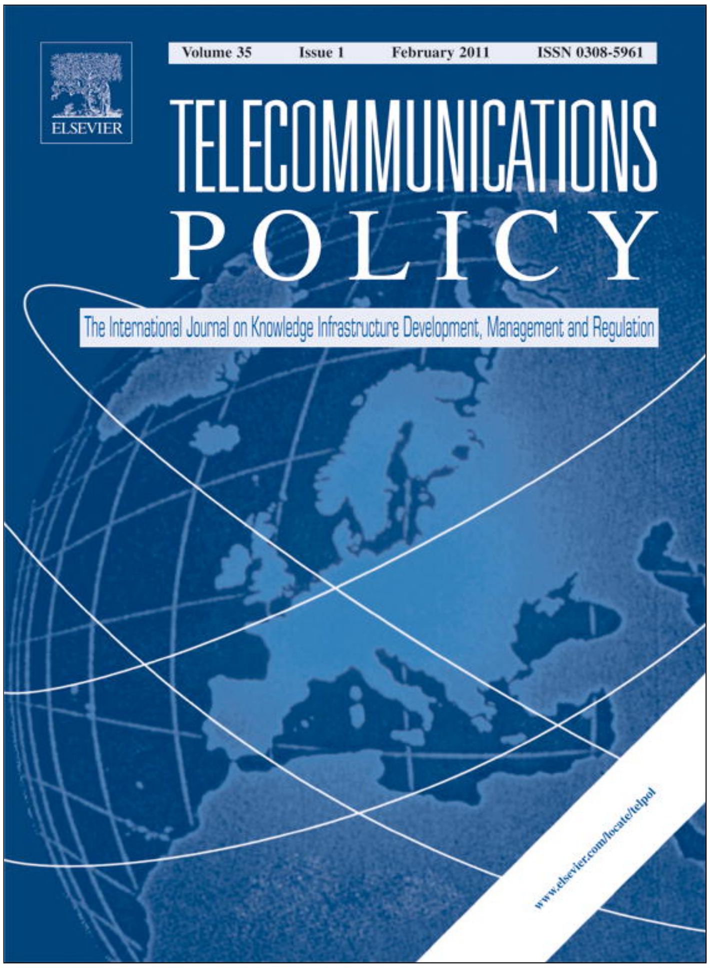
ELSEVIER
Volume 35
Issue 1
February 2011
ISSN 0308-5961
TELECOMMUNICATIONS
POLICY
The International Journal on Knowledge Infrastructure Development, Management and Regulation
www.elsevier.com/locate/telpol

Telecommunications Policy 36 (2012) 51–68

Contents lists available at SciVerse ScienceDirect

Telecommunications Policy

URL: www.elsevier.com/locate/telpol

Exploring the determinants of the OSS market potential: The case of the Apache web server

Spyridoula Lakka*, Christos Michalakelis, Dimitris Varoutas, Draculis Martakos

University of Athens, Department of Informatics and Telecommunications, Panepistimioupolis, Ilisia 15784, Athens, Greece

ARTICLE INFO

Keywords:
Diffusion of innovations theory
Institutional theory
Endogenous growth theory
Exogenous growth theory
Open source software

ABSTRACT

A theoretical framework is proposed for the better understanding of the OSS global diffusion. Following a case study approach, the Apache web server's market potential is estimated, forecasted and examined in terms of the socio-economic factors determining its diffusion, across different economic environments in developed versus developing countries. Market saturation is explored under the prism of three theoretical perspectives: the institutional, the endogenous and the exogenous growth theories. Findings suggest that Apache market saturation levels depend on both endogenous and exogenous to a country factors and that institutional quality plays an important role to the market potential. Implications for theory and public policy are discussed.

1. Introduction

Over the last years open source software (OSS) has moved to mainstream, creating a rapidly evolving ecosystem (i.e., open communities, standards and technologies). It has matured to a point that a growing number of business solutions exist, delivering real business value today. According to Forrester, 46% of North American companies now use open source software (Hammond, 2010) and 31% of large European companies have completed open source software roll-outs (Peyret, 2005).

As a result, the study of the OSS diffusion and the factors associated with it has been an objective of extensive research, as for instance Dedrick and West (2003, 2004), Bonaccorsi and Rossi (2003), Ågerfalk, Deverell, Fitzgerald, and Morgan (2005), Glynn, Fitzgerald, and Exton (2005), Goode (2005), Morgan and Finnegan (2007), Gallego, Luna, and Bueno (2008), Whitmore, Choi, and Arzrumtsyan (2009), Ven, Verelst and Mannaert (2008). However, current research still fails to answer questions like, how OSS is diffused across countries and what are the factors that enable this diffusion. Motivated by this gap in the literature, this research follows a case study approach and explores the diffusion capabilities of a successful OSS innovation, the Apache web server. Case study provides an appropriate research methodology to explore a situation in its natural setting and allows the researcher to get a deeper understanding of the object analyzed. Thus, it is very useful for theory building research (Yin, 2003).

The study, taking into account the international nature of the Apache server, proposes a theoretical framework that can provide better understanding of the diffusion process and its underlying factors. More particularly, the study aims to:

(i) Estimate and forecast the market potential (or market saturation) for the Apache web server. The estimation and forecasting is performed by applying a diffusion model, selected so as to reflect the rapid diffusion of the web servers

* Corresponding author. Tel.: +30 210 7275186; fax: +30 210 7275601.
E-mail addresses: lakka@di.uoa.gr (S. Lakka), michalak@di.uoa.gr (C. Michalakelis), arkas@di.uoa.gr (D. Varoutas), martakos@di.uoa.gr (D. Martakos).

0308-5961/$ - see front matter
doi:10.1016/j.telpol.2011.11.018

market. More specifically, the model assumes a time variant market potential that depends on the total population growth of the web servers market. Results provide critical information for the interpretation of the diffusion process, namely the current stage of the market with respect to its saturation level and its inflection point.

(ii) Explore the possible socio-economic, country-level factors that affect the Apache's market potential and ultimately diffusion. Taking into account the international dimension of Apache web server as an infrastructure of Internet technology, the study examines the market saturation behavior across different economic environments in developed versus developing countries. The factors are drawn out of three theoretical perspectives: the exogenous growth theory, the endogenous growth theory and the institutional theory. Using the estimated values of the Apache's market potential obtained in (i), an econometric analysis is performed in order to explore the critical, country-level factors that determine the Apache's diffusion process. Results also suggest the theoretical perspectives that can best interpret the phenomenon.

The selected case study presents a remarkable penetration over the last years and a constantly first position in the web servers market shares (Netcraft, 2010). The Apache HTTP server is an OSS technology serving as an infrastructure component of the web. The nature of diffusion of web servers, and Apache in particular, has important implications for the production strategies of companies dealing with host based services, web applications, social networks, various applications on demand and so forth. This, together with the dramatically increasing rates of Internet adoption among organizations, the public sector and individuals, makes the web servers market a sector of high interest.

In addition, even though the developed methodology was evaluated over data describing the diffusion of the Apache web server, it could be also be applied and evaluated for other rapidly diffused OSS in markets exposing similar characteristics.

Research results can provide helpful input for both researchers and practitioners. Research would benefit by the provision of new directions towards the development of a theoretical framework for the interpretation of the diffusion of OSS technology. This, in turn, would help to better understand the diffusion shapes of the software market. For practitioners, research findings can be very useful to strategic planning and decision making, in a continuously increased competitive environment, since more accurate a-priori estimates of the diffusion pattern can be derived.

The structure of the paper is as follows: Section 2 describes the theoretical framework of the study. Section 3 presents the conceptual model as well as the hypotheses considered under the prism of the three theoretical perspectives. Section 4 estimates and forecasts the Apache diffusion and market potential curves. Discussion of results are also provided. Section 5 provides with the econometric analysis for the identification of critical factors for the Apache's market potential. Implications of the corresponding results are presented. Section 6 summarizes conclusions together with limitations of current study and directions for future research.

2. Theoretical framework

The study of Apache's diffusion patterns is based on the mathematical modeling drawn from the diffusion of innovations (DoI) theory. The theory has been explicitly described by Rogers (1995) as the "process by which an innovation is absorbed by a social system". The latter is widely used to describe the process of the diffusion of innovations through a social system, seeking explanations in terms of how innovations diffuse through it, by the means of appropriate mathematical models, the so called diffusion models.

The diffusion process for most innovations follows an S-shaped curve with three main stages of the innovation life cycle: introduction, take off and growth and finally the maturity of the market. During the first stage (introduction), the adoption rate is relatively low, followed by the next stage (the take off), described by a high rate of adoption, until the peak of the bell curve is reached, which corresponds to the inflection point of the cumulative adoption. After that time, adoption rate decreases until the market saturation level is asymptotically met and the maximum number of adopters is reached. Although the diffusion path of most innovations can be represented in the general S-shaped fashion, different types of innovations can result in different evolutionary patterns. As a result, apart from the basic model of Bass (1969), a number of variations in the mathematical modeling exists, (e.g., Mahajan, Muller, & Bass, 1990; Meade & Islam, 2006). Also, a number of researchers have questioned the assumption that market saturation remains constant through time. In these case, market saturation is considered as a function of time and is frequently explored in terms of the factors affecting its value, (e.g., Chow, 1967; Chaddha & Chitgopekarm, 1971; Islam & Meade, 1996; Islam, Fiebig, & Meade, 2002; Karshenas & Stoneman, 1992; Mahajan & Peterson, 1978; Meade, 1985; Talukdar, Sudhir, & Ainslie, 2002).

Technology innovation diffusion models have helped to explain the adoption and diffusion of different types of new Information and Telecommunication Technologies (ICT). For instance, Liang (1986) for decision support systems, Kauffman and Techatassanasoontorn (2009) for the digital wireless phones, Krcmar and Lucas (1991) for strategic information systems, Standing and Vasudavan (2000), Gurbaxany (1990) for Bitnet, Kim and Galliers (2004), Rai, Ravichandran, and Samaddar (1998), Press (1997) for the Internet, Whitmore et al. (2009) for OSS diffusion, Chen (1998) for the web diffusion and others.

