

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА**

**Інститут інформаційних технологій в економіці
Кафедра штучного інтелекту, моделювання та статистики**

Освітньо-професійна програма

Економічна кібернетика

Галузь знань

05 Соціальні та поведінкові науки

Спеціальність

051 Економіка

Форма навчання: дистанційна

**КВАЛІФІКАЦІЙНА БАКАЛАВРСЬКА
РОБОТА**

**на тему: «Прогнозування цін на сільськогосподарську продукцію (на
прикладі цін на пшеницю)»**

(назва теми)

здобувача Турченко Богдани Василівни

(ПІБ, підпис)

Науковий керівник: к. е. н. доцент Галіцина Ольга Володимирівна

(науковий ступінь, учене звання, ПІБ)

(підпис)

**Робота допущена до захисту перед екзаменаційною
комісією з атестації здобувачів вищої освіти (ЕК)**

**Завідувач кафедри: к. ф.-м. н., професор
Галина ВЕЛИКОІВАНЕНКО**

(підпис)

Київ 2026

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА

Інститут інформаційних технологій в економіці
Кафедра штучного інтелекту, моделювання та статистики

Освітньо-професійна програма
Галузь знань
Спеціальність

Економічна кібернетика
05 Соціальні та поведінкові науки
051 Економіка

ПОГОДЖЕНО

Керівник проектної групи (гарант)
освітньо-професійної програми
О.М. Притоманова

(підпис)

20__ р.

ЗАТВЕРДЖУЮ

Завідувач кафедри математичного
моделювання та статистики
Г.І. Великоіваненко

(підпис)

20__ р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

здобувача вищої освіти Турченко Богдани Василівни
(прізвище, ім'я, по батькові)

дистанційної форми навчання
очної (денної), дистанційної

на підготовку кваліфікаційної бакалаврської роботи

на тему

«Прогнозування цін на сільськогосподарську продукцію (на прикладі цін на пшеницю)»

Тему затверджено наказом ректора Університету від «06» квітня 2026 р .№ 477 ст.

Кваліфікаційна бакалаврська робота виконується на матеріалах
з сайту <https://finance.yahoo.com>

План кваліфікаційної бакалаврської роботи

Розділ 1	Теоретичні та методичні засади прогнозування цін на аграрну продукцію
(назва розділу)	
Розділ 2	Аналіз динаміки цін на пшеницю за допомогою часових рядів
(назва розділу)	
Розділ 3	Порівняльний аналіз ефективності економетричних та нейромережевих моделей прогнозування цін на пшеницю
(назва розділу)	
Об'єкт дослідження:	Ціни на ринку аграрної продукції.
Предмет дослідження:	Методи та інструменти прогнозування цін на сільськогосподарську продукцію (або пшеницю)
Мета кваліфікаційної бакалаврської роботи:	Визначення найбільш ефективного методу прогнозування цін на аграрну продукцію.

Конкретні завдання, які здобувач повинен виконати для досягнення поставленої мети:

У розділі 1: 1) дослідити еволюцію методів прогнозування на товарних ринках від класичних авторегресійних моделей до сучасних нейромережевих архітектур; 2) розглянути методологію оцінювання прогнозованої якості та статистичного порівняння моделей; 3) визначити економічні чинники та макроекономічні детермінанти формування цін на зернові культури.

У розділі 2: 1) проаналізувати динаміку та структуру часових рядів цін на пшеницю; 2) провести діагностику властивостей часових рядів та перевірку на стаціонарність; 3) виявити структурні зміни та здійснити попередню підготовку вхідних параметрів для моделювання.

У розділі 3: 1) розробити та провести діагностику економетричних моделей ARIMA та VAR; 2) здійснити навчання та оптимізацію нейромережевих архітектур MLP та LSTM; 3) провести компаративний аналіз точності прогнозів отриманих моделей та сформулювати практичні рекомендації щодо їх застосування.

Завдання підготував
науковий керівник

(підпис)

Ольга ГАЛІЦИНА

(ім'я, прізвище)

«__» _____ 2026 р.

Завдання одержав
здобувач

(підпис)

Богдана ТУРЧЕНКО

(ім'я, прізвище)

«__» _____ 2026 р.

Реферат

Кваліфікаційна бакалаврська робота складається зі вступу, трьох розділів, висновків та переліку використаних джерел, що містить 34 найменування на 5 сторінках та 2 додатки на 50 сторінках. Загальний обсяг роботи 120 сторінок друкованого тексту, з яких 65 сторінок основної частини. Робота містить 14 таблиць, 29 рисунків, 15 формул.

«Прогнозування цін на сільськогосподарську продукцію (на прикладі цін на пшеницю)»

Об'єктом дослідження є ціни на ринку аграрної продукції.

Предметом дослідження є методи та інструменти прогнозування цін на сільськогосподарську продукцію.

Метою дослідження є визначення найбільш ефективного методу прогнозування цін на аграрну продукцію

Відповідно до поставленої мети визначені такі завдання:

1. Створити теоретичну основу для проведення порівняльного аналізу моделей прогнозування товарних ринків.
2. Провести статистичний аналіз часових рядів цін на пшеницю та макроекономічних показників (Brent, DXY, VIX, S&P 500) за період з 2010 по 2026 рік.
3. Перевірити стаціонарність та структурні розриви за допомогою тестів ADF, KPSS та Чоу.
4. Розробити та оцінити моделі ARIMA(p,1,q) та VAR(p) з фіктивними змінними для структурних шоків.
5. Навчити дві різні нейронні мережі, MLP та LSTM, та провести аналіз чутливості гіперпараметрів.
6. Порівняти моделі на основі показників MAE, RMSE, MAPE у тестовій вибірці з використанням тесту Діболда–Маріано.
7. Дослідити процес формування та динаміку цін на пшеницю на міжнародному ф'ючерсному ринку під впливом макроекономічних та геополітичних факторів.

Теоретична, методична та практична значущість отриманих результатів. У роботі систематизовано підходи до прогнозування цін на товарних ринках, досліджено еволюцію методів від класичних авторегресійних моделей до нейромережових архітектур. Застосовано комплекс економетричних методів для діагностики часових рядів і виявлення структурних розривів, зокрема пов'язаних із пандемією COVID-19 та повномасштабним вторгненням росії в Україну. Побудовано та оцінено чотири класи моделей в єдиній методологічній системі; запропоновано відтворюваний алгоритм оцінювання (послідовний прогноз + тест Діболда–Маріано + динаміка MAPE), який може бути поширений на інші товарні ринки.

Рік виконання кваліфікаційної бакалаврської роботи – 2026.

Рік захисту роботи – 2026.

Ключові слова: прогнозування цін, пшениця, часові ряди, ARIMA, VAR, MLP, LSTM, структурні злами, тест Діболда–Маріано, нейронні мережі, аграрний ринок.

В і д г у к
про кваліфікаційну бакалаврську роботу
здобувача освітньо-професійної програми «Економічна кібернетика»
«Інститут інформаційних технологій в економіці»

Турченко Богдани Василівни
(ПІБ)

на тему: «Прогнозування цін на сільськогосподарську продукцію (на прикладі
цін на пшеницю)»
(назва теми)

1. Актуальність теми: Кваліфікаційна бакалаврська робота є самостійним дослідженням, яке присвячено прогнозуванню цін на пшеницю з урахуванням сучасних геополітичних та макроекономічних шоків, що має особливе значення для агропромислового сектору України.
2. Позитивні риси кваліфікаційної бакалаврської роботи: У роботі продемонстровано доцільність застосування сучасних методів прогнозування цін на пшеницю (ARIMA, VAR, MLP та LSTM) з урахуванням структурних зламів для підвищення якості прогнозів та прийняття ефективних управлінських рішень в агропромисловому секторі.
3. Наявність самостійних розробок автора: Результати проведеного дослідження дозволили порівняти економетричні (ARIMA, VAR) та нейромережеві (MLP, LSTM) моделі прогнозування з урахуванням структурних зламів і сформулювати практичні рекомендації щодо їх застосування залежно від розмірів горизонту прогнозування.
4. Цінність теоретичних висновків та практичних рекомендацій: Результати дослідження обумовлюють практичну значущість роботи, що визначається у доцільності використання запропонованого інструментарію для прогнозування цін на сільськогосподарську продукцію.
5. Наявність недоліків: Кваліфікаційна бакалаврська робота містить незначну кількість недоліків стилістичного характеру, які не впливають на загальну якість виконаної роботи.
6. Загальна оцінка кваліфікаційної бакалаврської роботи та її допущення до захисту перед ЕК:
Тема роботи розкрита, основні положення та висновки достатньо обґрунтовані. Рівень виконання кваліфікаційної бакалаврської роботи високий. Робота Турченко Богдани Василівни на тему «Прогнозування цін на сільськогосподарську продукцію (на прикладі цін на пшеницю)» може бути допущена до захисту перед ЕК і заслуговує високої позитивної оцінки.

Науковий керівник доцент кафедри штучного інтелекту, моделювання та статистики к.е.н. доцент

(посада, учене звання, науковий ступінь)

(підпис)

Галіцина О.В.

(прізвище, ініціали)

“ ” 20 26 р.

ЗМІСТ

ВСТУП	5
РОЗДІЛ 1 ТЕОРЕТИЧНІ ТА МЕТОДИЧНІ ЗАСАДИ ПРОГНОЗУВАННЯ ЦІН НА АГРАРНУ ПРОДУКЦІЮ	8
1.1 Еволюція методів прогнозування на товарних ринках: від класичних авторегресійних моделей до нейромережевих архітектур	8
1.2 Методологія оцінювання прогнозової якості та статистичного порівняння моделей	10
1.3 Економічні чинники та макроекономічні детермінанти формування цін на зернові культури	12
РОЗДІЛ 2 АНАЛІЗ ДИНАМІКИ ЦІН ЗА ДОПОМОГОЮ ЧАСОВИХ РЯДІВ	16
2.1 Динаміка та структура часових рядів цін на пшеницю	16
2.2 Діагностика властивостей часових рядів та перевірка на стаціонарність	21
2.3 Обробка структурних зламів та попередня підготовка вхідних параметрів	27
РОЗДІЛ 3 ПОРІВНЯЛЬНИЙ АНАЛІЗ ЕКОНОМЕТРИЧНИХ ТА НЕЙРОМЕРЕЖЕВИХ МОДЕЛЕЙ	34
3.1 Розробка та діагностика економетричних моделей ARIMA та VAR	34
3.2 Навчання та оптимізація нейромережевих архітектур MLP та LSTM	46
3.3 Порівняльний аналіз точності прогнозів	54
ВИСНОВКИ	64
ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ	66
ДОДАТКИ	71

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

ACF	Autocorrelation Function — функція автокореляції
ADF	Augmented Dickey–Fuller — розширений тест Дікі–Фуллера (тест на одиничний корінь)
AIC	Akaike Information Criterion — інформаційний критерій Акайке
AR	Autoregressive — авторегресійна (компонента моделі)
ARDL	Autoregressive Distributed Lag — авторегресійна розподілена модель лагів
ARIMA	Autoregressive Integrated Moving Average — авторегресійна інтегрована модель ковзного середнього
ARIMAX	ARIMA with eXogenous variables — ARIMA із зовнішніми (екзогенними) змінними
BIC	Bayesian Information Criterion — байєсівський інформаційний критерій
BPTT	Backpropagation Through Time — зворотне поширення помилки у часі
Brent	ICE Brent Crude Oil — ф'ючерс на нафту марки Brent
BRICS	Brazil, Russia, India, China, South Africa — міжнародне об'єднання країн (Бразилія, росія, Індія, Китай, ПАР)
CBOT	Chicago Board of Trade — Чиказька товарна біржа
CBOE	Chicago Board Options Exchange — Чиказька біржа опціонів
CRB	Commodity Research Bureau Index — товарний індекс Бюро досліджень сировинних ринків
CV	Coefficient of Variation — коефіцієнт варіації
DM	Diebold–Mariano — тест Діболда–Маріано для порівняння точності прогнозів
DXY	US Dollar Index — індекс долара США
ECM	Error Correction Model — модель корекції помилок
ENSO	El Niño–Southern Oscillation — Ель-Ніньо / Південне коливання (кліматичний індекс)
ETF	Exchange-Traded Fund — біржовий інвестиційний фонд
FAO	Food and Agriculture Organization — Продовольча та сільськогосподарська організація ООН
FEVD	Forecast Error Variance Decomposition — декомпозиція дисперсії похибки прогнозу
HAC	Heteroskedasticity and Autocorrelation Consistent — оцінка, стійка до гетероскедастичності та автокореляції
HQIC	Hannan–Quinn Information Criterion — інформаційний критерій Ханнана–Квінна
I(1)	Integrated of order 1 — інтегрований процес першого порядку

