How to Overcome the Digital Divide? The Determinants of Internet Diffusion

Shchetinin, Oleg and Baptiste, Massenot

Toulouse School of Economics, Toulouse School of Economics - GREMAQ

June 2008

How to Overcome the Digital Divide? The Determinants of Internet Diffusion *

Baptiste Massenot[†] Oleg Shchetinin [‡]

May 2008

Abstract

We document the existence and the persistence of the digital divide and investigate the determinants of the Internet diffusion in both developing and developed countries. Our study innovates on the following: i) we use a data set that covers more countries and years than the earlier studies ii) We use the GMM estimator which requires milder assumptions to be consistent than the traditionally used panel data estimators in technology diffusion studies.

We find that i) the digital divide is likely to persist over time, ii) the Internet diffusion process is dynamic which makes static estimators inconsistent, iii) Internet adoption starts later but goes faster in developing countries, iv) inflows of the foreign investment and better human capital boost the diffusion of Internet for the developing countries only and v) GDP per capita has a negative impact on Internet diffusion in the developing countries and a positive impact in developed countries. This last finding seems surprising but it is consistent with the conditional convergence hypothesis as well as with the resource curse theory.

1 Introduction

It is well documented now that access to Internet, personal computers and other information technologies is highly unequal in the world (see, e.g. [21] and [5]). Moreover, the gap between developed and developing countries is increasing over time (see [2]). This gives rise to the debate on the so called digital divide which refers to the information and communication technology gap. [1]

*We are grateful to Thierry Magnac for attention and support in our work and to the participants of the Seminar at Toulouse School of Economics for comments.

†Toulouse School of Economics and GREMAQ

‡Toulouse School of Economics and GREMAQ

[1] The literature points out that the digital divide exists at different levels. In particular, it distinguishes individual, organizational and global digital divide. As it follows from its name, the individual digital divide reflects the fact that different individuals use IT with different intensities due to geographical, sociological and economic characteristics. The digital divide at

Understanding the nature of the digital divide is an important policy concern of the international organizations as documented in [21] and [5]. Better access to Internet in the developing countries will for example enhance education opportunities through e-learning, improve the business environment for industries who could sell or advertise their products on the web, reduce drastically the communication costs etc. All these factors can also contribute to economic growth.

Related literature

There exists already a number of studies on Internet penetration. One of the first empirical studies was conducted by Hargittai (1999, [14]). It focuses on the Internet penetration in 18 OECD countries, using OLS regression for a cross-section of countries. The main finding of the paper is the importance of wealth and market competition for Internet diffusion. However, clear limitations of the paper are the limited size of the data set which includes only 18 countries but estimations include up to seven explanatory variables.

Susmita Dasgupta et al. [11] conducted an empirical study of Internet diffusion based on data for 44 countries. They use data for growth in Internet penetration from 1990 to 1997 in a cross-section of countries and estimate a model using OLS. They found that wealth is not statistically significant (on the contrary of the previous paper) and their significant factors are size of urban population and index of competition policy. However there are a number of differences between the two papers. The former deals with OECD countries only whereas the latter with both developed and developing countries. Secondly, the time periods differ. Third, the first paper is based on levels and the second one on growth data. The differences mentioned above suggest that comparison of the results should be made with a lot of care.

Sampsa Kiiski and Matti Pohjola [16] studied Internet penetration by using three frameworks. First, with only 23 OECD countries. They used OLS cross-country estimation based on growth between 1995 and 2000. Second, they added 37 developing countries. They provide both cross-country and panel data estimation. Third, they used a simultaneous equation framework by considering an equation for GDP and access cost (proxied by telephone access cost) in addition to an Internet diffusion equation. In accordance to the previous studies, the wealth level is found to be significant for Internet diffusion. Internet access cost have negative impact in OECD countries whereas they have no impact for a larger set of countries. However, in the simultaneous equation model they found a negative impact for OECD countries. Education appears to be significant for the larger set of countries but not for OECD countries.

Menzie D. Chinn and Robert W. Fairlie [9] studied the Internet penetration in 1999-2001 in 161 countries . They conducted a similar study for 2002-2004 [10]. They use panel data random effect estimation, i.e. GLS estimation. They

the organizational level reflects different uses of IT by firms. Finally, the global digital divide refers to cross-country differences. The critical survey of the three levels of digital divide is presented in Dewan and Riggins ([13]). Our paper focuses on the global digital divide.

found wealth is important for Internet penetration. Also important is the rule of law. Somewhat surprisingly, the share of urban population has a negative impact for the earlier periods.

Similar but more extensive research was conducted by Sanjeev Dewan et al. [12]. They use different measures of IT penetration - the usual measures in per capita terms and in per income terms. On top of this, they consider three generations of IT - mainframes, PCs and Internet. The time period is 1985-2001 which is much larger than in the previous papers. The paper compares the three generations of IT and concludes that Internet penetration has different dynamics compared to the two other IT considered, as the set of statistically significant regressors are different. For the Internet penetration, if per capita measures are used, phone line cost (usually considered as a proxy for Internet access cost) has a negative impact, wealth has a positive impact as well as years of schooling. When per-GDP measures are used, phone line density appears to have a positive impact, Internet access cost has a negative impact as before, wealth is no more significant, years of schooling still have a positive impact and trade in goods as share of GDP appears to have a positive impact.

All the previous studies focused on OECD countries only or on OECD and developing countries in one sample. However, since the nature of technology diffusion (including Internet) in developing countries is suggested to be different, it is worth considering developing countries in a separate sample. One such study was conducted by Scott Wallsten [22]. The paper focuses on regulation and institutional issues of Internet adoption in the developing countries. It found like almost all other papers a positive effect of wealth. The empirical approach is a cross-section estimate of panel data.

The literature contains a lot of studies on other information and communication technologies penetration. One of the first and largely cited was conducted by Francesco Caselli and Wilbur John Coleman ([8]). They consider a sample of 89 countries (both developed and developing) from 1970 to 1990. They restricted their data on 5-year gaps (i.e. 1970, 1975, etc.) because of limited data accessibility so that finally the panel includes only 6 observations per country. The estimation was made with GLS panel data random effect estimator. The estimation however does not include the lagged dependent variable. In some estimations the number of observations is very small.

The studies mentioned above are interesting by themselves but they often have contradictory results. Our opinion is that this is, at least partially, because of drawbacks in the estimation techniques. In this paper, we pay special attention to the right choice of the estimation method.

Our Contribution

First, we use a data set larger both in time interval and in the number of countries than the previous studies.

Second, we discriminate between developing and developed countries. Our results indeed suggest that the patterns of the Internet diffusion are different for the two groups of countries.

