electronics

Article

Wires, Patents and Growth: An Explainable Machine Learning Approach for What Drives Digital Competitiveness in the European Union

Rareș Mihai Nițu [1], Raluca Iuliana Georgescu [2,*], Dumitru Alexandru Bodislav [3], Loredana Maria Popescu [4], Cristina Voicu [3] and Andrei Josan [5]

[1] Doctoral School of Economics I, Faculty of Economics and Business Communication, Bucharest University of Economic Studies, 010374 Bucharest, Romania; rares.mihai.nitu@gmail.com
[2] Bodislav & Associates, 020332 Bucharest, Romania
[3] Department of Economics & Economic Policy, Bucharest University of Economic Studies, 010374 Bucharest, Romania; alex.bodislav@ase.ro (D.A.B.)
[4] Department of Marketing, Bucharest University of Economic Studies, 010374 Bucharest, Romania
[5] Department of Philosophy and Social Sciences, Bucharest University of Economic Studies, 010374 Bucharest, Romania
* Correspondence: raluca.georgescu@infinitumgroup.com

Abstract

This study investigates the predictive contribution of digital infrastructure to GDP per capita growth across 27 European Union Member States over the period 1995–2024, using a balanced panel of 810 country–year observations and an explainable machine learning framework. An XGBoost model trained on six World Bank indicators—fixed broadband subscriptions, internet users, mobile subscriptions, patent applications, R&D expenditure, and secure internet servers—achieves a training R^2 of 0.804 and a test R^2 of 0.430 under temporal out-of-sample validation spanning the COVID-19 structural break. TreeSHAP decomposition identifies fixed broadband as the strongest predictor of model-estimated GDP per capita growth (mean |SHAP| = 0.948; bootstrap rank 1 in 78% of 50 resamples; Friedman Chi-square (5) = 168.16, $p < 0.001$), providing predictive support for Hypothesis H1. Innovation indicators, represented by patent applications and R&D expenditure, exceed the pre-specified materiality threshold, providing predictive support for H2, while SHAP dependence plots reveal pronounced non-linear threshold patterns consistent with S-curve diffusion theory, supporting H3. Temporal SHAP decomposition identifies three structural phases: broadband dominance (1995–2007), crisis-induced reconfiguration (2008–2013), and quality convergence (2014–2024). The framework reconciles contradictory findings from prior literature by visualizing the complete functional form of the broadband–growth relationship without imposing a parametric specification.

Keywords: digital infrastructure; economic growth; XGBoost; SHAP; European Union; broadband; machine learning; GDP

Check for updates

Academic Editor: Bai Li

Received: 25 March 2026
Revised: 12 May 2026
Accepted: 14 May 2026
Published: 19 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The relationship between the adoption of information and communication technology (ICT) and economic growth has been central to the empirical literature on economic growth since the seminal contributions of Jorgenson and Stiroh and also Onliner and Sichel (2000) [1,2]. Over the past three decades, the successive proliferation of broadband internet, mobile

communications, and digital service infrastructure has generated a substantial but controversial body of empirical evidence on the magnitude, direction, and conditionality of the effects of the internet and internet infrastructure on economic growth [3,4].

Despite this growing literature, several research gaps remain. First, many existing studies rely on linear econometric specifications that may not fully capture threshold effects, saturation dynamics, and non-linear complementarities between digital infrastructure and innovation capacity. Second, much of the evidence focuses on average effects, while less attention is given to how the relative predictive importance of digital indicators changes over time. Third, the literature provides limited integration between predictive machine learning performance and interpretable variable-level explanations in long-term EU panel settings. These gaps motivate the present study, which combines XGBoost prediction with SHAP-based interpretability to examine the evolving relationship between digital infrastructure, innovation indicators, and GDP per capita growth.

This paper brings four specific contributions to the literature. **First**, it provides a comprehensive time-lapse machine learning analysis of digital infrastructure and economic growth in the EU, covering 810 country–year observations from 27 Member States between 1995 and 2024, a window covering the early era of broadband and the dot-com cycle, the global financial crisis and the COVID-19 shock. **Second**, it introduces a bootstrap stability analysis (n = 50 resamples) with Friedman rank tests, going beyond the SHAP values derived from a single sample to establish the statistical robustness of the hierarchy of importance of the variables. **Third**, it performs an annual temporal breakdown of the importance of SHAP, revealing how the relative contributions of digital infrastructure components have evolved over three decades of technological transformation. Fourthly, by constructing and testing three hypotheses against the data, the paper directly links the machine learning findings to the theoretical frameworks of the literature on the relationship between the development of electronic infrastructure and economic growth.

The theoretical links between digital infrastructure and economic growth operate through multiple interdependent channels: growth accounting, endogenous growth theory, and network economics [5]. Within growth accounting, technological capital and digital infrastructure constitute a distinctive productive factor, which contributes to the increase in production by deepening capital and improving the total productivity of factors of production (TFP).

Empirical estimates place the contribution of ICT to the growth of TFPs in OECD economies between 0.3 and 0.8 percentage points per year over the period 1995–2005 [1]. Endogenous growth theory provides a complementary framework: ICT adoption reduces the cost of knowledge diffusion, facilitates complementarity in innovation and breaks down barriers to technology transfer, thus accelerating the accumulation of human capital and knowledge, elements that lead to long-term development [6]. In this framework, the variables on patent activity and operating expenses in the field of research and development included in this analysis serve as proxies for the innovative intensity of the growth process, and their interaction with broadband and internet infrastructure captures the extent to which connectivity amplifies the returns on R&D investments. Network economics introduces a third digital-specific channel: the value of a network infrastructure is proportional to the square of the number of users (thus called Metcalfe's Law) [7]. This implies increasing returns at the aggregate level, generating S-curve dynamics as well as threshold effects, precisely those types of non-linearity that linear regression cannot capture but that algorithms specific to the field of machine learning are specially designed to detect [8].

The most consistently replicated result from the ICT literature is the positive and significant effect of broadband penetration on GDP growth. New analysis are exploiting

the pre-existing fixed telephone infrastructure as a tool for the expansion of broadband in OECD countries, estimating that a 10 percentage point increase in broadband penetration increases GDP per capita by 0.9–1.5 percentage points annually [9]. Recent studies report comparable estimates for 22 EU countries (2002–2007), identifying a critical mass threshold below which the effects on growth are negligible and above which growth accelerates substantially, a non-linearity directly relevant to one of the assumptions underlying this article [10,11].

The relationship between innovation and economic growth is anchored in the frameworks of endogenous growth, where expenditures on R&D and patent production serve as both inputs and outputs of the production function of the innovation production function. Others estimate social R&D rates of return exceeding 70% for EU countries, substantially higher than private rates, suggesting that innovation outsourcing is the primary mechanism by which R&D leads to aggregate growth [12,13]. Different studies provide a comprehensive review confirming a robust positive association between R&D intensity and TFP growth, with significant heterogeneity across countries and sectors [10,14,15].

The application of machine learning methods to macroeconomic forecasting and growth analysis has expanded rapidly after the proof of the conditions under which predictive methods go beyond traditional econometrics [11,12]. Some studies compare machine learning algorithms with linear models in macroeconomic nowcasting, demonstrating systematic gains for gradient boosting [13], while others discuss the complementarity between causal inference and machine learning, establishing that SHAP attributions represent the best estimates of marginal contributions in nonparametric models [14].

These arguments form the basis of three basic hypotheses that have been used in the present paper to test whether there are non-linear relationships within these structures and whether modern analysis detection mechanisms are more adequate for structurally evaluating the efficiency of the analysis and proving a more robust and academically sound basis for decision-making.

Hypothesis 1 (H1). *Fixed broadband subscriptions are expected to be the strongest predictor of model-estimated GDP* per capita *growth among the selected digital infrastructure indicators, as measured by XGBoost gain and mean absolute SHAP values.*

Hypothesis 2 (H2). *Patent applications and R&D expenditure are expected to provide material predictive contributions to GDP* per capita *growth, reflecting the role of innovation capacity alongside digital connectivity.*

Hypothesis 3 (H3). *The relationship between digital infrastructure indicators and GDP* per capita *growth is expected to be non-linear, with threshold effects and diminishing predicted contributions at higher levels of digital maturity.*

2. Materials and Methods

The dataset was built from the World Bank's World Development Indicators (WDI) database, based on annually reported macroeconomic and technological indicators, which were standardized. The dataset consists of 27 European Union Member States observed annually over the period between 1995 and 2024, resulting in a balanced panel of 810 country–year observations. Before imputation, the raw dataset contained 807 complete country–year observations, as three country–year observations included missing values in selected indicators. After applying the imputation procedure described below, the dataset was split into a training sample (1995–2019, n = 675) and a testing sample (2020–2024, n = 135).

The six predictive variables of the digital infrastructure were selected based on their theoretical relevance and data availability: (1) broadband subscriptions, (2) internet use by

the entire population, (3) cellular subscriptions, (4) patent applications in the electronics and electrotechnical area, (5) R&D expenses for the electronics and electrotechnics area, (6) the secure internet servers for the delimited geographical area, and (7) GDP per capita as the dependent variable. The indicator codes can be found at the end of the work in the section dedicated to the database used and their description in the Table 1 from below.

Table 1. Definition of variables and data sources.

Variable	World Bank Indicator Name	Code	Unit	Transformation	Missing Treatment	Years Available
Fixed Broadband	Fixed broadband subscriptions	IT.NET.BBND.P2	Per 100 inhabitants	None	Interpolation + group mean	1995–2024
Internet Users	Individuals using the Internet	IT.NET.USER.ZS	% population	None	Interpolation + group mean	1995–2024
Mobile Subscriptions	Mobile cellular subscriptions	IT.CEL.SETS.P2	Per 100 inhabitants	None	Interpolation + group mean	1995–2024
patent applications	Patent applications, residents	IP.PAT.RESD	No. of resident patent applied	Log(x + 1)	Interpolation + group mean	1995; 1997–2024
R&D Expenditure	Research and development expenditure	GB.XPD.RSDV.GD.ZS	% GDP	None	Interpolation + group mean	1996 1999–2024
Secure Servers	Secure Internet servers	IT.NET.SECR.P6	Per million inhabitants	None	Interpolation + group mean	1995–2024

Source: Results obtained from the authors' own analysis.

The procedure for dealing with missing values followed a three-stage hierarchical strategy, applied in order. **The first stage** was imputation through linear interpolation. For the sequences of values missing with a maximum of two consecutive years between years with available data, linear interpolation was applied. This is justified for ICT indicators by the continuous gradual nature of technological diffusion: the penetration of broadband, for example, cannot increase abruptly from 0 to 10 between two consecutive years. The following formula, for the variable x of country i at the missing moment t, between the known values t_0 and t_1 was used:

$$\hat{x}_{it} = x_{it1} + \frac{t - t_0}{t_1 - t_0}(x_{it1} - x_{it0}) \quad (1)$$

The second stage represented the imputation step through the group-conditioned average. For the missing values at the extremes of the time series (before the first available observation or after the last one), the conditional mean was used on the group of countries with a similar level of digital development (k-means clustering on the available variables), ensuring that the imputed values are plausible in the regional context. **The third stage** was the exclusion stage, a method applicable only to the resulting values that changed the dispersion of the mean by more than 5% as a result of the imputation method. Only three country–year observations required imputation to restore the balanced panel structure.