A milestone of this study is the mathematical modeling of the diffusion process of the selected case. This will enable accurate estimation and forecasting of diffusion elements such as the expected market saturation, the current stage of the market with respect to its saturation level and its inflection point. For the case of Apache, the hypothesis that the market saturation level $\overline{N}$ is not constant, but is a function of time throughout the diffusion process, is evaluated. This assumption is based on the fact the Apache exhibits a rapid pace of adoption, a characteristic which cannot be captured by a constant saturation level.

The proposed method suggests a mathematical representation of the diffusion, based on Mahajan and Peterson (1978), as well as Mahajan, Peterson, Jain, and Malhotra (1979) research. They introduced a diffusion model (dynamic model), which assumes a time variant market saturation. In particular, the market potential can be expressed as $\overline{N}t = F(S(t))$, where $S(t)$ represents the vector of all relevant exogenous and endogenous factors affecting $\overline{N}(t)$. Examples of such factors can be socio-economic conditions, competition, social system population increases or decreases, government actions, marketing efforts, etc. The dynamic diffusion model takes into account the influence of one of the aforementioned factors, namely the market population growth (denoted by $P(t)$), that is, $\overline{N}(t) = F(P(t))$.

The contribution of this study is to extend the knowledge derived by the dynamic diffusion model and explore the impact of more factors on the time variant market potential. For that reason, market potential is further analyzed under the prism of three socio-economic theories: the exogenous growth, the endogenous growth and the institutional theories. The proposed theories have been widely used in Information Systems research to explain growth of technological innovations across countries. A brief overview of these theories is presented in the following subsections.

2.1. *Endogenous growth theory*

According to this theory, economic growth is generated from within a system as a direct result of internal processes Romer (1986, 1994). More specifically, the theory notes that the enhancement of a nation's human capital will lead to economic growth by means of the development of new forms of technology, as well as efficient and effective means of production. Supporters of endogenous growth theory argue that the productivity and economies of today's industrialized countries, as compared to the same countries in pre-industrialized eras, are evidence that growth was created and sustained from within the country and not through trade. Endogenous growth theory suggests that technological innovations are driven by the profit motives of agents within an economy and that government policies on technological innovation can affect long-run growth (Grossman & Helpman, 1991; Aghion & Howitt, 1992).

2.2. *Exogenous growth theory*

The concept of exogenous growth grew out of the neoclassical growth model and the works contributed by Solow (1956). It outlines how a steady economic growth rate will be accomplished with the proper amounts of the three driving forces: labor, capital and technology. The theory states that when a new technology becomes available, labor and capital need to be adjusted to maintain growth equilibrium.

Exogenous growth theory accounts for economic growth based on sustained technological advances, a constant capital-output ratio and labor share and stable interest rates on capital (Kaldor, 1961). It assumes that economic prosperity is primarily determined by external rather than internal factors, such as the flow of goods, ideas, capital and technology innovations.

2.3. *Institutional theory*

Institutionalism attends to the deeper and more resilient aspects of social structure. It considers the processes by which structures, including schemas, rules, norms and routines, become established as authoritative guidelines for social behavior. It inquires into how these elements are created, diffused, adopted and adapted over space and time and how they fall into decline and disuse (Scott, 2004). The theory exhibits three main streams in the view of the institutions in society: rational choice theorists, who stress regulative elements (e.g., Moe, 1984; North, 1990); early sociologists, who favor normative elements (Hughes, 1958; Parsons, 1990); and more recent organizational sociologists and cultural anthropologists, who emphasize cultural-cognitive elements (e.g., DiMaggio & Powell, 1991; Scott, 2001).

This study focuses on regulative aspects and the role of formal institutions in society. Formal institutions are explicitly created usually by law and government directives and include written rules, regulations, laws and contracts that represent the choices made by a society to give structure to their relations with others (North, 1990). Property rights, checks and balances in the branches of government that promote political stability, laws that protect freedom of the press are all examples of formal institutions.

3. Conceptual model

A country is conceptualized as an economic system within which Apache growth occurs. The model is based on the idea that the forces of growth to an economic system comprise of institutional, endogenous and exogenous factors and is specified as

$$\overline{N^i}(t) = F(X^{it}, Y^{it}, Z^{it}, \vartheta^i) \tag{1}$$

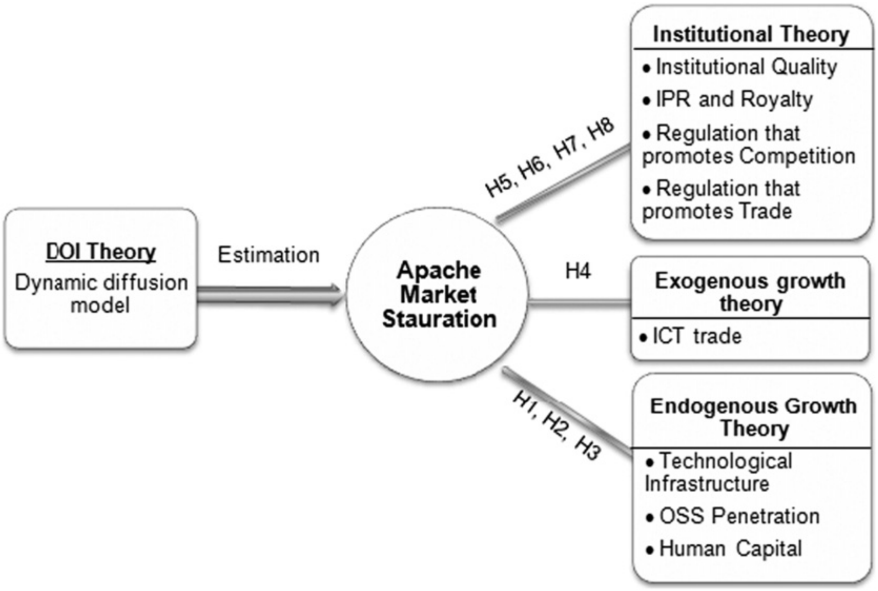

Fig. 1. Conceptual framework.

where X^{it} is a vector of all factors relevant to endogenous growth theory, Y^{it} a vector of all factors relevant to exogenous growth theory, Z^{it} a vector of all factors relevant to institutional theory, for each country *i*, at time *t*. In addition, ϑ^{i} a country specific variable, which determines developed and developing countries. The categorization has been based on whether a country belongs to the Organization for Economic Co-operation and Development (OECD)[1] group, or not.

As no previous research relevant to the determinants of cross country OSS diffusion has been detected, factors were selected on the grounds of specific hypotheses set under the prism of:

(i) The theoretical framework of the three socio-economic theories, as described in Section 2.
(ii) Apache's special characteristics, that are common to most OSS technologies.
(iii) Previous research in diffusion of technological innovations.

The conceptual model is illustrated in Fig. 1. The following paragraphs describe the hypotheses setting and the associated factors for each theoretical perspective.

3.1. Endogenous factors

Variables relevant to the theory are social, economical, environmental, organizational and technological factors that describe the conditions within a country. For instance, technological infrastructure, the number of Internet users, human capital (as used in Barro (1991)), research and development (R&D) level, the average level of education (Kiiski & Pohjola, 2002), etc.

As Apache is an infrastructure of the web, technological factors related to telecommunications advances and infrastructure are considered important elements for its diffusion. Technology readiness was also found significant for information systems adoption by a number of studies (e.g., Zhu & Kraemer, 2005; Armstrong & Sambamurthy, 1999). Thus, it is assumed that *technological infrastructure has a positive impact on Apache diffusion* (Hypothesis H1).

Another technological factor is the OSS penetration. However, its impact on Apache diffusion should not be regarded merely in a technological context. More importantly, it is assumed as an indirect network externality effect (Katz & Shaphiro, 1994), that is, the installed base of OSS adopters may influence the decision of potential web server adopters, towards zpache. This comes in accordance with Talukdar et al. (2002) who showed that the installed base of complementary products can influence adoption. In addition, the more people become familiar to OSS technology, the more people are probable to opt for a well established OSS product, such as Apache. Therefore, it is hypothesized that *OSS penetration plays an important role on Apache diffusion* (Hypothesis H2).

Finally, human capital is recognized as a key input to the development of knowledge, new ideas and products associated with technological progress. Prior studies demonstrate that the average level of education (Kiiski & Pohjola, 2002) and the quality of human capital (Caseli & Coleman, 2001) are influential drivers for individual technology adoption. In addition, a number of studies (Dedrick & West, 2003; Morgan & Finnegan, 2007; Krishnamurthy, 2003) showed that skilled and experienced users are

[1] Retrieved from http://www.oecd.org.

more probable to adopt OSS, as its complexity becomes an inhibitory factor. As a result, the third hypothesis assumes that *human capital skills and education have a positive impact on Apache growth* (Hypothesis H3).

3.2. Exogenous factors

Taking into account the international dimension of OSS technologies and the Apache web server in particular, and how these technologies are diffused through the web and globally collaborated communities, knowledge spillovers and technological flows variables may prove to affect diffusion.

A number of studies consider trade as a channel for the transfer of technological knowledge (Rivera-Batiz & Romer, 1991; Grossman & Helpman, 1991). Thus, the variable ICT goods ratio of imports plus exports to GDP is considered as a proxy for the technological flows and spillovers among countries, an approach followed also by (Barro & Sala-i-Martin, 2004). It is thus assumed that *ICT trade positively affects saturation* (Hypothesis H4).

3.3. Institutional factors

Governments have the capabilities to take actions and perform interventions in the markets, play an important role in the diffusion of new technologies. Actions and policies that promote OSS in public sector have been recorded worldwide. Comino and Manenti (2005) refer some examples where the adoption of OSS is subsidized either through tax savings or hardware discounts. A third policy followed by US, is to support OSS through promotional campaigns, such as public education consortiums. Finally, in Europe the European Commission has created the open source license called European Union Public License[2] (EUPL). It also financially supports programs like The Open Source Observatory and Repository for European public administrations,[3] encouraging the collaborative development and re-use of publicly-financed OSS applications developments for use in European public administrations.

Motivations for these actions can be economical (OSS is cost effective), technological (many OSS technologies like Linux and Apache, have proved their quality and consistency through years of testing and validation) and ideological, as OSS technologies transfer the ideas, believes and notions of freedom and openness. As a result, it is of great interest to investigate possible relation between the type of governance in a country and the diffusion of OSS technologies. Institutional theory has been previously extensively applied in the Information Systems field for the study of the impact of institutions on ICT adoption (e.g., Avgerou, 2000; Aguila, Bruque, & Padilla, 2002; Bjorck, 2004; Gibbs & Kraemer, 2004, Martinez & Williams, 2010). This study focuses on four aspects of governance: (i) institutional quality, (ii) property rights and royalty fee payments, and regulatory policies towards (iii) competition and (iv) trade barriers.

