.74
Int_sk		7.00	47.00	25.06	10.18
Price		0.16	0.74	0.36	0.12
Com_exp		1.00	3.20	1.88	0.67
Dens		2.90	481.90	109.80	92.97

References

[1] ITU, Digital.life, International Telecommunication Union, Geneva, 2006. Available: http://www.itu.int/ITU-D/ict/.
[2] J.Z. Cheng, H.-C. Yu, W.D. Sincoskie, Meeting the broadband access infrastructure demands: the promise of Gigabit Ethernet, Technol. Forecast. Soc. Change 72 (2005) 1–10.
[3] European Commission, i2010 — a European information society for growth and employment. Available: http://ec.europa.eu/information_society/eeurope/i2010/index_en.htm 2005.
[4] F. Höffler, Cost and benefits from infrastructure competition. Estimating welfare effects from broadband access competition, Telecommun. Policy 31 (2007) 401–418.
[5] K. Flamm, A. Chaudhuri, An analysis of the determinants of broadband access, Telecommun. Policy 31 (2007) 312–326.
[6] C. Michalakelis, G. Dede, D. Varoutas, T. Sphicopoulos, Impact of cross-national diffusion process in telecommunications demand forecasting, Telecommun. Syst. 39 (Sep 2008) 51–60.
[7] A. Robertson, D. Soopramanien, R. Fildes, Segmental new-product diffusion of residential broadband services, Telecommun. Policy 31 (2007) 265–275.
[8] L. Andrés, D. Cuberes, M. Diouf, T. Serebrisky, The diffusion of the Internet: a cross-country analysis, Telecommun. Policy (2010), http://dx.doi.org/10.1016/j.telpol.2010.01.003.
[9] A.E. Mack, H.T. Grubesic, Forecasting broadband provision, Inf. Econ. Policy 21 (2009) 297–311.
[10] Z. Papacharissi, A. Zaks, Is broadband the future? An analysis of broadband technology potential and diffusion, Telecommun. Policy 30 (2006) 64–75.
[11] J. James, Measuring the global digital divide at the level of individuals, Curr. Sci. 96 (2009) 194–197.
[12] R. LaRose, L.J. Gregg, S. Strover, J. Straubhaarc, S. Carpenter, Closing the rural broadband gap: promoting adoption of the Internet in rural America, Telecommun. Policy 31 (2007) 359–373.
[13] P. Naik, M.R. Hagerty, C.L. Tsai, A new dimension reduction approach for data-rich marketing environments: sliced inverse regression, J. Mark. Res. 37 (2000) 113–124.
[14] H. Akaike, Information theory and an extension of the maximum likelihood principle, In: 2nd International Symposium on Information Theory, Akademia Kiado, Budapest, 1973, pp. 267–281.
[15] G. Schwartz, Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.
[16] B. Gompertz, On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies, Philos. Trans. R. Soc. Lond. 115 (1825) 513–585.
[17] J. Choudrie, Y.K. Dwivedi, Investigating the research approaches for examining technology adoption issues, J. Res. Pract. 1 (2005).

