**Professor Stelian STANCU, PhD (corresponding author)**
**E-mail: stelian.stancu@csie.ase.ro**
**Department of Economic Informatics and Cybernetics**
**Bucharest University of Economic Studies, Romania**

**Andreea PERNICI, PhD Candidate**
**E-mail: andreea.pernici@csie.ase.ro**
**Bucharest University of Economic Studies, Romania**

**Denisa Elena BĂLĂ, PhD Candidate**
**E-mail: baladenisa16@csie.ase.ro**
**Bucharest University of Economic Studies, Romania**

**Associate Professor Daniela Elena MARINESCU**
**E-mail: daniela.marinescu@csie.ase.ro**
**Department of Economic Informatics and Cybernetics**
**Bucharest University of Economic Studies, Romania**

# LINKING E-GOVERNMENT TO SOCIO-DEMOGRAPHICS A MULTI-MODEL APPROACH USING PANEL DATA, REGRESSION, AND NEURAL NETWORKS

***Abstract.*** *The emergence of the Internet as a central phenomenon of the post-industrial era has generated structural changes in all socio-economic processes, including public administration. On this note, e-government has become the new practice, aimed to reshape the relationship between state and citizen, but showing considerable differences in the way it is adopted worldwide. Therefore, this paper proposes a methodology for assessing whether socio-demographic characteristics such as the urban population, the average years spent in school, or the overall median age of the population can successfully explain the level of e-government adaptation (measured through the EGDI). To observe it, we will use data from six years, from 2010 to 2020, and 130 countries of the UN. As tools, we have employed a multi-model approach that consists of regression and neural network models, with two main directions of analysis: individual yearly methods and longitudinal panel data.*

***Keywords****: E-government, EGDI, Socio-Demographics, Panel Data, Regression, Neural Networks*

**JEL Classification: C31, C33, C45, J1, O57, D73**



DOI: 10.24818/18423264/57.2.23.04

## 1. Introduction

The social context has always been closely linked to any major technological development, so the new information era has profoundly altered all the productive and administrative systems of a state, from the micro level to the macro one. Without making a distinctive mark, governments will make considerable efforts to synchronise themselves with the accelerated pace of innovation, by constantly implementing information and communication technologies through the capitalisation of the Internet. Consequently, public administrations have now switched to a new way of providing services, namely e-government. Since this transition generates bidirectional relations, the aim of this paper would be to study whether the social repartition of the population results in differences in the overall e-government performance. This topic has been constantly present in the related literature, as well as in the international agenda of the last two decades, still maintaining still a high level of relevancy. Therefore, before describing the methodologies employed, we will attempt to define e-government, connect it to socio-demographics through concepts such as the digital divide or digital user, and illustrate the contributions brought to the prior research efforts.

## 2. Conceptual model

### *2.1. E-government definition frameworks*

The concept of e-government started to gain momentum in the early 1990s, along with the development of several information and communication technologies (ICTs). Although it was first designed as a complementary way to address and provide public solutions, over the past few years it managed to replace a lot of physical, inefficient bureaucratic interactions, gaining new functionalities as we speak. Therefore, the concept has been dynamic, having multiple interpretations and definitions. We will briefly present some of them.

First, (Carter and Belanger, 2005) will interpret e-government as *the use of information technologies to enable and improve the efficiency with which government services are provided to citizens, employees, businesses, and agencies*. In this sense, the role assumed by e-solutions will be to enhance and facilitate interactions between the government and the public, by offering several advantages: the integration of new workflows and processes, efficient management of data and information, an augmented service delivery, or the expansion of public engagement through different channels (United Nations, 2014). In a similar, but whatsoever more simplistic overview, we can define e-government as the adaptation of public administrations to the new information society, by integrating digital services to achieve increased public accountability, improved efficiency and cost-effectiveness, and greater overall participation.

More recent efforts have switched the perspective, focusing on aspects such as public value. An example would be (Twizeyimana and Andersson, 2019) that have identified six dimensions that e-government is supposed to improve: *public services, administrative efficiency, open government capabilities, ethical behavior and professionalism, trust and confidence, and social value and well-being.* Another perspective has been to focus on the future, and in this case (Malodia et al., 2021) have stated that there are three underlying dimensions: *empowered citizenship, hyper-integrated network, and evolutionary system architecture.* Moreover, modern concepts such as blockchain (Kassen, 2022), sustainable growth (Osman and Zablith, 2021), or even COVID-19 (Dammak et al., 2023) have begun to emerge in congruence with e-government. However, although most perspectives focus on efficiency, there is also much concern regarding the negative social impact of this new way of providing public services. In that matter, one of the most common would be the disproportionate use and access of electronic means, known as the digital divide.

### *2.2. The digital divide and the e-government user profile*

The digital divide has been a subject of interest in the last two decades, with the related literature presenting several theorisation approaches, from determinants to policies, practices, impact, or even privileges. When limiting the field of research from the digital to the e-government divide, we can observe that although these kinds of services improve public administration quality and accountability, in developing countries they might fail to reach their desired goals due to challenges such as budgeting, human resources, managerial issues, or the digital culture (Meiyanti et al., 2018). This very last digital culture could be the most complex concept, since it is difficult to ensure a successful national strategy, without understanding the social factors behind it. What helps in this context is that the classical profile of a digital user can be susceptible to generalisations.