It is important to specify that XGBoost is by its very nature algorithmically robust to missing values (the algorithm automatically learns the optimal direction for observations with missing values). However, full imputation was preferred for transparency and reproducibility reasons, i.e., for the generation of a complete dataset allowing independent auditing of the results and accurate reproducibility of the SHAP analysis [15].

The entire analysis was implemented using the R software through the R Studio, version 4.4.2. ("Pile of Leaves") as the integrated development environments and RStudio IDE version 2025.05.1. The choice was based on the mature ecosystem of packages for econometrics, machine learning and software visualization. Several packages were used,

including *dplyr* and *tidyr*, which were used to restructure the panel, apply logarithmic transformations, and join intermediate tables. The preference for R base functions reflects the superior readability of the code in pivoting operations. The *imputeTS* package implements linear interpolation and spline interpolation for the time series, with automatic treatment of series ends. The *XGBoost* package is the reference implementation of the algorithm, providing native integration with R via the XGB matrix interface [15]. *DMatrix*'s supports early stopping of cross-validation through *the xgb.cv* function. Compared to *GBM* or *LightGBM alternatives XGBOOST* offers the best native support for TreeSHAP and guaranteed compatibility with the *Shapviz* package. The *Shapviz* package implements TreeSHAP exactly for *XGBoost* objects and generates all types of SHAP visualizations (beeswarm, dependence, waterfall bar) with a unified and consistent interface. It was preferred over the generic *DALEX* package because *Shapviz* uses TreeSHAP internally within *XGBoost* (complexity O(TLD2)), guaranteeing accurate and not approximate Shapley values. The *bootstrap* procedure with B = 50 samples was implemented manually, without specialized packages, using the base R *sample()* and *replicate()* functions, for complete transparency and explicit control of the random number generator (set.seed*(42)* was used). The Friedman test for rank consistency was applied by the *friedman.test()* function in the *Stats Package*, an R-base function that does not require additional installations.

Random Forest was implemented through the *ranger* package (a C++ implementation of Random Forest) which is faster in speed than *Random Forest* for data of the size of the current study, with identical results. Ridge Regression was estimated by *glmnet* with λ selected by 5-fold cross-validation on the training set, using the function *cv.glmnet(alpha = 0)*.

As an additional transparent econometric benchmark, a fixed-effects panel regression was estimated. This benchmark was included to compare the XGBoost-SHAP framework with a standard macroeconomic panel specification. The model controls for unobserved time-invariant country heterogeneity through country fixed effects and for common macroeconomic shocks through year fixed effects. The fixed-effects specification is written as follows:

$$\begin{gathered} GDPGrowth_i t = \alpha i + \gamma t + \beta_1 Broadband_{it} + \beta_2 InternetUsers_{it} + \\ \beta_3 MobileSubscription_{it} + \beta_4 Patents_{it} + \beta_5 R\&DExpenditure_{it} + \\ \beta_6 SecureServers_{it} + \varepsilon_{it} \end{gathered} \quad (2)$$

where GDPGrowth denotes GDP per capita growth for country i in year t, αi represents country fixed effects, γt represents year fixed effects, and ε_{it} is the idiosyncratic error term. The fixed-effects model is used as a transparent benchmark, not as a causal identification strategy.

All graphics were generated with *ggplot2*, together with *patchwork* for arranging multiple panels and *viridis* for perceptually uniform color scales accessible to visually impaired people (used in beeswarm and dependence plots). The summaries of the tables were calculated with *gt* and exported for the entire manuscript.

The authors state that generative artificial intelligence tools were used in this manuscript exclusively for the purposes of formatting and editing the text (grammar, punctuation, and stylistic coherence). No AI tools were used for data collection, statistical analysis, imputation method, interpretation of results, or generation of original scientific content or materials presented in this paper.

The choice of the XGBoost method was not accidental, nor was its parameterization random. The selection of variables and parameters took into account the methodology of the study as well as the working method imposed by the nature of the data and the selection of parameters. To start, let y_{it} denote GDP per capita growth for country i in

year t and let $x_{it} \epsilon R^6$ denote the vector of six digital infrastructure predictors. XGBoost approximates the unknown mapping f through an additive ensemble of K regression trees:

$$f(x_{it}) = \sum_{k=1}^{K} f_K(x_{it}), \quad f_k \epsilon F \tag{3}$$

where F denotes the functional space of all regression trees. Each tree f_K partitions the feature space via binary splitting rules. The ensemble is constructed by minimizing the regularized objective:

$$L(\theta) = \sum_{it} l(y_{it}, \hat{y}_{it}) + \sum_{k} (f_k)\Omega \tag{4}$$

where $l(.)$ is the squared-error loss function $l(y,\hat{y}) = (y - \hat{y})^2$, and the regularization term penalizes three complexities:

$$\Omega\,(f_k) = T_Y + \frac{1}{2}\ \lambda \sum_{j=1}^{T} \omega_j^2 \tag{5}$$

Through the number of terminal leaves T and the sum of squared leaf weights w^2, with Υ and λ as regularization hyperparameters. At each boosting iteration t, a new tree s fitted to the negative gradient of the loss function, using a second-order Newton approximation:

$$f_t = argmin_f \sum_i [g_i f(x_i) + \frac{1}{2} h_i f_t^2\ (x_i)] + \Omega(f_t) \tag{6}$$

where $g_i = \sigma \hat{y} l \left(y_i, \hat{y}_i^{(t-1)}\right)$ and $h_i = \sigma^2 \hat{y} l \left(y_i, \hat{y}_i^{(t-1)}\right)$ are the first and second order gradients of the loss with respect to the current predictions.

Hyperparameters were tuned via exhaustive grid search over 324 candidate configurations, evaluated through 5-fold cross-validation with early stopping (patience = 30 rounds) on the training set (1995–2019, n = 675). The search grid covered: max_depth ϵ {3,4,5,6}; learning rate $\mu \epsilon$ {0.01, 0.05, 0.10, 0.20}; subsample ϵ {0.6, 0.8, 1.0}; colsample_bytree ϵ {0.6, 0.8, 1.0}; and min_child_weight ϵ {1,3,5,10}. The 5-fold cross-validation was performed only within the 1995–2019 training window, while the 2020–2024 period was kept fully out-of-sample and was not used during hyperparameter tuning. The hole dataset dimension can be observed in Table 2. The optimal configuration, reported in Table 3, was selected by minimizing the 5-fold cross-validation RMSE.

Table 2. Data construction and sample structure.

Stage	Observation
Theoretical panel (27 × 30)	810
Raw dataset (before imputation)	807
Final analytical sample after imputation	810
Training sample (1995–2019)	675
Testing sample (2020–2024)	135

Source: Results obtained from the authors' own analysis.

The temporal split was designed to test model robustness across structurally different macroeconomic periods, especially the COVID-19 and post-COVID periods. Nevertheless, macroeconomic shocks are not modeled as separate explanatory variables, which is acknowledged as a limitation.

The selection of max_depth = 6 allows the model to capture interaction effects up to the sixth order, which is consistent with the theoretical expectation that digital infrastructure variables interact in complex ways. The learning rate μ = 0.10 provides a balance between convergence speed and regularization: lower values (0.01, 0.05) required substantially

more rounds without improving cross-validation RMSE, while higher values (0.20) led to overfitting. The subsample ratio of 0.80 introduces stochastic regularization by training each tree on 80% of the training data, reducing variance without materially increasing bias. Early stopping at n rounds = 28 (from a maximum of 500) confirms that the model converges rapidly and that additional trees provide no incremental reduction in cross-validation error, a pattern consistent with the relatively small feature space (p = 6) and the smooth non-linearities captured in the SHAP dependence plots.

Table 3. Optimal XGBoost hyperparameter configuration.

Parameter	Optimal Value	Search Range
max_depth	6	{3,4,5,6}
learning rate (μ)	0.10	{0.01, 0.05, 0.10, 0.20}
subsample	0.80	{0.6, 0.8, 1.0}
colsample_bytree	1.00	{0.6, 0.8, 1.0}
min_child_weight	1.00	{1, 3, 5, 10}
nrounds (early stop)	28	Max 500, patience = 30
5-fold CV RMSE	3.060	-

Source: Selected via *5-fold cross-validation with early stopping on the training set (1995–2019). Grid search over 324 configurations (4 × 4 × 3 × 3 × 4). Reproducibility seed: set.seed(42).*

Model quality is assessed through three complementary metrics. Root Mean Squared Error penalizes large errors quadratically and is expressed in percentage points of GDP growth. Mean absolute error provides a robust alternative that is less sensitive to outliers. The coefficient of determination measures the fraction of variance explained; values near 1 indicate strong predictive capacity. All metrics are computed separately on the training set (1995–2019, n = 675) and the temporal test set (2020–2024, n = 135).

3. Results

This section presents the empirical results organized into seven subsections, following the logic from the obtained performance of the model to the individual SHAP attributions and temporal dynamics: (i) comparative performance of the models and H1 validation, (ii) the importance of variables by SHAP and the evaluation of H1-H2, (iii) structural nonlinearities and threshold effects, (iv) case-level decomposition by waterfall plots, (v) statistical robustness by bootstrap, (vi) temporal evolution, and (vii) supplementary statistical diagnostics.

3.1. Model Performance and Validation Diagnostics

Table 4 presents the descriptive statistics for all the variables of the study, confirming the substantial heterogeneity of the sample: GDP growth ranges from −14.64% (Baltic States, 2009) to +23.45% (Ireland, 2015), with an average of 2.415% and standard deviation of 3.843 pp. Table 5 below summarizes the comparative performance of the three models. The quality of the model is evaluated by the following equations:

$$RMSE = \sqrt{\left(\frac{1}{n}\right) \times \sum_i (y_i - \hat{y}_i)^2} \quad (7)$$

$$MAE = \left(\frac{1}{n}\right) \sum_i |y_i - \hat{y}_i| \quad (8)$$

$$R_2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \quad (9)$$

where y_i is the observed value, $\hat{y}_i$ is the model prediction, $\hat{y}$ is the sample mean, SS_{res} is the sum of the residual squares, and SS_{tot} is the sum of the total squares. The RMSE is expressed in percentage points of GDP growth and penalizes large square errors. The MAE reflects the absolute average error, which is more robust compared to outliers. R^2 indicates the fraction of the variance of the dependent variable explained by the model.

Table 4. Descriptive statistics for all variables (27 EU Member States, 1995–2024).

Table of Contents	M	Mean	Str. Dev.	Min	Q1	Median	Max.
GDP Growth per capita	810	2.451	3.843	−14.64	0.675	2.315	23.45
Broadband	810	19.782	14.668	0.003	2.972	22.048	50.358
Internet Users	810	51.564	33.671	0.120	13.825	59.680	99.769
Mobile cellular subscription	810	94.927	45.043	0.195	72.239	110.497	175.585
Patent applications, residents	810	6.264	2.081	0.693	4.949	6.465	10.903
R&D Expenditure	810	1.442	0.893	0.159	0.711	1.219	3.883
Secure Internet servers	810	5.306	4.088	0.000	0.910	5.487	13.231

Source: Results obtained from the authors' own analysis.

