European Research Studies Journal
Volume XXVIII, Issue 4, 2025

pp. 1786-1794

Investigating the Relationship Between Nominal GDP and Adult HIV/AIDS Prevalence Using Machine Learning Methods

Submitted 19/10/25, 1st revision 14/11/25, 2nd revision 26/11/25, accepted 15/12/25

Anna Landowska[1], Marek Landowski[23]

Abstract:

Purpose: *The aim of this article is to analyze and model the relationship between nominal gross domestic product (GDP) and adult HIV/AIDS prevalence by machine learning methods.*
Design/Methodology/Approach: *The study utilized publicly available GDP and adult HIV/AIDS prevalence data. To achieve the research objective, machine learning and statistical regression models were used.*
Findings: *Using machine learning models and methods, it is possible to model the relationship between nominal GDP and adult HIV/AIDS prevalence. Selected indicators examining the differences between actual and predicted values indicated the best fit for the Ensemble Boosted Trees model. The relationship between nominal GDP and adult HIV/AIDS prevalence is negative and statistically significant.*
Practical Implications: *Possibilities of modeling adult HIV/AIDS prevalence and nominal GDP using machine learning models and methods.*
Originality/Value: *This article makes a significant contribution to the development of knowledge on the relationship between nominal GDP and adult HIV/AIDS prevalence. Furthermore, it demonstrates the feasibility of using machine learning methods to model this relationship.*

Keywords: *Nominal gross domestic product, HIV/AIDS prevalence, regression models.*

JEL codes: *C1, C3, E44, C8.*

Paper type: *Research article.*

Research funding: *Co-financed by the Minister of Science under the "Regional Excellence Initiative".*

[1]*Ph.D., University of Szczecin, Institute of Spatial Management and Socio-Economic Geography, Poland, e-mail: anna.landowska@usz.edu.pl;*
[2]*D.Sc. Ph.D., Maritime University of Szczecin, Faculty of Computer Science and Telecommunications, Poland, e-mail: m.landowski@pm.szczecin.pl*
[3]*D.Sc. Ph.D., Pomeranian Medical University in Szczecin, Independent Research and Biostatistics Laboratory, Department of Social Medicine, Poland, e-mail: marek.landowski@pum.edu.pl*

1. Introduction

The article presents research on the relationship between nominal gross domestic product (GDP) and adult HIV/AIDS prevalence. This topic remains relevant due to the continued rise in HIV infections worldwide. Many scientific papers deal with the impact of the HIV/AIDS epidemic on economic factors. In paper (Bidzha *et al.,* 2024), the authors analyze data covering the years 1995–2019 to examine the impact of antiretroviral treatment on the relationship between HIV/AIDS and GDP in South Africa.

They use various econometric approaches (linear and nonlinear). They also examine the relationship between HIV/AIDS and GDP by gender. The analysis and estimation of the impact of HIV/AIDS on GDP in the WHO African region is presented in paper (Kirigia *et al.,* 2002). The analysis was carried out using an econometric model and data from 45–46 countries. Article (Levi *et al.,* 2018) compares the results of implementing preventive measures such as the HIV treatment cascade (90-90-90 target) with GDP, corruption, and conflict in selected countries.

In the paper (Essig *et al.,* 2015), the authors analyze African countries based on the impact of the growth of HIV infections on the GDP per capita. They assume that the increase in infections negatively impacts human capital, and consequently, a country's GDP per capita.

The broader impact of the HIV epidemic was examined in paper (Jones-Smith *et al.,* 2007). This paper analyzed the impact of the HIV epidemic on global economic development, taking into account data from 2003. In the paper (Amiri *et al.,* 2012), the bilateral effect of HIV/AIDS mortality on GDP was examined. Data from 1990-2009 covering 44 African countries was selected for the study. Additionally, the impact of HIV/AIDS infections on the economy was discussed in the following articles (Schwefel *et al.*, 1990; Flessa *et al.*, 2009; Gonzalo *et al.*, 2009; Leidl, 1990; Lamontagne *et al.*, 2019; Canning, 2006; Bhargava *et al.*, 2008; Haakenstad *et al.*, 2019; Grima *et al.,* 2020; Nguyen *et al.,* 2022).

This article analyzes real-world data representing GDP in millions of US dollars and adult HIV/AIDS prevalence in percentage terms for 69 countries worldwide. The research was motivated by the need to verify the existence of a relationship between GDP and the number of HIV/AIDS infections.

