

**Дзюбановська Н.В.**

кандидат економічних наук,  
доцент кафедри прикладної математики  
Тернопільського національного економічного університету

**Dziubanovska Nataliia**

Ph.D. (Economics),  
Associate Professor of the Department of Applied Mathematics  
Ternopol National Economic University

## МОДЕЛІ MAR SPLINES ОЦІНЮВАННЯ ТОРГІВЛІ РЕГІОНІВ УКРАЇНИ

## MAR SPLINES MODELS FOR EVALUATION OF UKRAINIAN REGIONS' TRADE

*У статті застосовано багатомірний адаптивний регресійний аналіз для оцінювання торгівлі регіонів України засобами пакета програм STATISTICA 10. Проаналізовано те, як демографічна ситуація у регіонах України, добробут населення, ситуація на ринку праці, ВРП у розрахунку на одну особу та індекс промислової продукції впливають на показники експорту та імпорту на душу населення. Під час дослідження використано статистичні дані регіонів України Державної служби статистики України протягом 2010–2016 рр. За результатами аналізу побудовано MAR Splines моделі для встановлення взаємозалежності між показниками експорту та імпорту на душу населення у регіонах України та деякими демографічними й економічними ідентифікаторами. Отримано два рівняння мультифакторної регресії для показників експорту та імпорту на душу населення, що описують відповідні залежності.*

**Ключові слова:** інтелектуальний аналіз даних, моделювання, регресійний аналіз, сплайни, торгівля.

*В статье применен многомерный адаптивный регрессионный анализ для оценивания торговли регионов Украины средствами пакета программ STATISTICA 10. Проанализировано то, как демографическая ситуация в регионах Украины, благосостояние населения, ситуация на рынке труда, ВРП в расчете на одного человека и индекс промышленной продукции влияют на показатели экспорта и импорта на душу населения. При исследовании использованы статистические данные регионов Украины Государственной службы статистики Украины в течение 2010–2016 гг. По результатам анализа построены MAR Splines модели для установления взаимозависимости между показателями экспорта и импорта на душу населения в регионах Украины и некоторыми демографическими и экономическими идентификаторами. Получены два уравнения мультифакторной регрессии для показателей экспорта и импорта на душу населения, которые описывают соответствующие зависимости.*

**Ключевые слова:** интеллектуальный анализ данных, моделирование, регрессионный анализ, сплайны, торговля.

*In the article the multidimensional adaptive regression analysis for estimation of Ukrainian regions' trade was applied with the STATISTICA 10 software package. There are large number of studies using data mining techniques, however, the problem of identifying hidden relationships between main trade indicators and the factors influencing them is relevant, and the received results give the opportunity to choose the right tools for further trade developments. The main purpose of the study is to construct the MAR Splines models for evaluation trade of the regions of Ukraine based on the use of multivariate adaptive regression analysis in order to establish correlations between the indicators of trade and some economic and demographic characteristics of the regions of Ukraine. The advantage of using multidimensional adaptive regression analysis is its simplicity and high efficiency for solving problems with large data sets. The study used statistical data from the regions of Ukraine of the State Statistics Service of Ukraine during 2010–2016. The dependent variables are  $Y_1$  – volume of exports per capita of the regions of Ukraine and  $Y_2$  – volume of imports per capita of the regions of Ukraine. The impact of demographics in the regions of Ukraine, population's well-being, labor market situation, GRP per capita and industrial production index on export and import indicators per capita was analyzed. According to the results of the analysis, MAR Splines models were constructed in order to establish the correlation between exports and import per capita in the regions of Ukraine and some demographic and economic identifiers. Two multivariate regression equations for export and import per capita, describing the respective dependencies were obtained. On the basis of the multidimensional adaptive regression analysis, there is a strong correlation between the main indicators of trade (export and import per capita) and such factors as: migration of population, employment of the population aged 15–70 years, unemployment, average monthly wage, GRP per person calculation, industrial production index.*

