

УДК 519.7

Д.А. ОМЕЛЬЯНЧИК

## ЕМПІРИЧНА ВАЛІДАЦІЯ АГЕНТНО-ОРІЄНТОВАНИХ МОДЕЛЕЙ НА ПРИКЛАДІ МОДЕЛІ ПРОМИСЛОВОГО ВИРОБНИЦТВА У ПОЛЬЩІ

***Анотація.** Обговорюються основні проблеми емпіричної валідації та прогнозування в агентно-орієнтованих моделях. Запропоновано модель залежності обсягів промислового виробництва від розмірів грошових субсидій та кількості працівників на базі реальних статистичних даних. Проведено обчислювальний експеримент щодо оцінки якості різних сценаріїв роботи моделі за допомогою критеріїв  $R^2$  та MAPE. Проаналізовано чутливість використаних підходів до збурень даних.*

***Ключові слова:** математичне моделювання, агентно-орієнтоване моделювання, математична економіка, інтелектуальні агенти.*

### Вступ

Гетерогенність, обмежена раціональність, здатність до навчання – ці властивості економічних агентів стали основою критики традиційних (зокрема, неокласичних) підходів до моделювання економічних процесів та підставою для існування альтернативного методу – агентно-орієнтованого моделювання. Великі очікування щодо можливостей агентного підходу під час опису та пояснення найрізноманітніших процесів призвели до його стрімкого поширення в суспільних науках протягом останніх двадцяти років. Однак його слабкі сторони – складність емпіричної валідації та спірність придатності до кількісного прогнозування – знижують можливості практичних застосувань порівняно із моделями, які базуються на неокласичному та кейнсіанському підходах.

Мета статті – висвітлити основні питання, що виникають під час емпіричної валідації агентно-орієнтованих моделей, а також на прикладі моделі мікроповедінки, розробленої на основі реальних статистичних даних, оцінити ступінь відтворення нею макрозалежностей, проаналізувати її чутливість до можливих збурень та порівняти отримані результати із результатами традиційних економетричних підходів.

Показано, що можливості агентно-орієнтованого підходу не вичерпуються якісним аналізом, а, у випадку орієнтованості на фактичні дані, його припустимо застосовувати і як засіб кількісного моделювання.

Під емпіричною валідацією розуміють перевірку відповідності моделі реальним (емпіричним) даним. Тобто дані, згенеровані за допомогою обчислювальної моделі, що спрощено відображає реальне явище, порівнюються із деяким набором результатів емпіричного спостереження цього явища та оцінюється якість отриманої апроксимації.

Джерелом проблем під час емпіричної валідації агентно-орієнтованих моделей є нелінійність та випадковість індивідуальних поведінок агентів та міжагентних взаємодій, неможливість аналітичного аналізу складних

стохастичних процесів утворення змінних, зворотні зв'язки між мікро- та макрорівнем тощо.

Найбільш відомими підходами до емпіричної валідації агентно-орієнтованих моделей є непряме калібрування, метод Веркера-Бреннера та історичний метод. Кожний з них пропонує свої способи скоротити кількість модельних параметрів та обмежити простір досліджуваних «штучних світів» [7]. Так, історичний метод обмежує параметри, взаємодії та правила прийняття рішень у моделі згідно з конкретною, емпірично спостережуваною історією певної сфери економічної діяльності. Інші два підходи не накладають попередніх обмежень на параметри, але опосередковано використовують емпіричні дані для визначення підобластей у просторі потенційних параметрів. В рамках цих підобластей модель має відтворювати деякі значимі статистичні закономірності або «відомі факти»<sup>1</sup> [2, 3, 5].

Ще один підхід до перевірки моделей запропоновано у [6]. Його ідея полягає в побудові бази правил, яка описує структуру реальних статистичних даних. Далі ці правила можна використовувати для моделювання індивідуальної поведінки агентів та валідації агрегованої поведінки моделі.

Важливим наслідком питання про можливість емпіричної валідації моделі є її придатність до генерування кількісних прогнозів. Якщо вдається емпірично валідувати модель індивідуальної поведінки агентів на мікрорівні і встановити, як вона впливає на розвиток агрегованої макрозмінної, чи достатньо такої моделі для побудови прогнозу значень цієї змінної? Наскільки стійкими до збурень початкових умов чи параметрів моделі є отримані прогнози? Як вони залежать від застосованих методів агрегування макрозмінної [4]?

