Ecological footprint impact factors forecasts using VAR model: decision making case study from Ukraine

Andrii Verstiak
Department of Economic modelling and business informatics
Chernivtsi National University
Chernivtsi, Ukraine
a.verstyak@chnu.edu.ua

Vasyl Hryhorkiv
Department of Economic modelling and business informatics
Chernivtsi National University
Chernivtsi, Ukraine
v.hryhorkiv@chnu.edu.ua

Lesia Buiak
Department of Computer Science
Ternopil National Economic University
Ternopil, Ukraine
lesyabuyak@ukr.net

Mariia Hryhorkiv
Department of Economic modelling and business informatics
Chernivtsi National University
Chernivtsi, Ukraine
m.hryhorkiv@chnu.edu.ua

Oksana Verstiak
Department of Economic cybernetics and international economics
Chernivtsi Institute of Trade and Economics of Kyiv National University of Trade and Economics
Chernivtsi, Ukraine
oks1982@gmail.com

***Abstract* - In this article we investigated main factors affecting ecological footprint in Ukraine. Ecological footprint helps to improve sustainability and well-being, to optimize public project investments and to understand their impact in the planet. Deep analysis of researches for other countries has been adopted. We concluded that such factors as GDP per capita, exports of goods and services and population have an impact at ecological footprint. Further, we simulated data by Vector Autoregression model and used it as a forecasting algorithm. We chose this model because there are various time series influencing each other and EF. We've got statistically significant forecast values that can be used by policy and decision makers at the country level.**

Keywords—ecological footprint, sustainable development, impact factors, VAR, forecast

I. INTRODUCTION

The Ecological Footprint (EF) is a well-known metric for calculating how much natural resources we have and how much we consume. It helps to improve sustainability and well-being, to optimize public project investments and to understand their impact on the planet [1]. This measure allows us to compare the needs of the individual, family, community, nation and civilization as a whole in natural capital with the amount of environmental resources available, as well as with opportunities for their restoration. Analysis of ecological footprint assesses area of ecosystem that is required to support human populations [2]. The analysis of EF is universally by governments and scientists to support sustainable goals assessments. The indicator is calculated both for an individual and for a group of people and represents the area of biologically productive surface of land and water needed both for the supply of natural resources consumed by a person or group of people and for the absorption of waste associated with this consumption. Levels of EF varies greatly between regions and countries, with 29 high-income countries wanting more than eight times the per-person capacity requested by 55 low-income countries. [3].

Comparing consumption and lifestyle, as well as testing for biopotential - nature's ability to support that consumption – is what ecological footprint per capita is all about. The tool can help policymakers determine if a country consumes more (or less) than what is available on its soil, or if a country's way of life will be replicated globally. It should also be an effective tool for teaching individuals about excessive consumption and encouraging them to modify their habits. Many present lifestyles can be argued to be unsustainable using environmental footprints. This worldwide comparison also clearly demonstrates the disparities in resource consumption on this planet at the turn of the twenty-first century.

Therefore, researchers are now interested in the factors that affect the EF [4]. There are several interesting findings. Decun Wu [5] pointed that the major driving causes of EF evolution were population increase and affluence level, whereas technical development may successfully restrict EF expansion. The influence of energy consumption, urbanization, and economic growth on developing nations' ecological footprints from 1971 to 2014 is examined in another article [6]: the empirical estimation suggests that energy consumption has a positive and significant impact on the ecological footprint. Deep analysis was provided in research paper [7] indicating that (1) among 39 sectors essential factors influencing EF change are economic impact, population impact and footprint intensity with corresponding rates to total EF of 59.4%, 31.0%, and 7.7%, respectively. The effect of the industrial structure tends is equal 0; and (2) among 9 industries covered, population impact and economic impact boost the growth of the EF of 9 sectors.

From the other side there are several papers investigating EF impact factor depending on country development level. Thus, Sh-T. Chen et al. [4] classified the countries into three income groups to discussed whether the effects of the different factors on the EF change for different income groups. L. Charfeddine [8] extended the work of Al-Mulali and Ozturk [9]. Re-examining the Environment Kuznets Curve (EKC) hypothesis for 15 MENA (Middle East and North

African) nations using the Ecological Footprint as a proxy for environmental deterioration for the period 1975–2007 builds on the work of Al-Mulali and Ozturk [9]. The findings demonstrate that energy usage worsens environmental footprint, but real GDP per capita in oil-exporting nations has an inverted U-shaped connection with EF.