Therefore, there have been many attempts to find, with a degree of probability, certain characteristic demographics that best describe the information society, and implicitly, the e-government user. The enabler of this is Rogers (1962) and its framework illustrated in *The Diffusion of Innovation Theory*. He will describe how a new technology gains momentum and is ultimately adopted by the population, through five stages: *innovators, early adopters, early majority, late majority, and laggards.* Another way to classify them would be by using the terms: *pioneers, transitioners, and performers* (Pernici and Stancu, 2023). Both frameworks can be easily replicated in the e-government concept, since the diffusing of electronic public services has been irregular over time and with heterogenous results across the world. On this note, socio-demographics could be part of the solution since they can explain the divergence and provide insights on how to reduce it.

Consequently, several attempts have become representative of the socio-demographic profile of the e-government user. (Thomas and Streib, 2003) describe them as white, of higher income, more educated, and relatively younger. (Hart and Teeter, 2003) will present a similar description: white, college graduate, and professional. Taipale (2013) will also prove that education, income, and even the size of the residence impact the profile. Urban distribution has also been studied numerous times (Goldfinch et al., 2009), Reddick (2011), adding a new dimension. More recent approaches tend to be complementary, with education, age, gender, or occupation being studied in relation to e-participation (Zheng and Shachter, 2017) or citizen satisfaction (Ma and Zheng, 2019). However, these efforts are subject to certain general trends and limitations that we will expose moving forward.

### *2.3. Contributions to the related literature*

One of the most noticeable directions found in the related literature is the qualitative one, since many studies will be based on empirical data collected from the population through field studies, surveys, and questionnaires, being later analysed through a regression technique. A useful summary of these endeavors has been shaped by (Pérez-Morote et al., 2020), with a preference for cross-sectional data being observable. Therefore, most of the studies refer to a specific country, such as Indonesia (Sabani, et al., 2019), the USA (Nam and Sayogo, 2011), or Finland (Taipale, 2013). One exception will be represented by (Park et al., 2013) who will use longitudinal data from Korea from 2003 to 2005 to explore the connection between e-readiness and e-government use.

Regarding our model, of all the elements illustrated, we have chosen three, namely age, urban distribution, and education, that could explain the overall performance of e-government, both in historical and national contexts. There will be two main points of differentiation versus the previously illustrated approaches: the employment of panel data and their respective regression models and the introduction of neural network methods. Since we briefly mentioned the field of machine learning, we have identified two papers that have computed related methods. First is the study written by Sharma et al (2015), in which both a regression and a neural network algorithm will be employed to investigate the determinants of e-government adoption in Oman. Similarly, Mostafa and El-Massry (2013) have built a complex model related to Egypt’s user profile, starting from social-demographic, cognitive, and attitudinal variables and using data-mining techniques.

## 3. E-Government Development Index (EGDI)

Lastly, before describing the methodology behind our study, we need to illustrate the framework used to define e-government. Therefore, the EGDI is a model composed by the UN Organisation, through the Social and Economic Affairs Department, since 2001, being the only indicator that evaluates the level of e-government worldwide, with a frequency of two years. It is calculated based on a complex survey of the online resources used to deliver public services. Using the data from that survey, a composite indicator is created that will equally weigh three dimensions: Telecommunication Infrastructure (TII), Human Capital (HCI), and Online Services Availability (OSI). Each of them can be interpreted as an individual measure with explanatory potential.

$$EGDI = \frac{1}{3}\,(TII_{normalised} + HCI_{normalised} + OSI_{normalised}) \tag{1}$$

When prospecting the EGDI related literature, we have not found a model that links this index to socio-demographics in a panel approach on an international scale, so we can consider it as an additional contribution to the domain.

## 4. Dataset and methodology

As mentioned previously, the current model aims to describe the relationship between the E-Government Development Index (EGDI) and the socio-demographic distribution of the population. The methodology will consist of two stages, based on the type of data we have: yearly and panel. In the second case, as a time of reference, we have considered the last decade, with the six distinct moments for which the EGDI has been calculated: 2010, 2012, 2014, 2016, 2018, and 2020. For each of the stages, we included 130 UN countries in the analysis, removing those that had a null EGDI score at any given moment. Lastly, regarding the socio-demographic variables, we have included three indicators, extracted from the World Bank Data portal and the Human Capital Index Data, with their respective codification visible in Table 1.

**Table 1. Variables' description**

| Variables | Description | Source |
|---|---|---|
| URBAN | Percentage of population that lives in the urban area | World Bank Data |
| MEDAGE | Median age of the population | World Bank Data |
| MEANYEARS | Average number of years spent in school | HCI Index Data |
| EGDI | E-Government Development Index | United Nations |

No further initial processing has been applied to the data, as each of the methods will necessitate a different preliminary data setting. In the next segment, each stage will be split into various steps through which we will present the regression and neural network models applied.

### I. YEARLY DATA
### Step 1. Multiple linear regression models

The first step in the yearly analysis is to split the data into six distinct subsets, one for each year, with the 130 observations remaining unchanged. After that, we applied the normalisation procedure by using the *max-min* criteria and we defined our train and test set, with a ratio of 90% to 10%.