Створення агентних моделей зазвичай мотивується прагненням зрозуміти якісні, а не кількісні характеристики деякого явища. Більше того, розробники агентних моделей часто свідомо відмовляються від традиційних функціональних підходів до побудови економічних моделей, в тому числі від використання виробничих функцій, функцій корисності, явного вигляду функціональних залежностей між обсягами попиту та пропозиції. Зазначимо, що подібний підхід не означає принципову неможливість застосування відомих детермінованих залежностей між макроекономічними показниками під час розробки агентно-орієнтованих моделей, а, швидше, свідчить про бажання дослідників відтворити ці залежності, а можливо і знайти якісь нові, шляхом моделювання поведінки економічних агентів на мікрорівні.

Саме тому застосування загальноприйнятих економетричних підходів часто може перевершувати кількісні результати агентних моделей [1]. Запропонована нами агентно-орієнтована модель обсягів промислового виробництва в Польщі, побудована з урахуванням реальних даних, об'єднує класичний регресійний аналіз із інструментарієм агентно-орієнтованого підходу і має потенціал до покращення кількісних результатів стандартних методів прогнозування.

---

<sup>1</sup> Під «відомими фактами» тут розуміються стійкі структурні властивості та співвідношення, що підтверджуються у різних джерелах емпіричних даних.

## 1. Агентно-орієнтована модель залежності обсягів реалізованої промислової продукції від отриманих субсидій

Розглянемо агентно-орієнтовану модель, яка призначена для відтворення та аналізу залежності обсягів промислового виробництва від розмірів грошових субсидій, отриманих підприємствами. Окремо моделюється також механізм підбору нових працівників (ринок праці) та розподілу субсидій.

Нехай активними агентами цієї моделі виступають  $N$  фірм-виробників промислової продукції із параметрами  $S_i^t$  – обсяг субсидій, отриманих  $i$ -ю фірмою на ітерації  $t$ , та  $L_i^t$  – відповідна кількість працівників, тоді предметом аналізу буде динаміка значень загальної виручки агентів

$$Y^t = \sum_{i=1}^N Y_i^t = \sum_{i=1}^N f(S_i^t, L_i^t).$$

На кожній ітерації послідовно виконуються такі дії: регулювання кількості працівників, розподіл субсидій, розрахунок отриманої виручки. Кожна з цих дій може бути реалізована різними способами та незалежно від інших дій. Нами розглянуті наступні варіанти виконання таких дій.

Число працівників регулюється на основі аналізу статистичних даних про різницю між сукупною чисельністю працівників ( $\hat{L}$ ) у попередньому та поточному періоді  $\hat{L}^t - \hat{L}^{t-1}$ . Якщо  $\hat{L}^t - \hat{L}^{t-1} > 0$ , то підприємства набирають працівників, якщо ж  $\hat{L}^t - \hat{L}^{t-1} < 0$  – звільняють. Фірма-кандидат на зміну кількості працівників вибирається випадковим чином для кожного нового працівника. Цей процес триває доти, доки всіх нових працівників не буде розподілено між існуючими фірмами, тобто  $\sum_{i=1}^N L_i^t = \hat{L}^t$ . Зауважимо, що у випадку від'ємної різниці, що відповідає необхідності звільняти працівників, фірма-кандидат з одним працівником вилучається з моделі.

Результати роботи моделі в залежності від застосування різних сценаріїв розподілу субсидій є найбільш цікавими для аналізу державної підтримки підприємств. Загальний обсяг субсидій  $\hat{S}^t = \sum_{i=1}^N S_i^t$  визначається зі статистичної звітності. На даному етапі розробки моделі розглядається лише два сценарії розподілу субсидій: величина субсидій або однакова для всіх підприємств моделі ( $S_i^t = \hat{S}^t/N$ ), або встановлюється на основі використання нормального розподілу із заданими математичним сподіванням та дисперсією ( $S_i^t \sim N(ES_i^t, DS_i^t)$ ). В подальшому набір сценаріїв планується розширити за рахунок включення до розгляду регіональних особливостей.

