CONTENTS 1/2025

The Role of Sentiment Analysis and Central Bank Interest Rate Decisions in Forecasting Inflation: A Bayesian Non-Parametric Approach for Czech Republic and Romania

Mihaela Simionescu (mihaela_mb1@yahoo.com, mihaela.simionescu@unibuc.ro)
Faculty of Business and Administration, University of Bucharest, Romania
Academy of Romanian Scientists, Bucharest, Romania
Institute for Economic Forecasting, Romanian Academy, Bucharest, Romania

Alexandru-Sabin Nicula (sabin.nicula@ubbcluj.ro)
Academy of Romanian Scientists, Bucharest, Romania
Romanian Academy, National Institute for Economic Research "Costin C. Kiriţescu", Mountain Economy Center, Vatra Dornei, Romania
Centre for Research on Settlements and Urbanism, Faculty of Geography, Babeș-Bolyai University, Cluj-Napoca, Romania

ABSTRACT

Most Eastern European countries experienced high inflation because of the war in Ukraine, which makes more difficult to get accurate inflation forecasts necessary for policymakers, central bank and business environment. The main aim of this paper is to provide accurate forecasts for inflation rate in Romania and Czech Republic by using non-parametric Bayesian models and generalized regression neural networks (GRNN) that include sentiment index determined using central banks official reports. The forecasts based on Bayesian linear regression models outperformed the ones based on Bayesian linear regression model with LASSO prior, Bayesian linear regression model with stochastic search variable selection (SSVS) and GRNN on short-term horizon 2023: Q1- 2023: Q4. The approach based on difference-in-differences estimators suggested that the strategy of the Czech National Bank (ČNB) based on proactive increased interest rates in 2021-2022 to control inflation was effective and it reduced expected inflation, but had no significant impact on unexpected component. The expected inflation in Czechia decreased by 0.972 percentage points relative to the case when interest rate would have not increase sharply. These empirical findings bring

contribution for providing better short-term inflation forecasts by taking into account experts opinion on future evolution of inflation and their interventions to reduce the phenomenon.

Keywords*: inflation; non-parametric Bayesian models; generalized regression neural networks; interest rate*

1. INTRODUCTION

The war in Ukraine significantly impacted inflation in the EU, causing it to rise to record levels in 2022. Several EU countries experienced hyperinflation in 2022, with the Baltic states and some Eastern European countries often facing the highest levels. In this context dominated by uncertainty, the task to provide accurate inflation forecasts becomes more difficult than ever.

When it comes to forecasts, inflation rate is a key index of economic activity and a practical problem, being of utmost importance to predict it in different conditions. But inflation prediction is not an easy and simple task.

Nowadays there are at hand a wide range of methods than one can appeal to in order to increase the accuracy of predictions, ranging from the use of different econometric models to machine learning applications (ML). The scientific discourse from recent years was focused on the use of ML in the field of econometry (Athey and Imbens, 2019; Athey, 2018; Mullainathan and Spiess, 2017) and to compare its performance to traditional econometric models (Zhang et al., 2018). In this respect Pérez-Pons et al. (2021) steered to find experimental evidence to demonstrate the superiority of ML algorithms to conventional econometric models. The findings of their study, by appealing to systematic literature mapping, showed that in the majority of cases the econometric models are surpassed by ML, but in others, the finest performance was accomplished by compiling the both, leading to the development and implementation of hybrid models. It is important to note that even if at the moment the performance of ML in prediction is always superior, there is no guarantee that in an emerging field that is continuously changing (Standford Index, 2018; 2019) things will remain the same compared to traditional econometric models (Pérez-Pons et al., 2021).

Accurate forecasts of inflation are crucial for various economic players, especially during periods of significant inflation shifts. First, inflation forecasts directly impact monetary and other policy choices, influencing the overall economic landscape. Second, central banks rely on accurate forecasts to design timely and appropriate responses to inflation changes, ultimately guiding inflation back to target levels. Third, businesses utilize inflation forecasts to anticipate market changes and make informed decisions (Joseph et al., 2022).

Considering this recent economic context and the necessity to surpass it and provide accurate forecasts, the main objective of this paper is to propose accurate inflation forecasts for Romania and Czech Republic on the horizon 2023: Q1- 2023: Q4, two Eastern countries that were deeply affected by inflation. Only a short forecast horizon is considered, because the paper tackles a specific period dominated by the war in Ukraine that generated high inflation. Moreover, the ability of various methods to generate accurate forecasts is checked only in the particular context when the year before the forecast horizon is characterized by hyperinflation and the central banks implemented in that year strategies based on increasing interest rate.

The proposal of accurate forecasts depends on many factors, forecasting methods and central banks' interventions being crucial. Therefore, this paper is based on non-parametric models that are more suitable for capturing non-linearities. In addition, the efficiency of banks strategy to control inflation by rising interest rate is deeply analyzed in a comparative form using difference-in-differences estimators, including the approach that addresses the impact on inflation components (expected/unexpected inflation). From this point of view, the main research question tackles the necessity to get accurate inflation forecasts by selecting best prediction method and best strategy to control for inflation.

Only few studies have employed non-parametric models to predict inflation by assuming linear or non-linear connection. For example, the nonlinear relationship between inflation and various predictors might be captured by using Gaussian and Dirichlet processes (Clark et al., 2022). A sticky infinite hidden Markov model was employed by Jochmann (2014) to explain the US inflation rate. From this point view, the novelty of this study relies on the contribution in empirical forecasting by implementing suitable non-parametric approaches. First, the paper proposes non-parametric Bayesian models to forecast inflation (Bayesian linear regression model, Bayesian linear regression model with LASSO prior and Bayesian linear regression model with stochastic search variable selection (SSVS)). Moreover, generalized regression neural networks (GRNNs) are also employed to make forecasts and the evaluation of forecast accuracy allow us to determine the best predictions.

As a novelty for the inflation empirical forecasting, these methods capture the opinions of experts in central bank related to future evolution of inflation by introducing sentiment index as explanatory variable in the models. Sentiment analysis is becoming increasingly recognized as a valuable tool for forecasting inflation (Simionescu and Nicula, 2024). This approach has been previously employed to predict inflation rate in Romania using machine learning techniques (Simionescu, 2022). As a novelty for literature,

IntelliDockers software is employed to compute sentiment indexes based on official reports of Czech and Romanian central banks.

The forecast accuracy also depends on the anticipated future effects of central bank strategy to reduce inflation. This paper makes a step forward in the specific context of Romanian and Czech economies in the period of hyperinflation generated by the war in Ukraine and internal factors. The strategy based on raising interest rates was differently applied by the two states: the Czech bank made a sudden increase of interest rate, while the Romanian bank adopted a gradual increase in interest rate. It is clear that inflation in Czech Republic reduced more, while Romania also faced this issue in an acute form at the beginning of 2024. Therefore, a difference-in-differences estimator approach is applied to assess the contribution of the Czech strategy to inflation control compared to the case when the Romanian strategy would have been applied. Moreover, a deeper analysis makes us to check which component of inflation was targeted by this strategy (expected /unexpected inflation). This additional analysis allows us to understand why the forecasts provided by the Czech national bank were more accurate. The results confirmed the capacity of Bayesian linear regression model using sentiment index and interest rate as predictors to provide the most accurate forecasts for both countries. The predictions made for Czechia were more accurate, because its strategy to control for inflation was more effective. However, the strategy reduced the expected inflation, but had no significant impact on unexpected inflation.

All in all, the contributions of this paper consist of the use on various non-parametric approaches that are more suitable to model inflation, the use of sentiment analysis to capture experts opinions on future evolution of inflation, the evaluation of forecast accuracy and insights in the mechanisms that could improve predictions by assessing the impact of strategies implemented by Czech and Romanian central banks to control for overall inflation, expected and unexpected inflation using difference-in-differences estimators. These contributions are a step forward in designing the best policy proposals to control for inflation in the future.

All these contributions are integrated in the paper that follows a canonical structure. After this introduction, some insights in literature are provided with main focus on non-parametric methods. The next sections of the article describe methodology, data, results with discussion and propose pertinent conclusions.

2. LITERATURE REVIEW

Numerous methods have been used in literature to forecast inflation, the most utilized being: Phillips curve-based models (Dotsey et al., 2018), univariate unobserved component models (Stock et al., 2016), DSGE models (Cardani et al., 2022), aggregation of forecasts (Hubrich, n.d.), Bayesian VARs (Cimadomo et al., 2022), dimensionality reduction (Kim and Swanson, 2008). Most of these methods assume a linear relationship between inflation and other variables even if non-linear or non-parametric one would be more suitable in same cases. While most inflation forecasting research has focused on linear models, a growing body of work explores non-linear and non-parametric approaches.

Some studies suggest that the relationship between economic activity and inflation might be non-linear, with stronger effects emerging during periods of high economic growth (Babb & Detmeister, 2017). The results provided by linear and non-linear models are different at least for a couple of years.

Recent research highlights the potential of machine learning techniques, particularly random forests (Breiman, 2001), for macroeconomic forecasting, including inflation (Medeiros et al., 2021) These methods have shown promising results even during crisis periods. Medeiros et al. (2021) investigated the potential of machine learning (ML) methods for U.S. inflation forecasting. Despite skepticism in previous research, the authors demonstrate that ML models utilizing a large number of covariates outperform traditional benchmarks. Among these, the random forest model stands out for its superior accuracy. This success is attributed to its specific variable selection approach and the potential for non-linear relationships between past macroeconomic variables and inflation. Another study investigated the effectiveness of various machine learning techniques for forecasting inflation in Nigeria. The analysis compared ridge regression and artificial neural networks (ANNs) with other methods like LASSO, elastic net, and PLS. The results revealed that ridge regression and ANNs significantly outperformed the other models in terms of forecasting accuracy (Medeiros et al., 2021). Furthermore, the study identified key drivers of inflation in Nigeria, including food inflation, core inflation, prime lending rate, maximum lending rate, and the inter-bank rate.

Ülke et al. (2018) approached and compared in their research ML models and time series models (univariate and multi variate) in order to forecast inflation in the USA in 16 different horizons. The outcomes showed that ML models are more exact in seven conditions compared to time series models which are more accurate in nine. If compared, multivariate models are better in proving forecasts than univariate time series models (Stock and Watson, 1999).

ML models (e.g. Artificial Neural Network (ANN); Support Vector Machines (SVM); k-nearest neighbors (k-NN)) seem to be infrequently used for inflation forecasting (Ülke et al., 2018), but work better for other macroeconomic variables (exchange rates, income, stock market return volatility) according to Mizrach (1992), Rodriguez et al. (1999), Guegan and Rakotomarolahy (2010), Chen et al. (2010). Based on the findings of Ülke et al. (2018) there is not an unanimously accepted best model to forecast inflation because the way of approach depends on the research objective.

Influenced by previous works that appealed to ML methods for inflation forecasting (Chakraborty and Joseph, 2017; Garcia et al., 2017) the works of Aras and Lisboa (2022), Pavlov (2020), Baybuza (2018), Choudhary and Haider (2012), Marcek and Marcek (2006) aim to prove theirs' viability compared to conventional methods and to offer valuable solutions. The results are summarized as follows: ML models provide a different way by considering non-linear relationships, ML models are responsible for more accurate predictions, ML models represent a promising forecast inflation instrument, especially for short- and medium-term. If compared to conventional benchmarks, ML methods forecast in the same manner, no worse, no better. Non-linear ML models are indicated to perform weakly in the forecasting process. Baybuza (2018) employed neural networks and support vector machines to forecast Russian inflation. Additionally, Shapley value decomposition was employed to provide economic insights into the neural network's predictions. The performance of these machine learning models was then benchmarked against more traditional approaches - autoregression and regularized linear regression (ridge regression). The empirical results suggest that both machine learning models deliver forecasts comparable to, if not better than, the conventional benchmarks. Furthermore, Shapley decomposition proves to be a valuable framework for interpreting the neural network's forecasts in an economically meaningful way. To illustrate ML's applications in central banking, three case studies were presented by Chakraborty and Joseph (2017). The first tackles modeling alert detection in financial institution balance sheets for banking supervision. The second focuses on projecting UK CPI inflation on a two-year horizon, introducing a basic training-testing framework for time series analysis. Finally, the third case study investigates funding patterns of technology startups to identify potentially disruptive innovators in financial technology. While machine learning models often outperform traditional approaches in prediction tasks, open questions remain regarding their causal inference capabilities.

In the contemporary era of forecasting there is a growing interest in studying the Artificial Neural Networks (ANN). Haider and Hanif (2009)

compared the ANN ML methodology with conventional univariate time series forecasting models (AR (1) and ARIMA based models) and concluded that the ANN forecast performance is better, facts sustained also by Estiko and Wahyuddin (2019), Yusif et al. (2015), Akhter (2013). Similarly, Takur et al. (2016) proposed an ANN model to predict inflation and came to the conclusion that its accuracy is decent. When it comes to developing countries with volatile inflation, Hanif and Malik (2015), the policy makers need to adopt those prediction models that are best suited for a given state. Other researches focused on the efficacy of SARIMA models in forecasting inflation (Lidiema, 2017; Gikungu et al., 2015; Saz, 2011) with good outcomes when it comes to prediction accuracy. Other well-fitted stable prediction models, like symmetric nonlinear ARDL and asymmetric NADRL were studied in the works of DeLuna Jr. et al. (2021) and Pahlavani and Rahimi (2009). Estiko and Wahyuddin (2019) utilized monthly year-on-year inflation data for Indonesia from December 2006 to December 2018, obtained from Bank Indonesia (BI) and the Indonesian Central Bureau of Statistics (CBS). The results suggest that the NN model outperforms ARIMA in forecasting inflation for all three data series. Additionally, the empirical findings indicate that the influence of short-term lagged inflation (past inflation data) on future inflation appears to diminish in the more recent data sets compared to the earlier period. Yusif et al. (2015) showed that ANN models achieved better accuracy compared to the traditional econometric models (AR models, VAR models). This suggests that, based on this evaluation metric, forecasts generated by ANNs exhibit greater accuracy for Ghanaian inflation.

Beyond classical methods, Bayesian techniques offer alternative avenues. Jochmann (2015) proposes an infinite hidden Markov model to analyze US inflation dynamics, identifying a secular decline in volatility and distinct inflation regimes. Clark et al. (2021) utilize Bayesian additive regression tree (BART) models for inflation forecasting, demonstrating improved performance during volatile periods like the COVID-19 pandemic. However, a potential limitation of BART and other non-parametric techniques (e.g., Gaussian processes) lies in their assumption of Gaussian shocks. If empirical evidence suggests non-Gaussian features like heavy tails in the innovations, the model's flexibility might lead to capturing these features within the conditional mean, potentially misrepresenting the true non-linear relationships between inflation dynamics and its predictors.

The paper of (Tierney, 2019) expands the field of nonparametric forecasting by introducing three novel local nonparametric methods. These methods are specifically designed for the nonparametric exclusion-from-core inflation persistence model and can leverage revised real-time personal

consumption expenditure (PCE) and core PCE data for 62 vintages. Local nonparametric forecasting offers several advantages: a) data flexibility: It allows for a more nuanced analysis by incorporating low inflation periods into other low inflation segments and vice versa. b) real-time outlier detection: When applied to real-time data, these models can assist policymakers in identifying potential outliers, abnormal economic events, or data inconsistencies like volatility shifts. The most efficient method among the three is the third model. It leverages the flexibility of non-parametric approach by making forecasts conditional on the predicted value, enabling counterfactual analysis.

A particular attention was given to the debate between econometric models and neural networks, which are also the subject of this paper. The study of (Moshiri & Cameron, 2000) investigates the effectiveness of Back-Propagation Artificial Neural Network (BPN) models for inflation forecasting, comparing them with traditional econometric approaches. The main advantages of BPN models are their ability to handle complex relationships and their freedom from linearity assumptions commonly used in traditional methods. The study compares BPN models with four econometric models: a structural reduced-form model, an ARIMA model, a vector autoregressive model, and a Bayesian vector autoregression model. Each econometric model is paired with a hybrid BPN model using the same set of variables. Dynamic forecasts are then generated for three horizons: one, three, and twelve months ahead. The quality of forecasts is evaluated using root mean squared errors and mean absolute errors. The results demonstrate that hybrid BPN models perform as well as all traditional econometric models, and even outperform them in certain cases. Another study compared the effectiveness of Artificial Neural Networks (ANNs) and the random walk model for forecasting inflation in 28 OECD countries (Choudhary, n.d.). For short-term predictions, ANNs were more accurate in 45% of the countries, while the random walk model was better in 23%. Additionally, combining multiple ANN models proved to be a promising approach for inflation forecasting.

Other papers investigated if a specific type of neural network, the Long Short-Term Memory (LSTM) model, could improve inflation forecasting compared to other methods. Focusing on monthly US CPI inflation, (Almosova & Andresen, 2023) found that LSTM slightly outperformed traditional models like autoregressive (AR), basic neural networks (NN), and Markov-switching models. However, its performance was comparable to the seasonal autoregressive model (SARIMA), suggesting similar accuracy. Additionally, they analyzed the sensitivity of the model to different settings and explored how the network learns through a novel technique.

For a deep knowledge of the inflation mechanisms other studies focused on the individual components within the Consumer Price Index

(CPI) rather than just the overall headline inflation. For example, a novel approach to predicting inflation was introduced (Hierarchical Recurrent Neural Network (HRNN) model), which leverages information from higher levels of the CPI hierarchy to improve predictions at more volatile lower levels (Barkan et al., 2023). Evaluated using a large dataset from the US CPI, the HRNN model significantly outperforms existing methods for predicting various CPI components. This opens up new possibilities for policymakers and market players to gain insights into price changes at a more granular level. Overall, the general conclusion, regarding which models hold the supremacy in forecasting, the Machine learning or conventional econometric models is that neither one of them can not 100% predict the inflation rate or has the best accuracy, because there are a lot of unknown variables that can inevitably lead to the ups and downs of inflation rate. As Takur et al. (2016) highlighted, it is mandatory to take into consideration more factors specific to each country's economy.

Despite the multitude of neural networks used to predict inflation, none of the studies considered generalized regression neural networks (GRNNs), which is a non-parametric approach. The method was previously employed to forecast exchange rate focusing on the British pound, Canadian dollar, and Japanese yen, and this method outperformed other neural network and econometric methods in forecasting monthly exchange rates (Leung et al., 2000). This suggests that GRNNs have the potential to be valuable tools for financial forecasting. Given this gap in literature, this article provides predictions for inflation rate in Romania and Czech Republic using GRNNs and Bayesian non-parametric models to determine the most accurate forecasts. Moreover, the analysis goes deeper to understand better the efficiency of mechanisms used to reduce inflation. Since these countries implemented different strategies to control for inflation it is necessary to compare the efficiency of one strategy compared to other and to check to what component of inflation it addresses (expected and/or unexpected inflation).

3. METHODOLOGY AND DATA

First, inflation forecasts are made using non-parametric Bayesian models: Bayesian linear regression model, Bayesian linear regression model with LASSO prior and Bayesian linear regression model with stochastic search variable selection (SSVS). The prior values used in this study are selected after starting from initial values indicated by Karabatsos (2018) and trying more combinations as to retain the model with the highest coefficient of determination (R-square). Second, inflation forecasts are also made using

generalized regression neural networks (GRNN) that is implemented in R using ***tsfgrnn*** package. The two types of forecasting methods are described in this section.

Let us start from the general form of a Bayesian linear regression:

$y_t = \sum_{k=1}^{p} \beta_k x_{tk} + \varepsilon_t$, k=1,2,…,p

β_k- parameter (coefficient);

ε_t- error (disturbance)

σ^2- error (disturbance) variance

Bayesian linear regression model

$$y_t|x_t \sim f(y|x_t), t = 1,2,\dots,n$$
$$f(y|x) = n(y|x^T\beta, \sigma^2)$$
$$\beta_0 \propto 1$$
$$\beta_k|\sigma^2 \sim N(0, \sigma^2 v_\beta), \text{ k=1,2,…,p}$$
$$\sigma^2 \sim IG(a_0/2, a_0/2)$$

In this application, the prior parameters are given by v_β=1000 and a_0=0.001.

Bayesian Linear regression model with LASSO prior

$$y_t|x_t \sim f(y|x_t), t = 1,2,\dots,n$$
$$f(y|x) = n(y|x^T\beta, \sigma^2)$$
$$\beta_0 \propto 1$$
$$(\beta_k)_{k=1}^{p}|\lambda, \sigma^2 \sim \prod_{k=1}^{p} Laplace(\beta_k|0, \sigma/\lambda)$$
$$\lambda \sim Gamma(\lambda|r, \delta)$$
$$\sigma^2 \sim IG(\sigma^2|0,0)$$

In this application, the prior parameters take the values r=1 and δ =2.

Bayesian Linear regression model with stochastic search variable selection (SSVS)

$$y_t|x_t \sim f(y|x_t), t = 1,2,\dots,n$$
$$f(y|x) = n(y|x^T\beta, \sigma^2)$$
$$\beta_0 \propto 1$$
$$\beta_k|\sigma^2, \gamma_k \sim N(0, \sigma^2\{v_1\gamma_k + v_0(1-\gamma_k)\})$$
$$\gamma_k \sim Ber(w_\gamma), k = 1,\dots,p$$
$$\sigma^2 \sim IG(a_0/2, \lambda a_0/2)$$

In this application, the prior parameters are $v_1 = 1,\ v_0 = 0.1,$ $w_\gamma = 0.5, a_0 = 3, \lambda = 14.$

Generalized regression neural networks (GRNN)

A GRNN is a radial basis function (RBF) network with only one fast pass learning. It includes a hidden layer and neurons from RBF. The number of neurons in hidden layer equals the number of training examples. The neuron's center is represented by the corresponding training example. The neuron's output is a measure of the distance between training example and input vector. The neuron is based on a multivariate Gaussian function, when *x* is the input vector, x_t is neuron's center and σ is the smoothing parameter:

$$G(x, x_t) = e^{-\frac{\|x-x_t\|^2}{2\sigma^2}}$$

The smoothing parameter shows the number of targets that are relevant for the weighted mean. When the weights are similar, the result has a value around the average of training targets and this happens when smoothing parameter is large. In case of a small value of this parameter, significant weights are assigned only to those training targets that are close to input vector.

Considering *n* training patterns that forms a training set given by the vector $\{x_1, x_2, \dots, x_n\}$ and n targets given by normally scalars $\{y_1, y_2, \dots, y_n\}$, two steps should be considered to compute the output knowing the input pattern *x*.

Step 1: the computation of weights that are a measure of closeness of *x* to training patterns;

$$\omega_t = \frac{e^{-\frac{\|x-x_t\|^2}{2\sigma^2}}}{\sum_{j=1}^{n} e^{-\frac{\|x-x_j\|^2}{2\sigma^2}}}$$

Properties of weights: they decrease with distance to the training pattern; their sum is one and it represents the contribution of any training pattern to the final output.

Step 2: the calculation of GRNN output layer as a weighted mean of the training targets.

$$\hat{y} = \sum_{t=1}^{n} \omega_t y_t$$

Sentiment index and monetary policy interest rate are used as exogenous variables to explain inflation rate in non-parametric Bayesian models. The calculation of sentiment index is based on natural language

processing (NLP), the IntelliDockers software being used. This program employs Recurrent Neural Networks, a specific algorithm for machine learning. Quarterly inflation rate forecasts and actual values for Romania are provided by the *Inflation Reports* published by the National Bank of Romania.

The text in the section *Inflation outlook* of the quarterly reports is introduced in IntelliDockers to calculate sentiment index for the period 2006: Q1-2022: Q4. The proposed forecasting methods are used to predict inflation rate on the horizon 2023: Q1-2023: Q4.

The forecasts related to inflation rate in Czech Republic are published in Monetary Policy Reports for each season (winter, spring, summer and autumn) since 2021. Given the short quarterly inflation forecasts in these reports, the monthly inflation forecasts were taken from monthly reports called Global Economic Outlook. The quarterly inflation rate is computed as an average of corresponding monthly inflation rates. The quarterly sentiment index is computed also as average of sentiment indexes for the corresponding months using the reports in English and text available in the sections *Inflation* or *Focus*, depending on the structure of the reports. The monthly sentiment indexes are calculated using IntelliDockers and the text in Monetary Policy Reports for the period January: 2011- December: 2022. The actual inflation rate with quarterly frequency in Czechia is provided by the Federal Reserve Bank of St. Louis and the monetary policy interest rate is provided by Czech National Bank. The descriptive statistics for inflation reported in Table 1 suggest that inflation rate tends to be lower in Czech Republic compared to Romania. The highest inflation rate was observed in Romania at the end of the analysed period (2022: Q4) because of the post-pandemic shocks and because of the war in Ukraine. On the other hand, Czechia experienced the highest inflation in the first quarter of 2022. The war in Ukraine disrupted the global energy market, causing prices to skyrocket. Since Czechia relies heavily on energy imports, it felt this pinch more acutely. Beyond the global issue, some argue that internal factors in the Czech Republic, like domestic wage and markup dynamics, also played a role in amplifying inflation compared to other countries in the region. The normal distribution hypothesis is not supported for any of the time series.

Table 1. Descriptive statistics on inflation rate

Statistics	Romania	Czechia
Mean	4.91%	0.93%
Maximum	16.4%	6.27%
Minimum	-1.70%	-0.49%
Standard deviation	3.78%	1.51%
Shapiro-Wilk stat. (p-value in brackets)	5.391 (<0.01)	8.89 (<0.01)

Source: own calculations

Besides inflation forecasts, this paper also assesses the efficiency of Czech central bank strategy to reduce inflation compared to the one implemented by Romanian central bank by implementing difference-in-difference estimator. The analysis is also conducted on expected and unexpected inflation. Since there is not a clear methodology to calculate these components of inflation, an autoregressive integrated moving average is constructed (ARMA). The estimated values based on ARMA model represents the expected inflation, while the residuals represent the unexpected values of inflation (Kim & Lin, 2023).

Given these insights on methods and data, the next section presents empirical evidence related to regression models that explain the inflation rate in both countries. The main hypothesis that is checked is related to the capacity on non-parametric Bayesian regression models to provide more accurate forecast for 2023 compared to GRNN.

4. RESULTS AND DISCUSSION

The time series were seasonally adjusted and the presence of unit root is checked for each series using unit root with break test. The results in Table 2 supports the hypothesis that all the time series are stationary in level at 5% significance level.

Table 2. The results of unit root with break test

Country	Variable	Data series	Stat.	p-value	Decision
Romania	Inflation rate	Series in the second difference	-13.3	<0.01	no unit root
		Series in the first difference	-8.3	<0.01	no unit root
		Series in level	-4.8	0.02	no unit root at 5% significance level
	Sentiment index	Series in the second difference	-20.6	<0.01	no unit root
		Series in the first difference	-15.1	<0.01	no unit root
		Series in level	-8.4	<0.01	no unit root
	Interest rate	Series in the second difference	-12.5	<0.01	no unit root
		Series in the first difference	-10.3	<0.01	no unit root
		Series in level	-9.8	<0.01	no unit root
Czech Republic	Inflation rate	Series in the second difference	-10.3	<0.01	no unit root
		Series in the first difference	-9.2	<0.01	no unit root
		Series in level	-8.6	<0.01	no unit root
	Sentiment index	Series in the second difference	-20.4	<0.01	no unit root
		Series in the first difference	-16.3	<0.01	no unit root
		Series in level	-9.9	<0.01	no unit root
	Interest rate	Series in the second difference	-11.8	<0.01	no unit root
		Series in the first difference	-9.7	<0.01	no unit root
		Series in level	-8.9	<0.01	no unit root

The empirical findings in Table 3 suggest a negative impact of sentiment index and interest rate on inflation. This means that experts from national banks optimistic expectations related to the evolution of evolution translated into less inflation. The rise in interest rate contributed to inflation reduction in both countries.

Table 3. The results of estimations based on nonparametric Bayesian regression models 2023: Q1-2023: Q4

Country	Parameters	Bayesian linear regression model	Bayesian linear regression model with LASSO prior	Bayesian linear regression model with SSVS
Romania	Prior parameters of model	$v_\beta = 1000$ $a_0 = 0.001$	r=1, $\delta = 2$	$v_1 = 1, v_0 = 0.1,$ $w_\gamma = 0.5, a_0 = 3,$ $\lambda = 14$
	Parameters	Mean (standard deviation in brackets)		
	β_0	13.221 (2.793)	11.541 (2.843)	13.156 (2.256)
	$\beta_{sentiment}$	-12.006 (3.967)	-8.970 (4.165)	-11.914 (3.427)
	$\beta_{interest}$	-0.155 (0.243)	-0.179 (0.241)	-0.153 (0.192)
	λ	-	0.698 (0.297)	-
	$\lambda_{sentiment}$	-	-	0.986 (0.119)
	$\lambda_{interest}$	-	-	0.028 (0.164)
	σ^2	12.714 (2.25)	13.048 (2.303)	8.226 (1.098)
Czech Republic	β_0	5.663 (1.335)	4.221 (1.033)	5.558 (1.228)
	$\beta_{sentiment}$	-1.376 (0.456)	-0.995 (0.032)	-1.274 (0.337)
	$\beta_{interest}$	-1.003 (0.342)	-0.788 (0.223)	-0.995 (0.257)
	λ	-	0.557 (0.225)	-
	$\lambda_{sentiment}$	-	-	0.448 (0.206)
	$\lambda_{interest}$	-	-	0.142 (0.044)
	σ^2	5.667 (1.334)	4.556 (2.072)	3.221 (0.994)

Source: own calculations.

Two strategies are employed to make multi step ahead forecasts using GRNN:

- Multiple Input Multiple Output strategy known as MIMO based on training targets vectors with successive values of the chronological

predicted series, the dimension of a vector being given by the forecast horizon;

- Recursive strategy.

The forecast accuracy assessment reported in Table 4 shows that inflation predictions for Romania based on non-parametric models are better than those based on GRNN models, according to MAPE and SMAPE. DM test also confirms the superiority of non-parametric models compared to GRNN model with recursive forecasting strategy. On the other hand, all the accuracy indicators support the hypothesis that non-parametric Bayesian models that include sentiment index outperform GRNN approach for the horizon 2023: Q1-2023: Q4. Bayesian linear regression model determined the most accurate forecast for inflation rate in Czechia. The ability of non-parametric Bayesian models to provide accurate inflation forecasts is also supported by the study of Clark et al. (2022).

Table 4. Inflation forecast accuracy for Romania and Czech Republic

Country	Forecasting method	RMSE	MAE	MAPE	SMAPE	DM test (comparisons with forecasts based on GRNN model- recursive strategy)
Romania	Bayesian linear regression model	4.918197	4.223005	0.3813077	0.4921606	14.30*
	Bayesian linear regression model with LASSO prior	5.109041	4.391575	0.3969654	0.5168944	11.34*
	Bayesian linear regression model with SSVS	4.922759	4.226795	0.3816435	0.492689	14.28*
	GRNN model-MIMO strategy	3.162278	3.000000	4.236205	4.338796	1.18
	GRNN model-recursive strategy	2.236068	2.000000	2.807458	2.857449	-
Czech Republic	Bayesian linear regression model	2.343786	1.894773	0.296345	0.3227	7.89*
	Bayesian linear regression model with LASSO prior	2.998453	2.673492	0.359827	0.383385	8.56*
	Bayesian linear regression model with SSVS	2.445889	2.036992	0.317559	0.364593	7.97*
	GRNN model-MIMO strategy	4.783483	4.674593	5.045636	5.143844	1.87
	GRNN model-recursive strategy	3.349823	3.284595	3.884529	4.036738	-

*Source: own calculation. * means p-value less than 0.01.*

In the case of Romania, sigma (smoothing parameter) is 0.2196321 for recursive strategy and 0.1936133 for MIMO, while for Czech Republic the values are 0.334223, and 0.2983467 respectively.

Autoregressive models of order one (AR(1)) were built for both countries to explain inflation rate in order to calculate expected and unexpected inflation. The estimated values based on AR(1) represent the expected inflation series, while the residuals reflect the unexpected inflation values. The results of estimations are reported in Table 5.

Table 5. Autoregressive models used to calculate expected and unexpected inflation in Czechia and Romania (2011: Q1-2024: Q2)

	Romania				Czech Republic			
	Coefficient	Std. error	DW stat.	White test	Coefficient	Std. error	DW stat.	White test
Constant	4.235**	1.975	1.93	7.49 (0.44)	0.886***	0.296	2.162	10.05 (0.35)
AR(1)	0.832***	0.078			0.309**	0.137		

*Source: own calculations. p-values in brackets, ** means significance at 5% and *** shows significance at 1%.*

A key monetary policy tool used to combat inflation is raising interest rates. The Czech National Bank (ČNB) proactively increased interest rates in 2021-2022 in anticipation of rising inflation. This helped curb inflation to some extent. The National Bank of Romania (NBR) adopted a more wait-and-see approach**,** keeping interest rates lower for a longer period compared to the Czech Republic. This decision aimed to support economic recovery after the pandemic. By comparing the two approaches, one may observe the potential effect of interest rate hikes on inflation. Though inflation did rise due to global factors, the Czech Republic's proactive approach with interest rates might have helped moderate the increase compared to what could have been without them. Romania's decision to delay rate hikes might have contributed to a potentially steeper rise in inflation initially. However, they did raise rates later in 2022 to address the situation.

It seems that ČNB strategy is more efficient in reducing inflation. In the last quarter of 2021 and in the first two quarters of 2022, high increase in interest rate was observed. To check this hypothesis that the ČNB strategy was more suitable to reduce inflation compared to Romania, an approach based on difference-in-difference estimator is employed for the period 2011: Q1-2024: Q2 using panel data.

Parallel-trends test (pretreatment time period) is applied under the null hypothesis that the linear trends are parallel. Granger causality test supposes no effect in anticipation of treatment under null hypothesis.

Prior to the implementation of ČNB strategy, both countries followed a parallel path as Table 6 suggests. The assumption of no behavior change prior to treatment is not rejected at 5% significance level. These findings support the validity of the ATET (the average treatment effect on the treated) estimate.

Table 6. The results of approach based on difference-in-difference estimator

ATET Strategy (1 vs 0)	Coefficient	Robust std. error	p-value	Parallel-trends test (stat. and p-value)	Granger causality test (stat. and p-value)
Inflation	-0.302	0.039	<0.01	0.53 (0.47)	0.25 (0.81)
Expected inflation	-0.972	0.489	<0.01	0.45 (0.51)	0.18 (0.89)
Unexpected inflation	0.760	2.280	0.739	0.64 (0.33)	0.23 (0.82)

Source: own calculation.

The results indicate a decrease in inflation in Czechia by 0.302 percentage points compared to the case when the strategy based on raising interest rates had not been implemented. Moreover, the expected inflation decreases more, by 0.972 percentage points relative to the case when interest rate would have not increase sharply. On the other hand, this strategy determined a higher unexpected inflation, but this increase is not statistically significant. These empirical findings suggest that the proactive increase in interest rates made by ČNB contributed to reduction in expected inflation, but without control on unexpected inflation.

Lower and more stable inflation creates a more predictable economic environment. This can improve business confidence and lead to better investment decisions. Lower inflation can pave the way for lower interest rates in the future. This can stimulate economic activity, investment, and potentially lead to a more stable economy overall. However, raising interest rates can also have drawbacks. First, higher interest rates can discourage borrowing and investment, potentially slowing economic growth. Second, businesses may be hesitant to hire new employees if the economy weakens due to rising interest rates.

5. CONCLUSIONS AND POLICY PROPOSALS

Predicting inflation is a complex task. Structural economic models and basic economic reasoning suggest that inflation should be predictable using various indicators. These indicators include measures of domestic and international economic activity, import prices or exchange rates, cost measures like wage growth, and oil prices. However, sentiment index as a measure of experts' perception on future inflation is neglected in all of these studies, but this approach might improve the accuracy of inflation forecasts. This paper explores the capacity of sentiment analysis to provide indexes for non-parametric Bayesian models that can improve forecast accuracy. The results highlight that Bayesian linear regression models provide better inflation predictions than GRNN model on the horizon 2023: Q1- 2023: Q4. The forecasts for inflation in Czech Republic were more accurate than predictions for Romania and the strategy for controlling inflation applied by Czech central bank reduced more the inflation. Actually, the expected inflation was reduced by rising interest rate, while this strategy had no significant effect on unexpected inflation.