Third, we use the Generalized Method of Moments (GMM) estimator which is better designed to account for endogeneity problems. Although GMM is widely used in other areas of economics and is proved to be a powerful estimation tool[2], to our knowledge, it has not been used for IT penetration analysis so far. GMM may be considered as a generalization of the traditionally used OLS. The disadvantage of the OLS is that a number of restrictive assumptions have to be made to guarantee its consistency. However, these assumptions are often violated in practice which leads to the biased estimations. Indeed, for some of the variables, the coefficients' estimates found in the previous studies vary a lot. To some extent, this may be due to different samples used in the estimations. But if the model behind the technology diffusion process is the same, the argument is no more valid. Another explanation is inconsistency of some of the the estimations. The advantage of GMM it that it doesn't require these restrictive assumptions, so it is more flexible. By employing it, we introduce a contemporary econometric technique to the study of the information technology diffusion. This paper is just a first step in this direction.

Most of the papers mentioned in the literature review used static estimators despite the fact that the process is clearly dynamic. This leads to the omitted variable bias. To avoid it, the lag of the Internet diffusion has to be included into the model. We then use a dynamic estimator.

The endogeneity of some regressors may also be an issue and needs to be tested before concluding about the consistency of the estimates. We thus provide endogeneity tests for our regressors whenever we suspect them to be caused by Internet diffusion.

The rest of the paper is organized as follows. In the next section, we describe our data set. The third section presents evidence of the existence of a digital divide. The fourth section looks for the determinants of the digital divide using the adequate econometric tool and the fifth section presents consistency checks. Sixth and seventh sections respectively give interpretations of our findings and conclude by drawing policy implications.

2 Data

The literature on technology diffusion and particularly on Internet penetration suggests which channels influence the diffusion process.

First, there are economic factors which include country wealth and cost of technology adoption.

Second, human capital is important. Although the ability to read is probably the only necessary skill to use the Internet, installing and sustaining a network may require additional skills. On top of this, the need for communication as well as for the information search are important incentives to use Internet. These needs are closely related to the higher level of human capital.

[2]see, for example, [17] for the number of citations of the papers where the method is introduced

Third, openness can be important for the Internet diffusion, especially for the developing countries because the diffusion is initiated from the more advanced ones. It is driven by the imitation of the new technology rather then by developing the new technologies inside the developing countries. Openness may be measured by trade volumes as well as by financial flows.

Fourth, there is a *hardware* factor - the new technology may be based on other already existing technologies. In the case of Internet the preceding technology is personal computers.

Fifth, infrastructure factors such as access possibility may be important. For the Internet these are measured by the share of urban population, telephone lines accessibility etc.

In our investigation we use data from the World Development Indicators database [5]. The data are collected from different sources. The description of the data can be found in the Appendix A. The data set covers the period of 1991-2002 for 66 developing countries and 23 developed countries on a yearly basis. The two panels are unbalanced. The availability of the data is summarized in the Appendix A.

The most important variables are discussed below. Internet diffusion is measured by the number of Internet users per thousand inhabitants. Although this measure does not take into account the intensity of Internet use, there is no other data with comprehensive coverage. Country wealth is measured by GDP per capita in PPP terms in constant 2000 US dollars. We use the adult literacy rate as a measure of human capital. Country openness is measured by the share of gross Foreign Direct Investment (FDI) to GDP. Alternatively, we use trade openness (import plus export) as share of GDP. Computer penetration is measured as the number of personal computers per thousand inhabitants.

As FDI is highly fluctuating from year to year in a given country, we construct a smoother measure of openness. For each country we compute the linear trend of FDI over time and use these values as a measure of openness. This measure will be referred to as "smoothed FDI". The trend of FDI is a noise-refined measure of capital inflow in a country. It is used as a proxy to the imitation of new technology. As an illustration of the role of FDI in Internet diffusion, observe that substantial part of the Internet in South America was installed by the Spanish firm *Terra* [3] and a part of the Internet in Africa and Latin America was installed by the French firm *France Telecom* [4]. FDI can also be a proxy for institutional quality of a country as countries with high quality institutions are more likely to attract foreign investors.

A few illustrations are useful to see some features of the data. Figures 1 and 2 describe the evolution of the diffusion of Internet respectively in the Netherlands and in Brazil.

The example of the Netherlands represents the typical path for developed countries and Brazil is typical for developing countries. Both paths are S-shaped:

[3] Terra operates both as a web portal and/or an internet access provider in the US, Spain, and 16 Latin American countries.

[4] One of France Telecom's most important subsidiaries was Telecom Argentina. France telecom owns shares of many African Internet providers and is also active in Eastern Europe

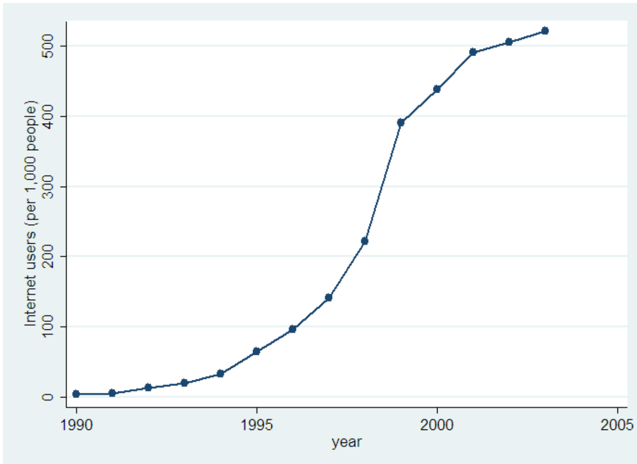

Figure 1: Internet diffusion in the Netherlands

a slow initial growth rate which increases afterwards, and finally slows down again as the diffusion process draws to completion. The Netherlands are close to the maximal diffusion level whereas Brazil just started its way on the S.These figures imply that there exists a digital divide and that the developing countries lag in the diffusion process.

However, we cannot infer from these observations that the digital divide is a serious policy issue. Indeed, it may well be that the developing countries will catch up within a few years without any need for policy intervention. The next section suggests though that the digital divide is likely to persist over time.

3 S-shaped dynamics

S-Shape Models

The literature typically uses two kinds of models for ICT diffusion: Gompertz and logistic. Both represent S-shaped diffusion paths. Below we use a slightly generalized version of these two models. [5]

The logistic model is described by the differential equation

$$\dot{y} = \frac{a}{c}(y - y_0)(c - (y - y_0)) \tag{1}$$

the solution of which is given by

$$y(t) = y_0 + \frac{c}{1 + e^{-a(t-t_0)}} \tag{2}$$

The parameters have the following interpretation:

[5] The paper [15] provides a pedagogical presentation of Gompertz and logistic equations generalizations and discusses limitations of *standard* equations in economic modeling.