Отже, володіючи інформацією про розмір наданих субсидій  $S_i^t$  та поточну чисельність працівників  $L_i^t$  на заданій ітерації  $t$ , кожне підприємство  $i$  може розрахувати власну виручку від реалізації продукції  $Y_i^t$  за загальною формулою  $Y_i^t = f_i(S_i^t, L_i^t)$ . Оскільки мета цієї статті порівняти результати моделювання за допомогою традиційних та агентно-орієнтованих методів, в

якості функції  $f_i$  вибрано лінійну комбінацію двох змінних параметрів фірми  $Y_i^t = \alpha_i^t S_i^t + \beta_i^t L_i^t$ , причому параметри  $\alpha_i$  та  $\beta_i$  визначені на основі зібраних статистичних даних за допомогою лінійної або байєсівської регресії. В рамках обчислювального експерименту розглянуто декілька варіантів побудови регресійних моделей.

## 2. Схема обчислювального експерименту

Емпірична валідація розробленої моделі здійснена з використанням масиву статистичних даних Республіки Польща [8]. Ця країна була вибрана, по-перше, через високу якість статистичної інформації у відкритому доступі, а по-друге, тому що з країн-сусідів України саме Польща відома значними обсягами грошових субсидій, спрямованих на розвиток та підтримку місцевих підприємств за державними програмами та грантами ЄС.

За результатами аналізу наявних даних в якості реального аналогу для отриманих субсидій вибрано показник «Інші операційні доходи – субсидії» (other operating revenues – subsidies), а для поточної чисельності працівників – показник «Середня кількість штатних працівників» (average paid employment in industry). Виручка від реалізації промислової продукції підприємств відображена показником «Доходи від продажу продукції, товарів та матеріалів» (revenues from sale of products, goods and materials). Зазначені дані надаються в кварталному розрізі, починаючи з 2008 року. До уваги бралася інформація стосовно всіх промислових підприємств та економіки Польщі на національному рівні. В подальшому планується розширити можливості моделі з урахуванням регіональної специфіки.

Для побудови початкового розподілу працівників по фірмах розглянуто три групи підприємств: малі (1–49), середні (50–250) та великі (понад 250 штатних працівників). Кількість підприємств кожної групи у моделі є одним з параметрів сценарію. В проведеному обчислювальному експерименті розглядалися такі варіанти розподілу: 10 (малі) – 10 (середні) – 10 (великі), 1000 – 10 – 10, 100 – 100 – 100, 1000 – 1000 – 1000, 5000 – 5000 – 5000, 10000 – 1000 – 100, 10000 (1-9 працівників) – 1000 (10-49) – 100 (50-249) – 10 (250 і більше). Особливу цікавість має сценарій 169162 – 6156 – 1510, оскільки таким є фактичний розподіл підприємств за кількістю працівників у Польщі.

Інші сценарні параметри обчислювального експерименту асоційовані зі способами наближення функції  $f$ , яка пов'язує отриману виручку від реалізації із субсидіями та чисельністю працівників. Апроксимувати цю функцію пропонується за допомогою апарату байєсівської або лінійної регресії. Крім того, розглянуто два набори даних для побудови регресії: абсолютні (фактичні) або усереднені (нормовані на загальну кількість підприємств у кожному періоді). В останньому випадку регресійна модель створюється, по суті, для якогось «середнього» польського промислового підприємства.

Важливо проаналізувати чутливість розробленої моделі до малих збурень параметрів регресії та даних. Для моделювання сценарію малих збурень параметрів кожне підприємство отримувало коефіцієнти регресії із випадковим відхиленням в інтервалі  $\pm 5\%$ . Щоб змоделювати збурення

отриманих даних, виручка фірми-агента обчислювалася з внесенням випадкових відхилень в такому ж інтервалі.

Щоб перевірити відтворюваність результатів та зменшити системну помилку, для кожного з 298 сценаріїв обчислювальний експеримент здійснювався по 100 разів. Виключення було зроблено для сценаріїв із близько 200 000 підприємств через значну обчислювальну складність розрахунків, недоступну для швидкої перевірки на персональному комп'ютері.

Для оцінки якості розробленої моделі використано два критерії:  $R^2$  (частка поясненої дисперсії) та MAPE (Mean Absolute Percentage Error, або середня абсолютна відсоткова помилка). Результати роботи різних сценаріїв моделі порівнюються з аналогічними оцінками звичайної лінійної регресії для загальних даних.

### 3. Аналіз результатів обчислювального експерименту

Проведений обчислювальний експеримент мав декілька цілей: по-перше, порівняти результати, отримані за допомогою регресійної та агентно-орієнтованої моделей, по-друге, проаналізувати ефект від застосування різних сценаріїв, по-третє, оцінити чутливість моделі до малих збурень вхідної інформації.