**Keywords:** data mining, modeling, regression analysis, splines, trade.

**Постановка проблеми.** Торгівля країн загалом має значний вплив на національну економіку країни, сприяє економічному зростанню та національному добробуту, тому важливо підібрати правильний підхід до оцінювання та вимірювання цього процесу. У наш час, коли комп'ютерні інформаційні технології стрімко розвиваються, щоразу більшої популярності набуває інтелектуальний аналіз даних як один з інструментів дослідження та моделювання торгівлі. Задля проведення ефективного аналізу даних, пошуку кореляційних взаємозв'язків використовують технології Data Mining, що призначені для пошуку закономірностей між об'єктами баз даних. Основним завданням інтелектуального аналізу даних переважно є пошук прихованих, об'єктивних закономірностей, що є корисними для практичного застосування у великих інформаційних масивах даних. До методів інтелектуального аналізу даних належить багатовимірний адаптивний регресійний аналіз.

**Аналіз останніх досліджень і публікацій.** Чимало робіт відомих учених присвячено інтелектуальному аналізу даних, зокрема застосуванню методів багатовимірних адаптивних регресійних сплайнів під час моделювання соціально-економічних процесів. Серед них слід відзначити таких науковців, як Г.В. Горбань [1; 2], О.О. Барсеґян, М.С. Купріянов, В.В. Степаненко, І.І. Холод [3], І.В. Турський [4], В.В. Стрижов, Р.А. Сологуб [5].

**Виділення не вирішених раніше частин загальної проблеми.** Незважаючи на велику кількість досліджень із використанням методів інтелектуального аналізу даних, проблема виявлення прихованих взаємозв'язків між основними показниками торгівлі та факторами впливу на ці показники є актуальною, а одержані висновки дадуть змогу правильно вибрати інструменти для подальших розробок у сфері торгівлі.

Формулювання цілей статті (**постановка завдання**). Метою статті є побудова MAR Splines моделей для оцінювання торгівлі регіонів України на основі використання багатовимірної адаптивної регресійної моделі для встановлення взаємозалежності між основними показниками торгівлі та деякими економічними й демографічними характеристиками регіонів України.

**Виклад основного матеріалу дослідження.** Вперше новий метод гнучкого регресійного моделювання масивів багатовимірних даних запропонував американський статистик Джером Фрідман (Jerome Harold Friedman) у своїй праці «Багатовимірні адаптивні регресійні сплайни» (1991 р.) [6], де він описав процедуру розв'язування регресійних задач і задач класифікації, в яких потрібно оцінити взаємозв'язок між наборами залежних і незалежних змінних. Ця процедура вмотивована рекурсивним підходом до регресії та

– для експорту:

$$\begin{aligned}
 Y_1 = & 1599,4 + 0,021 \cdot \max(0; X_9 - 22509,3) - 0,0102 \cdot \max(0; X_5 - 135599) + \\
 & + 0,016 \cdot \max(0; 135599 - X_5) - 7,88 \cdot \max(0; X_2 - 1399,8) - \\
 & - 3,56 \cdot \max(0; 1399,8 - X_2) + 13,3 \cdot \max(0; X_2 - 1279) + 29,32 \cdot \max(0; 22,8 - X_3) - \\
 & - 3,62 \cdot \max(0; X_2 - 773,5) - 0,1 \cdot \max(0; X_4 - 4559) + 0,024 \cdot \max(0; X_7 - 26943) - \\
 & - 0,05 \cdot \max(0; 26943 - X_7) + 0,042 \cdot \max(0; X_1 - 4982) + 0,093 \cdot \max(0; 4982 - X_1) + \\
 & + 17,81 \cdot \max(0; 112,8 - X_6);
 \end{aligned}$$

володіє більшою потужністю й гнучкістю для моделювання взаємозалежності між множинами змінних.

Перевагами застосування багатовимірної адаптивної регресійної моделі є простота й висока ефективність під час розв'язування задач з масивами даних великої розмірності.

У попередніх дослідженнях ми вже застосували MAR Splines моделі для оцінювання міжнародної торгівлі країн Європейського Союзу [7]. Доведемо, що цю методику багатовимірної адаптивної регресійної моделі можна застосувати для оцінювання теперішнього стану торгівлі регіонів України.