There are many others researches relayed with EF impact factors for other group of countries, but we found out the lack of info for Eastern European countries ex.gr. Ukraine; as biggest country in EEC has to be taken into consideration.

By this research we will discover impact factors that influencing EF in Ukraine and will provide its forecasting as multivariate time series where each variable depends not only on its past values but also has dependency on other factors and variables; further we can use this dependency for forecasting future values of impact factors.

II. Data

To carry up the EF study, we employed National Footprint and Biocapacity Accounts (NFAs). Nations' natural resource consumption and resource capacity are measured over time by the NFA [9].

According to [9] ecological footprint can be described as follows:

$$EF_c = EF_p + \left(EF_i - EF_e\right), \qquad (1)$$

where EF_c is ecological footprint of consumption indicating the consumption of biocapacity; EF_p - ecological footprint of production indicating the consumption of biocapacity resulting from production processes; EF_i and EF_e - imports and exports respectively indicating the use of biocapacity within international trade. The findings of this study offer insight on a country's environmental effect. If a country's Footprint is less than its biocapacity, it has an ecological reserve; otherwise, it has an ecological deficit. The former are known as ecological creditors, while the latter are known as ecological debtors.

Ukraine has strong ecological deficit (Fig. 1) with 2A data quality. Except for the most recent data year, time series have outcomes that are highly unreliable or highly improbable; the overall EF time series outcomes are not considerably altered by improbable data. [9]:

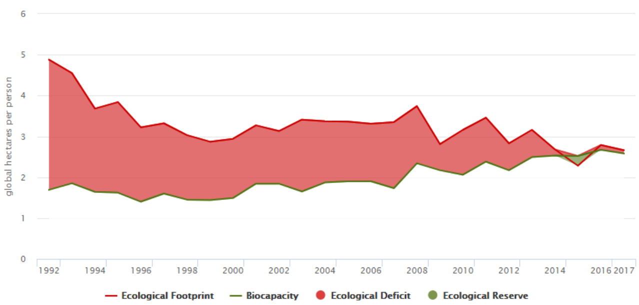

Fig. 1. Ecological Footprint vs Biocapacity (gha per person): Ukrainian trend [9]

Summarizing EF impact factors from above section we have chosen the next variables for Ukrainian case: Ecological Footprint of Consumption (*EF*, global hectares per person, depended variable), GDP per capita (*GDP_pc*, current US$), Urban population (*Urban*, % of total population), Exports of goods and services (*Exports*, % of GDP), Population, total (*Population_all*), Foreign direct investment, net inflows (*Invest*, % of GDP). Data was extracted from World Data Bank [10]. All numbers are presented in natural logarithm.

III. Methodology

We used Vector Autoregression (VAR) as an algorithm to make forecasts. When two or more time series interact, this approach can be utilized; in other words: relationship between selected variables is considered as bi-directional. VAR is an econometric model that generalizes one-dimensional AR models and describes the evolution and interdependence of many time series. VAR considers all variables symmetrically, including each equation variable, explaining its evolution using its own lags (prior period values) and the lags of all other variables in the model.

In fact, VAR is a system of econometric equations, each of which is autoregressive distributed lags (ADL). Let's assume that y^i, $i = 1,...,k$ time series *ADL(p,p)*-model for *i* can written as follows:

$$j_t^i = \alpha_0^i + \sum_{j=1}^{k} a_{1j}^i y_{t-1}^j + \sum_{j=1}^{k} a_{2j}^i y_{t-2}^j + \ldots + \sum_{j=1}^{k} a_{pj}^i y_{t-p}^j + \varepsilon_t^i, \qquad (2)$$

However, the vector-matrix notation of the model is more convenient and compact. For this, a vector of time series is introduced as $y_t = \left(y_t^1, y_t^2, ..., y_t^k\right)$. Then the above equations for each time series can be written in one equation in vector form:

$$y_t = a_0 + A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + \varepsilon_t = a_0 + \sum_{m=1}^{p} A_m y_{t-m} + \varepsilon_t, \qquad (3)$$

where A_m - matrices of elements a_{mj}^i.