Policy proposals could be provided starting from the empirical results. First, sentiment analysis from central bank reports could be integrated into inflation forecasting models. This can be achieved by developing sentiment analysis tools to analyze the tone and content of central bank reports. Second, central banks should be proactive in adjusting interest rates to control inflation. Central banks should closely monitor inflation trends and be prepared to adjust interest rates quickly. Transparent communication from central banks about their inflation expectations and policy decisions is crucial. Third, policymakers and businesses should prioritize short-term inflation forecasts for effective decision-making. Resources should be allocated towards developing and refining short-term forecasting models. Fourth, policymakers and central banks should continuously monitor and adapt their strategies based on evolving economic conditions. Policymakers and central banks should regularly assess the effectiveness of their inflation control strategies. The war in Ukraine has significantly impacted inflation dynamics, highlighting the need for flexibility. This suggests that they should be prepared to adjust their policies as needed in response to changing circumstances. Fifth, collaboration between central banks in Eastern Europe can be beneficial for sharing best practices and coordinating inflation control efforts. Addressing underlying structural issues in the economy, such as supply chain bottlenecks and energy dependence, can contribute to long-term inflation control. By implementing these policy proposals, Eastern European countries can improve their ability to manage inflation and navigate the current economic challenges.

This research presents few limitations like the inclusion of few predictors of inflation, the use of several forecasting methods and few countries in the analysis. Therefore, in a future study more inflation predictors will be considered by adding exchange rate, unemployment rate etc. Other forecasting methods might be taken into account like non-linear models. The sample of countries could be extended to more Eastern European countries like Hungary, Poland, Slovak Republic.

Acknowledgement: The authors gratefully acknowledge funding from the Academy of Romanian Scientists, in the "AOȘR-TEAMS-III" Project Competition EDITION 2024-2025, project name "Improving forecasts inflation rate in Romania using sentiment analysis and machine learning". The opinions expressed in this document are the sole responsibility of the authors and do not necessarily represent the official position of the Academy of Romanian Scientists.

References

1. Akhter, T. (2013). Short-term forecasting of inflation in Bangladesh with seasonal ARIMA processes.
2. Almosova, A., & Andresen, N. (2023). Nonlinear inflation forecasting with recurrent neural networks. *Journal of Forecasting*, *42*(2), 240–259. https://doi.org/10.1002/for.2901
3. Aras, S. & Lisboa, P.J. (2022). Explainable inflation forecasts by machine learning models. Expert Systems with Applications, 207, 117982.
4. Athey, S. & Imbens, G.W. (2019). Machine learning methods that economists should know about. Annual Review of Economics, 11, 685-725.
5. Athey, S. (2018). The impact of machine learning on economics. In The economics of artificial intelligence: An agenda (pp. 507-547). University of Chicago Press.
6. Babb, N. R., & Detmeister, A. K. (2017). Nonlinearities in the Phillips Curve for the United States: Evidence Using Metropolitan Data. *Finance and Economics Discussion Series*, *2017*(070). https://doi.org/10.17016/feds.2017.070
7. Barkan, O., Benchimol, J., Caspi, I., Cohen, E., Hammer, A., & Koenigstein, N. (2023). Forecasting CPI inflation components with Hierarchical Recurrent Neural Networks. *International Journal of Forecasting*, *39*(3), 1145–1162. https://doi.org/10.1016/j.ijforecast.2022.04.009
8. Baybuza, I. (2018). Inflation forecasting using machine learning methods. Russian Journal of Money and Finance, 77(4), 42-59.
9. Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.
10. Cano, A. (2018). A survey on graphic processing unit computing for large□scale data mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(1), e1232.
11. Cardani, R., Croitorov, O., Giovannini, M., Pfeiffer, P., Ratto, M., & Vogel, L. (2022). The euro area's pandemic recession: A DSGE-based interpretation. *Journal of Economic Dynamics and Control*, *143*. https://doi.org/10.1016/j.jedc.2022.104512
12. Cardani, R., Paccagnini, A. & Villa, S. (2019). Forecasting with instabilities: An application to DSGE models with financial frictions. Journal of Macroeconomics, 61, 103133.

13. Chakraborty, C. & Joseph, A. (2017). Machine Learning at Central Banks. Bank of England Working Papers, N 674.
14. Charpentier, A., Flachaire, E. & Ly, A. (2018). Econometrics and machine learning. Economie et Statistique, 505(1), 147-169.
15. Chen, S., Hardle, W. & Jeong, K. (2010). Forecasting volatility with support vector machine based GARCH model. Journal of Forecasting, 29, 406-433.
16. Choudhary, M. A., & Haider, A. (2012). Neural network models for inflation forecasting: an appraisal. *Applied Economics*, *44*(20), 2631-2635.
17. Cimadomo, J., Giannone, D., Lenza, M., Monti, F., & Sokol, A. (2022). Nowcasting with large Bayesian vector autoregressions. *Journal of Econometrics*, *231*(2), 500-519.
18. Clark, T. E., Huber, F., Koop, G., & Marcellino, M. (2022). *Forecasting US Inflation Using Bayesian Nonparametric Models*. http://arxiv.org/abs/2202.13793
19. Del Negro, M. & Schorfheide, F. (2013). DSGE model-based forecasting. In Handbook of economic forecasting (Vol. 2, pp. 57-140). Elsevier.
20. Deluna Jr, R.S., Loanzon, J.I.V. & Tatlonghari, V.M. (2021). A nonlinear ARDL model of inflation dynamics in the Philippine economy. Journal of Asian Economics, 76, 101372.
21. Döhrn, R. & Schmidt, C.M. (2011). Information or institution? On the determinants of forecast accuracy. Jahrbücher für Nationalökonomie und Statistik, 231(1), 9-27.
22. Dotsey, M., Fujita, S., & Stark, T. (2018). *Do Phillips Curves Conditionally Help to Forecast Inflation? **.
23. Estiko, F.I. & Wahyuddin, S. (2019). Analysis of Indonesia Inflation Using ARIMA and Artificial Neural Network. Economics Development Analysis Journal, 8(2), 151-162.
24. Garcia, M.G., Medeiros, M.C. & Vasconcelos, G.F. (2017). Real-Time Inflation Forecasting with High-Dimensional Models: The Case of Brazil. International Journal of Forecasting, 33(3), 679-693.
25. Gikungu, S.W., Waititu, A.G. & Kihoro, J.M. (2015). Forecasting inflation rate in Kenya using SARIMA model. American Journal of Theoretical and Applied Statistics, 12(1), 15-18.
26. Granger, C.W.J. (2004). Time series analysis, cointegration, and applications. American Economic Review, 94(3), 421-425.
27. Guegan, D. & Rakotomarolahy, P. (2010). Alternative methods for forecasting GDP. In Jawadi, F., Barnett, W.A., & Group, E. (Eds.), Nonlinear modelling of economic and financial time series (pp. 161-187). Emerald Group Publishing.
28. Haider, A. & Hanif, M.N. (2009). Inflation forecasting in Pakistan using artificial neural networks. Pakistan Economic and Social Review, 123-138.
29. Hak Kim, H., Swanson, N. R., Corradi, V., Hendry, D., Koop, G., Landon-Lane, J., Li, F., & Tkacz, G. (2008). *Mining Big Data Using Parsimonious Factor, Machine Learning, Variable Selection and Shrinkage Methods*. Kim and Swanson.
30. Hanif, M.N. & Malik, M.J. (2015). Evaluating performance of inflation forecasting models of Pakistan.
31. Hubrich, K. (n.d.). *Standard-Nutzungsbedingungen*. http://www.ecb.
32. Joseph, A., Potjagailo, G., Chakraborty, C. & Kapetanios, G. (2024). Forecasting UK inflation bottom up. International Journal of Forecasting.
33. Joseph, A., Potjagailo, G., Kalamara, E., Chakraborty, C., & Kapetanios, G. (2022). *Staff Working Paper No. 915 Forecasting UK inflation bottom up*. www.bankofengland.co.uk/working-paper/staff-working-papers
34. Karabatsos, G. (2018). Marginal maximum likelihood estimation methods for the tuning parameters of ridge, power ridge, and generalized ridge regression. *Communications in Statistics-Simulation and Computation*, *47*(6), 1632-1651.

35. Kim, D. H., & Lin, S. C. (2023). Income inequality, inflation and financial development. *Journal of Empirical Finance*, *72*, 468–487. https://doi.org/10.1016/j.jempfin.2023.04.008
36. Kripfganz, S. & Schneider, D.C. (2023). ardl: Estimating autoregressive distributed lag and equilibrium correction models. The Stata Journal, 23(4), 983-1019.
37. Leung, M. T., Chen, A. S., & Daouk, H. (2000). Forecasting exchange rates using general regression neural networks. *Computers & Operations Research*, *27*(11-12), 1093-1110.
38. Li, W. & Kockelman, K.M. (2022). How does machine learning compare to conventional econometrics for transport data sets? A test of ML versus MLE. Growth and Change, 53(1), 342-376.
39. Lidiema, C. (2017). Modelling and forecasting inflation rate in Kenya using SARIMA and Holt-Winters triple exponential smoothing. American Journal of Theoretical and Applied Statistics, 6(3), 161-169.
40. Marcek, M. & Marcek, D. (2006). Application of support vector machines to the modelling and forecasting of inflation. In Applied Artificial Intelligence (pp. 259-266).
41. Markus Jochmann. (2014). Modeling U.S. Inflation Dynamics: A Bayesian Nonparametric Approach. *Econometric Reviews*, *34*(5), 537–558.
42. Medeiros, M. C., Vasconcelos, G. F. R., Veiga, Á., & Zilberman, E. (2021). Forecasting Inflation in a Data-Rich Environment: The Benefits of Machine Learning Methods. *Journal of Business and Economic Statistics*, *39*(1), 98–119. https://doi.org/10.1080/07350015.2019.1637745
43. Mizrach, B. (1992). Multivariate nearest-neighbour forecasts of EMS exchange rates. Journal of Applied Econometrics, 7, 151-163.
44. Moshiri, S., & Cameron, N. (2000). Neural network versus econometric models in forecasting inflation. *Journal of forecasting*, *19*(3), 201-217.
45. Mullainathan, S. & Spiess, J. (2017). Machine learning: an applied econometric approach. Journal of Economic Perspectives, 31(2), 87-106.
46. Pahlavani, M. & Rahimi, M. (2009). Sources of inflation in Iran: An application of the ARDL approach. International Journal of Applied Econometrics and Quantitative Studies, 6(1), 61-76.
47. Pavlov, E. (2020). Forecasting inflation in Russia using neural networks. Russian Journal of Money and Finance, 79(1), 57-73.
48. Pérez-Pons, M.E., Parra-Dominguez, J., Omatu, S., Herrera-Viedma, E. & Corchado, J.M. (2021). Machine learning and traditional econometric models: a systematic mapping study. Journal of Artificial Intelligence and Soft Computing Research, 12(2), 79-100.
49. Pesaran, M.H., Shin, Y. & Smith, R.J. (2001). Bounds testing approaches to the analysis of level relationships. Journal of Applied Econometrics, 16(3), 289-326.
50. Petropoulos, F., Apiletti, D., Assimakopoulos, V., Babai, M.Z., Barrow, D.K., Taieb, S.B., ... & Ziel, F. (2022). Forecasting: theory and practice. International Journal of Forecasting, 38(3), 705-871.
51. Rodriguez, F., Rivero, S.S. & Felix, J.A. (1999). Exchange rate forecasts with simultaneous nearest-neighbour methods: evidence from EMS. International Journal of Forecasting, 15, 383-392.
52. Rossi, S. (2023). The distributional impacts of inflation-targeting strategies. In Central Banking, Monetary Policy and Income Distribution (pp. 261-273). Edward Elgar Publishing.
53. Saz, G. (2011). The efficacy of SARIMA models for forecasting inflation rates in developing countries: The case for Turkey. International Research Journal of Finance and Economics, 62, 111-142.
54. Shobana, G. & Umamaheswari, K. (2021, January). Forecasting by machine learning techniques and econometrics: A review. In 2021 6th international conference on inventive computation technologies (ICICT) (pp. 1010-1016). IEEE.

55. Simionescu, M. (2022). Econometrics of sentiments- sentometrics and machine learning: The improvement of inflation predictions in Romania using sentiment analysis. *Technological Forecasting and Social Change*, *182*, 121867. https://doi.org/10.1016/j.techfore.2022.121867.
56. Simionescu, M., & Nicula, A. S. (2024). Sentiment Analysis as Innovation in the Inflation Forecasting in Romania. *Marketing and Management of Innovations*, *15*(2), 13–25. https://doi.org/10.21272/mmi.2024.2-02.
57. Smets, F. & Wouters, R. (2007). Shocks and frictions in US business cycles: A Bayesian DSGE approach. American Economic Review, 97(3), 586-606.
58. Standford. (2018). Index 2018. Available at: https://hai.stanford.edu/ai-index-2018.
59. Standford. (2019). Index 2019. Available at: https://hai.stanford.edu/research/ai-index-2019.
60. Stock, J. H., Watson, M. W., Bauer, M., Bjørnland, H., Chodorow-Reich, G., Kilian, L., Plagborg-Møller, M., Ramey, V., & Anders, L. (2016). *Factor Models and Structural Vector Autoregressions in Macroeconomics*. https://research.stlouisfed.org/econ/mccracken/fred-databases/
61. Stock, J.H. & Watson, M.W. (1999). Forecasting inflation. Journal of Monetary Economics, 44(2), 293-335.
62. Thakur, G.S.M., Bhattacharyya, R. & Mondal, S.S. (2016). Artificial neural network based model for forecasting of inflation in India. Fuzzy Information and Engineering, 8(1), 87-100.
63. Tierney, H. L. R. (2019). Forecasting with the Nonparametric Exclusion-from-Core Inflation Persistence Model Using Real-Time Data. *International Advances in Economic Research*, *25*(1), 39–63. https://doi.org/10.1007/s11294-019-09726-7
64. Ülke, V., Sahin, A. & Subasi, A. (2018). A comparison of time series and machine learning models for inflation forecasting: empirical evidence from the USA. Neural Computing and Applications, 30, 1519-1527.
65. Wang, L. (2022). Research on the dynamic relationship between China's renewable energy consumption and carbon emissions based on ARDL model. Resources Policy, 77, 102764.
66. Woodford, M. (2007). The case for forecast targeting as a monetary policy strategy. Journal of Economic Perspectives, 21(4), 3-24.
67. Yusif, M.H., Eshun Nunoo, I.K. & Effah Sarkodie, E. (2015). Inflation forecasting in Ghana-artificial neural network model approach.
68. Zhang, X., Xue, T. & Stanley, H.E. (2018). Comparison of econometric models and artificial neural networks algorithms for the prediction of Baltic dry index. IEEE Access, 7, 1647-1657.

The Interplay of Demographic and Socioeconomic Factors in Financial Inclusion Across Romania's Regions

Stefan Johnson
Department of Demography and Geodemography, Faculty of Science, Charles University, Prague

ABSTRACT

This study examines the temporal evolution of financial inclusion across Romania's 42 counties from 2011 to 2021, analyzing regional disparities through a multidimensional lens. Using the Mahalanobis Distance-Based Financial Inclusion Index (IFIMd), the Second Demographic Transition Behavioral Index (SDT1), and hierarchical clustering, this research assess the relationship between financial access, demographic changes, and socioeconomic factors. Our findings reveal persistent inequalities, with GDP, pensioner-to-elderly ratio, and physician-to-adult population ratio emerging as key predictors of financial inclusion at the panel level. NUTS3-specific analyses underscore the influence of urbanization and aging populations, with Bucharest consistently exhibiting the highest inclusion levels. Linear regression out-performs Holt's exponential smoothing in forecasting financial inclusion trends, capturing linear growth patterns. These insights highlight the need for targeted policy interventions to bridge regional financial inclusion gaps and foster equitable economic participation. By providing the first subnational application of SDT1 alongside the first application of IFIMd longitudinally, this research advances understanding of financial inclusion dynamics in post-socialist contexts and informs strategies for addressing regional disparities.

Keywords: *Financial Inclusion; Economic Demography; Romania; NUTS3; Demographic Change*

JEL: *R, J*

INTRODUCTION

Romania's accession to the European Union (EU) in 2007, alongside Bulgaria, marked a pivotal moment in the EU's enlargement strategy, aimed at expanding economic integration and fostering regional stability (Emmert & Petrovi, 2014). This transition symbolized not only the aspirations of post-socialist nations for economic modernization but also their commitment to democratic and institutional reforms. However, Romania's integration was uniquely complex, shaped by a legacy of centralized planning, systemic

corruption, and institutional inefficiencies (Ciobanu, 2007). While early reforms, particularly under the 2004 Alliance of Truth and Justice government, sought to address these structural weaknesses, external shocks—most notably the 2008 global financial crisis—hindered Romania's trajectory toward economic convergence with its EU counterparts (Constantin, et al., 2011). This dual narrative of progress and persistent inequality remains central to understanding Romania's development, particularly in relation to financial inclusion.

Financial inclusion (FI), defined as access to and usage of financial services, has emerged as a crucial driver of economic development, social mobility, and regional cohesion. Yet, in Romania, progress in financial inclusion has been slow and uneven, trailing both global and regional benchmarks. In 2011, only 48% of the population had access to formal financial services, compared to 51% globally and 71% in Europe and Central Asia (Demirguc-Kunt & Klapper, 2013). While financial inclusion in Romania improved to 69% by 2021, it remained significantly below the global average of 76% and the regional average of 90% (Demirgüç-Kunt, et al., 2022). These figures underscore both Romania's advances and its continued struggle to bridge the gap with its European peers, reinforcing the role of financial exclusion as a barrier to economic and social development.

Despite increasing attention on financial inclusion, much of the existing research focuses on national-level trends, often overlooking the significant subnational disparities that characterize financial access in Romania. Aggregated national statistics fail to capture the localized economic, demographic, and institutional factors that shape financial behaviors across regions. This study seeks to address this gap by examining financial inclusion at the county (NUTS3) level over the period from 2011 to 2021. By incorporating a broad set of demographic and socioeconomic indicators, this research provides a deeper understanding of the spatial and temporal dynamics influencing financial access across Romania's 42 counties.

To achieve this, the study employs a comprehensive methodological framework, integrating forecasting, correlation analysis, regression models, and hierarchical clustering. Specifically, Holt's Exponential Smoothing, Simple Linear Regression, and Stepwise Regression are used to model financial inclusion trends and assess the predictive power of regional variables. Two key indices serve as the foundation for this analysis: the Mahalanobis Distance-Based Financial Inclusion Index (IFIMd) and the Second Demographic Transition Behavioral Index (SDT1). The SDT1 index, which captures evolving family structures and social behaviors through indicators such as mean age at first marriage, extramarital birth rates, and divorce rates, offers a novel lens to examine demographic change. Notably, Romania's SDT1 trajectory has diverged from that of other post-socialist

nations, reflecting a slower pace of transition. This study represents the first application of the SDT1 index at a subnational level, offering new insights into how demographic shifts intersect with financial access.

Beyond forecasting financial inclusion trends, hierarchical clustering is employed to identify regional patterns and structural similarities among counties. This method allows for a more granular exploration of subnational financial dynamics, revealing distinct clusters of counties that share similar inclusion profiles and underlying determinants. By integrating these analytical approaches, this study provides a more detailed understanding of financial disparities in Romania, highlighting the interplay between demographic change, economic development, and financial access.

This research contributes to the growing discourse on financial inclusion by situating Romania—one of the EU's most financially excluded member states—within the broader conversation on regional inequality. By focusing on the NUTS3 level, this study offers policy-relevant insights into the structural barriers that shape financial access and usage, providing a foundation for targeted interventions. Moreover, it is the first study to examine the temporal evolution of financial inclusion at the county level in Romania, addressing a critical gap in the literature. By integrating financial, demographic, and socioeconomic perspectives, this study seeks to advance the academic understanding of financial inclusion while offering practical insights for policymakers working to bridge regional disparities, promote financial accessibility, and reduce social exclusion.

LITERATURE REVIEW

This study aims to investigate the interplay and predictability of financial inclusion through a comprehensive examination of demographic and socioeconomic indicators. The demographic variables in this analysis include the Second Demographic Transition Behavioral Index (SDT1), life expectancy (LE), the urban-to-rural population ratio (URR), the mean age of the population (MAGE), net migration (NMIG), and the old-age dependency ratio (OADR). Concurrently, socioeconomic factors such as the unemployment rate (UNEMP), the number of pensioners receiving social security (PENS), GDP per capita in RON (GDP), the and physician-to-population ratio (PHSR). Indicators included in the SDT1 index are: mean age at first childbirth (MAFB), fertility rate of those under 20 (TEENFERT), total first marriage rate (TFMR), percent of nonmarital births (NONMAR), mean age at first marriage (MAFM), and total divorce rate (TDR). These nine indicators and SDT1 provide a nuanced lens to assess Romania's demographic evolution and the socioeconomic transformations shaping its contemporary economy and society.

The research framework is structured around three distinct but interconnected sections. The initial section reviews existing literature on financial inclusion, with a dual focus on national and sub-national scales, specifically at the NUTS3 level within the Romanian context. The subsequent section critically examines demographic variables, highlighting their role in shaping financial inclusion trends. Finally, the third section contextualizes these findings against a backdrop of Romania's socioeconomic shifts over the past decade. By exploring these dimensions, the study can establish a robust foundation for understanding the relationship between demographic and socioeconomic indicators against financial inclusion.

Kandari et al. (2021) found that gender and income were insignificant in account ownership in underdeveloped regions, echoing findings by Borooah & Iyer (2005). Significant relationships were found, however, between respondents' age and financial literacy. in particular, as respondents' age increased, so did bank ownership and financial literacy. Amari & Anis (2021), Kandari et al. (2021), Thu & Dao (2022), Tran et al. (2020), and Allen et al. (2016) offer insight into the impact of geography, culture, and socioeconomics on financial inclusion, and provide the importance of cultural nuance when exploring the demographic evolution and change upon a financial inclusion measure. Due to the limited discourse regarding post-socialist nations, and Romania in particular, in the context of financial inclusion and demographic change, this paper will be a novel exploration temporally at a NUTS3 level. By adopting a multidimensional approach, this study seeks to contribute to the broader discourse on financial inclusion and its interrelations with demographic and socioeconomic factors. In doing so, it endeavors to uncover novel insights into the subnational dynamics of financial inclusion, thereby addressing critical gaps in the existing body of literature, as discussed by Demirgüç-Kunt, et al. (2019).

Financial Inclusion. Financial inclusion (FI) refers to how individuals and businesses can access financial services—primarily savings, credit, and insurance—delivered responsibly and sustainably (World Bank, 2023; Demirguc-Kunt & Klapper, 2013). FI is recognized as a critical enabler of development, influencing nine Sustainable Development Goals (SDGs). However, its study has predominantly focused on three domains: education and literacy, income and wealth, and gender, collectively accounting for 64% of the existing research (Kara, et al., 2021). Research addressing additional demographic factors, such as age, race, social class, disability-related social exclusion, household size, and geographic location, remains comparatively sparse. This imbalance underscores a significant gap in understanding the broader demographic dimensions of financial inclusion.

Various frameworks and methodologies have been developed to estimate and measure financial inclusion. Among these, the World Bank's Global Financial Inclusion Database (FINDEX) is the most widely cited, providing a foundation for national-level analyses (Amari & Anis, 2021; Anzoategui, et al., 2014; Demirgüç-Kunt, et al., 2022; Khan, et al., 2020). While the FINDEX is comprehensive in question breadth, its triennial publication schedule results in data gaps during intermittent years. Before the FINDEX was implemented, Sarma (2008) proposed an indexing method later refined in 2012, adopting a Euclidean distance model for enhanced accuracy (Sarma, 2012). More recently, Li and Wang (2023) introduced a Mahalanobis distance-based model using the same indicators, but also accounts for variable correlations, marking a significant improvement in national-level financial inclusion metrics. Despite these advancements, little attention has been paid to the development of sub-national methods for measuring financial inclusion.

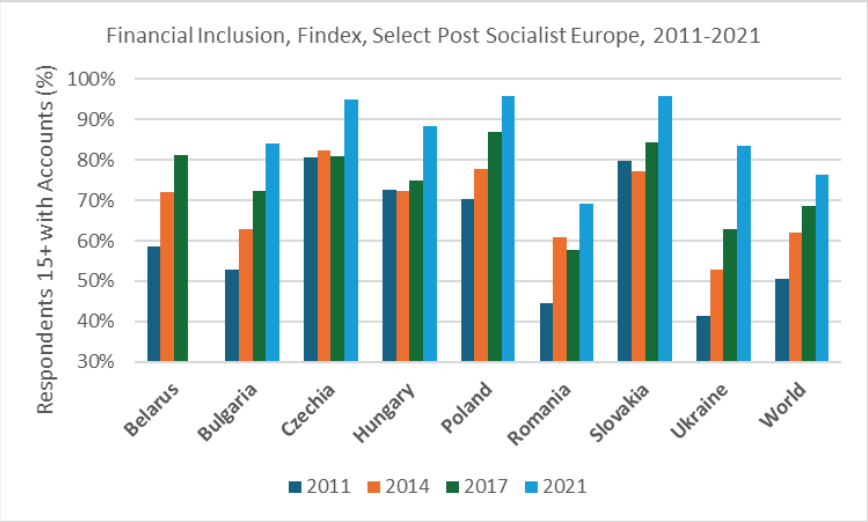
Though sparse in nature, sub-national financial inclusion indices have been explored through limited methodologies, including geometric means (del Carmen Dircio-Palacios-Macedo, et al., 2023; Gupte, et al., 2012) and Euclidean distancing (Yadav, et al., 2021). While these approaches offer their advantages, their application remains untested in the Romanian context. Moreover, these indices often rely on constrained parameters, limiting data availability and the robustness of measurement outcomes. A third approach, introduced at the NUTS3 level by Johnson (2025 Forthcoming) within the context of Romania. These considerations will be elaborated upon in subsequent sections to provide the context for this research's nuanced approach.

National-Level Financial Inclusion. The most common method of measuring financial inclusion has been the triennial World Bank Findex, a telephone-assisted survey administered to approximately 1,000 individuals aged 15+ in each of over 139 countries (Demirguc-Kunt & Klapper, 2013). Since 2011, Romania has exhibited notable variability in financial inclusion (FI), with trends reflecting periods of growth followed by decline. In 2011, only 45% of the population 15 and over were estimated to be included in the financial system (Demirguc-Kunt & Klapper, 2013). This figure rose to 61% in 2014 (Demirguc-Kunt, et al., 2015), decreased to 58% in 2017 (Demirgüç-Kunt, et al., 2018), and reached 69% by 2021 (Demirgüç-Kunt, et al., 2022). Although the overall trend has been positive, these fluctuations underscore Romania's status as the most under-banked country in Europe and one of the lowest within Europe and Central Asia (Demirgüç-Kunt, et al., 2019).

The global average for financial inclusion remains significantly higher than Romania across these years of study. *Figure 1*, based on World Bank

data, illustrates Romania's position relative to other countries in the region, including Czechia, Slovakia, and Poland. The differences are visibly unique in Romania from the rest of the region. Demirgüç-Kunt et al. (2019) highlighted this disparity and emphasized the importance of further investigation into Romania's unique FI landscape. In response, Khan et al. (2020) conducted a national-level analysis detailing Romania's banking environment. Their findings revealed that Romania surpasses the EU average in commercial banks per 100,000 adults (Romania: 28; EU: 25) and ATMs per 100,000 adults (Romania: 68; EU: 63). However, Romania lags significantly behind in the number of point-of-sale (POS) terminals per 100,000 adults (Romania: 1,177; EU: 2,819), a critical component for enabling digital and cashless transactions (Khan, et al., 2020). It should be noted that data utilized in the study was from 2018, prior to the COVID-19 pandemic.

Figure 1. Financial Inclusion, Findex, Select Post-Socialist Europe, 2011-2021



(Source: Author's Work, (World Bank, 2022))

These variances highlight Romania's distinct position within the Post-Socialist European financial landscape and disparities within the country itself (Khan, et al., 2020). Regional analyses based on NUTS3 divisions indicate significant within-country variation, further complicating national-level assessments of levels of inclusion (Khan, et al., 2020). Despite these limited observations from this data from 2018, there is a lack of discourse addressing the underlying causes of regional disparities. Existing research has

primarily focused on national-level metrics, which fail to capture the nuanced and complex dynamics at sub-national levels.

There are several merits associated with the Findex, including real-world responses, and a broad set of additional data, including respondent income quintile, employment status, age, gender, reasons for use, means of use, and reasons for exclusion – both voluntary and involuntary (Demirgüç-Kunt, et al., 2022; World Bank, 2022). In addition to these merits, there are several inherent issues associated with the Findex – primarily due to time and scope limitations. Though it is a commonly accepted measure of financial inclusion, it suffers from an underlying selection bias based on phone availability. Additionally, it is not possible to determine the locality of the respondents; thus, it is unknown how scattered or clustered respondents are geographically. Finally, estimations are based on gender and age weights; thus, the true inclusion levels are potentially skewed.

The subsequent subsection will explore methodologies for measuring sub-national financial inclusion. Romania's unique financial inclusion challenges, characterized by low overall inclusion rates despite an above-average density of ATMs and bank branches, underscore the limitations of national-level indices. These findings suggest the need for a more granular approach to understand and address the underlying factors shaping financial inclusion within the country.

Sub-Nation Level Financial Inclusion. National-level financial inclusion measures provide a broad overview of financial penetration across populations; however, as Khan et al. (2020) noted, such metrics fail to capture the nuanced realities within individual counties. Over the past decade, several methodologies have emerged to assess sub-national financial inclusion rates. Common approaches include relative indexing through geometric means (Gupte, et al., 2012; del Carmen Dircio-Palacios-Macedo, et al., 2023) and Euclidean distancing (Sarma, 2008; 2012; Yadav, et al., 2021). While each method presents distinct advantages, none have yet been applied to evaluate subnational financial inclusion rates in a European context and have been generally limited to only few geographical subregions. This section explores the literature and methodologies associated with subnational financial inclusion, identifying gaps and potential applications.

Gupte et al. (2012) introduced a geometric mean approach utilizing four key variables: outreach, usage, ease of transaction, and cost of transactions. These variables were normalized based on bank branch and ATM penetration and the number of loans per 1,000 adults. Ease of transaction incorporated directly related metrics (e.g., the number of accessible deposit

or loan account locations) and inversely related metrics (e.g., account opening costs, documentation requirements, and loan processing times). Each variable in this model was assigned an equal weight of 20%, regardless of its relative impact. While this approach provides a structured framework, it has notable limitations, including the data complexity and limitations in weights.

del Carmen Dircio-Palacios-Macedo et al. (2023) proposed a more comprehensive geometric mean model, the Index of Financial Inclusion Geometric (IFIG), incorporating 32 variables. These variables encompass a wide range of data, including correspondent banks, transaction accounts per 10,000 adults, payroll credits, point-of-sale transactions, and ATM usage per 10,000 adults. Unlike the Gupte et al. (2012) model, the IFIG employed a Benefit of Doubt (BoD) weighting method, allowing for dynamic weighting based on variable impact. Although this model has an increased robustness, its practicality is limited by the exhaustive data requirements, which may not be readily available for many geographies.

In contrast, Yadav et al. (2021) utilized Euclidean distancing to calculate financial inclusion indices in Indian sub-regions. Inspired by Sarma's (2012) approach, this method draws on data from sources such as the UNDP and parallels the Human Development Index (HDI). Yadav, et al.'s model, computes separate supply-side and demand-side indices to provide a comprehensive inclusion measure. The supply-side index comprises three components: availability (e.g., banks per 100,000 adults, banks per 1,000 km^2, bank employees per customer), penetration (e.g., accounts per bank per 1,000 adults), and usage (e.g., credit and deposit volumes as a proportion of GDP). From the demand-side, the model observes the access to savings, number of small borrowers, and the proportion of households that have access to credit. Despite its robustness, the model's complexity and reliance on extensive regional data limit its applicability beyond specific geographic contexts, such as India.

Johnson (forthcoming) proposed adopting the Mahalanobis distancing (RIFIMd) method developed by Li & Wang (2023) for assessing sub-national financial inclusion. While addressing its limitations, this approach incorporates elements of Sarma's (2012) Euclidean distancing, which includes penetration, usage, and access. Unlike Euclidean distancing, which assumes no relationships between variables, Mahalanobis distancing accounts for correlations among variables, providing a multidimensional analysis with a far more efficient data requirement than that of the previous sub-national methods. This model offers ease of use, moderate data requirements, and sophisticated inter-variable consideration, making it particularly suitable for subnational analyses, and can be applied throughout Europe's NUTS3 regions, including Romania.

Financial inclusion continues to be a pivotal issue in development, influencing nine of the seventeen Sustainable Development Goals (SDGs) (Demirgüç-Kunt, et al., 2022; Kara, et al., 2021). While national-level metrics have been central to understanding macroeconomic shifts (Demirguc-Kunt & Klapper, 2013), they fall short of capturing regional disparities, particularly in countries like Romania, with relatively low FI rates (Demirgüç-Kunt, et al., 2019). For high-income countries, sub-national variances in FI may be minimal. However, for Romania and other nations with significant regional disparities, exploring financial inclusion at the NUTS3 level is critical for addressing gaps and fostering equitable development.

Demographic Indicators. This research examines several demographic indicators that broadly align with themes of the second demographic transition (SDT), health outcomes, patterns of aging, social security, and urbanization. These factors collectively present a unique demographic profile of Romania and offer a foundation for exploring the extent to which these indicators interact with the RIFIMd. This section focuses on the current state of these indicators and reviews existing literature on their relationships with financial inclusion, both within Romania and in broader contexts. The demographic indicators include life expectancy (LE), the urban-rural ratio (URR), the mean age of the population (MAP), net migration (NMIG), and the old-age dependency ratio (OADR). These can be broadly themed as the Second Demographic Transition, health outcomes, and internal and external migration patterns.

Second Demographic Transition. The second demographic transition (SDT) is characterized by shifts in societal behaviors and attitudes related to marriage, fertility, divorce, and cohabitation (Sobotka, 2008; Rychtarikova, 1999). Initially conceptualized by van de Kaa and Lesthaeghe (1987) in Western Europe during the 1960s, SDT captured trends such as delayed first marriages and births, increased divorce rates, and a rise in cohabitation (Johnson, 2024; Zaidi & Morgan, 2017). In contrast, the transition within post-socialist countries (PSCs) did not begin until the end of the socialist era in Central and Eastern Europe (Rychtarikova, 1999). Following the collapse of socialism, many PSCs entered a period of economic and social upheaval, delaying these demographic shifts relative to Western counterparts (Johnson, 2024).

Sobotka (2008) developed the Second Demographic Transition Behavioral Index (SDT1), which measures temporal changes in teen fertility, divorce rates, mean age at first marriage, proportion of extramarital births, total first marriage rate, and mean age of mothers at first birth. Johnson (2024) further utilized this index,

comparing PSCs—Romania, Czechia, Slovakia, and Poland—to Austria. The findings indicated that Romania exhibited the slowest temporal change among the countries studied, reinforcing the dual-path framework of demographic transition described by Sobotka (2008), elucidating that economically sound countries develop on a "traditional" path; whereas, countries in economic crisis develop in somewhat of a "crisis", differing in path but presumably the "destination" remains the same. From 2004 to 2021, these indices highlighted significant disparities between post-modernist and post-communist countries (Johnson, 2024).

The SDT1 index has been positively associated with GDP per capita (GDP) and inversely associated with the Gini coefficient of income inequality (GINI) across various PSCs. However, Romania emerged as an outlier; stepwise regression analysis demonstrated that GINI was not a significant predictor of SDT1 for Romania (Johnson, 2024). Given the established relationship between SDT1 and GDP, particularly in Romania, examining the influence of SDT1 on changes in financial inclusion at the NUTS3 level becomes crucial for this research. A marker for SDT1 is the social normalization of women in the workforce and working in non-traditional roles. These changes are often marked by changing demographic behaviour, as illustrated by SDT1. Because the second demographic transition and the associated index, SDT1, have not yet been observed in the context of financial inclusion, this work will be the first of its kind to explore the potential relationship between this financial inclusion index (IFIMd) and SDT1.