After the data processing step is completed, we will compute the multiple linear regression model. In formula (2) we can observe the regression equation, with $\beta_0, \beta_1, \beta_2,$ and $\beta_3$ being the coefficients that will be estimated and the $u_i$ the independent error terms.

$$EGDI_i = \beta_0 + \beta_1\, URBAN_i + \beta_2\, MEDAGE_i + \beta_3\, MEANYEARS_i + u_i \tag{2}$$

The regression results for the six years can be seen in Table 2. All the models and most of the coefficients will be highly significant, except for 2010, 2012, 2014, and 2016 intercepts, highlighted by the (!) symbol in the below table. When this happens, we can conclude that we do not have enough statistical evidence that the intercept is different from zero, which could indicate that in the absence of the socio-demographic variables, the EGDI will not vary.

**Table 2. Linear Regression Models Results, 2010-2020**

| Model | R2 | Equation | MSE |
|---|---|---|---|
| 2010 | 84.1% | EGDI = 0.0021 (!)+ 0.2485*URBAN + 0.352*MEDAGE + 0.286*MEANYEARS | 0.0088 |
| 2012 | 87% | EGDI = 0.0116 (!)+ 0.273*URBAN + 0.35*MEDAGE + 0.342*MEANYEARS | 0.0057 |
| 2014 | 85.2% | EGDI = -0.0405 (!)+ 0.2884*URBAN + 0.35*MEDAGE + 0.358*MEANYEARS | 0.0107 |
| 2016 | 85.2% | EGDI = -0.0081 (!)+ 0.252*URBAN + 0.385*MEDAGE + 0.367*MEANYEARS | 0.0051 |
| 2018 | 85.6% | EGDI = 0.095+ 0.218*URBAN + 0.382*MEDAGE + 0.349*MEANYEARS | 0.0058 |
| 2020 | 84.8% | EGDI = 0.114 + 0.171*URBAN + 0.362*MEDAGE + 0.381*MEANYEARS | 0.0034 |

*Source: Authors' own processing*

Regarding the goodness of fit, the results show an overall good performance, evaluated through the *R-Squared* values. 85%, the average value obtained in our case indicates that we are accounting for a large portion of the variance, the model having a strong explanatory capacity. More than that, we can observe a generally small degree of deviation between years, so the relationship remains robust throughout the decade. To further explore the accuracy of prediction, we have also computed the *Mean Squared Error (MSE)*, measuring how close the regression line is to the data points of the test set. In our case, for all six years, the MSE will take very small values, again showing a very good fit.

$$MSE = \frac{1}{13} * \Sigma(EGDI - EGDI_p)^2 \quad (3)$$

As an economic interpretation, we can see that all the coefficients are positive and significant, showing a direct correlation with the EGDI and proving that the social profile of a nation will translate into the e-government's overall performance. Further observations worth mentioning:

- The highest coefficients will be registered either for the MEDAGE or the MEANYEARS variables, showing a stronger contribution.
- The URBAN coefficients will decrease over the years, showing that in the future this characteristic might not contribute that much to the user profile.
- There is a positive relation between the average age and e-government since as the value of MEDAGE increases, the EGDI will also increase.

Therefore, if for the urban distribution and the average years spent in school, the results are similar to the considerations exposed in the theoretical part, we could have expected that for the median age of the population, the relationship would be reversed, as stated by other authors: young population tends to adopt new technologies faster. However, from our results, this hypothesis has been rejected.

**Step 2. Neural Network Model**

To study if we can better explain the socio-demographics and e-government relation, we have computed a simple neural network (NN) algorithm. This method has become one of the most popular computational systems nowadays, inspired by the brain structure and its biological characteristics. Neural network methods interconnect layers of small units called nodes that will ultimately detect patterns in data, encompassing elements such as inputs, weights, bias, or activation functions. However, the most important element will be the *neuron*, characterised in machine learning as the processing unit. The field has grown exponentially, and the advances have led to the study of interconnection between neurons, which ultimately generates one of the most complex architectures capable of self-learning, adaptivity, and non-linearity problems (Abiodun, et al., 2018).

For this case, we will compute a NN method to see if we can increase the accuracy of prediction. We will use the *neuralnet* function in R and the same train and test set from the linear regression above. We will set up the hidden layer to 1, the hidden neurons to 2 (Figure 1), and the number of repetitions to 10. The structure of our neuron is represented in Figure 1.