The aim of this paper is to demonstrate the existence of a relationship between GDP and HIV/AIDS infections and to demonstrate the feasibility of modeling this relationship. Machine learning methods were used to model the relationship using regression functions of the interaction between GDP and adult HIV/AIDS prevalence. The results were compared with statistical regression models.

The rest of the article is organized as follows. Section 2 presents the data analyzed and the methods used to conduct the regression analysis, and describes the fit indices between the actual data and the forecasted data. Section 3 provides detailed results for 28 regression models and their assessment of forecast accuracy. Section 4 summarizes the research.

2. Data and Methods

The analysis used publicly available data on the nominal GDP of selected countries and their values.of adult prevalance of HIV/AIDS as a percentage of the country's population. The research data was obtained from: https://en.wikipedia.org/wiki/List_of_African_countries_by_GDP_(nominal), and https://en.wikipedia.org/wiki/HIV_adult_prevalence_rate.

Table 2 presents a list of countries and their nominal GDP in million US$ and adult HIV/AIDS prevalence in a percentage of the country's population. Depending on data availability, data covers the period of one year, 2023, 2024, or 2025.

Table 1. *Nominal GDP and adult prevalance of HIV/AIDS in the analyzed countries.*

Country	**Nominal GDP (milion US$) ***	**Adult prevalance of HIV/AIDS ****	**Country**	**Nominal GDP (milion US$) ***	**Adult prevalance of HIV/AIDS ****
Angola	80397	0.021	Kazakhstan	288406	0.004
Argentina	633267	0.0065	Kenya	136014	0.053
Australia	1775628	0.001	Latvia	43521	0.0085
Bahamas	16393	0.011	Liberia	5039	0.015
Barbados	44662	0.009	Libya	46636	0.0102
Belize	3297	0.03	Lithuania	84869	0.0045
Bolivia	45135	0.004	Malaysia	421972	0.003
Cambodia	46353	0.011	Mali	25591	0.03
Cameroon	49279	0.04	Mexico	1852723	0.0075
Canada	2142471	0.0021	Montenegro	7530	0.0012
Central African Republic	2752	0.04	Morocco	179612	0.0155
Chile	330267	0.02	Myanmar	74080	0.012
Colombia	418542	0.006	Nepal	42914	0.0092
Republic of the Congo	15720	0.0377	Netherlands	1227544	0.005
Croatia	84391	0.0014	Nigeria	188271	0.022
Djibouti	4086	0.015	Norway	483727	0.0045
Dominican Republic	124282	0.009	Panama	86260	0.013
Estonia	46763	0.002	Paraguay	44458	0.011
Ethiopia	163698	0.014	Philippines	461618	0.0019

France	3162079	0.003	Portugal	308683	0.005
Gabon	21455	0.045	Russian Federation	2008419	0.015
Gambia	2401	0.024	Saint Vincent and the Grenadines	1091	0.016
Georgia	33776	0.01	Senegal	32267	0.023
Germany	4525704	0.0025	South Korea	1839058	0.00029
Ghana	111963	0.02	Spain	1722746	0.007
Guatemala	120850	0.0201	Thailand	526411	0.0131
Guinea-Bissau	1841	0.0365	Togo	9926	0.052
Haiti	19841	0.022	Uganda	51023	0.05
Hong Kong	380812	0.0016	Ukraine	178757	0.011
Hungary	212657	0.0011	United Arab Emirates	537079	0.002
India	3912686	0.002	United Kingdom	3643834	0.0017
Indonesia	1396300	0.005	Uruguay	80962	0.011
Ireland	577389	0.0055	Venezuela	1161	0.007
Italy	2372775	0.003	Vietnam	476388	0.007
Jordan	56157	1.7E-06			

Source: *Publicly available data, available at:*
** https://en.wikipedia.org/wiki/List_of_African_countries_by_GDP_(nominal)*
*** https://en.wikipedia.org/wiki/HIV_adult_prevalence_rate*

The Spearman rank correlation between the 69 samples for nominal GDP (US$ million) and adult HIV/AIDS prevalence is negative at the -0.56 level. This relationship is statistically significant ($p < 0.001$).

In order to perform the regression analysis, the regression models included in the Matlab Regression Learner application were used, such as: regression trees, support vector machines models, ensembels of trees, neural networks, kernel approximation regression models, Gaussian process regression models, efficiently trained linear regression models. To compare the results, statistical models of regression analysis were used: linear, quadratic, exponential and logarithmic model.

Regression Learner trains regression models using various functions. Linear regression uses the Linear, Interactions Linear, and Robust Linear models. Tree regression uses the Fine Tree, Medium Tree, and Coarse Tree training models.

In regression using the support vector machines model, Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM and Coarse Gaussian SVM models are tested.