3.3.1. Institutional quality

A country's institutional strength or quality is operationalized by means of the Worldwide Governance Indicators (WGI) (Worldwide Governance Indicators, 2010). The project reports on six dimensions of governance for each country: (i) voice and accountability, that is, the level of the citizens' ability to participate in processes like selecting their government, the free media, etc., (ii) political stability and absence of violence, which reflects the likelihood that the government will not be destabilized or overthrown, (iii) government effectiveness, that is, the quality of public services, the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government, (iv) regulatory quality, that is, policies and regulations that permit and promote private sector development, (v) rule of law expresses the confidence and coherence to the rules of society, the police, the courts, as well as the likelihood of crime and violence and (iv) control of corruption reflects the control over situations where public power is exercised for private gain. More information on the data and methodology can be found in Kaufmann, Kraay, and Mastruzzi (2010).

The data have been previously used by a number of scholars in order to proxy the country-level institutional strength (e.g., Bovaird & Loffler, 2003; Globerman & Shapiro, 2002; May, Pyle, & Sommers, 2002) Although each of these indicators separately may not be directly related to OSS, this study considers the mean value of the six governance indicators, in order to create a variable reflecting a country's Institutional quality (an approach followed also by Martinez and Williams (2010)). As a result, a country's *high institutional quality is expected to have a positive impact on Apache potential* (Hypothesis H5).

3.3.2. Property rights and royalty fee payments

These factors are closely related to the Apache and other OSS technologies' characteristic, like absence of property rights and royalty fee payments. For that reason, it is anticipated that such factors may impact saturation (Hypothesis H6).

[2] Retrieved from http://www.osor.eu/eupl
[3] Retrieved from http://www.osor.eu

3.3.3. Competition

Following Nicoletti and Scarpetta (2003), the study also considers actions, policies and regulations taken by a governments that shape competition intensity. This would, in turn, give an indication of the competition conditions within a country.

As a measure of such policies, the variable Business Regulation from the Economic Freedom Network report (Gwartney, Hall, & Lawson, 2010) was obtained. The variable is designed to identify the extent to which regulations and bureaucratic procedures restrain entry and reduce competition and consists of the following metrics: (i) price controls imposed, (ii) administrative requirements that firms have to comply with, (iii) bureaucracy costs, (iv) amount of time and money it takes to start a new business, (v) undocumented extra payments or bribes firms should budget, (vii) licensing restrictions, that is, days and monetary costs required to obtain a license to construct a standard warehouse and (viii) cost of tax compliance, that is, time required per year for a business to prepare, file, and pay taxes on corporate income, value added or sales taxes, and taxes on labor. More details can be found in (Gwartney et al., 2010).

Higher index score corresponds to better competition conditions. As competition is a factor that impacts the diffusion of technological innovations, *it is supposed that it also has an impact on Apache* (Hypothesis H7).

3.3.4. Trade barriers

This factor, though regulatory, it also encloses the openness and flow of ideas and knowledge which is closely related to trade (Leamer, 1988). The assumption is that regulation that enables trade, could have an impact on globally diffused technologies that also promote openness and flow of code and knowledge, like OSS and Apache (Hypothesis H8).

4. Estimation and forecasting of the Apache web server diffusion

4.1. The dynamic model

For the mathematical modeling of the Apache diffusion the diffusion model of Mahajan and Peterson (1978) and Mahajan et al. (1979) is applied. The model assumes that the maximum value of the cumulative number of adopters' diffusion is expected to reach, that is the saturation level $\overline{N}(t)$, is not a constant but is a function of time and can be expressed as $\overline{N}(t)=f(S(t))$, where $S(t)$ represents the vector of all relevant exogenous and endogenous factors affecting $\overline{N}(t)$. The model, was initially defined by Mahajan and Peterson (1978) and explores the influence of the total market population growth (denoted by $P(t)$) on the market saturation, assuming that the rate of increase in the market saturation with respect to the total market population $P(t)$, at any time t, is a constant. That is

$$\frac{d\overline{N}(t)}{dP(t)}=k_2 \tag{2}$$

Integration of Eq. (2) yields (3), where k_1 is the integration constant and k_2 is the growth rate of market saturation with respect to total market population

$$\overline{N}(t)=f(S(t))=f(P(t))=k_1+k_2P(t) \tag{3}$$

The final formulation of the dynamic model is given by

$$\frac{dN(t)}{dt}=(a+bN(t))(\overline{N}(t)-N(t))=(a+bN(t))(k_1+k_2P(t)-N(t))$$
$$N(t=t_0)=N_0=0, \quad \overline{N}(t_0)=f_0 \tag{4}$$

where $N(t)$ refers to the cumulative number of adopters at time t. Also, a and b are the parameters that capture the power of innovation and imitation of the diffusion process, respectively. The initial number of adopters at the beginning of the process, is assumed to be $N(t=t_0)=N_0$ and the quantity $\overline{N}(t)-N(t)$ corresponds to the remaining market potential, the non-adopters. N_0 represents the number of adopters at time t_0 and f_0 is the initially estimated saturation level at time $t=t_0$ such that, $\overline{N}(t_0)=f(S(t_0))=f_0\neq 0$. It should be noted that (4) reduces to the logistic model, if $\overline{N}(t)=f(P(t))=\overline{N}$ (a constant).

The solution of the differential (4) gives the number of adopters of Apache, at each point of time t, and is given by the following equation:

$$N(t)=\frac{\overline{N}(t)-f_0\exp(-at-bM(t))}{1+f_0(b/a)\exp(-at-bM(t))} \tag{5}$$

where,

$$M(t)=\int_{t_0}^{t}\overline{N}(t)dt=\frac{k_2}{m_2}\ln z(t)+(k_1+k_2\overline{P})t-\frac{k_2}{m_2}\ln\left(1+\frac{m_2}{m_1}\overline{P}\right) \tag{6}$$

and

$$z(t)=1+\frac{m_2}{m_1}\overline{P}\exp\left[-(m_1+m_2\overline{P})t\right] \tag{7}$$

Detailed derivation of Eqs. (5)–(7) can be found in (Mahajan & Peterson, 1978).

Moreover, $P(t)$ is assumed as a time-variant function that can be estimated by the logistic diffusion model and thus is formed as follows:

$$\frac{dP(t)}{dt} = (m_1 + m_2 P(t))(\overline{P} - P(t)),$$
$$P(t = t_0) = P_0 = 0 \tag{8}$$

where $\overline{P}$ is the population saturation level and the parameters m_1, m_2 are defined in accordance to the parameters a, b of (4). The solution for (8) is given by (10). Derivation of (9) can be found in Mahajan and Schoeman (1977)

$$P(t) = \frac{\overline{P}(1 - \exp[-(m_1 + m_2\overline{P})t])}{1 + \overline{P}(m_2/m_1)\exp[-(m_1 + m_2\overline{P})t]} \tag{9}$$

The parameters $\overline{P}$, m_1, m_2 are estimated using the Nonlinear Least Squares (NLS) method. The estimation of parameters a, b, k_1, k_2 is based on the discrete regression analog of (4) and NLS. The regression coefficients x_1, x_2, x_3, x_4, x_5 give the estimates for the parameters of the model

$$N(t+1) = x_1 + x_2 P(t) + x_3 N(t) + x_4 N(t) P(t) + x_5 N^2(t)$$
$$x_1 = ak_1,\ x_2 = ak_2,\ x_3 = k_1 b - a + 1,\ x_4 = k_2 b,\ x_5 = -b \tag{10}$$

4.2. The dataset

The cumulative number of Apache web servers (denoted by $N(t)$) are extracted from Netcraft's Web Server Survey (Netcraft, 2010). The data span from year 1996 to the first half of the year 2010 and are on a six months basis. The total population size of the market is the total population of all possible web server adopters and can be measured by the number of Internet users. This can be justified by the fact that a potential adopter of a web server, should firstly install an Internet connection. The data for the Internet users penetration (denoted by $P(t)$) are derived from the United Nations (UN) database (United Nations Statistics, 2010). Data span from year 1996 to 2009 (were on an annual basis, but were extrapolated on a six months basis). A summary of the descriptive statistics for each of $N(t)$ and $P(t)$, is provided in Table 1.

4.3. Estimation and forecasting results

The process of estimation and forecasting follows three steps. Firstly, the model's (5) parameters are estimated based on a subset of the first 24 observations of the two datasets of $N(t)$ and $P(t)$. The rest of the dataset is left out deliberately for the evaluation of the forecasting (holdback sample). Based on the estimated parameters and Eqs. (3), (5)–(7), the Apache penetration $N(t)$ as well as the market potential $\overline{N}(t)$are estimated for time segments $t=1$–24. Evaluation of the estimation is performed by comparing the estimated versus the observed values for $N(t)$ in terms of the goodness-of-fit statistical measures: the Coefficient of Multiple Determination (R^2), the Mean Square Error (*MSE*) and the Mean Absolute Percentage Error (*MAPE*).

Secondly, the estimated parameters are used to forecast the Apache penetration (as well as market potential) at time segments $t=25$–29. Forecasting accuracy is evaluated by comparing the forecasted against the observed values (hold back sample). The evaluation is performed in terms of the same goodness-of-fit statistical measures: R^2, *MSE* and *MAPE*, as in the estimation process. Finally and still based on the values of the estimated parameters, the Apache diffusion $N(t)$ together with the market potential $\overline{N}(t)$are forecasted for time segments $t>29$ and until the time of the inflection point, which corresponds to the point where the maximum growth rate of the diffusion is reached.

Before estimating the parameters of (5)–(7), the parameters m_1, m_2 of (8) should be estimated first, by applying NLS to (9). Results (*F-ratio*, *t-ratios*) are significant at $p<0.01$ and fitting exhibits satisfactory results in terms of R^2, *MSE* and *MAPE*, as shown in Table 2. $P(t)$ can be calculated by incorporating the estimated parameter values in (9).

Table 1
Descriptive statistics for $N(t)$, $P(t)$.