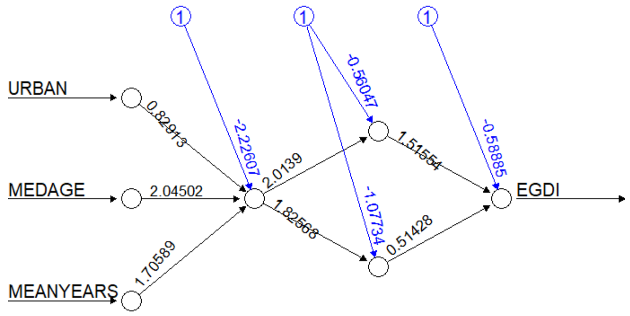


**Figure 1. Neural Network graphical representation, 2020**

*Source: Authors' own processing*

**Table 3. Neural Network Models Results compared to LM, 2010-2020**

| Year | MSE NN | MSE LM | Best Model |
|---|---|---|---|
| 2010 | 0.0091 | 0.0088 | LM |
| 2012 | 0.0064 | 0.0057 | LM |
| 2014 | 0.0104 | 0.0107 | NN |
| 2016 | 0.005 | 0.0051 | NN |
| 2018 | 0.0056 | 0.0058 | NN |
| 2020 | 0.0037 | 0.0034 | LM |

*Source: Authors' own processing*

From Table 3, we can observe that the performance of the prediction has the same evolution, with similar MSE values as in the linear model approach. However, another test that we can perform is to compare the predicted values obtained by both models, for the 13 countries in the test set. In Table 4 and Figure 2, we can see that the results are extremely close, with the linear model and neural network sharing the spot as the best predictor. Therefore, we can conclude that the relationship between e-government and socio-demographics can be described by a linear function, with the three social-demographic variables being capable of estimating EGDI results.

**Table 4. Example of predicted values, LM and NN methods, 2020**

| Country | EGDI | EGDI-LM | EGDI-NN | Best Model |
|---|---|---|---|---|
| BAHRAIN | 0.82 | 0.8 | 0.82 | NN |
| BELARUS | 0.81 | 0.87 | 0.87 | - |
| BARBADOS | 0.73 | 0.75 | 0.77 | LM |
| CZECH REPUBLIC | 0.81 | 0.9 | 0.89 | NN |
| ESTONIA | 0.95 | 0.9 | 0.89 | LM |
| FIJI | 0.66 | 0.72 | 0.73 | LM |
| FRANCE | 0.87 | 0.87 | 0.87 | - |
| ITALY | 0.82 | 0.88 | 0.88 | - |
| SERBIA | 0.75 | 0.85 | 0.86 | LM |
| SLOVENIA | 0.85 | 0.87 | 0.88 | LM |
| TOGO | 0.43 | 0.53 | 0.52 | NN |
| THAILANDA | 0.76 | 0.76 | 0.78 | LM |
| SOUTH AFRICA | 0.69 | 0.75 | 0.75 | - |

*Source: Authors' own processing*

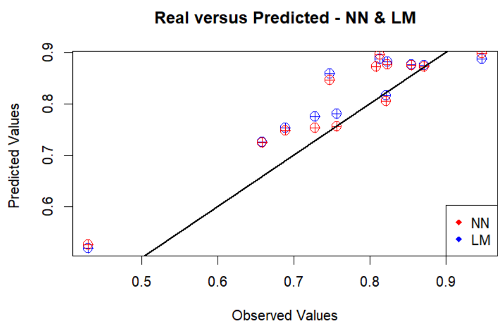


**Figure 2. Real versus Predicted Values, LM and NN methods, 2020**
*Source: Authors' own processing*

## II. PANEL DATA

Panel Data has been increasingly used in econometrics in recent years, due to its three-dimensional potential to gain insights. This type of data will be structured in a set that is collected over a period, for multiple entities, which is a combination of cross-sectional and longitudinal data. In other words, it will *provide information on both the intertemporal dynamics and the individuality of the entities, being better* *at controlling the effects of missing or unobserved variables* (Hsiao, 2022). For the current model, the dataset refers to the same 130 countries and 6 years described before. In a panel dataset, each observation (or row) is called a unit so, in our case, the countries will be the units, with all the respective data for the six moments. The panel is balanced, so no data is missing from any of the countries or years collected. The variables included are the *E-government Development Index (EGDI)* and the same three socio-demographic variables: URBAN, MEDAGE, and MEANYEARS. Before proceeding with the regression model description, we can evaluate the datasets from two points of view.

*Heterogeneity* will study the differences in parameters or the variance over individuals and time. This concept will be crucial since the main opportunity provided by panel data is that it will allow heterogeneity to exist, and more than that, it will capture any individual or time-specific effects**.** Before going into details about the heterogeneous models, we can get a glimpse of this characteristic from Figure 3.

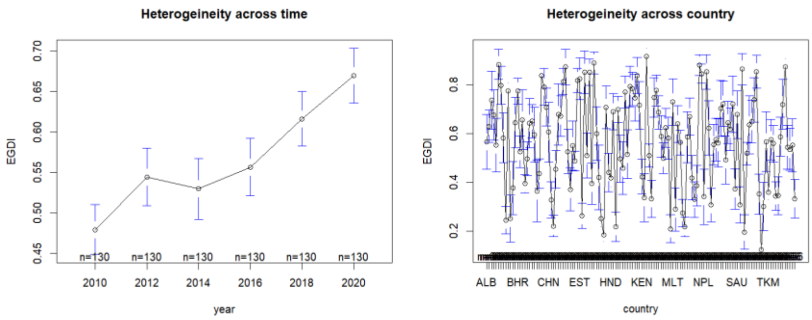


**Figure 3. Heterogeneity across time and individuals**
*Source: Authors' own processing*

*Stationarity* will show whether the dataset properties depend on time or are affected by trends and seasonality. To study this phenomenon, we will use the concept of a unit root, which will be a stochastic trend in a series that generates an unpredictable systematic pattern. There is a multitude of methods that can be used to implement the unit root test for panel data in R, but one of the most recurrent will be by computing the *purtest* function with the test parameter set to *levinlin*.