IRF	Impulse Response Function — імпульсно-реакційна функція
KDE	Kernel Density Estimation — ядрова оцінка щільності розподілу
KPSS	Kwiatkowski–Phillips–Schmidt–Shin — тест КПСС (тест на стаціонарність)
LB	Ljung–Box — тест Льюнга–Бокса (тест на автокореляцію залишків)
LSTM	Long Short-Term Memory — довгострокова короткочасна пам'ять (тип рекурентної нейронної мережі)
MA	Moving Average — ковзне середнє
MAE	Mean Absolute Error — середня абсолютна похибка
MAPE	Mean Absolute Percentage Error — середня абсолютна відсоткова похибка
MLE	Maximum Likelihood Estimation — метод максимальної правдоподібності
MLP	Multilayer Perceptron — багатошаровий перцептрон
MSE	Mean Squared Error — середньоквадратична похибка
OLS	Ordinary Least Squares — метод найменших квадратів
PACF	Partial Autocorrelation Function — часткова функція автокореляції
RMSE	Root Mean Squared Error — корінь із середньоквадратичної похибки
RNN	Recurrent Neural Network — рекурентна нейронна мережа
S&P 500	Standard & Poor's 500 — фондовий індекс 500 найбільших компаній США
S/U	Stocks-to-Use ratio — співвідношення запасів до споживання
USDA	United States Department of Agriculture — Міністерство сільського господарства США
VaR	Value at Risk — вартість під ризиком
VAR	Vector Autoregression — векторна авторегресія
VIX	CBOE Volatility Index — індекс волатильності CBOE ("індикатор паніки")
WASDE	World Agricultural Supply and Demand Estimates — Світові оцінки попиту та пропозиції на агропродукцію (звіт USDA)
WN	White Noise — білий шум
ФАО	Продовольча та сільськогосподарська організація ООН (FAO)

ВСТУП

Продовольча безпека має вирішальне значення в епоху кліматичних та геополітичних потрясінь. Повномасштабне вторгнення росії в Україну в лютому 2022 року спричинило різке зростання цін на ф'ючерси на пшеницю на ринку CBOT протягом семи тижнів; це продемонструвало вразливість глобальних ланцюгів постачання та обмеження традиційних методів прогнозування. Для України як одного з найбільших світових виробників та експортерів пшениці ефективні інструменти прогнозування цін є як науковою, так і практичною необхідністю. Незважаючи на великий масив літератури — від авторегресійних моделей до нейронних архітектур, — більшість із них не враховують структурні розриви 2020 та 2022 років і не містять уніфікованого порівняльного аналізу.

Актуальність теми зумовлена необхідністю вдосконалення існуючих підходів до прогнозування цін на пшеницю з урахуванням сучасних геополітичних та макроекономічних шоків, що має особливе значення для агропромислового сектору України.

Аналіз останніх досліджень та публікацій. Питання прогнозування цін на сировинні товари досліджується у працях вітчизняних та зарубіжних вчених. Класичні підходи на основі моделі ARIMA викладені в дослідженні Спіліотіса та інш.; застосування архітектур нейронних мереж до фінансових часових рядів досліджується Фішером та Краусом. Серед вітчизняних науковців методи аналізу та прогнозування часових рядів розглядаються у роботах Бідюка П.І.; динаміку та цінову волатильність ринку зерна України в умовах геополітичних шоків досліджено у праці Боднара О.В.. Однак більшість існуючих результатів розглядають окремі класи моделей без їхнього систематичного порівняння за наявності структурних зламів, що визначає відкриту частину проблеми.

Метою роботи є визначення найбільш ефективного методу прогнозування цін на аграрну продукцію, з урахуванням впливу макроекономічних показників та структурних зламів, а також визначення найбільш ефективного інструменту для практичного застосування.

Завдання дослідження:

1. Визначити теоретичну основу для проведення порівняльного аналізу моделей прогнозування товарних ринків.
2. Провести статистичний аналіз часових рядів цін на пшеницю та макроекономічних показників (Brent, DXY, VIX, S&P 500) за період з 2010 по 2026 рік.
3. Перевірити стаціонарність та структурні розриви за допомогою тестів ADF, KPSS та Чоу.
4. Розробити та оцінити моделі ARIMA(p,1,q) та VAR(p) з фіктивними змінними для структурних шоків.
5. Навчити дві різні нейронні мережі, MLP та LSTM, та провести аналіз чутливості гіперпараметрів.
6. Порівняти моделі на основі показників MAE, RMSE, MAPE у тестовій вибірці з використанням тесту Діболда–Маріано.

Об'єктом дослідження є ціни на ринку аграрної продукції.

Предметом дослідження є методи та інструменти прогнозування цін на сільськогосподарську продукцію.

Методи дослідження: тести стаціонарності ADF та KPSS; підхід Енгла–Грейнджера до коінтеграції; тест Чоу для виявлення структурних розривів; метод максимальної правдоподібності (MLE) для оцінювання параметрів ARIMA та VAR; зворотне поширення помилки з оптимізатором Adam при навчанні нейронних мереж; ковзне прогнозування; тест Діболда–Маріано для статистичного порівняння точності моделей.

Теоретична, методична та практична значущість отриманих результатів. Теоретичне значення полягає в систематичному порівнянні

чотирьох класів моделей у рамках однієї методологічної концепції з урахуванням гетерогенних структурних шоків; запропонований алгоритм оцінювання (ковзний прогноз + тест DM + динаміка MAPE) може бути застосований до інших товарних ринків. *Практичне значення:* моделі ARIMA рекомендуються для внутрішньоденних сигналів та оцінки VaR; VAR — для стратегічного міжринкового аналізу; ARIMA+LSTM — для оперативного виявлення змін у патернах. *Результати призначені* для використання сільськогосподарськими виробниками, фінансовими установами та регуляторами українського агроринку.

Інформаційна база: щоденні дані з Yahoo Finance (ZW=F, BZ=F, DX-Y.NYB, ^VIX, ^GSPC) за 2010–2024 роки; звіти USDA WASDE; дані ФАО; наукові статті в журналах Journal of Forecasting, Energy Economics, Journal of Agricultural Economics; дані конкурсу M4.

Робота складається з вступу, трьох розділів, висновку, списку використаних джерел та додатків. Вона містить 65 сторінок, 14 таблиць, 28 рисунків, перелік використаних джерел з 34 найменувань та 2 додатки.

РОЗДІЛ 1

ТЕОРЕТИЧНІ ТА МЕТОДИЧНІ ЗАСАДИ ПРОГНОЗУВАННЯ ЦІН НА АГРАРНУ ПРОДУКЦІЮ

1.1 Еволюція методів прогнозування на товарних ринках: від класичних авторегресійних моделей до нейромережових архітектур

Прогнозування цін на сільськогосподарську продукцію є складним завданням унаслідок одночасної дії геополітичних, кліматичних та фінансових чинників. Вибір між класичними та нейромережовими методами визначається горизонтом прогнозу, ступенем нелінійності ряду та обсягом даних.

ARIMA, як і раніше, залишається еталоном для одноденних прогнозів. На основі масштабного дослідження Макрідакіса, Спіліотіса та Ассімакопулоса [1] з понад 100 тисячами часових рядів можна сказати, що статистичні методи мають стабільно вищу ефективність, ніж методи ML, на горизонті 1–5 кроків у випадку часових рядів, які близькі до лінійної авторегресії після диференціювання. Однак суто лінійна природа ARIMA унеможливорює врахування структурних зламів та нелінійних залежностей — обмеження, що є важливим для ринку пшениці з його різкими коливаннями цін. Серед вітчизняних науковців методи аналізу та побудови моделей часових рядів, зокрема ARIMA та її модифікацій, систематизовано у статті Бідюка П.І. [2], що є методологічною базою для практичного застосування цих підходів в українській науковій традиції.

VAR необхідна для аналізу механізмів міжринкового перенесення. Як зазначають Чіфареллі та Паладіно [3], частка цін на нафту у варіативності цін на пшеницю зростає з ~5% у місячному горизонті до понад 20% у річному, що підкреслює корисність VAR для середньострокового структурного аналізу.

Імпульсно-реакційні функції та декомпозиція дисперсії (зокрема FEVD [4]) дозволяють кількісно аналізувати міжринкові взаємозв'язки, хоча для щоденного прогнозування їхня точність поступається простішим моделям.

Теоретично MLP може апроксимувати будь-яку нелінійну функцію. Однак Ахмед, Атія та Елгайяр [5], проаналізувавши 1 045 часових рядів, виявили, що MLP є менш ефективним методом, ніж ARIMA, для лінійних або майже авторегресійних рядів; цей висновок пояснюється ризиком перенавчання. Ван, Фенг та інші [6] відзначили зростання частки нейронних мереж у прогнозуванні в аграрній галузі з 12% (2000–2005 рр.) до 67% протягом досліджуваного періоду (2016–2020 рр.). Проте найбільшим обмеженням MLP є відсутність механізму пам'яті: вікно вхідних даних розглядається як статичний вектор, що не враховує динаміку часового ряду.

LSTM подолав це обмеження завдяки механізму стану комірок та шлюзам. Фішер та Краус [7] продемонстрували, що LSTM перевершує MLP та лінійні моделі у прогнозуванні S&P 500, зазначивши, що ця перевага є найбільш вираженою на горизонті 5–20 днів. Також підтверджена наступна закономірність: на 1-денному горизонті ARIMA є ефективнішою, тоді як на 5- та 20-денних горизонтах LSTM виявляється значно точнішою. Перспективним розширенням є LSTM із механізмом уваги, що забезпечує зниження RMSE на 12–18% на горизонті 5–10 днів [9]. Боднар О.В. [10] досліджував цінову волатильність ринку пшениці та вплив збройного конфлікту 2022 року на структуру цін, що підтверджує необхідність врахування структурних зламів у моделях прогнозування для вітчизняного аграрного ринку.

1.2 Методологія оцінювання прогнозової якості та статистичного порівняння моделей

Оцінка якості прогнозів є невід’ємною частиною порівняльного аналізу моделей: без формальних метрик та статистичних тестів висновки втрачають наукову обґрунтованість. У цьому розділі описано теоретичні основи показників MAE, RMSE та MAPE, а також процедуру тесту Діболда–Маріано для статистичного порівняння точності прогнозування моделей.

Оцінка якості прогнозування часових рядів зазвичай ґрунтується на трьох взаємодоповнювальних показниках, які враховують різні аспекти похибки. Їхній вибір для конкретного дослідження не є нейтральним, оскільки кожен показник по-різному реагує на розподіл похибок і має різне практичне значення [11].

Середня абсолютна похибка (MAE — Mean Absolute Error) визначається як:

$$MAE = (1/n) \cdot \sum_{t=1}^n |y_t - \hat{y}_t|, \quad (1.1)$$

де y_t — фактичне значення цільової змінної у момент t ;

$\hat{y}_t$ — прогнозоване значення;

n — кількість спостережень тестової вибірки.

MAE вимірює середню величину похибок у наборі прогнозів, не беручи до уваги їхній напрямок. Завдяки лінійному зважуванню похибок цей показник є стійким до викидів, що є особливо корисним для ринків, на яких спостерігаються випадкові екстремальні коливання. Гайндман і Келер [11] пропонують використовувати MAE як основний показник для порівняння моделей на певному рівні, наголошуючи на тому, що він є безпосередньо інтерпретованим.

Середньоквадратична похибка (RMSE — Root Mean Squared Error) обчислюється як:

$$RMSE = \sqrt{[(1/n) \cdot \sum_{t=1}^n (y_t - \hat{y}_t)^2]}, \quad (1.2)$$

де y_t — фактичне значення цільової змінної у момент t ;

$\hat{y}_t$ — прогнозоване значення;

n — кількість спостережень тестової вибірки.

RMSE використовує квадратичне покарання за похибку, а отже, приділяє більшу вагу більшим відхиленням: подвоєння похибки вчетверо збільшує її внесок. Як зазначають Дракслер і Таї [12], RMSE є природним вибором у випадках, коли ціна великих похибок є непропорційно високою (наприклад, при хеджуванні товарних ризиків). Практичним недоліком є те, що цей показник чутливий до окремих виняткових значень, які можуть суттєво спотворити оцінку загальної якості моделі.

Середня абсолютна відсоткова похибка (MAPE — Mean Absolute Percentage Error) є масштабно-незалежною метрикою:

$$MAPE = (100/n) \cdot \sum_{t=1}^n |y_t - \hat{y}_t| / |y_t|, \quad (1.3)$$

де y_t — фактичне значення цільової змінної у момент t ;

$\hat{y}_t$ — прогнозоване значення;

n — кількість спостережень тестової вибірки.

Показник MAPE виражає похибку у відсотках від фактичного значення, що дозволяє порівнювати результати на різних ринках та в різних часових горизонтах. Стандартна шкала інтерпретації [13]: <10% — висока точність, 10–20% — хороша, 20–50% — задовільна, >50% — незадовільна. Вплив метрики: чутливість до асиметрії похибок і невизначеність при $y_t = 0$.