Залежними змінними визначимо  $Y_1$  – обсяг експорту на душу населення, тис. дол. США;  $Y_2$  – обсяг імпорту на душу населення, тис. дол. США. Оцінимо залежність основних показників торгівлі від таких факторів, як  $X_1$  – міграційний рух населення, тис. осіб;  $X_2$  – зайнятість населення віком 15–70 років, тис. осіб;  $X_3$  – безробіття, тис. осіб;  $X_4$  – середньомісячна заробітна плата населення, грн.;  $X_5$  – доходи населення, млн. грн.;  $X_6$  – індекс споживчих цін, % до попереднього року;  $X_7$  – ВВП у розрахунку на одну особу, грн.  $X_8$  – індекс промислової продукції, % до відповідного періоду попереднього року;  $X_9$  – капітальні інвестиції, млн. грн.

Для аналізу використано статистичні дані регіонів України Державної служби статистики України [8] протягом 2010–2016 рр. (без урахування тимчасово окупованої території Автономної Республіки Крим, м. Севастополя та частини тимчасово окупованих територій у Донецькій та Луганській областях).

Використовуючи модуль Data Mining/MAR Splines Analysis програмного продукту STATISTICA 10, одержуємо результати аналізу (рис. 1) та матрицю коефіцієнтів моделі (рис. 2).

Model specifications	Model Summary (Spreadsheet1)	
	Value	
Independents	9	
Dependents	2	
Number of terms	15	
Number of basis functions	14	
Order of interactions	1	
Penalty	2,000000	
Threshold	0,000500	
GCV error	445686,8	
Prune	Yes	

Рис. 1. Результати MAR Splines Analysis  
Джерело: отримано автором засобами STATISTICA 10

Оскільки вибрано дві залежні змінні для аналізу, отримуємо два рівняння мультифакторної регресії:

## Держава та регіони

Model coefficients (Spreadsheet1)												
NOTE: Highlighted cells indicate basis functions of type max(0, independent-knot), otherwise max(0, knot-independent)												
Coefficients, knots and basis functions	Coefficients		Knots		Knots		Knots		Knots		Knots	
	Експорт на душу населення, Y1	Імпорт на душу населення, Y2	Міграційний рух, X1	Зайнятість у віці 15-70 років, X2	Безробіття, X3	Середньомісячна заробітна плата, X4	Доходи населення, X5	Індекс споживчих цін, X6	ВРП на душу населення, X7	Індекс промислової продукції, X8	Капіальні інвестиції, X9	
Intercept	1599,396	2345,754										
Term.1	0,021	0,119									22509,30	
Term.2	-0,012	-0,029					135599,0					
Term.3	0,016	0,006					135599,0					
Term.4	-7,879	-12,714		1399,800								
Term.5	-3,560	-3,201		1399,800								
Term.6	13,286	13,626		1279,000								
Term.7	29,318	65,516			22,80000							
Term.8	-3,619	-3,277		773,500								
Term.9	-0,100	-0,575				4559,000						
Term.10	0,024	0,006							26943,00			
Term.11	-0,053	-0,022							26943,00			
Term.12	0,042	0,153	4982,000									
Term.13	0,093	-0,023	4982,000									
Term.14	17,809	25,030								112,8000		

Рис. 2. Матриця коефіцієнтів моделі

Джерело: отримано автором засобами STATISTICA 10

– для імпорту:

$$\begin{aligned}
 Y_2 = & 2345,75 + 0,12 \cdot \max(0; X_9 - 22509,3) - 0,03 \cdot \max(0; X_5 - 135599) + \\
 & + 0,006 \cdot \max(0; 135599 - X_5) - 12,7 \cdot \max(0; X_2 - 1399,8) - \\
 & - 3,2 \cdot \max(0; 1399,8 - X_2) + 13,63 \cdot \max(0; X_2 - 1279) + 65,52 \cdot \max(0; 22,8 - X_3) - \\
 & - 3,28 \cdot \max(0; X_2 - 773,5) - 0,575 \cdot \max(0; X_4 - 4559) + 0,006 \cdot \max(0; X_7 - 26943) - \\
 & - 0,022 \cdot \max(0; 26943 - X_7) + 0,153 \cdot \max(0; X_1 - 4982) - 0,023 \cdot \max(0; 4982 - X_1) + \\
 & + 25,03 \cdot \max(0; 112,8 - X_6);
 \end{aligned}$$

Побудовані MAR Splines моделі для оцінювання обсягів експорту та імпорту регіонів України дають змогу виявити певні важливі залежності.