Eq. (3) is the *p* order *VAR (p)* vector autoregression model.

The presented model is a closed one in the sense that only lags of endogenous (explained) variables act as explanatory variables. The limitations of structural models are not present in VAR models. Nonetheless, the difficulty with VAR models is that as the number of examined time series and the number of delays grows, so does the number of parameters..

Our algorithm of proceeded simulations and forecasts includes the steps listed below:

- time series descriptive statistics analysis;
- causation test;
- time series stationarity test;
- differencing time series to make them stationary (if needed);
- finding optimal order *p* VAR-model;
- estimating of train and test datasets, VAR-model training;
- rolling back provided transformations (if needed);
- evaluating the model using test set and providing final forecasts.

Training and forecasting VAR model has been provided in Python using Statsmodels library.

IV. Results

Let's first provide descriptive statistics (Table I) of time series and its correlation (Table II). Here we have to stress that all of the time series have a similar trends over time. The correlations between all the variables are less than 1; additionally we may provide preliminary conclusion about EF impact factors: it is GDP per capita and exports levels.

TABLE I. DESCRIPTIVE STATICS OF TIME SERIES

	EF	GDP_pc	Exports	Population	Urban	Invest
Mean	1.380	7.399	3.796	17.687	4.217	0.625
Median	1.379	7.306	3.852	17.675	4.214	0.735
Standard Deviation	0.093	0.603	0.236	0.053	0.013	1.024
Kurtosis	0.991	-1.417	1.994	-1.403	-1.546	-0.721
Skewness	0.577	0.019	-1.463	0.351	0.327	-0.527
Minimum	1.223	6.455	3.177	17.618	4.202	-1.308
Maximum	1.627	8.301	4.134	17.770	4.238	2.205

TABLE II. CORRELATION OF VARIABLES

	EF	GDP_pc	Exports	Population	Urban
EF	1.0000				
GDP_pc	0.4467	1.0000			
Exports	-0.5952	-0.0404	1.0000		
Population	0.0616	-0.7466	-0.5596	1.0000	
Urban	0.0529	0.8326	0.3654	-0.9475	1.0000
Invest	-0.2132	0.4486	0.4878	-0.6430	0.4701

Let's go further with assessment of causality amongst variables. We have provided Granger causality and cointegration test for this purpose.

The basic idea of VAR is that each of variables in the model influences each other. To put it another way, we can forecast time series using their previous values and other time series in the model. The Granger causality test allows you to evaluate the link even before you fit a model (the rows of Table II represent the response Y and columns predictor series X):

TABLE III. GRANGER CAUSALITY

	EF	GDP_pc	Exports	Population	Urban
EF	1.0000	0.0000	0.0040	0.0002	0.0000
GDP_pc	0.0011	1.0000	0.0001	0.0005	0.0230
Exports	0.0000	0.0060	1.0000	0.0008	0.0000
Population	0.0034	0.0004	0.0005	1.0000	0.0001
Urban	0.0000	0.0000	0.0001	0.0000	0.0000
Invest	0.0000	0.0000	0.0004	0.0000	0.0000

We know if p-values is lower than significance level, then, the corresponding X variables induces Y. Remarkably, we may reject the null hypothesis and conclude all casualities from the Table III. Considering p-values from this table we reasonably observe that all variables in the model are correspondently producing each other. From the above we may conclude that proposed time series and variables are good avails to be assessed and forecasted by VAR model.

Let's go further with cointegration test. This test will aid in determining whether or not there is a statistically significant link between the time series under investigation. We used Johanssen [11] procedure to provide the cointegration test (Table IV).

TABLE IV. COINTEGRATION TEST

	Test statistics	*C(95\%)*	*Significance*
EF	216.84	83.9383	True
GDP_pc	125.83	60.0627	True
Exports	49.45	40.1749	True
Population	29.92	24.2761	True
Urban	11.26	12.3212	False
Invest	0.0	4.126	False

Proposed VAR model has been fitted on train data set (all data set minus 4 periods) and formerly used to forecast the next 4 observations.