For the Efficient Trained Linear regression model, Efficient Linear Least Squares and Efficient Linear SVM models were used. For the Ensemble of regression trees type, Boosted Trees and Bagged Trees models were used.

For Gaussian process regression models, the Squared Exponential GPR, Matern 5/2 GPR, Exponential GPR, and Rational Quadratic GPR models were used.

For regression using artificial neural networks, the Narrow Neural Network, Medium Neural Network, Wide Neural Network, Bilayered Neural Network, and Trilayered Neural Network models were tested. For Kernel approximation regression, two models were used: SVM Kernel and Least Squares Regression Kernel.

All simulations, calculations and analyses were performed using Matlab tools version R2025b. To compare the results of regression analysis, standard measures were used, such as: root mean squared error (RMSE), mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R^2). These indicators are defined by formulas (1) – (4), respectively,

$$\mathrm{RMSE} = \sqrt{\textstyle\sum_{i=1}^{n}(y_i - \bar{y})^2 / n} \tag{1}$$

$$\mathrm{MSE} = \textstyle\sum_{i=1}^{n}(y_i - \bar{y})^2 / n \tag{2}$$

$$\mathrm{MAE} = \textstyle\sum_{i=1}^{n} |y_i - \hat{y}_i| / n \tag{3}$$

$$R^2 = 1 - \textstyle\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / \sum_{i=1}^{n}(y_i - \bar{y})^2 \tag{4}$$

where: y_i – true value, $\hat{y}_i$ – predicted value, $\bar{y} = \sum_{i=1}^{n} y_i / n$ – mean value, n – number of variables, and $i \in \{1, \dots, n\}$.

3. Comparison of Regression Results Obtained Using Statistical and Machine Learning Methods

This section presents the results of a regression analysis between nominal gross domestic product (GDP) data and adult HIV/AIDS prevalence.

When analyzing the regression results for models using machine learning methods, taking into account the root mean squared error (RMSE) for the validation sample, the best result was obtained for the Ensemble Boosted Trees model. For this model, the RMSE value is the lowest, and is 0.011626. Model #16 Ensemble Boosted Trees has the best results in all four metrics. It has the lowest MSE = 0.000135 and MAE = 0.008303 values, and the highest R^2 = 0.252472.

Table 2 presents the results of regression analysis comparing the errors (RMSE, MSE, MAE, R^2) obtained for the validation data. Twenty-nine different regression models were used in the study.

Table 2. *Comparison of simulation results using Regression Learner for twenty-nine different methods used for regression analysis and accuracy rates of the obtained regression models for validation data.*

#	Model Type	Preset	Validation			
			RMSE	**MSE**	**MAE**	**R^2**
1	Linear Regression	Linear	0.012588	0.000158	0.009599	0.123731
2	Linear Regression	Interactions Linear	0.012588	0.000158	0.009599	0.123731
3	Linear Regression	Robust Linear	0.013135	0.000173	0.008917	0.045856
4	Stepwise Linear Regression	Stepwise Linear	0.012588	0.000158	0.009599	0.123731
5	Tree	Fine Tree	0.012533	0.000157	0.008944	0.13138
6	Tree	Medium Tree	0.011712	0.000137	0.008312	0.241457
7	Tree	Coarse Tree	0.013447	0.000181	0.010247	-2.22E-16
8	SVM	Linear SVM	0.013339	0.000178	0.009065	0.01608
9	SVM	Quadratic SVM	0.016169	0.000261	0.011419	-0.44582
10	SVM	Cubic SVM	229.6938	52759.26	53.54574	-2.9E+08
11	SVM	Fine Gaussian SVM	0.012541	0.000157	0.008514	0.130178
12	SVM	Medium Gaussian SVM	0.013345	0.000178	0.008672	0.015206
13	SVM	Coarse Gaussian SVM	0.013731	0.000189	0.008964	-0.0426
14	Efficient Linear	Efficient Linear Least Squares	0.012747	0.000162	0.008873	0.101434
15	Efficient Linear	Efficient Linear SVM	0.018207	0.000332	0.012965	-0.83326
16	Ensemble	Boosted Trees	0.011626	0.000135	0.008303	0.252472
17	Ensemble	Bagged Trees	0.011648	0.000136	0.008386	0.249659
18	Gaussian Process Regression	Squared Exponential GPR	0.012155	0.000148	0.009098	0.182919
19	Gaussian Process Regression	Matern 5/2 GPR	0.012362	0.000153	0.009178	0.154837
20	Gaussian Process Regression	Exponential GPR	0.012166	0.000148	0.009068	0.181471
21	Gaussian Process Regression	Rational Quadratic GPR	0.012327	0.000152	0.009161	0.159662
22	Neural Network	Narrow Neural Network	0.011965	0.000143	0.008771	0.208252
23	Neural Network	Medium Neural Network	0.012322	0.000152	0.009129	0.16033