За результатами аналізу доходимо висновку, що високі значення показників обсягів експорту й імпорту на душу населення регіонів України будуть досягнуті за максимальних значень показника зай-

нятості населення та мінімальних значень показника безробіття населення. Ця залежність підтверджується як для обсягів експорту, так і для обсягів імпорту. Наочно залежність ілюструється за допомогою сплайнових регресійних поверхонь, побудованих для невідомих  $Y_1, X_2, X_3$  (рис. 3) та для  $Y_2, X_2, X_3$  (рис. 4).

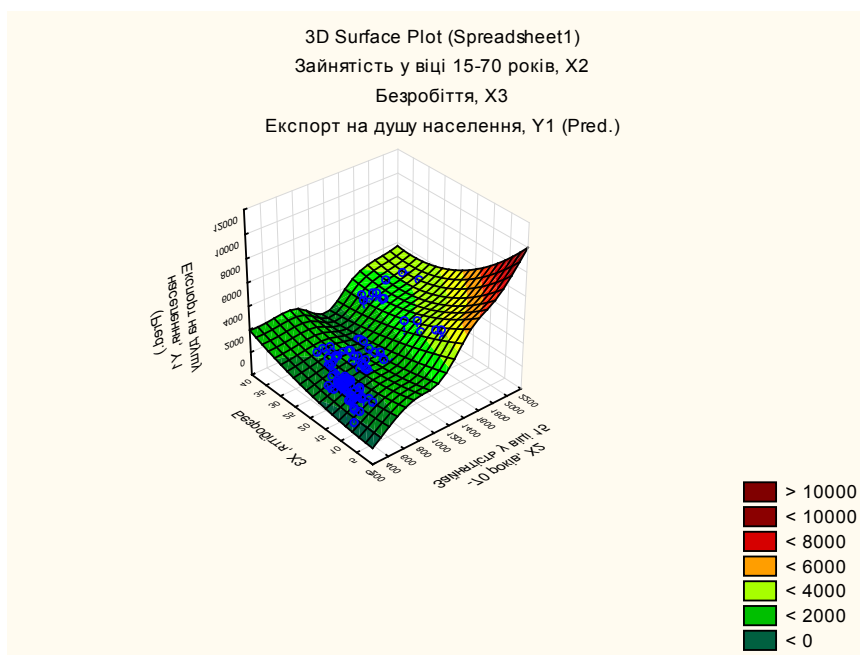


Рис. 3. Сплайнові регресійні поверхні (експорт на душу населення, зайнятість населення віком 15–70 років, безробіття населення)

