

Моделювання залежності валового регіонального продукту від сільського господарства України на основі лонгітюдних даних

Зомчак Л.М.

кандидат економічних наук,
доцент кафедри економічної кібернетики
Львівського національного університету імені Івана Франка

Умриш Г.Т.

магістрант
Львівського національного університету імені Івана Франка

У статті досліджено залежність валового регіонального продукту областей України від головних показників функціонування сільського господарства на основі статистичних даних за період з 2004 року по 2015 рік. Оскільки вхідні дані мають панельну структуру, то побудовано 5 моделей на основі лонгітюдних даних, серед яких виділено дві з найкращими характеристиками: модель з фіксованими ефектами та модель з випадковими ефектами. За допомогою тесту Хаусмана оцінено, що модель лонгітюдних даних з фіксованими ефектами краще описує залежність валового регіонального продукту від основних сільськогосподарських показників, коефіцієнт детермінації моделі становить 0,63562, що є цілком прийнятно для моделей на лонгітюдних даних.

Ключові слова: лонгітюдні дані, валовий регіональний продукт, сільське господарство, модель лонгітюдних даних з фіксованими ефектами, модель лонгітюдних даних з випадковими ефектами, тест Хаусмана, коефіцієнт детермінації.

Zomchak L.N., Umrysh G.T. МОДЕЛИРОВАНИЕ ЗАВИСИМОСТИ ВАЛОВОГО РЕГИОНАЛЬНОГО ПРОДУКТА ОТ СЕЛЬСКОГО ХОЗЯЙСТВА УКРАИНЫ НА ОСНОВАНИИ ЛОНГИТЮДНЫХ ДАННЫХ

В статье исследована зависимость валового регионального продукта областей Украины от главных показателей функционирования сельского хозяйства на основе статистических данных за период с 2004 года по 2015 год. Поскольку входные данные имеют панельную структуру, то построено 5 моделей на основе лонгитюдных данных, среди которых выделены две с лучшими характеристиками: модель с фиксированными эффектами и модель со случайными эффектами. С помощью теста Хаусмана оценено, что модель лонгитюдных данных с фиксированными эффектами лучше описывает зависимость валового регионального продукта от основных сельскохозяйственных показателей, коэффициент детерминации модели составляет 0,63562, что вполне приемлемо для моделей на лонгитюдных данных.

Ключевые слова: лонгитюдные данные, валовой региональный продукт, сельское хозяйство, модель лонгитюдных данных с фиксированными эффектами, модель лонгитюдных данных со случайными эффектами, тест Хаусмана, коэффициент детерминации.

Zomchak L.M., Umrysh H.T. MODELING OF GROSS REGIONAL PRODUCT DEPENDENCE OF THE AGRICULTURE OF UKRAINE ON THE BASIS OF LONGITUDE DATA

The article examines the dependence of the gross regional product of the regions of Ukraine on the main indicators of agriculture on the basis of statistical data for the period from 2004 to 2015. Since the input data have a panel structure, five models based on longitude data are specified, two are distinguished with better characteristics: a model with fixed effects and a model with random effects. With the help of the Hausman test, it is estimated that the model of longitudinal data with fixed effects better describes the dependence of the gross regional product on the main agricultural indicators, the coefficient of determination of the model is 0.63562, which is quite acceptable for longitudinal data models.

Keywords: longitude data, gross regional product, agriculture, model of longitude data with fixed effects, a model of longitude data with random effects, Hausman test, determination coefficient.

Постановка проблеми. У теорії часових рядів вхідні дані зазвичай зібрані впродовж певного проміжку часу для одного чи декількох об'єктів, тоді як крос-секційні (варіаційні) дані збирають для різних об'єктів у довільному порядку та без врахування часу. Панельні або лонгітюдні дані – це дані про функціонування

чи характеристики кількох об'єктів, зібрані за певний часовий період. Застосування панельних даних дозволяє врахувати недоліки і часових, і варіаційних рядів. Панельні дані дозволяють враховувати та впливати на змінні, які важко піддаються спостереженню та вимірюванню, наприклад, культурні відмін-

ності чи відмінності у веденні бізнесу тощо. Особливо актуальним є застосування лонгitudних даних для розроблення ефективної регіональної політики.