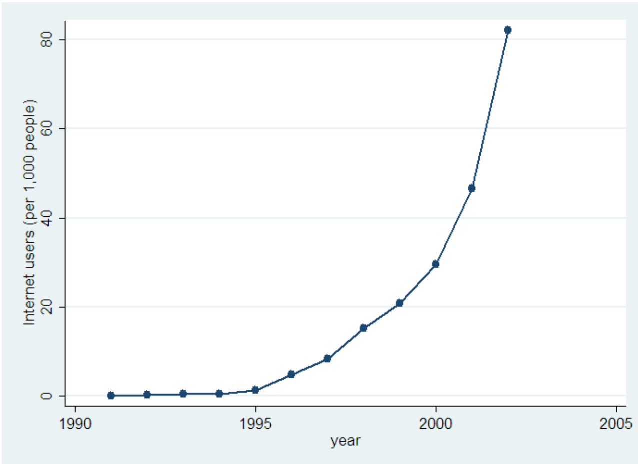

Figure 2: Internet diffusion in Brazil

- y_0 is the upper shifter of the solution. $c+y_0$ is the upper limit of the solution path or maximal penetration rate. $c + y_0 = \lim_{t\to\infty} y(t)$;
- a is the speed of convergence of $y(t)$ to its limit. It characterizes the curvature of the diffusion path or speed of diffusion;
- t_0 is the moment in time when $y(t_0) = \frac{c}{2} + y_0$. In other words, this is the moment in time when technology achieved half of its maximal level.[6]

With the change of variables $Y = y - y_0$, the differential equation 1 may be simplified to the standard form

$$\dot{Y} = \frac{a}{c} Y(c - Y) \tag{3}$$

The Gompertz model is described by the differential equation

$$\dot{y} = a(y - y_0)(\ln c - \ln(y - y_0)) \tag{4}$$

the solution of which is given by

$$y(t) = y_0 + ce^{-e^{-a(t-t_0)}} \tag{5}$$

The parameters have the following interpretation:

- y_0 is the upper shifter of the solution.
- $c+y_0$ is the upper limit of the solution path or maximal penetration rate. $c + y_0 = \lim_{t\to\infty} y(t)$;

[6] It is not exactly half if $y_0 \neq 0$. In this case it corresponds to half of the growth of the technology diffusion level.

- a is the speed of convergence of $y(t)$ to its limit. It characterizes the curvature of the diffusion path;
- t_0 is the moment in time when $y(t_0) = \frac{c}{e} + y_0$. In other words, when the technology achieved the share $\frac{1}{e}$ of its maximal level [7].

With the change of variables $Y = y - y_0$, the differential equation 4 may be simplified to the standard form

$$\dot{Y} = aY(\ln c - \ln Y) \tag{6}$$

The important feature of the Gompertz curve is that the diffusion goes faster at the beginning but becomes slower over time. This leads to a relatively short period of rapid expansion and to a relatively long period of gradual growth up to the maximal level. The logistic curve is more symmetric - the growth rate (measured as $\frac{\dot{y}}{y}$) is initially not as high as in the Gompertz curve and it declines more gradually (see [15]).

Notice that if the limit rate c in the two models is not constant (which is assumed in the diffusion models discussed below), solutions are no more given by (2) and (5) respectively.

Econometric models arise from equations (1) and (4) respectively. Assume that the maximal level $\log(c + y_0)$ is determined by the vector of economic parameters X: $\log(c + y_0) = f(X)$. Assume next, that the function f is linear: $f(X) = X\beta$. Then the empirical equation for country i at time t for the Gompertz model is

$$\log(y_{it}) - \log(y_{it-1}) = a(X_{it}\beta - \log y_{it-1})$$

which gives

$$\log y_{it} = (1 - a)\log y_{it-1} + X_{it} \cdot a\beta \tag{7}$$

The logistic model leads to non-linearity in econometric equations since the differential equation 1 may be rewritten either as $\frac{\dot{Y}}{Y} = (a/c)(c - Y)$ which implies the use of $\ln Y$ and Y simultaneously or as $\dot{Y} = (a/c)Y(c - Y)$ so that Y_t^2 or $Y_t Y_{t-1}$ appears in the econometric equation.

Moreover, the maximal diffusion level c enters twice into the equation which complicates the interpretation of the estimated coefficients. To sum up, estimation based on the Gompertz model has advantages as it is linear and simpler to interpret.

S-Shape Estimations

To demonstrate the very existence of the gap between developed and developing countries in Internet diffusion, we estimate equations which describe the diffusion paths for the developed and developing countries over time. A deeper analysis of the economic determinants of the diffusion process and of the differences between the two groups of countries is provided in the next section.

[7] see footnote 6

Estimations made here do not provide any economic insights of why the gap exists. They rather highlight the difference of the patterns of the Internet diffusion between the two groups of countries and provide some measures of the gap more appropriate than a simple comparison of levels at a given point in time.

The logistic path is given by

$$y(t) = y_0 + \frac{c}{1 + e^{-a(t-t_0)}}$$

The Gompertz path is given by

$$y(t) = y_0 + ce^{-e^{-a(t-t_0)}}$$

Estimations for the models with vertical shifters are given in the Appendix B. Since the vertical shifter in all the estimations is insignificant, it may be dropped and we can reestimate the models consistently. The results of these estimations are presented below.

Table 1: S-shaped dynamics estimation

	Logistic Model		Gompertz Model	
	Developing	Developed	Developing	Developed
	Dependent variable - Number of Internet users			
c	196.122***	453.338***	459.037	538.761***
	(3.693)	(18.246)	(1.179)	(9.594)
a	0.588***	0.655***	0.187**	0.335***
	(5.137)	(9.147)	(2.202)	(6.247)
t_0	2001.686***	1998.752***	2004.057***	1998.271***
	(1999.193)	(7794.760)	(496.052)	(5021.698)
N	831	324	831	324
R2	0.466	0.874	0.466	0.873
R2 adj.	0.464	0.873	0.464	0.872
AIC	9242.915	3789.265	9242.727	3792.132
BIC	9257.083	3800.607	9256.895	3803.474

*R*obust t-statistics in parenthesis
*Significance level: * - 10%, ** - 5%, *** - 1%*

Dropping the vertical shifter changes the quantitative results slightly. A number of observations arise from these results. First, the difference in estimates of t_0 suggests that there is a lag of approximately 3 years (if based on the logistic model) between developing and developed countries. Second, developing countries are not simply behind, they go to a lower maximal level of the Internet diffusion as indicated by the different estimates of c. Third, an important result is that the speed of convergence to the maximal level estimated by a is lower for the developing countries.

Although the estimations presented are rough and naive, they may be considered as an indication that something has to change in developing countries if they want to overcome the gap in Internet diffusion.

The insignificant coefficients on c in the Gompertz model is attributed to the outliers. It reflects the fact that the developing countries are much more dispersed than the developed ones.

The next section proceeds with more rigorous econometric and more meaningful economic analysis aiming at pointing out which factors are responsible for such a gap in the Internet diffusion.

4 Empirical Strategy

We estimate the model of the form

$$\log(I_{it}) = \alpha \log(I_{it-1}) + X_{it}\gamma + d_t + \varepsilon_{it}$$

where the error term ε_{it} is

$$\varepsilon_{it} = \eta_i + u_{it}$$

and η_i is the country effect of country i, u_{it} the idiosyncratic error and d_t are time dummies.