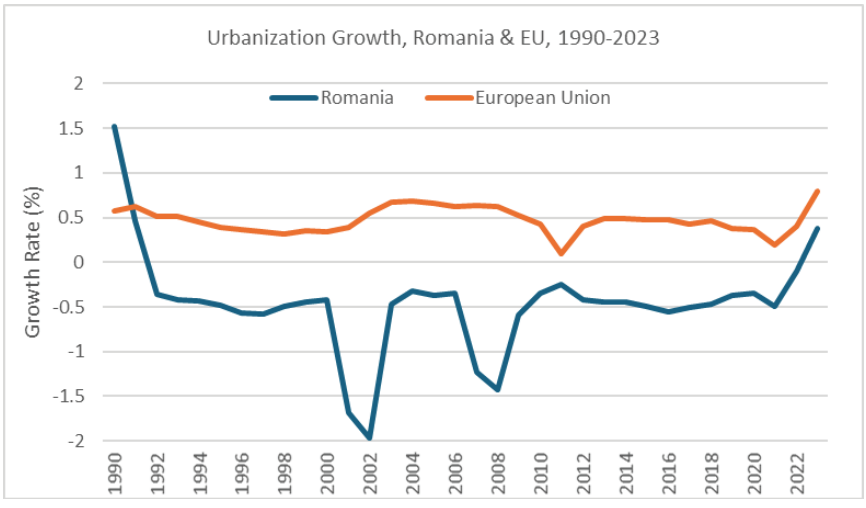
Life Expectancy. Life expectancy is integral to understanding demographic and health trends within a population. Across Europe, life expectancy has continued to increase, from 78.4 to 80.6 between 2004 and 2021. Much like other PSCs, Romania has also experienced improved life expectancy; however, life expectancy remains exceptionally low compared to other EU and PSC states (Muntele, et al., 2020). Romania's LE significantly increased over the observed years and outpaced the EU and several PSCs; however, it remains 5.5 years behind the European average.

Chowdhury & Chowdhury (2024) explored the relationship between the human development index and financial inclusion and uncovered a specifically unique relationship between financial inclusion and life expectancy in Bangladesh, India, and Pakistan. This corroborates the findings of Nanda & Kaur (2016), who found financial inclusion related to the human development index, including life expectancy across 68 countries. In contrast to those studies, Nica et al. (2023) discovered that there was some relationship between financial development and life expectancy in Eastern Europe, albeit the findings were insignificant – contrary to the findings of both Nanda & Kaur

(2016) and Chowdhury & Chowdhury (2024). These contrasting findings may be due to underlying differences in institutional quality (Nica, et al., 2023; Demirguc-Kunt & Klapper, 2013). For this reason, not only is life expectancy to be explored, but also the number of physicians per adult in each NUTS3 region. This will provide greater evidence of institutional quality and access in relation to both banking access and healthcare access.

Urbanization. Urban-rural migration in Romania appears to have intensified, driven in part by the escalating costs of urban housing, which were exacerbated by both population influx and the commercialization of urban spaces following the post-communist transition. This phenomenon has been partly attributed to a process of population de-concentration from urban centers to suburban areas (Grigorescu et al., 2012). Romania's transition from a centrally planned economy to a market-oriented one saw several phases of urbanization and ruralisation, primarily influenced by concerns over economic instability and the unaffordability of urban living (Ban, 2012; Ben-Ner & Montias, 1991). Notably, urban stagnation and even increased rural migration were most pronounced between 1990 and 2000, a period characterized by the dismantling of communist policies and Romania's efforts to stabilize and redefine its economic structure (Halbac-Cotoara-Zamfir et al., 2021). The only years during this period that saw positive urban growth were 1990, 1991, and 2023, as evidenced in *Figure 2*.

Figure 2. Urbanization Change, Romania and EU, 1990-2023



(Data Source: World Bank (2024))

The Harris-Todaro model provides a theoretical framework for understanding the motivations behind rural-to-urban migration, emphasizing the importance of identifying meaningful incentives for individuals to migrate from rural areas to urban centers (Harris & Todaro, 1970). This model was later expanded to account for not only income disparities but also the skills possessed by rural-urban migrants (Borjas & Bratsberg, 1994). Further elaborating on this, Todaro and Smith (2015) explained that migration is not solely driven by lower wages or the availability of skilled labor, but rather, in many contexts, by the perceived improvement in the standard of living. This theoretical approach offers valuable insight into the migration dynamics in Romania following the end of communism, where the relative standard of living, shaped by wage levels in relation to the cost of living, played a significant role in shaping urban migration patterns (Stănescu, 2018). This urban struggle was further compounded by a generally rising standard of living in rural Romania (Stănescu, 2021), creating a scenario in which individuals sought alternative living arrangements in response to these divergent conditions. Thus, urbanization is often times characterized by individuals and families seeking better opportunities and improvements in standard of living.

Li, Chen, and Hao (2024) examined the relationship between urbanization and financial inclusion, highlighting a nuanced interaction where increased urbanization was positively correlated with higher levels of financial inclusion in China. However, the study also revealed a paradox: as urbanization progressed, the utilization of borrowing and loans declined. This suggests a constraint in urban development driven by the limitations of the financial system, indicating a dual challenge—while urbanization can foster greater financial inclusion, the lack of sufficient financial inclusivity hampers further urbanization. As urban areas expand, the demand for accessible financial services rises, yet without comprehensive inclusion mechanisms, economic participation remains limited, hindering broader socio-economic development. Thus, urbanization may indeed impact financial inclusivity; however, it hinges on institutional quality, as discussed by Nice, et al. (2023) and Demirguc-Kunt & Klapper (2013).

Net Migration. Considerable research has been conducted regarding outward migration from Romania, as the problem has raised concerns regarding the long-term impact of depopulation (Otovescu & Otovescu, 2019). Goga & Ilie (2017) discussed the emigration issues associated with high reliance on remittance and the role that brain drain plays in the social structure of Romania. Over the decades since the transition out of communism, patterns of migration have shifted. Many migrants between 1990 and 2006 emigrated as low-skilled workers with less education; however, the current problem is that

many outbound migrants are high-skilled workers, such as doctors (Otovescu & Otovescu, 2019; Botezat & Moraru, 2020; Gavriloaia, 2020). This results in both an amplified reliance on remittances and a reduction in skilled, highly trained professions. An additional issue that this pattern of change introduces is individuals within the formal sector of work are leaving the country; unlike those who were more likely part of the informal economic environment in decades past.

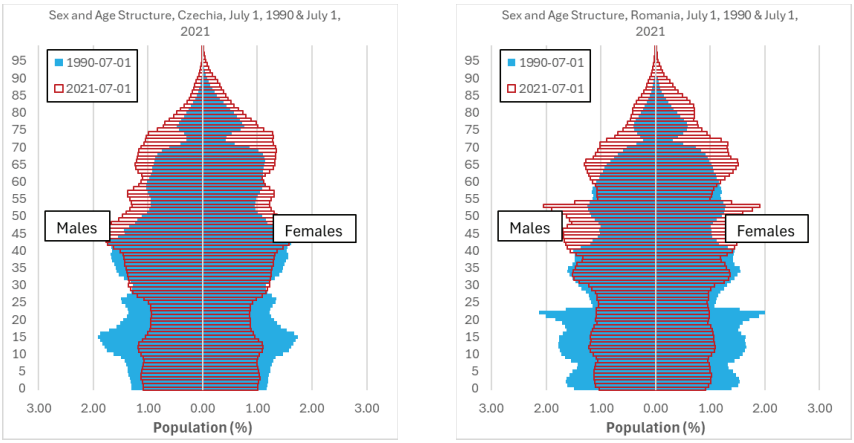
Johnson (2022) explored the relationships between financial inclusion, net migration (NMIG), educational attainment, and income quintile. Strong inverse relationships existed between financial inclusion and net migration in the age groups 25-29, 35-39, 40-44, and 45-49. This infers that as emigration increases in these select age groups, financial inclusion increases, corroborating the previous studies that associate increased migration with increased remittances (Tarsem, 2018; Sandu, 2010), and reducing the number of individuals within the informal sector. An additional element of the study was a strong positive relationship in the age groups 55-59, 60-64, and 65-69. As net migration increases in these age groups, so does financial inclusion – as foreign workers return to Romania in their old age, they presumably bring with them bank accounts and earnings.

Evidence suggests migration impacts financial inclusion, but this has only been explored at the national level limited to Findex data and has yet to be explored at the subnational level. This research will build on discourse and literature by Otovescu & Otovescu (2019), Botezat & Moraru (2020), Gavriloaia (2020), and Johnson (2022), where physician ratio, life expectancy, and old age dependency ratio will be explored beside net migration and financial inclusion. Combined, these characteristics will be explored to further understand and develop an inference mechanism that elucidates the likelihood of this nation-level change remaining the case under the consideration of subnational RIFIMd, and observing panel data in comparison to the Findex data.

Age Structure. Mean age is influenced by life expectancy but more so by fertility rates (Caselli, et al., 2006). Romania's age structure rapidly shifted at the end of the communist era, much like Bulgaria and Hungary (Muntele, 2024) due to several factors, including reduced fertility rates, increased outward migration, and since around 2000, an increased life expectancy relative to the communist era. In recent years, however, some indications show that mean ages have decreased in metropolitan areas despite a national-level shift. Certain ethnic populations have differing age structures – for instance, the Roma population maintains a lower mean age, whereas ethnic Romanians, Hungarians, Ukrainians, and Russians tend to show a higher mean age (Vasile & Dobre, 2015). Prior to the transition from communism in 1989, the share

of the elderly in Romania was approximately 10.3% of the population, with a median age of 32.6 years and a demographic dependency ratio of 51.5%. By 2020, 18.9% of the population was 65+, and the median age rose to 41.8 years. For context, *Figure 3* illustrates the significant structural transition that occurred in Romania against Czechia, a post-socialist country which exited communism in a more democratic fashion (Johnson, 2024).

Figure 3. Population Pyramids, Czechia and Romania, July 1, 1990 & July 1, 2021



(Source: Eurostat, Author's Own Calculation)

This research will explore the relationship these indicators have with banking and inclusion. This is especially important when considering the migration patterns of the 25-49 age groups and financial inclusion, notably, the 55-69 age groups. Old age groups return with bank accounts; younger professionals remit to family. The predicted issue arising from this is a population shrinking due to below-replacement fertility rates, emigration, and an aging population relying on pensions (Toma & Tuchilus, 2019; Roman, et al., 2018).

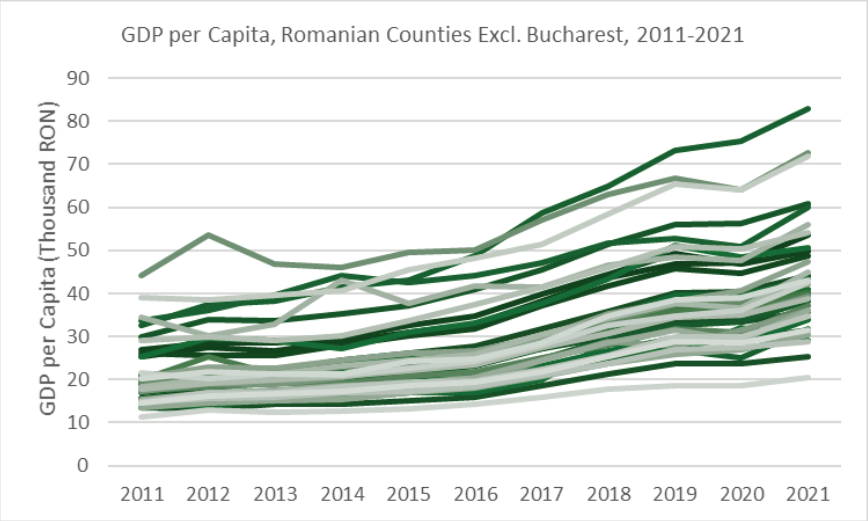
Economic Indicators. Romania's economy has undergone profound transformations since the fall of communism, marked by a challenging transition from a centralized economic model to an open-market system. While aligning closely with patterns observed in other post-socialist countries (PSCs), this transition brought about a series of economic crises as the population adjusted to the demands of a globalized economy (Johnson, 2024; Ban, 2012; Rychtarikova, 1999). Understanding these shifts is critical for analyzing the country's more recent economic trajectory, particularly between

2011 and 2021, as they shed light on the underlying dynamics of regional development and financial inclusion.

This section evaluates the economic factors that have driven temporal changes in Romania from 2011 through 2021, placing them within their broader historical and regional contexts. The analysis focuses on three key economic indicators: GDP, social security pensioners per 100 65+(PENS), and the unemployment rate (UMEMP). Each of these variables offers unique insights into Romania's development landscape. GDP serves as a comprehensive measure of economic performance across the NUTS3 regions, highlighting disparities and growth trends (Cristina, et al., 2021). Pensions reflect government spending on the elderly population, providing a lens into social support structures and their regional variations (Toma & Tuchilus, 2019). Lastly, the study of unemployment rates aims to uncover potential connections between social exclusion and financial inclusion, assessing whether these gaps in financial inclusivity are exacerbated by unemployment rates (Koku, 2015). Fernández-Olit et al. (2021) uncovered an opposing perspective, which found that FI and social exclusion were not related in a high-income country. This research will add to the literature discussing the relationship between financial inclusion levels and social exclusion by unemployment in the Romanian context.

Romania's economic disparities across regions are well-documented in literature. Russu and Ciuiu (2020) underscore significant variance in GDP among the NUTS3 regions, with Bucharest standing out as the most advanced and prosperous economy, followed distantly by Cluj. Other counties such as Timiș, Constanța, Brașov, and Ilfov also rank among the more economically developed regions, reflecting their industrial and infrastructural advantages. Conversely, counties such as Giurgiu, Vaslui, Botoșani, Teleorman, and Suceava have consistently recorded the lowest GDP figures, illustrating persistent challenges in regional development through 2021. Despite these disparities, there are signs of progress. Russu and Ciuiu (2020) identify a trend of convergence, where some of the poorest counties are gradually catching up to the national average, signaling advancements in economic cohesion and development.

Figure 4. GDP per Capita, Romanian Counties Excl. Bucharest, 2011-2021



(Source: National Institute of Statistics – Romania (2024))

The accelerated GDP growth observed in Romania's poorer regions is noteworthy, particularly compared to NUTS3 regions in other developed EU countries (Stawicki & Wojewódzka-Wiewiórska, 2023). This suggests that while Romania faces significant development hurdles, especially in its less economically advanced counties, the pace of growth has the potential to reduce long-standing inequalities (Russu & Ciuiu, 2020). Understanding these dynamics is crucial, as they shape the nation's socio-economic fabric and influence broader indicators such as financial inclusion and quality of life (Sen & Laha, 2021; Sakyi-Nyarko, et al., 2022; Demirguc-Kunt, et al., 2017). By examining GDP, pension expenditures, and unemployment, this study provides a nuanced view of Romania's economic evolution, emphasizing the interplay between regional development and national progress.

METHODOLOGY

This research examines the relationships between demographic and socioeconomic indicators with financial inclusion over time at the NUTS3 level in Romania. The financial inclusion index, RIFIMd, is calculated using the Mahalanobis distance model, as proposed by Li and Wang (2023) and further applied by Johnson (forthcoming). This model was chosen due to its associative properties, which allow for measuring inter-variable relationships,

unlike geometric mean or Euclidean distance methods. While theoretically robust, the geometric mean requires data not readily available in Romania and does not adequately weight variables by their correlation strength, making it unsuitable for this study.

Calculating IFIMd. The computation of IFIMd involves four steps, beginning with the standardization of variables using normalization, *P, A, U* are using *Equation 1*:

$$X_{ij}^{t} = \frac{x_{ij}^{t} - \min(x_j^t)}{Max(x_i^t) - \min(x_j^t)} \quad (1)$$

Where X_{ij}^{t} represents the standardized value of variable *j* in county *i* at time *t*. x_{ij}^{t} represents the non-standardized data for county *i*, for variable *j*, at time *t*. The min and max x_j^t denote the lowest and highest raw data under variable *j* at time *t*. The output of this is a standardized coefficient of $0 \leq X_{ij}^{t} \leq 1$. The mahalanobis distancing method is to specific reference points, rather than from mean as it provides a superior reference to perfect exclusion and perfect includion.

Equations 2, 3, and 4 outline the calculation of the Mahalanobis distance-based metrics to the specific reference point of perfect exclusion. Md_i^0 (*Equation 2*) measures the closeness to complete financial exclusion (*0,0,0*).

$$Md_i^0 = \sqrt{X_i^T \cdot \sum^{-1} \cdot X_i} \quad (2)$$

Similarly, Md_i^1 (*Equation 3*) measures the closeness to complete financial inclusion (*1,1,1*), where, *E* is the vector mean, *X* is the NUTS3 region vector, and $(E - X_i)^T$ represents the transposed deviation from the mean:

$$Md_i^1 = \sqrt{(E - X_i)^T \cdot \sum^{-1} \cdot (E - X_i)} \quad (3)$$

Finally, The normalized IFIMd is calculated as shown in *Equation 4*, demonstrating the relative distance from both reference points – perfect inclusion and exclusion:

$$IFIMd_i^t = \frac{Md_i^0}{Md_i^0 + Md_i^1} \quad (4)$$

This method provides robust clustering of observations due to the associative component embedded in the Mahalanobis distance calculation. All computations are conducted using SAS software.

SDT1 Behavoural Index. To complement the financial inclusion analysis, the SDT1 Behavioral Index, developed by Sobotka (2008), measures the progression of regions toward the second demographic transition. The SDT1 index incorporates six equally weighted variables: teen fertility rate, mean age at first live birth, percentage of extramarital births, total first marriage rate, mean age at first marriage, and total divorce rate. The RIFIMd and SDT1 indices have been adapted and calculated for all Romanian NUTS3 regions to evaluate their interplay. *Table 1* illustrates the maximum and minimum values for each of the variables.

Table 1. SDT1 Behavioural Index Upper and Lower Thresholds

Factor	Factor Abbreviation	SDT score = 0	SDT Score = 5	SDT Score = 10
Mean Age of Mother at First Birth	MAFB	<24	27	>30
Age Specific Fertility Rate Below Age 20 (per 1,000)	TEENFERT	>180	90	0
Percentage of Non-Marital Live Births	NONMAR	0	30	>60
Total First Marriage Rate	TFMR	>0.80	0.60	<0.40
Mean Age at First Marriage	MAFM	<23	27	>31
Total Divorce Rate	TDR	<0.15	0.35	>0.55

(Source: Sobotka (2008))

The rates are each scored based on the criteria in *Table 1*, aggregated, and averaged, leading to the SDT1 index score, where 0 is not considered to be transitioning, and 10 is regarded as an advanced second demographic transition. Sobotka added 0.5 to each of the indexed scores for countries where cohabitation counts for more than 10 percent of the total unions in the country. This type of analysis has not been conducted at a sub-national level; however, due to the association between more post-modernism and greater financial inclusion, this under-discussed index opens the study to a unique and novel perspective of financial inclusion, as well as regional variances in SDT change.

Correlative Testing and Simple Regression. The first stage of analysis explores pairwise correlations between the dependent variable (IFIMd) and independent variables using Pearson's correlation coefficient, and covariances amongst the independent variables. Pearson's correlation is suitable for this portion of the study due to the completeness of the panel data, allowing for a comprehensive overview of potential relationships. These correlations are tested indiscriminately, examining all relationships between dependent and independent variables to uncover preliminary insights. The 11 years in the panel data provide a foundational understanding of not only the

potential relationships between these variables, but importantly, the potential covariances that may occur amongst variables as well; thus, providing a more contextual analysis in further steps.

Pearson's r was the chosen method, as there are no gaps in the data, and remains continuous. The relationship and data for each of the variables is linear in nature – no significant outliers for the data. Because these are not ranked variables, Pearson's r is the optimal. Following correlation testing, the study evaluates the univariate predictability of IFIMd for each Romanian county using time-series models. This process will be further described in the following subsection.

Independent Predictability. This study compares simple linear regression and Holt's Exponential Smoothing with Trend to assess univariate predictability. Both methods offer distinct insights into the temporal evolution of IFIMd. Simple linear regression is a time-series approach to the associative technique used to capture linear trends. On the other hand, Holt's is a time-series model that introduces a trending component that accounts for acceleration or deceleration in predicted change. This is particularly important, as there have been observably accelerated changes in the Findex between 2011 and 2021. These methods will indicate how predictable IFIMd is when isolated. This will further aid in understanding how the data is trending and flowing over time.

Though not commonly used in demographic and geographic analysis, it has been recently used in Johnson's (2024) analysis of the predictability of SDT1 and total fertility rates across several post-socialist populations. Holt's is expressed in two functions: a level component (L)- the smoothing values of the series, and the trend component (T) – a representation of the accelerated or decelerated trend in the model. Observed values (Y) from the previous period(s) must be known to conduct this analysis. A smoothing constant for L is used in the level equation (α), and a trending (T) variable is added, along with a trending constant (β). The predicted future value ($\hat{Y}$) is then calculated. Equations (5), (6), and (7) illustrate the principal components.

Equation 1: Holt's Exponential Smoothing

$$L_t = \alpha \cdot Y_{t-1} + (1-\alpha) \cdot (L_{t-1} + T_{t-1}) \quad (5)$$

$$T_t = \beta \cdot (L_t - L_{t-1}) + (1-\beta) \cdot T_{t-1} \quad (6)$$

$$\hat{y}_t = L_t + T_t \quad (7)$$

The predictive performance of each method is evaluated using mean absolute deviation (MAD) and mean percentage error (MPE). Lower MAD and MPE values indicate better model performance. Holt's method, with its ability to capture non-linear trends, is particularly suited to the observed growth

patterns of IFIMd; however, this will be tested in this study to determine the performance of this method against the more traditionally implemented linear regression. This can be especially valuable due to the trended changes in the World Bank Findex between 2011 and 2021.

Stepwise Regression. After univariate testing, stepwise regression is used to identify the key predictors of IFIMd. Stepwise regression iteratively adds (p-in) and removes (p-out) variables based on their statistical significance ($p \leq 0.05$) retaining only those that contribute meaningfully to the model. The general form of the regression model is shown in *Equations 8* through *10*:

Equation 2: Multiple Regression, Stepwise

$$\hat{Y} = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 + \cdots + \beta_5 \times x_5 \quad (8)$$

$$\hat{Y}_{IFIMd} = \beta_0 + \beta_1 \times SDT1 + \beta_2 \times LE + \beta_3 \times URR \ldots + \beta_{10} \times PhysR \quad (9)$$

$$\hat{Y}_{IFIMd} = \beta_0 + \beta_1 \times MAGE + \beta_2 \times UNEMP + \beta_3 \times SDT1 \ldots + \beta_6 \times PhysR \quad (10)$$

The stepwise approach is employed for both panel data, which integrates all regions and years, and time-series data, which examines each region individually. This dual analytical framework allows for the identification of macro-level trends while simultaneously capturing regional variations. By adopting this two-pronged methodology to assess the predictive power of independent variables on IFIMd, the underlying factors contributing to regional disparities can be more effectively identified and analyzed.

To further refine the analysis, Ward's hierarchical clustering will be applied to group NUTS3 regions based on their IFIMd values and predictor variables, offering insights into the geographic and temporal dimensions of financial inclusion. All analyses will be conducted in SAS. These two hierarchical approaches will assess the predictive power of IFIMd and enable a comparative mapping of the results. At this stage, it remains uncertain whether the outcomes of these comparative hierarchical clusters will align; however, this approach will help clarify how different predictors of IFIMd may exert varying levels of influence across distinct IFIMd classifications.

RESULTS

This section presents the key findings from the analytical methods used to examine the relationship between financial inclusion (IFIMd) and demographic and economic indicators across Romania's NUTS3 regions. The results are organized into three core subsections. First, a correlative analysis assesses the strength and direction of relationships between IFIMd and its predictors, both at the national level and within individual counties. Second,

a comparative evaluation of forecasting methods—Holt's Exponential Smoothing and simple linear regression—examines the temporal predictability of IFIMd trends across the regions. Finally, stepwise regression identifies the most significant predictors, utilizing both panel and time-series data to capture trends at both macro and subnational levels. The section concludes with an analysis of regional clustering using Ward's hierarchical method.

The panel data results reveal strong associations between IFIMd and key indicators such as GDP, the urban-rural ratio (URR), and physician ratio, reinforcing previous findings in the literature. While Holt's Exponential Smoothing proved more effective in capturing non-linear growth patterns across counties, it exhibited a greater tendency to overestimate and underestimate financial inclusion levels compared to linear regression. Stepwise regression highlighted five principal predictors—GDP, SDT1, the physician ratio, URR, and the pensioner-to-elderly ratio—with an adjusted R^2 of 0.6824, explaining a substantial portion of the variation in IFIMd. These combined methods provide a robust framework for understanding both national trends and regional disparities, with a focus on policy-relevant insights to address financial inclusion gaps.

Table 2 summarizes the descriptive statistics of IFIMd values from 2011 to 2021 for Romania's 42 counties. As expected, Bucharest consistently ranks highest in financial inclusion, with an index score of 0.55 in 2011, rising to 0.87 in 2021. In contrast, Caraș-Severin remains at the bottom of the rankings throughout the study period. Notably, while Bucharest experienced a minor dip in 2018, Călărași's IFIMd increased slightly, albeit stagnating between 2018 and 2019. Despite these fluctuations, Bucharest's overall financial inclusion gains were substantially greater than those observed in low-ranked counties, reinforcing the persistent regional disparities that continue to define Romania's financial landscape.

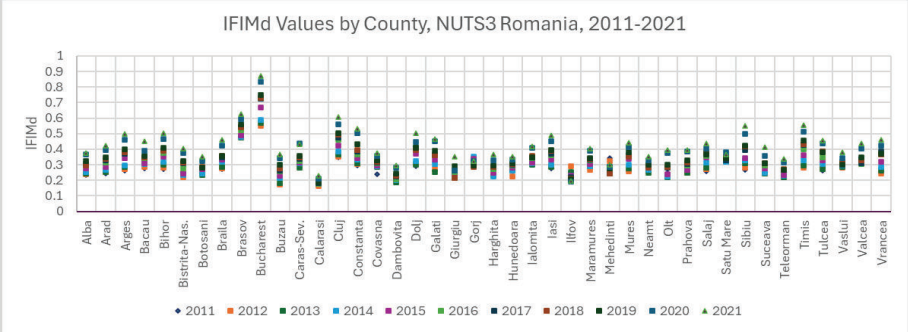
Table 2 Descriptive Statistics, IFIMd, NUTS3, Romania, 2011-2021

Descriptive Statistics - IFIMd 2011-2021, NUTS3 Romania											
	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
Mean	*0.28*	*0.28*	*0.28*	*0.29*	*0.31*	*0.34*	*0.34*	*0.34*	*0.35*	*0.40*	*0.43*
StdErr	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.01*	*0.02*	*0.02*
Median	*0.27*	*0.28*	*0.28*	*0.29*	*0.31*	*0.32*	*0.33*	*0.34*	*0.35*	*0.39*	*0.42*
Minimum	*0.17*	*0.18*	*0.18*	*0.18*	*0.18*	*0.18*	*0.18*	*0.20*	*0.19*	*0.19*	*0.19*
Maximum	*0.55*	*0.57*	*0.57*	*0.59*	*0.67*	*0.73*	*0.73*	*0.73*	*0.75*	*0.83*	*0.87*
Range	*0.38*	*0.39*	*0.39*	*0.40*	*0.49*	*0.55*	*0.55*	*0.53*	*0.56*	*0.64*	*0.68*
StdDev	*0.07*	*0.07*	*0.07*	*0.07*	*0.08*	*0.09*	*0.09*	*0.09*	*0.09*	*0.11*	*0.11*

(Source: Authors Calculations, Excel)

Bucharest consistently maintains the highest level of financial inclusion among all Romanian counties throughout the 11-year observation period. Following at a considerable distance, Brașov ranks second, with its IFIMd value remaining below 0.65, while Cluj follows in third, staying below 0.62 throughout the study period. Despite these rankings, financial inclusion outside of Bucharest exhibits relatively moderate variation across counties. *Figure 5* underscores a key pattern: while Bucharest is a clear outlier, the financial inclusion levels across the remaining 41 counties show less pronounced differences. The annual county-level averages for financial inclusion index values—0.28 in 2011 and 0.43 in 2021—illustrate a steady upward trajectory. However, these figures also highlight that most counties remain relatively close to the national average, with Bucharest standing as the singular exception, significantly outpacing all other regions.

Figure 5. IFIMd Values by County, NUTS3 Romania, 2011-2021



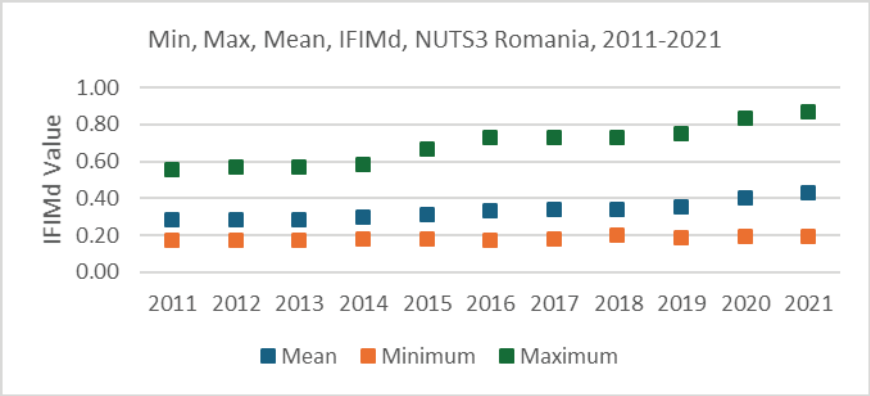
(Source: Author's Calculations, Excel)

Financial inclusion, as measured by the Findex, exhibited a decline between 2014 and 2021, with 2017 marking a particularly low point compared to both preceding and subsequent years. In contrast, the IFIMd values indicate a steady upward trend in financial inclusion over the study period, with only minor declines observed in select counties. The most notable decrease occurred in Bucharest, where IFIMd dropped by a modest 0.0028, alongside similar but small declines in Botoșani and Dolj. However, of these, only Bucharest recorded a reduction in bank accounts per 1,000 adults, falling from 1,066 to 1,039 (*Appendix 2*), though it remained significantly ahead of all other counties.

Figure 6 further highlights the stark contrast between the highest and lowest IFIMd values, emphasizing the persistent gap in financial inclusion across Romania. More notably, it underscores the substantial disparity between

Bucharest and the national mean, demonstrating the capital's exceptional financial accessibility relative to the rest of the country. These findings suggest that while financial inclusion has broadly improved, regional inequalities remain deeply entrenched.

Figure 6. Min, Max, and Mean for IFIMd in NUTS3 Regions, Romania, 2011-2021



(Source: Author's own calculations)

Despite Bucharest's consistently high IFIMd scores, which contribute to an overall increase in the national average, some counties continue to exhibit significantly lower levels of financial inclusion. Calarasi remains one of the lowest-scoring counties, with its IFIMd value exceeding 0.20 only in 2020. Buzău, which recorded the lowest IFIMd in 2017 at 0.17, surpassed Calarasi in 2018, while Ilfov briefly fell below Calarasi in 2020 and 2021. These trends underscore a critical observation: while financial inclusion has improved overall, regional disparities not only persist but have become more pronounced over time. The standard deviation of IFIMd increased from 0.07 to 0.11 by 2021, indicating widening inequality in financial access. Though the lowest-scoring counties may be statistical outliers, the trend suggests that gaps in financial inclusion are deepening, rather than converging. Moreover, Bucharest's exceptionally high IFIMd distorts the national average, increasing the mean change by 0.03 when included, highlighting the extent to which the capital's financial infrastructure differs from the rest of the country.

Alongside IFIMd, this study introduces a second key index—SDT1—which contributes to the literature on the second demographic transition (SDT). Originally conceptualized by Sobotka (2008) and later expanded upon by Johnson (2024), the SDT1 index has primarily been applied at the national level.

This study marks the first attempt to adapt SDT1 to a subnational (NUTS3) framework, offering a more granular perspective on demographic shifts within Romania. The subnational outputs of SDT1 provide a novel dimension to understanding demographic change, allowing for a more localized examination of its relationship with financial inclusion. A full breakdown of the NUTS3-level results is available in *Appendix 3*, while *Table 3* presents the descriptive statistics for Romania's 42 counties over the 11-year study period.

Table 3 Descriptive Statistics, SDT1 Index, NUTS3, Romania, 2017-2021

Descriptive Statistics - SDT1 2011-2021, NUTS3 Romania											
	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
Mean	*3.09*	*3.23*	*3.38*	*3.62*	*3.86*	*4.06*	*4.37*	*4.48*	*4.68*	*5.40*	*5.04*
Minimum	*2.13*	*2.24*	*2.36*	*2.60*	*2.88*	*3.19*	*3.70*	*3.32*	*3.75*	*4.74*	*4.20*
Maximum	*5.29*	*5.37*	*5.35*	*5.65*	*5.60*	*5.67*	*5.70*	*5.59*	*6.22*	*7.11*	*7.00*
Range	*3.16*	*3.13*	*2.98*	*3.05*	*2.72*	*2.48*	*2.00*	*2.27*	*2.47*	*2.37*	*2.80*
Standard Error	*0.42*	*0.10*	*0.09*	*0.09*	*0.09*	*0.08*	*0.07*	*0.07*	*0.07*	*0.07*	*0.08*
Median	*2.97*	*3.10*	*3.30*	*3.53*	*3.73*	*3.91*	*4.30*	*4.40*	*4.58*	*5.33*	*4.98*
StdDev	*0.63*	*0.62*	*0.59*	*0.60*	*0.58*	*0.51*	*0.47*	*0.45*	*0.48*	*0.48*	*0.53*

(Source: Author's Calculation)

A notable trend emerges in 2021, with nearly all counties experiencing a decline in SDT1, likely influenced by the social and economic disruptions caused by the COVID-19 pandemic. One particularly significant observation is the fluctuation of the highest and lowest SDT1 coefficients over time. While Bucharest, Romania's largest and most urbanized region, was expected to consistently lead in post-modernist demographic behaviors, it did not hold the highest SDT1 score in 2017 and 2018. This period coincided with a decline in financial inclusion, as measured by Findex, and a reduction in bank account ownership per 1,000 inhabitants. The SDT1 index serves as a valuable measure of shifting social and demographic norms, capturing Romania's transition toward non-traditional behaviors aligned with the second demographic transition (SDT). Its fluctuations over time highlight the evolving demographic landscape and its interaction with economic and institutional changes.

The interplay between financial inclusion and demographic change is central to this study, with both IFIMd and SDT1 providing complementary insights. Previous research has established a positive association between the second demographic transition and increasing economic equality, often linked to rising GDP (Johnson, 2022). Given that GDP is widely recognized as a key driver of financial inclusion, it is reasonable to expect that SDT1 will play a role in shaping IFIMd patterns as well. The following sections will analyze

SDT1 alongside nine other demographic and socioeconomic indicators, assessing both their correlation with and predictive influence on IFIMd. At its core, this study marks the first attempt to apply the SDT1 behavioral index at a subnational level while also introducing a regional financial inclusion index within a European context. By integrating these measures, this research offers new insights into Romania's persistent regional disparities in financial inclusion, extending the discourse initiated by Demirgüç-Kunt et al. (2019) and contributing to a broader understanding of financial and demographic dynamics in post-socialist economies.

Correlative Testing. The second stage of this research involves identifying the relationships between the dependent variable (IFIMd) and the demographic and socioeconomic indicators using Pearson's r correlative analysis. The primary focus is to evaluate the strength and direction of these relationships and covariances. This analysis was conducted in two phases: first, examining the correlations for panel data across all variables in all years with all counties, and second, assessing these relationships at the NUTS3 regional level, utilizing time series data for a temporal study for each of the 42 counties, across all variables and all 11 years. These two steps aim to understand macro trends within the population and gain greater content regarding the divergences amongst regions. This is especially important when considering the results of the IFIMd calculation, where there are clear divergences from region to region while maintaining a lower mean score.