Джерело: побудовано автором засобами STATISTICA 10

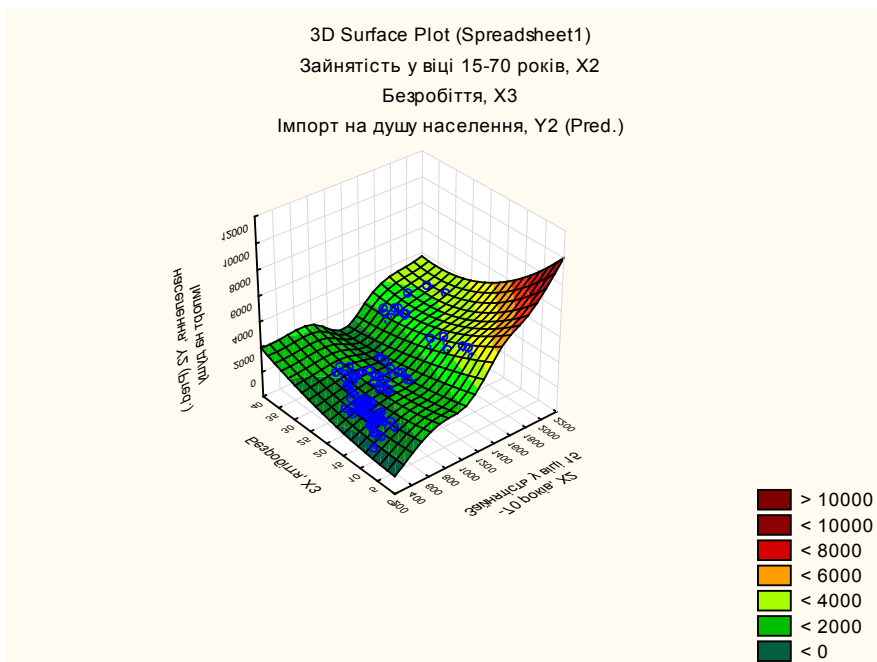


Рис. 4. Сплайнові регресійні поверхні (імпорт на душу населення, зайнятість населення віком 15–70 років, безробіття населення)

Джерело: побудовано автором засобами STATISTICA 10

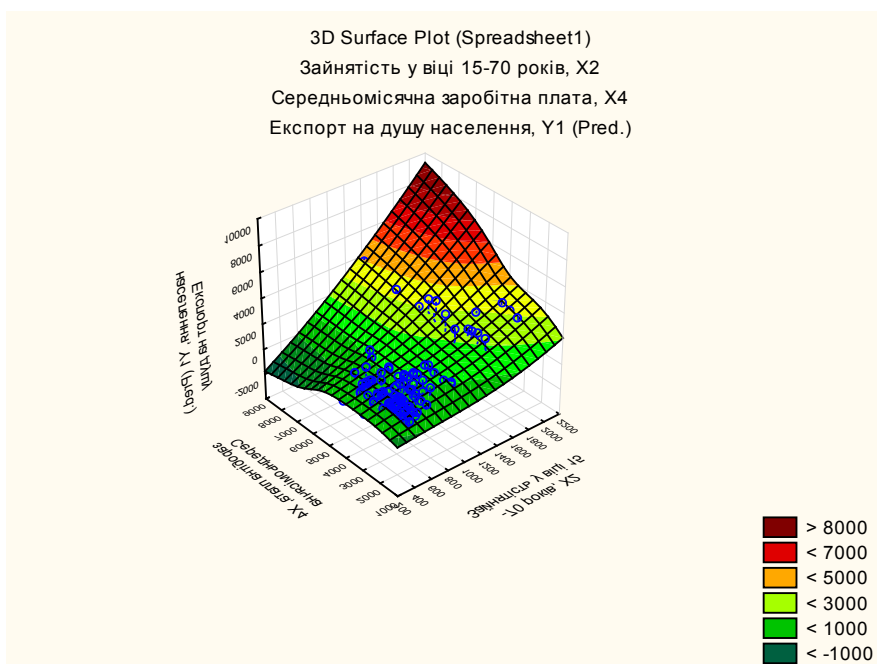


Рис. 5. Сплайнові регресійні поверхні (експорт на душу населення, зайнятість населення віком 15–70 років, середньомісячний дохід населення)

Джерело: побудовано автором засобами STATISTICA 10

Так, чим більше буде робочих місць, чим нижче буде рівень безробіття населення певного регіону, тим вище будуть значення основних показників торгівлі.

Така ж тенденція простежується для показників середньомісячної заробітної плати й безробіття населення. Чим більше буде матеріальне забезпечення населення, чим нижче буде рівень безробіття

певного регіону, тим вище будуть значення основних показників торгівлі.

Також дослідження підтвердило, що суттєвими будуть значення показників обсягів експорту й імпорту на душу населення за максимальних значень показників зайнятості населення та середньомісячного доходу населення. Отже, чим краще ситуація на ринку праці в певному регіоні та чим більш оплачу-

## Держава та регіони

вана праця населення, тим більш значущі показники торгівлі цього регіону. Це підтверджено сплайновими регресійними поверхнями побудованої MAR Splines моделі для показників експорту на душу населення ( $Y_1, X_2, X_4$ ) (рис. 5) і для показників імпорту на душу населення ( $Y_2, X_2, X_4$ ) (рис. 6).