The coefficient of the empirical model are related to the parameters of the Gompertz model in the following way:

$$\alpha = 1 - a$$

$$\gamma = a\beta$$

The vector of regressors X includes

- log(GDP per capita)
- Gross FDI as share of GDP, smoothed be linear trend over time for each country
- literacy rate
- log(number of PC per capita)
- other variables (see below)

Some of the regressors X_{it} may be endogenous and we provide tests for that. Potentially, there may be endogeneity in GDP. But it may be rather predetermined than endogenous because the current level of technology development may influence future production within the country even if this effect can be assumed to be very small. And even if there is reverse causality, it is lagged in time. Also, FDI may be endogenous. But as for GDP, the influence from Internet penetration in the country to FDI is also probably lagged in time and very small. All these exogeneity assumptions for these variables will be tested in a further section.

4.1 Assumptions

We make the usual assumptions on the error component:

$$E[\eta_i] = E[u_{it}] = E[\eta_i u_{it}] = 0$$

We do not require idiosyncratic errors to be *iid*. We allow for any autocorrelation of u_{it} within country. But we assume independence of errors between countries:

$$E[u_i u_j] = 0$$

where $u_i = (u_{i1}, ...u_{iT})^T$.

We do not assume orthogonality of the country effects η_i to the regressors. This assumption is required for consistency of OLS-based panel data estimators.

The main assumption to obtain moment conditions is exogeneity:

$$E[X_{it}\varepsilon_{is}] = 0$$

for all s, t.

4.2 Estimation Strategy

In our estimation we follow the Arellano-Bond dynamic panel estimation logic. The Arellano-Bond estimator was introduced in Arellano and Bond (1991, [3]) and then developed in Arellano and Bond (1995, [4]) and Blundell and Bond (1998, [7]). The estimator was then used in various studies. The estimation of the dynamic panel data is discussed in deep details in a very instructive paper [18].

To avoid possible endogeneity problems related to the correlation between country effects η_i and X_{it}, we take the first difference of the initial equation and obtain

$$\Delta \log(I_{it}) = \alpha \Delta \log(I_{it-1}) + \Delta X_{it}\gamma + (d_t - d_{t-1}) + \Delta u_{it} \tag{8}$$

This is the main equation to estimate. The well-known issue with this equation is endogeneity of the regressor $(\log(I_{it-1}) - \log(I_{it-2}))$ which is solved by instrumenting it with its lags $\log(I_{it'})$ starting from $t-2$: $t' = t-2, t-3, ..., t-p$. If p is too large we may face *weak instrument* as well as *too many instrument* problems. Both of them make the estimations less stable. We try different values of p in order to check what is the influence of increasing the number of instruments.

We also take care of the potential endogeneity of the other regressors. We provide Hansen and difference-in-Hansen tests for endogeneity (i.e. we test for relevance of the moment conditions). We also test for the autocorrelation in error terms by using the test developed by Arellano and Bond [3].

Instrumenting one of the variables X^j means excluding the moment condition $E\left[\Delta X_{it}^j \Delta u_{it}\right] = 0$ from the set of moment conditions and replacing it with the set of similar moment conditions $E\left[X_{it'}^{j}\Delta u_{it}\right]=0$ with $t'=t-2,t-3,....$ If regressors ΔX_{it} are in fact exogenous, this worsens the quality of estimation since the excluded moment condition is more informative. So, it is preferable to use variables themselves rather than instruments if there is no evidence for endogeneity.

4.3 Moment Conditions

The moment conditions to construct the GMM estimator are:

$$E\left[\log\left(I_{it-j}\right)\Delta u_{it}\right]=0, j=2,...,p, \tag{9}$$

with various values of p. We report the results for $p=2$ (this gives the same estimator as in Anderson-Hsiao (1981, [1]), $p=3$ and $p=T-2$ (i.e. the maximal p which gives the Arellano-Bond estimator). We try more values but all the phenomena we found are proved with these three values of p. We use *collapsed* instruments. [8]

Let the set of variables X consist of strictly exogenous variables $X^{(1)}$ and possibly endogenous variables $X^{(2)}$. For the first set we use IV-style moment conditions:

$$E\left[\Delta X_{it}^{(1)}\Delta u_{it}\right]=0$$

We also use the set of possibly endogenous variables as IV-style instruments, as if they were exogenous and then we test these moment conditions.

$$E\left[\Delta X_{it}^{(2)}\Delta u_{it}\right]=0$$

When tested, a potentially endogenous variable in the moment condition is replaced by its *instrument*, the second lag of the variable and this new moment condition is used in the estimation. The new estimate is used in a difference-in-hansen test for endogeneity.

4.4 Empirical Implication

Let Ω be the covariance matrix of the first differenced errors, H is its a priori estimate. It has block-diagonal structure and

$$H=I_N\otimes\begin{bmatrix}2 & -1 & & \\ -1 & 2 & -1 & \\ & -1 & 2 & \ddots \\ & & \ddots & \ddots\end{bmatrix} \tag{10}$$

[8] For a discussion on collapsed vs uncollapsed instruments, see [18]. Collapsing instruments means decreasing their number. This may improve the quality of estimation, especially with small samples - see [20],[23].

This estimate assumes homoscedasticity in non-differenced idiosyncratic errors and absence of serial correlation within individuals. The robust 1-step estimator relaxes these assumptions.

One can also set $H = I$ but this assumes homoscedasticity in differenced errors which is not the case. However, the GMM estimate based on this assumption is still consistent though not efficient.

Empirical implication of GMM leads to the estimator

$$\widehat{\beta} = (X'ZAZ'X)^{-1} X'ZAZ'Y,$$

where A is the covariance matrix of moments, Z is the matrix of instruments, X is the matrix of regressors.

For the 1-step GMM, it is equal to

$$A = (Z'HZ)^{-1},$$

for the 2-step GMM, it is equal to

$$A = \left(Z'\widehat{\Omega}Z\right)^{-1},$$

where $\widehat{\Omega}$ is the estimate of Ω based on the first step estimate. The well known issue with the 2-step GMM estimator is severe downward bias for the estimate of standard errors, so that too many regressors seem to be significant. There exists a correction for the 2-step estimator suggested by Windmeijer [23]. This correction gives more precise estimations for the standard errors.

So, in addition to different number of lags used for instrumenting first difference of $\log I_{it}$, i.e. different sets of moment conditions, we use different types of empirical implications of these moment conditions.

5 Empirical Results

This section first presents the empirical results for the sample of developing countries. We report and discuss equation estimation and relevant statistical tests. Then the results for the sample of developed countries are introduced.

For the developing countries the estimation results for the 1 step robust GMM and 2 step GMM can be found in Appendix C.

The main estimation results are presented in Table 5. Both the results for 1 step and 2 step GMM are included. We present results with different number of lags used to instrument the lagged first difference of Internet diffusion which is definitely endogenous.