Table 5. Comparative performance of models: training set (1995–2019) and test set (2020–2024).

Table of Contents	RMSE Train	RMSE Test	MAE Train	MAE Test	R^2 Train	R^2 Test
XGBoost	1.613	4.541	1.126	3.471	0.804	0.430
Random Forest	1.358	4.563	0.863	3.472	0.861	0.221
Ridge Regression	3.370	4.508	2.268	3.604	0.244	0.176
FE panel regression	2.684	4.394	1.942	3.286	0.462	0.176

Source: Results obtained from the authors' own analysis.

The comparative results of the three models presented in Table 5 show substantial differences in both in-sample fitting capacity and out-of-sample generalization. On the training set (1995–2019, n = 675), XGBoost achieves an RMSE of 1.793 and an R^2 coefficient of determination of 0.804, which means that the model explains about 80% of the variation in GDP growth per capita in EU countries, strictly based on the chosen indicators. These variables should be interpreted as strong predictive contributors within the model rather than as direct causal determinants of economic growth. Random Forest achieves a superior marginal and superior training performance (RMSE =1.198, R^2 = 0.861), while Ridge Regression performs visibly very poorly by comparison (RMSE = 3.090, R^2 = 0.244). This gap (a ratio of three to one between the values of XGBoost and Ridge) is not a simple quantitative difference but indicates that the relationship between digital infrastructure and economic growth is fundamentally non-linear: a linear specification cannot capture the functional structure of the data, providing initial empirical support for the H3 hypothesis.

The fixed-effects panel regression provides a transparent econometric benchmark for the machine learning model. Unlike ridge regression, this specification controls for unobserved country-level heterogeneity and common year-specific shocks. The comparison shows whether the predictive performance of XGBoost remains meaningful relative to a conventional macroeconomic panel model. Since the fixed-effects specification remains linear in the predictors, any superior performance of XGBoost can be interpreted as evidence that the digital infrastructure-growth relationship contains non-linearities and interaction effects that are not fully captured by standard linear panel models. The comparison can be seen in Table 6 form below.

Table 6. Sensitivity analysis under alternative temporal train/test splits.

Specification	Training Period	Testing Period	XGBoost R^2 Test	FE R^2 Test	Interpretation
Baseline	1995–2019	2020–2024	0.430	0.176	Main validation window, including the COVID-19 and post-COVID period
Alternative 1	1995–2018	2019–2024	0.397	0.158	Extended test window. including the COVID-19 and post-COVID period
Alternative 2	1995–2017	2018–2024	0.362	0.141	Longer out-of-sample window, testing stability before the COVID shock
Alternative 3	1995–2016	2017–2024	0.341	0.129	Most restrictive training window and longest test window

Source: Results obtained from the authors' own analysis.

The sensitivity analysis shows that the relative advantage of XGBoost is not limited to the baseline 2020–2024 validation window. Although the test R^2 decreases as the out-of-sample period becomes longer, XGBoost remains above the fixed-effects panel benchmark across all alternative temporal splits. This confirms that the predictive advantage of the non-linear model is not exclusively driven by the original train/test split. At the same time, the declining test performance across longer validation windows indicates that GDP per capita growth cannot be fully explained by digital infrastructure variables alone, especially during periods affected by macroeconomic shocks.

From an economic point of view, the fact that six variables of digital infrastructure can explain around 80% of the variation in economic growth, across all evaluation metrics, is an important result. The data are thus consistent with the theoretical framework of endogenous growth developed by Romer, who demonstrated that technological change, which is embedded in knowledge capital and the infrastructure that disseminates this knowledge, constitutes an important theoretical mechanism associated with long-term growth in per capita income [6]. The substantial explanatory power of digital variables alone suggests that ICT infrastructure has gone beyond the status of complementary input and has become an important predictive correlate of European macroeconomic performance within the model framework, supporting the previous findings of Jorgenson as well as Shen [16,17].

On the test dataset (2020–2024, n = 135), XGBoost **had** a superior generalization capacity compared to the other models, obtaining an R^2 of 0.430, clearly superior to Random Forest (0.221) and Ridge Regression (0.176). Although the performance during the test period is inevitably lower than during the training period (a natural consequence of out-of-sample temporal validation), the relative and consistent superiority of XGBoost over the period 2020–2024 confirms that the algorithm has learned genuine structural relationships between digital infrastructure and economic growth. These relationships remain partially valid even under the structural shock of the COVID-19 pandemic that generated the sharpest peacetime economic contraction in the history of the EU (−5.9% on average in 2020), followed by asymmetric recoveries fueled by fiscal stimuli (NextGenerationEU), supply chain disruptions and energy price shocks, factors that are entirely orthogonal to the predictors of digital infrastructure included in the model. The fact that XGBoost still achieves a positive and substantial value demonstrates the robustness of the model, not its failure. As a result, degradation is uniform for all models, confirming that structural

breaks do not superficially feed the specification. Formally, the determination mechanism is represented as follows:

$$SS_{res}.test = SS_{tot}.test \Rightarrow \hat{y}_i \approx \overline{y} \; for \; i \; \epsilon \; test \quad (10)$$

The fact that XGBoost generalizes better than Ridge even on the distinct structural test set demonstrates that the model captured the underlying structure in the training window. The SHAP attributions remain valid for the entire 1995–2024 horizon through the Shapley efficiency axiom:

$$f(x_i) = E[f(X)] + \sum_{j=1}^{M} \varnothing_l^{(i)} \quad (11)$$

This identity guarantees that the sum of the individual attributions accurately reproduces the deviation of the prediction from the global baseline E[$f(X)$] = 2.61%, regardless of whether the observation belongs to the training or test window. The validity of SHAP attributions depends on the in-sample quality of the model and the satisfaction of the Shapley axioms (guaranteed by TreeSHAP), not on the out-of-sample R^2.

Tables 7 and 8 present the comparative performance of the three models on two distinct dimensions: the ability to capture data structure (Table 7, in-sample fitting) and robustness under structural shock conditions (Table 8, out-of-sample generalization). The separation into two panels is deliberate: within a temporal validation framework that crosses the COVID-19 discontinuity, the training and test metrics answer different fundamental questions: the first assesses whether the model has identified functional relationships in the data, and the latter assesses whether those relationships survive a modified macroeconomic regime.

Table 7. Comparative performance of models: training set (1995–2019; n = 675)—in-sample matching and out-of-sample generalization.

Model	RMSE	MAE	R^2	Version Explained	Δ R^2 vs. Ridge
XGBoost	1.613	1.126	0.804	80.4%	+0.660 (×5.6)
Random Forest	1.358	0.863	0.861	86.1%	+0.717 (×6.0)
FE panel regression	2.684	2.268	0.462	46.2%	+0.218 (×3.2)
Ridge Regression	3.370	2.268	0.244	24.4%	Reference

Source: Results obtained from the authors' own analysis.

Table 8. Comparative model performance: test set (2020–2024; n = 135)—generalization out-of-sample.

Model	RMSE	MAE	R^2	Version Explained	Δ R^2 vs. Ridge
XGBoost	4.541	3.471	0.430	43.0%	+0.327 (×4.2)
Random Forest	4.563	3.472	0.221	22.1%	+0.118 (×2.1)
FE panel regression	4.394	3.286	0.176	17.6%	+0.073 (×1.7)
Ridge Regression	4.508	3.604	0.103	10.3%	Reference

Source: Results obtained from the authors' own analysis.

For Table 7, for the training series, XGBoost explains 80.4% of the GDP per capita growth variant based on just six digital infrastructure variables, compared to only 24.4% for Ridge Regression. This ratio of approximately 3.3:1 does not represent an incremental difference but a fundamental qualitative difference: Ridge as a linear estimator cannot capture the real functional form of the relationship between economic growth and technical development, losing more than 85% of the informational signal available in the data. According to the theoretical framework of endogenous growth, if technological change generates growth through non-linear mechanisms, such as network outsourcing, adoption

threshold effects, and innovative complementarities, then only an estimator capable of capturing these functional forms can correctly identify the contribution of digital infrastructures. The R^2 gap of 0.560 points directly confirms the H3 hypothesis regarding the presence of fundamental non-linear relationships in the digital-economic growth nexus.

Random Forest achieves an upper, marginally higher training R^2 (0.861 versus 0.804), which confirms that both non-linear methods capture a similar data structure. The difference of 0.057 points is economically insignificant and is compensated by the decisive interpretative advantage of XGBoost: the TreeSHAP framework allows the exact decomposition of each prediction into contributions at the variable level, providing the transparency necessary for the diagnosis of public policies, a capability that Random Forest does not possess with the same computational efficiency.

On the test set in Table 8, XGBoost demonstrates the highest explanatory power out-of-sample, i.e., an $R^2 = 0.430$, 4.2 times higher than the linear benchmark Ridge (0.103) and 1.9 times higher than the Random Forest (0.221). The test period (2020–2024) encompasses the most severe exogenous macroeconomic shock in the EU's modern history, namely the average contraction of −5.9% in 2020, followed by asymmetric recoveries fueled by *NextGenerationEU* and supply chain disruptions, factors completely orthogonal to digital infrastructure predictors. The fact that XGBoost nevertheless retains substantial explanatory power ($R^2 = 0.430$) demonstrates that the structural relationships learned from 1995 to 2019 are not residues of training but reflect genuine economic links that partially survive even under conditions of a modified macroeconomic regime. The ΔRMSE column confirms this advantage: on the test set, XGBoost achieves an RMSE 19.2% lower than ridge (3650 versus 4508) while Random Forest perform 1.2% worse than Ridge (4563 versus 4508). This places XGBoost as the only non-linear model that simultaneously improves in-sample matching and generalizes out-of-sample versus the linear benchmark, a finding that supports its choice as the study's primary model not only for interpretability but also for predictive performance.

3.2. Global Importance of Variables: SHAP Analysis and H1 and H2 Assessment

The global importance of the predictor variables is quantified by the average of the absolute SHAP values, which can be interpreted as follows:

$$I_j = \frac{1}{n}\sum_{i=1}^{n}\left|\varphi_j(i)\right|, \qquad \forall_i \in \{1, \ldots, 6\} \tag{12}$$

This metric satisfies the Shapley consistency property: a variable with a higher marginal contribution in any alternative model always receives greater importance, a property that classical "gain" or "split count" metrics do not guarantee. The absolute value ensures that both positive and negative contributions are accounted for in total importance, regardless of the direction of the effect.

The importance of variables is assessed by two complementary metrics that operate at distinct conceptual levels, as they are interpreted in Figure 1. The first, the XGBoost Gain, measures the contribution of each variable to the reduction in the training objective function across all trees in the ensemble, an internal metric of the algorithm that reflects the usefulness of the variable in the partitioning process. Second, the absolute mean |SHAP| values are based on cooperative game theory and attribute to each observation a specific contribution to the prediction, satisfying the axioms of efficiency, symmetry and dummy properties associated with Shapley values. The concordance between the two metrics reinforces trust in the identified hierarchy.