Крім того, заслуговує на увагу залежність обсягів експорту й імпорту на душу населення від індексу промислової продукції.

Бачимо, що найвищі значення показників обсягів експорту й імпорту на душу населення будуть досягнуті за максимальних значень індексу промислової продукції. Підтвердження цієї залежності проілюстровано за допомогою сплайнових регресійних поверхней, побудованих для показників експорту на душу населення ( $Y_1, X_4, X_8$ ) (рис. 7) і для показників імпорту на душу населення ( $Y_2, X_4, X_8$ ) (рис. 8).

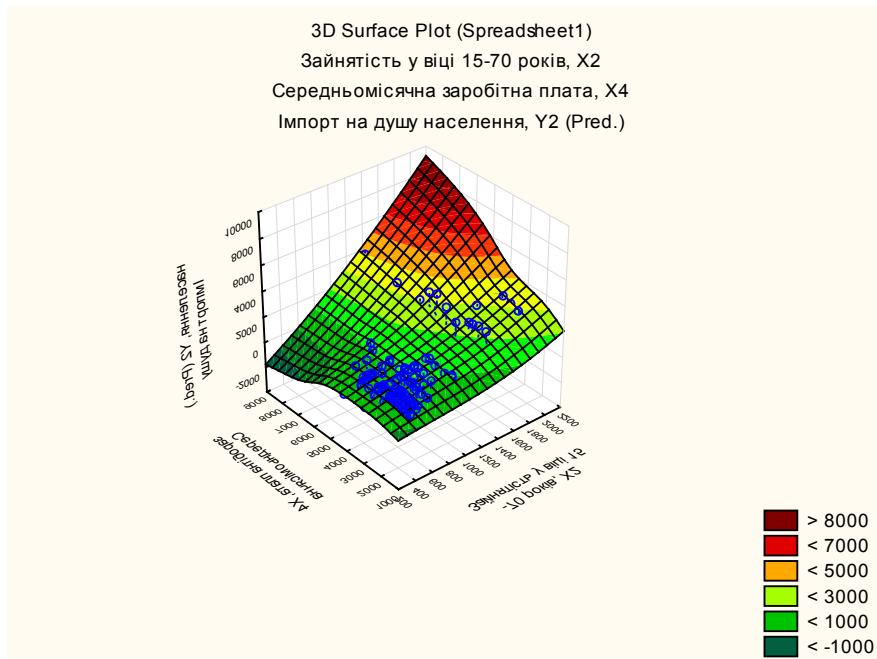


Рис. 6. Сплайнові регресійні поверхні (імпорт на душу населення, зайнятість населення віком 15–70 років, середньомісячний дохід населення)