**Table 5. Levin-Lin-Chu Stationarity Test**

| z | -15.704 | p-value | < 2.2e-16 * |
|---|---|---|---|
| Alternative hypothesis: stationarity | | | |

*Source: Authors' own processing*

As we can see from R output, our z statistic equals -15.704, with a p-value smaller than 2.2e-16, which means that we can reject the null hypothesis and accept the alternative one, that of a stationary dataset. As a general interpretation, that will mean that a shift in time will not cause a change in the distribution and the following models constructed can be significant.

After getting a glimpse of the data, we will employ three different regression approaches: Pooled OLS, Fixed Effects, and Random Effects.

**Step 1. Pooled OLS Regression Model**

The first model computed will be the Pooled OLS (Ordinary Least Squares), which can be used as a point of reference, since it will construct a regression without any cross-sectional or time effects. In other words, the method will ignore time and individual characteristics, focusing only on the dependencies between entities. In that way, the model is similar to linear regression, assuming that the intercept and slopes are constant regardless of the group or time period. The general panel regression model can be described mathematically as below.

$$EGDI_{it} = \beta_0 + \beta_1 URBAN_{it} + \beta_2 MEDAGE_{it} + \beta_3 MEANYEARS_{it} + u_{it} \quad (4)$$

Before applying the method in R, we have once again split the data into a train and a test set, with a 90%-10% distribution. Therefore, the *lm* function has been applied to 702 observations (from all 6 years). The dependent and explanatory variables will remain unchanged. The results can be seen in Table 6.

**Table 6. Pooled OLS Regression**

| Coefficient | Estimate | Std. Error | Pr (>\|t\|) |
|---|---|---|---|
| Intercept | -0.0685 | 0.0123 | 4.1e-08 *** |
| URBAN | 0.0022 | 0.0002 | <2e-16 *** |
| MEDAGE | 0.0094 | 0.0006 | <2e-16 *** |
| MEANYEARS | 0.0235 | 0.00221 | <2e-16 *** |
| **R-Squared** | **81.9%** | **Adj. R-Squared** | **81.8%** |
| **F-Statistic** | **1.049** | **p-value:** | **<2.2e-16** |

*Source: Authors' own processing*

The MSE (Mean Squared Error) for this method will be 0.009, showing a very good fit, while the regression line, as well as the observations can be seen in Figure 4. Regarding the economic interpretation, we can once again see that all the coefficients will be highly significant, showing a direct relation to the EGDI. The MEANYEARS coefficient will take the highest value, emphasising that regardless of time, the average years spent in school are weighing the most in the international e-government adoption equation.

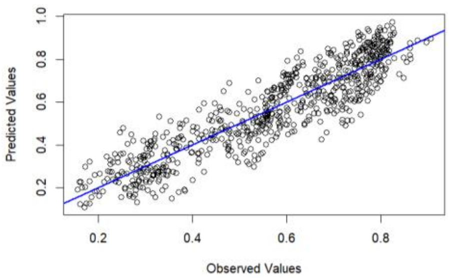


**Figure 4. Observed versus Predicted Values, Pooled OLS Method**
*Source: Authors' own processing*

Further on, circling back to the panel data definition and the *heterogeneity* concept, the next methods will showcase how to deal with the unobserved dependency of other independent variables not captured in the model. We can control this phenomenon, by acknowledging it as fixed or random, each with the respective family of models.

**Step 2. Fixed Effects Regression Model**

A fixed effects model will focus on the differences between entities, determining that the individual effects of unobserved variables are constant over time, allowing heterogeneity to be existent. In other words, when using the FE model, we assume that something within the individual might affect the predictor variables, so the entity's error term and outcome variables are correlated. As (Allison, 2009) further explains, in a FE *the model treats unobserved differences between variables as a set of fixed parameters that can be either estimated or partialed out of the estimating equations.*

There will be multiple ways of implementing the fixed-effect algorithm, but one of the most used is the *within* estimator. This method will look at the deviations from the group means, using the variation within each entity. The regression equation will be presented in (5), where $\beta_0$ is the intercept and the $Z_i$ will be the unobserved time-invariant heterogeneities between the individuals.

$$EGDI_{it} = \beta_0 + \beta_1\, URBAN_{it} + \beta_2\, MEDAGE_{it} + \beta_3\, MEANYEARS_{it} + \beta_4\, Z_i + u_{it}, \quad (5)$$

However, besides individual effects, this method can capture time-effects as well. For this case, the equation will look as below, where $\gamma_t$ is the time-fixed effect that will work as a dummy variable for each captured period. This type of effect can be explained by all the variables that are present for all individuals at a given time that might influence the outcome. In our case, for example, it might show some international trends that affect the EGDI variable globally.

$$EGDI_{it} = \alpha_i + \gamma t + \beta_1\, URBAN_{it} + \beta_2\, MEDAGE_{it} + \beta_3\, MEANYEARS_{it} + u_{it},$$
$$\text{where } \alpha_i = \beta_0 + \beta_4 Z_i \quad (5')$$

Further on, will compute the fixed-effects method on the same dataset as for the Pooled OLS, without splitting into train and test, using the *plm* function and the model parameter *within.* The results for the individual and time-effects methods are available in Tables 7 and 8. Almost all the coefficients will be highly significant, with a small difference in the case of MEANYEARS for the first model. Both will validate the F-statistic test.