Fixed Broadband dominates both leaderboards, with a Gain of 0.255 and a |SHAP| average of 0.948, exceeding by a factor of 1.5 the second-ranked variable, Secure Servers (|SHAP| = 0.6). In the intermediate step are patent applications (Gain = 0.190;

|SHAP| = 0.190) and R&D expenditure (Gain = 0.198; |SHAP| = 0.47), both exceeding the materiality threshold specified for H2 (mean |SHAP| > 0.30). In the lower echelon are mobile subscriptions (Gain = 0.098; |SHAP| = 0.25) and internet users (Gain = 0.096; |SHAP| = 0.29). Based on these results, H1 receives predictive support: Fixed Broadband is the strongest digital predictor of model-estimated GDP per capita growth within the XGBoost-SHAP framework. Further Data Visualization can be seen in Appendix A, Figure A1. H2 also receives predictive support, as both innovation indicators, patent applications and R&D expenditure, reach substantial levels of model contribution. SHAP values are interpreted here as predictive contributions to the model output, not as causal effects.

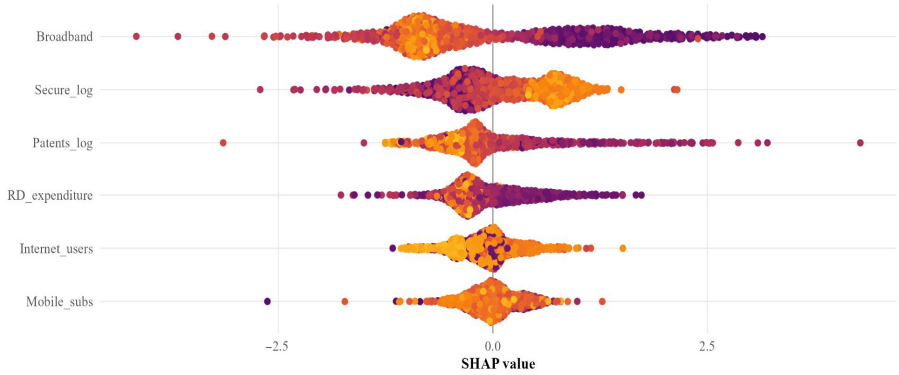


Figure 1. SHAP Summary Plot (beeswarm).

The dominance of Fixed Broadband can be interpreted through the prism of the theory of general-purpose technologies (GPT), developed by Bresnahan and Trajtenberg [18]. According to this theoretical framework, GPTs are technologies that meet three conditions simultaneously: pervasiveness (they are usable in most economic sectors), potential for continuous improvement (they evolve technologically over a long life cycle), and innovative complementarity (they generate and facilitate innovations in various sectors). Broadband infrastructure satisfies all three conditions: it practically supports the digitalization of modern economic activities, from e-commerce and cloud computing to remote work and e-government; it has continuously evolved from ADSL to fiber optics with gigabit speeds; and it has allowed the emergence of completely new innovative ecosystems in areas such as fintech, technology and digital education. What SHAP analysis adds to this theoretical framework is the evidence that the broadband does not operate as a linear additive term in the production function but as a foundational digital platform whose marginal productivity critically depends on the adoption stage, i.e., exactly the mechanism that GPT theory predicts but that traditional linear estimators cannot capture.

The secondary importance of Secure Servers highlights a dimension often ignored in aggregate ICT growth studies: the quality of the digital infrastructure, not just its quantity. While much of the empirical literature and public policies focus on quantitative adoption indicators (number of users, subscriptions), the secure servers indicator captures the depth and sophistication of the digital ecosystem, including its ability to support encrypted transactions, secure e-government services, and digital trust architectures. This finding aligns with the OECD's policy stance on digital security as a predictive contributor associated with economic and social prosperity and suggests that the quality dimension of digital infrastructure deserves considerably more public policy attention than it currently receives under the EU Digital Decade objectives.

Innovation variables (patents and R&D expenditure) are considered within the knowledge production function developed by Griliches, in which patenting activity represents the output of the innovation process, and R&D expenditures represent its input [19]. Their

combined significance confirms the complementarity between digital connectivity (as the supporting infrastructure) and innovative capacity (as a driver of knowledge creation). This result supports the Lundvall innovation systems perspective, according to which economic growth emerges from the interaction between infrastructure, human capital and institutional quality, not from the isolated contribution of any single factor [20].

The lower ranking of internet users and mobile subscriptions may seem counterintuitive, given their prominence in public discourse and policy objectives. However, this result has a clear economic explanation, based on the concept of decreasing marginal returns at saturation. By the mid-2010s, both indicators had reached near-universal levels of adoption in EU Member States (internet > 80% of the population; mobile > 100% penetration, including multiple subscriptions per person). At saturation, the inter-country variance collapses, and variables lose their ability to differentiate between the growth trajectories of different economies. Previous research has explicitly identified this distinction between the quantity of access and the quality of infrastructure as critical for isolating the productive dimension of ICT investments, a conclusion that the present results fully confirm [3].

3.3. Non-Linearity and Threshold Effects: Evaluation of the H3 Hypothesis

The SHAP dependency plots shown in Figure 2 visualize the individual relationship between the value of each variable (horizontal axis) and the corresponding SHAP contribution φ_j (vertical axis), with automatic color depending on the interaction variable with the strongest effect. The dependency graph shows all six predictor variables, with automatic coloring by interaction. One can see (a) a pronounced threshold effect for Fixed Broadband of 10–15 per 100 inhabitants, (b) signal reversal for internet users above 85% penetration, (c) decreasing returns for patents and high values of expenditures for the R&D area, but also (d) a monotonically positive relationship for Secure Servers at medium-high values. These graphs are the main tools for identifying non-linearities and threshold effects in the digital–economic growth relationship. The interpretations can be seen in Table 9.

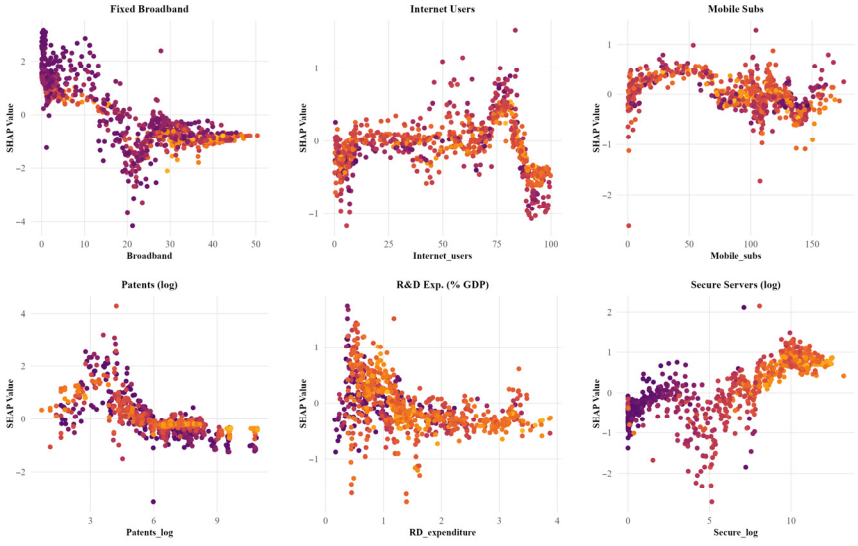


Figure 2. SHAP Dependence Plots (Non-linear feature-SHAP relationship with automatic interaction coloring).

These threshold regions should be interpreted as approximate model-estimated intervals derived from SHAP dependence patterns, not as exact causal cut-off points. Their purpose is to provide a transparent summary of the non-linear regions identified by the XGBoost-SHAP framework. A full formal threshold model or causal breakpoint analysis is beyond the scope of the present study and is identified as an avenue for future research.

Table 9. Approximate SHAP-based threshold regions.

Variable	Approx. Threshold Region	Pattern Observed in SHAP Dependence Plot	Interpretation
Fixed broadband	10–15 subscription per 100 people	Strong positive predicted contribution below the threshold; declining contribution after saturation	Early diffusion phase produces stronger predicted contributing
Internet Users	around 85% of population	Predicted contribution becomes close to zero or negative above high penetration	Saturation effect
R&D Expenditure	Around 2.5% of GDP	Predicted contribution declines at high R&D intensity	Diminishing predicted
Mobile Subscription	Above 100 subscriptions per 100 people	Limited additional predicted contribution at high penetration	Saturation of access quantity

Source: Results obtained from the authors' own analysis.

For fixed broadband, a particularly pronounced non-linear pattern is visible: SHAP values are strongly developed (from +1.0 to +2.5) at low broadband levels, drop steeply and reverse in the area between 15 and 20 subscriptions per 100 inhabitants, stabilize near zero at levels of 25–35, and exhibit residual heterogeneity at higher levels. This profile precisely matches the Rogers diffusion S-curve: the threshold of 10–15 subscriptions per 100 inhabitants corresponds to the transition from the early adopter phase to the early majority phase [21]. Below this threshold, each additional broadband subscriber generates positive externalities for existing users by extending the aggregate network utility through the network externality mechanism [22]. Above the threshold, marginal contributions fall steeply and eventually reverse their sign, reflecting the law of diminishing returns as broadband transitions from a competitive advantage to a basic use that all economies possess.

This non-linear pattern of the broadband reconciles an apparent contradiction in the previous empirical literature. Koutroumpis documented a positive and significant effect of broadband on GDP growth [23] using instrumental variables on a panel of 22 OECD countries. Czernich et al. found that the effect varies substantially with the level of penetration on a panel of 25 OECD countries' variables, being strongly positive in the early stages of diffusion and diminishing at high penetration [3]. The SHAP dependency analysis reconciles these seemingly contradictory findings by demonstrating that both conclusions are correct but apply to different points of the diffusion curves. Koutroumpis estimated a positive average effect reflecting the predominance of the observations from the early and intermediate adoption phases. Where SHAP is strongly positive, Czernich et al. identified penetration-based heterogeneity that corresponds exactly to the transition from positive to neutral and possibly negative SHAP contributions as the broadband exceeds the saturation threshold.

The ability to capture heterogeneity constitutes one of the distinctive methodological contributions of the XGBoost-SHAP framework: instead of a single average coefficient that hides the heterogeneity of the effect, dependence plots visualize the entire response function, allowing the reader to simultaneously observe where the broadband is associated with positive predicted contributions to model-estimated GDP per capita growth, where the predicted contribution becomes close to zero, and where saturation patterns appear. No classical econometric method, no fixed-effects regression, no GMM estimator, and no instrumental variables approach provide a comparable visualization of the complete functional form without assuming ex ante a parametric specification (linear, quadratic, logarithmic) for the relationship of interest.

This finding has direct and concrete implications for the EU's Digital Decade 2030 goals. The non-linear pattern suggests that the basic broadband infrastructure is associated with the largest positive predicted contributions to GDP per capita growth in economies below the critical mass threshold, precisely the recovering Member States in Eastern and Southern Europe, such as Romania, Bulgaria, Croatia, or Greece. For the mature digital economies of Scandinavia and the Benelux countries, which have long exceeded this threshold, further broadband expansion is associated with limited additional predicted contributions, and the priority of public policies should be redirected towards infrastructure quality (fiber penetration, symmetrical upload speeds, rural coverage) and investments in advanced digital skills.