[18] S. Lee, The determinants of the global broadband deployment: an empirical analysis, Presented at the Pacific Telecommunications Council's 08 Conference, 2007.
[19] Ç.A. Çilan, B.A. Bolat, E. Coşkun, Analyzing digital divide within and between member and candidate countries of European Union, Gov. Inf. Q. 26 (2009) 98–105.
[20] D.I. Giokas, G.C. Pentzaropoulos, Efficiency ranking of the OECD member states in the area of telecommunications: a composite AHP/DEA study, Telecommun. Policy 32 (2008) 672–685.
[21] N. Zupan, Using cellular automata to simulate electronic commerce receptivity in small organisations, Technol. Forecast. Soc. Chang. 74 (2007) 798–818.
[22] G. Ford, T. Koutsky, L. Spiwak, The Broadband Performance Index: a policy-relevant method of comparing broadband adoption among countries, Phoenix Center Policy Paper Number 29, 2007.
[23] G. Ford, T. Koutsky, L. Spiwak, The Broadband Efficiency Index: what really drives broadband adoption across the OECD, Phoenix Center Policy Paper Number 33, 2008.
[24] G. Fraquelli, D. Vannoni, Multidimensional performance in telecommunications, regulation and competition: analysing the European major players, Inf. Econ. Policy 12 (2000) 27–46.
[25] H. Kum, Global Broadband Diffusion: Identifying the Factors Affecting a Country's Broadband Deployment and a Government's Role in It, In: College of Information, Florida State University, 2008.
[26] L.J. Stanton, Factors influencing the adoption of residential broadband connections to the internet, In: 37th Annual Hawaii International Conference on System Sciences, 2004.
[27] T. Turk, Borka Jerman Blazic, P. Trkman, Factors and sustainable strategies fostering the adoption of broadband communications in an European Union, Technol. Forecast. Soc. Chang. 75 (2008) 933–951.
[28] S. Park, S.-H. Yoon, Separating early-adopters from the majority: the case of broadband Internet access in Korea, Technol. Forecast. Soc. Chang. 72 (2005) 301–325.
[29] S. Howick, J. Whalley, Understanding the drivers of broadband adoption: the case of rural and remote Scotland, J. Oper. Res. Soc. 59 (Oct 2008) 1299–1311.
[30] S. Lee, S.J. Brown, Examining broadband adoption factors: an empirical analysis between countries, Emerald 10 (2008).
[31] T. Modis, The end of the internet rush, Technol. Forecast. Soc. Chang. 72 (2005) 938–943.
[32] B.J. Horrigan, Broadband adoption at home in the United States: growing but slowing, Presented at the 33nd Telecommunications Policy Research Conference, , 2005.
[33] Office of Technology Policy, U.S. Department of Commerce, Understanding Broadband Demand: A Review of Critical Issues, 2002. ed.
[34] E.B. Whitacre, Factors influencing the temporal diffusion of broadband adoption: evidence from Oklahoma, Presented at the SAEA Meetings in Dallas, Texas, 2008.
[35] Z. Hulicki, Drives and barriers for development of broadband access — CE perspective, J. Univ. Comput. Sci. 14 (2008) 717–730.
[36] E.S. Stern, S. Gregor, A.M. Martin, S. Goode, J. Rolfe, A classification tree analysis of broadband adoption in Australian households, In: ICEC'04: Sixth International Conference on Electronic Commerce, 2004.
[37] BBC World Service, Four in five regard internet access as a fundamental right: global poll, GlobeScan, 2010..
[38] D. Pontikakis, C. Patrick, Exploring geographic variation in corporate broadband adoption; evidence from Irish small- and medium-sized enterprises (SMEs), Telecommun. Policy 34 (2010) 144–161.
[39] D. Vakratsas, C. Kolsarici, A dual-market diffusion model for a new prescription pharmaceutical, Int. J. Res. Mark. 25 (2008) 282–293.
[40] G.S. Maddala, Introduction to Econometrics, 2nd ed. Macmillan, 1992.
[41] R.L. Thomas, Introductory Econometrics, 2nd ed. Longman, 1993.
[42] Y.S. Chong, J.-L. Wang, Statistical modeling via dimension reduction methods, Nonlinear Anal. Theory Methods Appl. 30 (1997) 3561–3568.
[43] W.J. Conover, Practical Nonparametric Statistics, 3rd ed. John Willey & Sons, Inc., 1999.
[44] N. Duan, K.C. Li, Slicing regression: a link free regression method, Annu. Stat. 19 (1991) 505–530.
[45] W.S. Cleveland, Robust locally weighted regression and smoothing scatterplots, J. Am. Stat. Assoc. 74 (1979) 829–836.
[46] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control. 19 (1974) 716–723.
[47] Eurostat, Statistical Office of the European Communities. Available: http://epp.eurostat.ec.europa.eu 2009.
[48] OECD, Organisation for Economic Co-operation and Development. Available: http://www.oecd.org/statsportal/0,3352,en_2825_293564_1_1_1_1_1,00.html 2009.
[49] E. Hargittai, Digital na(t)ives? Variation in internet skills and uses among members of the "net generation", Sociol. Inq. 80 (2010) 92–113.
[50] R. Frieden, Lessons from broadband development in Canada, Japan, Korea and the United States, Telecommun. Policy 29 (2005) 595–613.
[51] J.H. Hanke, D.W. Wichern, Business Forecasting, Pearson International Edition, New Jersey, 2009.
[52] C.H. Chen, K.C. Li, Can SIR be as popular as multiple linear regression? Stat. Sin. 8 (1998) 289–316.
[53] P.M.E. Altham, Improving the precision of estimation by fitting a model, J. R. Stat. Soc. 46 (1984) 118–119.
[54] M.C. Davidson, J.M. Santorelli, Network Effects: an Introduction to Broadband Technology & Regulation, U.S. Chamber of Commerce, 2008.