**Table 7. Fixed-Effects Regression – individual effects**

| Coefficient | Estimate | Std. Error | Pr (>\|t\|) |
|---|---|---|---|
| URBAN | 0.0105 | 0.00169 | 1.089e-09 *** |
| MEDAGE | 0.038 | 0.00262 | <2.2e-16 *** |
| MEANYEARS | 0.0166 | 0.00623 | 0.007821 ** |
| **R-Squared** | **60.7%** | **Adj. R-Squared** | **52.7%** |
| **F-Statistic** | **295.3** | **p-value:** | **<2.22e-16** |

*Source: Authors' own processing*

As an interpretation direction, we can say that the coefficients will indicate how much EGDI changes over time, on average per nation, when the respective socio-demographic increases by one unit.

**Table 8. Fixed-Effects Regression – time effects**

| Coefficient | Estimate | Std. Error | Pr (>\|t\|) |
|---|---|---|---|
| URBAN | 0.0023 | 0.00018 | <2.2e-16 *** |
| MEDAGE | 0.0094 | 0.00057 | <2.2e-16 *** |
| MEANYEARS | 0.0220 | 0.00177 | <2.2e-16 *** |
| **R-Squared** | **84.5%** | **Adj. R-Squared** | **84.3%** |
| **F-Statistic** | **1.402.6** | **p-value:** | **<2.22e-16** |

*Source: Authors' own processing*

A relevant observation is that in the time approach, the R-Squared value is considerably higher, meaning that there might be some momentum influences that affect all the countries. For example, the focus put during the pandemic period on digitalisation and e-government from all international institutions might have had an impact on e-government performance.

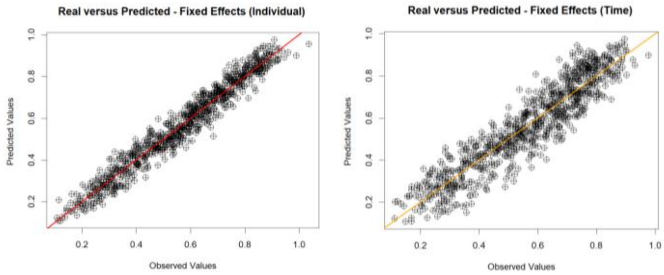


**Figure 5. Observed versus Predicted Values, Pooled OLS Method**
*Source: Authors' own processing*

Regarding the accuracy of prediction, we can see that for the individual effects, there is a more concentrated conglomerate of observations around the regression line, which will be a first indicator that this model manages to explain the relationship better.

**Step 3. Random Effects Regression Model**

To build on the first two models, in the random effects one, the focus will be on the variables that are constant across individuals, but that can be considered as random variables from an underlying process. Therefore, it assumes that the explanatory variables do have fixed effects on the outcome, but these effects may vary from observation to observation. For example, the international agenda of digitalisation might have affected all the countries involved, but the way it generated effects might have been different due to the lack of resources or infrastructure.

The equation will be similar to the fixed-effects one, but it will add a new term $\mu_{it,}$ that will show the variance introduced by the unit-specific effects.

$$EGDI_{it} = \beta_0 + \beta_1\, URBAN_{it} + \beta_2\, MEDAGE_{it} + \beta_3\, MEANYEARS_{it} + \beta_4\, Z_i + \mu_{it} + u_{it} \quad (6)$$

For the time-effects, once again the equation will be similar to the fixed-effects one, with the addition of $\mu_{it.}$

$$EGDI_{it} = \alpha_i + \gamma t + \beta_1\, URBAN_{it} + \beta_2\, MEDAGE_{it} + \beta_3\, MEANYEARS_{it} + \mu_{it} + u_{it}, \quad \text{where } \alpha_i = \beta_0 + \beta_4 Z_i \qquad (6')$$

**Table 9. Random-Effects Regression – individual effects**

| Coefficient | Estimate | Std. Error | Pr (>\|t\|) |
|---|---|---|---|
| INTERCEPT | -0.1867 | 0.02626 | 1.161e-12 *** |
| URBAN | 0.00180 | 0.00046 | 0.0001*** |
| MEDAGE | 0.01256 | 0.00133 | <2.2e-16 *** |
| MEANYEARS | 0.0292 | 0.00395 | 1.453e-13*** |
| **R-Squared** | **57.4%** | **Adj. R-Squared** | **57.3%** |
| **F-Statistic** | **1.047.3** | **p-value:** | **<2.22e-16** |

*Source: Authors' own processing*

**Table 10. Random-Effects Regression – time effects**

| Coefficient | Estimate | Std. Error | Pr (>\|t\|) |
|---|---|---|---|
| INTERCEPT | -0.0587 | 0.01344 | 1.238e-05 *** |
| URBAN | 0.0023 | 0.00018 | <2.2e-16 *** |
| MEDAGE | 0.0094 | 0.00057 | <2.2e-16 *** |
| MEANYEARS | 0.0222 | 0.0018 | <2.2e-16 *** |
| **R-Squared** | **84.1%** | **Adj. R-Squared** | **84%** |
| **Chisq** | **4.118.5** | **p-value:** | **<2.22e-16** |

*Source: Authors' own processing*

We can see that once again the time-effects model will have a higher $R^2$, but with overall weaker results than the ones registered for the fixed effects.