Різниця у значеннях MAE або RMSE для двох моделей сама по собі не є свідченням того, що їхня прогностична здатність відрізняється, оскільки це може бути результатом вибіркової мінливості. Для формальної перевірки цього питання Діболд і Маріано [14] розробили стандарт оцінки порівняльної

прогнозної здатності (comparative predictability) — тест непараметричної рівності прогнозних здібностей. Нульова гіпотеза тесту H_0 стверджує, що дві порівнювані моделі мають однакову очікувану прогнозну точність, тобто математичне сподівання диференціалу функцій втрат дорівнює нулю:

$$H_0 : E[d_t] = 0, \quad d_t = L(e_t^1) - L(e_t^2), \quad (1.4)$$

де $e_t^j = y_t - \hat{y}_t^j$ — похибка прогнозу j -ї моделі у момент t ($j = 1, 2$);

$L(\cdot)$ — функція втрат;

$L(e_t) = e_t^2$ (квадратична втрата, що відповідає порівнянню за критерієм MSE/RMSE).

Тестова статистика тесту будується на вибірковому середньому диференціалу втрат. За гіпотезою H_0 , тест Діболда–Маріано має асимптотичний стандартний нормальний розподіл. Негативні значення тесту вказують на те, що перша модель у парі є кращою, а позитивні — на те, що кращою є друга модель; двостороннє $p < \alpha$ також є підставою для відхилення H_0 . Харві, Лейборн та Ньюболд [15] запровадили модифіковану версію, яка враховує малі вибірки, а Кларк [16] показав, що асимптотична нормальність тесту Діболда–Маріано зберігається під час оцінювання параметрів моделі. Отже, методологія оцінки якості прогнозування в роботі відповідає попереднім підходам: порівняльний аналіз прогнозів шляхом формування результатів ковзного прогнозу тестової вибірки, три допоміжні показники (MAE, RMSE, MAPE) та тест Діболда–Маріано як статистична верифікація відмінностей між моделями.

1.3 Економічні чинники та макроекономічні детермінанти формування цін на зернові культури

Світова ціна на пшеницю визначається під сукупним впливом різних факторів: фундаментальних (попит і пропозиція, рівень запасів),

макроекономічних (курси валют, ціни на енергоносії), фінансових ринків (поведінка фінансових інвесторів) та геополітичних (торговельні обмеження, війни). Розуміння цих визначальних факторів є необхідною передумовою для теоретично обґрунтованого вибору вхідних змінних для моделей прогнозування.

Ключовий показник напруженості ринку — співвідношення запасів до споживання (S/U) — публікується Міністерством сільського господарства США (USDA) у щомісячному звіті WASDE. Ірвін та Сандерс [17] продемонстрували, що падіння S/U нижче критичного порогу $\sim 15\text{--}18\%$ є провідним індикатором стрибків цін: саме вичерпання запасів нижче цього рівня було домінуючим фундаментальним фактором, що спричинив продовольчі кризи 2007–2008 та 2010–2011 років. Торговельна політика урядів є, іншими словами, незалежною змінною. Роуз та ін. [18] задокументували, що блокування експорту зерна з України та росії у 2022 році зменшило доступні запаси приблизно на 30 відсотків, що призвело до зростання цін на пшеницю з ~ 800 до ~ 1325 центів за бушель протягом двох місяців.

Зв'язок між нафтою та пшеницею передається через два основні канали. Канал витрат: нафта є важливим фактором виробництва зерна через добрива, паливо для техніки та транспорту — 10-відсоткове зростання цін на нафту підвищує витрати на виробництво пшениці на 3–7% [19]. Канал біопалива: коли ціна на нафту перевищує 80 доларів за барель, біопаливо з зернових стає прибутковим і створює додатковий попит на зерно з боку енергетичної галузі [20]. Чоудхурі та ін. [21] виявили асиметрію цієї залежності: 10-відсотковий шок зростання цін на нафту підвищує ціну на пшеницю на 2,8%, тоді як аналогічний шок зниження цін підвищує її лише на 1,4%. Ця нелінійність є аргументом на користь врахування нафти марки Brent як окремого фактора в багатofакторних моделях.

Оскільки на біржі CBOT пшениця котирується в доларах, курс долара безпосередньо впливає на відносну конкурентоспроможність американських

виробників та доступність продукції для покупців з інших валютних зон. Зміцнення долара (зростання індексу DXY) призводить до зниження зовнішнього попиту (та рівноважної ціни на CBOT). Ріс [22] показав, що 1% подорожчання долара пов'язане зі зниженням індексу цін на сировинні товари на 0,6–1,0%. Чен та Жанк [23] виявили, що DXY пояснює від 15 до 25 відсотків квартальних коливань ціни на пшеницю, що є найвищим показником серед нефундаментальних детермінант і вагомою причиною для включення його до набору предикторів.

VIX виступає проксі-показником глобальної схильності до ризику. Зі зростанням VIX інституційні інвестори скорочують позиції в ризикових активах, таких як товарні ф'ючерси, і переносять капітал у «безпечні гавані». Менсі та ін. [24] продемонстрували, що VIX є статистично значущим предиктором щотижневих змін індексу CRB із затримкою від 1 до 5 торгових днів. Ця залежність для пшениці є нелінійною та залежить від режиму: вона статистично незначуща для $VIX < 20$, тоді як для $VIX > 30$ (кризові режими) кореляція між VIX та ціною на пшеницю стає сильно негативною [25]. І саме ця нелінійна поведінка, що залежить від режиму, спонукає до використання моделей нейронних мереж, які можуть адаптуватися до різних ринкових умов.

Вплив індексу S&P 500 на ціну пшениці є наслідком зростаючої фінансизації товарних ринків — процесу, в рамках якого фінансові інвестори (пенсійні фонди, ETF, хедж-фонди) вийшли на ринки товарних ф'ючерсів, розглядаючи їх як клас активів для диверсифікації. Тан і Шен [26] задокументували, що з 2004 року кореляція між товарними ф'ючерсами та прибутковістю акцій зросла з нульового рівня до +0,40–0,50: у ситуаціях «уникнення ризику» фінансові інвестори ліквідують обидва класи активів одночасно, штучно створюючи позитивний зв'язок між ринками. Ірвін і Сандерс [17] показали, що 10-відсоткове збільшення чистих довгих позицій трейдерів товарних індексів пов'язане з підвищенням ціни на пшеницю на 1,5–3 відсотки, навіть за відсутності змін у фундаментальних показниках.

Досвід 2022 року продемонстрував, що геополітичні події можуть нівелювати прогнозу силу будь-яких моделей протягом кількох тижнів. Девадосс і Рідлі [27] за допомогою просторової рівноважної моделі показали, що шок пропозиції внаслідок вторгнення підвищив світові ціни на пшеницю приблизно на 2%, тоді як в Україні вони впали майже на 27%

Ці шоки є принципово непередбачуваними для будь-якого класу моделей, але їхній ефект можна врахувати за допомогою певних фіктивних змінних, що кодують структурні розриви. У дослідженні побудовано дві такі змінні: D1_КОВІД (лютий–вересень 2020 р.) та D2_Вторгнення (від 24.02.2022 р.), які разом із п'ятьма кількісними факторами утворюють теоретично обґрунтований набір вхідних змінних, що охоплюють усі основні канали впливу на ціну пшениці.

РОЗДІЛ 2

АНАЛІЗ ДИНАМІКИ ЦІН ЗА ДОПОМОГОЮ ЧАСОВИХ РЯДІВ

2.1 Динаміка та структура часових рядів цін на пшеницю

Джерелом даних, використаних у дослідженні, є п'ять наборів щоденних часових рядів, завантажених із платформи Yahoo Finance [28] за допомогою бібліотеки `yfinance` (Python). Зведену характеристику наборів даних наведено в додатку А, код програмної реалізації — у додатку Б

Часовий проміжок охоплює період з 4 січня 2010 року по 30 березня 2026 року (понад 4000 торгових днів).

Нижня межа вибірки (2010 рік) була обрана з огляду на вплив фінансової кризи 2008–2009 років, щоб виключити надмірну структурну неоднорідність. Цей набір даних охоплює два значні зовнішні шоки: пандемію COVID-19 (2020 рік) та повномасштабне вторгнення росії в Україну (лютий 2022 року).

Ціни на пшеницю наведені за ф'ючерсними контрактами CBOT ZW=F у центах за бушель — це стандартна одиниця виміру Чиказької товарної біржі. Як макроекономічні прогнознi змінні ми використовуємо чотири показники, що зазвичай застосовуються в дослідженнях цін на сировинні товари:

— ціна на нафту марки Brent ($BZ=F$, дол. США/барель) — відображає логістичні та енергетичні витрати у сільськогосподарському виробництві та транспортуванні;

— індекс долара США (DXY) (NYB , DXY) — відображає валютні фактори: зміцнення долара зазвичай робить активи, що котируються в доларах, дорожчими для іноземних покупців і знижує попит на американське зерно;

— індекс волатильності CBOE ($\wedge$ VIX) — «індикатор паніки» на світових ринках, що відображає рівень очікуваної невизначеності та схильність інвесторів до ризику;

— індекс S&P 500 ($\wedge$ GSPC) — об’єктивний фактор стану американської економіки, який також відображає стан світової економіки та впливає на ринок сировинних товарів через ефект перетікання між активами.

Відсутні значення, спричинені розбіжностями в календарях торгів (британські святкові дні не збігаються з торговими днями на ринку нафтових ф'ючерсів), були заповнені за допомогою методу «forward fill» — тобто заповнення попереднім спостереженим значенням. Після такої обробки остаточна вибірка не містить відсутніх значень. Для первинного аналізу властивостей вибірки застосована описова статистика, результати якої наведено у Таблиці 2.1.

Таблиця 2.1 — Описова статистика часових рядів (2010–2026)

Показник	Середнє	Ст.відх.	Мін	Макс	Коеф. варіації %	Асиметрія	Ексцес
Пшениця	598,18	139	361	1425,25	23,24	1,23	2,28
Нафта	77,69	23,46	19,33	127,98	30,2	0,13	-0,89
DXU	92,82	9,33	72,93	114,11	10,05	-0,34	-0,98
VIX	18,44	6,84	9,14	82,69	37,12	2,38	11,04
SP500	3003,49	1552,82	1022,58	6978,6	51,7	0,8	-0,31

Джерело: розроблено автором на основі [28]

Середня ціна на пшеницю становить 598,18 центів за бушель (стандартне відхилення 139,0), коефіцієнт варіації (CV) = 23,24% — амплітуда коливань помірна. Найнижча ціна становила 361 цент за бушель (2016 р.), найвища — 1 425,25 центів за бушель (квітень 2022 р.). Значення асиметрії 1,23 вказує на правобічну асиметрію; критерій Харке–Бера [4] підтверджує ненормальний розподіл. Коефіцієнт варіації нафти Brent є досить високим і становить 30,2%. VIX характеризується коефіцієнтом варіації 37,12% та ексцесом 11,04 — це зумовлено різкими кризовими коливаннями (березень 2020 року). Індекс S&P 500 має найбільшу абсолютну амплітуду коливань

(стандартне відхилення 1 552,82 пункти), що пояснюється сильним висхідним трендом, а не хаотичними коливаннями (коефіцієнт варіації $CV = 51,7\%$).

Динаміку усіх п'яти часових рядів упродовж 2010–2026 рр. відображає рисунок 2.1. Для кращого сприйняття взаємозв'язків між змінними графіки суміщені по єдиній часовій осі, а зони структурних шоків (COVID-2020 та Вторгнення 2022) виділені кольоровим затіненням.

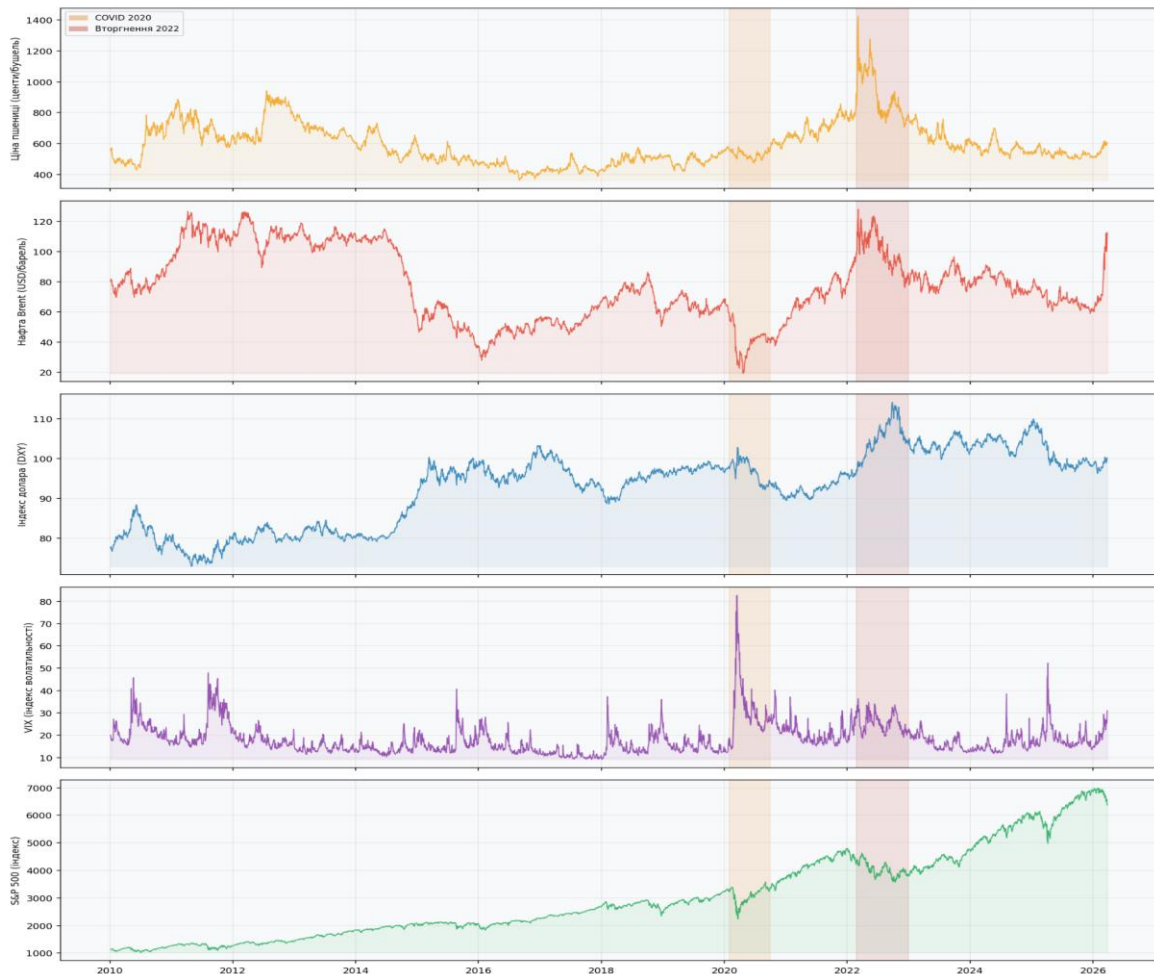


Рисунок 2.1 — Динаміка часових рядів (2010–березень 2026)