Since VAR model desires that time series have be stationary it is conventional to proceed time series in the model for stationarity. We used ADF-test for this reason and got all times series stationary with the 2nd difference.

To find the optimal order of VAR model we proceeded with fitting of increasing orders and selected the order with the lowest AIC. Still the common procedure is AIC comparison, we additionally estimated BIC, and HQIC. Accordingly, the best lag was studied at the lag order of 4. Regression summary modelled by OLS method is expressed in the article due to long output. We proceeded rather with serial correlation of residuals to analyze existence of extra in the errors/residuals. If the correlation still exists in residuals we may conclude that some pattern in time series is still left and has to be explained additionally by model. Correlation matrix of residuals is expressed in Table V.

TABLE V. CORRELATION MATRIX OF RESIDUALS

	EF	GDP_pc	Exports	Population	Urban
EF	1	0.900138	-0.083772	-0.514605	0.233376
GDP_pc	0.9001	1.0000	-0.0178	-0.4836	0.1909
Exports	-0.0838	-0.0178	1.0000	-0.0341	0.3297
Population	-0.5146	-0.4836	-0.0341	1.0000	0.2642
Urban	0.2334	0.1909	0.3297	0.2642	1.0000
Invest	-0.3425	-0.2556	-0.2781	0.7242	0.1823

To assess this procedure the typical algorithm is either to change the order of model, quantity of predictors in the system or take another method of time series simulations. Remarkably, valuation of serial correlation is to establish that a model is adequately explains deviations and patterns in simulated time series. A well-known way of valuation of serial correlation of residuals is Durbin Watson’s Statistic. All values of this statistic (not mentioned in the paper) are closer to value 2, implying that there is no substantial serial connection in our situation. Let's get started with the forecast.

Finally, we received forecasts based on VAR model that supposes lag order value in time series from its past values. It is related to the fact that terms of the model are originally the lags of different time series. In this case it is important to get an equal quantity of past values to the lag order that was used in the model fitting. Forecast results of train and test data are shown in Fig. 2.

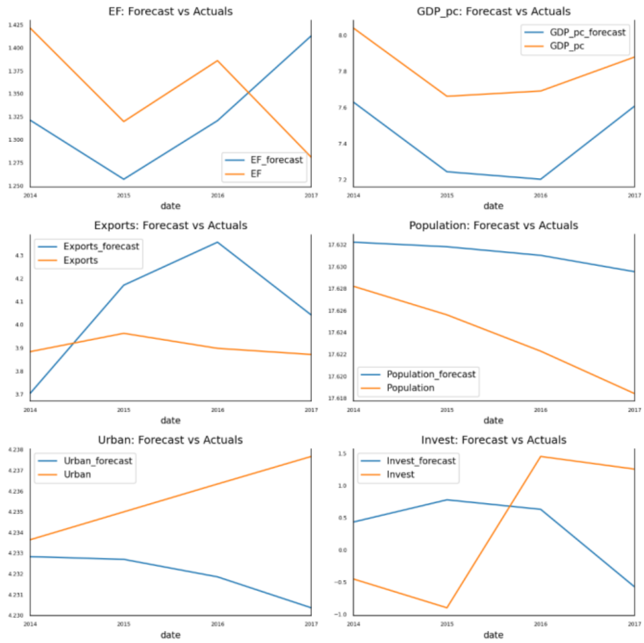

Fig. 2. Forecasts: test and train data

Let’s compare forecasts against actual values of test data. With this purpose, we evaluated various accuracy metrics such as MAE, MAPE, MPE, ME, RMSE, Pearson product-moment correlation coefficient and minmax (Table VI).