The last step of the analysis would be to observe the prediction accuracy, by randomly extracting 4 countries from the dataset and fitting the values for the years 2018 and 2020 (once again chosen randomly). In 5 out of 8 observations, the most accurate prediction has been computed through the fixed, individual effects model, so this will be considered the most efficient.

**Table 11. Example of predicted values, LM and NN methods, 2020**

| Country | Year | EGDI | EGDI Pooled OLS | EGDI FE Ind | EGDI FE Time | EGDI RE Ind | EGDI RE Time |
|---|---|---|---|---|---|---|---|
| BAHRAIN | 2018 | **0.81** | 0.67 | **0.82** | 0.7 | 0.77 | 0.69 |
| BAHRAIN | 2020 | **0.82** | 0.69 | 0.87 | 0.77 | **0.8** | 0.74 |
| ITALY | 2018 | **0.82** | 0.77 | **0.8** | **0.80** | 0.79 | 0.79 |
| ITALY | 2020 | **0.82** | 0.78 | 0.84 | 0.86 | 0.80 | **0.83** |
| THAILANDA | 2018 | **0.65** | 0.60 | **0.64** | 0.63 | 0.60 | 0.62 |
| THAILANDA | 2020 | **0.76** | 0.62 | **0.7** | 0.69 | 0.63 | 0.67 |
| SOUTH AFRICA | 2018 | **0.66** | 0.58 | **0.61** | **0.61** | 0.58 | 0.6 |
| SOUTH AFRICA | 2020 | **0.69** | 0.60 | 0.66 | **0.68** | 0.61 | 0.65 |

*Source: Authors' own processing*

However, since this comparative method could seem empirical, multiple statistical tests can support the decision to choose the optimal model.

*Poolability Test – Chow.* This method will calculate an *F* test of stability for the coefficients of a panel model. Using it, we can select between the Pooled OLS and FE models. The results can be seen in Table 12, with the p-value being placed below the 0.05 limit, so the FE model can be considered better for the current dataset.

**Table 12. Poolability Test**

| | | | |
|---|---|---|---|
| F-statistic | 15.011 | p-value | < 2.2e-16 |
| Alternative hypothesis: significant effects | | | |

*Source: Authors' own processing*

*Hausman Test.* This test will be based on the differences between the vectors of coefficients of two models, calculating the quadratic formula developed by (Hausman, 1978)[1]. It can be used to select between a Random-Effects and a Fixed-Effects model. Since once again the p-value is very small, we can reject the null hypothesis, so the FE model is more suitable.

**Table 13. Hausman Test**

| | | | |
|---|---|---|---|
| Chisq | 383.85 | p-value | < 2.2e-16 |
| Alternative hypothesis: one model is incosistent | | | |

*Source: Authors' own processing*

*The Breutsch-Pagan Lagrange Multiplier Test.* This method will determine if there are significant random effects in the panel datasets, helping to decide between OLS and Random-Effects[2]. As before, the p-value will be below 0.05, so we will reject the null hypothesis, with the RE model being the most appropriate one, with evidence of significant differences across countries.

**Table 14. Lagrange Multiplier Test**

| | | | |
|---|---|---|---|
| Chisq | 513.63 | p-value | < 2.2e-16 |
| Alternative hypothesis: significant effects | | | |

*Source: Authors' own processing*

Therefore, we can conclude that the best model for panel data is the fixed effects with the individual method applied. Similar to the yearly analysis stage, we can see that the coefficients will be positive and highly significant, ultimately showing that besides the socio-demographics, there will be some country characteristics that will not be captured in the model, but that will be relevant to a nation's profile.

[1] https://search.r-project.org/CRAN/refmans/plm/html/phtest.html. Last accessed January 2023.
[2] https://www.princeton.edu/~otorres/Panel101R.pdf. Last accessed January 2023.

## 5. Conclusions

Overall, all the models have been proven suitable to predict with good accuracy the level of e-government performance. The three indicators, age, urban distribution, and education, are significant and have a positive relationship with the EGDI values. For the yearly data, the 85% explained variance is the first proof that socio-demographics do influence the way e-government is adopted regardless of regional specificity.

Going into detail, when studying the individual effects, we have seen that the variance is dropping, to an average of 60%, which is still a good level, proving that the three social indexes manage to explain most of the dynamics in a panel data, but there are some uncaptured variables that will remain particular to the countries. The practical implication of this will be the potential to personalise international strategies to the local action plans, in order to enhance them, construct a complete user profile, and ultimately reduce the digital divide. Knowing that governments can address at-risk groups, for example, rural or primary and secondary educated populations, by allocating resources for the development and propagation of personalised solutions.