24	Neural Network	Wide Neural Network	0.012194	0.000149	0.009269	0.177763
25	Neural Network	Bilayered Neural Network	0.012248	0.00015	0.009045	0.170435
26	Neural Network	Trilayered Neural Network	0.012298	0.000151	0.009073	0.163612
27	Kernel	SVM Kernel	0.013204	0.000174	0.009584	0.035915
28	Kernel	Least Squares Regression Kernel	0.012135	0.000147	0.009076	0.185604

Source: *Authors' calculations.*

The resulting Ensemble Boosted Trees model for nominal gross domestic product (GDP) and adult HIV/AIDS prevalence data was compared with statistical regression models such as linear, quadratic, exponential and logarithmic. The results, in the form of RMSE, MSE, MAE, and the coefficient of determination R^2, are presented in Table 3.

Table 3. *Comparison of regression analysis results for selected statistical models and the Ensemble Boosted Trees machine learning model.*

Model Type	**RMSE**	**MSE**	**MAE**	**R^2**
Linear Regression	0.0124	1.5415e-04	0.0094	0.1400
Quadratic Regression	0.0120	1.4513e-04	0.0091	0.1903
Exponential Regression	0.0118	1.3834e-04	0.0088	0.2282
Logarithmic Regression	0.0116	1.3417e-04	0.0083	0.2514
Regression Learner: Ensemble Boosted Trees	0.0094	8.7947e-05	0.0064	0.5093

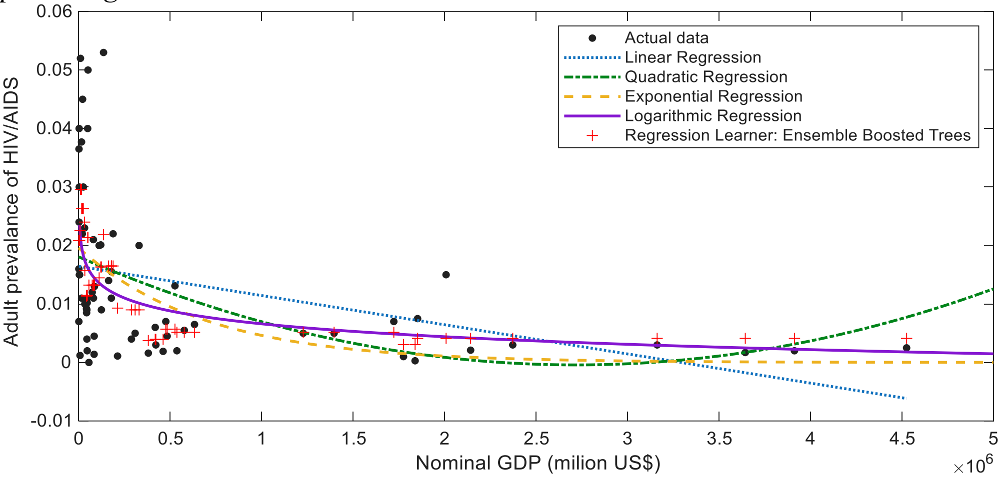
Source: *Authors' calculations.*

In this case, errors and R^2 were calculated based on the actual and predicted values for each model. Comparing the results of each of the presented indices (RMSE, MSE, MAE, R^2) for the regression models, we find that the Ensemble Boosted Trees model has the best fit to the actual data.

In this case, the RMSE = 0.0094, MSE = 8.7947e-05, and MAE = 0.0064 errors are the lowest, and the coefficient of determination ($R^2 = 0.5093$) is the highest. Among the presented models, the linear regression model predicts the actual data worst. For this regression model, all errors are the highest, and the coefficient of determination (R^2) is the lowest.

Figure 1 presents the actual data and the linear, quadratic, exponential, and logarithmic regression models and the Ensemble Boosted Trees regression model for the nominal gross domestic product (GDP) and adult prevalence of HIV/AIDS in percentages.

Figure 1. *Regression results using statistical and machine learning methods for the nominal gross domestic product (GDP) and adult prevalence of HIV/AIDS in percentages.*



Source: *Authors' calculations.*

4. Conclusions

The article presents the use of machine learning methods to build regression models demonstrating the relationship between nominal gross domestic product (GDP) and adult HIV/AIDS prevalence.