Джерело: побудовано автором засобами STATISTICA 10

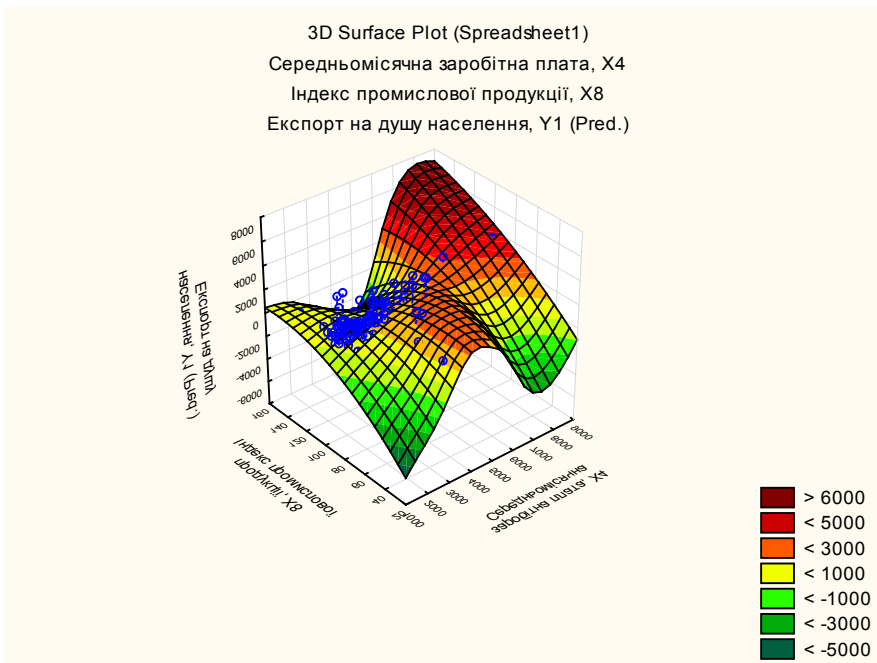


Рис. 7. Сплайнові регресійні поверхні (експорт на душу населення, середньомісячний дохід населення, індекс промислової продукції)

Джерело: побудовано автором засобами STATISTICA 10

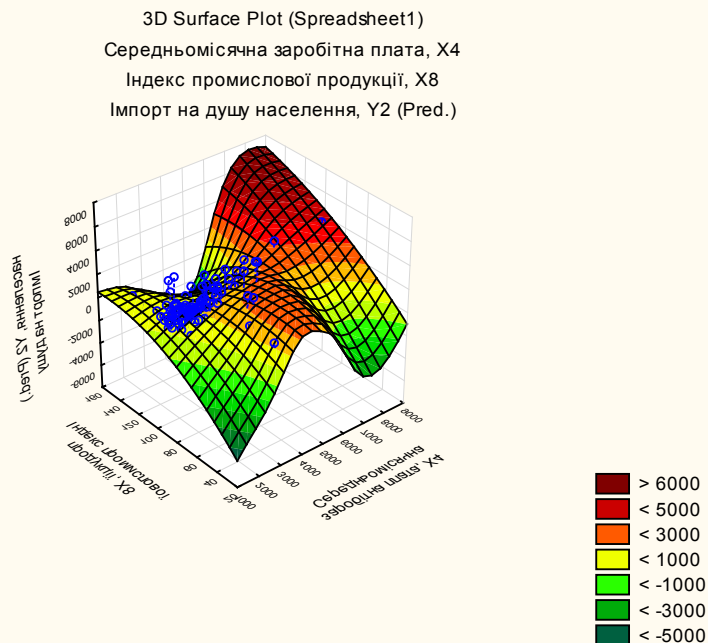


Рис. 8. Сплайнові регресійні поверхні (імпорт на душу населення, середньомісячний дохід населення, індекс промислової продукції)

Джерело: побудовано автором засобами STATISTICA 10

Такий же вплив на показники обсягів експорту й імпорту на душу населення має ВРП у розрахунку на одну особу. Крім того, спостерігається досить сильний зв'язок між показниками торгівлі регіонів України та міграційним рухом населення.

**Висновки.** Таким чином, проведений багатовимірний адаптивний регресійний аналіз дає змогу стверджувати, що забезпечення достатньої кількості робочих місць та високий дохід населення сприятимуть зменшенню рівня безробіття, підвищенню індекса промислової продукції, збільшенню ВРП, а також зростанню основних показників торгівлі.

За побудованими MAR Spline моделями отримано два регресійних рівняння, що описують лінійну залежність основних показників торгівлі від вищеприписаних факторів. З огляду на різні результати регресійного моделювання задля оцінювання впливу всіх факторів економічного розвитку для підвищення показників країни у сфері торгівлі доцільно буде застосувати комбіновані методи інтелектуального аналізу даних.

#### Бібліографічний список:

1. Горбань Г.В. Характеристики значущості міжвимірних асоціативних правил у багатомірних даних та їх обчислення. Миколаїв : вид-во ЧДУ ім. Петра Могили, 2015. Т. 2. С. 97–100.
2. Горбань Г.В. Пошук можливих асоціативних правил у багатомірних даних за допомогою генерації їх шаблонів. *Досвід та тенденції розвитку суспільства в Україні: глобальний, національний та регіональний аспекти* : збірник тез Всеукраїнської науково-методичної конференції «Могілянські читання – 2015». Т. 1. Миколаїв : вид-во ЧДУ імені Петра Могили, 2015. С. 75–77.
3. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И. Методы и модели анализа данных: OLAP и Data Mining. Санкт-Петербург : БХВ-Петербург, 2004. 336 с.