Джерело: розроблено автором на основі [28]

Графік демонструє ключову закономірність: у період з 2012 по 2016 рік ціни на пшеницю мали тенденцію до зниження (приблизно з 900 до 360 центів США за бушель) і стабілізувалися у 2017–2019 роках. Під час пандемії COVID-19 (2020 р.), незважаючи на різке падіння цін на нафту та стрибок індексу VIX, ціни на пшеницю помірно зросли, що відображає її властивості «безпечного активу». На початку 2022 року одночасно відбулося різке

зростання різних показників (спотові ціни на пшеницю досягли приблизно 1 425 центів за бушель, нафта марки Brent перевищила 130 доларів за барель, а індекс VIX піднявся вище 35), що підкреслило системний характер геополітичних потрясінь. Індекс S&P 500 демонстрував стійку тенденцію до зростання протягом усього періоду спостереження, навіть на тлі хаотичної ситуації на сировинному ринку.

Більш детальний аналіз розподілу змінних представлений на рисунку 2.2, де для кожної змінної побудовано гістограму з накладеною кривою щільності (KDE) та вертикальними лініями середнього й медіани.

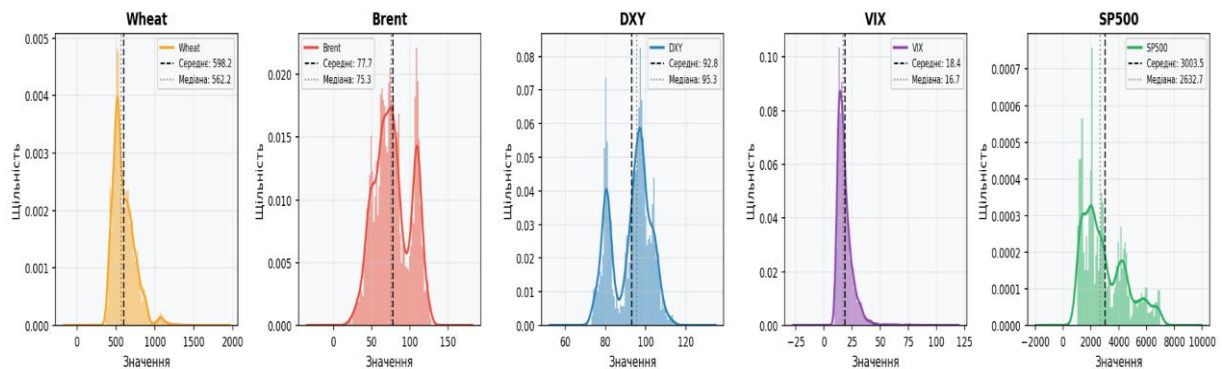


Рисунок 2.2 — Розподіл значень змінних (гістограма + KDE)

Джерело: розроблено автором на основі [28]

Ціни на пшеницю та нафту марки Brent мають правосторонню асиметрію; індекс VIX — найбільш виражену правосторонню асиметрію (спалахи паніки); індекс DXY наближається до симетричного розподілу; індекс S&P 500 демонструє бімодальність (два стійкі рівні цін до та після відновлення). Жодна з цих змінних не підпорядковується нормальному розподілу, що зумовлює необхідність застосування нелінійних та робастних методів моделювання.

Волатильність ціни пшениці, обчислена як 30-денне ковзне стандартне відхилення прибутковості, представлена на рисунку 2.3.

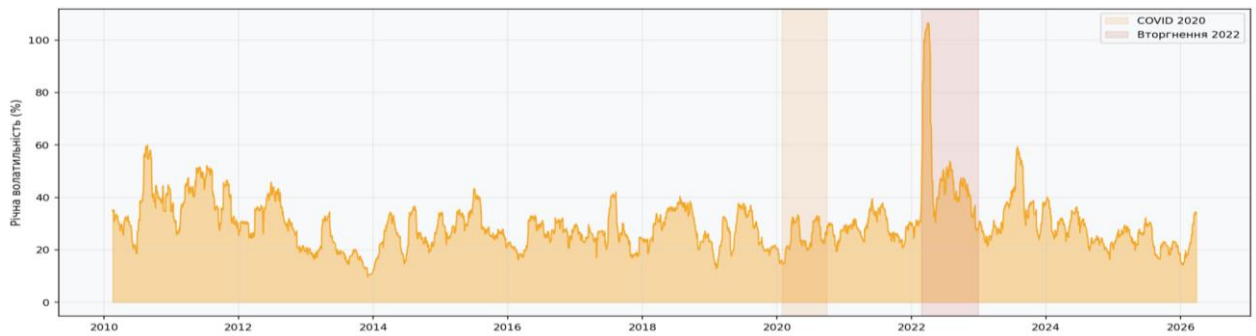


Рисунок 2.3 — Волатильність цін на пшеницю (30-денне ковзне стандартне відхилення)

Джерело: розроблено автором на основі [28]

Аналіз волатильності (рис. 2.3) показує три основні стрибки: помірний у 2010–2012 роках (продовольча криза), незначний у 2014–2015 роках (геополітична нестабільність в Україні) та рекордний у першому кварталі 2022 року (річна волатильність перевищила 100%). Під час пандемії COVID-19 (2020 р.) волатильність пшениці зросла лише незначно, що відповідає її відносно «безпечним» характеристикам. Після середини 2022 р. вона повернулася до фонового рівня 15–35% на рік.

Для оцінки лінійних взаємозв'язків між змінними побудовано матриці кореляцій Пірсона як для рівнів цін, так і для денних прибутковостей (перших логарифмічних різниць), що наведені на рисунку 2.4.

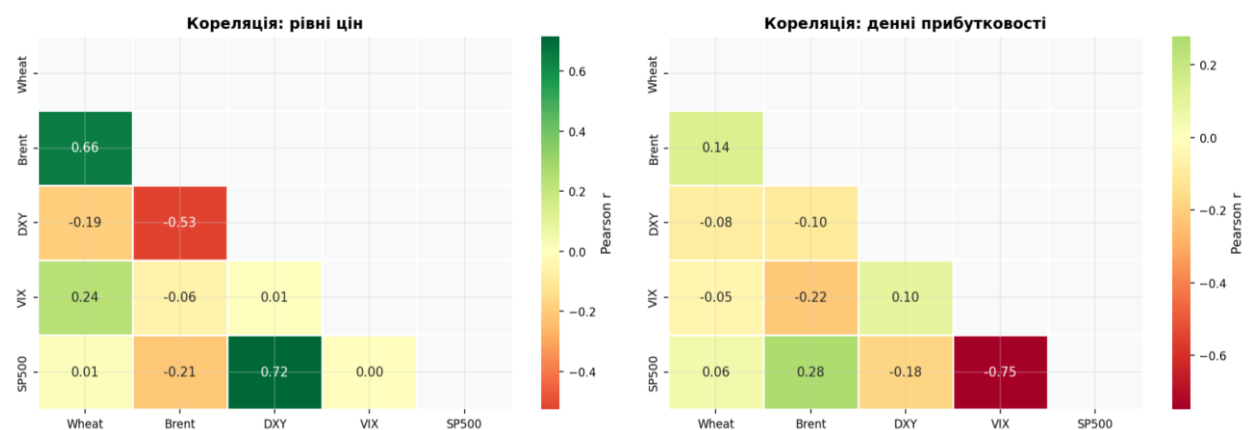


Рисунок 2.4 — Матриця кореляцій між змінними (рівні цін та денні прибутковості)

Джерело: розроблено автором на основі [28]

Що стосується цін, найсильніша позитивна кореляція спостерігається між пшеницею та нафтою марки Brent ($r = 0,66$); індекс долара США (DXY) має негативну кореляцію з цінами на пшеницю ($r = -0,19$) та нафту марки Brent ($r = -0,53$) — це відображає механізм «зміцнення долара — падіння цін на сировинні товари». Кореляція щоденних дохідностей значно слабшає: коефіцієнт кореляції пшениці з нафтою Brent знижується до $r = 0,14$, а кореляція пшениці з VIX стає від'ємною ($r = -0,05$), що означає нижчу залежність цін на пшеницю від макроекономічних факторів у короткостроковій перспективі. Найбільш значуща кореляція між дохідностями спостерігається між індексом S&P 500 та VIX ($r = -0,75$).

Отже, попередній статистичний аналіз даних підтверджує, що часові ряди досліджуваних змінних мають такі характеристики: ненормальний розподіл, виражену волатильність із часовими кластерами, наявність структурних розривів, а також кардинально різні моделі короткострокової та довгострокової коваріації. Ці особливості означають, що перед моделюванням необхідно провести більш ретельну попередню обробку даних, а також надають теоретичне обґрунтування для поєднання класичної економетричної моделі та методів нейронних мереж, які будуть представлені в наступних розділах [29].

2.2 Діагностика властивостей часових рядів та перевірка на стаціонарність

Перевірка на стаціонарність є необхідним етапом підготовки даних часових рядів: використання неперетворених нестаціонарних рядів призводить до «хибної регресії», коли значущі коефіцієнти відображають лише загальні тенденції, а не справжні економічні взаємозв'язки [30].

Майже всі фінансові та товарні ціни є нестационарними за рівнем і добре описуються процесами з одиничним коренем $I(1)$: їхні перші різниці задовольняють припущення про стаціонарність. У цьому підрозділі проводяться тести ADF та KPSS, аналіз корелограм ACF/PACF, а також аналіз коінтеграції за методом Енгла–Грейнджера.

Першим кроком діагностики є візуальний огляд рядів у рівнях та після першого диференціювання. Рисунок 2.5 демонструє ціну пшениці у рівнях (верхній лівий панель) та її першу різницю ΔWheat (верхній правий панель), а також відповідні розподіли значень (нижній рядок).