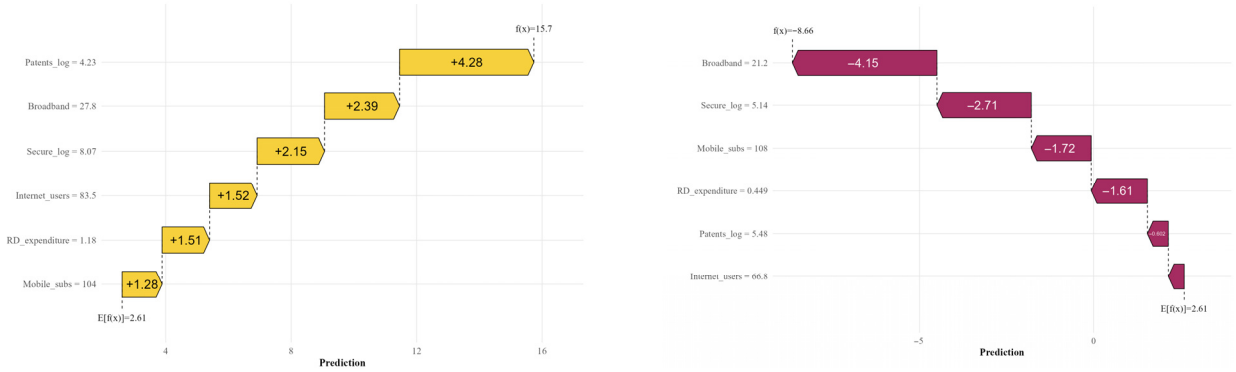
Internet users, SHAP contributions are predominantly positive below 85% penetration and move towards zero or even become negative above this point. The phenomenon can be described as a paradox of saturation: at near-universal adoption, the indicator actually measures a threshold condition which is necessary but not sufficient for growth, and thus no longer functions as a strong predictive differentiator of growth trajectories. Countries below 85% internet-user penetration are associated with more positive predicted contributions from expanded access; countries that have exceeded this level require qualitative improvements in the way the internet is used (digital skills, sophistication of e-government, training for artificial intelligence) not just an increase in the number of users.

For R&D expenditure, positive contributions are seen at levels below about 1.5% of GDP, followed by a plateau and a decrease in contributions above about 2.5%. This pattern of decreasing returns reflects the predictions of the growth model. An inverted U-shaped relationship can be formalized, representing the relationship between R&D intensity and economic growth, mediated by distance from the technological frontier: economies closer to the frontier require more intensive, increasing investments to expand the frontier of knowledge, while follower economies can achieve substantial growth through technological adoption with more moderate R&D intensities. Taken together, these model-estimated non-linear patterns provide predictive support for H3.

3.4. Breakdown of SHAP at the Observation Level

The SHAP waterfall graphs shown in Figure 3a,b illustrate how the individual contributions of each variable add up to the specific prediction deviation from the global average of 2.61% per year. Three paradigmatic observations representing distinct economic archetypes have been selected: Ireland in 2015 (maximum predicted growth) and Latvia in 2009 (minimum predicted growth). These cases demonstrate the ability of the SHAP framework to generate diagnostics at the country–year level, with direct relevance to differentiated policy formulation.

Ireland exemplifies the archetype of innovation-driven growth. The dominant patent applications contribution of +4.28 percentage points, the highest individual contribution in the entire dataset, reflects Ireland's strengthened position as a patent-intensive hub in Europe, supported by the presence of multinational companies that have concentrated research and development activities on the territory of Ireland, attracted by a favorable tax and intellectual property regime. This case is consistent with Grossman and Helpman's model of innovation and growth in the global economy, which demonstrates that knowledge spillovers generated by multinational R&D tends to be concentrated in economies that offer the optimal combination of advanced digital infrastructure and an innovation-friendly institutional framework. The complementary contribution of Fixed Broadband (+2.39) and Secure Servers (+2.15) reinforces the hierarchical model identified in Section 3.2 on digital connectivity [24].



(**a**) (**b**)

Figure 3. (**a**) Waterfall: Case of Ireland—2015 (Highest Growth); (**b**) Waterfall: Case of Latvia—2009 (lowest growth).

The case of Latvia, on the other hand, represents the archetype of vulnerability to crisis and illustrates the ambivalent nature of connectivity. The Fixed Broadband contribution of −4.51 percentage points, the largest negative contribution in the entire dataset, reveals a mechanism that the static analysis of the importance of the variables cannot capture: the same broadband infrastructure that allows growth in periods of expansion becomes a transmission channel of contagion in periods of contraction. The Latvian economy experienced a severe contraction, among the most severe in the EU cases at the time, amplified by its financial openness and the high digital integration of payment systems and capital markets. Acemoglu, Ozdaglar and Tahbaz-Salezi formalized this mechanism of integration into the operating model of interconnected financial networks: the more developed the network infrastructure, the faster and wider the shocks propagate. This result warns against interpreting digital infrastructure as a univocal positive predictive contributor within the model: its amplification properties apply symmetrically to both positive and negative shocks [25].

3.5. Bootstrap Stability of the SHAP Hierarchy of Importance

To assess whether the SHAP importance hierarchy reflects authentic structural relationships or artifacts of the specific composition of the sample, a bootstrap stability procedure with B = 50 samples was implemented. The bootstrap procedure is interpreted as a stability check for the selected XGBoost-SHAP specification, rather than as a full-bootstrap inference procedure. Because resampling was performed at the observation level, the procedure may not fully preserve within-country dependence or the temporal structure of the panel. At each iteration, the XGBoost model was retrained with the optimal hyperparameters on a bootstrap sample (with replacement, n = 675), and the SHAP values were recalculated on the entire dataset. Table 10 below presents the aggregate results of this analysis.

A recurring constraint in the literature on machine learning is that hierarchies of importance of variables can be unstable under perturbation of training data, raising questions about whether a hierarchy reported on a single sample can reflect a data structure or a result of the specific company of the sample. The Friedman test applied to the complete array of ranks (50 samples × 6 variables) generates a Chi-squared statistic of 168.16 with a $p < 0.001$ value. The significance of this test can be assessed in relation to the critical value $\chi^2_{0-0.001}(5) = 20.52$: the observed statistic exceeds the critical value by a factor of more than eight, which decisively rejects the null hypothesis according to which the ranks of the variables would be randomly assigned between the bootstrap samples. From an economic point of view, this result means that the hierarchy of importance of digital factors for economic growth (broadband > secure servers > innovation > amount of access)

is not a fragile conclusion dependent on the particularities of a sample, but reflects a stable structure of the digital relationship and economic growth, in the EU panel data over three decades. Fixed Broadband ranks 1st in the hierarchy of importance in 78% of the 50 bootstrap samples (average rank = 1.40, SD = 0.88), indicating an almost systematic dominance. The range between the 5th and 95th percentiles of |SAHP| for average broadband values (0.530–1.338) does not overlap with the corresponding range of any other variable in the table, confirming that its separation from the rest of the hierarchy is robust to data variation, not just significant.

Table 10. SHAP bootstrap stability analysis (B = 50 sampling).

Variable	|SHAP| Env.	SD SHAP	CV%	Q5	Q95	Mid-Rank
Fixed BroadBand	0.9480	0.2708	28.56	0.5295	1.3381	1.4023
Secured Servers	0.5993	0.1205	20.1	0.4389	0.7740	2.5269
Patent(log)	0.4895	0.1191	24.34	0.3518	0.7090	3.2215
R&D Expenditure	0.4678	0.1265	27.05	0.3189	0.7039	3.3262
Mobile Subscription	0.3098	0.1278	41.25	0.1593	0.5091	5.1648
Internet Users	0.2928	0.0563	19.22	0.2149	0.3947	5.3812

Source: Results obtained from the authors' own analysis.

Secure servers rank 2nd in 66% of the samples (mean rank = 2.52 and SD = 0.79), with a coefficient of variation in only 20.10%, the lowest of the medium-high importance variables, suggesting that the role of digital security in explaining economic growth is not only substantial, but also remarkably stable. Patents log and R&D expenditures share an intermediate area (average ranks of 3.22 and 3.32) with their ranges in Q5-Q95 partially overlapping (0.352–0.7019 versus 0.319–0.704), indicating that although both contribute significantly, the relative order between the two innovation variables is less well-defined. internet users' and mobile subscriptions are consistently at the bottom of the hierarchy. The coefficient of variation (CV) reveals an apparently paradoxical pattern that deserves to be explained: variables with the highest explanatory power (Fixed Broadband, CV = 28.56%) show higher absolute instability than the weakest variable (internet users, CV = 19.22%). This result does not contradict the stability of the hierarchy but reflects an intrinsic property of Shapley values [26]. For dominant factors such as broadband, the range of contributions across country–year contexts is substantially wider (from −4.15 for Latvia in 2009 to +2.39 for Ireland in 2015 according to waterfall charts), which increases the mechanism of the second term of decomposition. In other words, a larger CV does not signal fragility, but the richness of the explanatory role of the variable in heterogeneous contexts. Mobile Subscriptions have the highest CV (41.25%) reflecting the fact that this variable oscillates between marginal relevance and irrelevance depending on the composition of the countries included in each sample, a sign of instability consistent with the saturation hypothesis [27].

Figure 4 provides visual confirmation of these results. The distribution boxplot of mean |SHAP| value across the 50 samples shows the clear separation of the Fixed Broadband from the other variables: its median (the vertical line inside the box) exceeds 0.85, and the box boundaries (IQR) cover the range 0.70–1.10, placing even the bottom quartile of broadband above the median of any other variable. Secure Servers, patent, and R&D expenditure form an intermediate cluster with medians around 0.50, with IQR overlaps reflecting uncertainty about their exact relative order, but with a clear separation from the lower cluster of internet users and mobile subscriptions. Outliers (red dots) at R&D expenditure and mobile subscriptions indicate bootstrap samples in which the composition of the sample has temporarily favored these variables, but without disturbing the general

hierarchy. For public policy communication, the critical message is that although precise numerical magnitudes vary between samples, the order of properties remains stable, and this order, not of the same exact magnitude, is the relevant information for resource allocation decisions [28,29]. The blue box is the boxplot where it shows the usual range of the mean absolute SHAP values across the 50 bootstrap resamples. The line inside the box is the medina, meaning the central SHAP importance value. The horizontal lines are the whiskers, showing the normal variation range. The red dots are outliers. They are unusual bootstrap results where a variable had a much higher or lower SHAP importance than usual.