Аналіз останніх досліджень і публікацій. Дослідженню можливостей застосування методів аналізу та моделювання на основі лонгitudних даних до проблем сільськогосподарського виробництва присвячені дослідження як вітчизняних, так і зарубіжних вчених. Серед вітчизняних виділимо роботу Н.О. Окселевка [1], який намагався вдосконалити процес управління оборотними активами сільськогосподарських підприємств із використанням моделей лонгitudних даних та розробив систему економетричних ANCOV-моделей для сільськогосподарських підприємств. Також проблема моделювання лонгitudних даних була висвітлена у праці О.Ю. Червак-Смерічко, яка застосувала панельні дані в сучасних умовах розвитку економіки при моделюванні соціально-економічних та політичних процесів на регіональному рівні [2].

Серед закордонних публікацій назвемо статтю С. Кумбахар [3] зі співавторами, у якій застосовано моделі панельних та крос-секційних даних для моделювання діяльності норвезьких виробників зерна за період 2004-2008 рр. та рекомендовано в подальших емпіричних дослідженнях застосовувати саме панельні дані. Прямі та непрямі ефекти впливу клімату на сільське господарство Тунісу досліджено у статті О. Зауабі та Н. Періді [4]. Схоже дослідження, а саме впливу клімату на родючість ґрунтів Індії, представлено у статті [5]. Л. Галіндо та ін. [6] досліджували вплив кліматичних змін та іригації на сільськогосподарську активність Мексики, використовуючи рикардіанську модель на основі панельних даних. Публікація М. Кассі [7] зі співавторами присвячена дослідженню впливу зміни технологій на фермерському та ринковому рівнях у Ефіопії, дослідження реалізовано на панельних даних. Автори дійшли висновку, що цілеспрямовані зусилля щодо розширення та інтеграції технологій ефективні на різних рівнях. Роль комунікативних технологій у сільському господарстві Австралії досліджено на основі панельних даних у статті Р. Салім та ін. [8]. Якщо попередні статті присвячені застосуванню панельного аналізу на рівні країн, то публікація К. Антоніні та Дж. Англієс-Бош [9] присвячена його застосуванню для моделювання продуктивності та екологічних втрат від інтенсифікації сільськогосподарського виробництва в країнах Європейського Союзу.

Виділення невирішених раніше частин загальної проблеми. Дослідження впливу різних чинників на обсяг валового регіонального продукту можна проводити за допомогою класичних економетричних методів та моделей, наприклад, множинної регресії, однак, ці інструменти дозволяють будувати моделі для кожного регіону зокрема та не враховують взаємозв'язок між регіонами країни. Для усунення цього недоліку доцільно скористатись методами лонгitudного аналізу, які дозволяють досліджувати дані одночасно і в регіональному, і в часовому розрізі.

Формулювання цілей статті (постановка завдання). Метою роботи є виявлення та моделювання залежності між валовим регіональним продуктом та головними показниками сільського господарства, при чому вхідні дані подані у розрізі областей та в часовому розрізі.

Виклад основного матеріалу дослідження. Для побудови моделі використано економічні показники: валовий регіональний продукт, середня урожайність основних сільськогосподарських показників (включаючи зернові та зернобобові культури, картопля, овочеві культури, плодові та ягідні культури) та загальна кількість сільськогосподарських тварин (включаючи велику рогату худобу, свині, птиця, вівці та кози). Усі показники взято у розрізі областей України (22 області, без урахування тимчасово окупованої території та АР Крим) та в розрізі часу (проміжок з 2004 по 2015 роки) із офіційного сайту Державної служби статистики України [10].

За допомогою програмного забезпечення R [11], а також його графічного редактор RStudio [12], були оцінені п'ять моделей, що будуються на панельних даних (pooled model, between estimator model, first differences estimator model, fixed effects model та random effects model). Кожна з цих моделей має свої характеристики (такі як коефіцієнт детермінації, t-статистика, F-статистики тощо), які дають змогу зрозуміти наскільки модель є адекватною і чи знайденні параметри коефіцієнтів моделі мають значення для даної моделі.

У загальному вигляді регресійну модель панельних даних можна виразити [13]:

$$Y_{it} = a + X_{it}'b_{it} + \varepsilon_{it}, \quad i = \overline{1, N}, \quad t = \overline{1, T},$$

де i – індекс економічної одиниці (фірми, країни, регіони тощо),

t – індекс часу,

Y_{it} – значення досліджуваного показника в період часу t для економічної одиниці.;

$X'_{it} = \{X_{1it}, X_{2it}, \dots, X_{kit}\}$ вектор порядку $(k \times 1)$ пояснюючих змінних (факторів);

a – скаляр;

β_{it} – параметри моделі, що вимірюють часткові ефекти від зміни X'_{it} період часу t для економічної одиниці i ;

ε_{it} – збурення для i -го об'єкта в період часу t .