The panel data confirms findings in prior literature concerning the relationship between financial inclusion and GDP. These results illustrate that financial inclusion rates in Romania, as measured by IFIMd, align with well-established trends in other global studies (Yadav, et al., 2021; Amari & Anis, 2021; Gupte, et al., 2012; World Bank, 2023). Specifically, IFIMd demonstrated the strongest correlation with GDP (r=0.74), followed by the ratio of physicians to adults (r=0.66) and urban-rural ratio (r=0.61). These panel data trends indicate a unique insight into the simultaneous change between the increasing urbanization across the 42 counties, increasing GDP, and the increased number of physicians in the country. There are several key takeaways that provide context for future exploration.

Figure 7. Correlation Testing: Panel Data, NUTS3 Romania, 2011-2021

	IFIMd	*SDT1*	*LifEx*	*URR*	*PENS*	*GDP*	*Mean Age*	*NMIG*	*OADR*	*PhysRat*	*UMEMP*
IFIMd	1										
SDT1	0.59	1									
LifEx	0.53	0.61	1								
URR	0.61	0.33	0.42	1							
PENS	-0.02	0.27	-0.06	-0.32	1						
GDP	0.74	0.60	0.60	0.64	-0.08	1					
MAGE	0.28	0.50	0.30	0.18	0.30	0.33	1				
NMIG	0.04	0.06	0.34	0.05	-0.03	0.01	-0.11	1			
OADR	0.12	0.42	0.22	-0.06	0.61	0.11	0.84	-0.08	1		
PhysRat	0.66	0.31	0.45	0.57	-0.22	0.67	0.10	0.06	-0.09	1	
UMEMPL	-0.45	-0.44	-0.42	-0.27	0.15	-0.65	-0.05	-0.05	0.11	-0.45	1

(Data Source: Author's Own Calculations, Excel)

Goga and Ilie (2017) highlighted Romania's significant brain drain, a trend that is further reflected in key counties when examining the descriptive data from 2011 to 2021. Notably, between 2014 and 2017, the physician-to-adult (65+) population ratio declined in the lowest-scoring counties, mirroring a temporary outflow of healthcare professionals. However, since 2017, this ratio has steadily increased across all NUTS3 regions, suggesting a stabilization in the availability of medical practitioners. The median county-level physician ratio also dipped between 2014 and 2015, further reinforcing Goga and Ilie's (2017) findings. This shift aligns with broader demographic changes, where an aging population coexists with a gradually increasing physician ratio, indicating a potential reversal of previous trends in medical workforce migration.

The relationship between financial inclusion and net migration presents a more complex picture. While prior studies (Johnson, 2022; Goga & Ilie, 2017; Sandu, 2010) identified a strong association between net migration and financial inclusion, as measured by the Findex database, this study's panel data (2011–2021) using IFIMd suggests no significant correlation at the national level. This discrepancy implies that financial inclusion dynamics may be more regionally dependent than previously assumed. The findings indicate that different counties exhibit distinct financial behaviors and structural determinants, necessitating a more localized approach to understanding financial inclusion. Additionally, potential methodological biases in survey-based measures such as the Findex may contribute to these differences. Given that Findex surveys rely on a sample size of only 1,000 respondents per population, it becomes challenging to capture county-level nuances, unlike IFIMd, which provides a more granular assessment based on administrative and financial data. The differences in measurement technique may explain the variance in financial inclusion outcomes between the two approaches.

At the NUTS3 level, these trends become even more pronounced. Table 4 presents time-series descriptive statistics of Pearson's r coefficients for IFIMd and ten key socioeconomic variables across Romania's 42 counties. Strong correlations were observed between IFIMd and indicators such as SDT1, GDP, mean age, pension-recipients, old-age dependency ratio, and unemployment. Specifically, these variables exhibited strong correlations in 36, 36, 35, 35, 30, and 25 counties, respectively. In contrast, only two regions demonstrated a strong correlation between net migration and IFIMd, diverging from Johnson's (2022) national-level findings, which suggested a robust relationship. Notably, age-related indicators emerged as the most consistently associated factors, with 71% of NUTS3 regions displaying $r \geq 0.7$ between IFIMd and the old-age dependency ratio (OADR). This reinforces the positive link between financial inclusion and individuals aged 55 and older in Romania (Johnson, 2022). Conversely, unemployment exhibited an inverse correlation with IFIMd, where 25 of the 42 counties demonstrated a consistent negative relationship ($r \leq -0.7$). These patterns underscore the demographic and economic factors shaping financial inclusion at a regional scale, providing critical insights for policymakers aiming to address disparities in banking access across Romania.

Table 4 Descriptive Statistics, Pearson's r Coefficients by Variable, IFIMd, NUTS3 Regions, Romania, 2011-2021

	SDT1	*LifEx*	*URR*	*PENS*	*GDP*	*Mean Age*	*NMIG*	*OADR*	*PhysRat*	*UMEMP*
Mean	0.78	0.56	-0.26	0.79	0.79	0.82	-0.25	0.67	0.52	-0.66
Min	-0.90	-0.90	-0.98	-0.97	-0.89	-0.37	-0.71	-0.72	-0.94	-0.97
Max	0.97	0.98	0.98	0.99	0.97	0.98	0.76	0.98	0.99	0.91
Median	0.93	0.68	-0.74	0.94	0.91	0.92	-0.32	0.85	0.81	-0.77
StdDev	0.39	0.38	0.81	0.41	0.39	0.29	0.36	0.42	0.55	0.36
Count	42	42	42	42	42	42	42	42	42	42

(Source: Author's Own Calculations, Excel)

These correlation tests underscore the relationships between IFIMd and key demographic variables, such as the old-age dependency ratio (OADR) and mean age. While these associations offer valuable initial insights, they represent only a surface-level understanding of the factors shaping financial inclusion. To move beyond these correlations, further analysis through time-series models and stepwise regression is necessary to uncover deeper patterns and more precise predictors of IFIMd.

To strengthen the assessment of predictability and causality, univariate and multivariate regression approaches were employed alongside hierarchical

clustering techniques to examine the geographic distribution of financial inclusion trends. This multifaceted approach allows for a more comprehensive evaluation of how IFIMd varies across counties and how spatial patterns influence financial accessibility. The regional distribution of Pearson's correlation coefficients, illustrating the strength of these relationships at the NUTS3 level, is provided in *Appendix 5*.

Holt's Exponential Smoothing and Linear Regression Compared. Holt's exponential smoothing with a trend and simple linear regression were employed to assess the predictability of IFIMd as a univariate, both at the panel level and across the time-series data for each of Romania's 42 counties. These methods were chosen to evaluate whether financial inclusion follows a more structured trend or if relationships detected through linear regression risk overfitting the data. The comparison aimed to determine which approach provides a more reliable forecast of IFIMd changes over time.

The findings indicate that Holt's exponential smoothing generally outperforms linear regression across all NUTS3 regions, particularly in capturing underlying trends in IFIMd. As detailed in **Table 5**, Holt's method demonstrated stronger predictive consistency, though linear regression exhibited a slightly lower mean absolute deviation (MAD), suggesting greater precision in certain cases (**Appendix 5**). The average MAD for Holt's method was 0.025, compared to 0.022 for linear regression, reinforcing the high predictability of IFIMd's progression. However, while linear regression offers a tighter fit, Holt's method provides a more adaptive representation of financial inclusion trends. The key distinction lies in the level component of Holt's method, which, while effective in modeling long-term movement, tends to respond more slowly to changes, occasionally leading to slight overestimation. These results highlight the structured yet gradual nature of IFIMd's evolution, emphasizing the value of incorporating trend-based approaches for understanding financial inclusion dynamics at the subnational level.

Table 5. Descriptive Statistics, MAD and MPE, Forecasting Methods, Romanian NUTS3 IFIMd

	Holt's		*Linear Regression*	
	MAD	*MPE*	*MAD*	*MPE*
Mean	0.025	4.8%	0.022	4.2%
Median	0.024	4.0%	0.022	4.0%
Standard D	0.010	6.2%	0.007	3.3%
Range	0.044	28.1%	0.028	14.3%
Minimum	0.006	-9.8%	0.007	-3.0%
Maximum	0.050	18.4%	0.035	11.3%
Count	42	42	42	42

(Source: Author's Own Calculations, Excel)

When $\alpha = 1.0$ for all 42 NUTS3 regions, the predictive accuracy of IFIMd improves significantly compared to linear regression, though the level function ultimately retains a linear predictive component. Holt's method yielded an average mean percentage error (MPE) of 4.8%, indicating relatively stable forecasting performance. In contrast, linear regression produced a slightly lower MPE of 4.2%, suggesting less variability and fewer instances of over- or underestimation than Holt's method. These results align with the panel data findings, reinforcing the observation that IFIMd at the NUTS3 level follows a largely linear growth trajectory, with financial inclusion steadily increasing across most regions. While Holt's method effectively captures these patterns, linear regression remains the more precise approach. These findings further highlight Romania's ongoing progress in financial inclusion, signaling a steady convergence toward broader regional and global financial norms despite the country's historical lag.

Stepwise Regression. Building upon the initial correlative and univariate analyses, stepwise regression was employed to isolate the most significant predictors of IFIMd, allowing for a more refined understanding of the factors driving financial inclusion at the county level. This method was particularly well-suited for the analysis, as Pearson's r testing revealed that while several variables exhibited moderate to strong correlations with IFIMd, only a select few demonstrated consistent predictive power. Using panel data spanning all NUTS3 regions, five independent variables emerged as the most influential contributors to the model: GDP, physician-to-adult ratio, SDT1, urban-rural ratio, and pensioner-to-elderly ratio. Collectively, these variables accounted for 68.2% of the variation in IFIMd (adjusted $R^2 = 0.6824$), underscoring their central role in shaping financial inclusion

dynamics over time. Notably, four of the five predictors displayed strong statistical significance ($p < 0.01$), reinforcing their robustness as explanatory variables. These findings highlight the interplay between economic structures, demographic patterns, and institutional factors in determining access to financial services, offering valuable insights for policymakers seeking to address regional disparities.

Table 6 Stepwise Results, Panel Data, Romania, y=IFIMd

Step	*Variable*	*Partial R2*	*Model R2*	*F-Value*	*Pr > F*	*Adj R2*
1	**GDP***	*0.5575*	*0.5575*	*565.67*	*<.0001*	
2	**PhysRat***	*0.0448*	*0.6023*	*50.51*	*<.0001*	
3	**SDT1***	*0.0507*	*0.6531*	*65.35*	*<.0001*	
4	**URR***	*0.0272*	*0.6803*	*37.98*	*<.0001*	
5	**PENS****	*0.0057*	*0.6860*	*8.05*	*0.0048*	*0.6824*

* <0.001; **<0.05

(Source: Author's own calculations, SAS)

The results of this study offer several critical insights into the determinants of financial inclusion at both the national and subnational levels. At the panel level, GDP, pension accessibility, and the physician-to-population ratio emerged as the most significant predictors of IFIMd across Romania's 42 counties. However, when analyzed at the county level, a more complex and regionally variable picture emerges. Key factors such as GDP, mean age (MAGE), net migration (NMIG), old-age dependency ratio (OADR), and physician ratio (PHYS) were identified as statistically significant predictors in more than half of the NUTS3 regions. These findings highlight the intricate interplay between economic conditions, demographic transitions, and access to essential services in shaping financial inclusion across Romania. Importantly, the fact that GDP is a significant predictor in both national and subnational analyses indicates its broad influence, but the emergence of demographic and institutional variables as dominant factors at the county level suggests that financial inclusion is not solely driven by economic prosperity but is also shaped by population dynamics and public service infrastructure.

The stepwise regression analysis further underscores the importance of geographic and temporal variability in these relationships. By systematically assessing the most influential predictors across both dimensions, this approach reinforces the necessity for a region-specific analytical framework. The heterogeneity in the significance of predictors between counties suggests that national-level policy interventions may not be sufficient in addressing

disparities in financial inclusion. Instead, more localized, targeted strategies are required to accommodate the distinct demographic, economic, and institutional realities of different regions. For instance, while GDP and physician ratio may be critical determinants in some areas, in others, factors such as the pensioner-to-elderly ratio or net migration may exert greater influence on IFIMd. This variability underscores the need for tailored interventions that reflect the specific financial behaviors and constraints of different populations.

Stepwise regression applied to county-level time-series data (Table 7) further highlights these regional differences in financial inclusion predictors. The results demonstrate that while certain counties align with broader national trends, many diverge significantly, exhibiting unique patterns in how financial inclusion evolves over time. This reinforces the importance of understanding localized dynamics within Romania's financial landscape, as counties with similar macroeconomic indicators may still experience vastly different trajectories of financial inclusion due to variations in age structure, migration patterns, and institutional accessibility. The adjusted R^2 scores (full steps and additional R^2 values provided in Appendix 6) illustrate how these predictors function differently across regions, providing empirical support for the need to move beyond national-level assessments when designing policies aimed at improving financial access.

Ultimately, these findings reinforce the necessity of disaggregated, subnational analyses to fully capture the drivers of financial inclusion. By identifying region-specific predictors, policymakers can craft more effective, targeted strategies that address the unique conditions of individual counties. Nationally aggregated statistics, while valuable for broad trend analysis, risk overlooking the localized barriers and opportunities that shape financial inclusion on the ground. As such, this study contributes to a growing body of literature emphasizing the importance of granular, place-based approaches to understanding and improving financial inclusion, ensuring that interventions are responsive to the diverse socioeconomic realities present across Romania's NUTS3 regions.

Table 7 Stepwise Results, NUTS3 Time Series Data, Romania, y=IFIMd

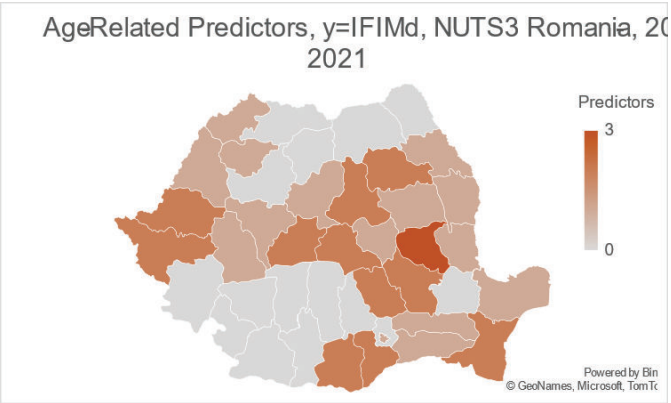
County	Step/ Variable	Adjusted r2	County	Step/ Variable	Adjusted r2	County	Step/ Variable	Adjusted r2	County	Step/ Variable	Adjusted r2
Alba	PENS	0.9446	Harghita	PENS	0.9600	Cluj	URR	0.9664	Salaj	LifEx	0.9734
Arad	PENS	0.9902		MAGE			PhysRat			URR	
	LifEx		Hunedoara	PhysRat	0.9930	Constanta	LifEx	0.9989	Satu Mare	PENS	0.8552
Arges	PhysRat	0.9800		LifEx			UNEMP			PhysRat	
	SDT1		Ialomita	PhysRat	0.9593		MAGE		Sibiu	PENS	0.9893
Bacau	URR	0.9660		LifEx		Covasna	URR	0.9995		LifEx	
	MAGE			SDT1			OADR		Suceava	GDP	0.7864
Bihor	PENS	0.9878	Iasi	PENS	0.9863		SDT1		Teleorman	PENS	0.9682
	SDT1			UNEMP			NMIG			LifEx	
Bistrita-Nasaud	GDP	0.7245	Ilfov	PhysRat	0.8643		GDP			UNEMP	
	UNEMP			NMIG		Dambovita	GDP	0.9265	Timis	PENS	0.9927
Botosani	URR	0.9203	Maramures	UNEMP	0.9249	Dolj	PhysRat	0.9824		LifEx	
Braila	URR	0.9203	Mehedinti	NMIG	0.5382		URR		Tulcea	PENS	0.9642
Brasov	PENS	0.9723	Mures	MAGE	0.8271		GDP		Vaslui	PENS	0.7916
	MAGE		Neamt	PhysRat	0.9957	Galati	PENS	0.9840	Valcea	GDP	0.6763
Bucharest	PENS	0.9439		MAGE			GDP			NMIG	
Buzau	MAGE	0.9853		URR		Giurgiu	URR	0.9928	Vrancea	MAGE	0.9908
	PENS			OADR			LifEx			PENS	
Caras-Severin	SDT1	0.8931	Olt	URR	0.8469		GDP			LifEx	
Calarasi	GDP	0.6252	Prahova	PENS	0.9619		PENS			SDT1	
	LifEx			LifEx		Gorj		0.0000			

0.05 Probability-in

(Source: Author's own calculations, SAS)

Interestingly, while the panel data highlights mean age as a strong predictor of financial inclusion, the NUTS3 time-series analysis identifies eight counties that exhibit MAGE as a predictor variable—Bacau, Brasov, Buzau, Constanta, Harghita, Mures, Neamt, and Vrancea. This serves as a minor yet illustrative example of the substantial variability across Romania's 42 counties. Notably, despite mean age being limited to 8 counties, population aging emerges as a critical factor in shaping financial inclusion at the national level, as depicted in *Figure 8*. Stepwise regression reveals that aging-related variables—such as pensioner-elderly ratio (18 counties), life expectancy (12 counties), and mean age (8 counties)—significantly predict IFIMd coefficients in 28 counties. This finding underscores the pivotal role of demographic shifts in influencing financial inclusion dynamics across the country.

Figure 8 Aging-Related Predictors of IFIMd, NUTS3 Romania, 2011-2021



(Source: Author's Own Calculation)

A central finding of this study is the role of aging in shaping the Index of Financial Inclusion at the Regional Level (IFIMd), particularly in Romania's Northwestern and Eastern border regions. Notably, while age-related variables emerge as key predictors in several areas, the southern and northern border counties exhibit minimal influence from aging-related dynamics. One striking case is Iași, which has the lowest mean age in the country at 38.9 years. However, Olt presents an interesting contrast, as its pensioner ratio has shown a consistent year-over-year decline, pointing to shifting social support structures and potentially altering financial inclusion trends in the long run. Despite the broad assumption that aging might drive financial inclusion, old age dependency did not emerge as a significant predictor of IFIMd, playing an explanatory role in only two counties—Covasna (R^2=0.0448) and Neamț (R^2=0.0168). This limited explanatory power suggests that while age structure shifts are occurring, their direct impact on financial inclusion is contingent on institutional mechanisms.

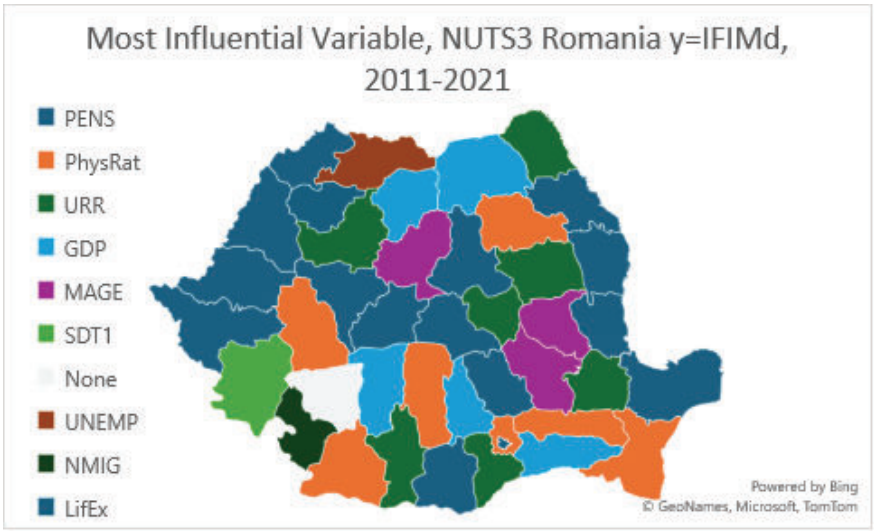
At the national level, GDP, physician-adult ratio, SDT1, urban-rural ratio, and pensioner ratio were identified as statistically significant predictors ($P < 0.05$) of IFIMd across Romania's 42 counties, as revealed through panel data analysis. However, while these relationships hold at a macro scale, the subnational variations provide a more nuanced perspective. The interplay between counties complicates the broader narrative, as panel data hides unique NUTS3 trends that shape financial inclusion. SDT1, for instance, exhibits a strong correlation with IFIMd at the national level, but the county-specific predictors reveal diverse trajectories of financial inclusion over time, highlighting both persistent disparities and evolving financial behaviors within different demographic and economic contexts.

This study's findings align with previous work by Johnson (2024), who identified the Gini coefficient of income disparity as a key driver of SDT1 at the national level, while GDP did not emerge as a statistically significant factor. A similar pattern emerges here: GDP does not serve as a primary predictor of financial inclusion at the national scale, yet at the county level, it exerts influence in nine regions, demonstrating a measurable—though uneven—impact. Its explanatory power is particularly strong in Bistrița-Năsăud (R^2=0.8960), Dâmbovița (R^2=0.9339), and Suceava (R^2=0.8077), where GDP emerges as a dominant determinant of financial inclusion. In the remaining six counties where GDP plays a role, it functions as a secondary explanatory variable, reinforcing but not driving changes in financial inclusion.

Stepwise regression analysis offers deeper insights into the specific factors that influence IFIMd over time. Among the ten variables analyzed—spanning age-related, migration-related, health-related, and economic factors—the pensioner-to-elderly ratio emerged as the strongest predictor of financial inclusion. This variable reflects the extent to which elderly populations engage with formal financial systems through pension disbursements. Of the 42 counties analyzed, 15 demonstrated that the pensioner-to-elderly ratio had the highest predictive power for financial inclusion. Another seven counties exhibited IFIMd changes most influenced by shifts in the urban-rural ratio, capturing the demographic movement from rural to urban areas, including suburban metropolitan regions. Meanwhile, six counties identified the physician-to-adult ratio as the primary predictor, suggesting a direct link between access to healthcare and financial system participation.

Collectively, three key variables—pensioner-to-elderly ratio, urban-rural ratio, and physician-to-adult ratio—accounted for the primary predictors in 28 of the 42 counties, underscoring the diverse yet interconnected pathways influencing financial inclusion across Romania. While GDP remains an important consideration, the evidence suggests that localized demographic and institutional factors—such as pension accessibility, migration trends, and healthcare infrastructure—play a more direct role in shaping financial inclusion at the county level. Figure 9 illustrates these regional variations, providing a clear spatial representation of how predictor variables influence IFIMd across the country.

Figure 9 Most Influential Predictor of IFIMd, NUTS3 Romania, 2011-2021



(Source: Author's Calculations, Excel)

Figure 9 highlights the most influential predictor of financial inclusion as determined through stepwise regression. Several key observations emerge from this visualization. Notably, the distribution of pensioners to elder ratio (PENS) extends across the country, particularly along the Moldovan and Hungarian borders. A distinct pattern also appears in Alba, Sibiu, Brașov, and Prahova, which may be influenced by the geographic constraints posed by the Carpathian Mountains and the associated settlement patterns. These findings reinforce the role of demographic factors in shaping financial inclusion at the NUTS3 level in Romania.

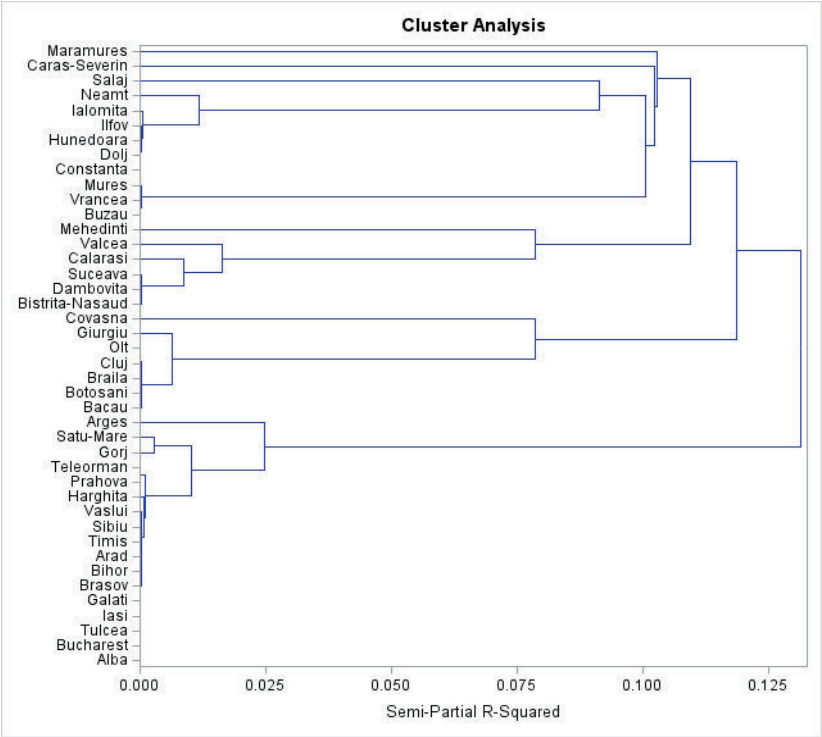
While *Figures 8* and *9* effectively illustrate the impact of individual predictors on IFIMd, the purpose of multivariate regression is to evaluate how multiple factors collectively shape financial inclusion. Stepwise regression offers a more refined approach to explaining IFIMd changes over the study period (2011–2021). To further explore these patterns, Ward's hierarchical clustering (Ward's) was applied using R^2 outputs from the stepwise regression model. The data used for this analysis is provided in *Appendix 7*, with further details in *Appendix 6*.

Ward's clustering approach sheds light on the underlying geographic structures of financial inclusion. The resulting dendrogram (*Figure 10*) presents a series of initial groupings, with the most coherent clusters emerging at a depth of five clusters (*Table 8*). These clusters are primarily associated

with key predictors: (1) urban-rural ratio (URR), (2) pension-related variables (PENS), (3) mean age (MAGE) and life expectancy (LifEx), (4) second demographic transition (SDT1), GDP, and net migration (NMIG), and (5) unemployment (UNEMP).

Table 8 further details these classifications, highlighting the extensive reach of Cluster 2, which includes 17 counties, compared to Cluster 5, which consists of a single outlier—Maramureș. This distinction is particularly noteworthy, as Maramureș stands apart as the only county where unemployment rate emerged as the primary predictor of IFIMd. Figure 10 visually reinforces this point, showing that Maramureș remains isolated from the other clusters until reaching the five-cluster depth. This deviation suggests that while pension-driven inclusion dominates in most counties, employment-related factors can still play a unique role in shaping financial participation in specific local contexts.

Figure 10. Dendrogram, Ward's Clustering, Romanian NUTS3 Regions, Predictors of IFIMd



(Source, Author's own calculation, SAS)

Table 8. Ward's Clustering, 5 Clusters, Stepwise Regression Results, y=IFIMd, NUTS3 Romania, 2011-2021

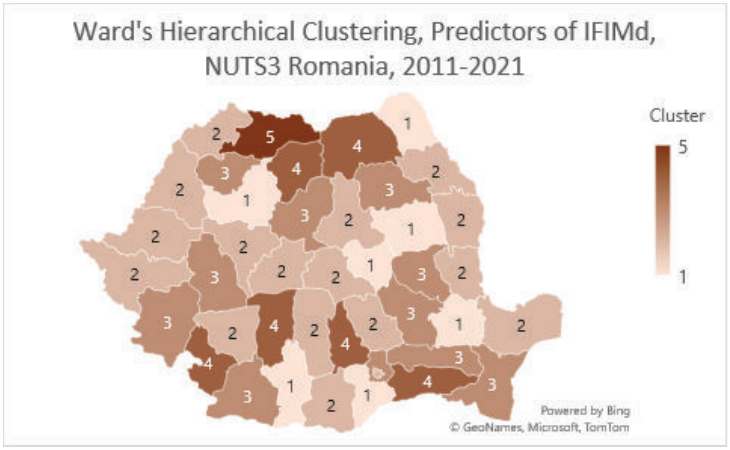
Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Bacau	*Alba*	*Buzau*	*Bistrita-Nasaud*	*Maramures*
Botosani	*Arad*	*Caras-Severin*	*Calarasi*	
Braila	*Arges*	*Constanta*	*Dambovita*	
Cluj	*Bihor*	*Dolj*	*Mehedinti*	
Covasna	*Brasov*	*Hunedoara*	*Suceava*	
Giurgiu	*Bucharest*	*Ialomita*	*Valcea*	
Olt	*Galati*	*Ilfov*		
	Gorj	*Mures*		
	Harghita	*Neamt*		
	Iasi	*Salaj*		
	Prahova	*Vrancea*		
	Satu-Mare			
	Sibiu			
	Teleorman			
	Timis			
	Tulcea			
	Vaslui			

(Author's Work based on Ward's Hierarchical Clustering, SAS, Excel)

One of the most compelling aspects of the findings is the geographic clustering of financial inclusion predictors, which underscores distinct regional patterns in the determinants of financial access. The spatial distribution reveals clear groupings, with Cluster 2 predominantly concentrated in Western Romania, while Cluster 7 is largely found in the South-East and South-West regions. Cluster 1 forms a stark division, effectively splitting the country into eastern and western halves, whereas Cluster 4 is primarily located in the eastern counties.

Among these, Cluster 2 stands out due to its widespread presence across the country, closely mirroring the patterns observed in *Figure 9*. This broad geographic dispersion suggests that a unique set of predictors—primarily related to pension systems (PENS)—play a critical role in shaping financial inclusion trends in ways that extend beyond regional economic and demographic differences. *Figure 10* illustrates the clustering results from stepwise regression of predictors, while *Figure 11* presents a geographic representation of clustering based on changes in IFIMd values between 2011 and 2021. Together, these visualizations reinforce a key finding: pensioners are a driving force behind financial inclusion in Romania.

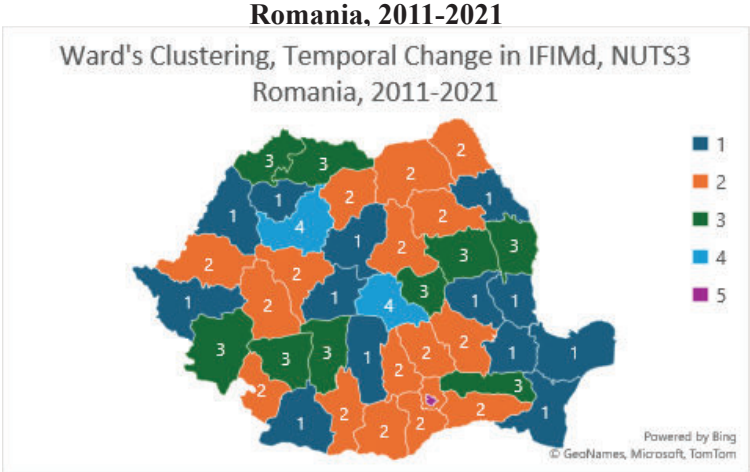
Figure 11 Ward's Clustering, Predictors of IFIMd, NUTS3 Romania, 2011-2021



(Source: Author's Calculations, Excel)

One of the most striking aspects of the comparison between financial inclusion levels and their predictors, when analyzed through Ward's hierarchical clustering, is the stark contrast between the two cartograms (*Figures 11* and *12*). While the spatial patterns of IFIMd predictors tend to follow a west-east distribution, the temporal change in IFIMd over the 11-year period exhibits a distinct north-south trend. This divergence highlights a crucial insight into financial inclusion in Romania—despite similarities in predictors, the actual levels of financial inclusion vary significantly, and the presence of a given predictor does not necessarily determine the trajectory of change. *Figure 13* further illustrates the hierarchical clustering of NUTS3 counties based on IFIMd values over time (2011–2021), while *Appendices 8* and *9* provide a detailed breakdown of the NUTS3 regions and their associated clusters for both IFIMd and its predictors.

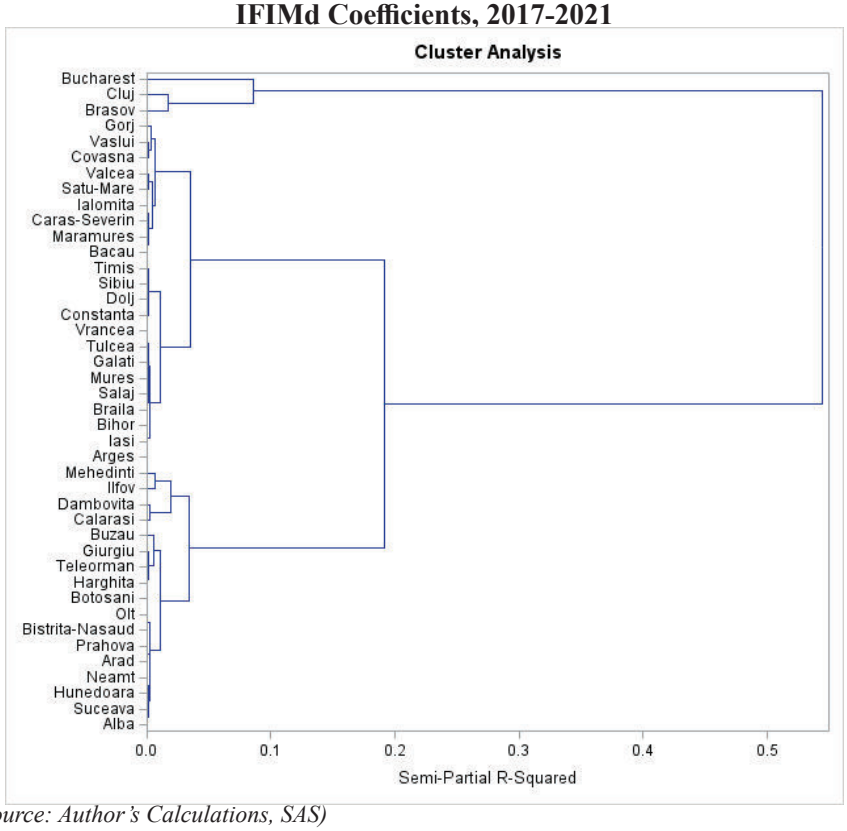
Figure 12 Ward's Clustering, Temporal Change in IFIMd, NUTS3 Romania, 2011-2021



(Source: Author's Calculation, Excel)

One of the most notable patterns emerges in Romania's westernmost region, where both IFIMd values and their key predictors align closely (Cluster 5). Meanwhile, Cluster 1 in IFIMd is geographically dispersed across multiple NUTS2 regions, with a particularly strong concentration in the southeastern counties, as illustrated in Figures 7 and 8. Similarly, the western region forms a nearly contiguous cluster, reinforcing the notion that certain areas share common characteristics. However, despite some regional consistencies, these findings underscore that Romania's NUTS2 regions are far from uniform—mirroring the disparities observed at the national level. By analyzing financial inclusion and its key determinants at the NUTS3 level, policymakers can develop more targeted strategies that address the specific needs of communities with varying levels of banking access.

Figure 13 Dendrogram, Ward's Clustering, NUTS3 Regions, IFIMd Coefficients, 2017-2021



(Source: Author's Calculations, SAS)

Despite the distinct temporal variations in IFIMd across Romania, understanding the underlying drivers of this change remains essential. Long-standing regional disparities make it unsurprising that IFIMd dynamics differ by county and do not strictly mirror the same structural shifts as their predictors. Nonetheless, while IFIMd shows no sudden surges, univariate analysis suggests it advances in a fairly steady, year-over-year manner at the NUTS3 level. Additionally, despite moderate to strong correlations with numerous independent variables, only a handful of socioeconomic and demographic factors clearly influence both inclusion and change. Among these, age-related indicators feature most prominently—life expectancy, pensioner-to-elderly ratio, mean age, and old-age dependency ratio collectively reflect the growing engagement of older cohorts in the financial system. As Romania's age structure continues to evolve, pensioners and retirees alike will increasingly turn to bank account access for government payments and remittances. In this context, the present study offers essential insights into the structure and

progression of IFIMd at the county level and establishes a valuable reference point for policymakers to broaden support for financially excluded populations.