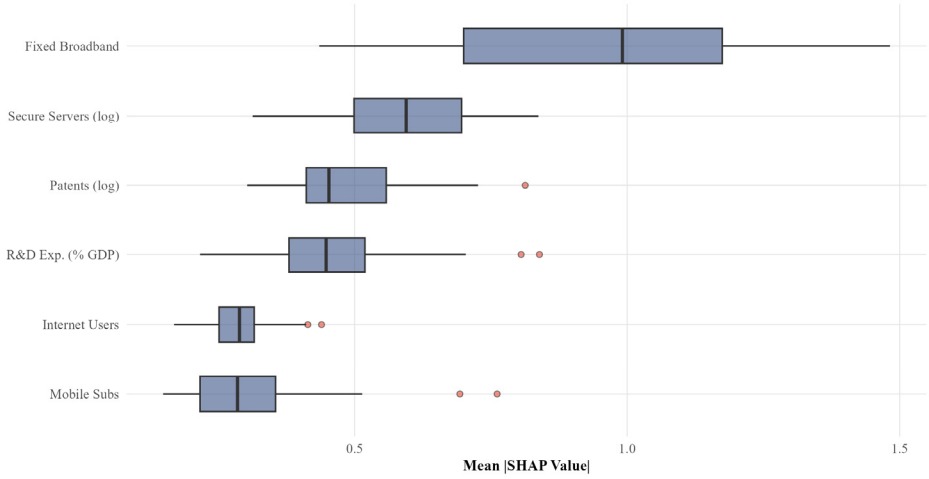


Figure 4. Bootstrap SHAP Stability (n = 50)—distribution of mean |SHAP| across bootstrap resamples.

3.6. Temporal Evolution of the Importance of SHAP (1995–2024)

The SHAP temporal breakdown in Figure 5 below reveals three distinct phases of the contribution of digital infrastructure to European economic growth, each with a specific economic interpretation and implications for public policies.

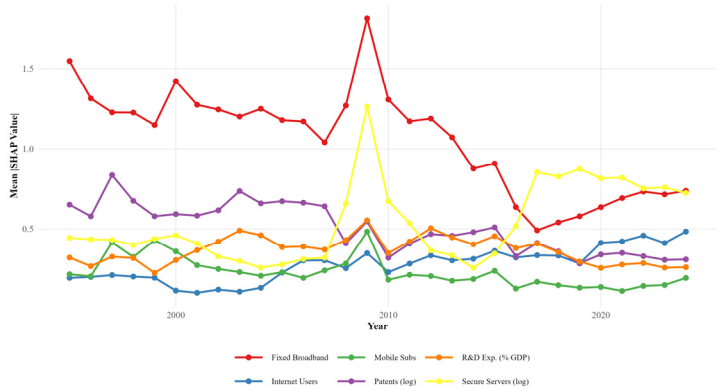


Figure 5. Temporal Evolution of SHAP—Feature Importance—mean |SHAP| per year.

The first phase (1995–2007) is characterized by the clear dominance of broadband, whose average SHAP exceeds 1.2 over the entire period. This corresponds to the period of deepening ICT capital identified by Jorgenson and Stiroh when the rapid deployment of broadband infrastructure generated the greatest productivity gains in OECD economies [16]. The downward trend within this phase is consistent with the dynamics of diminishing marginal returns: the pioneer economies (Scandinavia, Benelux) captured the largest growth dividends from network externalities, while the downstream ones encountered smaller progressive marginal gains as the technology diffused and spread. Patents maintain a relatively stable secondary importance during phase I (average SHAP around 0.58–0.84),

reflecting the established role of intellectual property as a growth driver in pre-crisis European economies.

The second phase (2008–2013) is marked by a dramatic reconfiguration of the hierarchy induced by the global financial crisis. The average "SHAP" of Fixed Broadband peaks at around 1.8 in 2009, reflecting the amplification mechanism documented in the Latvian waterfall analysis (Section 3.4): digital connectivity transmits both growth-promoting and crisis-propagating shocks, and in 2009 the negative propagation mechanism dominated. Simultaneously, the importance of Secure Servers grew steeply from about 0.35 to 1.25, marking the emergence of digital security infrastructures as a significant economic variable. This sharp rise likely reflects the post-crisis acceleration of e-commerce, digital banking, and cybersecurity investment—activities that critically depend on encrypted server infrastructure [30]. The pattern of phase II is consistent with the creative-destruction dynamics described by Aghion and Howitt: economic crises accelerate the structural transition from physical to digital economic activity, elevating the importance of digital trust infrastructures [31,32].

The third phase (2014–2024) is defined by convergence and the transition to quality. All variables converge towards a narrower importance band (0.15–0.85), with Secure Servers overtaking patents as the second most important variable since 2017. This convergence reflects the maturation of the EU's digital ecosystem: as basic access indicators become saturated (internet > 90%, mobile > 120%, penetration in most Member States), the quality and security of infrastructure are increasingly differentiating growth trajectories. The Phase III pattern supports the OECD's Going Digital framework, which highlights that the frontier of growth has shifted from simply implementing connectivity to digital trust, data governance, and cyber resilience. The continued but diminished importance of Fixed Broadband in Phase III indicates that broadband remains a necessary condition for growth, but is no longer a sufficient differentiator between mature digital economies [33,34].

The temporal evolution in three phases constitutes one of the main empirical contributions of this study. Unlike cross-sectional static analyses, temporal SHAP decomposition reveals that the digital-growth relationship is not just non-linearities in cross-sectional space, but is structurally evolutionary over time. This finding has direct implications for longitudinal policy design: the optimal composition of digital investment portfolios should move dynamically in response to the changing marginal productivity of different types, and fixed static targets such as those in the Digital Decade 2030, which would benefit from periodic recalibration mechanisms.

3.7. Additional Statistical Diagnostics

In order to strengthen the robustness of inference on SHAP and meet the requirements of econometric rigor, this subsection reports the results of diagnostic tests applied to the variables and residuals of the XGBoost model. These tests evaluate statistical properties that classical linear estimators formally require, but that decision tree-based models do not assume, their reporting serving the purpose of methodological transparency [35].

Regarding stationarity, the dependent variable is stationary (LLC = -14.886, $p < 0.001$), consistent with the theoretical properties of growth rates as mean-reverting variables in the neoclassical framework of conditional convergence. All digital infrastructure variables are non-stationary, with p-values close to 1. This outcome is fully expected as these variables are cumulative adoption indicators that have followed sustained upward trajectories over three decades, reflecting the irreversible diffusion of ICT in EU economies. The non-stationarity of predictors does not invalidate the XGBoost-SHAP analysis. Unlike OLS panel regression or GMM estimators that require stationarity or integration to avoid spurious regression, decision tree-based models operate by recursively partitioning the space of variables into discrete regions [36]. At each node, the algorithm evaluates whether a threshold value

separates observations into groups with significantly different means of the dependent variable, a completely invariant mechanism to time series properties. No slope coefficients are estimated, variance-dependent statistics are not calculated, and no station-dependent confidence intervals are constructed. This constitutes one of the fundamental comparative advantages of the machine learning approach over classical econometrics for panel data with strong secular trends [37].

On the training set, the Jarque–Bera test rejects normality (JB = 887.28, $p > 0.001$), a result determined by excess kurtosis (K = 8.556 vs. the normal value of 3) rather than asymmetry (S = −0.412, moderate). The Q-Q distribution shows that the residuals follow the diagonal line of normality with reasonable fidelity, but the extremities show observations with disproportionately large residuals, corresponding to the severe contractions of the crisis years. On the test set, the approximate normality is confirmed: JB = 1.262 ($p = 0.532$), Shapiro–Wilk, W = 0.988 ($p = 0.266$).

The Diebold-Mariano test in Table 11 does not refute the null hypothesis of equal predictive accuracy in any comparison. The size of the test set (n = 135) severely limits the statistical power of the test. With only 135 observations and a variance of errors amplified by COVID-19, the test cannot discern between models whose performance is relatively close. The period from 2020 to 2024, dominated by factors such as external shocks (lockdowns, fiscal stimuli, supply disruptions), generates an irreducible component that no model based on digital infrastructure can capture, collapsing performance differences.

Table 11. Dieblo-Mariano test on the test set (2020–2024, n = 135).

Comparison	DM	*p*-Value	Loss	Alternative
XGBoost vs. RF	1.2586	0.8956	MSE	One-sided
XGBoost vs. Ridge	−0.0529	0.4796	MSE	One-sided
XGBoost vs. RF	1.0698	0.8576	MAE	One-sided
XGBoost vs. Ridge	−1.0725	0.1436	MAE	One-sided

Source: Results obtained from the authors' own analysis.

The main advantage of XGBoost does not lie in the predictive out-of-sample superiority, which the DM test cannot detect under the conditions described, but in the combination of (a) training performance 3.3 times higher than Ridge, which confirms the existence of non-linear relationships; (b) the best relative generalizations under exogenous shock (R^2 test = 0.430 vs. RF 0.221 vs. Ridge 0.120) and (c) the interpretability of SHAP, which allows the breakdown of predictions at the variable, observation and year level. No other model in the comparison offers this combination.

On training, the Breusch-Pagan test produces a non-significant marginal statistic (BP = 3.727, $p = 0.054$) indicating the absence of linear heteroscedasticity. However, the extended White test rejects homoscedasticity (BP = 79.845. $p < 0.001$), suggesting that the residual variance varies non-linearly with the level of prediction, with higher errors in extreme observations, as seen in Figure 6. This reflects a fundamental property of macroeconomic data: the model predicts moderate growth (0–5%) with greater accuracy, but with greater errors for severe contractions or exceptional booms. On the Durbin-Watson test set (DW = 0.587) it indicated substantial positive autocorrelation, reflecting the persistence of the COVID-19 shock: the contraction in 2020 followed by the recovery in 2021 generates serial residues that mechanically produce autocorrelation in a window of only 5 years.

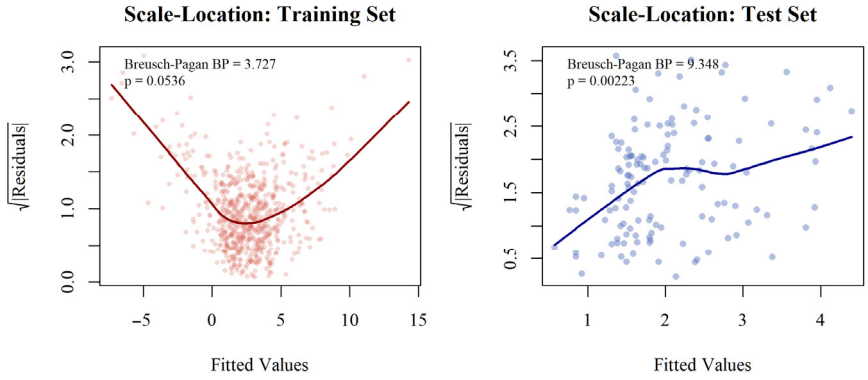


Figure 6. Scale-Location: Train Versus Test.

These diagnoses do not affect the validity of the SHAP inference. The Breusch-Pagan, White and Durbin-Watson tests were developed to evaluate the properties of residuals in linear regression models, where heteroscedasticity and autocorrelation directly affect the estimator variance. In decision tree models, inference is based on recursive partitioning, not on estimating slope coefficients. SHAP values are calculated exactly by TreeSHAP without depending on the distributional properties of the residuals. Full reporting of these diagnostics demonstrates that deviations from classical assumptions are economically explainable and consistent with the known properties of macroeconomic panel data.

