



УДК 330.3:519.86

УСУНЕННЯ МУЛЬТИКОЛІНЕАРНОСТІ ЗМІННИХ У ПРОЦЕСІ ІДЕНТИФІКАЦІЇ МОДЕЛЕЙ СОЦІАЛЬНО-ЕКОНОМІЧНИХ ПРОЦЕСІВ

О. Белз

*Львівський національний університет імені Івана Франка
79008 м. Львів, проспект Свободи, 18
E-mail: homepc@list.ru*

Досліджено вплив мультиколінеарності економічних показників на прогнозні характеристики економетричних моделей. Запропоновано для вибору суттєвих факторів моделі застосовувати методи статистичного аналізу даних.

Ключові слова: ідентифікація моделей, моделювання соціально-економічних процесів, прогнозування значень економічних показників.

Постановка проблеми. У процесі побудови моделей макроекономічних процесів важливим є вдалий вибір методів моделювання, які б давали змогу врахувати складність та динамічність зміни взаємозв'язків між об'єктами економічної системи. Процес ідентифікації моделі системи (побудови моделі системи за даними її функціонування) є надзвичайно складним.

Аналіз останніх досліджень і публікацій та невирішені частини проблеми. Дж. Бокс та Г. Дженкінс зазначають, що «питання про те, які типи моделей зустрічаються на практиці і у яких обставинах – це риса поведінки фізичного світу, і вона не може бути розв'язана чисто математичними міркуваннями» [1, с. 193]. Складність процесу ідентифікації моделей макроекономічних процесів поглиблюють обмежені можливості формування статистичної вибірки з огляду на те, що існує визначена періодичність формування макроекономічних показників, а збільшення статистичної вибірки за рахунок розширення періоду спостереження не завжди є коректним через зміну методики розрахунку таких показників.

Однією із проблем у процесі побудови економетричних моделей є наявність мультиколінеарності незалежних (пояснювальних) змінних. Ю. Толбатов зазначає, що у випадку строгої мультиколінеарності неможливо отримати оцінки параметрів моделі методом найменших квадратів, а у випадку нестрогої мультиколінеарності оцінки параметрів малонадійні [8]. С. Наконечний, Т. Терещенко та Т. Романюк серед наслідків мультиколінеарності виокремлюють падіння точності оцінювання параметрів моделі [5]. В. Доля вказує на те, що мультиколінеарність не дає змоги обґрунтовано оцінити параметри рівняння регресії [2]. Н. Паклін і В. Орешков зазначають, що мультиколінеарність спричиняє нестійкість та неоднозначність регресійної моделі [6]. Проте цілком об'єктивним є те, що макроекономічні показники пов'язані між собою і мають певний ступінь кореляції.

І. Й. Яремко, проаналізувавши стан розвитку економетричної науки, стверджує, що «оцінюючи сучасну проблематику адекватності економетричних моделей в



економіці, можемо зробити висновок, що створення конструктивної комплексної методики, яка враховує як об'єктивні особливості змодельованих об'єктів, так і особливості їхнього пізнання, є одним з найактуальніших завдань економіко-математичних досліджень» [9].

Мета дослідження. Метою цієї роботи є дослідження впливу мультиколінеарності пояснювальних змінних на прогнозні характеристики моделей соціально-економічних процесів.

Вхідними даними для дослідження є помісячні значення чотирнадцяти макроекономічних показників з січня 2006 р. по грудень 2013 р. (на підставі даних Держкомстату України [7]).

Інструментарієм дослідження є програмні засоби Statgraphics та MS Excel.

Основні результати дослідження. Для дослідження мультиколінеарності незалежних змінних застосовано алгоритм Фаррара-Глобера [5], який підтвердив наявність мультиколінеарності. Для усунення мультиколінеарності обрано метод головних компонентів (МГК), так як цей метод призначений для оцінювання моделей великого розміру, якщо до таких моделей входять мультиколінеарні змінні [5; 6]. У роботі проведено компонентний аналіз за допомогою програмного засобу Statgraphics.

Ефективність застосування МГК у процесі побудови моделей запропоновано порівняти з методами статистичного аналізу даних, за допомогою яких здійснюють вибір набору пояснювальних (екзогенних) змінних моделей.

З метою дослідження прогнозних характеристик побудованих моделей із статистичної вибірки виокремлено три частини: навчальну (дані з січня 2006 р. по грудень 2011 р.), тестову (дані з січня 2012 р. по грудень 2012 р.) та контрольну (дані з січня 2013 р. по грудень 2013 р.). На підставі даних навчальної вибірки розраховують коефіцієнти моделей, на підставі даних тестової вибірки оцінюють якість прогнозування, а на підставі даних контрольної вибірки оцінюють тенденцію збереження якості прогнозування.