TABLE VI. CORRELATION MATRIX OF RESIDUALS

	EF	GDP_pc	Exports	Population	Urban	Invest
MAE	0.067	0.051	0.0009	0.0651	1.4564	0.0004
MAPE	-0.0241	-0.3978	-0.0037	0.1639	-0.022	0.0075
MPE	0.09	0.3978	0.0037	0.2541	1.3018	0.0075
ME	-0.015	-0.051	-0.0009	0.0419	-1.456	0.0004
RMSE	0.0943	0.4055	0.0045	0.2803	1.3786	0.008
Corr	-0.364	0.9327	-0.9372	0.3752	-0.519	0.9832
Minmax	0.0646	0.051	0.0009	0.0609	1.5647	0.0004

CONCLUSIONS

The major source of a person's ecological footprint is their daily activities. A country's ecological footprint is made up of socioeconomic characteristics, income levels, food, products and services consumed, and waste created. The degree of humanity's influence on the earth is determined by how much energy and water we spend, how much rubbish we toss away, what food (and in what packaging) we eat, and what furniture and clothes we pick. No norms, bans, or regulations can help people halt environmental degradation and attain peace with nature unless they change their habits and behavior.

By this investigations we showed main factors influencing ecological footprint in Ukraine such as GDP per capita, Exports of goods and services, Population total, Urban population, Foreign direct investment net inflows. We concluded that that first 4 factors influence EF in Ukraine.

Further we used Vector Autoregression as a forecasting algorithm. We chose this model because there are more than two time series influence each other and EF; relationship between selected variables is considered as bi-directional. By this article, we provided VAR simulation technique including Ganger test, assessment of optimal *p*-order, checking serial correlation and computing accuracy metrics. We’ve got statistically significant forecast values that can be used by policy and decision makers at the country level.

REFERENCES

[1] Global Footprint Network, 2021 [Online], Available: https://www.footprintnetwork.org/our-work/ecological-footprint/. Accessed 1 Feb. 2021.

[2] W. E. Rees, Ecological Footprint, Concept of, Editor(s): Simon A Levin, Encyclopedia of Biodiversity (Second Edition), Academic Press, 2013, pp. 701-713, https://doi-org.skaitykla.mruni.eu/10.1016/B978-0-12-384719-5.00037-X

[3] M. Wackernagel, J. Kitzes, Ecological Footprint, Editor(s): Sven Erik Jørgensen, Brian D. Fath, Encyclopedia of Ecology, Academic Press, 2008, pp. 1031-1037. https://doi.org/10.1016/B978-008045405-4.00620-0.

[4] Sheng-Tung Chen, Hui-Ting Chang. Factors that affect the ecological footprint depending on the different income levels[J]. AIMS Energy, 2016, 4(4), pp. 557-573. doi:10.3934/energy.2016.4.557

[5] Decun Wu, Spatially and temporally varying relationships between ecological footprint and influencing factors in China’s provinces Using Geographically Weighted Regression (GWR), Journal of Cleaner Production, Volume 261, 2020, pp. 121089, https://doi.org/10.1016/j.jclepro.2020.121089.

[6] Danish, Zhaohua wang, Investigation of the ecological footprint’s driving factors: What we learn from the experience of emerging economies, Sustainable Cities and Society, Volume 49, 2019, pp. 101626, https://doi.org/10.1016/j.scs.2019.101626.

[7] Lingna Liu, Yalin Lei, Jianping Ge, Kejia Yang. Sector screening and driving factor analysis of Beijing's ecological footprint using a multi-model method, Journal of Cleaner Production, Volume 191, 2018, pp. 330-338, https://doi.org/10.1016/j.jclepro.2018.04.132.

[8] Usama Al-Mulali, Ilhan Ozturk. The effect of energy consumption, urbanization, trade openness, industrial output, and the political stability on the environmental degradation in the MENA (Middle East and North African) region, Energy, Volume 84, 2015, pp. 382-389, https://doi.org/10.1016/j.energy.2015.03.004.

[9] National Footprint and Biocapacity Accounts 2021 edition (Data Year 2017); GDP, World Development Indicators, The World Bank 2020; Population, U.N. Food and Agriculture Organization.

[10] World data Bank indicators,2021 [Online], Available: https://databank.worldbank.org/reports.aspx?source=world-development-indicators. Accessed 1 Feb. 2021.

[11] Johansen, Søren. “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models.” Econometrica, vol. 59, no. 6, 1991, pp. 1551–1580. JSTOR, www.jstor.org/stable/2938278. Accessed 15 Feb. 2021.