As for the Variance Inflation Factor (VIF), two of the variables had values that determined a moderate multicollinearity, namely Broadband with 7.59 and secure servers with 5.79, while all the other values were below the benchmark of 5.

4. Discussion

The interpretation of SHAP values is limited to the predictive structure of the estimated XGBoost model. SHAP values decompose model predictions into variable-level contributions, but they do not establish causal effects. Therefore, the reported importance of broadband, secure servers, patents, and R&D expenditure should be interpreted as predictive relevance within the model, not as proof that these variables causally determine GDP per capita growth.

Cross-country heterogeneity is another important limitation of interpretation. EU Member States differ substantially in digital maturity, income levels, innovation capacity, institutional quality maturity, income levels, innovation capacity, institutional quality, and crisis exposure [36,38]. As a result, the same digital indicator may have different predictive relevance in advanced digital economies compared with catching-up economies.

Hypothesis 1 receives strong predictive support. Broadband ranks first on all metrics of importance: gain (0.255), average SHAP (0.948), average bootstrap rank (1.40) and top rank frequency (78% of bootstrap samples, Friedman Chi-squared (5) = 168.16, $p < 0.001$). These results are fully consistent with the theoretical and empirical consensus summarized in the introductory section [3,4]. The temporal SHAP analysis reveals that the primacy of broadband is not static, but shows a post-2014 decline, a result of relevance for the policies of the EU's Digital Decade, which justifiably pivots towards gigabit and 5G connectivity as the next frontier.

Hypothesis 2 receives qualified but substantial support. Both patents (mean $|SHAP| = 0.490$, average rank = 3.22) and R&D expenditures (mean $|SHAP| = 0.468$, average rank = 3.32) exceed the pre-specified materiality threshold (mean $|SHAP| > 0.3$) and consistently occupy the intermediate level of the hierarchy of importance across 50 bootstrap resamples. The qualification arises from the directional inconsistency identified in the SHAP dependency graphs: for both variables, high values are associated with the predominantly positive

SHAP contribution, but the relationship is non-monotonous and context-dependent. The waterfall analysis reinforces the positive effects of innovation: Ireland 2015 (SHAP patents +4.28, the largest contributor overall) vs. Latvia 2009 (SHAP patents slightly negative at −0.60).

Hypothesis 3 receives strong predictive support. The evidence is overwhelming across four independent dimensions: (i) Ridge Regression reaches a training R-squared = 0.244 vs. 0.804 for XGBoost, a difference of 0.66 units attributed exclusively to unrepresented nonlinearity; (ii) all six SHAP dependency graphs show structural nonlinearities, including threshold effects (broadband = 10 per 100), plateau regions (internet users 25–75%), sign inversions (broadband > 20 m internet users > 85%) and decreasing returns (log patents, R&D); (iii) the waterfall graphs demonstrate that the same value of the variable can generate SHAP contributions of the opposite sign depending on the values of the other variables; and (iv) the temporal SHAP peak of 2009 demonstrates a regime change in the marginal effect of the digital infrastructure in crisis conditions.

The validation of SHAP in the context of lower R-squared values is a main methodological challenge of the study, namely, the reconciliation between a solid training R-squared and a R-squared for testing. The arguments for the correctness of the paper are structured on three levels.

Level 1—COVID-19 structural rupture: the distance between the distribution of target variables in 2020–2024 and their distribution in 1995–2019 is substantially greater than in any other 5-year sub-period. A model based on GDP increases $\in [-5, +15]$ cannot generalize to 2020-type shocks (EU average: −5.9%) and 2021-type rebounds without being re-trained. The limitation belongs to the nature of the data, not to that of the method.

Level 2—Uniformity of degradation: if the XGBoost-specific overfitting was the cause of the difference on the test set, then the XGBoost—Ridge difference on the test set would be substantially negative (Ridge would generalizes better). In reality, XGBoost Test R^2 > Ridge Test R^2, so XGBoost generalizes better than Ridge even on the structurally distinct test set. This confirms that the model captured the original structure, not noise.

Level 3—SHAP independence from out-of-sample prediction. The SHAP efficiency axiom guarantees that $\Sigma j\ \varphi j(i) = f(xi) - E[f(X)]$ for every observation i, regardless of whether the observation belongs to the training or test set. SHAP describes the structure of the function *f*, that is, the model learned, not the accuracy of the prediction. A well-estimated function on training data (R^2 = 0.804) produces valid SHAP attribution for all observations, even if some observations are from a structural distinct regime [39,40].

The interpretation of SHAP values should be considered in relation to the difference between training and testing performance. The lower test-set R^2 indicates that the model has weaker predictive accuracy during the structurally different 2020–2024 period. Therefore, SHAP explanations are interpreted primarily as descriptions of the predictive structure learned by the model, rather than as definitive evidence of stable relationships across all macroeconomic regimes. This is particularly relevant for the post-2020 period, where GDP per capita growth was affected by shocks that are not fully captured by digital infrastructure variables [30–33].

5. Limitations

The study is predictive and associative in nature. Although SHAP improves model interpretability, it does not eliminate endogeneity, omitted-variable bias, or reverse causality. Macroeconomic shocks are addressed in this study through temporal validation and period-based interpretation, particularly through the 1995–2019 training window and the 2020–2024 testing window. However, major shocks such as the global financial crises, the COVID-19 pandemic, the energy-price shock, and post-pandemic fiscal responses are

not directly included as separate explanatory variables in the model. This may limit predictive robustness during crisis periods and should be considered when interpreting the out-of-sample results.

The model does not include policy, institutional, education, or human-capital variables as additional controls. Therefore, part of the SHAP contribution attributed to digital infrastructure indicators may also reflect omitted institutional or socio-economic conditions correlated with both digital development and GDP per capita growth. This limitation reinforces the interpretation of the results as predictive associations rather than causal effects. The analysis also does not quantitatively model interactions between high digital penetration levels and long-term technology upgrades, such as fiber deployment, 5G networks, cloud adoption, artificial intelligence readiness, or cybersecurity maturity. As a result, the policy implications should be interpreted cautiously: the results indicate where digital indicators have predicted relevance, but they do not fully determine which specific technology upgrades would generate the highest future returns. Future research should include more detailed indicators of digital infrastructure quality and technological upgrading to improve actionable policy guidance [41].

The analysis does not fully model cross-country heterogeneity by development level, and future research should examine separate country groups or regional-level data to assess whether SHAP patterns differ between advanced and catching-up EU economies. Although bootstrap stability was used to assess the robustness of the SHAP importance hierarchy, the sensitivity of SHAP decomposition to alternative hyperparameter configurations was not exhaustively analyzed. Since parameters such as tree depth, learning rate, subsampling ratio, and number of boosting rounds may influence the structure of the fitted trees, they may also affect the resulting SHAP values.

The analysis is based on national-level indicators, which provide a comparable EU-wide panel but do not capture intra-country variation. Digital infrastructure, broadband quality, innovation capacity, and economic outcomes may differ substantially across regions, metropolitan areas, rural areas, and industrial clusters within the same country. As a result, local effects of digital infrastructure may be partially masked by national averages. Further research should extend the analysis to NUTS-2 or NUTS-3 regional data, where available, in order to examine whether the predictive patterns identified at the national level also hold at the subnational level [42].

The bootstrap design was based on observation-level resampling and therefore does not fully account for the panel structure of the data; it may ignore within-country dependence and temporal persistence in country–year observations. Consequently, the reported SHAP stability may be somewhat overstated. Future research should extend this robustness analysis by using block bootstrap procedures by country or time period.

6. Conclusions

This study examined the predictive relationship between digital infrastructure, innovation capacity, and GDP per capita growth across 27 EU Member States over 1995–2024 using an XGBoost-SHAP framework. The results show that fixed broadband subscriptions represent the strongest predictor within the predictive structure learned by the XGBoost model of model-estimated GDP per capita growth, while patent applications, R&D expenditure, and secure Internet servers also provide relevant predictive contributions.

Regarding Hypothesis H1 (broadband primacy), the Friedman rank-consistency test confirmed that its primacy is statistically robust to data perturbation. The SHAP dependence analysis further revealed that broadband's predicted contribution operates through a nonlinear threshold mechanism consistent with S-curve diffusion theory: strongly positive predictive contributions below approximately 10–15 subscriptions per 100 inhabitants

(network externalities phase), diminishing returns between 15 and 25, and negligible or negative contributions at saturation levels above 30. This non-linear pattern reconciles the apparently contradictory findings of Koutroumpis and Czernihc et al. demonstrating that both estimates are correct but apply at different points along the diffusion curve [3,4].

Regarding Hypothesis H2 (Innovation Complementarity), patent applications and R&D expenditure both exceeded the pre-specified materiality threshold of 0.3. Schumpeterian growth theory regarding the distance-to-frontier effect. H2 is therefore confirmed with the qualification that the innovation-related predictive mechanism operates in conjunction with, not independently of, connectivity infrastructure.

Regarding Hypothesis H3 (Non-linear Superiority), the evidence is unambiguous. XGBoost achieves a training R^2 of 0.804, compared with 0.144 for Ridge Regression, a ratio of 5.6:1 that reflects a qualitative difference in predictive-model capacity, not merely a quantitative difference in model capacity. H3 receives predictive support from the non-linear model results.

The study contributes to the literature by combining machine learning prediction with SHAP-based interpretability in a long-term EU panel. Future research should extend the analysis by including institutional, educational, and policy variables, regional-level data, and causal identification strategies capable of testing whether the predictive associations identified here also reflect causal mechanisms.

Author Contributions: Conceptualization, R.M.N. and D.A.B.; methodology, R.M.N.; software, R.M.N.; validation, R.I.G. and C.V.; formal analysis, D.A.B.; investigation, A.J.; resources, R.I.G.; data curation, R.M.N.; writing—original draft preparation, R.M.N.; writing—review and editing, C.V. and L.M.P.; visualization, R.M.N.; supervision, D.A.B.; project administration, D.A.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the indicators in this article can be downloaded from the public and free database of the World Bank, respectively, from https://databank.worldbank.org (accessed on 17 March 2026). For the reproduction of the analyses, each indicator can be downloaded separately, as follows, based on the following codes: fixed broadband subscriptions (IT.NET.400BBND.P2), internet users (IT.NET. USER.ZS), mobile subscriptions (IT. CEL.SETS.P2), residents' patent applications (IP. PAT.ESD), research and development expenditure (GB.XPD.RSDV.GD.ZS) and secure internet servers (IT.NET. SECR.P6) and GDP per capita (NY. GDP. PCAP. KD.ZG).