DISCUSSION AND CONCLUSION

This research offers new insights into how demographic and socioeconomic factors influence financial inclusion in Romania at a granular NUTS3 scale. Two key indices—the SDT1 Behavioral Index and the Mahalanobis distance-based IFIMd—helped illustrate the unevenness in both demographic change and access to financial services across the nation's 42 counties. In applying the SDT1 index at a subnational level, the analysis shows that regional disparities in marriage, fertility, and divorce patterns can remain invisible if only aggregated national data are used. While earlier studies (Sobotka, 2008; Johnson, 2024) outline the second demographic transition in Romania from a broader lens, the county-specific SDT1 suggests that some regions exhibit more "modern" behaviors—like higher mean age at first marriage and more extramarital births—while others have retained traditional patterns. Understanding how local economies interact with changing social norms is thus a critical new line of inquiry.

Parallel to this, the Mahalanobis-based financial inclusion index (IFIMd), adapted from Li and Wang (2023) and Johnson (2025), maps the distribution of banking and credit usage across Romanian counties from 2011 through 2021. Five distinct IFIMd clusters emerged, revealing how certain areas—Bucharest foremost among them—maintain notably higher inclusion rates, while counties along the southern and northeastern borders consistently lag. Interestingly, counties adjacent to higher-GDP neighbors such as Hungary and Serbia tended to score higher on IFIMd, suggesting that cross-border dynamics in commerce and labor flows might shape local banking access.

Beyond establishing these macro-level patterns, the study confirms that GDP strongly correlates with IFIMd at the national level. Yet, when viewed on a county-by-county basis, demographic variables tied to aging—such as pension recipients, mean age, and life expectancy—surpass GDP in explaining local financial adoption. This is especially evident in regions with direct deposit pension transfers, which appear to draw older cohorts into formal banking. The exception is Maramureș, where the unemployment rate is the top inclusion predictor, pointing to the possibility that fewer jobless claims (and their associated benefits) similarly facilitate financial uptake. Overall, the minimal role of unemployment in most other counties highlights how local physical accessibility (e.g., ATMs and bank branches) or perceived trust in banks may be more decisive factors for financial inclusion.

Clustering results reinforce these findings by grouping counties not simply by geography but by shared socioeconomic conditions. Regions dominated by older pensioners demonstrate that government transfers alone can boost account ownership, whereas areas clustering around higher physician-to-adult ratios or strong healthcare systems could reflect broader institutional quality that fosters trust in formal financial channels. From a policy perspective, this underscores the importance of county-level or even local-level interventions: older, pension-reliant populations might benefit from direct outreach and financial literacy programs; counties lacking robust health or banking infrastructures might require integrated institutional capacity-building to instill trust and overcome structural barriers.

Persistent disparities remain, indicating that Romania continues to rank among the most financially excluded economies in Europe (Li & Wang, 2023). Although annual improvements in IFIMd show an overall upward trend, wide internal gaps persist. Bucharest stands apart at the top, while counties like Caraș-Severin struggle with low scores, possibly reflecting limited bank branch density or weaker pull from the urban core. Policies focusing only on the national average thus risk oversimplifying the complexities of local realities. Tailored measures—such as targeted deposit programs, better public transport to financial centers, or advanced digital infrastructures—would more effectively reduce persistent inequalities.

Several open questions merit further research. First, the role of informal economies across different counties remains understudied, and unreported informal labor might affect both the accuracy of unemployment data and the real extent of exclusion. Second, investigating how bank and ATM networks, as well as digital financial tools, intersect with regional demographics could clarify whether technology adoption can bridge disparities where physical infrastructure is thin. Third, exploring the inverse of inclusion—exclusion—through a lens that distinguishes between voluntary and involuntary exclusion would shed light on how certain communities remain unreachable by existing channels. Lastly, comparing Romania's internal clusters to those in other post-socialist or emerging EU contexts could illuminate whether these patterns are idiosyncratic or indicative of broader Eastern European phenomena.

Taken as a whole, this paper demonstrates that Romania's financial inclusion landscape is shaped by a mix of demographic forces, institutional structures, and cross-border effects. Integrating the SDT1 index with the Mahalanobis-based IFIMd allows for a more nuanced perspective: demographic transitions, such as delayed marriages and higher mean age at first birth, coincide with socioeconomic variations in banking access. Policymakers, therefore, can leverage these county-level findings to target financial programs more effectively, from direct deposit expansions for pensioners to digital

banking initiatives in underserved regions. Although Romania still lags behind many European counterparts in financial access, the steady IFIMd increases signal a promising trajectory: even the most financially isolated counties show incremental gains. These findings enrich ongoing discussions on narrowing regional divides and offer a conceptual roadmap for future investigations into how demographic shifts and economic forces jointly shape the adoption of formal financial services in post-socialist settings.

References

1. Allen, F., Demirgüç-Kunt, A., Klapper, L. & Peria, M. S. M., 2016. The foundations of financial inclusion: Understanding ownership and use of formal accounts. *Journal of Financial Intermediation,* pp. 1-30.
2. Amari, M. & Anis, J., 2021. Exploring the impact of socio-demographic characteristics on financial inclusion: empirical evidence from Tunisia. *International Journal of Social Economics,* pp. 1331-1346.
3. Anzoategui, D., Demirgüç-Kunt, A. & Martínez Pería, M. S., 2014. Remittance and Financial Inclusion: Evidence from El Salvidor. *World Development,* pp. 338-349.
4. Ban, C., 2012. Sovereign Debt, Austerity, and Regime Change : The Case of Nicolae Ceausescu's Romania. *East European Politics & Societies,* pp. 743-776.
4. Ben-Ner, A. & Montias, J. M., 1991. The Introduction of Markets in a Hypercentralized Economy: The Case of Romania. *Journal of Economic Perspectives,* pp. 163-170.
5. Borjas, G. J. & Bratsberg, B., 1994. *Who Leaves? The Outmigration of the Foreign-Born,* Cambridge, MA: National Bureau of Economic Research.
6. Borooah, V. K. & Iye, S., 2005. Vidya, Veda, and Varna: The Influence of Religion and Caste on Education in Rural India. *The Journal of Development Studies,* pp. 1369-1404.
7. Botezat, A. & Moraru, A., 2020. Brain Drain from Romania: What do we know so far about the Romanian medical diaspora?. *Eastern Journal of European Studies,* pp. 309-334.
8. Caselli, G., Vallin, J. & Wunsch, G., 2006. Population Models. In: *Demography Analysis and Synthesis.* Burlington: Elsevier, pp. 249-267.
9. Chowdhury, E. K. & Chowdhury, R., 2024. Role of Financial Inclusion in Human Development: Evidence from Bangladesh, India and Pakistan. *Journal of the Knowledge Economy,* pp. 3329-3354.
10. Ciobanu, M., 2007. Romania's Travails with Democracy and Accession to the European Union. *Europe-Asia Studies,* pp. 1429-1450.
11. Constantin, D., Goschin, Z. & Danciu, A., 2011. The Romanian Economy from Transition to Crisis. Retrospects and Prospects. *World Journal of Social Sciences,* pp. 155-171.
12. Cristina, I. O. M., Nicoleta, C., Cătălin, D. R. & Margareta, F., 2021. Regional development in Romania: Empirical evidence regarding the factors for measuring a prosperous and sustainable economy. *Sustainability,* 13(7), p. 3942.
13. del Carmen Dircio-Palacios-Macedo, M., Cruz-García, P., H.-T. F. & Tortosa-Ausina, E., 2023. Constructing a financial inclusion index for Mexican municipalities. *Finance Research Letters,* p. 103368.
14. Demirgüç-Kunt, A., Hu, B. & Klapper, L., 2019. *Financial Inclusion in the Europe and Central Asia Region - Recent Trends and a Research Agenda,* Washington DC: World Bank Group.
15. Demirguc-Kunt, A. & Klapper, L., 2013. *Measuring Financial Inclusion: The Global Findex Database,* Washington DS: The World Bank.

16. Demirguc-Kunt, A., Klapper, L. & Singer, D., 2017. *Financial Inclusion and Inclusive Growth - A Review of Recent Empirical Evidence,* Washington DC: World Bank Group.
17. Demirgüç-Kunt, A. et al., 2018. *The Global FINDEX Database 2017: Measuring Financial Inclusion and the Fintech Revolution,* Washington DC: The World Bank.
18. Demirgüç-Kunt, A., Klapper, L., Singer, D. & Ansar, S., 2022. *The Global Findex Database 2021: Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19,* Washington DC: The World Bank.
19. Demirgüç-Kunt, A., Klapper, L., Singer, D. & Ansar, S., 2022. *The Global Findex Database. Financial Inclusion, Digital Payment, and Resilience in the Age of COVID-19,* Washington DC: The World Bank.
20. Demirguc-Kunt, A., Klapper, L., Singer, D. & Oudheusden, P. V., 2015. *The Global Findex Database 2014: Measuring Financial Inclusion around the World,* Washington DC: The World Bank.
21. Emmert, F. & Petrovi, S., 2014. The Past, Present, and Future of EU Enlargement. *Fordham International Law Journal.*
22. Fernández-Olit, B., Paredes-Gázquez, J. D. & de la Cuesta-González, M., 2021. Are Social and Financial Exclusion Two Sides of the Same Coin? An Analysis of the Financial Integration of Vulnerable People. *Social Indicators Research,* Issue 135, pp. 245-268.
23. Gavriloaia, G.-C., 2020. The impact of the brain drain process on Romania - possible solutions in order to increase resilience. *CES Working Papers.*
24. Goga, C. I. & Ilie, V., 2017. From "brain drain" to "brain gain". Where does Romania Stand?. *Revista de Stiinte Politice,* Issue 54, pp. 90-103.
25. Grigorescu, I., Mitrică, B., Mocanu, I. & Ticană, N., 2012. Urban sprawl and residential development in the Romanian Metropolitan Areas. *Romanian Journal of Geography,* pp. 43-59.
26. Gupte, R., Venkataramani, B. & Gupta, D., 2012. Computation of financial inclusion index for India. *Procedia - Social and Behavioral Sciences,* pp. 133-149.
27. Halbac-Cotoara-Zamfir, R., Ferreira, C. S. S. & Salvati, L., 2021. Long-term urbanization dynamics and the evolution of green/blue areas in eastern europe: Insights from Romania. *Sustainability,* 13(24), p. 14068.
28. Harris, J. R. & Todaro, M. P., 1970. Migration, Unemployment and Development: A Two-Sector Analysis. *The American Economic Review,* 60(1), pp. 126-142.
29. Immurana, M., Iddrisu, A. A., Boachie, M. K. & Dalaba, M. A., 2021. Financial inclusion and population health in Africa. *Journal of Sustainable Finance & Investment,* pp. 1-16.
30. Johnson, S., 2022. Relationship Between Net Migration and Financial Inclusion in Romania. *Forum Geografic,* pp. 5-22.
31. Johnson, S., 2024. From Economic Turbulence to Demographic Change: Tracing the Pathways of the Second Demographic Transition in Post-Socialist Contexts. *AUC Geographica,* pp. 1-22.
32. Johnson, S., Forthcoming. Unveiling Regional Disparities: A Comparative Analysis of Financial Inclusion Using Mahalanobis and Euclidean Distance Indices in the case of Romania. *Eastern European Economics.*
33. Kandari, P., Bahuguna, U. & Salgotra, A. K., 2021. Socio-Economic and Demographic Determinants of Financial Inclusion in Underdeveloped Regions: A Case Study in India. *Journal of Asian Finance, Economics and Business,* pp. 1045-1052.
34. Kara, A., Zhou, H. & Zhou, Y., 2021. Achieving the United Nations' sustainable development goals through financial inclusion: A systematic literature review of access to finance across the globe. *International Review of Financial Analysis,* p. 101833.

35. Khan, I. Z. et al., 2020. *Financial Inclusion in Romania : Issues and Opportunities,* Washington D.C.: World Bank Group.
36. Koku, P. S., 2015. Financial exclusion of the poor: a literature review. *International Journal of Bank Marketing,* 33(5), pp. 654-668.
37. Li, D.-D. & Wang, Z.-X., 2023. Measurement Methods for Relative Index of Financial Inclusion. *Applied Economics Letters,* pp. 827-833.
38. Li, Q., Chen, L. & Hao, T., 2024. Unlocking Urbanization: The symbiotic relationship between inclusive finance and urban development in China. *Heliyon,* p. e27457.
39. Maria, H. A., 2019. *Money Saving - The Chance for a Happy Old Age.* Warsaw, Faculty of Management University of Warsaw, pp. 281-291.
40. Muntele, I., 2024. The Evolution of the Structure by Age Groups and the Ageing of the Population of Romania between 1992-2021. *Analele Universităţii din Oradea, Seria Geografie,* 34(2), pp. 123-138.
41. Muntele, I., Istrate, M., Bănică, A. & Horea-Serban, R.-I., 2020. Trends in Life Expectancy in Romania between 1990 and 2018. A Territorial Analysis of Its Determinants. *Sustainability.*
42. Nanda, K. & Kaur, M., 2016. Financial Inclusion and Human Development: A Cross-country Evidence. *Management and Labour Studies,* pp. 127-153.
43. Nica, E. et al., 2023. The impact of financial development, health expenditure, CO2 emissions, institutional quality, and energy Mix on life expectancy in Eastern Europe: CS-ARDL and quantile regression Approaches. *Heliyon,* p. e21084.
44. Otovescu, C. & Otovescu, A., 2019. The Depopulation of Romania–Is It an Irreversible Process?. *Revista de Cercetare si Interventie Sociala,* Volume 65, pp. 370-388.
45. Roman, M. D., Toma, G. C. & Tuchiluş, G., 2018. Efficiency of pension systems in the EU countries. *Romanian Journal of Economic Forecasting,* 21(4), pp. 161-173.
46. Russu, C. & Ciuiu, D., 2020. *Convergence of the most/least developed counties/ NUTS3 regions in Romania/Euro area.* s.l., s.n., pp. 130-146.
47. Rychtarikova, J., 1999. Is Eastern Europe Experiencing a Second Demographic Transition?. *Geographica,* pp. 19-44.
48. Sakyi-Nyarko, C., Ahmad, A. H. & Green, C. J., 2022. The role of financial inclusion in improving household well-being. *Journal of International Development,* 34(8), pp. 1606-1632.
49. Sandu, D., 2010. 13 Modernising Romanian Society Through Temporary Work Abroad. *A Continent Moving West?,* p. 271.
50. Sarma, M., 2008. *Index of Financial Inclusion,* New Delhi: Indian Council for Research on International Economic Relations.
51. Sarma, M., 2012. *Index of Financial Inclusion–A measure of financial sector inclusiveness,* Delhi: School of International Studies Working Paper Jawaharlal Nehru University.
52. Sen, A. & Laha, A., 2021. Financial Inclusion and Quality of Life: Empirical Evidences from Indian States with Special Reference to West Bengal. *Management and Labour Studies,* 47(2), pp. 139-164.
53. Sobotka, T., 2008. The diverse faces of the Second Demographic Transition in Europe. *Demographic Research,* pp. 171-224.
54. Stănescu, I., 2018. Quality of life in Romania 1918–2018: An overview. *Calitatea vieţii,* 29(2), pp. 107-144.
55. Stănescu, I., 2021. Living Conditions in Rural Areas in Romania From 1990 to 2020. *Calitatea vieţii,* 32(2), pp. 1-22.
56. Stawicki, M. & Wojewódzka-Wiewiórska, A., 2023. Regional Differentiation of GDP at the NUTS-3 Level in Selected European Countries after their Accession to the European Union. *Ekonomika regiona / Economy of regions,* 19(4), pp. 1224-1236.

57. Tarsem, L., 2018. Impact of financial inclusion on poverty alleviation through cooperative banks. *International Journal of Social Economics,* pp. 807-827.
58. Thu, N. H. & Dao, L. K. O., 2022. How do socio-demographic characteristics influence the probability of financial inclusion? evidence from a transitional economy. *Dalat University Journal of Science,* pp. 42-59.
59. Todaro, M. P. & Smith, S. C., 2015. Urbanization and Rural-Urban Migration. In: *Economic Development 12e.* s.l.:Pearson, pp. 330-374.
60. Toma, G. C. & Tuchilus, G., 2019. Pensioners Versus Employees in Romania: A Regional Study. *European Scientific Journal,* 15(19), p. 1857 – 7881.
61. Tran, H. S., Nguyen, T. L. & Nguyen, V. K., 2020. Mobile Money, Financial Inclusion and Digital Payment. *International Journal of Financial Research,* pp. 417-424.
62. van de Kaa, D. & Lesthaeghe, R., 1987. Europe's Second Demographic Transition. *Population Bulletin No. 42.*
63. Vasile, V. & Dobre, A. M., 2015. Overview of Demographic Evolution in Romania. *Romanian Statistical Review,* Volume 4, pp. 27-45.
64. World Bank, 2023. *Financial inclusion overview.* [Online]
Available at: https://www.worldbank.org/en/topic/financialinclusion/overview
65. World Bank, 2024. *Urban population growth (annual %) - Romania, European Union.* [Online]
Available at: https://data.worldbank.org/indicator/SP.URB.GROW?end=2023&locations=RO-EU&start=1990
[Accessed 2024].
66. Yadav, V., Pratap, S. B. & Nirmala, V., 2021. Multidimensional financial inclusion index for Indian states. *Journal of Public Affairs,* p. e2238.
67. Zaidi, B. & Morgan, S. P., 2017. The Second Demographic Transition Theory: A Review and Appraisal. *Annual Review of Sociology,* pp. 473-492.

Appendices

Appendix 1. IFIMd Value, NUTS3 Romania, 2011-2021

IFIMd Values, NUTS3, Romania, 2011-2021											
County	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
Alba	0.24	0.24	0.24	0.25	0.28	0.30	0.30	0.30	0.33	0.37	0.37
Arad	0.25	0.25	0.25	0.27	0.28	0.32	0.33	0.34	0.35	0.39	0.42
Arges	0.26	0.27	0.28	0.29	0.34	0.37	0.37	0.38	0.40	0.46	0.50
Bacau	0.28	0.29	0.31	0.31	0.31	0.34	0.35	0.35	0.36	0.39	0.45
Bihor	0.27	0.28	0.30	0.32	0.35	0.37	0.39	0.39	0.41	0.47	0.50
Bistrita-Nas.	0.25	0.22	0.24	0.24	0.25	0.28	0.31	0.31	0.32	0.38	0.41
Botosani	0.24	0.25	0.23	0.25	0.27	0.28	0.28	0.28	0.28	0.32	0.35
Braila	0.27	0.28	0.29	0.32	0.33	0.35	0.35	0.36	0.36	0.42	0.46
Brasov	0.49	0.48	0.48	0.48	0.49	0.52	0.55	0.54	0.56	0.60	0.63
Bucharest	0.55	0.55	0.57	0.59	0.67	0.73	0.73	0.73	0.75	0.83	0.87
Buzau	0.17	0.17	0.18	0.22	0.23	0.26	0.26	0.28	0.30	0.34	0.37
Caras-Sev.	0.29	0.30	0.28	0.31	0.32	0.32	0.34	0.35	0.36	0.44	0.43
Calarasi	0.21	0.17	0.18	0.18	0.18	0.18	0.18	0.20	0.19	0.21	0.23
Cluj	0.35	0.35	0.37	0.38	0.42	0.46	0.47	0.49	0.50	0.56	0.61
Constanta	0.30	0.31	0.31	0.33	0.35	0.37	0.38	0.39	0.43	0.51	0.53
Covasna	0.24	0.29	0.28	0.29	0.30	0.31	0.31	0.32	0.33	0.36	0.37
Dambovita	0.19	0.19	0.19	0.20	0.21	0.21	0.22	0.23	0.24	0.28	0.29
Dolj	0.29	0.31	0.30	0.32	0.37	0.41	0.41	0.39	0.41	0.44	0.50
Galati	0.26	0.27	0.25	0.30	0.33	0.36	0.37	0.36	0.39	0.45	0.47
Giurgiu	0.24	0.25	0.21	0.23	0.24	0.25	0.26	0.22	0.29	0.26	0.35
Gorj	0.31	0.30	0.32	0.35	0.34	0.32	0.31	0.29	0.29	0.32	0.32
Harghita	0.26	0.23	0.23	0.23	0.25	0.27	0.28	0.28	0.29	0.33	0.36
Hunedoara	0.26	0.23	0.26	0.26	0.28	0.30	0.30	0.31	0.31	0.33	0.35
Ialomita	0.32	0.32	0.30	0.31	0.32	0.34	0.35	0.36	0.35	0.41	0.42
Iasi	0.28	0.29	0.29	0.30	0.33	0.36	0.37	0.39	0.40	0.45	0.49
Ilfov	0.25	0.29	0.25	0.23	0.23	0.22	0.22	0.21	0.21	0.19	0.19
Maramures	0.27	0.27	0.29	0.29	0.31	0.33	0.33	0.34	0.34	0.39	0.40
Mehedinti	0.34	0.33	0.29	0.26	0.27	0.26	0.24	0.24	0.28	0.28	0.30
Mures	0.26	0.26	0.28	0.30	0.34	0.42	0.36	0.35	0.38	0.41	0.44
Neamt	0.28	0.26	0.25	0.26	0.28	0.30	0.29	0.28	0.30	0.33	0.35
Olt	0.25	0.23	0.22	0.22	0.24	0.28	0.28	0.29	0.30	0.38	0.39
Prahova	0.28	0.25	0.25	0.28	0.27	0.30	0.31	0.31	0.33	0.39	0.40
Salaj	0.26	0.27	0.28	0.32	0.35	0.36	0.37	0.38	0.37	0.40	0.44
Satu Mare	0.34	0.34	0.32	0.32	0.35	0.34	0.33	0.34	0.35	0.38	0.36
Sibiu	0.27	0.28	0.31	0.33	0.34	0.41	0.40	0.42	0.42	0.50	0.55
Suceava	0.26	0.25	0.25	0.25	0.28	0.29	0.31	0.31	0.31	0.36	0.41
Teleorman	0.23	0.22	0.22	0.23	0.24	0.27	0.27	0.27	0.27	0.32	0.34
Timis	0.30	0.28	0.29	0.33	0.36	0.40	0.43	0.43	0.46	0.51	0.55
Tulcea	0.26	0.28	0.27	0.30	0.33	0.35	0.38	0.39	0.38	0.44	0.46
Vaslui	0.29	0.28	0.29	0.30	0.30	0.31	0.31	0.31	0.30	0.34	0.38
Valcea	0.35	0.33	0.33	0.34	0.33	0.30	0.31	0.33	0.35	0.40	0.44
Vrancea	0.26	0.24	0.26	0.29	0.32	0.37	0.39	0.37	0.37	0.42	0.46

(Source: Author's Calculation, SAS)

Appendix 2. Non-Standardized PUA Inputs

	2011			2012			2013			2014			2015			2016		
	U	A	P	U	A	P	U	A	P	U	A	P	U	A	P	U	A	P
Alba	0.59	2.86	490.24	0.59	3.22	499.88	0.66	4.65	480.54	0.69	5.36	493.79	0.76	5.37	498.16	0.88	9.33	497.46
Arad	0.56	20.21	521.66	0.59	20.78	517.14	0.60	20.78	515.24	0.69	21.07	508.52	0.77	21.09	528.01	0.89	24.03	531.88
Arges	0.72	4.60	501.05	0.78	4.81	461.43	0.83	5.64	458.63	0.89	8.76	477.65	1.04	8.78	480.02	1.15	11.32	478.00
Bacau	0.83	0.78	320.40	0.86	1.37	336.96	0.97	1.94	315.53	0.95	2.13	315.30	0.97	2.71	320.86	1.09	2.71	310.36
Bihor	0.77	1.59	413.99	0.80	2.04	409.06	0.89	2.72	408.34	0.94	3.40	434.48	1.06	3.40	434.04	1.17	7.06	436.99
Bistrița-Nă	0.71	1.81	405.83	0.69	9.01	404.14	0.76	10.34	380.40	0.74	10.32	385.19	0.78	12.11	400.20	0.87	12.05	400.86
Botoșani	0.71	0.33	286.39	0.75	0.33	264.83	0.70	0.33	285.45	0.74	0.66	272.06	0.86	2.30	267.21	0.88	2.30	265.59
Buzău	0.87	6.77	1205.70	0.87	7.07	1196.27	0.83	8.80	1213.10	0.91	10.27	1168.06	1.00	14.05	1155.71	1.15	14.43	1189.33
Braila	0.83	4.00	365.42	0.86	4.03	346.17	0.92	6.26	355.14	1.02	6.30	340.85	1.08	7.85	338.32	1.13	8.67	345.44
Brașov	0.54	3.25	254.03	0.55	3.90	260.74	0.56	3.91	271.37	0.68	3.70	258.52	0.74	8.30	265.28	0.83	13.37	250.95
Călărași	0.91	6.66	397.46	0.95	7.12	420.04	0.88	7.59	394.65	0.96	8.07	393.88	0.99	8.56	412.95	1.02	12.69	440.22
Caraș-Seve	0.60	0.40	285.13	0.50	0.81	277.32	0.52	1.22	299.05	0.54	1.23	279.13	0.55	2.48	278.23	0.55	2.91	275.53
Cluj	0.83	0.00	610.70	0.80	33.86	654.93	0.87	33.74	666.62	1.00	33.86	689.71	1.23	33.54	678.80	1.38	34.11	728.86
Constanța	0.65	4.32	651.78	0.68	4.86	662.36	0.66	8.10	720.94	0.71	9.36	769.14	0.88	13.18	698.74	0.99	13.92	683.46
Covasna	0.75	9.37	393.38	0.79	23.83	373.95	0.74	24.50	378.51	0.81	24.54	373.23	0.84	24.00	379.51	0.89	24.69	387.92
Dâmbovița	0.48	0.53	413.21	0.47	0.53	409.73	0.48	0.53	390.70	0.53	1.06	385.35	0.55	1.06	404.93	0.63	6.12	399.53
Dolj	0.80	0.19	414.26	0.87	1.17	399.67	0.88	1.94	397.44	1.03	9.13	407.00	1.24	15.20	399.86	1.38	16.80	388.49
Galați	0.81	14.22	351.25	0.88	14.35	338.54	0.79	14.05	349.15	0.98	13.76	345.80	1.10	13.77	323.94	1.23	13.11	310.70
Giurgiu	0.78	4.99	288.38	0.80	5.48	253.10	0.67	5.48	292.89	0.74	6.49	272.80	0.78	6.49	279.24	0.84	6.50	262.85
Gorj	0.79	0.00	507.61	0.70	0.76	550.80	0.82	0.76	536.71	1.00	1.13	477.33	0.91	1.13	507.13	0.89	4.14	506.37
Harghita	0.72	0.00	396.70	0.66	1.71	385.93	0.69	3.84	383.16	0.72	8.94	379.87	0.80	9.80	369.92	0.86	15.80	375.93
Hunedoara	0.81	12.82	378.20	0.66	13.95	450.05	0.78	19.02	376.14	0.76	19.35	384.04	0.84	21.14	359.52	0.88	21.56	369.67
Ialomița	0.95	0.00	356.42	0.94	0.49	349.94	0.90	3.42	366.97	0.96	4.40	346.38	0.98	5.40	352.47	1.09	5.43	350.28
Iași	0.75	22.38	409.31	0.76	24.43	415.24	0.73	24.44	423.88	0.83	24.42	428.23	1.04	23.89	417.45	1.17	23.51	413.21
Ilfov	0.00	1.09	751.71	0.00	1.76	861.89	0.00	1.71	748.52	0.00	2.65	706.37	0.00	4.18	719.98	0.00	10.83	687.48
Maramureș	0.80	0.82	350.05	0.77	0.81	362.55	0.84	1.08	355.96	0.86	1.35	367.51	0.90	2.43	383.30	1.01	3.78	369.37
Mehedinți	1.07	0.00	294.09	1.00	0.47	296.19	0.88	1.89	286.47	0.84	15.62	284.01	0.84	15.73	288.37	0.84	15.79	293.99
Mureș	0.80	14.28	409.54	0.76	17.15	429.43	0.79	20.47	420.70	0.88	21.48	437.17	1.05	22.73	440.50	1.40	22.53	422.77
Neamț	0.83	0.00	268.90	0.80	1.22	275.52	0.77	1.96	281.03	0.81	2.69	274.63	0.90	4.89	282.88	0.97	6.36	282.77
Olt	0.76	3.26	296.62	0.71	4.16	310.61	0.69	4.46	313.08	0.70	5.66	315.76	0.78	5.97	304.90	0.90	6.89	305.13
Prahova	0.51	2.65	659.01	0.64	5.80	543.55	0.59	6.98	587.46	0.52	11.84	734.17	0.69	14.11	608.09	0.76	18.27	635.67
Sălaj	0.86	0.00	566.98	0.82	0.00	571.42	0.77	1.74	591.12	0.76	1.74	591.35	0.88	1.74	590.08	0.87	2.32	589.49
Satu Mare	0.76	0.00	234.25	0.83	0.00	242.70	0.86	0.00	253.58	0.96	0.00	262.54	1.09	0.00	260.50	1.15	0.73	259.34
Sibiu	0.70	16.44	568.25	0.77	19.47	552.71	0.90	19.44	537.94	1.00	19.10	532.57	1.03	19.11	558.47	1.27	20.63	586.29
Suceava	0.81	2.51	320.64	0.77	4.57	316.49	0.78	4.54	320.38	0.78	4.51	321.25	0.87	5.31	310.83	0.93	5.68	311.85
Teleorman	0.71	3.31	281.66	0.68	3.66	303.19	0.68	4.02	300.73	0.71	4.39	309.07	0.74	4.44	307.08	0.87	4.83	290.29
Timiș	0.61	12.20	739.33	0.62	16.35	686.31	0.68	16.81	698.15	0.85	18.73	684.41	0.95	18.52	724.52	1.10	20.87	726.84
Tulcea	0.77	4.44	419.34	0.86	4.47	372.58	0.83	4.50	387.92	0.93	4.53	369.12	1.03	5.12	371.49	1.10	6.28	374.76
Vâlcea	0.89	0.34	247.61	0.86	0.67	268.77	0.88	1.01	256.37	0.93	3.01	254.26	0.98	3.68	250.13	1.02	5.01	256.83
Vaslui	1.01	0.00	388.57	0.97	0.67	392.03	0.99	1.65	376.80	1.03	1.59	345.98	1.00	2.51	362.57	0.93	3.71	344.54
Vrancea	0.79	2.53	292.59	0.75	2.53	306.96	0.80	2.54	312.46	0.87	2.54	306.82	0.99	2.91	311.49	1.17	2.92	308.46
Bucharest	1.38	12.62	1020.82	1.43	20.13	997.36	1.49	25.85	1039.23	1.55	28.91	1045.18	1.84	33.22	1102.80	2.11	36.72	1082.75

	2017			2018			2019			2020			2021		
	U	A	P	U	A	P	U	A	P	U	A	P	U	A	P
Alba	0.90	11.17	506.65	0.89	11.23	511.36	0.98	11.65	496.91	1.13	12.43	499.57	1.17	13.26	504.69
Arad	0.94	24.98	531.12	0.98	24.74	523.21	1.04	24.53	527.44	1.25	24.35	500.05	1.40	23.33	477.59
Arges	1.16	12.42	490.08	1.19	12.90	485.01	1.28	13.20	479.51	1.50	13.73	470.80	1.67	13.45	456.28
Bacau	1.12	3.28	301.58	1.10	3.87	305.89	1.15	4.06	311.82	1.29	4.65	326.99	1.52	4.86	310.93
Bihor	1.24	8.90	439.11	1.26	8.92	435.13	1.31	8.71	441.79	1.54	9.64	446.05	1.71	9.68	436.02
Bistrița-Nă	0.99	12.05	386.84	1.02	12.03	375.71	1.06	12.03	380.80	1.24	12.91	378.32	1.37	12.97	362.17
Botoșani	0.89	3.27	269.94	0.90	4.22	270.67	0.92	5.75	268.08	1.06	6.02	266.38	1.18	6.65	252.60
Buzău	1.22	18.08	1242.85	1.20	18.54	1265.30	1.24	18.74	1304.96	1.42	18.92	1289.50	1.56	19.14	1326.96
Braila	1.14	9.15	354.92	1.15	9.23	354.65	1.18	8.94	358.67	1.42	9.41	347.37	1.58	13.49	350.93
Brașov	0.86	11.45	252.29	0.91	12.59	260.07	0.98	17.47	259.58	1.16	16.82	264.52	1.25	16.00	268.34
Călărași	1.09	12.34	423.01	1.15	12.45	398.39	1.16	12.57	403.44	1.45	12.67	404.52	1.44	12.79	393.35
Caraș-Seve	0.58	3.77	271.03	0.63	4.21	274.31	0.61	5.08	274.01	0.69	5.53	254.14	0.73	6.43	285.30
Cluj	1.40	34.09	771.33	1.47	33.16	753.40	1.51	32.57	773.30	1.70	31.26	800.89	1.87	29.60	788.73
Constanța	1.11	15.24	639.08	1.17	15.68	621.44	1.34	15.41	577.70	1.62	15.51	556.69	1.72	16.19	588.15
Covasna	0.89	25.41	392.74	0.92	26.13	398.76	0.97	26.89	400.61	1.12	27.50	403.12	1.20	27.57	401.68
Dâmbovița	0.69	7.47	393.68	0.72	7.49	377.50	0.76	8.58	384.23	0.88	9.16	392.97	0.95	10.06	395.33
Dolj	1.36	16.47	395.07	1.32	16.53	405.25	1.37	17.63	419.72	1.52	18.13	421.64	1.76	17.71	410.85
Galați	1.24	14.46	310.58	1.21	13.35	316.12	1.33	13.96	316.82	1.60	13.92	300.37	1.65	13.67	321.80
Giurgiu	0.86	6.54	266.64	0.67	6.57	337.12	0.92	6.13	276.55	0.84	6.20	343.39	1.15	6.30	279.70
Gorj	0.88	4.52	501.34	0.78	6.03	506.31	0.79	6.06	536.92	0.93	7.22	505.81	1.01	11.48	469.92
Harghita	0.86	19.71	394.14	0.86	19.79	395.05	0.88	20.72	407.27	1.03	21.19	408.90	1.19	20.83	398.63
Hunedoara	0.88	22.60	370.05	0.89	24.24	369.36	0.90	24.46	369.20	1.04	22.89	373.77	1.15	22.89	366.67
Ialomița	1.14	8.96	355.09	1.17	9.54	350.96	1.12	9.65	359.43	1.35	10.26	350.56	1.40	10.39	378.86
Iași	1.21	22.76	436.62	1.28	22.30	422.42	1.32	19.82	430.11	1.54	18.53	433.98	1.69	17.47	432.21
Ilfov	0.00	11.05	688.56	0.00	10.71	671.81	0.00	10.96	648.22	0.00	10.39	622.10	0.00	9.82	631.99
Maramureș	1.03	4.05	377.77	1.05	4.86	370.52	1.09	6.21	379.83	1.24	6.49	380.93	1.31	6.51	380.58
Mehedinți	0.74	15.86	325.76	0.74	15.95	337.13	0.88	16.08	293.78	0.91	16.18	331.26	0.97	16.38	338.66
Mureș	1.15	22.82	415.63	1.11	23.11	433.84	1.26	22.70	424.20	1.36	23.70	451.42	1.49	22.36	468.66
Neamț	0.95	6.86	284.49	0.90	7.13	286.94	0.96	5.90	282.13	1.09	6.16	291.89	1.17	6.19	292.44
Olt	0.90	8.45	319.10	0.93	10.03	324.97	1.00	12.26	325.08	1.29	12.36	300.83	1.35	14.40	330.59
Prahova	0.87	18.60	551.74	0.92	19.09	549.18	1.00	19.24	517.36	1.21	21.96	506.52	1.25	23.05	531.75
Sălaj	0.84	3.49	607.14	0.89	3.50	594.57	0.87	3.52	605.23	1.05	7.06	614.81	1.02	9.49	610.07
Satu Mare	1.19	0.73	268.51	1.22	1.84	263.80	1.19	2.95	264.02	1.35	2.95	254.91	1.47	3.33	277.64
Sibiu	1.24	20.60	570.68	1.31	22.10	557.30	1.35	22.37	559.17	1.65	22.66	554.96	1.88	21.16	530.87
Suceava	1.00	5.84	309.09	0.99	5.94	306.52	1.00	6.09	314.27	1.16	6.46	321.42	1.38	6.66	300.49
Teleorman	0.86	5.95	291.45	0.87	6.39	292.39	0.85	6.50	311.08	1.02	6.60	309.31	1.10	7.48	320.08
Timiș	1.25	24.84	682.21	1.29	25.87	685.53	1.38	26.19	695.47	1.59	25.80	684.54	1.75	24.92	688.11
Tulcea	1.17	0.00	379.39	1.18	0.00	387.25	1.24	7.63	374.04	1.44	7.68	385.04	1.50	7.77	406.58
Vâlcea	1.05	9.70	252.72	1.04	10.37	253.73	1.00	10.40	244.59	1.18	10.07	246.82	1.30	10.43	241.43
Vaslui	0.98	3.94	338.09	1.01	3.85	350.97	1.11	4.65	344.75	1.31	4.58	342.95	1.44	4.86	335.38
Vrancea	1.225	0.00	303.42	1.20	5.15	309.44	1.19	5.18	317.83	1.42	6.31	313.64	1.60	6.35	309.04
Bucharest	2.16	38.62	1066.93	2.22	39.39	1039.17	2.23	37.23	1068.57	2.50	36.00	1074.83	2.84	33.80	1076.43