Результати моделювання відображалися за допомогою графіка типу «ящик з вусами» («box-and-whiskers plot») (рис. 1). Тут кожний «ящик» («box») відображає розподіл оцінки MAPE, отриманої в результаті обчислювального експерименту для різних сценаріїв і відповідного розподілу фірм. Пунктирною лінією подано значення похибки MAPE для базової моделі – лінійної регресії за сукупними (загальними) даними без жодних збурень.

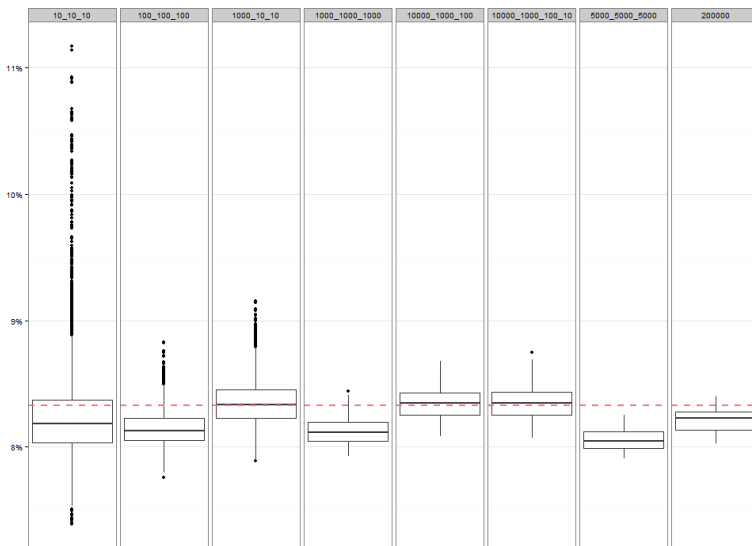


Рисунок 1 – MAPE для моделей з різними розподілами фірм

У п'яти з восьми розглянутих сценаріїв розподілу фірм середнє значення MAPE було меншим, ніж аналогічний показник для базової моделі, причому

незначно гірший результат спостерігався і для фірм з нерівномірним розподілом на групи. При цьому відмічено, що включення в модель більшої кількості фірм суттєво зменшує інтерквартильний розмах оцінки, а кращі екстремальні значення загалом спостерігаються для малої кількості фірм.

Детальні результати для частки поясненої дисперсії  $R^2$  наведено у табл. 1. Результати перевірки стійкості класичних моделей до різних збурень відображає варіант «1» (одна фірма), вони обчислювалися на базі тих же сценаріїв, що й для інших розподілів.

Таблиця 1 – Результати для частки поясненої дисперсії  $R^2$

Розподіл фірм	Максимум	Середнє	Мінімум	Розмах	t-статистика
1	48,42	16,94	-206,58	255,00	-
10_10_10	46,97	38,53	15,12	31,85	15,47
100_100_100	44,24	40,53	35,14	9,10	-50,96
1000_10_10	41,46	39,47	35,76	5,70	16,18
1000_1000_1000	41,77	40,76	39,28	2,49	-129,76
10000_1000_100	40,37	39,64	38,56	1,82	6,33
10000_1000_100_10	40,56	39,63	38,44	2,12	6,54
5000_5000_5000	41,76	41,01	39,95	1,82	-165,61
200000	40,70	40,36	39,78	0,93	-54,92

У більшості розглянутих розподілів фірм середнє значення  $R^2$  перевищувало базовий рівень (39,87%). Незважаючи на те, що покращення в середньому складо близько 1%, значення t-статистики свідчать на користь статистичної значимості цієї різниці, а максимальне досягнуте значення перевищує базовий результат більше ніж на 7%. Зауважимо, що результати обчислювального експерименту для однієї фірми характеризуються найбільшим розмахом і, відповідно, найбільшою чутливістю до вхідних параметрів моделі та їхніх збурень.

Результати обчислювального експерименту з точки зору ефективності використання тих чи інших сценарних параметрів подано на рис. 2, де наведено розподіл  $R^2$  за різними видами регресії. Тут і далі для кращої візуалізації існуючих особливостей з графіка вилучено результати обчислювального експерименту для розподілу фірм 10\_10\_10.

Як бачимо, у розробленій простій моделі загальна регресія (за фактичним набором даних) дає кращі результати, ніж усереднена, а лінійний підхід в середньому кращий за байєсівський. Проте причиною такого результату щодо байєсівської регресії могла бути відсутність донавчання підігнаної моделі на нових результатах для кожного агента-фірми.