Для практичного застосування на параметри цієї моделі вводять додаткові обмеження.

В результат специфікації моделі об'єднаних панелей, моделі між групового оцінювання, моделі оцінювання перших різниць зроблено висновок про недоцільність їхнього застосування.

Наступну модель, яку побудовано – модель з фіксованими ефектами. Цей тип моделей, мабуть, найбільш поширений серед моделей панельних даних для застосування в сфері економіки та фінансів. Фактично це особливий випадок моделі найменших квадратів з фіктивними змінними за припущення, що випадкові змінні (коваріати) мають однаковий вплив на залежну змінну. Цю модель застосовують тоді, коли потрібно дослідити динаміку змінних, які коливаються в часу.

Результати моделювання відображено в табл. 1.

Порівнюючи результати цієї моделі із усіма попередніми, слід зазначити, що остання за попередніми висновками є найкращою за більшістю показниками. По-перше, p -значення t -статистики для обох змінних є меншими за 0,01:

$$\Pr(t - value_1) = 1.045e - 15,$$

$$\Pr(t - value_2) = 4.769e - 05,$$

що вказує на те, що у цій моделі з 99,9% ймовірністю обидві змінні є статистично значущі. По-друге, коефіцієнт детермінації є теж вищим, ніж в інших моделях:

$$R^2 = 0.63562.$$

По-третє, p -значення F -статистики є також доволі низьким:

$$P - value(F - Statistics) = 2.22e - 16$$

що дозволяє зробити припущення про те, що із 99,9% ймовірністю ця модель є адекватною.

Дослідимо ще одну модель – модель панельних даних з випадковим ефектом.

Модель випадкових ефектів, на відміну від моделі фіксованих ефектів, виходить із припущення, що відхилення у характеристиках об'єктів випадкові та не корелюють із предиктором чи залежними змінними, включеними у модель. Для оцінювання параметрів такої моделі не можна застосовувати метод най-

Таблиця 1
Специфікація моделі панельних даних з фіксованим ефектом
(fixed effect model)

Модель з фіксованими ефектами					
Код:					
plm(formula = Y ~ X, data = paneldata, model = "within")					
Збалансована панель: n = 22, T = 12, N = 264					
Залишки:					
	Мін	1-й Кв.	Медіана	3-й Кв.	Макс
	-75876,246	-5316,824	-11,723	5541,004	95277,653
Коефіцієнти:					
	Оцінка	Ст.похибка	t.value	Pr(> t)	
X1	579,94456	67,43318	8,6003	1,045E-15	***
X2	1,73051	0,41778	4,1422	4,769E-05	***

Проміжки значень:					
0	****	0,001	***	0,01	**
			0,05	.	0,1
				'	'
					1
Заг. сума квадратів:				9,6828E+10	
Залишкова сума квадратів:				6,0779E+10	
Коеф. детермінації:				0,63562	
Кор. коеф. детермінації:				0,61214	
F-стат.:	71,1726 на 2 та 240 DF,			p-value:	< 2,22e-16

менших квадратів, оскільки він дає зміщені та неконсистентні оцінки. Більш доцільно використовувати узагальнений метод найменших квадратів, який дозволяє врахувати специфічну структуру матриці коваріацій похибок. Модель випадкових відхилень доцільно застосовувати тоді, коли нема підстав припускати, що відхилення у характеристиках об'єктів дослідження мають вплив на залежну змінну. Перевагою моделі випадкових ефектів є можливість включення незалежних від часу змінних (наприклад, гендеру).

Результати реалізації моделі панельних даних із випадковими ефектами наведено у табл. 2.

На перший погляд результати цієї моделі нічим не гірші, ніж результати моделі з фіксованими ефектами:

$$\Pr(t - value_1) = 3.256e - 14 ,$$

$$\Pr(t - value_2) = 4.4769 - 07 ,$$

$$R^2 = 0.61347 ,$$

$$P - value(F - Statistics) = 2.22e - 16$$

Тому потрібно з'ясувати, яка ж із представлених моделей краща, яку модель варто використовувати при оцінюванні валового регіонального продукту через основні сільськогосподарські показники. При порівнянні моделі з фіксованими ефектами та моделі з випадковими ефектами використовують тест Хаусмана (Hausman test).