(Author's Calculations, Data Source: U: BNR (2024); A: Open Street Maps (2024); P: BNR (2024))

Appendix 3. SDT1 Index Values, NUTS3 Romania, 2011-2021

County	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Alba	2.97	3.06	3.23	3.49	3.99	3.98	4.38	4.48	4.61	5.62	5.25
Arad	2.93	2.97	2.96	3.10	3.30	3.57	3.96	4.09	4.20	5.14	4.75
Arges	3.02	3.04	3.13	3.33	3.53	3.79	4.33	4.31	4.73	5.34	5.00
Bacau	2.85	3.06	3.34	3.57	3.83	4.15	4.42	4.55	4.65	5.57	4.97
Bihor	2.91	3.14	3.30	3.46	3.66	3.85	4.12	4.16	4.27	5.21	4.54
Bistrita-Nas.	2.65	2.70	2.69	3.22	3.33	3.78	3.96	4.28	4.46	5.29	4.83
Botosani	2.90	3.23	3.42	3.59	3.96	4.21	4.64	4.80	5.08	5.89	5.24
Braila	3.13	3.38	3.55	3.79	4.38	4.50	4.82	5.18	5.40	5.77	5.58
Brasov	3.22	3.27	3.31	3.56	3.71	4.02	4.28	4.41	4.52	5.52	5.01
Bucharest	20.29	5.37	5.35	5.65	5.60	5.67	5.70	5.59	6.22	7.11	7.00
Buzau	3.55	3.66	3.88	3.90	4.12	4.24	4.41	4.52	4.52	4.93	4.66
Caras-Sev.	2.39	2.54	2.70	2.98	3.33	3.62	3.93	4.18	4.72	5.15	5.25
Calarasi	2.80	2.88	3.10	3.20	3.49	3.73	4.01	4.17	3.96	4.77	4.34
Cluj	4.52	4.51	4.57	4.97	5.10	5.12	5.30	5.05	5.22	6.26	6.07
Constanta	3.00	3.05	3.17	3.37	3.52	3.93	4.26	4.33	4.48	5.32	4.72
Covasna	3.72	3.86	4.08	4.16	4.19	4.38	4.54	4.71	4.62	5.22	4.87
Dambovita	2.43	2.58	2.80	3.07	3.35	3.65	3.90	4.27	4.24	4.91	4.55
Dolj	2.42	2.52	2.81	3.09	3.38	3.66	3.97	3.98	4.24	4.81	4.62
Galati	2.93	3.26	3.37	3.81	4.43	4.49	4.73	4.87	4.88	5.67	5.11
Giurgiu	2.13	2.28	2.36	2.65	2.96	3.19	3.70	3.32	3.75	4.74	4.20
Gorj	2.82	2.86	3.00	3.44	3.68	3.86	4.25	4.46	4.75	5.62	5.19
Harghita	4.26	4.41	4.43	4.65	4.79	5.06	5.31	5.24	5.33	5.70	5.84
Hunedoara	3.64	3.61	3.77	4.05	4.20	4.51	4.94	4.72	5.13	6.08	5.66
Ialomita	2.68	2.79	3.04	3.31	3.44	3.70	4.02	4.16	4.27	4.83	4.41
Iasi	2.97	3.06	3.25	3.37	3.57	3.69	4.28	4.19	4.45	5.79	5.18
Ilfov	3.08	3.14	3.51	3.60	3.67	3.54	3.73	4.07	4.49	5.49	5.10
Maramures	2.92	2.91	3.00	3.18	3.34	3.91	4.38	4.31	4.71	5.15	4.80
Mehedinti	2.17	2.24	2.45	2.60	2.88	3.28	3.77	3.80	4.01	4.99	4.53
Mures	3.21	3.38	3.55	3.70	3.96	4.01	4.29	4.39	4.29	5.00	4.79
Neamt	3.57	3.66	3.73	4.18	4.39	4.63	4.99	4.83	5.38	5.78	5.39
Olt	2.50	2.73	3.03	3.12	3.56	3.90	4.03	4.26	4.40	5.10	4.67
Prahova	3.51	3.61	3.69	3.99	4.12	4.49	4.70	4.93	5.17	5.92	5.59
Salaj	2.65	2.82	3.07	3.29	3.48	3.82	4.03	4.11	4.31	4.83	4.49
Satu Mare	2.36	2.84	2.83	3.09	3.38	3.58	3.87	4.02	4.25	4.79	4.54
Sibiu	2.88	2.95	3.08	3.11	3.50	3.66	4.05	4.27	4.31	5.08	4.88
Suceava	3.14	3.19	3.18	3.57	3.76	3.91	3.99	4.34	4.55	5.50	4.59
Teleorman	3.11	3.28	3.47	3.54	3.85	3.92	4.47	4.54	4.56	4.98	5.00
Timis	3.73	3.87	3.78	3.98	4.00	3.81	4.32	4.43	4.89	5.84	5.72
Tulcea	3.48	3.72	3.91	4.08	4.17	4.55	4.71	5.13	5.18	5.63	5.13
Vaslui	2.37	2.77	3.29	3.52	3.77	4.00	4.31	4.63	4.94	5.55	5.25
Valcea	3.91	4.29	4.42	4.79	5.29	5.02	5.31	5.48	5.53	5.66	5.64
Vrancea	3.09	3.31	3.57	3.69	3.97	4.24	4.51	4.59	4.77	5.20	4.87

(Source: Author's Calculations, Excel)

Appendix 4. Descriptive Statistics, NUTS3 Romania, 2011-2021

Appendix 4.1 Descriptive Statistics, Physician-Adult Ratio, NUTS3 Romania, 2011-2021

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Mean	2.37632	2.430997	2.466222	2.493233	2.542346	2.585237	2.630406	2.712812	2.801992	2.878434	2.999936
Standard Er	0.198137	0.200405	0.196898	0.202327	0.212171	0.218079	0.23024	0.241061	0.259017	0.282133	0.303456
Median	1.824995	1.940802	1.968158	1.982534	1.956616	2.042658	2.048212	2.109944	2.151014	2.169835	2.234902
Standard De	1.284074	1.298776	1.276043	1.311226	1.375025	1.413315	1.492128	1.562257	1.67862	1.828429	1.966622
Range	5.060574	5.228445	5.149836	5.470228	5.636128	5.92731	6.124343	6.170584	6.669526	7.033592	7.52365
Minimum	1.105542	1.116924	1.169557	1.17701	1.141337	1.133513	1.046333	1.196377	1.176559	1.222485	1.329623
Maximum	6.166116	6.345369	6.319393	6.647238	6.777465	7.060823	7.170677	7.366961	7.846085	8.256077	8.853273
Count	42	42	42	42	42	42	42	42	42	42	42

(Source: Author's Calculations, National Institute of Statistics – Romania (2024), Excel)

Appendix 4.2 Descriptive Statistics, SDT1, NUTS3 Romania, 2011-2021

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Mean	3.090837	3.232917	3.384677	3.615098	3.856197	4.062287	4.37232	4.47961	4.677566	5.39819	5.043162
Standard Er	0.097496	0.095613	0.090511	0.092941	0.089098	0.079274	0.072373	0.069347	0.074624	0.073921	0.082346
Median	2.972239	3.099835	3.29513	3.527805	3.734808	3.914685	4.302712	4.398693	4.584811	5.32608	4.981871
Standard De	0.631847	0.619642	0.586575	0.602324	0.577419	0.513751	0.469033	0.449419	0.483619	0.479062	0.533663
Range	3.15509	3.128895	2.984177	3.050961	2.723964	2.480107	1.998316	2.273451	2.473581	2.37224	2.800658
Minimum	2.133365	2.240902	2.36478	2.602967	2.8807	3.189184	3.701552	3.318972	3.749772	4.739283	4.201912
Maximum	5.288454	5.369797	5.348957	5.653927	5.604665	5.669291	5.699868	5.592423	6.223353	7.111522	7.00257
Count	42	42	42	42	42	42	42	42	42	42	42

(Source: Author's Calculation, Excel)

Appendix 4.3 Descriptive Statistics, Life Expectancy, NUTS3 Romania, 2011-2021

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Mean	73.8325	74.3375	74.792	75.043	75.00675	75.19	75.3275	75.4305	75.533	75.61625	75.523
Standard Er	0.165292	0.153376	0.148074	0.152625	0.154814	0.146453	0.150176	0.156578	0.16478	0.158042	0.150376
Median	73.845	74.215	74.695	75.005	74.94	75.18	75.23	75.21	75.365	75.59	75.43
Standard De	1.0454	0.970035	0.936505	0.965288	0.979132	0.926252	0.949798	0.990289	1.042163	0.999548	0.951059
Range	5.12	5.05	4.63	4.74	4.58	4.24	4.4	4.43	4.7	5.16	4.21
Minimum	71.91	72.2	72.91	73.09	73.19	73.43	73.44	73.68	73.56	73.18	73.65
Maximum	77.03	77.25	77.54	77.83	77.77	77.67	77.84	78.11	78.26	78.34	77.86
Count	42	42	42	42	42	42	42	42	42	42	42

(Source: Author's Calculation, National Institute of Statistics - Romania (2024), Excel)

Appendix 4.4 Descriptive Statistics, Urban-Rural Ratio, NUTS3 Romania, 2011-2021

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Mean	1.483777	1.468669	1.460812	1.449208	1.443962	1.430442	1.423893	1.420027	1.41489	1.409081	1.398055
Standard Er	0.281005	0.271802	0.264208	0.254213	0.248295	0.239951	0.2335	0.229567	0.225661	0.221652	0.216711
Median	1.079773	1.072798	1.072659	1.07172	1.070433	1.063825	1.063874	1.063937	1.064469	1.063622	1.057093
Standard De	1.777234	1.719027	1.671001	1.607783	1.570356	1.517582	1.476786	1.45191	1.427206	1.401848	1.370601
Range	11.1335	10.73219	10.39277	9.944401	9.671553	9.296526	8.998474	8.815105	8.634317	8.447535	8.21851
Minimum	0.255446	0.253624	0.253293	0.252419	0.252454	0.250399	0.24998	0.249192	0.248379	0.246842	0.244513
Maximum	11.38894	10.98581	10.64606	10.19682	9.924007	9.546926	9.248454	9.064297	8.882696	8.694377	8.463022
Count	42	42	42	42	42	42	42	42	42	42	42

(Author's Calculation, National Institute of Statistics – Romania (2024), Excel)

Appendix 4.5 Descriptive Statistics, Pensioner Ratio, NUTS3 Romania, 2011-

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Mean	0.705033	0.713439	0.720316	0.734301	0.747722	0.762486	0.773601	0.785706	0.798062	0.818295	0.82665
Standard Er	0.013782	0.013137	0.012451	0.012017	0.011662	0.011272	0.010722	0.010075	0.009456	0.00913	0.008915
Median	0.690159	0.699267	0.708723	0.722364	0.737106	0.752341	0.765526	0.778992	0.792984	0.813594	0.828349
Standard De	0.087164	0.083083	0.078747	0.076002	0.073759	0.071291	0.067814	0.063723	0.059807	0.057744	0.056385
Range	0.366379	0.3457	0.317493	0.308347	0.299372	0.296788	0.283739	0.275731	0.265715	0.263057	0.258558
Minimum	0.566084	0.579568	0.59303	0.606976	0.615377	0.627629	0.639118	0.652249	0.665695	0.68351	0.688261
Maximum	0.932463	0.925268	0.910523	0.915323	0.914749	0.924417	0.922857	0.92798	0.93141	0.946568	0.946819
Count	42	42	42	42	42	42	42	42	42	42	42

(Author's Calculation, National Institute of Statistics – Romania (2024), Excel)

Appendix 4.6 Descriptive Statistics, GDP, NUTS3 Romania, 2011-2021

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Mean	22571.15	24081.48	24232.81	25552.03	27113.89	28753.95	32772.58	37058.34	40780.01	40847.1	45719.84
Standard Er	1588.7	1713.779	1718.844	1857.688	2002.235	2111.437	2329.449	2525.818	2835.328	2800.894	3125.966
Median	19515.97	20570.2	20775.76	21212.71	22926.67	24749.53	28530.29	32206.62	35464.4	36130.3	40253.4
Standard De	10047.82	10838.89	10870.92	11749.05	12663.25	13353.9	14732.73	15974.67	17932.19	17714.41	19770.34
Range	53477.76	54558.43	58747.4	62443.06	70806.98	73113.71	82462.5	90452.3	103717.7	104264.9	117289.4
Minimum	11302.92	13029.98	12462.08	12762.22	13182.62	14289.5	15883	17842.76	18663.3	18618.81	20358.53
Maximum	64780.68	67588.41	71209.48	75205.28	83989.6	87403.2	98345.5	108295.1	122381	122883.7	137647.9
Count	42	42	42	42	42	42	42	42	42	42	42

(Author's Calculation, National Institute of Statistics – Romania (2024), Excel)

Appendix 4.7 Descriptive Statistics, Mean Age, NUTS3 Romania, 2011-2021

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Mean	39.655	39.9075	40.185	40.475	40.6425	40.8875	41.1025	41.315	41.52	41.735	41.865
Standard Er	0.190914	0.19066	0.192005	0.195158	0.199579	0.20347	0.206295	0.21242	0.219124	0.222559	0.225421
Median	39.5	39.7	40	40.4	40.5	40.8	41	41.2	41.4	41.6	41.7
Standard De	1.207445	1.205838	1.214348	1.234286	1.26225	1.286854	1.304723	1.343465	1.385863	1.407589	1.425689
Range	5.7	5.7	5.7	5.8	5.9	5.9	5.9	6.1	6.2	6.2	6.4
Minimum	37.4	37.6	37.8	38	38	38.2	38.4	38.5	38.6	38.8	38.8
Maximum	43.1	43.3	43.5	43.8	43.9	44.1	44.3	44.6	44.8	45	45.2
Count	42	42	42	42	42	42	42	42	42	42	42

(Author's Calculation, National Institute of Statistics – Romania (2024), Excel)

Appendix 4.8 Descriptive Statistics, Net Migration, NUTS3 Romania, 2011-2021

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Mean	-0.2098	-0.16467	-0.42264	-1.31621	-3.07918	-3.78392	-2.94289	-2.45753	-0.98901	-2.12284	-1.17658
Standard Er	0.058387	0.221804	0.286962	0.441544	0.25324	0.199838	0.372297	0.534952	0.517051	0.28817	0.40884
Median	-0.22293	-0.35675	-0.64782	-2.00622	-3.40894	-4.0293	-3.67505	-3.57261	-2.07071	-2.62227	-1.86808
Standard De	0.369273	1.40281	1.814905	2.792568	1.601629	1.263884	2.354612	3.383335	3.270119	1.82255	2.585732
Range	2.068618	8.902428	13.07464	17.71124	10.34441	7.309128	10.54267	15.6445	14.00208	9.072769	11.21545
Minimum	-0.93759	-1.39071	-3.25043	-2.84244	-4.1444	-5.25693	-4.96283	-5.64751	-4.02066	-4.21973	-3.90161
Maximum	1.131026	7.511716	9.824213	14.8688	6.200006	2.052194	5.57984	9.996996	9.981425	4.853043	7.31384
Count	42	42	42	42	42	42	42	42	42	42	42

(Author's Calculation, National Institute of Statistics – Romania (2024), Excel)

Appendix 4.9 Descriptive Statistics, Old Age Dependency Ratio, NUTS3 Romania, 2011-2021

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Mean	15.0187	15.30833	15.73409	16.41119	17.17884	17.63246	18.0021	18.34631	18.33251	18.56264	18.54303
Standard Er	0.314001	0.324232	0.33809	0.341168	0.354278	0.347143	0.335207	0.331914	0.328234	0.332185	0.332861
Median	14.73246	15.19796	15.65708	16.54971	17.26902	17.8387	18.2277	18.39736	18.34567	18.78318	18.76276
Standard De	1.985918	2.050623	2.138269	2.157736	2.240648	2.195524	2.120036	2.099205	2.075933	2.100921	2.105195
Range	9.829479	10.31814	10.96099	11.29761	12.00854	11.83386	11.81631	11.71759	11.89149	11.92767	11.90868
Minimum	12.03165	12.28483	12.69506	12.93332	13.40303	13.91645	13.98924	14.08389	13.80545	13.74667	13.55946
Maximum	21.86112	22.60297	23.65605	24.23093	25.41157	25.75031	25.80555	25.80148	25.69694	25.67434	25.46814
Count	42	42	42	42	42	42	42	42	42	42	42

(Author's Calculation, National Institute of Statistics – Romania (2024), Excel)

Appendix 4.10 Descriptive Statistics, Unemployment Rate, NUTS3 Romania, 2011-2021

	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
Mean	5.8525	6.065	6.375	6.1	5.7425	5.62	4.68	3.8825	3.4275	3.9325	3.65
Standard Er	0.316855	0.328859	0.377301	0.393977	0.419507	0.435186	0.390713	0.323004	0.279147	0.279248	0.30498
Median	5.75	5.95	6.2	5.95	5.5	5.35	4.15	3.4	3.1	3.55	3.45
Standard De	2.003969	2.079885	2.386259	2.49173	2.653192	2.75236	2.471084	2.04286	1.765479	1.766118	1.928863
Range	8.2	8.3	8.9	9.9	10.4	11	9.8	8	7.1	7.2	8.2
Minimum	1.6	1.8	1.9	1.5	1.2	1	0.6	0.6	0.4	0.7	0.4
Maximum	9.8	10.1	10.8	11.4	11.6	12	10.4	8.6	7.5	7.9	8.6
Count	42	42	42	42	42	42	42	42	42	42	42

(Author's Calculation, National Institute of Statistics – Romania (2024), Excel)

Appendix 5. Forecasting Accuracy, IFIMd, Holt's & Linear Regression, NUTS3 Romania, 2011-2021

	Holt's		LinReg			Holt's		LinReg	
County	**MAD**	**MPE**	**MAD**	**MPE**	**County**	**MAD**	**MPE**	**MAD**	**MPE**
Alba	0.01	2.7%	0.02	3.6%	**Harghita**	0.04	14.6%	0.03	9.7%
Arad	0.01	3.2%	0.02	4.3%	**Hunedoara**	0.01	2.2%	0.01	2.7%
Arges	0.02	0.5%	0.02	3.0%	**Ialomita**	0.03	8.9%	0.02	5.9%
Bacau	0.03	-4.8%	0.02	1.4%	**Iasi**	0.02	3.2%	0.02	4.2%
Bihor	0.02	-1.1%	0.02	2.1%	**Ilfov**	0.01	-2.1%	0.01	-0.8%
Bistrita-Na	0.03	9.2%	0.03	9.4%	**Maramures**	0.01	-0.2%	0.01	2.3%
Botosani	0.02	6.7%	0.02	3.9%	**Mehedinti**	0.05	18.4%	0.03	11.3%
Braila	0.02	0.6%	0.02	1.8%	**Mures**	0.03	-0.9%	0.03	-0.5%
Brasov	0.03	4.6%	0.02	4.2%	**Neamt**	0.03	10.6%	0.02	5.4%
Bucharest	0.03	2.5%	0.03	2.5%	**Olt**	0.04	15.9%	0.03	11.0%
Buzau	0.02	4.8%	0.02	4.8%	**Prahova**	0.04	12.7%	0.03	8.1%
Caras-Seve	0.03	7.5%	0.02	5.7%	**Salaj**	0.01	-2.5%	0.02	-0.9%
Calarasi	0.03	16.7%	0.02	9.7%	**Satu Mare**	0.02	6.5%	0.01	3.4%
Cluj	0.02	1.7%	0.02	3.3%	**Sibiu**	0.03	-3.8%	0.03	1.8%
Constanta	0.02	3.3%	0.02	5.2%	**Suceava**	0.03	10.4%	0.03	7.6%
Covasna	0.03	-9.8%	0.01	-2.1%	**Teleorman**	0.02	7.8%	0.02	5.7%
Dambovita	0.02	7.4%	0.02	6.5%	**Timis**	0.03	7.5%	0.03	5.9%
Dolj	0.02	1.7%	0.03	1.6%	**Tulcea**	0.02	2.4%	0.02	2.7%
Galati	0.03	7.8%	0.03	4.2%	**Vaslui**	0.02	3.4%	0.02	2.6%
Giurgiu	0.04	13.4%	0.03	7.0%	**Valcea**	0.03	7.4%	0.03	7.2%
Gorj	0.02	-3.5%	0.03	-3.0%	**Vrancea**	0.03	4.8%	0.03	3.3%

(Author's own calculations, Excel)

Appendix 6. Stepwise Outputs, Steps and Added r, Adjusted r2, NUTS3 Romania, 2011-2021

County	Step/ Variable	r2	Adjusted r2
Alba	PENS	0.9501	0.9446
Arad	PENS	0.9675	0.9902
	LifEx	0.0247	
Arges	PhysRat	0.9370	0.9800
	SDT1	0.0470	
Bacau	URR	0.9047	0.9660
	MAGE	0.0681	
Bihor	PENS	0.9604	0.9878
	SDT1	0.0119	
Bistrita-Nasaud	GDP	0.8960	0.7245
	UNEMP	0.0572	
Botosani	URR	0.9283	0.9203
Braila	URR	0.9283	0.9203
Brasov	PENS	0.9450	0.9723
	MAGE	0.0329	
Bucharest	PENS	0.9495	0.9439
Buzau	MAGE	0.9557	0.9853
	PENS	0.0325	
Caras-Severin	SDT1	0.9038	0.8931
Calarasi	GDP	0.4798	0.6252
	LifEx	0.2204	
Cluj	URR	0.9698	0.9664
Constanta	PhysRat	0.9852	0.9989
	LifEx	0.0062	
	UNEMP	0.0051	
	MAGE	0.0027	
Covasna	URR	0.9432	0.9995
	OADR	0.0448	
	SDT1	0.0081	
	NMIG	0.0028	
	GDP	0.0009	
Dambovita	GDP	0.9339	0.9265
Dolj	PhysRat	0.9199	0.9824
	URR	0.0359	
	GDP	0.0320	
Galati	PENS	0.9637	0.9840
	GDP	0.0235	
Giurgiu	URR	0.5861	0.9928
	LifEx	0.2124	
	GDP	0.1064	
	PENS	0.0569	
Gorj			0.0000
Harghita	PENS	0.8397	0.9600
	MAGE	0.1282	
Hunedoara	PhysRat	0.9445	0.9930
	LifEx	0.0499	
Ialomita	PhysRat	0.8597	0.9593
	LifEx	0.0830	
	SDT1	0.0268	
Iasi	PENS	0.9794	0.9863
	UNEMP	0.0096	
Ilfov	PhysRat	0.9343	0.8643
	NMIG	0.0371	
Maramures	UNEMP	0.9324	0.9249
Mehedinti	NMIG	0.5844	0.5382
Mures	MAGE	0.8444	0.8271
Neamt	PhysRat	0.8072	0.9957
	MAGE	0.1564	
	URR	0.0171	
	OADR	0.0168	
Olt	URR	0.8622	0.8469
Prahova	PENS	0.8956	0.9619
	LifEx	0.0740	
Salaj	LifEx	0.9510	0.9734
	URR	0.0277	
Satu Mare	PENS	0.5548	0.8552
	PhysRat	0.3294	
Sibiu	PENS	0.9475	0.9893
	LifEx	0.0396	
Suceava	GDP	0.8077	0.7864
Teleorman	PENS	0.8678	0.9682
	LifEx	0.0783	
	UNEMP	0.0317	
Timis	PENS	0.9762	0.9927
	LifEx	0.0180	
Tulcea	PENS	0.9678	0.9642
Vaslui	PENS	0.8125	0.7916
Valcea	GDP	0.4838	0.6763
	NMIG	0.2551	
Vrancea	MAGE	0.9124	0.9908
	PENS	0.0466	
	LifEx	0.0320	
	SDT1	0.0060	

(Author's own calculations, SAS & Excel)

Appendix 7. R^2 Values of Presdictors, y=IFIMd, Stepwise Regression, NUTS3 Romania, 2011-2021

	SDT1	*LifEx*	*URR*	*PENS*	*GDP*	*MAGE*	*NMIG*	*OADR*	*PhysRat*	*UNEMP*
Alba	0.0000	0.0000	0.0000	0.9501	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Arad	0.0000	0.0247	0.0000	0.9675	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Arges	0.0470	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9370	0.0000
Bacau	0.0000	0.0000	0.9047	0.0000	0.0000	0.0681	0.0000	0.0000	0.0000	0.0000
Bihor	0.0119	0.0000	0.0000	0.9604	0.0000	0.0000	0.0000	0.0000	0.0000	0.0192
Bistrita-Nasaud	0.0000	0.0000	0.0000	0.0000	0.8960	0.0000	0.0000	0.0000	0.0000	0.0572
Botosani	0.0000	0.0000	0.9283	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Braila	0.0000	0.0000	0.9283	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Brasov	0.0000	0.0000	0.0000	0.9450	0.0000	0.0329	0.0000	0.0000	0.0000	0.0000
Bucharest	0.0000	0.0000	0.0000	0.9495	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Buzau	0.0000	0.0000	0.0000	0.0325	0.0000	0.9557	0.0000	0.0000	0.0000	0.0000
Caras-Severin	0.9038	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Calarasi	0.0000	0.2204	0.0000	0.0000	0.4798	0.0000	0.0000	0.0000	0.0000	0.0000
Cluj	0.0000	0.0000	0.9698	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Constanta	0.0000	0.0062	0.0000	0.0000	0.0000	0.0027	0.0000	0.0000	0.9852	0.0051
Covasna	0.0081	0.0000	0.9432	0.0000	0.0009	0.0000	0.0028	0.0448	0.0000	0.0000
Dambovita	0.0000	0.0000	0.0000	0.0000	0.9339	0.0000	0.0000	0.0000	0.0000	0.0000
Dolj	0.0000	0.0000	0.0359	0.0000	0.0320	0.0000	0.0000	0.0000	0.9199	0.0000
Galati	0.0000	0.0000	0.0000	0.9637	0.0235	0.0000	0.0000	0.0000	0.0000	0.0000
Giurgiu	0.0000	0.2124	0.5861	0.0569	0.1064	0.0000	0.0000	0.0000	0.0135	0.0218
Gorj	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Harghita	0.0000	0.0000	0.0000	0.8397	0.0000	0.1282	0.0000	0.0000	0.0000	0.0000
Hunedoara	0.0000	0.0499	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9445	0.0000
Ialomita	0.0268	0.0830	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.8597	0.0000
Iasi	0.0000	0.0000	0.0000	0.9794	0.0000	0.0000	0.0000	0.0000	0.0000	0.0096
Ilfov	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0371	0.0000	0.9343	0.0000
Maramures	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.9324
Mehedinti	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.5844	0.0000	0.0000	0.0000
Mures	0.0000	0.0000	0.0000	0.0000	0.0000	0.8444	0.0000	0.0000	0.0000	0.0000
Neamt	0.0000	0.0000	0.0171	0.0000	0.0000	0.1564	0.0000	0.0168	0.8072	0.0000
Olt	0.0000	0.0000	0.8622	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Prahova	0.0000	0.0740	0.0000	0.8956	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Salaj	0.0000	0.9510	0.0277	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Satu Mare	0.0000	0.0000	0.0000	0.5548	0.0000	0.0000	0.0000	0.0000	0.3294	0.0000
Sibiu	0.0000	0.0396	0.0000	0.9475	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Suceava	0.0000	0.0000	0.0000	0.0000	0.8077	0.0000	0.0000	0.0000	0.0000	0.0000
Teleorman	0.0000	0.0783	0.0000	0.8678	0.0000	0.0000	0.0000	0.0000	0.0000	0.0317
Timis	0.0000	0.0180	0.0000	0.9762	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Tulcea	0.0000	0.0000	0.0000	0.9678	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Vaslui	0.0000	0.0000	0.0000	0.8125	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Valcea	0.0000	0.0000	0.0000	0.0000	0.4838	0.0000	0.2551	0.0000	0.0000	0.0000
Vrancea	0.0060	0.0320	0.0000	0.0466	0.0000	0.9124	0.0000	0.0000	0.0000	0.0000

(Source: Author's Work, Excel)

Appendix 8. Ward's Hierarchical to 5 Clusters by NUTS3 Romania, Stepwise Regression y=IFIMd, 2011-2021

Obs	County	Cluster	Obs	County	Cluster
1	Botosani	1	22	Bacau	1
2	Braila	1	23	Bistrita-Nasaud	4
3	Alba	2	24	Dambovita	4
4	Bucharest	2	25	Vaslui	2
5	Tulcea	2	26	Mures	3
6	Arad	2	27	Ilfov	3
7	Timis	2	28	Suceava	4
8	Iasi	2	29	Ialomita	3
9	Galati	2	30	Harghita	2
10	Cluj	1	31	Gorj	2
11	Sibiu	2	32	Satu-Mare	2
12	Brasov	2	33	Giurgiu	1
13	Bihor	2	34	Calarasi	4
14	Prahova	2	35	Neamt	3
15	Teleorman	2	36	Valcea	4
16	Constanta	3	37	Arges	2
17	Dolj	3	38	Covasna	1
18	Buzau	3	39	Mehedinti	4
19	Vrancea	3	40	Salaj	3
20	Olt	1	41	Caras-Severin	3
21	Hunedoara	3	42	Maramures	5

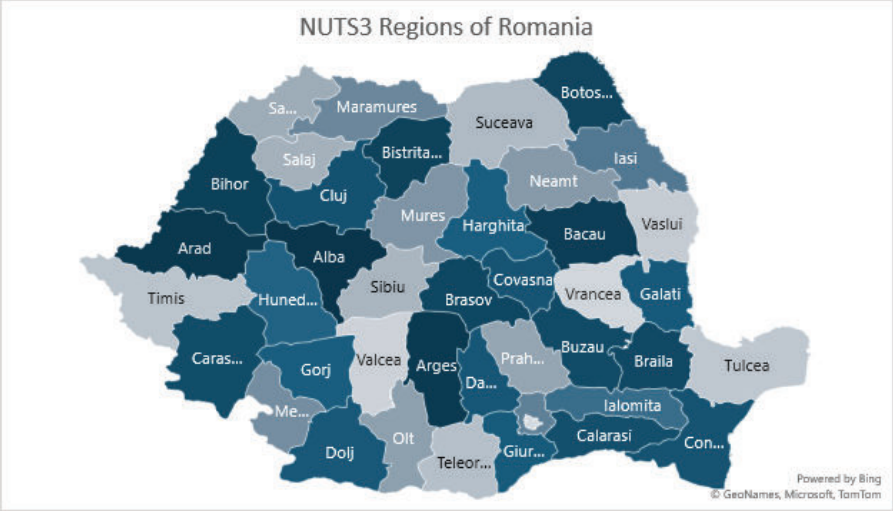
(Author's Calculations, SAS)

Appendix 9. Ward's Hierarchical Clustering to 5 Clusters by NUTS3 Romania, Stepwise Regression y=IFIMd, 2011-2021

Obs	County	CLUSTER	Obs	County	CLUSTER
1	Galati	1	**22**	Maramures	3
2	Tulcea	1	**23**	Vrancea	1
3	Arges	1	**24**	Sibiu	1
4	Iasi	1	**25**	Timis	1
5	Braila	1	**26**	Covasna	3
6	Salaj	1	**27**	Vaslui	3
7	Alba	2	**28**	Constanta	1
8	Suceava	2	**29**	Dolj	1
9	Botosani	2	**30**	Mures	1
10	Harghita	2	**31**	Satu-Mare	3
11	Bihor	1	**32**	Valcea	3
12	Bistrita-Nasaud	2	**33**	Giurgiu	2
13	Olt	2	**34**	Calarasi	2
14	Hunedoara	2	**35**	Dambovita	2
15	Neamt	2	**36**	Gorj	3
16	Teleorman	2	**37**	Buzau	2
17	Arad	2	**38**	Ilfov	2
18	Prahova	2	**39**	Mehedinti	2
19	Caras-Severin	3	**40**	Brasov	4
20	Ialomita	3	**41**	Cluj	4
21	Bacau	3	**42**	Bucharest	5

(Author's Calculations, SAS)

Appendix 10. Cartogram, NUTS3 Regions of Romania



(Source: Author's own work)

Individual Determinants of the Fixed Internet Adoption in Romania

Eugenia OANA (oanaeugenia19@stud.ase.ro)[1]
Faculty of Cybernetics, Statistics and Economic Informatics, Bucharest University of Economic Studies, Romania

Monica ROMAN (monica.roman@csie.ase.ro)
Faculty of Cybernetics, Statistics and Economic Informatics, Bucharest University of Economic Studies, Romania

Emanuelle Perta (munteanemanuelle23@stud.ase.ro)
Faculty of Cybernetics, Statistics and Economic Informatics, Bucharest University of Economic Studies, Romania

ABSTRACT

The aim of this paper is to analyze the association of internet adoption, represented by having fixed internet services, to relevant sociodemographic variables among Romanian population. The data source is a representative survey with a sample of 1437 respondents undertaken in 2022. Using several specific statistical tests and regression models, the paper has determined the relevant factors affecting the internet adoption in Romania. Our results indicate a positive association between the education level, the income level, the employment and the number of persons in the household and having fixed internet services. In elderly population instead the probability to adopt fixed internet services decreases. Our findings are relevant in identifying the gaps in fixed internet services adoption between several categories of population and designing the configuration of appropriate public policies and commercial strategies for setting incentives to adopt the services.

Keywords: *Fixed Internet, Survey, Romania, Logistic regression*
JEL Classification: *O20, O31, C10*

1. INTRODUCTION

The access to internet has come a long way from its incipient form to the scarce resource and then to the everyday fuel of connectivity, productivity and creativity. The internet is an important investment that provides information as data or knowledge, communication between people or between machines and increasing categories of services.

1. Corresponding author

The digital divide is one concern that is addressed between countries, regions or various categories of population because it reduces the benefits for end users, companies and the internal market. At the European level, the importance of internet availability is best reflected in the Directive (EU) 2018/1972 of the European Parliament and of the Council of 11 December 2018 establishing the European Electronic Communications Code that underlines the benefits across the Union through the connectivity and the participation in the digital economy.

The Romanian situation is particularly relevant because the digital economy indicators are placed at extremes and this raises the questions of what are the relevant factors that might benefit the development and how could these be enhanced in order to obtain better connectivity. The high-level analysis of the internet use by the natural persons in 2022 at European level reveals that Romania has fallen behind other countries, although it has managed to narrow the gap and the divide between countries has reduced. The data produced by Eurostat shows that while the EU average fixed internet utilization at the European level in 2022 was 91.14%, in Romania the percentage was 88.86% (Annex 1).

Although the mobile internet is more widely spread than the fixed internet, the latter is the one that supports more traffic. The data available in the Statistical data report of the National Authority for Management and Regulation in Romania (ANCOM) shows that a bit less than 25% of the internet connections provide almost 90% of the traffic (Annex 2).

The aim of this paper is to analyze the gap within the sociodemographic factors that could impact the adoption of the fixed internet services in the Romanian households. Previous studies have shown that there are differences between sociodemographic groups with respect to the adoption and use of the internet and the devices (Tsetsi, 2017) and even when the internet adoption increases despite those differences, the speed of connections makes a difference in the accessibility of information (DiMaggio et al., 2001).