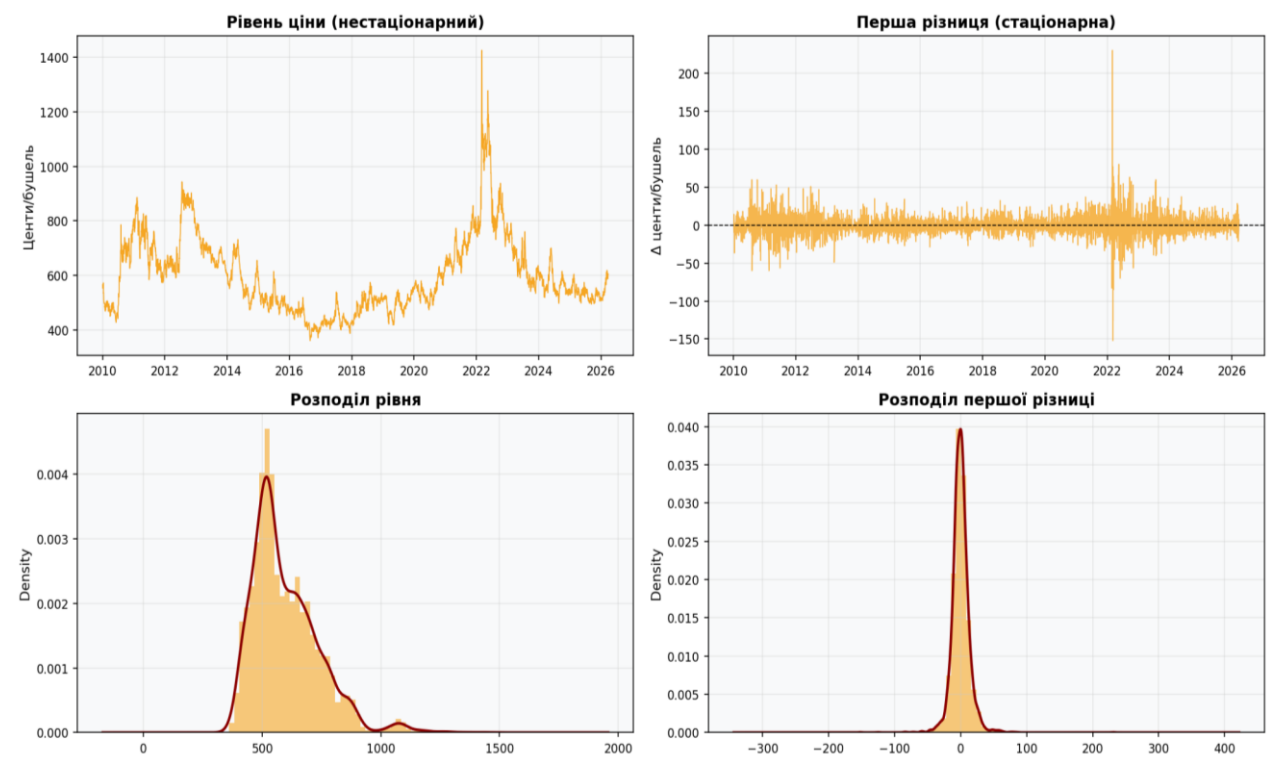


Рисунок 2.5 — Рівень та перша різниця ціни пшениці (2010–2026)

Джерело: розроблено автором на основі [28]

Ціни на пшеницю мають класичні ознаки процесу $I(1)$: відсутній детермінований рівень повернення, спостерігаються тривалі підтренди, а дисперсія є змінною. Натомість $\Delta \text{Пшениця}$ коливається біля нуля без трендової компоненти та має лептокуртичний розподіл, за винятком значного викиду в січні-березні 2022 р. (реакція на військове вторгнення). Це підтверджує $I(1)$ -характер вихідного ряду.

ADF-тест (H_0 : наявність одиничного кореня) проводився з автоматичним вибором лагів за AIC та включенням константи і тренду. Результати для рядів у рівнях — Таблиця 2.2.

Таблиця 2.2 — ADF тест: рівні (2010–2026)

Показник	ADF stat	p-value	Крит. 1%	Крит. 5%	Крит. 10%	Стаціонарність
Пшениця	-3,0862	0,0276	-3,432	-2,8622	-2,5671	Так
Нафта	-1,7429	0,4091	-3,432	-2,8622	-2,5671	Hi
DXY	-1,7227	0,4194	-3,432	-2,8622	-2,5671	Hi
VIX	-6,3228	0	-3,432	-2,8622	-2,5671	Так
SP500	0,6079	0,9878	-3,432	-2,8622	-2,5671	Hi

Джерело: розроблено автором на основі [28]

Результати для рівнів неоднозначні. Пшениця: ADF = $-3,0862$, $p = 0,0276$ — граничний результат, що потребує підтвердження KPSS. VIX: ADF = $-6,3228$, $p = 0,000$ — стаціонарний за природою (mean-reverting). Brent ($p = 0,409$), DXY ($p = 0,419$), SP500 ($p = 0,988$) — H_0 не відхиляється, ряди нестаціонарні. Результати наведено у Таблиці 2.3.

Таблиця 2.3 — KPSS тест: рівні (2010–2026)

Показник	KPSS stat	p-value	Крит. 5%	Стаціонарність
Пшениця	0,8837	0,01	0,463	Hi
Нафта	2,2081	0,01	0,463	Hi
DXY	7,6802	0,01	0,463	Hi
VIX	0,5161	0,038	0,463	Hi
SP500	9,3594	0,01	0,463	Hi

Джерело: розроблено автором на основі [28]

Тест KPSS відхилив нульову гіпотезу H_0 про стаціонарність для всіх п'яти змінних (критичне значення 0,463): пшениця — 0,8837; Brent — 2,2081; DXY — 7,6802; VIX — 0,5161; SP500 — 9,3594. Навіть для VIX, який тест ADF визначив як стаціонарний, KPSS зафіксував нестаціонарність, що свідчить про зміни режиму волатильності. Загальний висновок: усі п'ять часових рядів є нестаціонарними. Наступним кроком є верифікація того, що перші різниці всіх рядів є стаціонарними, тобто змінні є інтегрованими першого порядку $I(1)$. Результати ADF-тесту для перших різниць наведено у таблиці 2.4

Таблиця 2.4 — ADF тест: перші різниці (2010–2026)

Показник	ADF stat	p-value	Крит. 1%	Крит. 5%	Крит. 10%	Стаціонарність
Δ Пшениця	-14,6273	0	-3,432	-2,8623	-2,5671	Так
Δ Нафта	-27,184	0	-3,432	-2,8622	-2,5671	Так
Δ DXU	-19,2933	0	-3,432	-2,8622	-2,5671	Так
Δ VIX	-26,749	0	-3,432	-2,8622	-2,5671	Так
Δ SP500	-21,1713	0	-3,432	-2,8622	-2,5671	Так

Джерело: розроблено автором на основі [28]

Результати підтверджують, що перші різниці є стаціонарними, а отже всі ряди мають порядок інтегрованості $I(1)$. Цей висновок є принциповим для подальшого моделювання: у моделях VAR використовуватимуться перші різниці змінних, тоді як для ARIMA застосовується параметр $d = 1$ (одне диференціювання). ACF/PACF корелограми (рисунок 2.6) ідентифікують структуру автокореляції і слугують основою для вибору порядків p та q моделі ARIMA

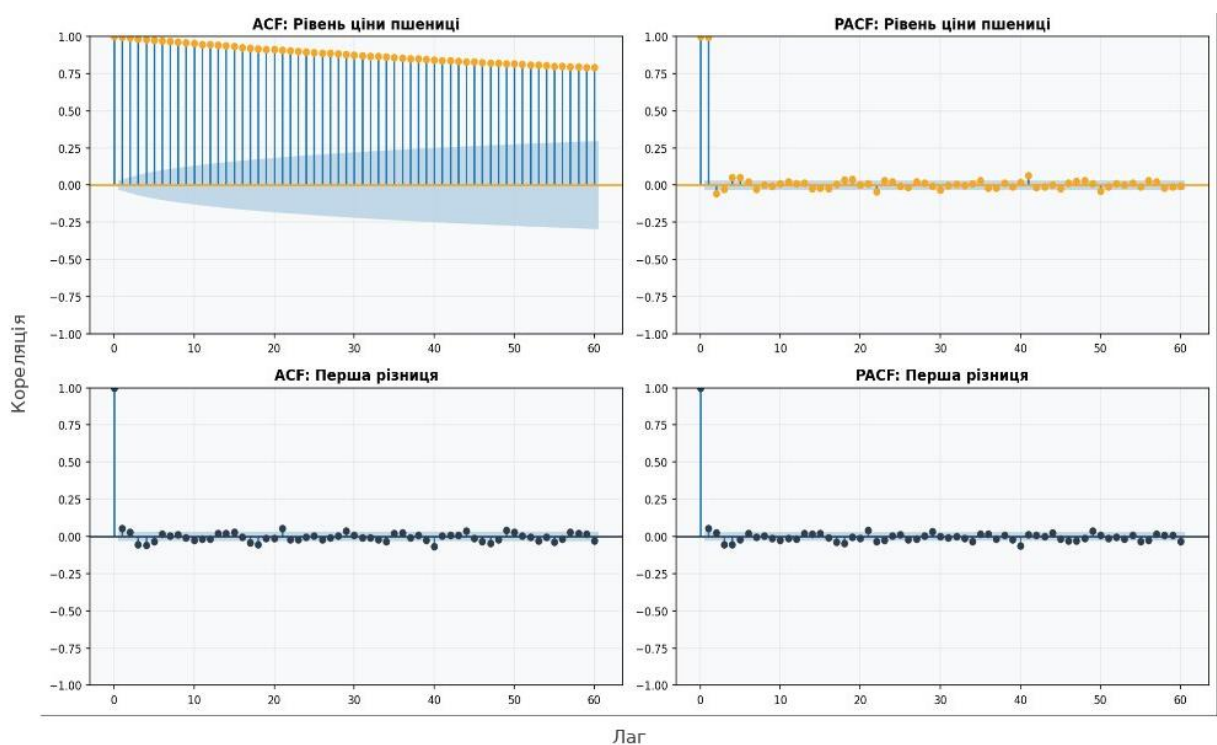


Рисунок 2.6 — Корелограми ACF та PACF для ціни пшениці (рівні та перша різниця)

Джерело: розроблено автором на основі [28]

Коефіцієнт автокореляції рівнів (ACF) спадає повільно і є значущим для лагів 60+ — це типова ознака моделі $I(1)$. Частковий коефіцієнт автокореляції (PACF) рівнів має значущий пік лише на лазі 1, а потім дорівнює нулю (характеристика моделі $AR(1)$ або «випадкового блукання»). Після диференціювання ACF та PACF для Δ Пшениця не демонструють жодних значущих лагів — це профіль білого шуму, що узгоджується з гіпотезою про слабку форму ринкової ефективності.

Аналогічний аналіз проведено для перших різниць макроекономічних факторів — Δ Brent, Δ DXY, Δ VIX та Δ SP500. Результати наведено на рис. 2.7.

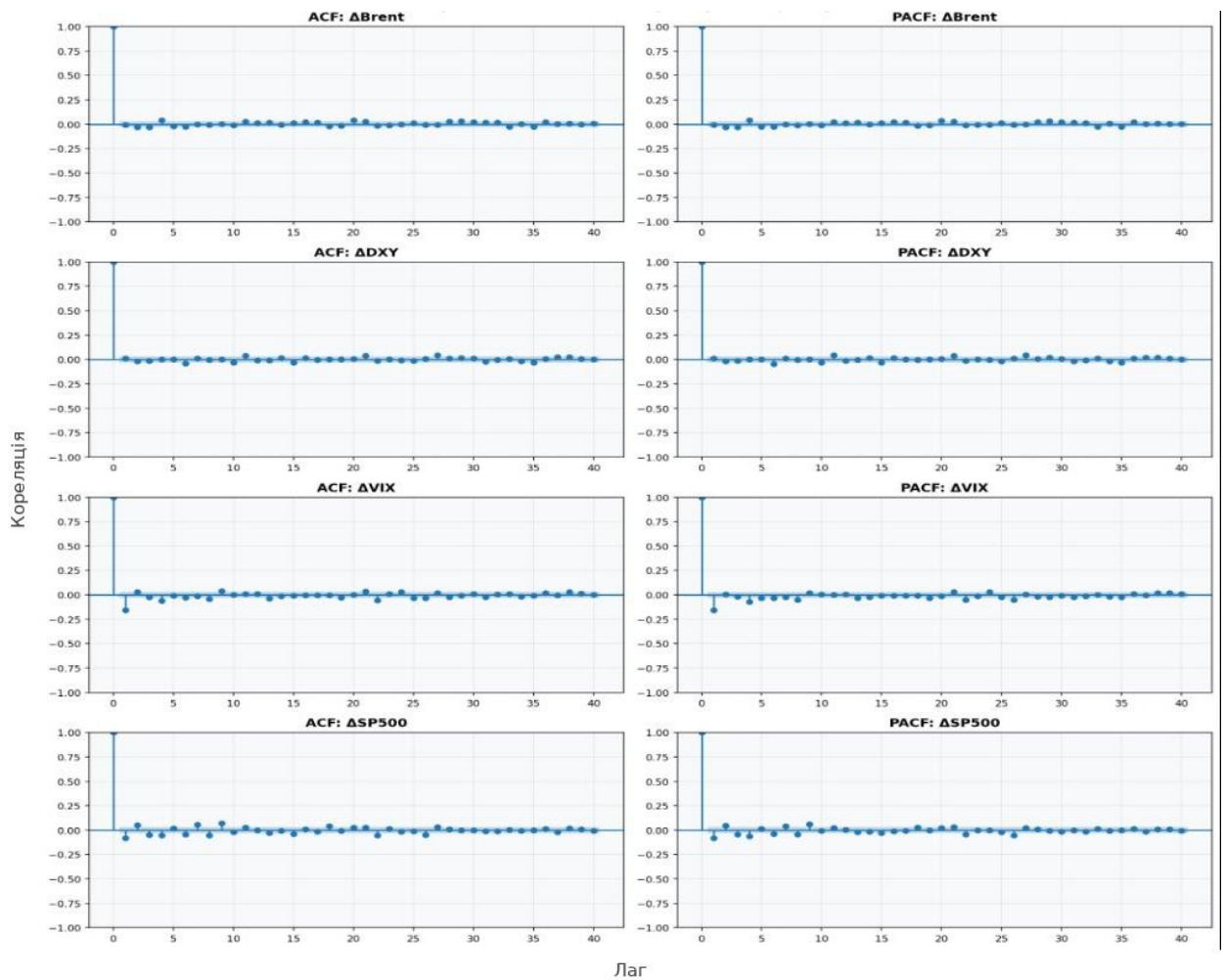


Рисунок 2.7 — ACF та PACF для перших різниць макроекономічних факторів

Джерело: розроблено автором на основі [28]

Перші похідні всіх чотирьох факторів (рис. 2.7) не мають значущих піків ACF та PACF, що підтверджує їхню стаціонарність. Виняток: ΔVIX демонструє слабкий негативний пік PACF на першій затримці (близько $-0,10$), що є типовим для рядів волатильності. Відсутність лінійної автокореляції означає, що довжина лагу для VAR визначається за критеріями AIC/BIC; водночас набір моделей нейронних мереж (MLP, LSTM) здатен виявляти приховані нелінійні залежності, недоступні для лінійних специфікацій.