У процесі побудови моделей соціально-економічних процесів використано підхід К. Сімса, запропонований у 1980 р., який має назву VAR або векторна авторегресія [3]. К. Сімс запропонував подавати кожну із компонент багатовимірного випадкового процесу як лінійну комбінацію від попередніх значень усіх змінних. Цей підхід передбачає побудову системи рівнянь такого виду:

$$\begin{cases} \tilde{X}_t^1 = \alpha_1 + \beta_{11} \cdot X_{t-1}^1 + \beta_{12} \cdot X_{t-1}^2 + \dots + \beta_{1m} \cdot X_{t-1}^m \\ \tilde{X}_t^2 = \alpha_2 + \beta_{21} \cdot X_{t-1}^1 + \beta_{22} \cdot X_{t-1}^2 + \dots + \beta_{2m} \cdot X_{t-1}^m \\ \dots \\ \tilde{X}_t^m = \alpha_m + \beta_{m1} \cdot X_{t-1}^1 + \beta_{m2} \cdot X_{t-1}^2 + \dots + \beta_{mm} \cdot X_{t-1}^m \end{cases} \quad (1)$$

де $\tilde{X}_t^1, \tilde{X}_t^2, \dots, \tilde{X}_t^m$ – теоретичні значення макроекономічних показників у часі t ; $X_{t-1}^1, X_{t-1}^2, \dots, X_{t-1}^m$ – фактичні значення макроекономічних показників у часі $(t-1)$; $\alpha_1, \alpha_2, \dots, \alpha_m$ та $\beta_{11}, \beta_{12}, \dots, \beta_{mm}$ – коефіцієнти побудованих моделей, розраховані за допомогою методу найменших квадратів; m – кількість макроекономічних показників.

Для ідентифікації моделей соціально-економічних процесів методами статистичного аналізу даних запропоновано введення пояснювальних змінних у



кожне рівняння системи (1) здійснювати поступово по мірі спадання абсолютних значень парних коефіцієнтів кореляції між фактичним значенням кожного з макроекономічних показників у часі $(t-1)$ та фактичним значенням обраного макроекономічного показника у часі t . Як результат для кожного з чотирнадцяти макроекономічних показників побудовано по одній парній лінійній регресійній моделі та по тринадцять множинних лінійних регресійних моделей (усього 196 економетричних моделей).

Для ідентифікації моделей соціально-економічних процесів із застосуванням компонентного аналізу було виокремлено головні компоненти на підставі як нормалізованих, так і ненормалізованих значень змінних. Як результат для кожного з чотирнадцяти макроекономічних показників на підставі головних компонентів з нормалізованих значень змінних побудовано по одній парній лінійній регресійній моделі та по тринадцять множинних лінійних регресійних моделей (усього 196 економетричних моделей). Також для кожного з чотирнадцяти макроекономічних показників на підставі головних компонентів з ненормалізованих значень змінних побудовано по одній парній лінійній регресійній моделі та по тринадцять множинних лінійних регресійних моделей (усього 196 економетричних моделей).

Для оцінення якості побудованих моделей Г. Г. Канторович у [4] рекомендує застосовувати коефіцієнт множинної детермінації R^2 .

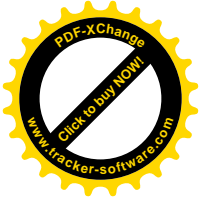
Ефективність застосування МГК та методу статистичного аналізу даних порівнювано з моделлю, на якій досягається максимуму значення коефіцієнта множинної детермінації R^2 , розрахованого на підставі даних навчальної вибірки. Після проведення необхідних розрахунків виявлено, що значення коефіцієнта множинної детермінації R^2 на підставі даних навчальної вибірки для кожного з чотирнадцяти макроекономічних показників по мірі введення у модель кожної наступної пояснювальної змінної покращується. Найкраще значення коефіцієнта множинної детермінації R^2 на підставі даних навчальної вибірки досягається, коли у економетричну модель уведено всі чотирнадцять макроекономічних показників. Тому для кожного з чотирнадцяти макроекономічних показників побудовано по одній множинній лінійній регресійній моделі (1), у яку входять всі чотирнадцять незалежних змінних.