Descriptive statistics	$N(t)$	$P(t)$—Internet users
Observations	29	29
Mean	35.38	3.67
Standard error mean	6.53	0.62
Standard deviation	35.17	3.26
Variance	1237.05	10.63
Minimum	0.01	0.15
Maximum	112.16	
Skewness	0.89	0.71
Kurtosis	−0.26	−0.75

Table 2
Estimation results for $P(t)$—logistic model.

Number of observations: 24 $R^2 = 0.99$ $F(2, 21) = 1451.71$***			$MSE = 0.006$ $MAPE = 0.078$
Parameters	Value	Standard error	t-ratio
$\overline{P}$	25.75	3.402	7.56***
m_1	0.0031	0.00027	11.23***
m_2	0.004706	0.00087	5.43***

Notes: Significance levels are: * = $p < 0.10$, ** = $p < 0.05$.
*** $p < 0.01$.

Table 3
Estimation results for $N(t)$—dynamic model.

Number of observations: 24 $R^2 = 0.98$ $F(4, 19) = 895.33$***					
Coefficients	Values	Standard error	t-ratio	Parameters	Values
x_2	5.669	0.156	1.57**	α	0.79
x_3	0.067	0.23	7.78***	b	−0.00476
x_4	−0.034	0.0034	3.22***	k_1	30
x_5	0.048	0.0028	4.34***	k_2	7.15

Notes: Significance levels are as follows: * = $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

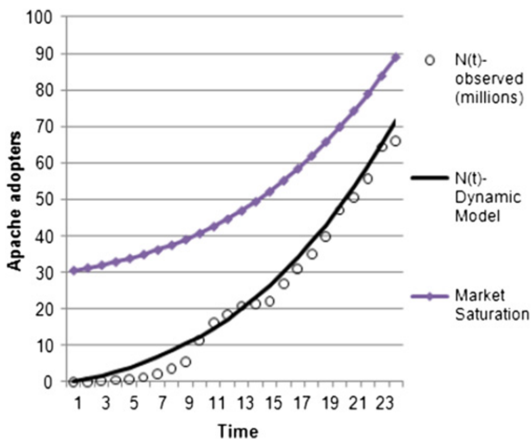

Fig. 2. Estimation of Apache diffusion with the Dynamic model-24 observations.

In the next step, the regression of Eq. (10) is performed. The regression results for x_1 gave a significant probability of accepting the zero hypothesis (*t-ratio* 0.22 with *p-value* 0.61), thus the regression was run a second time, for the rest of the coefficients, leaving x_1 out. The final regression results are given in Table 3.

As it is shown, the F-statistic and all t-statistics are significant at the $p < 0.01$ level. Consequently, the regression parameters and the regression have statistical significance and can be accepted. Using the values of x_i, $i = 2, \ldots, 5$ and (10), the parameters a, b, k_1, k_2, of the dynamic model (5) can be derived. Their values are also presented in Table 3. The positive values of k_1, k_2, and Eq. (3) imply that the Internet penetration positively affects the market potential and, consequently, diffusion.

By substituting the estimated parameters m_1, m_2, a, b, k_1, k_2 in (5)–(7), the diffusion curve as well as the time variant market potential of Apache are derived and illustrated in Fig. 2.

The evaluation of the estimation is presented in Table 4. The values of the statistical measures (R^2, *MSE* and *MAPE*) show good fitting results, confirming that the growth of the Apache server can be effectively described by the dynamic model. As shown in Fig. 2, there is a co-evolution of the $N(t)$ and $\overline{N}(t)$ curves, which shows how the evolution of the market saturation affects diffusion. This confirms that the Internet penetration $P(t)$ positively affects the diffusion process.

Table 4
Evaluation of the estimation and forecasting for $N(t)$.

Number of observations	Estimation		Forecasting	
	24		29	
	Dynamic model	*Logistic*	*Dynamic model*	*Logistic*
R^2	0.98	0.98	0.988	0.81
MSE	9.4	9.21	15.4	78.3
MAPE	1.3	2.145	1.068	2.808

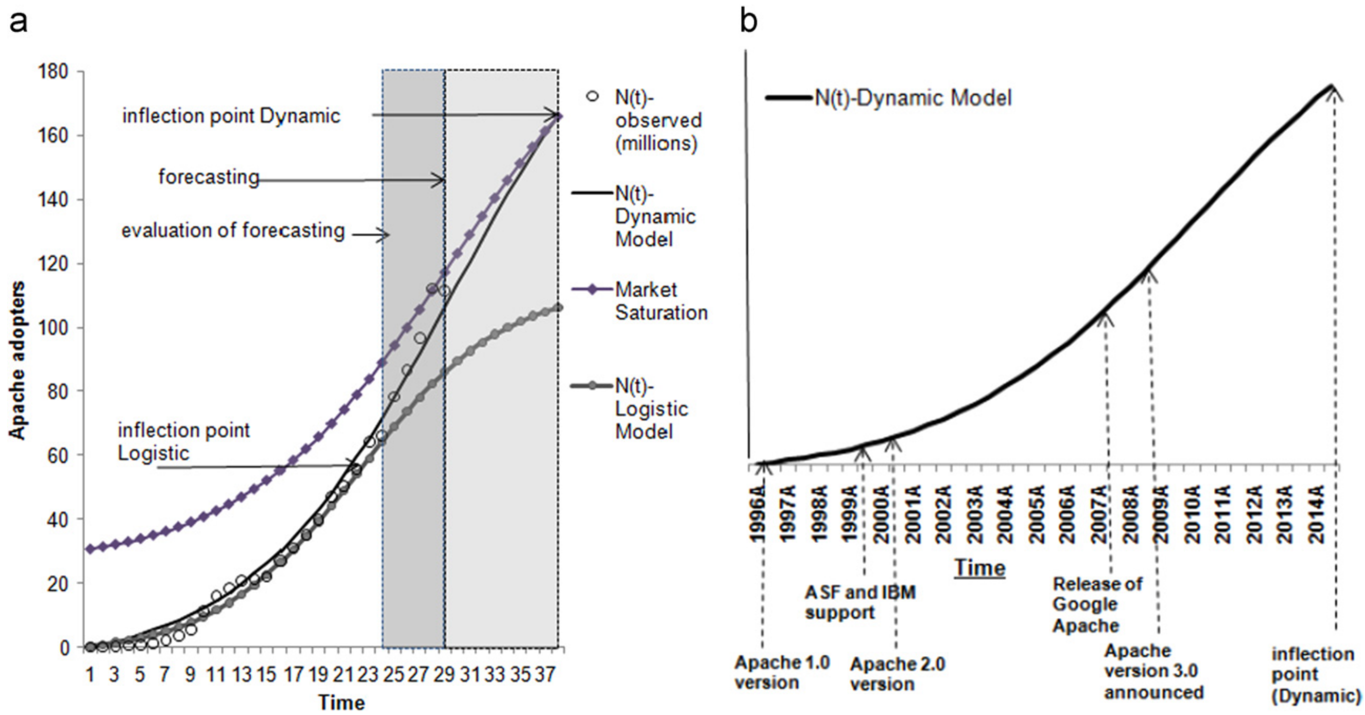

Fig. 3. (a) Diffusion and forecasting of Apache. (b) Matching historical benchmarks in the Apache diffusion.

In the next step, the evaluation of the forecasting is performed for the 29 time segments, by means of the statistical measures R^2, *MSE* and *MAPE.* As shown in Table 4, the measures confirm the dynamic model's effectiveness in fitting, for both estimation and forecasting. In order to better illustrate the dynamic model's effectiveness, Table 4 also presents the estimation and forecasting statistical measures (R^2, *MSE* and *MAPE*) obtained by applying the logistic model to the same dataset. It can be elicited that both the dynamic and logistic models can give good estimation results, yet the dynamic model has a superior forecasting ability as compared to the logistic. This is mainly due to the time variant market potential, which captures the growth of the market and shifts the diffusion curve up to higher values.

Finally, the forecasting of the Apache diffusion is estimated until the time of the inflection point. The latter is an important element of the diffusion curve, where a local extremum (maximum) occurs and the maximum growth rate is reached. As this is a maximum point, it is calculated by setting Eq. (4) equal to zero. After performing the required mathematical calculations, it is derived that (4) equals to zero, if and only if $\overline{N}(t) = -(\alpha/b)$. By setting (3) equal to $-(\alpha/b)$ and substituting (9) into (3), the time t^* of the inflection point for the dynamic model can be derived and is given by

$$t^* = \frac{\ln((1)/(S))}{m_1 + m_2\overline{P}} \text{ where } S = \frac{a + k_1 b + k_2 b\overline{P}}{k_2 b\overline{P} - \overline{P}(a + k_1 b)((m_2)/(m_1))} \tag{11}$$

From Eq. (11) it can be derived that $t^* = 37.94$, or approximately at time segment 38. As time segments are on a six months basis, time segment 38 corresponds at the second half of year 2014. It is notable that according to the logistic model,[4] the time of inflection point has already been reached (first half of 2007). Fig. 3a illustrates the forecasted diffusion curve and the different saturation levels $\overline{N}(t)$ at each time t.

[4] The logistic model is assumed to be normally distributed, thus inflection point can be easily derived by setting to the solution of $N(t)$, $N(t) = (\overline{N}/2)$. The time t^* is given by $t^* = (\ln(1)/(S))/a + b\overline{N}$, where $S = 2 + \overline{N}(b/a)$.

Table 5
Historical benchmarks in the evolution of Apache.

Time	Event
1980–1990	Tim Berners-Lee implements "Enquire" and the "Information Management" system which led to the first HTTP server
1990–1995	The National Center for Supercomputing Applications (NCSA) at the University of Illinois releases the first version of the NCSA HTTP server
April 1995	First public release of Apache HTTP server (version 0.6.2)
December 1995	First major release of Apache (version 1.0)
June 1999	Formation of the ASF with support of IBM
April 2000	Second major release of Apache (version 2.0 Alpha 1)
2007	Release of the Google web server
2008	Announcement of the third major release of Apache (version 3.0)

The shaded part of the diagram shows the forecasted parts of the curves. It can be deduced that the low value of the constant saturation level $\overline{N}$ (estimated at 112.81 million) of the logistic model shapes a downward slope that increases the gap between the last observations of the data. On the contrary, the time variant saturation level $\overline{N}(t)$ (estimated at 167.52 at $t=38$) of the dynamic model shifts the curves up resulting in much better fitting, especially for the forecasted values (time segments 25–29). This confirms the necessity for a non-constant market saturation for the rapidly diffused Apache web server.