Comparison of the 1-step GMM estimations shows that the magnitudes and the significance of the coefficients on lag of Internet and GDP per capita change slightly over different variants of estimation. The coefficient estimate for the lagged value of Internet users varies between 0.5 and 0.53. Surprisingly, the coefficient on $\log(GDP\ pc)$ is negative and varies between -1.3 and -1.33. Literacy rate is significant in all 1 step GMM estimations but it loses its significance

Table 2: Estimation results

	1 step robust GMM			2 step GMM	
	$p = 2$	$p = 3$	$p = \max$	$p = 3^9$	$p = \max$
	Dependent variable - log(*Internet users*)				
L.log_Internet	0.505***	0.527***	0.535***	0.522***	0.532***
	(7.588)	(8.114)	(8.180)	(8.070)	(8.891)
log_GDP_pc	-1.300**	-1.334**	-1.317**	-1.330**	-1.474***
	(-2.554)	(-2.558)	(-2.496)	(-2.551)	(-2.908)
FDI_gross_hat	0.057*	0.061*	0.058*	0.058*	0.058*
	(1.729)	(1.821)	(1.719)	(1.731)	(1.783)
literacy_rate	0.025*	0.027*	0.033*	0.024	0.025
	(1.702)	(1.756)	(1.973)	(1.602)	(1.584)
Year dummies	YES	YES	YES	YES	YES
Observations	453	453	453	453	453
Number of id	66	66	66	66	66
F-statistic	433.364	451.434	455.482	452.446	615.531
sigma2	0.134	0.138	0.140	0.137	0.141
instrum. number	15	16	24	16	24
AR1 test	-3.911	-3.848	-3.861	-3.872	-3.962
AR2 test	-0.899	-0.932	-0.945	-0.916	-0.930

Robust t-statistics in parenthesis
*Significance level: * - 10%, ** - 5%, *** - 1%*

in 2 step GMM except in the exact identification case with only 1 lag employed as instrument. But this coincides with the 1 step GMM. The magnitude of the coefficient is around 0.03. Finally, gross FDI is significant when the number of instruments is not too high, i.e. for estimations with collapsed instruments. It keeps significance for a small number of lags used as instruments for estimations with non-collapsed instruments. Somewhat unexpectedly, the number of personal computers per capita is always insignificant with very low t-statistics. It is not included in the regressions but some regressions with it are reported in the robustness check results.

These results serve as an illustration of the *too many instrument* problem. In many regressions the magnitude of the coefficients changes considerably for non-collapsed instruments for a large number of lags p.

The 2 step GMM is known to bias downward standard errors, or equivalently, to inflate t-statistics. To correct for small sample size and to capture non-linearity in minimized sum of empirical moments, Windmeijer correction for variance-covariance matrix is used. This doesn't change coefficient estimates. The only change is in standard error estimations and related statistics for the whole regression.

The empirical results obtained in the paper illustrate the inflation property of 2 step GMM. In particular, comparison of the estimates with Windmeijer correction with those of uncorrected 2 step GMM shows that t-statistics are higher in the latter case. Moreover, the inflation magnitude is greater the higher is the number of instruments. The result is absolutely expected.

Economic intuition and possible explanation of the results obtained are discussed in section 6. The following section proceeds with a number of tests.

Tests for Exogeneity

To test the validity of the moment conditions, we provide Hansen tests which may be considered as checks for the joint validity of the set of instruments. We also use difference-in-Hansen tests to check the validity of each regressor. Despite the well-known drawbacks of these tests they are generally employed in empirical studies. The test does not depend on the type of estimation, i.e. whether 1 step robust or 2 step estimator was used because it assumes efficient estimations. Tests differ when the number of lags used in (9) varies. A set of tests corresponding to different numbers of lags is reported in the Appendix D in Table D. All the joint hansen tests do not reject exogeneity.

We also provide individual exogeneity tests for all the variables of interest. These are difference-in-Hansen tests. The are reported in the Appendix D in table D. These tests do not reject exogeneity of the three variables: log_GDP_pc, smoothed FDI and literacy rate with a small number of instruments.

Tests for Autocorrelation

We used standard Arellano-Bond test for autocorrelation in errors. As for endogeneity tests, we went through a series of tests. Not surprisingly, there is residual autocorrelation of first order. We should focus on second order autocorrelation. In a series of model specifications we considered, second order autocorrelation in residuals is rejected (see "AR1 test" and "AR2 test" line in the tables in Appendix C).

Robustness Checks

We tried a number of other variables including PC, FDI in level, share of urban population, number of phone lines, number of phones or openness as measured by the sum of imports and exports. However, they appear to be insignificant. In Appendix E, we report the 1 step and 2 step GMM estimates with the maximal number of lags of the dependent variable used as instruments. A lower number of lags gives similar results.

Estimation for the Developed Countries

All the estimates were repeated for the developed countries. Since the literacy rate varies only slightly within this sample, it is not included. Ireland was excluded from the sample because it has a gross FDI much higher than the rest of the countries and creates an outlier bias in the estimations. Only 2 step GMM with the Windmeijer correction are reported in the Appendix F as other estimation technics give similar results. The only significant variables are the lagged value of Internet users and log(*GDP per capita*), which now has a positive sign.

6 Discussion

The most surprising result we obtained is the negative influence of wealth on the level of Internet diffusion for the developing countries. This strikingly differs from the previous studies. To interpret this coefficient, remember that it is related to the Gompertz model parameters and determinants of the maximal Internet penetration level c. A negative sign means that the richer the countries, the lower the level of Internet penetration they converge to, holding other parameters (literacy rate and FDI) constant. This may be supportive for the conditional convergence hypothesis within the group of developing countries. Recall that the model is specified in such a way that the actual regressors are GDP growth rather than levels. Under the conditional convergence hypothesis, GDP is negatively correlated to GDP growth, more especially in a sample including developing countries. This hypothesis may explain the negative impact of GDP on Internet for developing countries and positive for developed countries.

The result is also consistent with the resource curse theory (see [19]). The richest of the developing countries generally have resource-oriented economies. According to this theory, these countries would be less democratic and the other sectors of their economy should be underdeveloped. Less democratic countries are more likely to restrict the use of Internet as it is done e.g. in China. Notice also that economies oriented towards natural resources should be less Internet-intensive than economies oriented towards services. As a consequence, the resource-intensive countries should be less interested in Internet.

We found that literacy rate (as a proxy for human capital) is important for the developing countries but has no impact for the developed ones. This is a consequence of the fact that the developed countries have almost one hundred percent literacy rate, so there is no variation between them and the coefficient cannot appear to be significant in the regression. On top of this, adult literacy rate may not be the most relevant measure of human capital for the developed countries. The human capital in the developed countries is qualitatively different from developing countries and should rather be measured with e.g. years of higher education.

Smoothed FDI was found to be significant for the developing countries and insignificant for the developed ones. The positive impact of FDI has a natural interpretation. Foreign capital flows into those developing countries which are most attractive, i.e. with good infrastructure and institutions, and then these countries are more likely to benefit from more advanced technologies.