Lastly, regarding the time effects, we can see a higher R-Square, so the explanatory power is stronger since we refer to movements that are generated on an international scale in certain moments. This proves that e-government is indeed a global subject, a reality for most countries, with major events affecting all of them.

## REFERENCES

[1] **Abiodun, O.I., Jantan, A., Omolara, A.E., Dada, K.V., Mohamed, N.A., Arshad, H. (2018),** ***State-of-the-art in artificial neural network applications: A survey.*** *Heliyon,* 4(11), e00938;

[2] **Carter, L., Belanger F. (2005),** ***The utilization of e‐ government services: citizen trust, innovation and acceptance factors.*** *Information Systems Journal*, 15(1), 5-25;

[3] **Dammak, S., Mbarek, S., Moalla, M. (2023),** ***E-government, political system and COVID-19 in Africa: lessons for future disruptive shocks.*** *Transforming Government: People, Process and Policy.* Ahead-of-print;

[4] **Goldfinch, S., Gauld, R., Herbison, P. (2009),** ***The participation divide? political participation, trust in government, and e‐ government in Australia and New Zealand.*** *Australian Journal of Public Administration*, 68(3), 333-350;

[5] **Hart, D.P., Teeter, M. (2003),** ***The new e-government equation: Ease, engagement, privacy and protection.*** *Hart-Teeter Research;*

[6] **Hsiao, C. (2022),** ***Analysis of panel data.*** No. 64. *Cambridge University Press;*

[7] **Kassen, M. (2022),** ***Blockchain and e-government innovation: Automation of public information processes.*** *Information Systems*, 103, 101862;

[8] **Ma, L., Zheng, Y. (2019),** ***National e-government performance and citizen satisfaction: a multilevel analysis across European countries***. *International Review of Administrative Sciences*, 85(3), 506-526;

[9] **Malodia, S., Dhir, A., Mishra, M., Bhatti, Z.A. (2021),** ***Future of e-Government: An integrated conceptual framework.*** *Technological Forecasting and Social Change*, 173;

[10] **Meiyanti, R., Utomo, B., Sensuse, D.I., Wahyuni, R. (2018),** ***e-Government challenges in developing Countries: A literature review.*** In *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, 1-6, IEEE;
[11] **Mostafa, M.M., El-Masry, A.A. (2013),** ***Citizens as consumers: Profiling e-government services' users in Egypt via data mining techniques***. *International Journal of Information Management*, 33(4), 627-641;
[12] **Nam, T., Sayogo, D.S. (2011),** ***Who uses e-government? Examining the digital divide in e-government use.*** In *Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance*, 27-36;
[13] **Osman, I.H., Zablith, F. (2021),** ***Re-evaluating electronic government development index to monitor the transformation toward achieving sustainable development goals.*** *Journal of Business Research*, 131, 426-440;
[14] **Park, S., Choi, Y-T., Bok H-S. (2013),** ***Does better e-readiness induce more use of e-government? Evidence from the Korean central e-government.*** *International Review of Administrative Sciences*, 79(4), 767-789;
[15] **Pernici, A., Stancu, S. (2023),** ***Performing different clustering methods for mapping the European Union member states using green energy, digitalization, and R&D indicators. A five-year comparison: 2016-2020.*** In *Proceedings of the Computing Conference 2023.* To be published;
[16] **Pérez-Morote, R., Pontones-Rosa, C., Núñez-Chicharro, M. (2020),** ***The effects of e-government evaluation, trust and the digital divide in the levels of e-government use in European countries.*** *Technological forecasting and social change*, 154, 119973;
[17] **Rogers, E.M. (1962),** ***Diffusion of Innovations.*** *Free Press*, New York;
[18] **Reddick, C.G. (2011),** ***Citizen interaction and e- government: Evidence for the managerial, consultative, and participatory models.*** *Transforming Government: People, Process and Policy*, 5(2), 167-184;
[19] **Sabani, A., Deng, H., Thai, V. (2019),** ***Evaluating the development of E-government in Indonesia.*** In *Proceedings of the 2nd International Conference on Software Engineering and Information Management*, 254-258;
[20] **Sharma, S.K., Govindaluri, S.M., Gattoufi, S. (2015),** ***Understanding and predicting the quality determinants of e-government services: A two-staged regression-neural network model.*** *Journal of Modelling in Management*, 10(3), 325-340;
[21] **Taipale, S. (2013),** ***The use of e-government services and the Internet: The role of socio-demographic, economic and geographical predictors.*** *Telecommunications Policy*, 37(4-5), 413-422;
[22] **Thomas, J.C., Streib, G. (2003),** ***The new face of government: Citizen-initiated contacts in the era of egovernment***. *Journal of Public Administration Research and Theory*, 13(1), 83-102;
[23] **Twizeyimana, J.D., Andersson, A. (2019),** ***The public value of E-Government – A literature review.*** *Government information quarterly*, 36(2), 167-178;
[24] **Zheng, Y., Schachter, H.L., (2017***),* ***Explaining citizens' E-participation use: the role of perceived advantages.*** *Public Organization Review*, 17(3), 409–428;
[25] **United Nations E-Government Survey (2014)** *https://publicadministration.un.org/publications/content/PDFs/UN%20E-Government%20Survey% 202014.pdf*. Last Accessed January, 2023.