4.4. Discussion of results

The estimated diffusion curve for the Apache server is further considered in terms of Apache's historical benchmarks as well as its current market position. The diffusion curve is forecasted until the second half of the year 2014 (time of inflection point) and is illustrated in Fig. 3b. The time line has been transformed to show the years that correspond to the model's time segments (1996–2010), while the first half of each year is denoted by the letter 'A'.

As shown, the most important benchmarks in the evolution history of Apache have been matched to the corresponding time segments of the curve. The historical benchmarks are also shown in Table 5. The information about the timing of the different Apache releases was extracted from Roy Fielding's (co-founder and Vice President of the Apache Project) presentation at the ApacheCon 2008 (Fielding, 2008).

As it can be deduced from the diffusion curve, Apache has passed the take off stage of the diffusion process at the time period 1999–2000. This can be related to two facts: firstly, in 1999 the Apache Software Foundation (ASF) was founded under the financial support of IBM. This led to the second major release of Apache in 2000. Secondly, the year 2000 coincided with the dot-com bubble, a time of a rigorous increase in investments on Internet business. Among them, large Internet actors like Yahoo!, eBay, Amazon and Google, which all used Apache http servers to offer web services and applications. This increased the popularity of Apache and consequently its adoption (network effects).

Year 2007 ($t=23$ and 24) is the time when Google released its own Apache version, the Google web server. Google represents a significant portion of the Internet, with a large number of http servers used to offer its services. The growth rate of Apache is shown to be positively influenced in Fig. 3b, as year 2007 is behind the inflection point (where the maximum growth rate is reached). This positive influence implies that competition positively affects the Apache's diffusion.

It should be noted that the inflection point is met after a long period (7 years or 14 time segments) since the time of the last observation ($t=24$) of the sample used for the estimation and forecasting of the diffusion curve. In general, long run predictions (more than 5 years) are not safe, with any diffusion model, especially when the inflection point has not been reached. Market factors, such as the new release of Apache 3.0 announced, technological improvements in the web servers market, competition, socio-economic factors, etc. may alter the market saturation and diffusion curves and shift inflection point to later time segments.

In the following sections, the market potential $\overline{N}(t)$is examined in terms of the factors impacting its expected values.

5. Socio-economic factors that determine the market potential for Apache

5.1. Methodology overview

The factors are evaluated for their impact on the Apache market saturation, by means of a panel data analysis with 25 countries over the period 2003–2008. The countries that participate in the analysis, were selected so that to represent different regions and economic status. The list of countries is presented in the corresponding Appendix A.

The full set of variables used in the econometric model, their related measures and sources are illustrated in Table 6. Some additional explanations on the derivation of the dependent variable and OSS penetration are provided in the following paragraphs.

5.1.1. Dependent variable

Market saturation $\overline{N}(t)$ of the Apache web server is the dependent variable, explored in terms of possible influencing factors. The expected market saturation $\overline{N_i}(t)$ for each country i, can be estimated by Eq. (3), that is,

$$\overline{N_i}(t) = k_1 + k_2 P_i(t) \quad (12)$$

where $P_i(t)$ is the number of Internet users and can be obtained at a per country level, by the UN database (United Nations Statistics, 2010). The parameters k_1, k_2 have been estimated in Section 3 (Table 3). By substituting the values of $P_i(t)$, k_1, k_2 in (12), the expected market saturation for each country and for each time period from 2003 to 2008 is obtained.

Table 6
Hypotheses, variables and measures.

Hypothesis	Measurement	Variable name
Endogenous growth theory variables—X^{it}		
H1. Technological infrastructures have a positive impact on Apache penetration	*Telephone lines*: Lines that connect a subscriber's terminal equipment to the telephone network and that have a port on a telephone exchange. Integrated services digital network channels and fixed wireless subscribers are included. *Source*: *World Development Indicators*(*WDI*)–*World Bank* (*WB*)[a] (World Bank Indicators, 2010)	*phone*
	ICT expenditure (*% of GDP*): Information and communications technology expenditures include computer hardware (e.g. computers, storage devices, peripherals); computer software (e.g. operating systems, utilities, applications, internal software development); computer services (e.g. consulting, computer and network systems integration, Web hosting, data processing services); communications services (i.e. voice and data) and wired and wireless communications equipment. *Source*: *WDI–WB*[a]	*ICTexp*
H2. OSS penetration plays an important role on Apache saturation	Number of subscribed users in the SourceForge portal. *Source*: *University of Notre Dame* (OSS research portal, 2010)	*OSS*
H3. Human capital skills and education have a positive impact on Apache potential	*Education expenditure* (*% GNI*): Education expenditure refers to the current operating expenditures in education, including wages and salaries and excluding capital investments in buildings and equipment. *Source*: *WDI–WB*[b]	*educGNI*
	Human capital index: The index relies on the United Nations Development Program (UNDP) 'education index' which is a composite of the *adult literacy rate* and the combined *primary*, *secondary and tertiary gross enrollment ratio* with two third weight given to adult literacy and one third to gross enrollment ratio. The index is expressed in values between 0 and 1. *Source*: *United Nations E-Government Development Database* (United Nations Public Administration Program, 2011)	*HCI*
Exogenous growth theory variables—Y^{it}		
H4. *ICTtrade* is expected to be positively related to the Apache saturation potential	*ICT goods ratio of exports plus imports per GDP:* ICT goods include telecommunications, audio and video, computer and related equipment; electronic components; and other information and communication technology goods. Software is excluded. *Source*: *WDI–WB*	*ICTtrade*
Institutional theory variables—Z^{it}		
H5. Institutional quality has an impact on Apache growth	*Institutional quality*: How effective a government is based on the presence of six dimensions in a society. The 2003–2008 average of six aggregate governance indicators measured in units ranging from −2.5 to +2.5. Higher values indicate higher institutional quality. *Source: WGI–WB* (World Governance Indicators, 2010)	*IQ*
H6. Royalty and property rights protection, may have an impact in an OSS technology such as Apache	*Protection of property rights*: A property right is the exclusive authority to determine how a resource is used, whether that resource is owned by government or by individuals. If the resource is owned by the government, the agent who determines its use has to operate under a set of rules determined. Protection of personal life and actions . Measured in the range of 1–10. *Source*: *Economic Freedom Network* (Gwartney et al., 2010)	*IPR*
	Royalty and license fees payments: Payments and receipts between residents and nonresidents for the authorized use of intangible, non-produced, nonfinancial assets and proprietary rights (such as patents, copyrights, trademarks, industrial processes, and franchises) and for the use, through licensing agreements, of produced originals of prototypes (such as films and manuscripts). Data are in current U.S. dollars. *Source*: *WDI–WB*[c]	*royalty*

Table 6 (*continued*)

Hypothesis	Measurement	Variable name
H7. Regulation policies for competition have an impact on Apache	*Business regulation*: Extent to which regulations and bureaucratic procedures restrain entry and reduce competition. Average of eight components, all measures ranging from 1 to 10. Higher values indicate, less restrictions and better competition conditions. *Source: Economic Freedom Network* (Gwartney et al., 2010)	*B_R*
H8. Regulation policies for trade are positively related to Apache potential	*Regulatory trade barriers:* Non-tariff trade barriers and compliance cost of importing and exporting. Their goal is the economic freedom of the world. Measured in the range of 1–10. Higher values indicate higher quality. *Source*: *Economic Freedom Network* (Gwartney et al., 2010)	*regTradeBar*
Qualitiative variable ϑ		
H9. Apache's penetration is higher in developed countries	Dummy variable (0 = Member of OECD and 1 = otherwise) Corresponds to ϑ variable of the theoretical model (1). *Source*: www.oecd.org	*OECD*

[a] World Bank staff estimates using data from the United Nations Statistics Division's Statistical Yearbook (http://unstats.un.org/unsd), and the UNESCO Institute for Statistics online database (http://stats.uis.unesco.org/).
[b] International Monetary Fund (IMF), Balance of Payments Statistics Yearbook and data files (http://www.imf.org/).
[c] World Bank national accounts data, and OECD National Accounts data files (http://www.oecd.org).

5.1.2. OSS penetration

The OSS penetration is measured in terms of the number of subscribed users/developers in the SourceForge.net portal (SourceForge.net, 2010). SourceForge is owned and operated by OSTG, Inc., and is the world's largest OSS development website, with the largest repository of OSS code and applications available on the Internet, offering free services to Open Source developers. The website is database driven and the supporting database includes historic and status statistics on projects and user activities. OSTG has shared certain SourceForge.net data with the University of Notre Dame (UND) for academic and scholarly research purposes.

The data used for this paper were extracted from the UND's platform (OSS research portal, 2010) under written permission. More particularly, the number of users of each country were calculated, by using the two fields *user_id* and *timezone*, of the table *user*. The first field corresponds to the user's unique id, and the second to the town in which each user lives. The field *timezone* mainly contains all countries' capital cities. Using these two fields, the number of subscribed users for each of the 25 countries of interest have been aggregated, for years 2003–2008.

5.2. Econometric analysis and results

The data described in Table 6 were initially transformed were necessary. The descriptive statistics for the dependent and independent variables, as well as their transformations are illustrated in Table 7. It should be noted that the sample contained few null values, as in some cases different countries did not have corresponding values in different variables. The small extend of the problem however, had no hindering effect on the regression.

The above results, as well as the rest of the tests and calculations were derived with the STATA software. The data are further refined, by performing a correlation analysis of the participating variables. Results showed evidence of correlations, for the variables *IPR* and *regTradeBar*, both significant at the $p<0.01$ and $p<0.05$ levels, respectively. All other correlations were lower than 80%, the criterion level suggested by Kennedy (2003). The correlation matrix is illustrated in Table 8.