The nature of the Internet penetration is different for the developed countries as illustrated by the insignificance of this variable in our regressions. In these countries Internet penetrates as a result of their own R&D activities rather than by adopting technologies from other countries. Our results are in line with the innovation-imitation theory of new technologies development (see [6], chapter 8). Finally, FDI may be a proxy for the quality of institutions in developing countries whereas it is less likely to be so in the developed world as investors are more likely to invest in countries with high quality institutions. So, the significance of FDI may indicate the significance of institutions for Internet adoption in developing countries.

Somewhat surprisingly, we found that the number of personal computers has no impact for the Internet diffusion for both groups of countries. This may be a consequence of the fact that the use of computers is more widely spread than the use of Internet and as a consequence Internet use is not bound by the the possession of a computer. Also, our study focuses on Internet users. In developing countries, Internet users are more likely to use Internet cafés, which makes the number of Internet users more independent of the number of computers.

This may be explained by the fact that Internet diffusion starts after PC adoption and at the moment when Internet diffusion starts the PCs are already abundant, consequently access to the PC is not a problem for those wanting to use Internet. In other words, the number of computers is usually high enough not to slow down the penetration of the Internet. In the process of internet adoption the number of PCs continue to be high enough and doesn't limit the penetration of the Internet.

Last, we found that the growth rate has a stronger autocorrelation in the developed countries. This means that the adoption process in the developing countries is determined by the inner dynamics of the process to a lower degree and there is more room for influencing the diffusion process through policies. As suggested by our study, these policies should be oriented towards creating better infrastructure and institutions to attract foreign investment and improve human capital quality.

7 Conclusion

The paper studies the Internet diffusion in the developing countries and compares it with the developed ones. The paper discusses two types of models: logistic and Gompertz. We confirm the existence of a digital divide between developing and developed countries. We use an empirical model based on Gompertz model to reveal the driving forces of the Internet penetration. We use a dynamic panel GMM estimator (Arellano-Bond estimator) which requires less assumptions to be consistent compared to other estimators. The paper argues this estimator is consistent whereas static panel estimators mostly used in the literature lead to an omitted variable bias. We use a large enough data set which improves the quality of our estimations.

We find, first, that the diffusion process is dynamic. Moreover, the growth rate of Internet diffusion has higher autocorrelation for the developed countries than for the developing ones. This means that the diffusion in the developing countries is determined by the inner dynamics of the process to a lower degree and as a consequence policies can play a greater role in influencing the Internet adoption in developing countries.

Second, the inflow of foreign investments boosts the diffusion for the developing countries only, which is consistent with the innovation-imitation theory of technology adoption. Put differently, the developed countries rather invent technologies whereas the developing ones adopt them. We also interpret this result as a positive impact of high quality institutions on Internet diffusion.

Then, better human capital is found to be conducive for the Internet diffusion in the developing countries only.

Finally, and surprisingly, the wealth of a country has a positive impact in the developed countries and a negative impact in the developing countries. This empirical result is puzzling but we proposed two possible explanations. First, the finding is in line with the conditional convergence hypothesis which implies a negative correlation between GDP and GDP growth. Another possible explanation is related to the resource curse theory. Into the group of developing countries, some of the richest ones have resource-oriented economies which according to the resource curse theory makes them more likely to have poor institutions and an under-developed manufacturing sector. This makes these countries less interested in Internet.

References

[1] T.W. Anderson and C. Hsiao. *Formulation and Estimation of Dynamic Models Using Panel Data.* Institute for Policy Analysis, University of Toronto, 1981.

[2] L. Andres, D. Cuberes, M.A. Diouf, and T. Serebrisky. Diffusion of the Internet: A Cross-Country Analysis. *World*, 2007.

[3] M. Arellano and S. Bond. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*, 58(2):277–297, 1991.

[4] M. Arellano and O. Bover. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics*, 68(1):29–51, 1995.

[5] The World Bank. *World development indicators.* World Bank, 2004.

[6] R.J. Barro and X. Sala-i Martin. *Economic growth.* MIT Press, 2004.

[7] R. Blundell and S. Bond. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, 87(1):115–143, 1998.

[8] Francesco Caselli and Wilbur John Coleman II. Cross-country technology diffusion: The case of computers. Nber working papers, National Bureau of Economic Research, Inc, February 2001.

[9] M.D. Chinn and R.W. Fairlie. The determinants of the global digital divide: a cross-country analysis of computer and internet penetration. *Oxford Economic Papers*, 2006.

[10] Menzie D. Chinn and Robert W. Fairlie. Ict use in the developing world: An analysis of differences in computer and internet penetration. Nber working papers, National Bureau of Economic Research, Inc, July 2006.

[11] Susmita Dasgupta, Somik Lall, and david Wheeler. Policy Reform, Economic Growth, and the Digital Divide: An Econometric Analysis. Working paper, Developement Reserach Group, World Bank, 2001.

[12] Sanjeev Dewan, Dale Ganley, and Kenneth L. Kraemer. Across the Digital Divide: A Cross-Country Analysis of the Determinants of IT Penetration. *Journal of the Association of Information Systems*, 2005.

[13] Sanjeev Dewan and Frederick J. Riggins. The digital divide: Current and future research directions. *Journal of Association for Information Systems*, 6(2):298–337, 2005.

[14] Eszter Hargittai. Weaving the Western Web: explaining differences in Internet connectivity among OECD countries. *Telecommunications Policy*, 23(10-11):701–718, 1999.

[15] G. Jarne, J. Sanchez-Choliz, and F. Fatas-Villafranca. S-shaped curves in economic growth. A theoretical contribution and an application*. *Evolutionary and Institutional Economics Review*, 3(2):239–259, 2007.

[16] Sampsa Kiiski and Matti Pohjola. Cross-country diffusion of the internet. *Information Economics and Policy*, 14(2):297–310, 2002.

[17] D. Roodman. A Short Note on the Theme of Too Many Instruments. *Center for Global Development*, 2007.

[18] David Roodman. How to do xtabond2: An introduction to difference and system gmm in stata. *Center for Global Development*, (Working Paper No. 103), 2006.

[19] J.D. Sachs and A.M. Warner. The curse of natural resources. *European Economic Review*, 45(4-6):827–838, 2001.

[20] JD Sargan. The Estimation of Economic Relationships using Instrumental Variables. *Econometrica*, 26(3):393–415, 1958.

[21] UNCTAD. *The Digital Divide Report: ICT Diffusion Index*. United Nations Publications, 2005.

[22] S.J. Wallsten. *Regulation and Internet Use in Developing Countries*. World Bank, Development Research Group, Investment Climate, 2003.

[23] F. Windmeijer. A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126(1):25–51, 2005.

Appendix

A Data

All the data are taken from the World Development Indicators database of the World Bank.