Based on data from 69 countries, 28 regression models were analyzed, taking into account measures of fit such as RMSE, MSE, MAE, and R^2. The Ensemble Boosted Trees model demonstrated the best fit. This model was also compared with four statistical regression models. In this case, the Ensemble Boosted Trees model also proved to be the best fit. This regression model allows for the prediction of previously unknown values.

Furthermore, the Spearman rank correlation revealed a statistically significant, negative relationship between nominal GDP and adult HIV/AIDS prevalence. Further research will be based on expanding the number of countries and attempting to model the relationship presented in the article.

References:

Adult prevalance of HIV/AIDS. 2025. Available at: https://en.wikipedia.org/wiki/HIV_adult_prevalence_rate.

Amiri, A., Gerdtham, U., Ventelou, B. 2012. HIV/AIDS-GDP Nexus? Evidence from panel-data for African countries. Economics Bulletin, AccessEcon, 32(1), 1060-1067.

Bhargava, A., Docquier, F. 2008. HIV pandemic, medical brain drain, and economic development in Sub-Saharan Africa. World Bank Economic Review, 22(2), 345-

366.
Bidzha, M.L., Ngepah, N., Greyling, T. 2024. The impact of antiretroviral treatment on the relationship between HIV/AIDS and economic growth. Economic Analysis and Policy, 81, 368-387.
Canning, D. 2006. The Economics of HIV/AIDS in Low-Income Countries: The Case for Prevention. Journal of Economic Perspectives, 20(3), 121-142.
Essig, A., Kang, S., Sellers, R. 2015. The Relationship between HIV Infection Rates and GDP Per Capita in African Countries. Georgia Institute of Technology, 1-16, Available at: https://repository.gatech.edu/server/api/core/bitstreams/6be76a0a-aa48-4070-a84e-4e1d1ed47b0a/content.
Flessa, S., Marschall, P. 2009. Socio-economic impact of antiviral intervention. Handbook of experimental pharmacology, 189, 347-374.
Gonzalo, T., García Goñi, M., Muñoz-Fernández, M. 2009. Socio-economic impact of antiretroviral treatment in HIV patients. An economic review of cost savings after introduction of HAART. AIDS reviews, 11(2), 79-90.
Grima, S., Dalli Gonzi, R., Thalassinos, E. 2020. The impact of COVID-19 on Malta and it's economy and sustainable strategies. Journal of Corporate Governance, Insurance, and Risk Management, 7(1), 53-73.
Haakenstad, A., Moses, M.W., Tao, T., Tsakalos, G., Zlavog, B., Kates, J., Wexler, A., Murray, Ch.J.L., Dieleman, J.L. 2019. Potential for additional government spending on HIV/AIDS in 137 low-income and middle-income countries: an economic modelling study. Lancet HIV, DOI: 10.1016/S2352-3018(19)30038-4.
Jones-Smith, K. 2007. The Impact of HIV on Economic Development. Master's Theses – Economics, Illinois State University, 1-18, Available at: https://ir.library.illinoisstate.edu/cgi/viewcontent.cgi?article=1007&context=mte.
Kirigia, J.M., Sambo, L.G., Okorosobo, T., Mwabu, G.M. 2002. Impact of HIV/AIDS on Gross Domestic Product (GDP) in the WHO Africa Region. African Journal of Health Sciences, 9(1-2), 27-39.
Lamontagne, E., Over, M., Stover, J. 2019. The economic returns of ending the AIDS epidemic as a public health threat. Health Policy, 123(1), 104-108.
Leidl, R. 1990. Model-Based Scenarios to Describe Economic Impacts of AIDS: The Role of Case-Mix. DOI:10.1007/978-3-642-84089-0_26.
Levi, J., Pozniak, A., Heath, K., Hill, A. 2018. The impact of HIV prevalence, conflict, corruption, and GDP/capita on treatment cascades: data from 137 countries. Journal of Virus Eradication, 4(2), 80-90.
Nguyen, T. D., Le, A.H., Thalassinos, E.I., Trieu, L.K. 2022. The impact of the COVID-19 pandemic on economic growth and monetary policy: An analysis from the DSGE model in Vietnam. Economies, 10(7), 159.
Nominal GDP (milion US$). 2025. Available at: https://en.wikipedia.org/wiki/List_of_African_countries_by_GDP_(nominal).
Schwefel, D., Leidl, R., Rovira, J., Drummond, M.F. 1990. Economic Aspects of AIDS and HIV Infection. Health Systems Research, Springer Berlin, Heidelberg.