Для оцінення прогнозних характеристик побудованих моделей будемо розраховувати середню помилку прогнозування (у відсотках) за формулою:

$$SP = \frac{\sum_{i=1}^N \frac{|y_i - \tilde{y}_i|}{y_i} \cdot 100}{N}, \tag{2}$$

де SP – середня помилка прогнозування; y_i – фактичне значення показника в i -ий період; \tilde{y}_i – теоретичне значення показника в i -ий період; N – кількість періодів.

Значення середньої помилки прогнозування розрахуємо окремо на підставі даних навчальної, тестової та контрольної вибірок (див. табл. 1).



Таблиця 1

Прогнозні характеристики моделей, $min SP$ на тестовій вибірці

Метод аналізу даних	Кількість пояснювальних змінних (компонент)	R^2 (на навчальній вибірці)	SP (на навчальній вибірці), %	SP (на тестовій вибірці), %	SP (на контрольній вибірці), %
1. Модель обсягу реалізованої промислової продукції					
Регресійна модель зі всіма незалежними змінними	14	0,6074	22,05	36,24	62,87
Статистичний	7	0,4723	20,53	15,95	36,09
МГК (ненормалізовані змінні)	1	0,0189	21,14	3,86	10,72
МГК (нормалізовані змінні)	1	0,2197	19,03	11,16	18,05
2. Модель обсягу продукції сільського господарства					
Регресійна модель зі всіма незалежними змінними	14	0,5476	59,49	98,29	172,28
Статистичний	2	0,3286	59,48	52,27	54,89
МГК (ненормалізовані змінні)	6	0,4257	57,39	65,31	98,38
МГК (нормалізовані змінні)	8	0,3918	60,40	71,09	67,19
3. Модель обсягу роздрібногo товарообороту					
Регресійна модель зі всіма незалежними змінними	14	0,9429	6,62	11,04	23,45
Статистичний	2	0,8262	9,69	7,51	9,06
МГК (ненормалізовані змінні)	5	0,8731	7,81	6,50	13,17
МГК (нормалізовані змінні)	12	0,9307	7,26	6,67	13,15
4. Модель грошової маси М3					
Регресійна модель зі всіма незалежними змінними	14	0,9976	1,23	1,61	1,00
Статистичний	8	0,9974	1,23	1,32	0,89
МГК (ненормалізовані змінні)	2	0,9919	1,90	1,17	0,71
МГК (нормалізовані змінні)	14	0,9976	1,23	1,61	1,00
5. Модель доходів зведеного бюджету					
Регресійна модель зі всіма незалежними змінними	14	0,7477	12,32	15,47	32,10
Статистичний	2	0,6665	13,69	12,24	16,23
МГК (ненормалізовані змінні)	6	0,7000	12,66	12,20	24,88
МГК (нормалізовані змінні)	12	0,7425	12,47	11,81	22,31



Продовження таблиці 1

6. Модель видатків зведеного бюджету					
Регресійна модель зі всіма незалежними змінними	14	0,8385	10,90	14,35	23,31
Статистичний	7	0,7559	12,59	11,80	16,75
МГК (ненормалізовані змінні)	2	0,5844	15,00	11,40	11,91
МГК (нормалізовані змінні)	3	0,5759	15,59	11,16	12,50
7. Модель індексу споживчих цін					
Регресійна модель зі всіма незалежними змінними	14	0,5068	0,56	0,57	0,40
Статистичний	2	0,4202	0,62	0,19	0,22
МГК (ненормалізовані змінні)	13	0,4923	0,58	0,28	0,39
МГК (нормалізовані змінні)	9	0,4813	0,59	0,15	0,43
8. Модель індексу цін виробників промислової продукції					
Регресійна модель зі всіма незалежними змінними	14	0,4632	1,07	1,71	3,24
Статистичний	2	0,3226	1,13	1,34	1,71
МГК (ненормалізовані змінні)	5	0,1179	1,29	1,29	2,28
МГК (нормалізовані змінні)	9	0,3755	1,09	1,12	1,84
9. Модель чисельності громадян, зареєстрованих як безробітні					
Регресійна модель зі всіма незалежними змінними	14	0,9275	5,66	22,76	93,79
Статистичний	3	0,8565	6,85	13,54	54,32
МГК (ненормалізовані змінні)	4	0,3407	18,25	18,78	87,45
МГК (нормалізовані змінні)	11	0,8881	6,87	13,98	37,80
10. Модель рівня безробіття					
Регресійна модель зі всіма незалежними змінними	14	0,9467	4,45	5,50	5,97
Статистичний	14	0,9467	4,45	5,50	5,97
МГК (ненормалізовані змінні)	14	0,9467	4,45	5,50	5,97
МГК (нормалізовані змінні)	14	0,9467	4,45	5,50	5,97
11. Модель середньомісячної номінальної заробітної плати працівників					
Регресійна модель зі всіма незалежними змінними	14	0,9971	0,97	1,08	1,07
Статистичний	6	0,9966	0,95	0,53	0,62
МГК (ненормалізовані змінні)	10	0,9969	0,92	0,48	0,96
МГК (нормалізовані змінні)	13	0,9970	0,95	0,52	0,58