This paper is a contribution to the scrutiny at the national regarding the incentives for the internet development and utilization. Its results provide the support for public policies aimed at the efficient network development and services provision. Private sector can benefit as well by identifying strategies to increase the adoption of fixed internet by narrowing the gap between the different categories of potential users.

Our study capitalizes on the available data set from 2022 on the attitudes toward the electronic communications that highlights the correspondence between the population characteristics and the adoption of the services. The data has been selected from the user survey undertaken by the

National Authority for Management and Regulation in Romania (ANCOM) and it includes various information related to having fixed internet personally or in the household, as well as socio-demographic characteristics: the age, the gender, the employment status, the income, the education, the number of persons in the household, the number of children in the household and the living area.

The novelty of the paper is specifically related to providing a closer perspective on the Romanian case, less covered by the national or international literature. The analysis of the large set of recent data is based on the representative survey amongst Romanian population and the application of several statistical tests and models to determine the significance of the socio-demographic factors, while identifying in the results the specificities of the Romanian context.

Considering the available data for the Romanian population and the relevant studies, in our analyses we have taken into account the following hypotheses: socio-economic factors have a significant impact on having fixed internet, positive (income, education, family size and employment) or negative (rural and age). We employ statistical methods to capture the influence of individual factors on owning fixed internet services.

2. LITERATURE REVIEW

The research on internet adoption is extensive and rightly so taking into account the impact on the economy of this technology. Many studies focus on the internet from the perspective of an independent variable, as the factor that has a strong influence on the life of the people and on the performance of the businesses.

Our contribution is related to the previous stage that refers to the factors that influence internet adoption where it is the dependent variable. For the purpose of including all the potentially relevant factors, this paper covers both fixed and mobile internet studies and the literature has revealed that the differences between social groups regarding the access to technology are diminishing, but the ones related to the skills and the access to the information remain relevant. This research provides the background for the formulation of our hypotheses and a challenge to determine the validation of previous work for the Romanian market.

The digital divide in terms of the dependence on the smartphones as unique means of access to the internet by different demographic groups (according to race, age, income and education) has been seen in the dependency of the minority, less educated and lower income population, while Caucasian,

more educated and higher income population has been more likely to be multi-modal (Tsetsi, 2017). The differences are widening when it is taken into account the manner of using the smartphones, low-income individuals engaging in less information activities; by the other hand, they use the smartphones mainly for social activities, which might widen the differences between the groups. Broadband adoption at households' level is also dependent on higher level of income and education (Whitacre et al., 2015).

Family income has shown to be one of the most significant factors even by comparison to other demographic characteristics (such as education, age, sex and race/ethnicity), pointing to the highest level of inequality together with nativity/citizenship (Martin and Robinson, 2007).

Even when internet services adoption increases although the differences in income and education persist, lower speed connections and low skills imply more time is needed in order to obtain the necessary information (DiMaggio et al., 2001). There are several dimensions of the online inequalities such as technical capacity, autonomy, skills, social support and goals/purposes. Some differences in payment determined by computer use have been determined, but these could also have been attributed to collateral factors.

A prerequisite for having fixed internet is at least the basic knowledge for accessing it. Considering the internet use skills as a vital resource that can be classified in four relevant categories: operational, formal, informative and strategic (Van Deursen and Van Dijk, 2011), there are significant differences in performance between the categories of the analyzed population that are determined by age and by the level of education. Further sensitivity in internet use by the elderly persons show a high heterogeneity of this group, including gender differences (Ramon-Jeronimo et al., 2013). It would also show an impact on the purpose of using the internet, such as online purchases (Cazacu et al., 2021).

The difference in knowledge is propagated at the gender level, men utilizing the internet more than women, nonetheless, access to the internet being independent of gender (Wasserman and Richmond-Abbott, 2005).

This difference in utilization has been noticed since adolescence, as a research that has included several countries, including Romania, has found that pathological internet use is higher in males than in females adolescents, while for maladaptive internet use this is reversed (Durkee et al., 2012). The authors have mentioned that they have been aware that dependency is more severe in males in general, which might partially explain the results.

In Romania and also for the younger population, the assessment of female and male university students has shown that the skills, the attitude and the use of the internet are higher for males in some studies (Durndell

and Haag, 2002), while in other studies the difference between young male and female participants was not significant, as compared to the age and the education level and subjects studied (Cazan et al., 2016).

Taking into account the higher investment needs in the rural areas in the telecom sector, it is understandable that remote areas have lower internet, including broadband penetration. This is accentuated by lower education levels, aging population, the structure of population in terms of ethnicity, higher unemployment rate and larger primary industry sectors (Sora Park, 2017). A study done for rural Mexico has confirmed that wealth and education are key factors for internet penetration, a positive impact being determined in the households with a greater number of students, while the age and the employment status introduce gaps between categories for internet access and use (Martinez-Dominguez and Mora-Rivera, 2020).

In addition to factors such as age, education level and perceived benefits of internet use, depending on national specificities other significant variables have been noted, such as family support in South Korea (Rhee and Kim, 2017).

In the debate concerning the sources of innovation as either technology push or demand pull, the bibliometric analysis has revealed the role of demand as one important factor (Di Stefano et al., 2012).

3. METHODOLOGY AND DATA

3.1. Data source

The micro data have been collected from the database of respondents obtained during the study undertaken by ANCOM amongst the end users of electronic communication services. It is a tracking study that has as objective to obtain relevant information on the attitude regarding the utilization of the fixed and mobile telephone, fixed and mobile internet and television services by the individuals. This study facilitates the highlighting of the evolution of telecommunication services and the users' attitude regarding relevant aspects such as the problems they meet, reasons for not using the services, the level of satisfaction, the devices owned, the consumption of the services.

The data has been collected during April-June 2022 and it includes 1437 respondents older than 16 years of age, directly involved in the decision to have electronic communications services and who live in Romania. The questionnaire has been filled in by computer assisted telephone interviewing (CATI). The study has used a simple random sample, stratified by locality size. The sample is representative at national level, and the margin of error is +/- 2.59% at a confidence level of 95%.

In order to reach our aim to analyze the factors that influence internet adoption, we have used the variable that shows if the respondents have fixed internet services personally or in the household. The variable has been set to 1 for having fixed internet services and 0 for not having fixed internet services.

The independent variables have been used either in the form that they have been collected from the respondents or in an aggregated form. Age has been measured in years, based on the respondents' direct answers. The family size has been measured by the number of persons in the household and the number of children under 18 years in the household. The gender, the living area and the employment status have been included as dichotomous variables where 1 represents male, urban and employed respectively and 0 represents female, rural and unemployed respectively. For the education the respondents have been offered 10 options increasing from no education to having a PhD and these have been aggregated into 3 categories (primary, secondary and tertiary education). The answers for household income have been in the form of 8 intervals, from no income to income above 10,000 RON and these have been aggregated into 4 categories with values from less than 2,000 RON to above 6,000 RON.

3.2. Methodology

We have started the data exploration with the descriptive analysis in order to determine the profile of the respondents. Afterwards, we have tested our hypotheses that income, education, employment and family size have a positive impact on having fixed internet, while rural areas and age have a negative impact.

In this process, several non-parametric tests have been done, depending on the nature of the variables, in order to determine the association of the parameters with having fixed internet services. The chi-squared test has been used to determine the extent to which having fixed internet is influenced by employment and by gender respectively, taking into account that the analyzed variables are not normally distributed.

In order to verify whether there is a correlation between having fixed internet services and the education and the income interval respectively (that are ordinal variables in the database) Mann-Whitney tests have been undertaken.

The significance of the age has been assessed with Student's t-test for two independent samples (those who have fixed internet and those who don't have fixed internet).

All the parameters have been afterwards factored into the binary logistic regression model in order to determine the correlation between

sociodemographic factors and having fixed internet services. The dependent dichotomous variable of the model (y) is showing if the respondents have fixed internet services and it takes the value of 1 for the respondents who have fixed internet services and 0 for those who don't have these services.

The general form of the model is (Andrei and Bourbonnais, 2008):

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon, \text{ where p is P(y = 1 | x1,x2,…,xk)} \quad [1]$$

In the equation, when xi increases by one while other variables are constant, the logit (logarithm of OR) increases by β_i.

The model cand be rewritten as:

$$P(y=1/x_1,x_{2,\ldots}x_k) = \frac{\exp(\beta_0+\beta_1 1+\ldots\beta_k x_k)}{1+\exp(\beta_0+\beta_1 1+\ldots\beta_k x_k)} \quad [2]$$

From the calculation we extract:

$$\exp(\beta_0) = \frac{P(y=1/x_1=x_2=\ldots=x_k=0)}{P(y=0/x_1=x_2=\ldots=x_k=0)} \quad [3]$$

that is OR when all factors are set to 0.

For the β_i coefficient it results that:

$$\exp(\beta_i) = \frac{P(y=1/x_i=1,x_j=0 \text{ for } j\neq i)}{1-P(y=1/x_i=1,x_j=0 \text{ for } j\neq i)} \times \frac{1}{OR_{base}} = \frac{OR_{x_i=1,x_j=0}}{OR_{base}} \quad [4]$$

Taking into account the multiplicative nature of the model, we determine:

$$OR_{x_i,x_2,\ldots x_k} = \exp(\beta_0)\, x \prod_{i=1}^{k} \exp(\beta_i x_i) \quad [5]$$

that shows the contribution of the x_i factor in the explanation of the probability (as OR) of the event y=1, expressed by β_i.

Therefore, while setting $x_i = 1$, $\exp(\beta_i)$ will be the constant multiplicative factor, regardless of the values of other independent variables.

In case $\beta_i = 0$, there is no effect of the corresponding factor because the multiplication is by 1.

For values of β_i different to 0, the likelihood of the event y = 1 is reduced by the presence of the factor (when it is negative) or increased (when it is positive).

4. RESULTS AND DISCUSSIONS

4.1. Sample description

The descriptive analysis of the variables offers a first image of the sample. This shows that the great majority of the respondents (72%) have fixed internet.

With respect to the number of persons in the household, most respondents live in households formed by 2 persons and the highest percentages are the ones of the households formed by one to four persons, the rest of the households being represented in a lower number (less than 10%).

At the same time, most respondents (64.4%) don't have children and the households with one child are 21.78% of total households, which could have an influence on having fixed internet through the way in which the decisions are taken in the household or through taking into account the access to the technology skills of some members of the household.

Regarding the education, most respondents have an average level of education, high-school being the most frequent form (37%). One third of respondents have higher education; according to Bologna Implementation report 2024, the percentage of the persons with higher education in Romania is amongst the lowest in Europe and well below the average. This can be a factor that influences the skills level in internet usage for the jobs and therefore the need to have fixed internet.

Income is a relevant variable for our study, as the households with higher income may have a better internet access. The highest percentage of the respondents falls in the income category 4001-6000 RON per household, below the monthly average total income at national level, which was 6464 RON in 2022 (Source: INS press release no. 136 from 7 June, 2023).

With respect to the other socio-demographic data, such as age and gender, we have noticed that the average age of the respondents is 48 years and the gender is balanced at 52.26% women and 47.74% men. We have looked at gender to test if there is an influence on having fixed internet in Romania or the previous observations that there are no gender disparities is sustained.

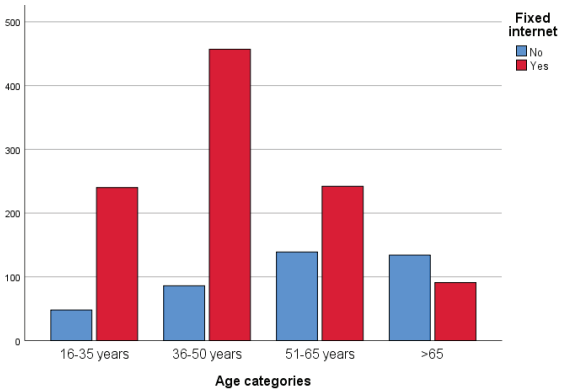
There is also a balance regarding the living area, 57.69% of the respondents being in urban areas.

The percentage of the employed persons is 66.4%. The other respondents are unemployed, retired, students, housewives, on maternity leave or in other situations that are taken into account cumulatively.

4.2. Identifying the factors related to the internet adoption using the statistical hypothesis testing

Looking at the age categories in Figure 1, these are reflected in the percentages in which the fixed internet services are owned, being evident that the respondents of ages 50 and lower enthusiastically adopt these services. After this age, the adoption decreases to the degree that the respondents of 65 years old and older predominantly don't have fixed internet. This is related probably to more specific factors, such as the interest that the respondents of this age show towards new technologies and the skills they have not had the opportunity to develop. The interaction with other age groups and the attitude towards technology could therefore be important factors for the development of the interest for the fixed internet services.

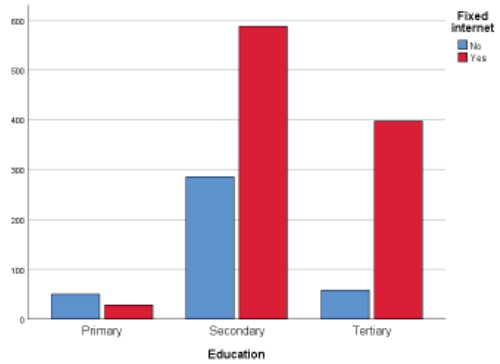
Figure 1. Fixed internet adoption by age



Source: produced by authors based on data from the ANCOM study

The association between education and having fixed internet services can be noticed in Figure 2 by the fact that, amongst the respondents with primary education, the percentage of the ones who have these services is smaller in total. This is reversed for the secondary and tertiary education. The explanation for those who do not have these services although they have tertiary education could be the substitutability with the mobile internet that is more accessible in some areas or maybe that after the completion of the studies the pursued careers do not require internet access in the household.

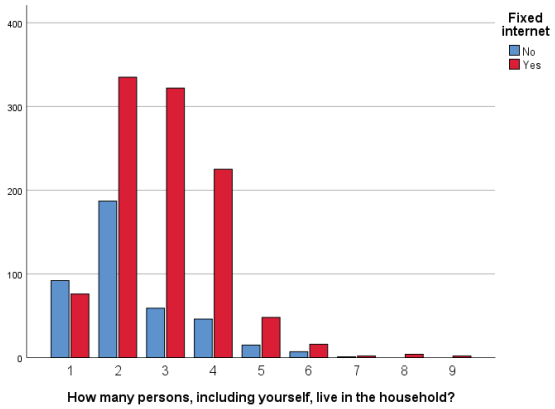
Figure 2. Fixed internet adoption by education



Source: produced by authors based on data from the ANCOM study

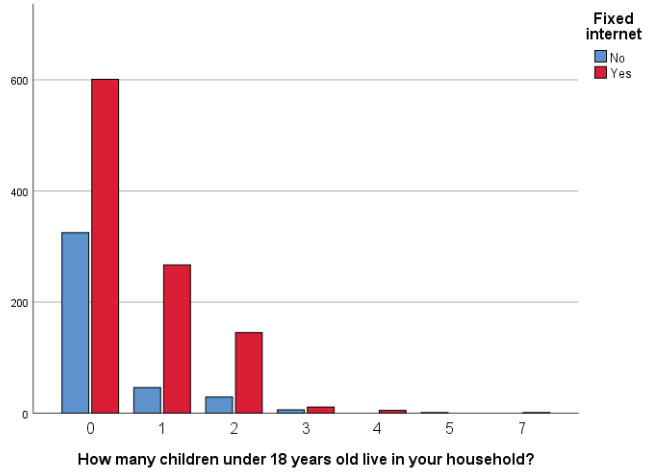
As it can be seen in the household structure in Figure 3 and Figure 4, the number of persons might be associated with having fixed internet services, while the number of children to a lesser extent. This could be a reasonable perspective taking into account suppositions such as the increased pressure from the households members who want internet on those who are less likely to adopt these services or the fact that the increase in the contribution to the expenses related to electronic communications that are fixed costs in nature decreases the individual burden or the need determined by the education process.

Figure 3. Fixed internet adoption by number of persons in the household



Source: produced by authors based on data from the ANCOM study

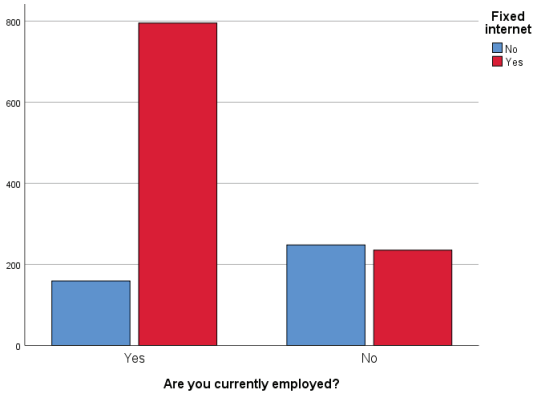
Figure 4. Fixed internet adoption by number of children in the household



Source: produced by authors based on data from the ANCOM study

There are differences between the employed and unemployed persons with respect to the proportion in which they have fixed internet services. Therefore, while there are no significant differences within the unemployed regarding having fixed internet services, within the employed the proportion of the ones who have fixed internet services is higher (figure 5).

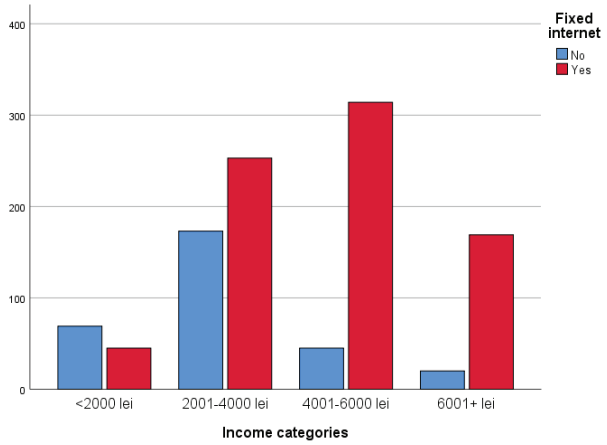
Figure 5. Fixed internet adoption by occupation



Source: produced by authors based on data from the ANCOM study

One possible explanation could be that the unemployed have lower income but, as the percentage of the unemployed respondents who have services is almost as high as the ones who don't, the cost doesn't seem to be an overarching factor, specifically taking into account that in Romania the tariffs are amongst the lowest in Europe. Furthermore, looking at the income categories in Figure 6, it can be noticed that, starting with the income above 2,000 RON, the situation regarding the fixed internet services becomes even and, for larger income categories, the proportion of the ones who have these services is much higher.

Figure 6. Fixed internet adoption by income



Source: produced by authors based on data from the ANCOM study

This observation supports the further investigation of the dependency on the income of the access to the fixed internet, although low income does not completely restrict the access.

Therefore, it could be assumed that the fixed internet is considered a relevant service by the employed individuals and it could influence the labor market specifically in the context of the continued implementation of the work from home during our year of analysis, 2022, that required adequate resources of communication. At the same time, the breakdown on occupation shows that the retired and the unemployed individuals have these services in a lower proportion, which might point to a lower accessibility determined by income levels and the skills needed to interact with the online environment.

In order to determine the extent to which having fixed internet is influenced by employment, a chi-squared test has been used taking into account that the analyzed variables are not normally distributed, as seen in

Table 1. This confirms that the employment has an association with the fixed internet services.

Table 1. The impact of employment on having fixed internet. Results of the chi-squared test

Variables	Pearson Chi-Square	df	P-value	Phi	Cramer's V
Employment	189.957	1.00	0.000	-0.364	0.364

Source: produced by authors based on data from the ANCOM study

During our research, we have undertaken other several statistical tests depending on the used variables with the purpose of identifying the degree to which having fixed internet is determined by certain factors.

In order to verify whether there is a correlation between having fixed internet services and the education and the income interval respectively that are ordinal variables in the database, we have undertaken Mann-Whitney tests.

For all the analyzed data there is a variation determined by having fixed internet services (Table 2). Having fixed internet services increases with the education and the income interval.

Table 2. The role of education and income on having fixed internet. Results of the Mann Whitney test

Variables	Yes (Mean rank)	No (Mean rank)	Z	P-value
Education	762.08	552.50	-10.190	0.000
Income interval	775.17	576.86	-8.319	0.000

Source: produced by authors based on data from the ANCOM study

The Student's t-test for two independent samples (those who have fixed internet and those who don't have fixed internet) for the age shows, in Table 3, that this variable is significantly different in terms of having fixed internet services. Looking further into the data, the average age of those who have fixed internet services is 45 years and for those who don't have these services is 56 years.

Table 3. The role of age on having fixed internet. Result of the Student' t-test

Variables	t	df	P-value
Age	12.103	631	0.000

Source: produced by authors based on data from the ANCOM study

Another non-parametric test done taking into account that the analyzed variables are not normally distributed is chi-squared (table no.4).

Table 4. The role of gender on having fixed internet services. Result of the chi-squared test

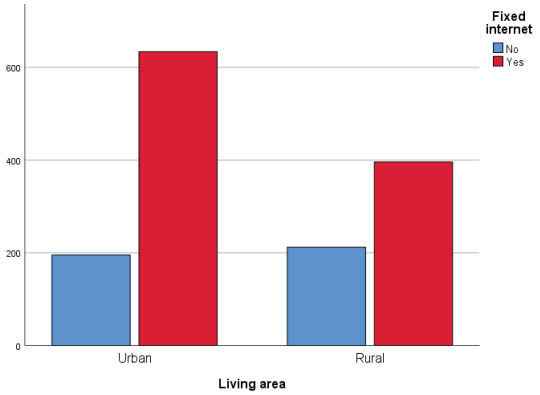
Variables	Pearson Chi-Square	df	P-value	Phi	Cramer's V
Gender	2.218	1.00	0.136	0.039	0.039

Source: produced by authors based on data from the ANCOM study

Based on the probability to reject the null hypothesis, it can be noticed that there is no association between the gender and having fixed internet.

Based on the received answers, there seems to be an association of having fixed internet services with the living area; the number of the individuals who don't have fixed internet is close in urban and rural areas, but the number of those who have fixed internet in the urban area is 3 times higher than for those who don't have these services (figure 7).

Figure 7. Fixed internet by living area



Source: produced by authors based on data from the ANCOM study

4.3. Explaining the fixed internet adoption with binary regression

The information provided by our sample can be better integrated by understanding the specific characteristics of the Romanian electronic communications and particularly fixed internet services. By one hand, according to the Digital Economy and Society Index (DESI) indicators 2023 (that refer to previous years 2022 or 2021, according to the availability of the data), the internet use in Romania was below the European average and basic

skills were almost half the European average and the lowest in the European Union. By the other hand, the percentage of the IT specialists women (as % in total specialists) was 25.2% compared to 18.9% in the European Union and the ITC graduates were 6.9% of the graduates compared to 4.2% in the EU.

Furthermore, Romania ranked first places in terms of fixed very high capacity networks, being the first in fibre to the premises coverage. In Romania, the alternative operators have the highest market share which is an indication that there are competitive conditions for their development. This aspect is reflected in the most competitive prices in the EU.

In this context, our analysis provides an explanation based on the population's attitude toward fixed internet services by using a regression analysis that is appropriate to the variables. The explained variation in having fixed internet services based on our model is 30%. The results in Table 5 show that our assumptions are confirmed for most of the variables.

Results from the binary logistic regression

Table 5

Variables		B	S.E.	Sig.	Exp(B)	95% C.I. for EXP(B)	
						Lower	Upper
Age	Age in years	-0.026	0.007	0.000	0.974	0.962	0.987
Gender	Female						
	Male	-0.047	0.158	0.764	0.954	0.700	1.300
Employed	No						
	Yes	0.684	0.192	0.000	1.981	1.360	2.885
Number of persons in the household	Number of persons	0.182	0.092	0.047	1.199	1.002	1.435
Number of children in the household	Number of children	-0.148	0.142	0.299	0.862	0.653	1.140
Living area	Rural						
	Urban	0.240	0.161	0.137	1.271	0.927	1.744
Income interval	<2000 lei			0.000			
	2001-4000 lei	0.302	0.250	0.227	1.353	0.828	2.209
	4001-6000 lei	1.374	0.298	0.000	3.951	2.203	7.086
	6001+ lei	1.253	0.366	0.001	3.503	1.710	7.174
Education	Primary			0.018			
	Secondary	0.500	0.339	0.140	1.648	0.849	3.200
	Tertiary	0.968	0.383	0.011	2.634	1.243	5.580

Source: produced by authors based on data from the ANCOM study

We have found age to be a significant factor in having fixed internet services, with a negative impact; the difference between average ages of those who have fixed internet and those who don't is almost 10 years. The general literature highlights the age gap is the result of the lack of interest and lack of skills, while few other studies have identified a further divide of the elderly depending on the gender. In one of the previous studies, family support has been identified as a significant factor and we have also determined that the number of persons in the household is statistically significant in the analysis of the adoption of the fixed internet services and it increases the probability to have fixed internet services by almost 20%.

The income and the education have a positive impact on internet adoption. Higher levels of income increase significantly the probability to have fixed internet services. The education is also important and the increase in the level of education is reflected accordingly in having fixed internet services, as it has been assumed and as it was reflected in our previous figure.

At the same time, being employed doubles the odds to have fixed internet. The association between having fixed internet and the employment could be seen in both ways: by one hand the persons who are employed have the necessary income to cover the expenses with communications services and they use the fixed internet services to work from home; by the other hand, as it has been identified in one of the analyzed studies, the access to the internet is the one that determines the increase in employment related both to the higher opportunities to seek and to select the candidates and to the better skills of the candidates.

There is a wide consensus that the rural areas are a challenge and this paper also shows that the urban areas include more persons with access to the fixed internet than rural areas, but this is not significant in our model. One of the reasons identified in previous studies could be the low satisfaction with the quality of the services, but this aspect is applicable to a lower degree in Romania where the respondents declare that they are mostly satisfied and very satisfied with the fixed internet, with low percentage differences between urban and rural.

The gender of the respondents is not a significant variable and therefore it doesn't influence having fixed internet services. In Romania, where the percentage of female ITC specialists is above the European average, the expectations regarding such a difference were even lower.

5. CONCLUSIONS

While the fixed internet deployment has advanced considerably in the latest years, there are still barriers to the access to these services that could be in both sides of the market (the offer and the demand). Our work has focused on the demand side and on the investigation of the relevant sociodemographic factors because in the end the population is the one who uses the services. The adoption of fixed internet services is a complex process that depends on circumstances such as availability, interest, skills, income and support from the family. Our results have clarified the association between the analyzed variables and having fixed internet services, but in some cases it is limited with respect to the underlying reasoning. Additional research could be employed in order to identify the cross-references between the variables or maybe the national or cultural specificities. The research advances constantly with the availability of the data and finds additional evidence to support or to revise previous assumptions.

Our analysis has confirmed the positive impact of the income, the education, the family size and the employment on having fixed internet services and the negative impact of the age, while there is no significance related to the living areas and to the gender.

Taking into account the high relevance of the fixed internet services for the decision factors in Romania determined by the fact that the digitalisation is a driver for the development of other economic sectors, it is utterly important to understand the areas where public policies could improve the adoption of the fixed internet services and what are the measures that can be employed to incentivize its development. The aim should be in fact to make some of the factors irrelevant in the future, while setting appropriate incentives for others. A suitable approach could be to ensure the affordability of the services for all the users and indifferent to the living area. The measures could be aimed at a wide availability of the services that decreases the unit cost or at appropriate subsidies. Additionally, taking into account that previous studies have shown that the behavioral intention is not the underlying factor for the use of internet services by the elderly population, the incentives to adopt the internet could be oriented towards factors such as the increase usefulness, ease of use and even enjoyment by manners that could circumvent the reluctance to embrace innovation. This is the more relevant considering the education has been found to be a significant factor in adopting fixed internet services. Being a long term investment, timely action is advisable in order to keep the pace with technology developments.

The private sector actors can also benefit from the results of the study and identify the possible strategies that could be applied in order to increase

the adoption of these services by narrowing the gap between the different categories of potential users.

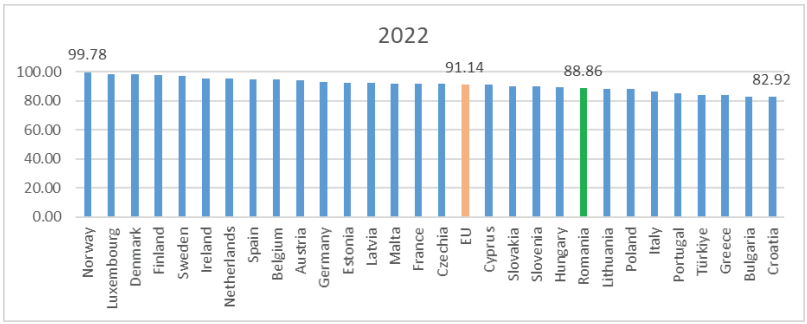
References

1. Alam, K., Mamun, S. (2017) "|Access to broadband Internet and labour force outcomes: A case study of the Western Downs Region", Queensland, Telematics and Informatics, 34(4): 73–84
2. Andrei, T., Bourbonnais, R. (2008) "Econometrie,", Editura Econimcă
3. Bhuller, M., Kostol, A., Vigtel, T. (2019) "How Broadband Internet Affects Labor Market Matching." Available at SSRN: https://ssrn.com/abstract=3507360
4. Cazacu, M., Țițan, E., Manea, D., Mihai, M. (2021) "Offensive Strategy Approach of Aging Population in the Context of a Digital Society", Romanian Statistical Review nr. 3/ 2021
5. Cazan, A-M., Cocoradă, E., Maican, C. I. (2016) "Computer anxiety and attitudes towards the computer and the internet with Romanian high-school and university students", ScienceDirect,
6. https://doi.org/10.1016/j.chb.2015.09.001
7. DiMaggio, P., Hargittai, E., Celeste, C., Shafer, S.(2001) "From Unequal Access to Differentiated Use: A Literature Review and Agenda for Research on Digital Inequality", Report prepared for the Russell Sage Foundation
8. Di Stefano, G, Gambardella, A., Verona, G. (2012) "Technology push and demand pull perspectives in innovation studies: Current findings and future research directions", ScienceDirect, https://doi.org/10.1016/j.respol.2012.03.021
9. Durkee et al. (2012) "Prevalence of pathological internet use among adolescents in Europe: Demographic and social factors", PubMed
10. Durndell, A., Haag, Z. (2002) "Computer self efficacy, computer anxiety, attitudes towards the Internet and reported experience with the Internet, by gender, in an East European sample", ScienceDirect, https://doi.org/10.1016/S0747-5632(02)00006-7
11. Martinez-Dominguez, M., Mora-Rivera, J. (2020) "Internet adoption and usage patterns in rural Mexico", ScienceDirect, https://doi.org/10.1016/j.techsoc.2019.101226
12. Steven, P.M., Robinson, J.P.(2007) "The Income Digital Divide: Trends and Predictions for Levels of Internet Use", Social Problems, vol. 54, no. 1, 2007, pp. 1–22
13. Park,S.(2017) "Digital inequalities in rural Australia: A double jeopardy of remoteness and social exclusion", Journal of Rural Studies, Volume 54, 2017, Pages 399-407, ISSN 0743-0167, https://doi.org/10.1016/j.jrurstud.2015.12.018.
14. Ramon-Jeronimo, M. A., Peral-Peral, B and Arenas-Gaitan, J. (2013) "Elderly Persons and Internet Use", Social Science Computer Review, https://doi.org/10.1177/0894439312473421
15. Kyung Yong Rhee, Wang-Bae Kim (2004) "The Adoption and Use of the Internet in South Korea", *Journal of Computer-Mediated Communication*, Volume 9, Issue 4, 1 July 2004, JCMC943, https://doi.org/10.1111/j.1083-6101.2004.tb00299.x
16. Tsetsi, E., Rains, S. (2017) "Smartphone Internet access and use: Extending the digital divide and usage gap.", Mobile Media & Communication. 5. 205015791770832. DOI: 10.1177/2050157917708329
17. Van Deursen, A., Van Dijk, J. (2011) "Internet skills and the digital divide", Sage Journals
18. Wasserman, I., M., Richmond-Abbott, M. (2005) "Gender and the Internet: Causes of Variation in Access, Level, and Scope of Use", Social Science Quarterly, https://doi.org/10.1111/j.0038-4941.2005.00301.x

19. Whitacre, B., Strover, S., Gallardo, T. (2015) "How much does broadband infrastructure matter? Decomposing the metro–non-metro adoption gap with the help of the National Broadband Map", Government Information Quarterly, Volume 32, Issue 3, 2015, Pages 261-269, ISSN 0740-624X, https://doi.org/10.1016/j.giq.2015.03.002
20. The European Higher Education Area in2024 - Bologna Process Implementation Report

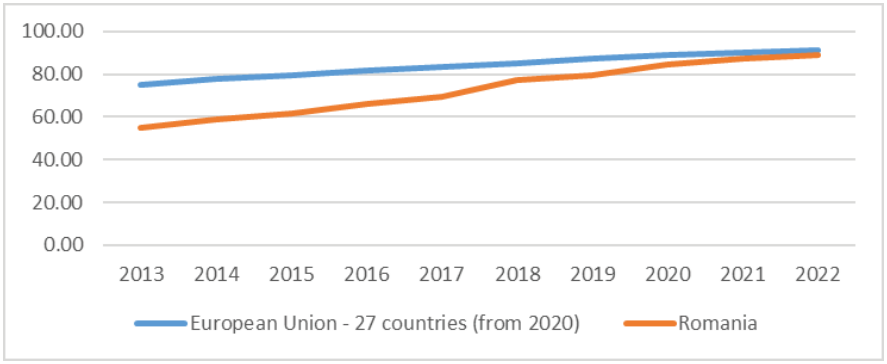
Annex 1

Figure 8. Fixed internet utilization at the European level in 2022



Source: produced by the authors based on Eurostat data

Figure 9. Fixed internet utilization evolution



Source: produced by the authors based on Eurostat data

Annex 2

Table 6. Internet data for Romania for 2022

Internet	Number of connections (millions)	Annual traffic (thousands PB)	Average monthly traffic per inhabitant (GB)	Connections in total	Traffic in total
Fixed internet	6.36	15.7	69	23%	89%
Mobile internet	21.37	1.85	8.1	77%	11%
Total	27.73	17.55	77.1	100%	100%

Source: produced by authors based on data from the ANCOM study

Characteristics of Purchasing Behavior of Food Items by Region Contained in the "Family Income and Expenditure Survey" Data

Atsushi Kimura (atkimura@nstac.go.jp, kimura@ta2.so-net.ne.jp)
National Statistics Center, Japan

ABSTRACT

This paper shows for the first time that regional characteristics of food purchasing behavior in Japan can be visualized by cluster analysis of published data from the "Family Income and Expenditure Survey" conducted by Statistics Bureau of Japan. We also show that these regional characteristics are common and stable structures by analyzing data from several different years of the household survey.

The "Family Income and Expenditure Survey" data contains no geographic information such as regional adjacency or distance information between regions, and the data set consists purely of purchase amounts. However, as this paper will show, regional characteristics are inherent in the food item purchasing behavior of Japanese households.

We also discuss the possibility of applying this stable regional characteristic inherent in the "Family Income and Expenditure Survey" to the Analyse phase (GSBPM: Analyse phase (6.2 Validate outputs sub-process)).

Finally, we introduce several new methods that are useful in interpreting the results of different survey-year data analyses. These methods make it possible to quantify the contribution of each variable to the dissimilarity between agglomerating clusters in a multivariate cluster analysis. It also greatly reduces the researcher's effort when intercomparing the results of analyses of data from different survey years.

Keywords: *Food item purchasing behavior, regional characteristics of food item purchasing behavior, LIV method, cluster agglomeration contribution, Ward method hierarchical cluster analysis, dissimilarity*
JEL Classification: *JEL: D10, C38*

1. INTRODUCTION

Today, the proliferation of social networking services such as YouTube and Instagram have made it possible to access the same quality of information in real time from anywhere in Japan using a smartphone. In addition, with the

use of various e-commerce platforms such as Amazon, we have entered an era of location-free shopping, where one can easily obtain the products, they want from anywhere in Japan. In this era, trends in fashion, music, movies, etc. can spread throughout Japan in the blink of an eye. However, even in this era, it is well known empirically that regional differences exist in daily food preferences in Japan. The existence of diverse and rich food culture in each region is also well known.

