

М'ЯКА КЛАСТЕРИЗАЦІЯ КРАЇН СВІТУ ЗА ПОКАЗНИКАМИ РОЗВИТКУ ЗЕРНОВИРОБНИЦТВА НА ОСНОВІ МОДЕЛЕЙ ГАУСОВИХ СУМІШЕЙ

У забезпеченні стабільності світового ринку зерна, який є одним з ключових елементів системи глобальної продовольчої безпеки, значну роль відіграє зерновий ринок України. За обсягами експорту Україна входить до п'ятірки найбільших експортерів зернових у світі. Понад 400 мільйонів людей на планеті залежать від поставок зерна з України. Україна бере участь у програмах ООН, спрямованих на боротьбу з голодом, наприклад, через Всесвітню продовольчу програму (WFP). Навіть після початку війни країна продовжує відігравати важливу роль у забезпеченні глобальної продовольчої безпеки. Все це обумовлює актуальність завдання щодо моделювання розвитку зернової галузі України в системі глобальної продовольчої безпеки.

Враховуючи наслідки війни, значні порушення економічних процесів, суттєві втрати у різних секторах, включаючи сільське господарство, моделювання на основі історичних даних України може бути некоректним. Для розв'язання цієї проблеми уявляється доцільним проведення дослідження на підґрунті даних по країнах світу, схожих з Україною за показниками розвитку зернового ринку. У попередньому дослідженні [1] на основі методу k-середніх (k-means) було здійснено групування країн світу за аграрно-економічними показниками розвитку зернового виробництва, що дозволило виокремити групу країн зі схожими з Україною умовами функціонування зернового сектору та створило підґрунтя для подальшого економіко-математичного моделювання розвитку галузі. Використання жорсткої кластеризації передбачає, що кожний об'єкт належить лише до одного кластера. Для складних соціально-економічних систем таке припущення не завжди є виправданим, оскільки реальна структура даних може характеризуватися частковим перекриттям кластерів, внаслідок чого для окремих країн однозначне віднесення до певної групи є недостатньо обґрунтованим. У зв'язку з цим доцільним є застосування методів м'якої кластеризації, які дозволяють враховувати невизначеність кластерної структури. У нечіткій кластеризації така невизначеність описується через ступені належності об'єкта до кожного кластера, у ймовірнісній кластеризації - через ймовірності потрапляння до латентних статистичних груп. Використання м'якої кластеризації дає змогу не лише визначити найбільш характерний для країни тип аграрного розвитку, але й оцінити ступінь впевненості такого віднесення та виявити країни, що знаходяться на межі між кластерами.

Серед методів м'якої кластеризації найбільш поширеними є нечіткий алгоритм c-середніх (Fuzzy C-Means) та модель гаусових сумішей (Gaussian Mixture Model, GMM). У першому випадку для кожного об'єкта визначаються ступені належності до всіх кластерів, які інтерпретуються як міра схожості. У другому випадку належність описується ймовірностями потрапляння до прихованих статистичних груп, що формують структуру даних [2, 3]. Для дослідження країн світу за аграрно-економічними показниками застосування моделі GMM уявляється особливо доцільним, оскільки дозволяє інтерпретувати кластери як латентні типи розвитку зерновиробництва та враховує різну форму, розмір і орієнтацію кластерів у багатовимірному просторі ознак.

Метою даного дослідження є аналіз кластерної структури країн світу за показниками розвитку зернового виробництва на основі моделі гаусових сумішей та порівняння отриманих результатів з результатами жорсткої кластеризації методом k-середніх і нечіткої кластеризації Fuzzy C-Means.

Дослідження проводилось за даними Світового банку [4]. Для кластерного аналізу були використані кількісні індикатори по 169 країнах світу за 2021 рік, що комплексно відображають аграрно-економічний профіль країн та їхню екологічну, демографічну й просторову специфіку: x_1 – урожайність зернових культур; x_2 – валовий внутрішній продукт на душу населення в доларах США; x_3 – частка орних земель у загальній площі держави; x_4 – обсяг внесення мінеральних добрив на гектар орних земель; x_5 – кількість орних земель на одну особу; x_6 – частка лісів у площі суші; x_7 – площа посівів зернових культур на одну особу; x_8 – середньорічна кількість атмосферних опадів; x_9 – частка сільського населення у загальній чисельності населення. Встановлено, що досліджувану сукупність країн доцільно розбивати на 6 груп. Опис отриманих кластерів країн та їх економічна інтерпретація представлені у праці [1].