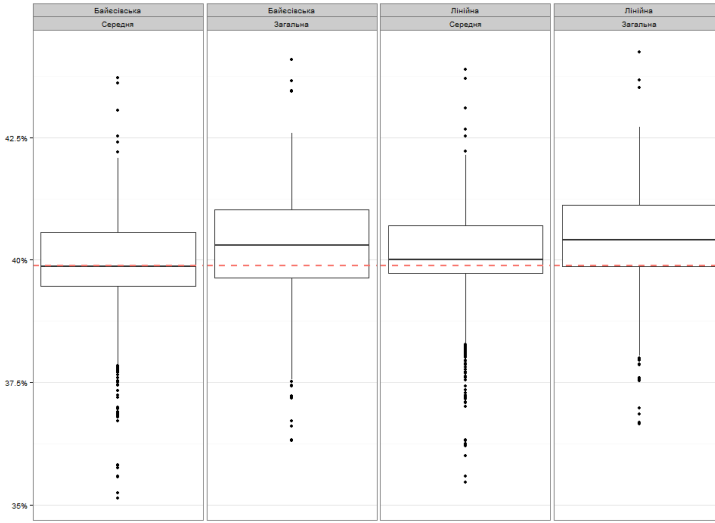


Рисунок 2 – «Ящик з вусами» для  $R^2$  за видами регресії

Особливу цікавість мають результати обчислювального експерименту для різних збурень (рис. 3). Зауважимо, що хоча оптимальні середнє та розмах і досягаються у випадку відсутності збурень, проте друге за величиною середнє спостерігається тоді, коли і результат, і коефіцієнти моделі збурюються. Але й величина розмаху у такому випадку є максимальною. Рис. 3 також ілюструє, що збурення коефіцієнтів без збурення результатів призводить до зниження якості моделі, а збурення результатів за незбурених коефіцієнтів проявляє тенденцію до її покращення.

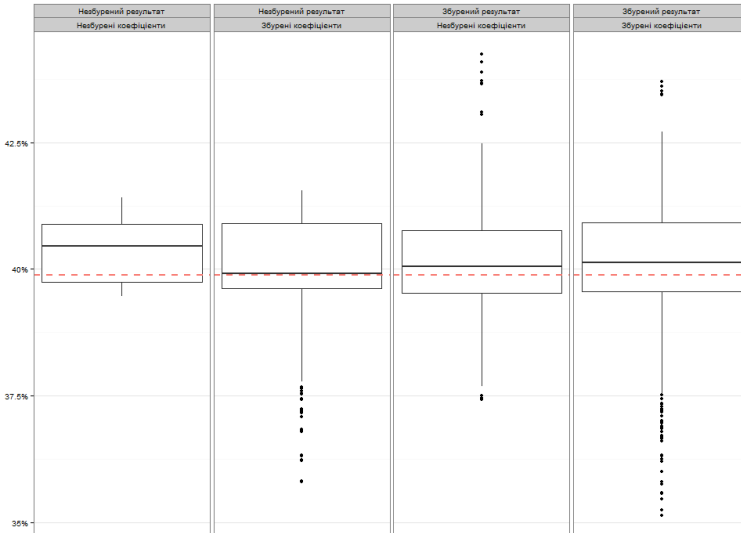


Рисунок 3 – Залежність  $R^2$  від різних видів збурень

Важливо також порівняти розглянуті два варіанти розподілу субсидій: випадковий і рівномірний. Із рис. 4 видно, що в рамках розробленої моделі

суттєвої різниці між двома підходами не виявлено. Можливою причиною цього є те, що для моделювання випадкового розподілу субсидій було обрано нормальний розподіл із значенням середнього, рівним розміру субсидій у рівномірному розподілі. Відповідно, різниця між підсумковими значеннями розподілів виявилася незначною.