Table 3: Data description

Variable name	Description
GDP_pc	GDP per capita in constant 2000 US dollars
PC	Number of personal computers per 1000 people
Internet users	Number of Internet users per 1000 people
Literacy rate	Adult literacy rate - % of people ages 15 and above
FDI_gross	Foreign Gross Direct Investment as share of GDP, %
Trade openness	Sum of country export and import as share of GDP, %
Urban population	Urban population, % of total
Phonelines	Telephone mainlines (per 1000 people)
Phones	Fixed line and mobile phones subscribers (per 1000 people)

Table 4: Data availability for developing countries

year	Internet users	GDP_pc	FDI_gross	literacy rate	PC
	Number of countries with available observations				
1991	5	80	61	69	34
1992	14	80	63	69	36
1993	24	81	70	69	42
1994	43	80	70	68	49
1995	72	82	71	68	56
1996	86	83	72	71	60
1997	88	83	73	68	64
1998	88	83	74	67	71
1999	87	82	73	65	74
2000	88	83	71	66	76
2001	86	81	70	64	77
2002	86	80	70	60	75
2003	52	77	60	11	14

Table 5: Data availability for developed countries

year	Internet users	GDP_pc	FDI_gross	PC
	Number of countries with available observations			
1991	21	22	22	22
1992-1999	22	22	22	22
2000	24	22	22	22
2001	23	22	22	22
2002	23	22	22	22
2003	14	22	22	3

B S-shaped dynamics estimation

Table 6: Estimation of the S-shaped dynamics with logistic and Gompertz models augmented by vertical shifter

	Logistic Model		Gompertz Model	
	Developing	Developed	Developing	Developed
	Dependent variable - Number of Internet users			
y_0	-1.839	2.176	1.020	9.908
	(-0.278)	(0.220)	(0.183)	(1.206)
c	210.394**	448.787***	411.363	505.526***
	(2.444)	(14.156)	(1.053)	(9.498)
a	0.549***	0.668***	0.201*	0.371***
	(3.114)	(7.059)	(1.720)	(5.748)
t_0	2001.885***	1998.751***	2003.557***	1998.249***
	(1405.026)	(7968.358)	(468.666)	(5864.031)
N	831	324	831	324
R2	0.307	0.775	0.307	0.774
R2 adj.	0.305	0.773	0.305	0.772
AIC	9244.825	3791.217	9244.696	3792.839
BIC	9263.716	3806.340	9263.586	3807.962

t-statistics in parenthesis

*Significance level: * - 10%, ** - 5%, *** - 1%*

C Dynamic Model estimation results

Table 7: 1 step robust GMM

	collapsed instruments			non-collapsed instruments		
	$p = 2$	$p = 3$	$p = \max$	$p = 2$	$p = 3$	$p = \max$
	Dependent variable - log_(*Internet users*)					
L.log_Internet	0.505***	0.527***	0.535***	0.513***	0.502***	0.522***
	(7.588)	(8.114)	(8.180)	(7.774)	(7.781)	(7.716)
log_GDP_pc	-1.300**	-1.334**	-1.317**	-1.300**	-1.327**	-1.020*
	(-2.554)	(-2.558)	(-2.496)	(-2.477)	(-2.542)	(-1.720)
FDI_gross_hat	0.057*	0.061*	0.058*	0.060*	0.053	0.036
	(1.729)	(1.821)	(1.719)	(1.776)	(1.611)	(1.071)
literacy_rate	0.025*	0.027*	0.033*	0.040**	0.039**	0.028*
	(1.702)	(1.756)	(1.973)	(2.187)	(2.170)	(1.709)
Year dummies	YES	YES	YES	YES	YES	YES
Observations	453	453	453	453	453	453
Number of id	66	66	66	66	66	66
F-statistic	433.364	451.434	455.482	448.203	465.759	353.312
sigma2	0.134	0.138	0.140	0.136	0.134	0.138
instrum. number	15	16	24	25	35	74
AR1 test	-3.911	-3.848	-3.861	-3.881	-3.743	-3.653
AR2 test	-0.899	-0.932	-0.945	-0.915	-0.938	-0.979

*R*obust t-statistics in parenthesis

*Significance level: * - 10%, ** - 5%, *** - 1%*

Table 8: 2 step GMM

	collapsed instruments			non-collapsed instruments		
	$p = 2$	$p = 3$	$p = \max$	$p = 2$	$p = 3$	$p = \max$
	Dependent variable - log_(*Internet users*)					
L.log_Internet	0.505***	0.522***	0.532***	0.512***	0.480***	0.478***
	(7.588)	(8.070)	(8.891)	(9.327)	(10.788)	(28.180)
log_GDP_pc	-1.300**	-1.330**	-1.474***	-1.319***	-1.346***	-0.837***
	(-2.554)	(-2.551)	(-2.908)	(-2.723)	(-3.088)	(-4.942)
FDI_gross_hat	0.057*	0.058*	0.058*	0.060*	0.048*	0.007
	(1.729)	(1.731)	(1.783)	(1.874)	(1.702)	(0.525)
literacy_rate	0.025*	0.024	0.025	0.027	0.027	0.029***
	(1.702)	(1.602)	(1.584)	(1.562)	(1.633)	(2.781)
Year dummies	YES	YES	YES	YES	YES	YES
Observations	453	453	453	453	453	453
Number of id	66	66	66	66	66	66
F-statistic	433.364	452.446	615.531	515.701	1000.435	530675.538
sigma2	0.134	0.137	0.141	0.137	0.132	0.130
instrum. number	15	16	24	25	35	74
AR1 test	-3.911	-3.872	-3.962	-4.206	-4.185	-4.456
AR2 test	-0.899	-0.916	-0.930	-0.810	-0.815	-0.877

Table 9: 2-step GMM with Windmeijer correction

	collapsed instruments			non-collapsed instruments		
	$p = 2$	$p = 3$	$p = \max$	$p = 2$	$p = 3$	$p = \max$
	Dependent variable - log_(*Internet users*)					
L.log_Internet	0.505***	0.522***	0.532***	0.512***	0.480***	0.478***
	(7.588)	(7.624)	(5.485)	(6.599)	(7.555)	(8.130)
log_GDP_pc	-1.300**	-1.330**	-1.474**	-1.319**	-1.346**	-0.837
	(-2.554)	(-2.471)	(-2.032)	(-2.093)	(-2.252)	(-1.566)
FDI_gross_hat	0.057*	0.058*	0.058	0.060	0.048	0.007
	(1.729)	(1.698)	(1.312)	(1.467)	(1.253)	(0.191)
literacy_rate	0.025*	0.024	0.025	0.027	0.027	0.029
	(1.702)	(1.557)	(1.427)	(1.417)	(1.059)	(1.426)
Year dummies	YES	YES	YES	YES	YES	YES
Observations	453	453	453	453	453	453
Number of id	66	66	66	66	66	66
F-statistic	433.364	441.873	377.226	357.755	420.335	366.065
sigma2	0.134	0.137	0.141	0.137	0.132	0.130
instrum. number	15	16	24	25	35	74
AR1 test	-3.911	-3.819	-3.438	-3.812	-3.848	-3.927
AR2 test	-0.899	-0.916	-0.926	-0.808	-0.813	-0.875