In this paper, we show that the regional characteristics of food purchasing behavior in Japan can be visualized by performing a Ward's method hierarchical cluster analysis of food purchase data by region extracted from the published data of the "Family Income and Expenditure Survey" conducted by Statistics Bureau of Japan. We also show that these regional characteristics are common and stable across different years of the "Family Income and Expenditure Survey" data.

The "Family Income and Expenditure Survey" data does not include any geographical information such as regional adjacencies or distances between regions, and the data set consists purely of purchase amounts. However, as this paper will show, regional characteristics are inherent in the food item purchasing behavior of Japanese households. By taking advantage of these characteristics, we also discuss the applicability of the GSBPM (Analyse phase (6.2 Validate outputs sub-process)) to analytical validation work.

In addition, we introduce a new method that is useful for interpreting the results of different survey-year data analyses. It is the LIV method, which makes the characteristics between clusters visually comprehensible. In cluster analysis using multivariate data, it is useful to understand the contribution of each variable in the differences between agglomerating clusters. It can also greatly reduce the researcher's effort when intercomparing the results of analyses of data from different survey years.

In this research, we used the R language because it allows us to use a variety of mathematical and statistical functions and is excellent at ensuring reproducibility. The dataset and R scripts used for the analysis in this paper can be freely downloaded and verified from the author's GitHub. (https://github.com/ibuchichi/R_function_2024.git)

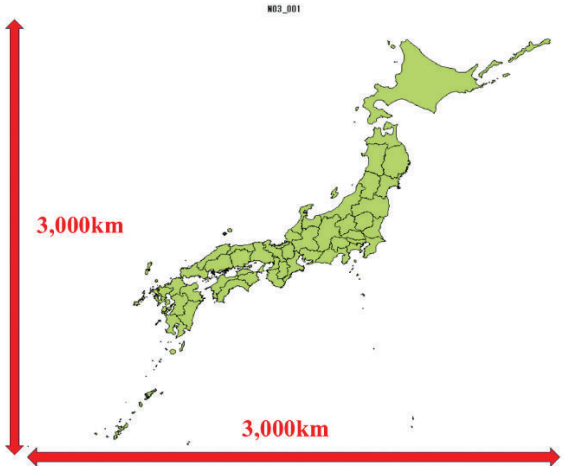
2. DATA AND METHODS USED IN THE ANALYSIS OF JAPANESE FOOD CULTURE

2.1 About Japan

Japan is an island nation stretching approximately 3,000 km from north to south and from east to west (Figure 1). It is well known that there are various local cuisines in Japan that are rich in regional characteristics, using

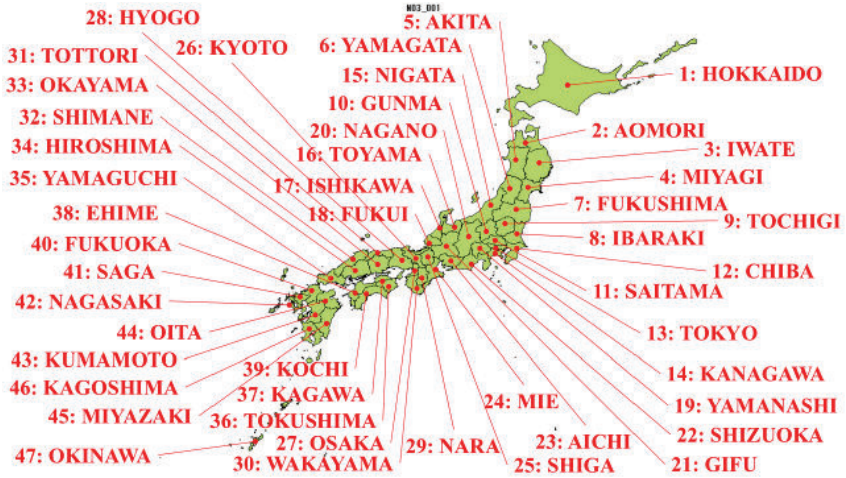
a variety of ingredients and seasonings according to the local region, and that there is a regional character to the food culture.

Figure 1: Size of the Japanese archipelago



Japan consists of a total of 47 prefectures, from Hokkaido in the north to Okinawa Prefecture in the south (Figure 2).

Figure 2 Japan is divided into 47 prefectures (The number in front of the prefecture name indicates the row number in the data set used for the analysis.)



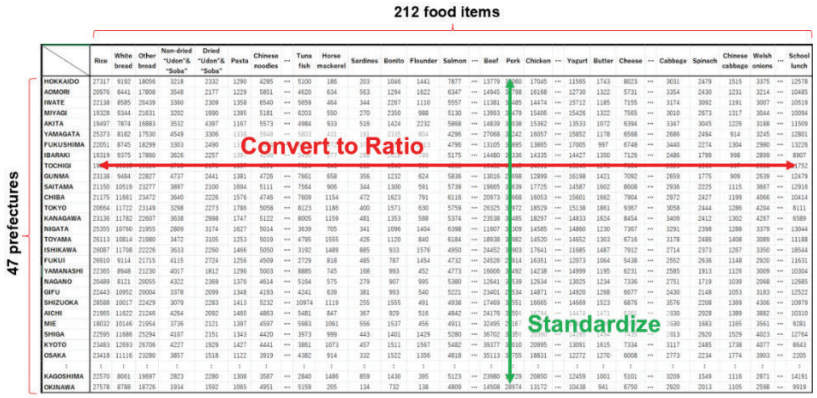
2.2. About the "Family Income and Expenditure Survey"

The "Family Income and Expenditure Survey" is one of Japan's official statistics compiled and published by the Statistics Bureau of Japan. The survey targets approximately 9,000 households (randomly selected from across the country using statistical methods) in 168 cities, towns, and villages nationwide. Survey households fill out a "household account book" regarding their daily income and expenditure. Online surveys are also conducted using computers and smartphones. Consumption expenditure is tallied by item. The total number of income and expenditure items is approximately 550. In this analysis, we will extract and use items related to food item expenditures from the "Family Income and Expenditure Survey" data. (https://www.stat.go.jp/english/data/kakei/index.html)

2.3 Data preprocessing and analysis methodology

We extract and analyze data on the amount spent on 212 food items for each of the 47 prefectural governments and cities (Figure 3).

Figure 3: Dataset used for analysis in this paper



Specifically, the data set consists of "households of two or more persons" from the Family Income and Expenditure Survey by prefecture and city, by item (all food items), and by annual expenditure per household (average value from 2020 to 2022).

The published values for the capitals of the 47 prefectures will be regarded as representative values for all 47 prefectures. Furthermore, food purchase expenditures were converted into ratios for each prefecture and standardized for each of the 212 food items (average 0, standard deviation 1). This is a preprocessing step to treat the variation among food items equally.

Using this data, we performed Ward's hierarchical cluster analysis (Ward, 1963), with the prefecture as the smallest unit.

3. ANALYSIS RESULTS

3.1. Analysis results and structural stability

Figure 4 Shows a tree diagram of the analysis results. The red boxes indicate the group when 12 clusters were aggregated.

Figure 4 Dendrogram

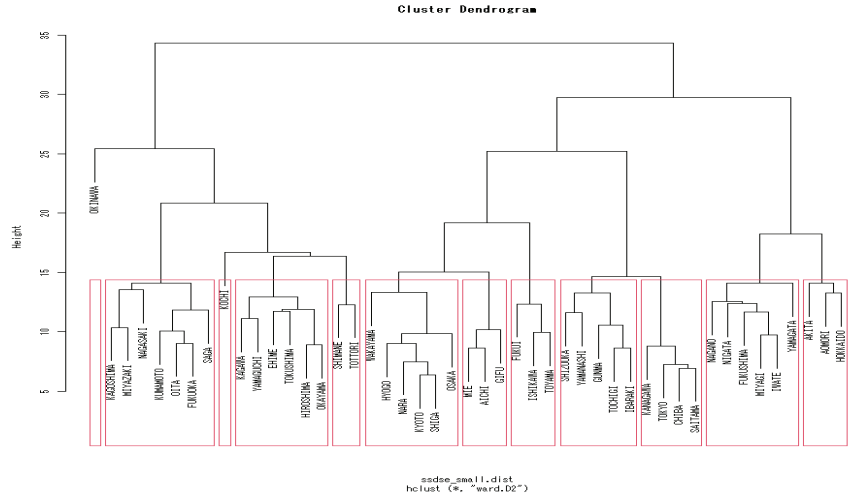
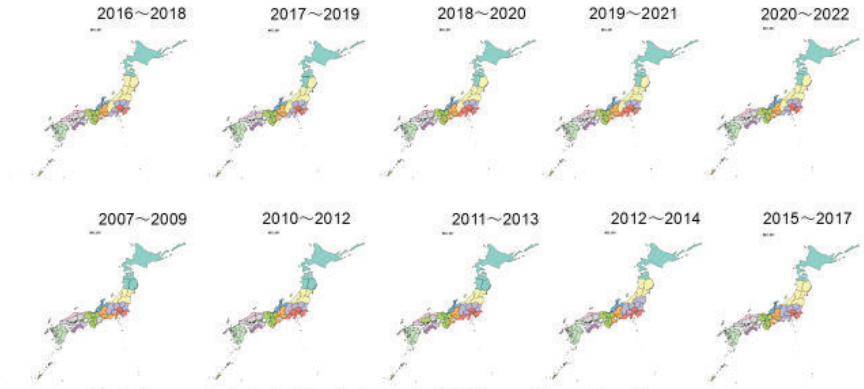


Figure 5 Shows the Japan region classification chart for the case of 12 clusters.

Figure 5: Map of Japan with 12 clusters

The data from the "Family Income and Expenditure Survey" is the value of purchases for each food item. It does not include any geographical information such as prefectural adjacencies as numerical information. Nevertheless, the emergence of geographically proximate cluster structures such as this one indicates the existence of unconscious regional characteristics in the food purchasing behavior of Japanese households. For verification purposes, we conducted a similar analysis using the "Family Income and Expenditure Survey" data (food items) from different survey time periods (2007-2016), and also confirmed that a regional cluster diagram similar to the present one is stably obtained, although some boundary fluctuations are observed (Figure 6).

Figure 6 Comparison of analysis results using household survey data from different survey years



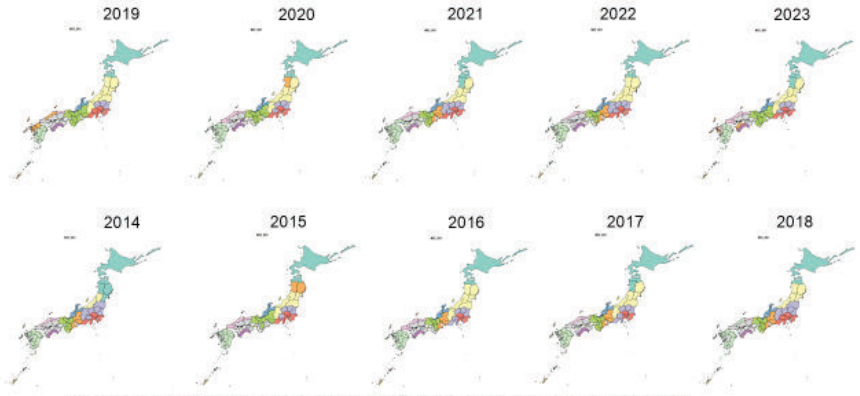
All clusters were indicated with a cluster count of 12. The results are almost the same.

Therefore, it is fair to say that this is a valid, stable, and essential regional classification that indicates the regional nature of daily food preferences.

As far as the author knows, research into regionality based on household survey data includes Yamashita (1992) who used 32 food items from the 1963 and 1990 the "Family Income and Expenditure Surveys" to create a map of Japan's regions using the Ward method. Fujino (2015) also used the Ward method to create a dendrogram using 30 food items from the 2009 the "National Survey of Family Income, Consumption and Wealth" (https://www.stat.go.jp/english/data/zenkokukakei/index.html), and then used the k-means method to create a map of Japan's regions. However, in neither of these studies was the geographically close clusters seen in this study. This is likely due to the fact that the number of food items used in the analysis was orders of magnitude smaller than in this study.

The most important difference between the results of this study and previous studies is that all previous studies used only a coarse set of around 30 food categories. In this study, 212 food items were used for classification, and we believe that the fact that the information indicating regional characteristics was used in the analysis with good resolution without omission is a major factor that brought about the beautiful results of this analysis.

Figure 7 Comparison of analysis results using the "Family Income and Expenditure Survey" data for a single year

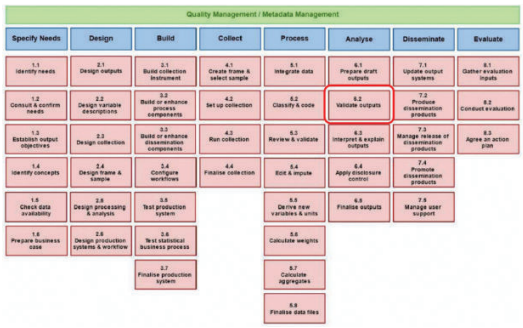


All clusters were indicated with a cluster count of 12. The results are almost the same.

The results of the analysis using data for a single year are shown below (Figure 7). It can be seen that regional characteristics are generally preserved in the single-year data.

3.2. Apply these regional characteristics to the GSBPM6.2 process

Figure 8. 6.2Validation outputs of GSBPM

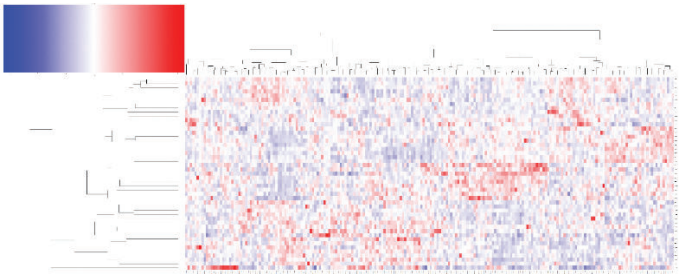


These regional characteristics of grocery purchasing behavior could be utilized in the GSBPM 6.2 (Validate outputs) process. If the results of the cluster analysis using the generated result table data do not show similar regional characteristics, it is possible that the result table data is incorrect. In this case, it can be said that there is a need for reconfirmation in the process of creating the result table data. (https://unece.org/sites/default/files/2023-11/GSBPM%20v5_1.pdf)

3.3 A method for calculating the contribution of food items to cluster agglomeration

In order to reconfirm the data in the process of creating the result table data, it is effective to understand which variables contributed strongly to cluster agglomeration. Although decision trees can be used as a method to identify the contribution of variables to cluster agglomeration, the algorithms for decision trees and hierarchical cluster analysis are inherently different and do not always yield highly convincing interpretations. The 2-way heat map is also used to understand the overall trend of the original data (Figure9), but its drawback is that it does not directly show the relationship between cluster agglomeration and observed variables, and in addition, it is difficult to understand intuitively when the number of variables in the original data is large.

Figure 9. 2-way heat map for the results of this analysis



The basic idea of the new "method for calculating the contribution rate by variable" proposed in this paper is to focus on the aggregation algorithm of the Ward method (Ward, 1963), which is often used in hierarchical cluster analysis, and to calculate the degree of contribution of each variable to the aggregation and division of clusters as the "Loss of Information rate by Variable (LIV). The LIV method is a method to quantitatively evaluate the degree of contribution of each variable to the aggregation and division of clusters using a numerical value named "Loss of Information rate by Variable (LIV)". Below is a brief explanation of the concept of the LIV method and the Ward method as a prerequisite knowledge.

3.3.1. Agglomeration algorithm of the Ward method

The algorithm for cluster agglomeration in the Ward method (Ward, 1963) is briefly described below. For the *i*-th cluster *i* in the agglomeration process, let *Ni* be the number of elements, $\boldsymbol{x_{ij}}$ be the *j*-th element vector of

cluster i, and $\boldsymbol{c_i}$ be the center of gravity vector of cluster i. If the number of observed data variables in each element is m, then both $\boldsymbol{x_{ij}}$ and $\boldsymbol{c_i}$ are m -dimensional vectors.

$$\boldsymbol{x_{ij}} = (x_{ij1}, x_{ij2}, \ldots, x_{ijm}), \quad \boldsymbol{c_i} = (c_{i1}, c_{i2}, \ldots, c_{im}) \qquad [1]$$

The ESS (Error Sum of Squares) of cluster i is defined as follows.

$$ESSi = \sum_{j=1}^{Ni} ||\boldsymbol{x_{ij}} - \boldsymbol{c_i}||^2 \qquad [2]$$

where$||x||$ is the length of vector $\boldsymbol{x}$ (Euclidean distance).

Ward (1963) considers the amount of change in ESS to be the amount of "loss of information" associated with the formation of the cluster, and proposes that the pair with the smallest increase in "loss of information" associated with agglomeration be selected as the agglomeration target. The amount of increase in "loss of information" when clusters p and q aggregate into cluster r is as follows.

$$ESSr - (ESSp + ESSq) = \sum_{j=1}^{Nr} ||\boldsymbol{x_{rj}} - \boldsymbol{c_r}||^2 - (\sum_{j=1}^{Np} ||\boldsymbol{x_{pj}} - \boldsymbol{c_p}||^2 + \sum_{j=1}^{Nq} ||\boldsymbol{x_{qj}} - \boldsymbol{c_q}||^2) = \{Np \times Nq / (Np + Nq)\} \times ||\boldsymbol{c_p} - \boldsymbol{c_q}||^2 \qquad [3]$$

3.3.2. Basic concept of the "Variable-Specific Contribution Calculation Method"

If the degree of contribution of each of the m observed data variables to the increase in "information loss" can be clarified, it is possible to determine the variables that affect cluster agglomeration and separation. For this purpose, we define a new value called "variable-specific information loss contribution rate (LIV). In the equation that expresses the amount of increase in "information loss," it is the "square distance term" that is affected by the values of the m variables in the observed data. If we extract only the "square distance term" from eqn[3], we obtain the following equation

$$||\boldsymbol{c_p} - \boldsymbol{c_q}||^2 = \sum_{i=1}^{m} (c_{pi} - c_{qi})^2 \qquad [4]$$

However,

$$\boldsymbol{c_p} = (c_{p1}, c_{p2}, \ldots, c_{pm}), \quad \boldsymbol{c_q} = (c_{q1}, c_{q2}, \ldots, c_{qm}) \qquad [5]$$

Let us define "Loss of Information rate by Variable (LIV)" as the ratio of the contribution of a particular variable among the m variables to the "distance between centers of gravity" that constitutes the "square distance term," and let us define it by the following eqn[6]. The LIV value of the n-th variable is denoted as LIV_n. $\boldsymbol{e_n}$ is the basis vector of variable n, and " $\cdot$ " is the inner product symbol.

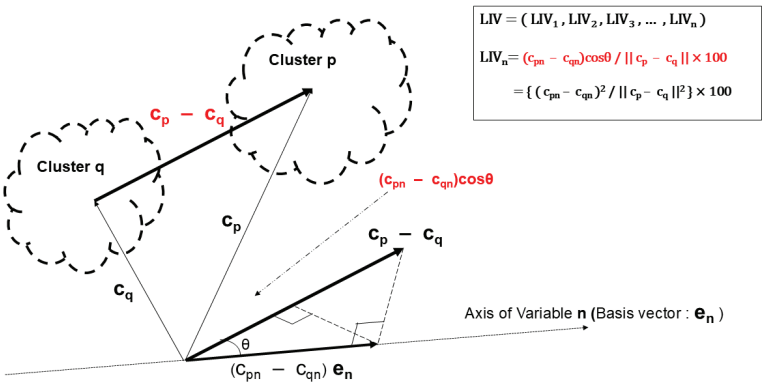
$$LIV_n = [\{(c_{pn} - c_{qn})\, \boldsymbol{e_n}^{\cdot}\, (\boldsymbol{c_p} - \boldsymbol{c_q})\} / ||\boldsymbol{c_p} - \boldsymbol{c_q}||^2] \times 100$$
$$= \{(c_{pn} - c_{qn})^2 / ||\boldsymbol{c}_{\cdot\cdot} - \boldsymbol{c_q}||^2\} \times 100$$
$$= \{(c_{pn} - c_{qn})^2 / \textstyle\sum_{i=1}^{m} (c_{pi} - c_{qi})^2\} \times 100 \qquad [6]$$

Also, the sum of all variables of LIV values is,

$$\textstyle\sum_{n=1}^{m} LIV_n = 100. \qquad [7]$$

To put the meaning of the LIV_n value in a simple and straightforward manner, it is "the contribution (%) of variable n to the 'distance between centers of gravity' of the cluster pair under agglomeration" (Figure 10).

Figure 10 Conceptual diagram of the LIV method



When the LIV_n value is large, the degree to which variable *n* contributes in the direction of increasing the "square distance term" is high. In other words, variable *n* with a large LIV_n value contributes to making clusters *p* and *q* less agglomerated. In other words, variable n with a large LIV_n value is a major factor that characterizes clusters *p* and *q*. When LIV_n is large, the following five patterns (hereinafter referred to as "LIV patterns") exist ($c_{pn} > c_{qn}$).

$c_{pn} > c_{qn} > 0$: cluster *p* has stronger positive characteristics

$c_{pn} > c_{qn} \doteqdot 0$: cluster *p* has positive characteristics

$c_{pn} > 0 > c_{qn}$: cluster *p* has positive features and cluster *q* has negative features

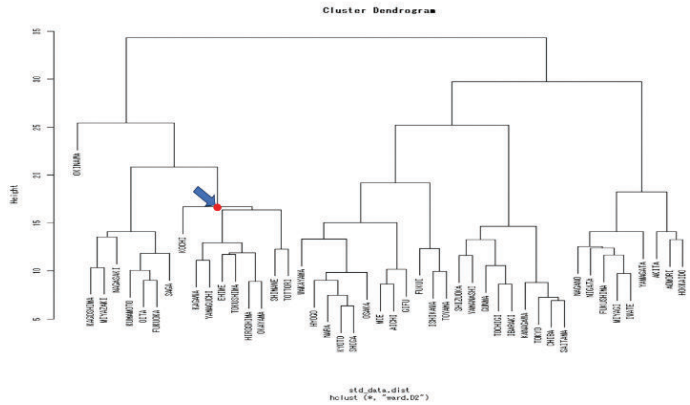
$0 > c_{pn} > c_{qn}$: cluster *q* has stronger negative characteristics

As with the Ward method, the LIV method can be applied to the center-of-gravity and median methods that use the square distance between cluster centers of gravity.

3.3.3 Specific application of the variable-specific contribution ratio calculation method

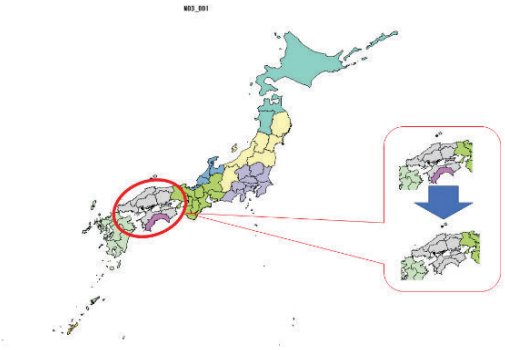
Let us apply the "Variable-Specific Contribution Ratio Calculation Method" to the actual cluster analysis results. Let us calculate the LIV value for the agglomeration process marked with a red circle in the dendrogram of the analysis results.

Figure 11 Cluster agglomeration process for which LIV values are calculated this time



The agglomerated clusters circled in red are shown on the map of Japan as follows. The gray and purple clusters indicated by the red ovals are the agglomeration process.

Figure 12. Agglomerated clusters for calculating LIV values



The LIV values in this agglomeration process are calculated and plotted in a Pareto plot (the lower part is omitted; the same applies hereafter). Figure 13 shows a Pareto plot of the LIV values calculated during the agglomeration process (the lower side is omitted; the same applies hereafter).

The graph in the lower half of the figure shows a Pareto plot of LIV values. The horizontal axis of the graph shows the names of 212 food items, the left vertical scale is the LIV value of each food item (bar graph), and the right vertical scale is the cumulative LIV value (line graph). The food items on the left side of the Pareto chart contribute more to the dissimilarity among the agglomeration clusters. An enlarged version is shown in the upper left of the figure. The food item with the largest contribution is the fish "skipjack. In fact, skipjack is a fish that is eaten in Kochi Prefecture, which is in the purple cluster.

Figure 13 Pareto chart of calculated LIV values

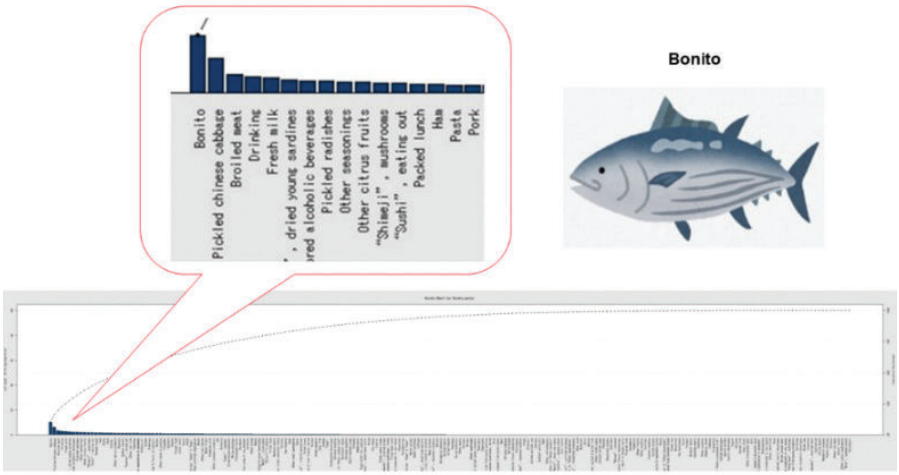


Figure 14. The hclust_LIV function

LIV Calculation : hclust_LIV()

hclust_LIV() function

usage:
hclust_LIV(std_data, LIV_matrix)
input: std_data
output: LIV_matrix

example:
# Draw LIV pareto chart
Library(qcc)
LIV_matrix <- hclust_LIV(std_data)

for(i in 1: nrow(LIV_matrix)){
Pareto_vector <- LIV_matrix[i,]
names(Pareto_vector) <- colnames(LIV_matrix)
png_Pareto_file_name <- paste("LIV_",i ,"_pareto_qcc.png", sep="")
png(png_Pareto_file_name, width=3800, height=770)
ylab_words <- paste("LIV about ",i ,"th agglomaration")
pareto.chart(Pareto_vector, ylab = ylab_words)
dev.off()
}

3.3.4. Useful tools for comparing the results of different survey year data analyses

Consider the case where multivariate data measured at different time points exist for the same observation target/variable. One of the data at multiple time points is defined as “reference data”, and the data at other time points are called “target data”. If the number of observations is *N* and the

number of variables is *P*, all multivariate data in this case can be represented by an *N*×*P* matrix with different element values. Let *RefD* denote the matrix of "reference data" and *D* the matrix of one of the "target data".

3.3.4.1. Ordered tuplet representation

When we divide *N* objects into *K* clusters, it is equivalent to divide a set of *N* objects into *K* subsets. Let *Y* be a set of *N* objects, and we assign an identification number to each of the *N* objects, a natural number from *1* to *N*.

In this case, it can be expressed as

$Y = \{1, 2, 3, \ldots, i, \ldots, j, \ldots, N\}$ [8]

Let sets *A* be the partition of set *Y* into *K* pieces by cluster analysis of multivariate data *RefD*, and let sets *B* be the partition of set *Y* into *K* pieces, also using multivariate data *D*.

Let A_i ($i=1$ to K) denote the partitioned sets that are elements of sets *A* and B_i ($i=1$ to K) denote the partitioned sets of sets *B*. Then it can be expressed as

$A = \{A_1, A_2, A_3, \ldots, A_i, \ldots, A_j, \ldots, A_K\}$, $B = \{B_1, B_2, B_3, \ldots, B_i, \ldots, B_j, \ldots, B_K\}$ [9]

In this case, it holds that

$A_i \cap A_j = B_i \cap B_j = \varphi\ (i \neq j)$, $Y = \bigcup_{i=1}^{K} A_i = \bigcup_{i=1}^{K} B_i$ [10]

Let ${}_nA$ be one of the ordered tuplets consisting of $A_1, A_2, A_3, \ldots, A_i, \ldots, A_j, \ldots$,and A_K. The total number of such ordered tuplets exists as *K!*. Similarly, let ${}_mB$ be an ordered tuplet consisting of $B_1, B_2, B_3, \ldots, B_i, \ldots, B_j, \ldots$,and B_K. The total number of such ordered tuplets also exists as *K!*.

Let us write the element sets of the ordered tuplets as ${}_nA_i$ ($i=1$ to K) and ${}_mB_i$ ($i=1$ to K), respectively.

${}_nA = ({}_nA_1, {}_nA_2, {}_nA_3, \ldots, {}_nA_i, \ldots, {}_nA_K)$, ${}_mB = ({}_mB_1, {}_mB_2, {}_mB_3, \ldots, {}_mB_i, \ldots, {}_mB_K)$ [11]

In this case, it can be expressed as

${}_nA_i \cap {}_nA_j = {}_mB_i \cap {}_mB_j = \varphi\ (i \neq j)$, $Y = \bigcup_{i=1}^{K} {}_nA_i = \bigcup_{i=1}^{K} {}_mB_i$ [12]

for this as well.

Also, if $|\ |$ is defined as the cardinality of a set, then

$N = \sum_{i=1}^{K} |\ {}_nA_i\ | = \sum_{i=1}^{K} |\ {}_mB_i\ |$. [13]

Identification of similar clusters is equivalent to selecting one specific ordered tuplet ${}_mB$ for a given ${}_nA$ according to a defined index and method from among the *K!* of ordered tuplet ${}_mB$.

The Jaccard coefficient (Jaccard,1901), which indicates the similarity between sets A and B, is defined by

$$| A \cap B | / | A \cup B | \qquad [14]$$

If we set ${}_mJAC_{ij}$, the Jaccard coefficient of ${}_{Ref}A_i$ and ${}_mB_j$ then

$${}_mJAC_{ij} = | {}_{Ref}A_i \cap {}_mB_j | / | {}_{Ref}A_i \cup {}_mB_j |. \qquad [15]$$

We shall also try an approximate method using the Jaccard coefficients, which are often used as set similarity. As an algorithm for cluster identification, we will identify clusters ${}_{Ref}A_i$ and ${}_mB_j$ in the same way, starting from the largest element of ${}_mJAC_{ij}$. Since this method is also an approximate method, the minimum "cluster difference degree" is not necessarily guaranteed. Instead, the computational complexity is on the order of K squared and can be solved in polynomial time.

Figure 15 Color scheme needs to be changed

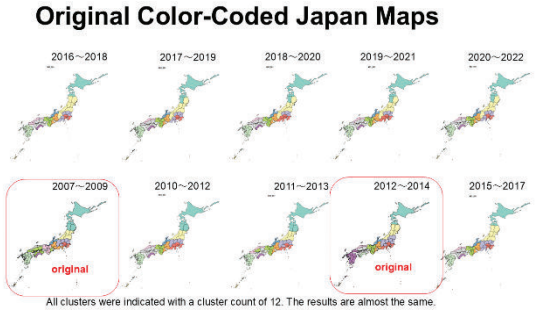


Figure 16 Color scheme replacement

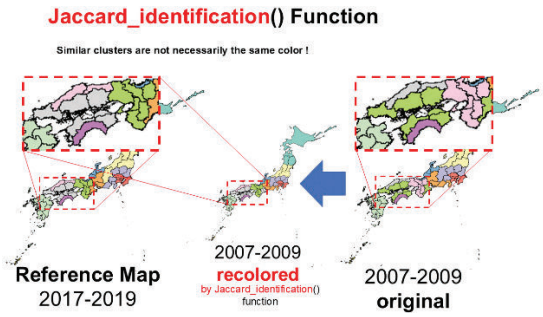
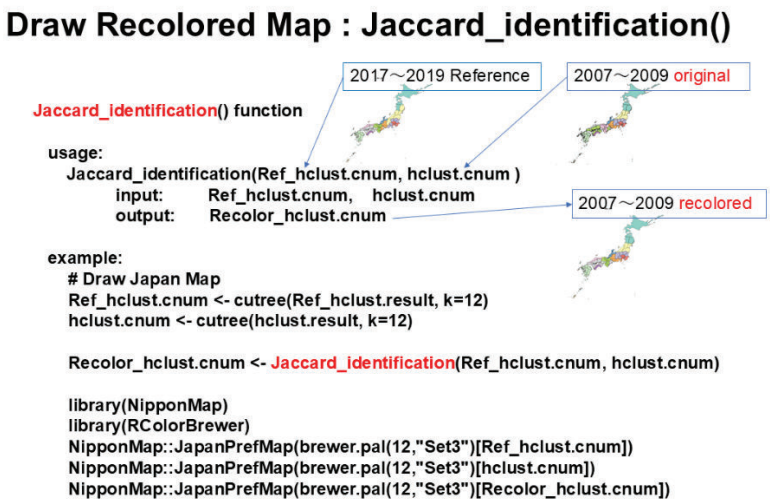


Figure 17 Jaccard_identification



4. CONCLUSION

In this paper, we have shown for the first time that regional characteristics of food item purchasing behavior in Japan can be visualized by cluster analysis of published data from the “Family Income and Expenditure Survey” conducted by Statistics Bureau of Japan. We also show that these regional characteristics are common and stable structures by analyzing data from several different years of the “Family Income and Expenditure Survey”.

The “Family Income and Expenditure Survey” data contains no geographic information such as regional adjacency or distance information between regions, and the data set consists purely of purchase amounts. However, as this paper will show, regional characteristics are inherent in the food item purchasing behavior of Japanese households.

We also point out that it is effective to apply the stable regional characteristics inherent in the “Family Income and Expenditure Survey” to the Analyse phase (GSBPM: Analyse phase (6.2 Validate outputs sub-process)).

We proposed the “variable-by-variable contribution ratio calculation method (LIV method)” for quantitative interpretation of the results of the Ward method hierarchical cluster analysis, and demonstrated its effectiveness by actually applying it to the results of the present analysis.

We also introduced a method that greatly reduces the researcher’s

effort when intercomparing the results of analyses of data from different survey years.

You can download the dataset and R script to verify the analysis of this paper from the author's GitHub. (https://github.com/ibuchichi/R_function_2024.git)

If you place the Japanese household survey data starting with "FIES-", the Shapefile of the Japanese map starting with "Japan_", and the sample script starting with "uRos2024_" in the same directory and run the sample script, it will generate a cluster analysis of the household survey data for the year specified in the script, plot the results, and generate the LIV values corresponding to all the aggregation processes and their Pareto charts.

References

1. Fujino, T., (2015), "Chapter8: Cluster Analysis", Yamamoto,Y. Fujino,T. Kubota,T., Co-authored, "Introduction to Data Mining with R", Ohm-sha, Japan, p108-112 (In Japanese)
2. GSBPM ver5.1, UNECE, https://unece.org/sites/default/files/2023-11/GSBPM%20v5_1.pdf
3. Jaccard, P.,(1901), "Distribution de la flore alpine dans le Bassin des Dranses et dans quelques régions voisines", Bulletin de la Societe Vaudoise des Sciences Naturelles · January 1901
4. Ward, J. H.,(1963) "Hierarchical Grouping to Optimize an Objective Function", Journal of the American Statistical Association, Vol. 58, No. 301(Mar., 1963), pp.236-244
5. Yamashita, M.,(1992). "Regional characteristics of food culture and its transformation in Japan", J. Fac. Edu. Saga. Univ. Vol.39, No.2(I)-1(1992) p115-133 (In Japanese)

- GitHub, ibuchichi/R_function_2024, https://github.com/ibuchichi/R_function_2024.git
- "Family Income and Expenditure Survey", Statistics Bureau of Japan, https://www.stat.go.jp/english/data/kakei/index.html
- "National Survey of Family Income, Consumption and Wealth", Statistics Bureau of Japan, https://www.stat.go.jp/english/data/zenkokukakei/index.html