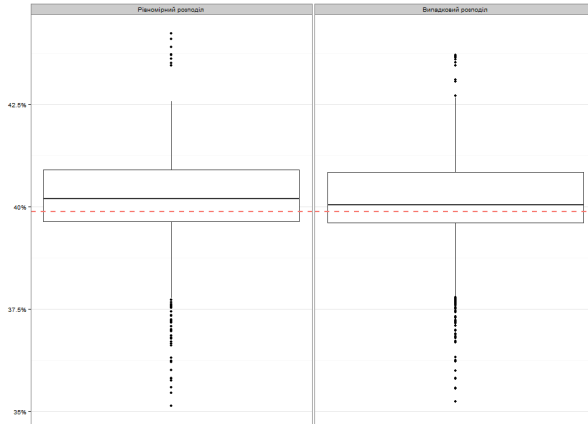


Рисунок 4 – Вплив способу розподілу субсидій на  $R^2$

Таким чином, результати обчислювального експерименту показали, що застосування агентно-орієнтованого підходу здатне покращити результати, отримані за допомогою відомих методів, і є більш стійким до збурень у даних. Отримані значення оцінок  $R^2$  та MAPE дають підставу для висновку про задовільне проходження моделлю процесу валідації на фактичних даних. Створення детальніших агентно-орієнтованих моделей на основі виявлення інших емпіричних закономірностей між існуючими статистичними показниками дасть змогу ще більше покращити отримані результати.

У подальшому планується розширити розроблену модель за допомогою врахування регіональних економічних особливостей. Додаткового дослідження потребує також питання щодо прогнозних можливостей моделі. Його пропонується вирішити шляхом моделювання індивідуальної поведінки фірм на базі регресії за неповним набором даних, або ж застосовуючи методи аналізу часових рядів, наприклад, ARIMA.

## Висновки

Агентно-орієнтований підхід є відносно новою парадигмою моделювання, яка виникла на «стику» комп'ютерних наук та економіки. Маючи широкі конструктивні можливості та надзвичайну гнучкість у застосуванні, він має і слабкі місця – складність валідації на емпіричних даних та спірну придатність до кількісного моделювання.

На основі реальних статистичних даних розроблено агентно-орієнтовану модель обсягів промислового виробництва у Польщі. Параметрами функції прийняття рішень індивідуальним агентом моделі вибрано розмір отриманих грошових субсидій та чисельність працівників. Для оцінки якості побудованої моделі використано статистики  $R^2$  та MAPE.



У рамках проведеного обчислювального експерименту розглянуто різні параметри (сценарії) роботи моделі: кількість і розподіл фірм за розміром, лінійна/байєсівська регресія, абсолютні/нормовані дані, випадковий чи рівномірний розподіл субсидій, збурення чи відсутність збурення коефіцієнтів та результатів тощо.

Результати експерименту показали, що застосування агентно-орієнтованого підходу має потенціал до покращення результатів класичних методів моделювання і є менш чутливим до збурень вхідної інформації. У подальшому планується дослідити можливості розробки агентно-орієнтованих версій відомих методів прогнозування.

## СПИСОК ЛІТЕРАТУРИ

1. Aschenwald J., Fink S., Tappeiner G. Brave new modeling: Cellular automata and artificial neural networks for mastering complexity in economics // Complexity. – 2001. – N 7. – P. 39–47.
2. Boero R., Squazzonu F. Does empirical embeddedness matter? Methodological issues on agent-based models for analytical social science // J. of Artificial Societies and Social Simulation. – 2005. – N 8(4). – 31 p. URL <http://jasss.soc.surrey.ac.uk/8/4/6.html>
3. Moss S. Alternative Approaches to the Empirical Validation of Agent-Based Models // J. of Artificial Societies and Social Simulation. – 2008. – N 11(1). – 16 p. URL <http://jasss.soc.surrey.ac.uk/11/1/5.html>
4. Hassan S., Arroyo J., Galán J.M., Antunes L., Pavón J. Asking the Oracle: Introducing Forecasting Principles into Agent-Based Modelling // J. of Artificial Societies and Social Simulation. – 2013. – N 16(3). – 17 p. URL <http://jasss.soc.surrey.ac.uk/16/3/13.html>.
5. Windrum P., Fagiolo G., Moneta A. Empirical validation of agent-based models: Alternatives and prospects // J. of Artificial Societies and Social Simulation. – 2007. – N 10(2). – 19 p. URL <http://jasss.soc.surrey.ac.uk/10/2/8.html>
6. Гуляницький Л.Ф., Омелянчик Д.А. Обучение агентов-фирм с помощью вывода иерархической базы правил // Компьютерная математика. – 2014. – № 2. – С. 14–21.
7. Омелянчик Д.А. Агентно-орієнтовані моделі обчислювальної економіки: особливості, переваги і недоліки // Математичне моделювання в економіці – № 1, 2015. – С. 41–53.
8. Центральне управління статистики Польщі [Central Statistical Office of Poland]. URL <http://stat.gov.pl/en/>.

*Стаття надійшла до редакції 11.11.2016*