Продовження таблиці 1

12. Модель середньозваженої ставки рефінансування за всіма інструментами					
Регресійна модель зі всіма незалежними змінними	14	0,8607	7,22	11,07	49,54
Статистичний	8	0,8438	7,54	5,65	27,48
МГК (ненормалізовані змінні)	12	0,8598	7,25	9,76	33,83
МГК (нормалізовані змінні)	9	0,8259	7,69	7,69	9,52
13. Модель внутрішнього кредиту					
Регресійна модель зі всіма незалежними змінними	14	0,9985	1,37	1,04	1,06
Статистичний	1	0,9966	1,96	0,81	0,80
МГК (ненормалізовані змінні)	14	0,9985	1,37	1,04	1,06
МГК (нормалізовані змінні)	13	0,9985	1,40	0,95	0,87
14. Модель депозитів резидентів, залучених депозитними корпораціями окрім Національного банку України					
Регресійна модель зі всіма незалежними змінними	14	0,9969	1,43	1,02	1,69
Статистичний	8	0,9966	1,42	0,85	2,08
МГК (ненормалізовані змінні)	8	0,9956	1,54	0,91	0,54
МГК (нормалізовані змінні)	14	0,9969	1,43	1,02	1,69

За результатами проведеного дослідження виявлено, що середня помилка прогнозування менше трьох відсотків на підставі даних навчальної, тестової та контрольної вибірок досягається лише для шістьох моделей (див. табл. 1). Зауважимо, що дані контрольної вибірки підтвердили збереження прогнозних характеристик побудованих моделей.

Також виявлено, що високе значення коефіцієнта множинної детермінації R^2 (0,83–0,95), розраховане на підставі даних навчальної вибірки, не гарантує хороших прогнозних характеристик моделі, зокрема, для моделей таких макроекономічних показників: обсяг роздрібногo товарообороту; рівень безробіття; середньозважена ставка рефінансування за всіма інструментами. Значення середньої помилки прогнозування перелічених моделей на підставі даних тестової вибірки є від 5,5% до 11,1%. Окрім того, моделі з низьким значенням коефіцієнта множинної детермінації R^2 (0,12–0,51), розрахованим на підставі даних навчальної вибірки, дають хороші прогнозні характеристики (значення середньої помилки прогнозування на підставі даних тестової вибірки є від 0,15% до 1,71%). Це моделі таких макроекономічних показників: індекс споживчих цін; індекс цін виробників промислової продукції.

За результатами виконаного дослідження проведемо порівняння ефективності застосування методів аналізу даних (див. табл. 2).

Як видно з таблиці 2, у жодному випадку не вдалося досягнути мінімального значення середньої помилки прогнозування тих моделей макроекономічних показників, у які входять всі незалежні змінні (і за цих умов досягається максимальне значення коефіцієнта множинної детермінації R^2 , розраховане на підставі даних



навчальної вибірки). Статистичний аналіз даних у більшості випадків дав кращі результати, аніж застосування компонентного аналізу як на підставі нормалізованих, так і на підставі ненормалізованих значень змінних. Переваги статистичного аналізу даних спостерігаються як для моделей, середня помилка прогнозування яких менше трьох відсотків, так і для моделей, середня помилка прогнозування яких більше трьох відсотків. Окрім того, такі переваги є для моделей, середня помилка прогнозування яких була розрахована як на підставі даних навчальної вибірки, так і на підставі даних тестової вибірки.