Для оцінювання результатів м'якої кластеризації було побудовано три моделі GMM, які відрізняються типом коваріаційної матриці: повною (модель GMM-full), діагональною (модель GMM-diagonal) та сферичною (модель GMM-spherical) [5]. В якості початкових значень використовувалися результати жорсткої кластеризації за методом k-середніх. У табл. 1 наведено показники якості кластеризації: коефіцієнт силуету (Silhouette); індекс Калінського-Гарабаша (CH); індекс Данна (Dunn Index), а також інформаційні критерії Акаїке (AIC) та байєсівський (BIC). Найкращі значення AIC та BIC отримано для моделі GMM-full. Це свідчить про те, що саме дана модель найбільш адекватно описує структуру досліджуваних даних. Водночас за внутрішніми показниками якості Silhouette, CH, Dunn Index вона поступається як методу k-means, так і моделі GMM-spherical, що пояснюється орієнтацією моделі GMM-full насамперед на моделювання щільності багатовимірного розподілу даних, а не на максимізацію геометричної відокремленості кластерів. Незважаючи на дещо нижчі значення внутрішніх індексів кластеризації, перевага моделі GMM-full полягає у відсутності жорстких обмежень на форму кластерів та можливості врахування кореляцій між ознаками, що особливо важливо для багатовимірних соціально-економічних даних. Порівняно невисокі значення Silhouette для всіх розглянутих моделей свідчать, що країни світу не формують чітко відокремлених компактних груп, досліджувана сукупність має скоріше континуальний характер, а межі між окремими типами аграрного розвитку є розмитими. За таких умов використання методів м'якої кластеризації видається більш обґрунтованим, оскільки дозволяє враховувати прикордонні випадки та невизначеність кластерної належності.

Таблиця 1

Порівняння моделей кластеризації за показниками якості розбиття та інформаційними критеріями

Модель кластеризації	Критерії якості кластеризації			Інформаційні критерії	
	Silhouette	CH	Dunn Index	AIC	BIC
k-means	0,221	40,2	0,069	-	-
GMM-full	0,140	21,8	0,064	2703	3733
GMM-diagonal	0,065	19,4	0,073	3045	3399
GMM-spherical	0,216	39,3	0,068	3581	3785

Джерело: розроблено авторами

Порівняння отриманих кластерів з результатами методу k-means (табл. 2) показує, що модель GMM-spherical практично повністю відтворює структуру, сформовану за методом k-means (частка однакових об'єктів для більшості кластерів перевищує 90%). Це закономірно через геометричні обмеження, оскільки сферична коваріаційна структура накладає на кластери обмеження, близькі до припущень методу k-means. Згідно з AIC та BIC найбільш адекватною є модель GMM-full, яка дозволяє виявити додаткові закономірності даних.

Таблиця 2

**Відповідність кластерів, отриманих за моделями GMM,
результатам кластеризації за методом k-means**

Модель кластеризації	Показник	Кластери (за методом k-means)					
		0	1	2	3	4	5
		кількість об'єктів в кластері					
		47	35	4	30	35	18
GMM-full	відповідний кластер за GMM	1	1	2	4	0	5
	частка однакових об'єктів, %	74,5	82,9	50,0	76,7	91,4	88,9
GMM-diagonal	відповідний кластер за GMM	0	1	2	3	4	5
	частка однакових об'єктів, %	48,9	80,0	75,0	43,3	40,0	100
GMM-spherical	відповідний кластер за GMM	0	1	2	3	4	5
	частка однакових об'єктів, %	95,7	100	100	83,3	100	94,4

Джерело: розроблено авторами

Найбільший інтерес становить модель GMM-full. На відміну від інших вона фактично об'єднує два кластери, виділені методом k-means: значна частка країн з кластерів 0 і 1 методу k-means формує спільний кластер GMM-full, що свідчить про високу подібність цих груп країн та відсутність між ними чіткої статистичної межі. Це вказує на те, що ймовірна модель дозволяє виявити латентні закономірності, які не повністю відображаються у жорсткому розбитті.

Додатковий аналіз ймовірностей потрапляння до кластерів (табл. 3) показав високий рівень впевненості класифікації для всіх розглянутих моделей. Середні значення максимальних ймовірностей для моделей GMM-full, GMM-diagonal та GMM-spherical знаходяться в межах 0,979–1, 0,906–0,989 та 0,906–1 відповідно. Отже, більшість країн мають чітко виражену належність до відповідних кластерів, і лише незначну їх частину можна розглядати як прикордонні випадки. При цьому для моделі GMM-full високі значення апостеріорних ймовірностей поєднуються зі зміною структури кластерів порівняно з результатами k-means, що свідчить про статистичну обґрунтованість виявленого об'єднання окремих груп країн.