Основне завдання тесту Хаусмана полягає у знаходженні емпіричного значення H -статистики, порівнянні його з критичним значенням χ_2 , та прийнятті (або відхиленні) нульової гіпотези, яка передбачає те, що саме модель з випадковими ефектами має перевагу і саме її слід брати до уваги. Результати застосування тесту Хаусмана відображено у табл. 3.

Знову ж таки, бачимо, що p -значення (p -value) є меншим, ніж 0,01. Тому відхиляємо

Таблиця 2

Специфікація моделі панельних даних з випадковим ефектом (random effect model)

Модель з випадковими ефектами					
Трансформація Свами-Аропа (Swamy-Arora`s transformation)					
Код:					
plm(formula = Y ~ X, data = paneldata, model = "random")					
Збалансована панель: n = 22, T = 12, N = 264					
Ефекти:					
		Відхилення	Ст.відхил.	Розподіл	
спеціальні		253247678	15914	0,394	
індивідуальні		389006009	19723	0,606	
Тета (θ)	0,7732				
Залишки:					
	Мін	1-й Кв.	Медіана	3-й Кв.	Макс
	-59756,5	-7068,7	-1618,2	5113,4	109835,4
Коефіцієнти:					
	Оцінка	Ст.похибка	t.value	Pr(> t)	
(Перетин)	-4,0986E+04	7,9202E+03	-5,1748	4,55E-07	***
X1	5,2629E+02	6,5504E+01	8,0345	3,256E-14	***
X2	1,9716E+00	3,8073E-01	5,1784	4,476E-07	***

Проміжки значень:					
	0	****	0,001	***	0,01
			**	0,05	' . ' 0,1 ' ' 1
Заг. сума квадратів:				1,0395E+11	
Залишкова сума квадратів:				6,7880E+10	
Коеф. детермінації:				0,61347	
Кор. коеф. детермінації:				0,5942	
F-стат.:	69,3485 на 2 та 261 DF,			p-value:	< 2,22e-16

Порівняння результатів random та fixed effects model тестом Хаусмана

Тест Хаусмана			
дані:	$Y \sim X$		
$\chi^2 =$	19,521	df = 2	p-value = 5,77E-05
Альтернативна гіпотеза:			Приймається

нульову гіпотезу і приймаємо альтернативну, яка свідчить про те, що модель лонгітюдних даних з фіксованими ефектами краще описує залежність валового регіонального продукту від основних сільськогосподарських показників, які були взяті для цієї роботи.

Отже, використавши знайдені оцінки параметрів моделі з фіксованими ефектами, можна записати модель у явному вигляді:

$$Y_{it} = 579.94457X_{1it} + 1.73051X_{2it} + u_{it},$$

де Y_{it} – валовий регіональний продукт в i -ій області в t -му році;

X_{1it} – середня урожайність основних сільськогосподарських культур в i -ій області в t -му році;

X_{2it} – загальна кількість сільськогосподарських тварин в i -ій області в t -му році.

Висновки. Отже, обидва фактори позитивно впливають на валовий регіональний продукт. Тобто збільшення середньої урожайності або/та збільшення кількості сільськогосподарських тварин призводить до збільшення валового регіонального продукту. Пропорції зростання описує модель лонгітюдних даних з фіксованими ефектами. Відношення детермінації для моделі становить 0,63562, що цілком прийнятно для моделі лонгітюдних даних, тому результати моделювання можуть бути використані для ухвалення ефективних рішень в регіональній політиці.

ЛІТЕРАТУРА:

1. Окселекко Н. О. Особливості застосування лонгітюдних даних для сільськогосподарських підприємств. Вісник Полтавської державної аграрної академії. Економіка. 2013. № 4. С. 117-120.
2. Червак-Смерічко, О. Ю. Моделі лонгітюдних даних в емпіричних дослідженнях. Науковий вісник Ужгородського університету : збірник наукових праць. 2013. Вип. 3(40). С. 183-186.
3. Kumbhakar S. C., Gudbrand L., Hardaker J. B. Technical efficiency in competing panel data models: a study of Norwegian grain farming. Journal of Productivity Analysis. 2014. № 41.2 P. 321-337.
4. Zouabi O., Peridy N. Direct and indirect effects of climate on agriculture: an application of a spatial panel data analysis to Tunisia. Climatic change. 2015. № 133.2. P. 301-320.
5. Kumar A., Sharma P., Joshi S. Assessing the impacts of climate change on land productivity in Indian crop agriculture: An evidence from panel data analysis. Journal of Agricultural Science and Technology. 2016. № 18.1. P. 1-13.
6. Galindo L. M., Reyes O. Climate change, irrigation and agricultural activities in Mexico: A Ricardian analysis with panel data. Journal of Development and Agricultural Economics. 2015. № 7.7. P. 262-273.
7. Kassie M. et al. Measuring farm and market level economic impacts of improved maize production technologies in Ethiopia: evidence from panel data. Journal of Agricultural Economics. 2018. № 69.1. P. 76-95.
8. Salim R. Mamun S, Hassan K. Role of communication technologies in broadacre agriculture in Australia: an empirical analysis using panel data. Australian Journal of Agricultural and Resource Economics. 2016. № 60.2. P. 243-264.
9. Antonini C., Argilés-Bosch J. Productivity and environmental costs from intensification of farming. A panel data analysis across EU regions. Journal of cleaner production. 2017. № 140. P. 796-803.
10. Державна служба статистики України. URL : <http://www.ukrstat.gov.ua/> (дата звернення 20.05.2018)
11. The R Project for Statistical Computing. URL : <https://www.r-project.org/> (дата звернення 20.05.2018)
12. RStudio – Open source and enterprise-ready professional software for R. URL^ <https://www.rstudio.com/> (дата звернення 20.05.2018)
13. Лук'яненко І. Г., Городніченко Ю. О. Сучасні економетричні методи у фінансах. Київ: Літера ЛТД. 2002. 352 с.

REFERENCES:

1. Okselenko N. O. (2013) Osoblyvosti zastosuvannia lonhitudnykh danykh dlia silskohospodarskykh pidpriemstv. *Visnyk Poltavskoi derzhavnoi ahranoi akademii. Ekonomika*, 4, 117-120.
2. Chervak-Smerichko, O. Ju. (2013) Modeli longhitjudnykh danykh v empirychnykh doslidzhennjakh. *Naukovyj visnyk Uzhghorodskogo universytetu : zbirnyk naukovykh pracj*, 3(40), 183-186.
3. Kumbhakar, S. C., Lien, G., & Hardaker, J. B. (2014). Technical efficiency in competing panel data models: a study of Norwegian grain farming. *Journal of Productivity Analysis*, 41(2), 321-337.
4. Zouabi, O., & Peridy, N. (2015). Direct and indirect effects of climate on agriculture: an application of a spatial panel data analysis to Tunisia. *Climatic change*, 133(2), 301-320.
5. Kumar, A., P. Sharma, and S. Joshi. "Assessing the impacts of climate change on land productivity in Indian crop agriculture: An evidence from panel data analysis." *Journal of Agricultural Science and Technology* 18.1 (2016): 1-13.
6. Kumar, A., Sharma, P., & Joshi, S. (2016). Assessing the impacts of climate change on land productivity in Indian crop agriculture: An evidence from panel data analysis. *Journal of Agricultural Science and Technology*, 18(1), 1-13.
7. Kassie, M., Marenya, P., Tessema, Y., Jaleta, M., Zeng, D., Erenstein, O., & Rahut, D. (2018). Measuring farm and market level economic impacts of improved maize production technologies in Ethiopia: evidence from panel data. *Journal of Agricultural Economics*, 69(1), 76-95.
8. Salim, R., Mamun, S. A. K., & Hassan, K. (2016). Role of communication technologies in broadacre agriculture in Australia: an empirical analysis using panel data. *Australian Journal of Agricultural and Resource Economics*, 60(2), 243-264.
9. Antonini, C., & Argilés-Bosch, J. M. (2017). Productivity and environmental costs from intensification of farming. A panel data analysis across EU regions. *Journal of cleaner production*, 140, 796-803.
10. State Statistics Service of Ukraine. URL : <http://www.ukrstat.gov.ua/> (date 20.05.2018)
11. The R Project for Statistical Computing. URL: <https://www.r-project.org/> (date 20.05.2018)
12. RStudio – Open source and enterprise-ready professional software for R. URL: <https://www.rstudio.com/> (date 20.05.2018)
13. Luk'janenko I. Gh., Ju. O. Ghorodnichenko (2002) *Suchasni ekonometrychni metody u finansakh*. Kyiv, Litera LTD, 352.