*R*obust t-statistics in parenthesis

*Significance level: * - 10%, ** - 5%, *** - 1%*

D Tests of the model

Table 10: Hansen tests - overall exogeneity tests

	collapsed instruments		non-collapsed instruments		
	$p = 3$	$p = \max$	$p = 2$	$p = 3$	$p = \max$
Hansen	0.739	14.286	11.586	17.920	45.276
Hansen prob.	0.390	0.113	0.314	0.593	0.906

Table 11: Difference-in-Hansen tests - Individual exogeneity tests

	log_GDP_pc	literacy_rate	FDI_gross_hat
collapsed instruments			
$p = 3$			
Difference in Hansen	.73872017	.73872017	.73872017
Difference in Hansen p	.39007116	.39007116	.39007116
$p = 5$			
Difference in Hansen	2.5263493	4.7797306	6.0094082
Difference in Hansen p	.11195898	.02879663	.0142298
$p = \max$			
Difference in Hansen	.00315476	1.011582	5.8463286
Difference in Hansen p	.95520858	.31452414	.01560958
non-collapsed instruments			
$p = 2$			
Difference in Hansen	.14278453	.77429345	5.4435393
Difference in Hansen p	.70552836	.37889211	.01964083
$p = 3$			
Difference in Hansen	.52115527	2.9717194	3.0484073
Difference in Hansen p	.47034929	.08473176	.08081636
$p = 5$			
Difference in Hansen	3.6821048	4.9844604	.54532198
Difference in Hansen p	.05499938	.02557596	.46023546
$p = \max$			
Difference in Hansen	-1.5916264	.75307947	-1.6325997
Difference in Hansen p	1	.385503	1

E Robustness check

Table 12: Robustness check with 1-step robust GMM estimation

	Dependent variable - log(*Internet users*)					
L.log_Internet	0.540***	0.467***	0.540***	0.537***	0.533***	0.537***
	(7.79)	(7.18)	(8.02)	(8.24)	(8.17)	(8.20)
log_GDP_pc	-1.574**	-1.181**	-1.336**	-1.346**	-1.291**	-1.313**
	(-2.56)	(-2.19)	(-2.56)	(-2.41)	(-2.42)	(-2.50)
literacy_rate	0.027*	0.019	0.028*	0.033*	0.032*	0.033*
	(1.70)	(1.35)	(1.80)	(1.88)	(1.88)	(1.98)
FDI_gross_hat	0.069**		0.073**	0.060*	0.064*	0.066*
	(2.14)		(2.01)	(1.74)	(1.80)	(1.71)
log_PC	0.044	0.028				
	(0.19)	(0.12)				
trade_open						0.023
						(0.059)
urban_population					0.021	
					(0.48)	
log_phonelines				0.028		
				(0.090)		
log_phones			0.206			
			(1.19)			
FDI_gross		0.003				
		(0.57)				
Observations	411	392	450	452	453	449
Number of id	63	61	66	66	66	66
F-statistic	493.8	507.1	384.3	413.2	430.4	408.9
hansen	10.46	8.973	15.11	14.83	14.24	13.64
hansen prob.	0.315	0.440	0.0880	0.0956	0.114	0.136
instrum. number	24	24	25	25	25	25
AR1 test	-3.743	-4.070	-3.842	-3.849	-3.846	-3.861
AR2 test	-0.0401	-0.587	-0.908	-0.953	-0.934	-0.938

t-statistics in parenthesis
*S*ignificance level: * - 10%, ** - 5%, *** - 1%

Table 13: Robustness check with 2-step GMM with Windmeijer correction estimation

	Dependent variable - log(*Internet users*)					
L.log_Internet	0.565***	0.500***	0.524***	0.532***	0.528***	0.539***
	(6.40)	(6.27)	(5.39)	(5.41)	(5.41)	(5.72)
log_GDP_pc	-1.887**	-1.418**	-1.474**	-1.518**	-1.453*	-1.490**
	(-2.36)	(-2.14)	(-2.04)	(-2.00)	(-1.98)	(-2.08)
literacy_rate	0.017	0.014	0.018	0.022	0.024	0.025
	(0.95)	(0.97)	(1.13)	(1.22)	(1.36)	(1.41)
FDI_gross_hat	0.061*		0.076	0.063	0.062	0.068
	(1.68)		(1.62)	(1.41)	(1.38)	(1.38)
log_PC	0.120	0.090				
	(0.53)	(0.42)				
trade_open						0.003
						(0.72)
urban_population					0.015	
					(0.32)	
log_phonelines				0.134		
				(0.36)		
log_phones			0.346*			
			(1.67)			
FDI_gross		0.007				
		(1.26)				
Observations	411	392	450	452	453	449
Number of id	63	61	66	66	66	66
F-statistic	395.4	392.3	306.0	327.9	362.5	340.5
hansen	10.46	8.973	15.11	14.83	14.24	13.64
hansen prob.	0.315	0.440	0.0880	0.0956	0.114	0.136
instrum. number	24	24	25	25	25	25
AR1 test	-3.723	-3.913	-3.383	-3.408	-3.407	-3.527
AR2 test	-0.0922	-0.555	-0.855	-0.912	-0.917	-0.926

t-statistics in parenthesis
*S*ignificance level: * - 10%, ** - 5%, *** - 1%

F Estimation for the developed countries

Table 14: 2-step GMM with Windmeijer correction

	collapsed instruments			non-collapsed instruments		
	$p = 2$	$p = 3$	$p = \max$	$p = 2$	$p = 3$	$p = \max$
	Dependent variable - log(*Internet users*)					
L.log_Internet	0.867***	0.815***	0.722*	0.747***	0.721***	0.600*
	(4.434)	(7.746)	(1.886)	(3.187)	(3.222)	(1.910)
log_GDP_pc	2.907**	2.829**	1.271	3.720***	-4.983	-6.887
	(2.322)	(2.357)	(0.112)	(3.174)	(-0.468)	(-0.562)
log_PC	-0.182	-0.172	-0.270	0.562	-0.112	-0.247
	(-0.643)	(-0.625)	(-0.247)	(0.762)	(-0.270)	(-0.174)
FDI_gross	0.002	0.002	0.001	0.002	0.004	0.070
	(1.083)	(1.097)	(0.148)	(0.931)	(1.423)	(0.351)
Year dummies	YES	YES	YES	YES	YES	YES
Observations	239	239	239	239	239	239
Number of id	22	22	22	22	22	22
F-statistic	935.490	981.386	244.044	1438.967	312.966	10.356
sigma2	0.053	0.050	0.046	0.048	0.050	44.161
instrum. number	16	17	27	27	37	82
AR1 test	-3.037	-3.611	-1.406	-2.100	-2.288	-0.407
AR2 test	-0.125	-0.103	-0.066	0.005	-0.383	-0.370

t-statistics in parenthesis

*Significance level: * - 10%, ** - 5%, *** - 1%*