Оскільки всі ряди є $I(1)$, перевірено наявність коінтеграції — стабільного лінійного зв'язку між нестационарними змінними — за двокроковим методом Енгла–Грейнджера: регресія рівнів пшениці на кожен фактор, потім ADF-тест залишків на стаціонарність. Результати — Таблиця 2.5.

Таблиця 2.5 — Тест на коінтеграцію Енгла–Грейнджера

Пара рядів	t-stat	p-value	Крит. 1%	Крит. 5%	Коінтегрованість
Пшениця ~ Нафта	-4,0871	0,0054	-3,8991	-3,3376	Так
Пшениця ~ DXY	-2,8979	0,1365	-3,8991	-3,3376	Ні
Пшениця ~ VIX	-3,4969	0,0327	-3,8991	-3,3376	Так
Пшениця ~ SP500	-2,8602	0,1473	-3,8991	-3,3376	Ні

Джерело: розроблено автором на основі [28]

Коінтеграція встановлена для пар «Пшениця–Brent» ($t = -4,0871$, $p = 0,0054$, рівень значущості 1%) та «Пшениця–VIX» ($t = -3,4969$, $p = 0,0327$, рівень значущості 5%): обидві пари мають спільні довгострокові рівноважні взаємозв'язки. Коінтеграція не виявлена для пшениці та DXY ($p = 0,1365$) і пшениці та S&P 500 ($p = 0,1473$), а вплив цих факторів є переважно короткостроковим. Оскільки горизонт прогнозування в дослідженні становить 1 день, ЕСМ не будується, але виявлений зв'язок теоретично обґрунтовує включення Brent та VIX до VAR-моделі.

Спільна діагностика підтвердила: (1) всі п'ять рядів є $I(1)$ — нестационарними за рівнями та стаціонарними за першими різницями (ADF + KPSS + візуальний огляд); (2) перші різниці не мають вираженої лінійної автокореляційної структури (близькість до білого шуму обмежує потенціал

лінійних моделей); (3) коінтеграція пшениці–Brent та пшениці–VIX підтверджує теоретичне обґрунтування їх включення до багатовимірних моделей. Нижче наведено специфікації, отримані в розділі 3: ARIMA з $d = 1$; VAR за першими різницями; моделі нейронних мереж (MLP, LSTM) — за нормалізованими рівнями.

2.3 Обробка структурних зламів та попередня підготовка вхідних параметрів

Підготовка даних охоплює три ключові аспекти: (1) виявлення та верифікацію структурних зламів; (2) конструювання дам-змінних; (3) нормалізацію вхідних параметрів для нейромереж.

Рисунок 2.8 ілюструє поведінку цін на пшеницю в зонах двох ключових шоків.

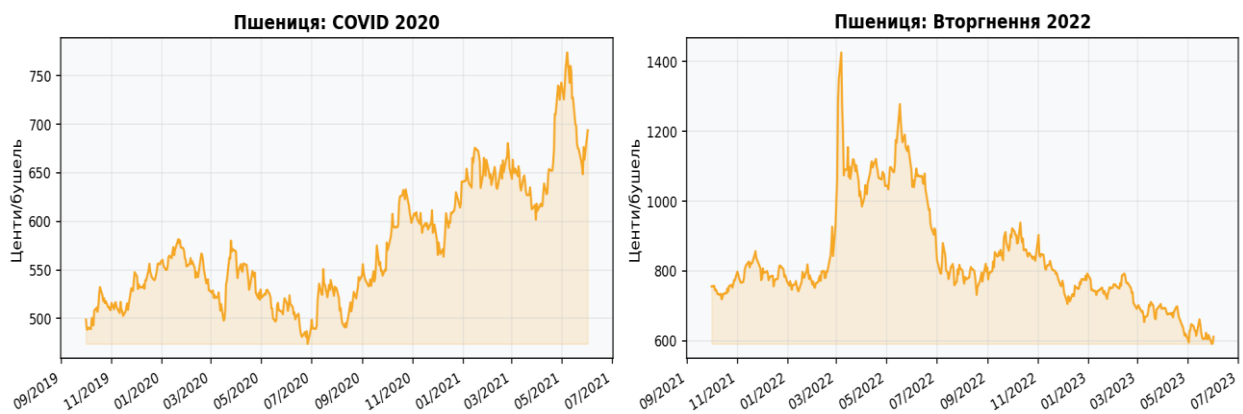


Рисунок 2.8— Поведінка цін на пшеницю під час структурних шоків (COVID-2020 та Вторгнення 2022)

Джерело: розроблено автором на основі [28]

Ліва панель відображає динаміку ціни пшениці у період COVID-19: від вересня 2019 до серпня 2021 р. Всупереч інтуїтивному очікуванню, початок пандемії (березень 2020) не спричинив ні обвалу, ні миттєвого стрибку цін —

зернові демонструють відносну стійкість завдяки стабільному базовому попиту. Натомість з літа 2020 р. розпочинається стійке зростання, що продовжується до середини 2021 р.: ціна піднялась з ~490 до ~750 ц/бушель (+53%). Це пов'язано з одночасним впливом кількох чинників: порушенням ланцюгів постачання, посухою у ключових виробничих регіонах (Канада, США), масштабними закупівлями Китаю та загальним сплеском інфляції на тлі монетарного стимулювання. Таким чином, шок COVID-2020 проявився не як раптовий ціновий стрибок, а як поступова зміна рівня цін і посилення волатильності.

Права панель відображає епізод Вторгнення 2022. Тут картина принципово інша: від моменту початку повномасштабного вторгнення (24 лютого 2022) ціна пшениці злетіла з ~800 до ~1 200–1 425 ц/бушель протягом менш ніж двох місяців — одне з найрізкіших короткострокових зростань в історії ринку. Причина — безпосередня: Україна і росія разом забезпечують понад 30% світового експорту пшениці, а блокування Чорноморських портів миттєво відрізало значну частину глобальної пропозиції. Відновлення цін до рівня ~600–700 ц/бушель відбулося поступово впродовж 2022–2023 рр. після підписання «зернової угоди» та адаптації ринку до нових умов. Цей епізод залишив стійкий структурний злам: рівень цін після 2022 р. суттєво відрізняється від докризового, а волатильність залишається підвищеною.

Тест Чоу (Ню: відсутність структурного злому) проводився для двох апріорно відомих дат — 11 березня 2020 р. (оголошення пандемії ВООЗ) та 24 лютого 2022 р. (початок вторгнення). Рисунок 2.9 ілюструє ці точки на часовому ряді.

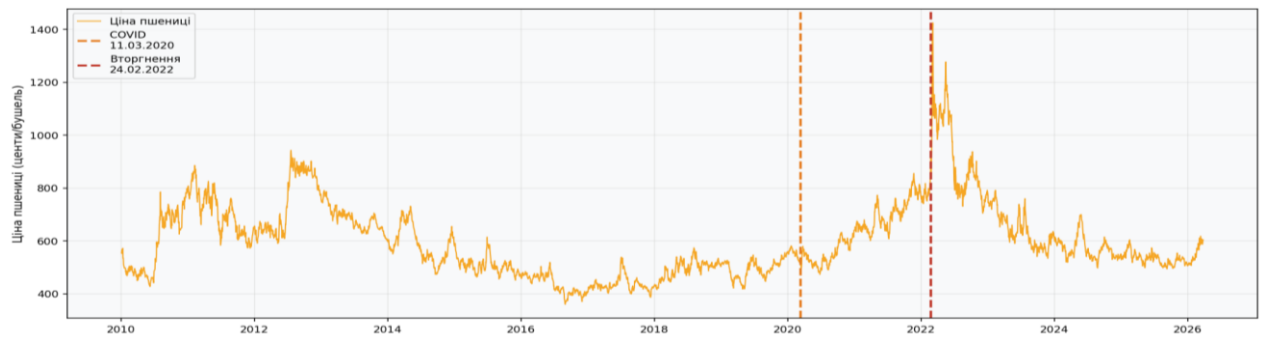


Рисунок 2.9 — Точки структурних зламів на ряді цін пшениці (2010–2026)

Джерело: розроблено автором на основі [28]

Дві вертикальні пунктирні лінії на графіку чітко позначають дві точки перегину. Перша (COVID, березень 2020 року) візуально збігається з початком висхідного тренду, що тривав понад рік. Друга (вторгнення, лютий 2022 року) збігається з найрізкішим стрибком ціни за всю історію ряду даних. Як бачимо, обидва ці періоди суттєво змінили траєкторію ряду — чи то за рівнем тренду, чи за нахилом, чи за дисперсією — і можна помітити, що вони відрізняються, навіть не дивлячись на графік.

Однак, щоб правильно сформулювати припущення про стаціонарність в економетриці, необхідний формальний тест, результати якого наведено в таблиці 2.6.

Таблиця 2.6 — Тест Чоу на структурні зміни

Подія	F-статистика	p-value	Структурний злам
COVID (2020-03-11)	641,698	0	Так
Вторгнення (2022-02-24)	562,888	0	Так

Джерело: розроблено автором на основі [28]

Результати тесту Чоу є однозначними: COVID (11.03.2020) — $F = 641,698$, $p = 0,000$; Вторгнення (24.02.2022) — $F = 562,888$, $p = 0,000$. Обидва зовнішні шоки спричинили статистично значущі структурні переломи; їх ігнорування у специфікації призведе до систематичних помилок. Оскільки дати шоків відомі, пошук невідомих точок розриву (Зібот–Ендрюс, Бай–Перрон) не є необхідним.

Дві бінарні фіктивні змінні створено для кодування зламів (рис. 2.10): $D1_KOBID = 1$ за період 01.02–30.09.2020 (гострий шок та адаптація, 8 місяців); $D2_Вторгнення = 1$ з 24.02.2022 до кінця даних як «сходінкова фіктивна змінна» (постійний структурний зсув). Обидві змінні включені до вхідних векторів ARIMA (як регресори ARIMAX), VAR (як детерміновані компоненти) та MLP/LSTM.

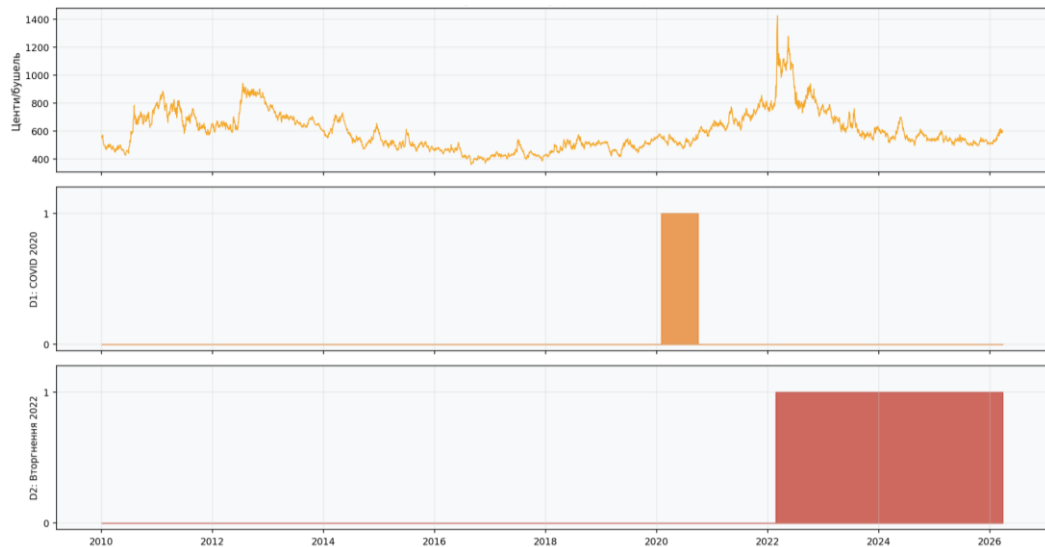


Рисунок 2.10 — Дамі-змінні структурних шоків (2010–2026)

Джерело: розроблено автором на основі [28]

Коефіцієнт кореляції ковзання (вікно 252 дні, див. рис. 2.11) свідчить про нестабільність кореляції між цінами на пшеницю та макрофакторами, що підтверджує висновок про перевагу методу нейронних мереж над лінійним методом.