The variables *IPR* and *regTradeBar*, which showed evidence of correlation are excluded and the econometric model is given by

$$\ln(Nbar_{it}) = a + b_1 \ln(phone_{it}) + b_2 ICT\exp_{it} + b_3 \ln(OSS_{it}) + b_4 educGNI_{it} + b_5 HCI_{it} + b_6 B\_R_{it} + b_7 ICTtrade_{it} + b_8 IQ_{it} + b_9 \ln(royalty_{it}) + b_{10} OECD_i + u_i + \varepsilon_{it} \quad (13)$$

For each country i and year t. Also, u_i is the country specific effect and ε_{it} is the idiosyncratic error. Firstly, the Hausman specification test (Hausman, 1978) between random and fixed effects model, clearly indicates that the fixed effects model should be preferred. ($\chi^2(9) = 46.05$, $p = 0$). Also, the Breusch and Pagan Lagrangian multiplier (LM) test for random effects (Breusch & Pagan, 1980) indicates that the variance of the country specific effect u_i, is significantly different from zero ($\chi^2(1) = 56.0$, $p = 0$, H_0: Var(u) = 0), showing the significance of the individual-specific effects.

The next specification test explores the case of endogeneity. A variable is endogenous, when it is correlated with the error term. The Durbin–Wu–Hausman (DWH) test (proposed by Durbin (1954) and separately by Wu (1973) and Hausman (1978)), showed evidence of endogeneity for the *IQ* variable ($\chi^2(1) = 10.44$, $p = 0.001$, under the null hypothesis that *IQ* is exogenous).

Table 7
Descriptive statistics and transformations of the dataset.

Variable	Transformation	Obs	Mean	Std. dev.	Min	Max
Nbar	LN( )	150	5.121	0.956	3.541	7.678
phone	LN( )	150	16.610	1.327	13.967	19.723
ICTexp		150	5.663	1.279	3.164	10.098
OSS	LN( )	150	9.474	1.374	3.892	13.210
educGNI		150	4.451	1.131	1.801	6.853
HCI		125	0.916	0.095	0.570	0.993
ICTtrade		148	20.225	13.358	6.343	59.473
IQ		150	0.692	0.761	−0.715	1.897
IPR		150	6.736	1.927	2.298	9.608
royalty	LN( )	136	21.038	1.747	15.642	24.005
B_R		150	4.99	2.12	0.00	9.00
regTradeBar		150	7.169	1.102	3.597	9.333
OECD		150	0.32	0.47	0	1

Table 8
Correlation matrix.

		1	2	3	4	5	6	7	8	9	10	11	12
***ln*(*phone*)**	1	1											
ICTexp	2	0.12	1										
***ln*(*OSS*)**	3	0.62***	0.19**	1									
educGNI	4	−0.58***	0.09	−0.18**	1								
HCI	5	−0.07	0.21**	0.33***	0.16*	1							
ICTtrade	6	0.43***	0.43***	0.13	−0.22***	0.10	1						
IQ	7	−0.21***	0.34***	0.29***	0.45***	0.61***	0.02	1					
IPR	8	−0.14*	0.42***	0.34***	0.43***	0.35***	0.02	0.82***	1				
***ln*(*royalty*)**	9	0.66***	0.46***	0.68***	−0.22**	0.43***	0.43***	0.50***	0.45***	1			
B_R	10	−0.19**	0.29***	0.29*	0.52***	0.23***	−0.13	0.65***	0.65***	0.24**	1		
reg TradeBar	11	−0.14*	0.24***	0.29***	0.36***	0.42***	0.00***	0.82**	0.74	0.43	0.69***	1	
OECD	12	0.00	0.03	0.28***	0.22***	0.61	0.17**	0.75	0.51	0.47	0.50	0.64***	1

Notes: Significance levels.
* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.

In the case of endogeneity, the most applicable econometric method is the two stage least squares method (2SLS) and generalized method of moments (GMM). According to Wooldridge (2002), the modern approach to system Instrumental Variables (IV) estimation is based on the principle of generalized method of moments (GMM). Hansen (1982) and White(1982) showed how the method of moments can be generalized to apply to a variety of econometric models, and they derived the asymptotic properties of GMM. The method also provides robust estimations in the case of arbitrary heteroscedasticity in inter- and intra-clusters and autocorrelation. A GMM estimator is a useful alternative to the fixed or random effects IV estimators, as the latter assume that correlations within groups are constant.

Apart from the above specification tests, panel models often violate standard Ordinary Least Squares (OLS) assumptions. Two post-estimation tests were performed, for serial correlation and heteroscedasticity for the 25 panels. The Wooldridge (2002) test for autocorrelation in panel data, under the null of no first-order autocorrelation, showed evidence of serial correlation in the idiosyncratic errors: $F(1,\ 22)=39.67$, $p=0$. The Pagan and Hall's (1983) test for heteroskedasticity for IV estimation, showed that disturbances are heteroscedastic ($\chi^2(11)=23.072$, $p=0.017$), under the null that disturbance is homoscedastic.

The above tests indicate that the optimal method choice is the two step feasible generalized method of moments (GMM estimator) with heteroscedastic and autocorrelation (HAC) consistent standard errors. Baum, Schaffer, and Stillman (2003) state that the two-step GMM gains in efficiency relative to the traditional IV/2SLS estimators, because of the use of the optimal weighting matrix, the overidentifying restrictions of the model and the relaxation of the violation of independently and identically distributed (i.i.d.) errors assumption.

However, the 2SLS results are valid, provided that the selected instrumental variables are also valid. More particularly, the instruments should be: (i) uncorrelated with the error term, (ii) uncorrelated with the rest of the exogenous variables and (iii) correlated with the endogenous variable *IQ*. Table 8 shows that the variables *IPR*, *regTradebar* satisfy (ii) and (iii), and thus could be selected as candidate instruments. The variable *royalty* is also added to the list of instruments, as it satisfies (ii) and also exhibits a fair correlation with the endogenous *IQ*, at a significant level ($p < 0.01$). The selected instruments are further validated with the appropriate tests, as follows.

Firstly, the orthogonality conditions (i) are tested. In the context of GMM and HAC errors, the overidentifying restrictions may be tested via the commonly employed *J* statistic of Hansen (1982), that is properly formulated to account for inter and intra cluster correlation. The *J* statistic tests for joint null hypothesis that (i) holds and that excluded instruments are correctly excluded. The *J* statistic results show clear acceptance of H_0 with $\chi^2(2)=2.78$, $p=0.24$.

Secondly, the under identification test is a test of whether the equation is identified, that is, that the excluded instruments are relevant, meaning correlated with the endogenous regressors. A rejection of the null indicates that the matrix is full column rank and the model is identified. For heteroskedastic, AC, HAC, or cluster-robust statistics, the LM and Wald versions of the Kleibergen–Paap rk statistic (Kleibergen & Paap, 2006), is used. The corresponding statistic's result rejects the null hypothesis, showing that the model is identified (Kleibergen–Paap rk LM statistic is $\chi^2(3)=12.52$, $p=0.005$, while the Kleibergen–Paap rk Wald statistic is $\chi^2(3)=72.76$, $p=0$).

Weak instrument problems, however, may still be present. Weak identification arises when the excluded instruments are correlated with the endogenous regressors but only weakly. Estimators can perform poorly when instruments are weak, and different estimators are more robust to weak instruments (Stock, Wright, & Yogo, 2002; Stock & Yogo, 2005). In the case of non i.i.d. errors, the corresponding, robust to these errors statistic, is the Kleibergen–Paap Wald rk F statistic. As a rule of thumb, the Kleibergen–Paap Wald rk *F* statistic should be at least 10 for weak identification not to be considered as a problem (Staiger & Stock, 1997). The reported *F*-statistic is $F(3,\ 100)=21.66$, $p=0$, indicating no problem of weak identification.

Finally, the significance of the endogenous regressors can be tested with the Anderson and Rubin (1949) statistic, which is robust to non-i.i.d. errors, as well. The statistic is also robust to weak identification, that is, as instruments become weak, the power of the test declines. The Anderson and Rubin statistic reports $\chi^2(3)=10.45$, $p=0.01$ under the null hypothesis that the coefficients of the endogenous regressors in the structural equation are jointly equal to zero.

From the above tests, it can be elicited that instruments are valid. As a result, the econometric model considered for the market saturation $\overline{N}$ is described by the following structural equations:

$$\ln(Nbar_{it}) = a + b_1\ln(phone_{it}) + b_2ICTexp_{it} + b_3\ln(OSS_{it}) + b_4educGNI_{it} + b_5HCI_{it} + \mathrm{b}_6B\_R_{\mathrm{it}} + \mathrm{b}_7ICTtrade_{it} + b_8IQ_{it} + b_9OECD_i + v_{it}IQ_{it} = c_0 + c_1IPR_{it} + c_2regTradeBar_{it} + c_3\ln(royalty_{it}) + \mu_{it} \tag{14}$$

where $v_{it}=u_i+\varepsilon_{it}$ and μ_{it}, are the error terms of each equation. *IQ* is the endogenous variable, identified by the three instrumental variables *IPR*, *regTradeBar* and ln(*royalty*).

As one last step, before the final model to be specified, is a simultaneity test. Simultaneity is another form of endogeneity and arises when one or more of the explanatory variables is jointly determined with the dependent variable, each of them having a ceteris paribus, causal interpretation (Wooldridge, 2002). If simultaneity exists, then both the dependent and explanatory variables are correlated with the error term, thus are endogenous. In (14), the only endogenous variable is *IQ*. Thus, it will be tested whether there is reverse causality, that is, ln(*Nbar*) and *IQ* are jointly determined. Although, there is no apparent reason why ln(*Nbar*) would have a causal effect on *IQ*, the test was mainly performed for reasons of completeness. In a case of simultaneity, the econometric model would be described by (14), apart from the structural equation describing *IQ*, which would become:

$$IQ_{it} = c_0 + c_1IPR_{it} + c_2regTradeBar_{it} + c_3\ln(royalty_{it}) + c_4\ln(Nbar_{it}) + \mu_{it} \tag{15}$$

The Wooldridge (2002) test for endogeneity, based on the significance of the residuals of the reduced equation was performed, taking into account heteroscedasticity and autocorrelation. As it was expected, there was no evidence of simultaneity (residuals where not significant: $z=0.36$, $p=0.717$). Rejecting the case of simultaneity, the final regression results of (14) of 2SLS with GMM and HAC errors are provided in Table 9

5.3. Discussion of results

Table 9 shows the positive impact and statistical significance of technological infrastructure variables, *phone* and *ICTexp.* In addition, the variable phone is the most significant variable of the regression. That was an expected result, since the market potential of web servers should depend on the existence of the necessary infrastructure for their functioning. Thus, H1 can be accepted. The same applies for H2, as the variable *OSS* is highly significant at the $p=0$ level. The positive coefficient (0.19) implies there is an indirect network externality effect of OSS growth on a specific OSS technology, such as the Apache web server.