Таблиця 2

Порівняння ефективності застосування методів аналізу даних

Дані для аналізу	Переваги застосування методу аналізу даних			
	Max R^2 (на підставі даних навчальної вибірки)	Статистичний	МГК (ненормалізовані змінні)	МГК (нормалізовані змінні)
Навчальна вибірка для всіх моделей	0	4	4	3
Навчальна вибірка для моделей, де середня помилка прогнозування менше 3 % (шість моделей)	0	3	1	1
Тестова вибірка для всіх моделей	0	5	4	4
Тестова вибірка для моделей, де середня помилка прогнозування менше 3 % (шість моделей)	0	2	2	2
Всього переваг	0	14	11	10

Висновки. Підсумовуючи зазначимо, що значення коефіцієнту множинної детермінації побудованих моделей на підставі даних навчальної вибірки по мірі введення у модель кожної наступної пояснювальної змінної покращується незалежно від того, які були застосовані методи аналізу даних. Високе значення коефіцієнту множинної детермінації моделі, розраховане на підставі даних навчальної вибірки, не гарантує хороших прогнозних характеристик такої моделі.

Застосування МГК та методів статистичного аналізу даних, за допомогою яких здійснюють вибір набору пояснювальних змінних, значно покращує прогнозні характеристики побудованих економетричних моделей. Так як у більшості випадків краща модель досягається на підставі методів статистичного аналізу даних та з огляду на обсяг обчислювальних ресурсів, які вимагає застосування компонентного аналізу, ідентифікувати моделі соціально-економічних процесів, усуваючи мультиколінеарність незалежних змінних МГК, не завжди є доцільним.



1. Бокс Дж. Анализ временных рядов. Прогноз и управление / Дж. Бокс, Г. Дженкинс ; пер. с англ. А. Л. Левшина ; под ред. В. Ф. Писаренко. – [вып. 1]. – М. : Мир. 1974. – 408 с.
2. Доля В. Т. Економетрія: навч. посібник / В. Т. Доля; Харк. нац. акад. міськ. госп-ва. – Х.: ХНАМГ, 2010. – 171 с.
3. Канторович Г. Г. Анализ временных рядов / Г. Г. Канторович // Экономический журнал ВШЭ. – 2002. – № 4. – С. 498–523.
4. Канторович Г. Г. Анализ временных рядов / Г. Г. Канторович // Экономический журнал ВШЭ. – 2002. – № 2. – С. 251–273.
5. Наконечний С. І. Економетрія: підручник / С. І. Наконечний, Т. О. Терещенко, Т. П. Романюк. – [вид. 2-ге, допов. та перероб.]. – К. : КНУ, 2000. – 296 с.
6. Паклин Н. Б. Бизнес-аналитика: от данных к знаниям (+CD): Учеб. пособие / Н. Б. Паклин, В. И. Орешков. – [2-е изд., перераб. и доп.]. – СПб.: Питер, 2010. – 704 с.
7. Статистичний бюлетень (електронне видання). – Режим доступу: http://www.bank.gov.ua/control/uk/publish/category?cat_id=57897
8. Толбатов Ю. А. Економетрика: Підручник для студентів екон. спеціальн. вищ. навч. закл / Ю. А. Толбатов. – К.: Четверта хвиля, 1997. – 320 с.
9. Яремко І.Й. 100-ліття витоків економетричної науки / І.Й. Яремко // Менеджмент та підприємництво в Україні: етапи становлення і проблеми розвитку : [збірник наукових праць] / відповідальний редактор О. Є. Кузьмін. – Львів : Видавництво Львівської політехніки, 2010. – 416 с. : іл. - (Вісник Національного університету «Львівська політехніка»; № 691). – С. 400–406. – Режим доступу: <http://ena.lp.edu.ua:8080/bitstream/ntb/8649/1/63.pdf>

ELIMINATE THE MULTICOLLINEARITY OF THE VARIABLES IN THE PROCESS OF IDENTIFYING MODELS OF SOCIO-ECONOMIC PROCESSES

O. Belz

*Ivan Franko National University of Lviv,
Prospekt Svobody 18, UA – 79008, Ukraine*

The author discovers problem of influence of multicollinearity of economic indicators on the prognostic characteristics of econometric models. The author suggests to use statistical analysis techniques for choice significant factors of model.

Keywords: identification of models, modeling social and economic processes, forecast values of economic indicators.



**УСТРАНЕНИЕ МУЛЬТИКОЛЛИНЕАРНОСТИ ПЕРЕМЕННЫХ В
ПРОЦЕССЕ ИДЕНТИФИКАЦИИ МОДЕЛЕЙ СОЦИАЛЬНО-
ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ**

А. Белз

*Львовский национальный университет имени Ивана Франко
79008 г. Львов, проспект Свободы, 18*

Исследовано влияние мультиколлинеарности экономических показателей на прогнозные характеристики эконометрических моделей. Предложено для выбора существенных факторов модели применять методы статистического анализа данных.

Ключевые слова: идентификация моделей, моделирование социально-экономических процессов, прогнозирование значений экономических показателей.