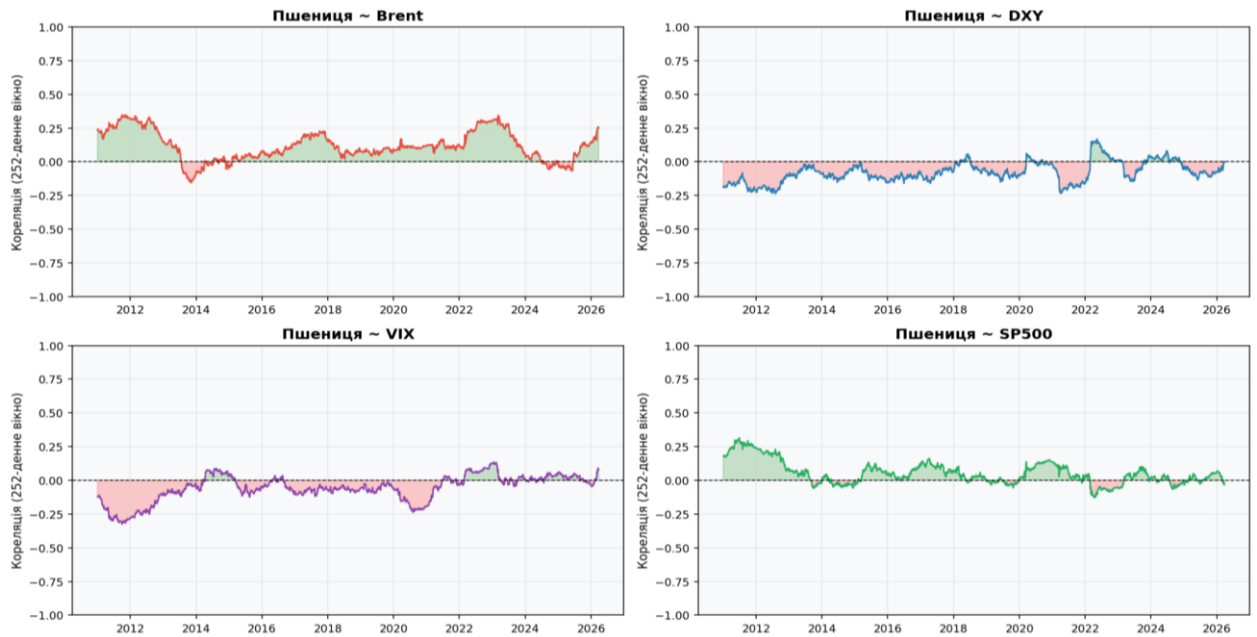


Рисунок 2.11 — Ковзна кореляція (252 дні) між пшеницею та макроекономічними факторами

Джерело: розроблено автором на основі [28]

Пшениця–Brent: найбільш мінлива кореляція (від $-0,25$ у 2016–2017 до $+0,60$ у 2022) — відображає різні режими (шейл-надвиробництво vs. геополітичний шок). Пшениця–DXY: переважно слабо від'ємна ($-0,1$ до $-0,3$), проте нестійка, що унеможливорює надійне лінійне використання. Пшениця–VIX: здебільшого від'ємна або нульова до 2019 р., непередбачувана після 2022 р. Пшениця–SP500: найслабша, навколо нуля — структурного зв'язку немає. Загальний висновок: всі чотири кореляції є нелінійними та режимозалежними; LSTM з механізмом пам'яті теоретично краще адаптується до таких умов, що буде перевірено у Розділі 3.

Моделі нейронних мереж (MLP, LSTM) чутливі до масштабу вхідних даних: різні діапазони значень змінних (пшениця $\sim 400\text{--}1400\$$, VIX $\sim 9\text{--}85\$$, SP500 $\sim 1000\text{--}6000\$$) призводять до нерівномірного зважування ознак у алгоритмах на основі градієнта та до насичення функцій активації [27VYR]. На рис. 2.12 показано порівняння вихідних даних із результатами, отриманими після нормалізації методами Min-Max та Z-score.

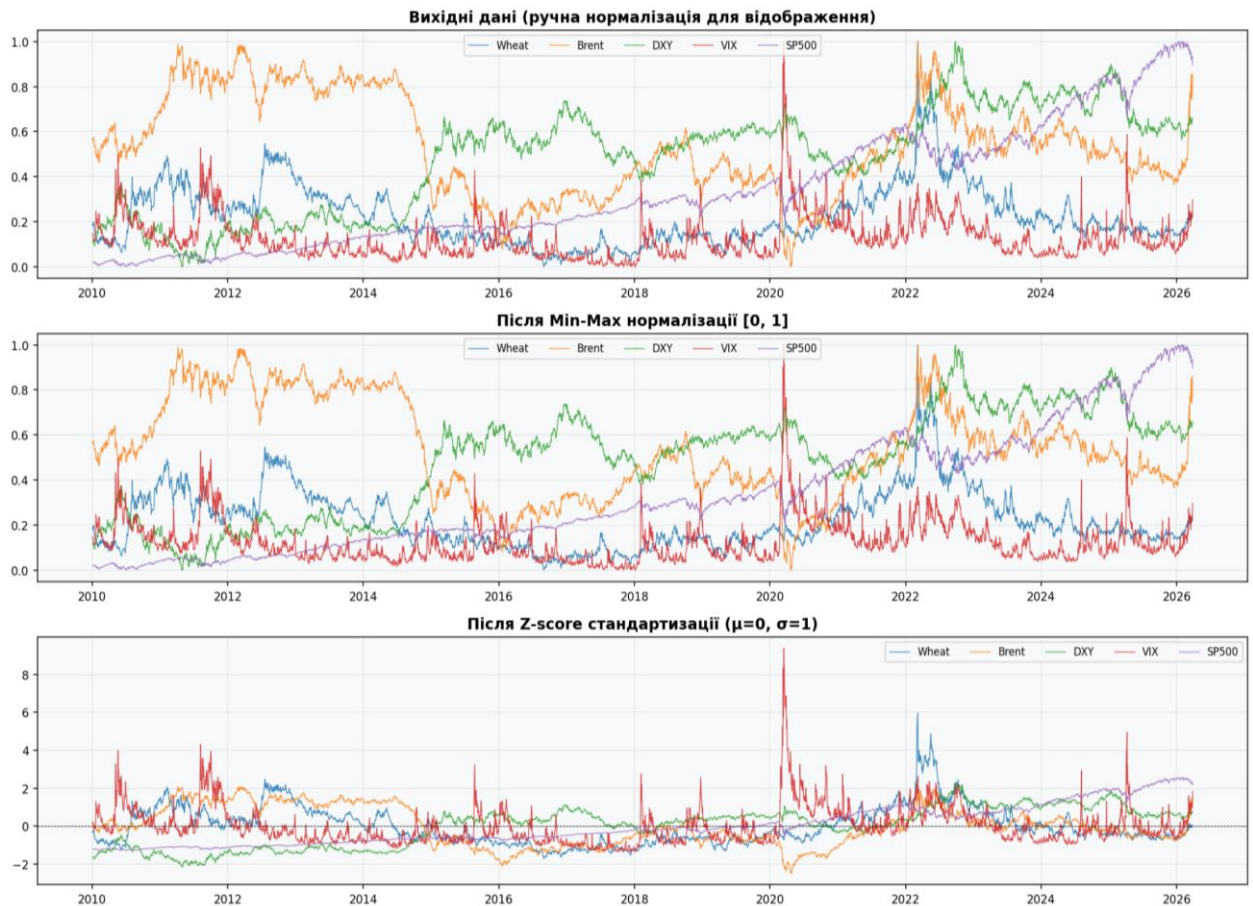


Рисунок 2.12 — Порівняння методів нормалізації даних (вихідні дані, Min-Max, Z-score)

Джерело: розроблено автором на основі [28]

Метод Min-Max нормалізує кожен ознаку до діапазону $[0, 1]$: $x' = (x - x_{\min}) / (x_{\max} - x_{\min})$, що забезпечує рівномірну участь у навчанні при збереженні динаміки даних. Z-score ($x' = (x - \mu) / \sigma$) має нульове середнє та одиничну дисперсію і надійно обробляє викиди (пік 2022 року обрізається як $\sim 5-6\sigma$ і не зводить усі інші значення до 0).

Обрано діапазон Min-Max $[0, 1]$, оскільки ReLU (MLP) та tanh (LSTM) демонструють найкращі результати в цьому інтервалі, а відносна структура між рядками даних зберігається; обернене перетворення з x' в $x = x' \cdot (x_{\max} - x_{\min}) + x_{\min}$ є точним і не потребує вибірових статистичних даних. Параметри нормалізації обчислюються виключно на основі навчальних даних (2010–2022), що запобігає витоків даних. Для ARIMA та VAR нормалізація вхідних даних не є обов'язковою, оскільки оцінювання параметрів методом

максимальної правдоподібності (MLE) та МНК є інваріантним до масштабування змінних [28].

Отриманий протокол підготовки даних відповідає таким вимогам: структурні розриви підтверджені тестом Чоу та закодовані у фіктивних змінних — D1_KOVID та D2_Вторгнення; вхідні дані для нейронної мережі нормалізовані за методом Min-Max $[0,1]$ на основі параметрів навчання (відсутність витоку даних); розподіл на тренувальний (2010–2022, 3390 спостережень) та тестовий (2023–2026, 814 спостережень) набори виконано з суворим дотриманням хронологічного порядку, що є достатнім для тесту Діболда–Маріано.

РОЗДІЛ 3

ПОРІВНЯЛЬНИЙ АНАЛІЗ ЕКОНОМЕТРИЧНИХ ТА НЕЙРОМЕРЕЖЕВИХ МОДЕЛЕЙ

3.1 Розробка та діагностика економетричних моделей ARIMA та VAR

Класичні лінійні моделі ARIMA та VAR є стандартом порівняння (benchmark) завдяки прозорості інтерпретації та надійним прогнозам на короткому горизонті. У підрозділі послідовно викладено специфікацію, оцінювання, діагностику та результати rolling прогнозу на тестовій вибірці 2023–2026 рр.

Модель ARIMA(p, d, q) враховує нестационарність ряду через d диференціювань: p — порядок AR-компоненти, q — порядок MA-компоненти. Оскільки ціна пшениці є I(1), параметр d = 1 фіксований.

Загальна форма моделі ARIMA(p, d, q) записується таким чином:

$$\Delta^d y_t = c + \sum_{i=1}^p \varphi_i \Delta^d y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t, \quad (3.1)$$

де y_t — значення ряду (ціна пшениці) у момент часу t ;

Δ^d — оператор взяття d -х різниць ряду: $\Delta y_t = y_t - y_{t-1}$;

c — константа (знесення/drift);

φ_i — коефіцієнти авторегресії (AR-частина), $i = 1, \dots, p$;

θ_j — коефіцієнти ковзного середнього (MA-частина), $j = 1, \dots, q$;

ε_t — білий шум: $\varepsilon_t \sim N(0, \sigma^2)$;

p, q — порядки AR- та MA-складових; для ціни пшениці порядок інтегрування $d =$

У розгорнутому вигляді для стаціонарного ряду (після d диференціювань) модель ARIMA(p,d,q) записується як:

$$\Delta^d y_t = \mu + \phi_1 \Delta^d y_{t-1} + \dots + \phi_p \Delta^d y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad (3.2)$$

де $\Delta^d y_t = (1 - L)^d y_t$ — d -та різниця ряду;

μ — константа (дрейф);

$\phi_1, \dots, \phi_p$ — коефіцієнти AR-компоненти, що відображають вплив минулих значень ряду;

$\theta_1, \dots, \theta_q$ — коефіцієнти MA-компоненти, що відображають вплив минулих шоків (інновацій);

ε_{t-j} — шоки (залишки) попередніх моментів часу.

Прогноз на один крок вперед: $\hat{y}_{t+1|t} = y_t + \Delta \hat{y}_{t+1}$, де $\Delta \hat{y}_{t+1}$ обчислюється за рівнянням (3.7). Параметри оцінюються MLE, 95% ДІ — за формулою (3.4).

$$\hat{y}_{t+1|t} = y_t + \hat{\phi}_1 \Delta y_t + \dots + \hat{\phi}_p \Delta y_{t-p+1} + \hat{\theta}_1 \varepsilon_t + \dots + \hat{\theta}_q \varepsilon_{t-q+1}, \quad (3.3)$$

де $\hat{y}_{t+1|t}$ — точковий прогноз ціни на крок $t+1$, обумовлений інформацією до моменту t ;

$\hat{\phi}_i, \hat{\theta}_j$ — оцінені параметри моделі;

ε_{t-j} — оцінені залишки попередніх кроків.

Процедура ідентифікації порядків p і q починається з візуального аналізу корелограм ACF і PACF диференційованого ряду Δ Пшениця, що наведені на рисунку 3.1.