Both the statistical measures of *educGNI* and *HCI*, confirm hypothesis H3 that Apache server adoption is affected by the level of skills and education of potential users. Thus, highly skilled users are more probable to opt for a more sophisticated software, such as Apache.

Technological flows and spillovers expressed by ICT trade per GDP, also exhibit a positive impact to Apache potential, with high statistical significance. This in turn confirms the hypothesis (H4), that ICT trade and the free exchange of code, knowledge and ideas offered by an OSS technology, like Apache, have a positive relation.

Institutional quality has a high coefficient in the regression (0.28) and is statistically significant, thus the assumption that high institutional quality helps expansion of innovative technologies like Apache (H5) is accepted. The result imply that stable and well organized societies promote the diffusion of new technologies, as they create the necessary conditions for the development of software markets and the raise of competition. Moreover, the freedom of expression and free media

Table 9
2SLS GMM **regression results for** $\overline{N}(t)$.

Second regression: dependent variable: $\ln(\overline{N}(t))$				First regression: dependent variable: *IQ*		
No. of observations: 112 **Centered** $R^2=0.93$ **Uncentered** $R^2=0.99$ $F(9, 102)=86.76$***				**No. of observations:** 112 **Centered** $R^2=0.93$ **Uncentered** $R^2=0.96$ $F(11, 100)=66.87$***		
Vars	**Coef.**	**Std. err.**	***z***	**Coef.**	**Std. err.**	***t***
ln(phone)	0.570	0.049	11.72***			
ICTexp	0.085	0.035	2.43**			
ICTtrade	0.015	0.003	4.32***			
HCI	0.011	0.007	1.91*			
ln(OSS)	0.190	0.040	4.73***			
educGNI	0.088	0.034	2.58***			
B_R	0.021	0.023	0.91			
OECD	−0.411	0.158	−2.60***			
cons	−5.901	0.785	−7.52***			
IQ	0.280	0.131	2.14**			
ln(royalty)				0.179	0.039	4.63***
reg TradeBar				0.137	0.043	3.16**
IPR				0.094	0.023	4.08***

Notes: Significance levels.
* $p<0.10$.
** $p<0.05$.
*** $p<0.01$.

complies with the ideas of open source software and open content. Hypotheses H6 and H8 could not be verified, as the corresponding variables were not tested directly on the Apache saturation, but served as instruments for the identification of institutional quality *IQ*. However, as shown in Table 9, all of the three variables have a very high impact on *IQ*. Thus, it can be regarded, that they positively affect Apache's expected saturation, but only indirectly.

Regulation that promotes competition, was found insignificant ($p>0.1$) and thus H7 cannot be accepted. This may be due to the fact that Apache is the incumbent in a duopolistic market structure, a fact with little fluctuation at a country level. In addition, Apache can be obtained for free, thus it is not affected by policies like price control. As a result, Apache's penetration has been increasing independently of the country's individual competition conditions.

Finally, hypothesis H9 that developed countries have higher saturation levels, is also accepted. The statistically significant, negative coefficient (−0.41) of *OECD*, shows that expected market saturation of developing countries (denoted by 1) is about 41% less than the developed countries, holding all other factors fixed. Although Apache can be obtained at a zero price, its adoption is hindered by the cost of effort and experience, that is higher for countries with lower levels of education, skills and technological advancements. The result comes in accordance with the problem of digital divide among developed and developing countries. Thus, developed countries, with higher technological and educational rates, are good incubators of new technologies like the Apache server, in which it can flourish.

6. Conclusion

This paper follows a case study approach aiming to shed light on the global diffusion of the OSS technologies and proposes a theoretical framework that can provide better understanding of the diffusion process and its underlying factors.

The proposed methodology consists of two discrete steps, each one contributing to the understanding of the diffusion mechanism. The first step includes the process of estimating and forecasting the global diffusion and the market potential of the Apache web server. A model with a time-variant saturation proved quite efficient and gave better prediction results, due to the fact that classical diffusion models cannot capture the changing effects of markets with rapid diffusion and changing market size characteristics. Results show that Internet penetration has a positive impact on the market potential for Apache and that the growth of Apache has not yet reached the maximum rate (inflection point). The latter means that the diffusion curve and the market potential are still growing and long run predictions are still not safe to be produced.

In the second step, socio-economic factors affecting the estimated from the previous step Apache market potential were explored. The theoretical model is empirically tested over a set of 25 countries, for a six year time period. Findings suggest that the diffusion of Apache depends on both endogenous and exogenous to a country factors, namely technological infrastructure, level of skills and education and ICT trade. Another important outcome is the impact of governmental factors, that is, the institutions, rules and laws of an organized society were found to positively affect the growth of the Apache technology. On the contrary, regulation that promotes competition does not appear to have any impact.

Table A1
List of countries.

	Country	OECD		Country	OECD
1	Argentina	No	14	Korea	Yes
2	Australia	Yes	15	Mexico	Yes
3	Belgium	Yes	16	Netherlands	Yes
4	Brazil	No	17	Norway	Yes
5	Canada	Yes	18	Romania	No
6	China	No	19	Russia	No
7	Finland	Yes	20	South Africa	No
8	France	Yes	21	Spain	Yes
9	Germany	Yes	22	Tunisia	No
10	Greece	Yes	23	Turkey	Yes
11	India	No	24	United Kingdom	Yes
12	Italy	Yes	25	United States	Yes
13	Japan	Yes			

Finally, Apache saturation levels are higher in developed versus developing countries, an outcome which is in accordance with other technological innovations and the problem of digital divide.

Although these results cannot be generalized to all OSS technologies, the proposed theoretical framework could be efficiently applied to other OSS technologies that are well established in the markets and share similar characteristics, as for instance the rest of the LAMP (Linux–Apache–MySQL–PhP) stack software, applications like JBoss and Mozilla, mobile platforms like Android, etc. The application of this framework can provide helpful inputs for strategic planning and decision making not only for administrators but also for policy makers who want to assess the opportunities of the rigorously emerging OSS technologies.

6.1. *Limitations and future research*

One of the limitations of this case study, is the fact that the results are limited in a sample of 25 countries. The sample however, has been selected, so that to be evenly distributed to all continents. Countries with large number of population were preferred, so as to cover greater part of the world's population. Moreover, as this was an initial evaluation of country specific factors that affect saturation of an OSS technology, missing factors may also exist. The findings of current research are still quite important, as they give an insight of the factors that positively affect the diffusion mechanism.

Future research, however, could explore more inhibitory or favoring factors for the market saturation of Apache, and/or other OSS technologies. In addition, other theories could also be deployed for the identification of the possible underlying factors of diffusion, like the TOE (Technological–Organizational–Environmental) framework, theories of technology acceptance, etc. Finally, of greatest research interest is the exploration of project-level OSS characteristics, able to affect its diffusion.

Appendix A

See Table A1.

References

Ågerfalk, P. J., Deverell, A., Fitzgerald, B., & Morgan, L. (2005). Assessing the role of open source software in the European secondary software sector: A voice from industry. In *The proceedings of the first international conference on open source systems held in Genova, Switzerland*.

Aghion, P., & Howitt., P. (1992). A model of growth through creative destruction. *Econometrica*, *60*(2), 323–351.

Aguila, A. R. d., Bruque, S., & Padilla, A. (2002). Global information technology management and organizational analysis: Research issues. *Journal of Global Information Technology Management*, *5*(4), 18.

Anderson, T. W., & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *Annals of Mathematical Statistics*, *20*, 46–63.

Armstrong, C. P., & Sambamurthy, V. (1999). Information technology assimilation in firms: The influence of senior leadership and IT infrastructures. *Information Systems Research*, *10*(4), 304–327.

Avgerou, C. (2000). IT and organizational change: An institutionalist perspective. *Information Technology & People*, *3*(4), 234.

Barro, R. J. (1991). Economic growth in a cross section of countries. *Quarterly Journal of Economics*, *106*(2), 407–443.

Barro, R. J., & Sala-i-Martin, X. (2004). *Economic growth*. Cambridge, MA: MIT Press.

Bass, F. M. (1969). A new product growth model for consumer durables. *Management Science*, *15*(5), 215–227.

Baum, C. F., Schaffer, M. E., & Stillman, S. (2003). Instrumental variables and GMM: Estimation and testing. *Stata Journal*, *3*(1), 1–31.

Bjorck, F. (2004). *Institutional theory: A new perspective for research into IS/IT security in organisations*. Paper presented at the 37th annual Hawaii international conference on system sciences (HICSS'04) held in Hawaii.

Bonaccorsi, A., & Rossi, C. (2003). Why open source software can succeed. *Research Policy*, *32*(7), 1243–1258.

Bovaird, T., & Loffler, E. (2003). Evaluating the quality of public governance: Indicators, models and methodologies. *International Review of Administrative Sciences*, *69*(3), 313–328.

Breusch, T. S., & Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies*, *47*, 239–253.