Таблиця 3

**Характеристики ймовірностей потрапляння об'єктів
до кластерів для різних моделей кластеризації**

Модель кластеризації	Показник	Номер кластера					
		0	1	2	3	4	5
GMM-full	кількість об'єктів у кластері	46	66	2	-	32	23
	середня ймовірність	0,983	0,979	1	-	0,979	0,988
GMM-diagonal	кількість об'єктів у кластері	41	45	3	17	30	33
	середня ймовірність	0,910	0,921	0,989	0,972	0,928	0,962
GMM-spherical	кількість об'єктів у кластері	47	37	4	26	38	17
	середня ймовірність	0,922	0,906	1	0,950	0,937	0,961

Джерело: розроблено авторами

Для перевірки стійкості результатів додатково застосовано алгоритм Fuzzy C-Means. Отримані результати підтвердили наявність розмитих меж між окремими групами країн та тенденцію до часткового перекриття кластерів. Тобто, результати нечіткої та ймовірнісної кластеризації є узгодженими й свідчать про відсутність чітко відокремлених типів розвитку зерновиробництва.

Таким чином, використання моделі GMM дозволило не лише оцінити ступінь невизначеності кластерної належності країн, а й показало, що окремі кластери методу k-means не мають чітких статистичних меж та можуть бути інтерпретовані як підгрупи ширших латентних типів розвитку зерновиробництва. Разом з тим порівняно невисокі значення Silhouette свідчать про складність кластерної структури досліджуваної сукупності та потребують подальшого аналізу альтернативних наборів ознак і моделей кластеризації.

Список використаних джерел

1. Піскунова, О. В., Савіна, С. С., & Дрончак, О. В. (2025). Застосування кластерного аналізу для визначення груп країн, схожих за умовами розвитку зерновиробництва. У *Матеріали III Міжнародної науково-практичної конференції «Цифрова економіка»*. Київський національний економічний університет імені Вадима Гетьмана. <https://doi.org/10.33111/978-966-926-559-3>
2. Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, 10(2–3), 191–203. [https://doi.org/10.1016/0098-3004\(84\)90020-7](https://doi.org/10.1016/0098-3004(84)90020-7)
3. Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer. <https://link.springer.com/book/9781493938438>
4. World Bank. (2025). *World Bank Open Data*. Retrieved June 2025, from <https://data.worldbank.org/>
5. McLachlan, G., & Peel, D. (2000). *Finite mixture models*. Wiley. <https://doi.org/10.1002/0471721182>

Піскунова О. В.,
д.е.н., професор,

Ернст Д. В.,
здобувач другого (магістерського) рівня вищої освіти,
Київський національний економічний університет імені Вадима Гетьмана

МОДЕЛЮВАННЯ СПЕКУЛЯТИВНОЇ КРИХКОСТІ ФІНАНСОВИХ АКТИВІВ З ВИКОРИСТАННЯМ GJR-GARCH ТА BNS-ПІДХОДУ

Актуальність: нестабільність фінансових активів набуває особливого значення в контексті безпрецедентного ШІ-буму, який докорінно змінив архітектуру ринкових очікувань. На відміну від класичних спекулятивних бульбашок, що описуються моделлю Кіндлбергера-Мінські[7] як наслідок монетарної експансії, сучасний технологічний перегрів живиться концентрацією капіталу у вузькому колі інфраструктурних лідерів за умов жорсткої монетарної політики – цей феномен не має загальноприйнятого формалізованого трактування. Зростання ціни акцій NVIDIA з \$3.34 (січень 2019 р.) до \$137.44 (грудень 2024 р.) на 3970% (split-adjusted, розраховано за даними Yahoo Finance [10]) супроводжувалося якісною зміною структури ризику: емпіричний аналіз свідчить про домінування дискретних структурних стрибків над дифузійною компонентою волатильності – явище, яке стандартні моделі волатильності систематично недооцінюють. Це актуалізує потребу у розробці та застосуванні спеціалізованого інструментарію – зокрема інтеграції моделей GJR-GARCH та BNS-декомпозиції, які здатні кількісно розмежувати фундаментальний технологічний драйвер та самопідсилювальну ендогенну волатильність, що і становить предмет цього дослідження.