4. Турський І.В. Моделювання залежності валового регіонального продукту від розвитку підприємництва в регіонах України. *Науковий вісник Херсонського державного університету*. 2017. Вип. 25. Ч. 2. С. 77–81.
5. Стрижов В.В., Сологуб Р.А. Индуктивное построение регрессионных моделей волатильности опционных торгов. *Вычислительные технологии*. 2009. Т. 14. № 5. С. 102–113.
6. Friedman J.H. Multivariate Adaptive Regression Splines. *The Annals of Statistics*. 1991. Vol. 19. No. 1. P. 1–67. URL: [www.jstor.org/stable/2241837](http://www.jstor.org/stable/2241837) (дата звернення: 10.11.2019).
7. Дзюбановська Н.В. Прагматизм оцінювання міжнародної торгівлі країн: методи і моделі : монографія. Тернопіль : ТНЕУ, 2019. 298 с.
8. Державна служба статистики України : офіційний сайт. URL: <http://www.ukrstat.gov.ua> (дата звернення: 18.12.2019).

#### References:

1. Horban H.V. (2015) Kharakterystyky znachushchosti mizh-vymirovnykh asotsiatyvnykh pravyl u bahatomirnykh danykh ta yikh obchyslennia [Characteristics of the Significance of Tnterdimensional Associative Rules in Multidimensional Data and Their Calculation]. *Mykolaiv : vyd-vo ChDU im. Petra Mohyly*. (in Ukrainian)
2. Horban H.V. (2015) Poshuk mozhylyvykh asotsiatyvnykh pravyl u bahatomirnykh danykh za dopomohoiu heneratsii yikh shabloniv [Finding Possible Associative Rules in Multidimensional Data by Generating Their Patterns]. *Vseukrainska naukovo-metodychna konferentsiia "Moghilianski chytannia – 2015: dosvid ta tendentsii rozvytku suspilstva v Ukraini: hlobalnyi, natsionalnyi ta rehionalnyi aspekty"* : zbirnyk tez. T. 1. Mykolaiv : vyd-vo ChDU imeni Petra Mohyly. (in Ukrainian)
3. Barsehian A.A., Kupryanov M.S., Stepanenko V.V., Kholod Y.Y. (2004) Metody i modeli analiza dannykh: OLAP y Data Mining [Data Analysis Methods and Models: OLAP and Data Mining]. *SPB. : BKhV-Peterburh* (in Russian)

4. Turskyi I.V. (2017) Modeliuvannia zalezhnosti valovoho rehionalnoho produktu vid rozvytku pidpriemnytstva v rehionakh Ukrainy [Modeling the Dependence of Gross Regional Product on Entrepreneurship Development in Ukrainian Regions]. *Naukovyi visnyk Khersonskoho derzhavnoho universytetu*. (in Ukrainian)
5. Stryzhov V.V., Solohub R.A. (2009) Ynduktivnoe postroenie rehressionnykh modelei volatylnosti optsiionnykh torhov [Inductive Construction of Regression Models of Option Trading Volatility]. *Vychislitelnye tekhnologii*. (in Russian)
6. Friedman J.H. Multivariate Adaptive Regression Splines. *The Annals of Statistics*, vol. 19, no. 1, 1991, pp. 1–67. Retrieved from: [www.jstor.org/stable/2241837](http://www.jstor.org/stable/2241837) (accessed: 10 November 2019).
7. Dziubanovska N.V. (2019) Prahmatyzm otsiniuvannia mizhnarodnoi torhivli krain: metody i modeli [The pragmatism of assessment countries' international trade: methods and models]. *Ternopil : TNEU*. (in Ukrainian)
8. Derzhavna sluzhba statystyky Ukrainy: ofitsiinyi sait [State Statistics Service of Ukraine: official site]. [in Ukrainian]. Retrieved from: <http://www.ukrstat.gov.ua> (accessed: 18 December 2019).