Caseli, F., & Coleman, W. J. (2001). Cross-country technology diffusion: The case of computers. *The American Economic Review*, *91*(2), 328–335.
Chaddha, R. L., & Chitgopekar, S. S. (1971). A ''generalization'' of the logistic curves and long-range forecasts (1966–1991) of residence telephones. *The Bell Journal of Economics and Management Science*, *2*(2), 542–560.
Chen, H. (1998). *Dual acceptance of web diffusion: A case of clients and servers*. Paper presented at the 26th annual conference of the Canadian association for the information science held in Ottawa, Canada.
Chow, G. C. (1967). Technological change and the demand for computers. *American Economic Review*, *57*, 1117–l 1130.
Comino, S., & Manenti, F. M. (2005). Government policies supporting open source software for the mass market. *Review of Industrial Organization*, *26*, 217–240.
Dedrick, J., & West, J. (2003). Why firms adopt open source platforms: A grounded theory of innovation and standards adoption. In *The proceedings of the workshop on standard making: A critical research frontier for information systems held in Seattle, Washington*.
Dedrick, J., & West, J. (2004). *An exploratory study into open source platform adoption.* Paper presented at the 37th Hawaii international conference on system sciences (HICSS 04) held in Hawaii.
DiMaggio, P. J., & Powell, W. W. (1991). Constructing an organizational field as a professional project: U.S. art museums, 1920–1940. In: P. J. DiMaggio, & W. W. Powell (Eds.), *The new institutionalism in organizational analysis* (pp. 267–292). Chicago: University of Chicago Press.
Durbin, J. (1954). Errors in variables. *Review of the International Statistical Institute*, *22*, 23–32.
Fielding, R. (2008). *Apache 3.0* (*a tall tale*). Paper presented at the ApacheCon Europe 2008 held in Amsterdam, The Netherlands.
Gallego, M. D., Luna, P., & Bueno, S. (2008). Designing a forecasting analysis to understand the diffusion of open source software in the year 2010. *Technological Forecasting and Social Change*, *75*, 672–686.
Gibbs, J. L., & Kraemer, K. L. (2004). A cross-country investigation of the determinants of scope of e-commerce use: An institutional approach. *Electronic Markets*, *14*(2), 124–137.
Globerman, S., & Shapiro, D. (2002). Global foreign direct investment flows: The role of governance infrastructure. *World Development*, *30*(11), 1899–1919.
Glynn, E., Fitzgerald, B., & Exton, C. (2005). Commercial adoption of open source software: An empirical study. In The *Proceedings of International Conference on Empirical Software Engineering held in Noosa Heads, Australia*.
Goode, S. (2005). Something for nothing: Management rejection of open source software in Australia's top firms. *Information & Management*, *42*, 669–681.
Grossman, G. M., & Helpman., G. (1991). *Innovation and growth in the global economy*. Cambridge, MA: MIT Press.
Gurbaxany, V. (1990). Diffusion in computing networks: The case of BITNET. *Communications of the ACM*, *33*(12), 65–75.
Gwartney, J.D., Hall, J.C., & Lawson, R. (2010). *Economic freedom of the world: 2010 annual report*. Economic Freedom Dataset, Economic Freedom Network. Retrieved from 〈http://www.freetheworld.com/release.html〉.
Hammond, J. (2010). *Open source adoption patterns in enterprise IT.* Paper presented at the LinuxCon held in Boston, MA.
Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*, 1029–1054.
Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*, 1251–1271.
Hughes, E. C. (1958). *Men and their work*. Glencoe: Free Press.
Islam, T., Fiebig, D. G., & Meade, N. (2002). Modelling multinational telecommunications demand with limited data. *International Journal of Forecasting*, *18*(4), 605–624, doi:10.1016/s0169-2070(02)00073-0.
Islam, T., & Meade, N. (1996). Forecasting the development of the market for business telephones in the UK. *Journal of the Operational Research Society*, *47*, 906–918.
Kaldor, N. (1961). Capital accumulation and economic growth. In: F. A. Lutz, & D. C. Hague (Eds.), *The theory of capital*. New York: St. Martin's Press.
Karshenas, M. D., & Stoneman, P. (1992). A flexible model of technological diffusion incorporating economic factors with an application to the spread of colour television ownership in the UK. *Journal of Forecasting*, *11*, 577–601.
Katz, M., & Shaphiro, C. (1994). Systems competition and network effects. *Journal of Economic Perspectives*, *8*(2), 93–115.
Kauffman, R. J., & Techatassanasoontorn, A. A. (2009). Understanding early diffusion of digital wireless phones. *Telecommunications Policy*, *33*, 432–450.
Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). *The worldwide governance indicators: Methodology and analytical issues* [World Bank Policy Research Working Paper no. 5430.]. SSRN Retrieved from 〈http://ssrn.com/abstract=1682130〉.
Kennedy, P. A. (2003). *A guide to economics* (5th edition). Cambridge, MA: MIT Press.
Kiiski, S., & Pohjola, M. (2002). Cross-country diffusion of the Internet. *Information Economics and Policy*, *14*(2), 297–310.
Kim, C., & Galliers, D. (2004). Toward a diffusion model for Internet systems. *Internet Research*, *14*(2), 155–166.
Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singularvalue decomposition. *Journal of Econometrics*, *127*, 97–126.
Krcmar, W. J., & Lucas, H. C. (1991). Success factors for strategic information systems. *Information & Management*, *21*, 137–145.
Krishnamurthy, S. (2003). A Managerial overview of open source software. *Business Horizons*, *46*(5), 47–56.
Leamer, E. (1988). Measures of openness. In: R. E. Baldwin (Ed.), *Trade policy issues and empirical analysis* (pp. 145–204). Chicago, USA: University of Chicago Press.
Liang, T. P. (1986). Critical success factors of decision support systems: An experimental study. *Data Base*, 3–17.
Mahajan, V., Muller, E., & Bass, F. M. (1990). New product diffusion-models in marketing—A review and directions for research. *Journal of Marketing*, *54*(1), 1–26.
Mahajan, V., & Peterson, R. A. (1978). Innovation diffusion in a dynamic potential adopter population. *Management Science*, *24*(15), 1589–1597.
Mahajan, V., Peterson, R. A., Jain, A. K., & Malhotra, N. (1979). New product growth-model with a dynamic market potential. *Long Range Planning*, *12*(4), 51–58.
Mahajan, V., & Schoeman, M. (1977). Generalised model for the time pattern of the diffusion process. *IEEE Transactions on Engineering Management*, *EM-24*, 12–18.
Martinez, C. A., & Williams, C. (2010). National institutions, enterpreneurship and global ICT adoption: A cross country test of competing theories. *Journal of Electronic Commerce Research*, *11*(1), 73–91.
May, J. W., Pyle, W., & Sommers, P. M. (2002). Does governance explain unofficial activity? *Applied Economics Letters*, *9*, 537–539.
Meade, N. (1985). Statistical forecasting. In: G. K. Rand, & R. W. Eglesee (Eds.), *Further developments in operational research* (pp. 57–75). Oxford: Pergamon.
Meade, N., & Islam, T. (2006). Modelling and forecasting the diffusion of innovation—a 25-year review. *International Journal of Forecasting*, *22*(3), 519–545.
Moe, T. M. (1984). The new economics of organization. *American Journal of Political Science*, *28*, 739–777.
Morgan, L., & Finnegan, P. (2007). *How perceptions of open source software influence adoption: An exploratory study*. Paper presented at the 15th European Conference Information Systems (ECIS 07), University of St. Gallen.
Netcraft Ltd. (2010). *The Apache web server survey*. Retrieved from 〈http://www.netcraft.com〉.
Nicoletti, G., & Scarpetta, S. (2003). Regulation, productivity and growth: OECD evidence. *Economic Policy*, *18*(36), 9–72.
North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge: Cambridge University Press.
OSS research portal (2010). Retrieved from 〈http://zerlot.cse.nd.edu〉.
Pagan, A. R., & Hall, D. (1983). Diagnostic tests as residual analysis. *Econometric Reviews*, *2*, 159–218.
Parsons, T. (1990). Prolegomena to a theory of social institutions. *American Sociological Review*, *55*, 319–339.
Peyret, H. (2005). *ObjectWeb: Trends for open source infrastructure.* Keynote session presented at fourth annual objectweb conference, Lyon, France. Retrieved from 〈https://wiki.ow2.org/ObjectWebCon05/Wiki.jsp?page=Sessions〉.
Press, L. (1997). Tracking the global diffusion of the Internet. *Communications of the ACM*, *40*(11), 11–18.
Rai, A., Ravichandran, T., & Samaddar, S. (1998). How to anticipate the Internet's global diffusion. *Communications of the ACM*, *41*(10), 97–106. 〈http://doi.acm.org/10.1145/286238.286253〉.
Rivera-Batiz, L. A., & Romer, P. M. (1991). Economic integration and endogenous growth. *Quarterly Journal of Economics (The MIT Press)*, *106*(2), 531–555.

Rogers, E. M. (1995). *Diffusion of innovations* (fourth Edition). New York: The Free Press.
Romer, P. M. (1986). Increasing returns and long run growth. *Journal of Political Economy*, *94*(5), 1002–1037.
Romer, P. M. (1994). The origins of endogenous growth. *Journal of Economic Perspectives*, *8*(1), 3–22.
Scott, W. R. (2001). *Institutions and organizations*. Thousand Oaks, CA: Sage.
Scott, W. R. (2004). Institutional theory. In: G. Ritzer (Ed.), *Encyclopedia of social theory*. Thousand Oaks, CA: Sage.
Solow, R. M. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics (The MIT Press)*, *70*(1), 65–94.
SourceForge.net (2010). Retrieved from 〈http://sourceforge.net/〉.
Staiger, D., & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*, 557–586.
Standing, C., & Vasudavan, T. (2000). *Diffusion of Internet technologies in travel agencies in Australia*. Paper presented at the 8th European Conference on Information Systems held in Austria.
Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business and Economic Statistics*, *20*, 518–529.
Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In: D. W. K. Andrews, & J. H. Stock (Eds.), *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg* (pp. 80–108). Cambridge: Cambridge University Press.
Talukdar, D., Sudhir, K., & Ainslie, A. (2002). Investigating new product diffusion across products and countries. *Marketing Science*, *21*(1), 97–114, doi:10.1287/mksc.21.1.97.161.
United Nations Public Administration Program (2011). Retrieved from 〈http://www2.unpan.org/egovkb/datacenter/CountryScore.aspx?ddl=4〉.
United Nations Statistics (2010). Retrieved from 〈http://data.un.org/〉.
Ven, K., Verelst, J., & Mannaert, H. (2008). Should you adopt open source software. *IEEE Software*, *25*(3), 54–59.
White, H. (1982). Instrumental variables regression with independent observations. *Econometrica*, *50*, 483–499.
Whitmore, A., Choi, N., & Arzrumtsyan, A. (2009). Open source software: The role of marketing in the diffusion of innovation. *Information Technology and Control*, *38*(2), 91–101.
Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge, Massachusetts: The MIT Press.
World Bank Indicators (2010). Retrieved from 〈http://data.worldbank.org/〉.
World Governance Indicators (2010). World Bank Group. Retrieved from 〈http://info.worldbank.org/governance/wgi/sc_country.asp〉.
Wu, D. M. (1973). Alternative tests of independence between stochastic regressors and disturbances. *Econometrica*, *41*, 733–750.
Yin, R. K. (2003). *Case study research: Design and methods* (3rd edition). Thousand Oaks, California: Sage Publications.
Zhu, K., & Kraemer, K. (2005). Post-adoption variations in usage and value of e-business by organizations: Cross-country evidence from the retail industry. *Information Systems Research*, *16*(1), 61–84.