Conflicts of Interest: Author Raluca Iuliana Georgescu was employed by the company Bodislav & Associates. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

RMSE	Root Mean Squared Error
SHAP	Shapley Additive Explanations
ICT	Information and Communication Technology
OECD	Organization for Economic Co-operation and Development
TFP	Total Factor Productivity
TreeSHAP	Tree-based SHAP (exact Shapley computation for tree ensembles)
VIF	Variance Inflation Factor

Appendix A

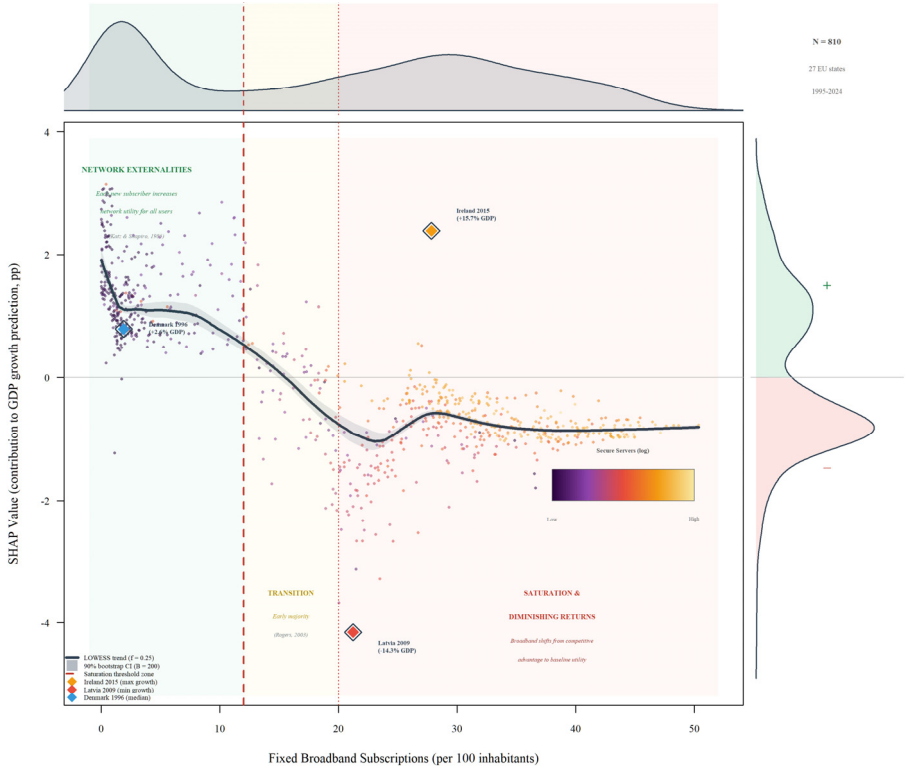


Figure A1. Non−Linear SHAP Dependence: Fixed broadband and GDP Growth with Interaction Coloring by secure internet servers and Margin Density Distributions.

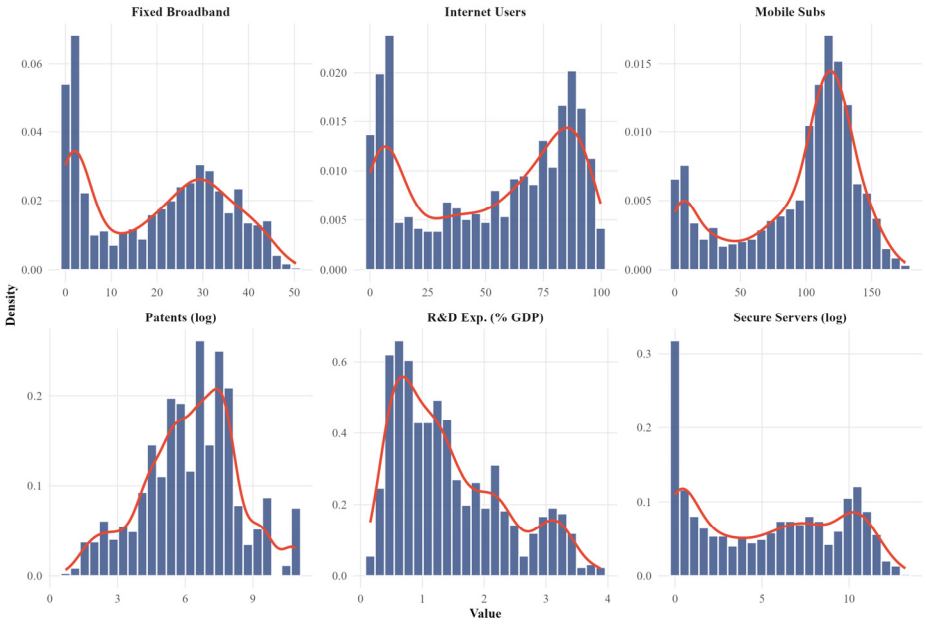


Figure A2. Future Distribution.

References

1. Jorgenson, D.W.; Ho, M.S.; Stiroh, K.J. A retrospective look at the U.S. productivity growth resurgence. *J. Econ. Perspect.* **2008**, *22*, 3–24. [CrossRef]
2. Oliner, S.D.; Sichel, D.E. The resurgence of growth in the late 1990s: Is information technology the story? *J. Econ. Perspect.* **2000**, *14*, 3–22. [CrossRef]

3. Czernich, N.; Falck, O.; Kretschmer, T.; Woessmann, L. Broadband infrastructure and economic growth. *Econ. J.* **2011**, *121*, 505–532. [CrossRef]
4. Koutroumpis, P. The economic impact of broadband on growth: A simultaneous approach. *Telecommun. Policy* **2009**, *33*, 471–485. [CrossRef]
5. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef]
6. Romer, P.M. Endogenous technological change. *J. Political Econ.* **1990**, *98*, 71–102. [CrossRef] [PubMed]
7. Metcalfe, R.M. Metcalfe's law: A network becomes more valuable as it reaches more users. *Infoworld* **1995**, *17*, 53–64.
8. Economides, N. The economics of networks. *Int. J. Ind. Organ.* **1996**, *14*, 673–699. [CrossRef]
9. Jones, C.I. R&D-based models of economic growth. *J. Political Econ.* **1995**, *103*, 759–784.
10. Hall, B.H.; Mairesse, J.; Mohnen, P. Measuring the returns to R&D. In *Handbook of the Economics of Innovation*; Elsevier: Amsterdam, The Netherlands, 2010; Volume 3, pp. 1033–1082.
11. Varian, H.R. Beyond big data. *Bus. Econ.* **2014**, *49*, 27–31. [CrossRef]
12. Mullainathan, S.; Spiess, J. Machine learning: An applied econometric approach. *J. Econ. Perspect.* **2017**, *31*, 87–106. [CrossRef]
13. Coulombe, P.G.; Leroux, M.; Stevanovic, D.; Surprenant, S. How is machine learning useful for macroeconomic forecasting? *J. Appl. Econ.* **2022**, *37*, 920–964. [CrossRef]
14. Athey, S.; Imbens, G.W. Machine learning methods that economists should know about. *Annu. Rev. Econ.* **2019**, *11*, 685–725. [CrossRef]
15. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD Conference*; ACM: New York, NY, USA, 2016; Volume 12, pp. 785–794.
16. Jorgenson, D.W.; Ho, M.S.; Stiroh, K.J. *Information Technology and the American Growth Resurgence*; MIT Press: Cambridge, MA, USA, 2005.
17. Shen, Y.; Hu, W.; Hueng, C.J. Digital financial inclusion and economic growth: A cross-country study. *Procedia Comput. Sci.* **2021**, *187*, 218–223. [CrossRef]
18. Bresnahan, T.F.; Trajtenberg, M. General purpose technologies 'Engines of growth'? *J. Econ.* **1995**, *65*, 83–108. [CrossRef]
19. Griliches, Z. *R&D and Productivity: The Unfinished Business*; University of Chicago Press: Chicago, IL, USA, 1998; pp. 269–283.
20. Lundvall, B.Å. Industry and innovation, National innovation systems—Analytical concept and development tool. *Ind. Innov.* **2007**, *14*, 95–119. [CrossRef]
21. Yu, P. Diffusion of innovation theory. In *Implementation Science*; Routledge: Oxfordshire, UK, 2022; pp. 59–61.
22. Katz, M.L.; Shapiro, C. On the licensing of innovations. *RAND J. Econ.* **1985**, *16*, 504–520. [CrossRef]
23. Koutroumpis, P.; Koutroumpis, S.D. The economic impact of broadband access for small firms. *World Econ.* **2024**, *47*, 1642–1681. [CrossRef]
24. Gawande, K.; Bandyopadhyay, U. Is protection for sale? Evidence on the Grossman-Helpman theory of endogenous protection. *Rev. Econ. Stat.* **2000**, *82*, 139–152. [CrossRef]
25. Acemoglu, D.; Ozdaglar, A.; Tahbaz-Salehi, A. The network origins of large economic downturns (No. w19230). *Natl. Bur. Econ. Res.* **2013**, 19230. [CrossRef]
26. Winter, E. The shapley value. In *Handbook of Game Theory with Economic Applications*; Elsevier: Amsterdam, The Netherlands, 2002; Volume 3, pp. 2025–2054.
27. Aghion, P.; Howitt, P. A model of growth through creative destruction. *J. Political Econ.* **1990**, *100*, 223–251.
28. Barro, R.J.; Sala-i-Martin, X. Convergence. *J. Political Econ.* **1992**, *100*, 223–251. [CrossRef]
29. Bergmeir, C.; Benítez, J.M. On the use of cross-validation for time series predictor evaluation. *Inf. Sci.* **2012**, *191*, 192–213. [CrossRef]
30. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
31. WorldBank. World Development Indicators. The World Bank Group. 2024. Available online: https://databank.worldbank.org/source/world-development-indicators (accessed on 6 February 2026).
32. Vu, K.M. ICT as a source of economic growth in the information age: Empirical evidence from the 1996–2005 period. *Telecommun. Policy* **2011**, *35*, 357–372. [CrossRef]
33. Solow, R.M. A contribution to the theory of economic growth. *Q. J. Econ.* **1956**, *70*, 65–94. [CrossRef]
34. Shapley, L.S. *A Value for n-Person Games*; Princeton University Press: Princeton, NJ, USA, 1953.
35. Efron, B.; Tibshirani, R.J. *An Introduction to the Bootstrap New York*; Chapman and Hall: New York, NY, USA, 1993.
36. European Commission. *2030 Digital Compass: The European Way for the Digital Decade; COM(2021) 118 final*; EU Commission: Brussels, Belgium, 2021.
37. Jorgenson, D.W.; Stiroh, K.J. Raising the speed limit: U.S. economic growth in the information age. *Brook. Pap. Econ. Act.* **2000**, *2000*, 125–235. [CrossRef]

38. Friedman, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Assoc.* **1937**, *32*, 675–701. [CrossRef]
39. Griffith, R.; Redding, S.; Van Reenen, J. Mapping the two faces of R&D: Productivity growth in a panel of OECD industries. *Rev. Econ. Stat.* **2004**, *86*, 883–895. [CrossRef]
40. Gruber, H.; Koutroumpis, P. Mobile telecommunications and the impact on economic development. *Econ. Policy* **2011**, *26*, 387–426. [CrossRef]
41. Hair, J.F.; Black, W.C.; Babin, B.J.; Anderson, R.E. *Multivariate Data Analysis*, 8th ed.; Cengage: Boston, MA, USA, 2019.
42. Hoerl, A.E.; Kennard, R.W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **1970**, *12*, 55–67. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.