Vagia Kyriakidou received her degree in Finance and Accounting from Athens University of Economics and Business (AUEB) in 2003. She holds a M.Sc. degree in Administration and Economics of Telecommunication Networks from National and Kapodistrian University of Athens (Interfaculty course of the Departments of Informatics and Telecommunications and Economic Sciences). Since 2006, she has been working with the Laboratory of Optical Communications in the Department of Informatics and Telecommunications (National and Kapodistrian University of Athens). Since 2007, she has been a PhD candidate in the University of Athens and her research interests include socio-economic analysis of the telecommunication sector, business modeling for optical networks, structural equation modeling, non-parametric analysis etc.

Christos Michalakelis holds a B.Sc. in Mathematics (University of Athens, Department of Mathematics), an M.Sc. degree in Software Engineering (The University of Liverpool, UK) and an M.Sc. degree in "Administration and Economics of Telecommunication Networks" from the National and Kapodistrian University of Athens (Department of Informatics and Telecommunications, Interfaculty course of the Departments of Informatics and Telecommunications and Economic Sciences). He also holds a Ph.D. (University of Athens, Department of Informatics and Telecommunications) degree in the area of technoeconomics especially in demand estimation and forecasting of high technology products.He has been working for 7 years with the Greek Ministry of Education, in the Managing Authority of Operational Program for Education and Initial Vocational Training, as an IT manager. Dr Michalakelis has participated in a number of projects concerning the design and implementation of information systems and he now participates in several technoeconomic activities for telecommunications, networks and services.Dr Michalakelis' research interests include the study of diffusion of high technology products and services, the application of economic theories over the telecommunication and the wider area of informatics and price modeling and indexing. He has authored a number of papers presented to conferences and he has ten published papers in scientific journals. He has also contributed as a co-author in three book chapters.

Thomas Sphicopoulos received the Physics degree from Athens University in 1976, both the D.E.A. degree and Doctorate in Electronics from the University of Paris VI in 1977 and 1980 respectively, the Doctorat Es Science from the Ecole Polytechnique Federale de Lausanne in 1986. From 1976 to 1977 he worked in Thomson CSF Central Research Laboratories on Microwave Oscillators. From 1977 to 1980 he was an associate researcher in the Thomson CSF Aeronautics Infrastructure Division. In 1980 he joined the Electromagnetism Laboratory of the Ecole Polytechnique Federal de Lausanne where he carried out research on applied electromagnetism. Since 1987 he has been with the Athens University engaged in research on broadband communications systems. In 1990 he was elected as an assistant professor of communications in the Department of Informatics and Telecommunications, in 1993 as associate professor and since 1998 he has been a professor in the same department.His main scientific interests are optical communication systems and networks and techno-economics. He has lead about 50 national and European R&D projects.Professor Sphicopoulos has more than 150 publications in scientific journals and conference proceedings and he is an advisor in several organizations.