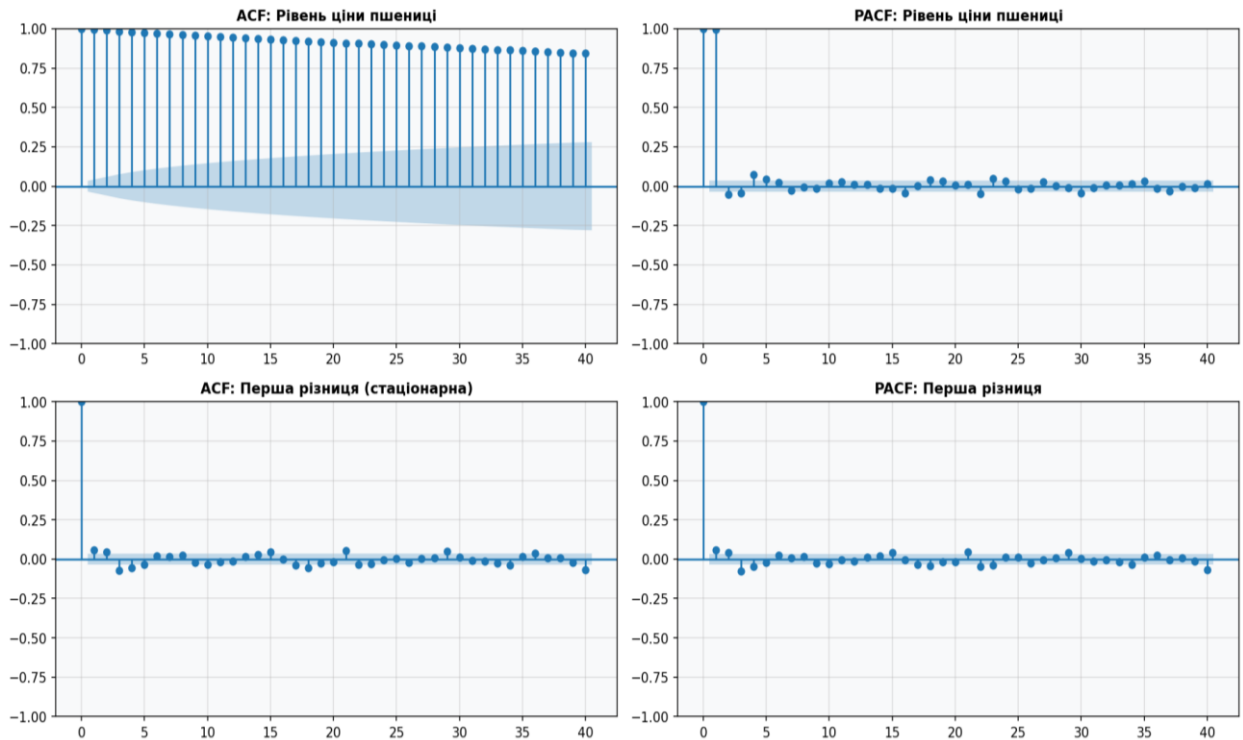


Рисунок 3.1 — ACF та PACF для ідентифікації ARIMA (рівні та перша різниця)

Джерело: розроблено автором на основі [28]

ACF/PACF ДПшениця (Рисунок 3.1) знаходяться в межах довірчого коридору на всіх лагах — відсутня чітка лінійна структура, тому порядки p і q підбираються формально за AIC/BIC шляхом перебору $p, q \in \{0 \dots 4\}$ при $d = 1$ (25 кандидатних моделей).

Для підбору специфікації проведено перебір $p \in \{0 \dots 4\}$, $q \in \{0 \dots 4\}$ при $d = 1$. Критерії AIC (3.5) та BIC (3.6) визначають оптимальну модель. Результати топ-10 — у Таблиці 3.1.

$$AIC = -2 \ln \hat{L} + 2k, \quad (3.4)$$

де $\ln \hat{L}$ — логарифм максимізованої функції правдоподібності;

k — кількість оцінюваних параметрів моделі ($p + q + 1$ для константи).

Перший доданок відображає якість підгонки (більша правдоподібність — менший AIC), другий — штраф за складність моделі. Модель з найменшим AIC є оптимальною. Байєсівський критерій (BIC) використовує більш жорсткий штраф:

$$\text{BIC} = -2 \ln \hat{L} + k \cdot \ln(n), \quad (3.5)$$

де n — кількість спостережень у вибірці. Оскільки $\ln(n) > 2$ при $n > 7$,

BIC сильніше карає за складність і схильний обирати більш парсимонійні моделі. Результати порівняння топ-10 специфікацій за AIC наведено у таблиці 3.1.

Таблиця 3.1 — Порівняння специфікацій ARIMA за AIC

Модель	AIC	BIC
ARIMA(3,1,4)	26618,37	26667,11
ARIMA(2,1,4)	26618,44	26661,09
ARIMA(4,1,3)	26618,75	26667,49
ARIMA(4,1,2)	26619,81	26662,47
ARIMA(3,1,2)	26627,28	26663,84
ARIMA(2,1,3)	26627,48	26664,04
ARIMA(3,1,3)	26628,97	26671,62
ARIMA(4,1,4)	26629,03	26683,87
ARIMA(4,1,0)	26639,6	26670,07
ARIMA(3,1,1)	26640,25	26670,72

Джерело: розроблено автором на основі [28]

Оптимальна специфікація — ARIMA(3,1,4) (AIC = 26 618,37, BIC = 26 667,11). Топ-3 моделі мають $\Delta\text{AIC} < 0,5$ і є рівноцінними за правилом Бернхема–Андерсона, проте ARIMA(3,1,4) обрана як основна.

ARIMA(3,1,4) оцінена на тренувальній вибірці 2010–2022 рр. методом MLE (алгоритм `innovations_mle`). Специфікація — рівняння (3.6); прогноз рівня відновлюється як $\hat{y}_{t+1} = y_t + \Delta\hat{y}_{t+1}$ (3.7).

$$\Delta y_t = \mu + \varphi_1 \Delta y_{t-1} + \varphi_2 \Delta y_{t-2} + \varphi_3 \Delta y_{t-3} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4}, \quad (3.6)$$

де $\Delta y_t = y_t - y_{t-1}$ — одноденна зміна ціни пшениці (ц/бушель);

$\varphi_1, \varphi_2, \varphi_3$ — коефіцієнти AR(3) компоненти;

$\theta_1, \theta_2, \theta_3, \theta_4$ — коефіцієнти MA(4) компоненти;

μ — константа (середньодобовий дрейф);

$\varepsilon_t \sim WN(0, \sigma^2)$ — процес білого шуму.

Прогноз ціни пшениці на наступний торговий день $t+1$ відновлюється з першої різниці:

$$\hat{y}_{t+1} = y_t + \Delta \hat{y}_{t+1}, \quad (3.7)$$

де $\Delta \hat{y}_{t+1}$ — прогнозована перша різниця з рівняння (3.6).

Залишки перевіряються на відповідність білому шуму: нульове середнє, постійна дисперсія, відсутність автокореляції. Результати діагностики — Рисунок 3.2.

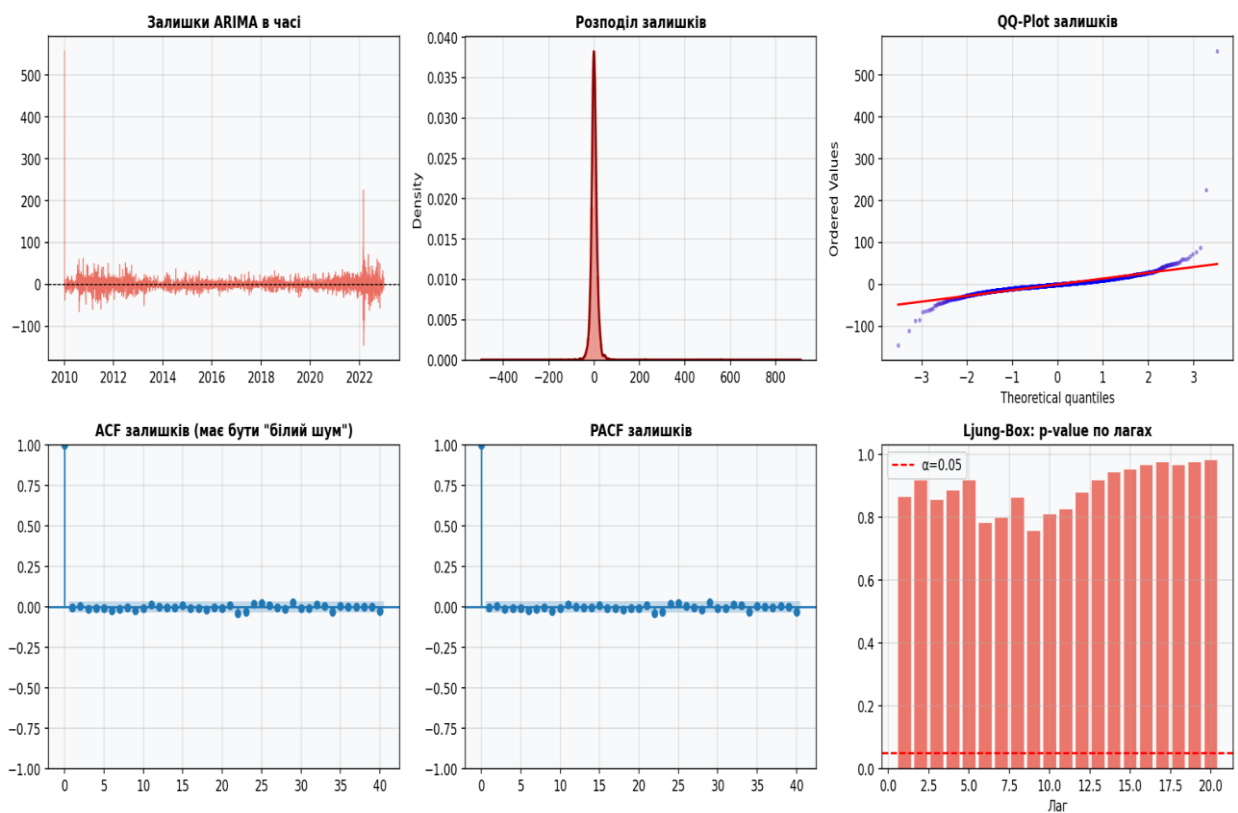


Рисунок 3.2 — Діагностика залишків ARIMA(3,1,4)

Джерело: розроблено автором на основі [28]

Графік у верхньому лівому куті — залишки у часі: коливаються симетрично навколо нуля без явного тренду чи гетероскедастичності, хоча у 2022 році спостерігається різке відхилення, спричинене шоком від вторгнення. Верхня середня панель: гістограма залишків з кривою KDE: розподіл симетричний і лептокуртичний (високий пік з товстими хвостами) — «товстохвоста» нормальність, типова для фінансових даних [31]. Вгорі

праворуч — Q-Q-діаграма: точки знаходяться близько до лінії теоретичного нормального розподілу в середній частині і відхиляються лише у двох хвостах, що є доказом умовно нормального розподілу залишків. Нижня ліва та середня панелі — ACF та PACF залишків: жоден із стовпчиків не виходить за межі довірчих інтервалів, що є ключовим доказом відсутності автокореляції залишків і, отже, правильності специфікації.

Для формального тестування відсутності автокореляції застосований тест Льюнга–Бокса (Ljung–Box), нульова гіпотеза якого H_0 : перші m автокореляцій залишків дорівнюють нулю (залишки є білим шумом).

Результати тесту Льюнга–Бокса для лагів 1–20 наведені у таблиці 3.2.

Таблиця 3.2 — Тест Льюнга–Бокса (ARIMA залишки)

Лаг	LB-статистика	p-value	Білий шум
1	0,029	0,8649	Так
2	0,1742	0,9166	Так
3	0,7765	0,8551	Так
4	1,1523	0,8859	Так
5	1,4692	0,9166	Так
6	3,2172	0,7811	Так
7	3,8334	0,7988	Так
8	3,9302	0,8634	Так
9	5,8222	0,7576	Так
10	6,0574	0,8104	Так
11	6,6684	0,8253	Так
12	6,6782	0,8781	Так
13	6,6998	0,917	Так
14	6,8025	0,9421	Так
15	7,1817	0,9524	Так
16	7,3641	0,9656	Так
17	7,5394	0,9754	Так
18	8,769	0,9649	Так
19	8,9018	0,9751	Так
20	9,1589	0,981	Так

Джерело: розроблено автором на основі [28]

Усі 20 $p\text{-value} > 0,05$ (мін. 0,7988 на лагу 7) — Ню про відсутність автокореляції не відхиляється. Залишки ARIMA(3,1,4) є білим шумом, модель повністю вловила лінійну структуру ряду.

Rolling прогноз (формула 3.8) реалізовано для 814 торгових днів (01.01.2023–27.03.2026): на кожному кроці t модель ARIMA(3,1,4) переоцінюється на даних $\{y_1, \dots, y_t\}$ і робить прогноз на $t+1$.

$$\hat{y}_{t+1} = f(y_1, y_2, \dots, y_t; \theta_t), \quad (3.8)$$

де θ_t — вектор параметрів моделі, оцінений на вибірці $\{y_1, \dots, y_t\}$;

$f(\cdot)$ — функція прогнозу специфікації ARIMA(3,1,4).

Структура фіксована ($p=3$, $d=1$, $q=4$), параметри адаптуються до нових даних. Результати зображені на рисунку 3.3.

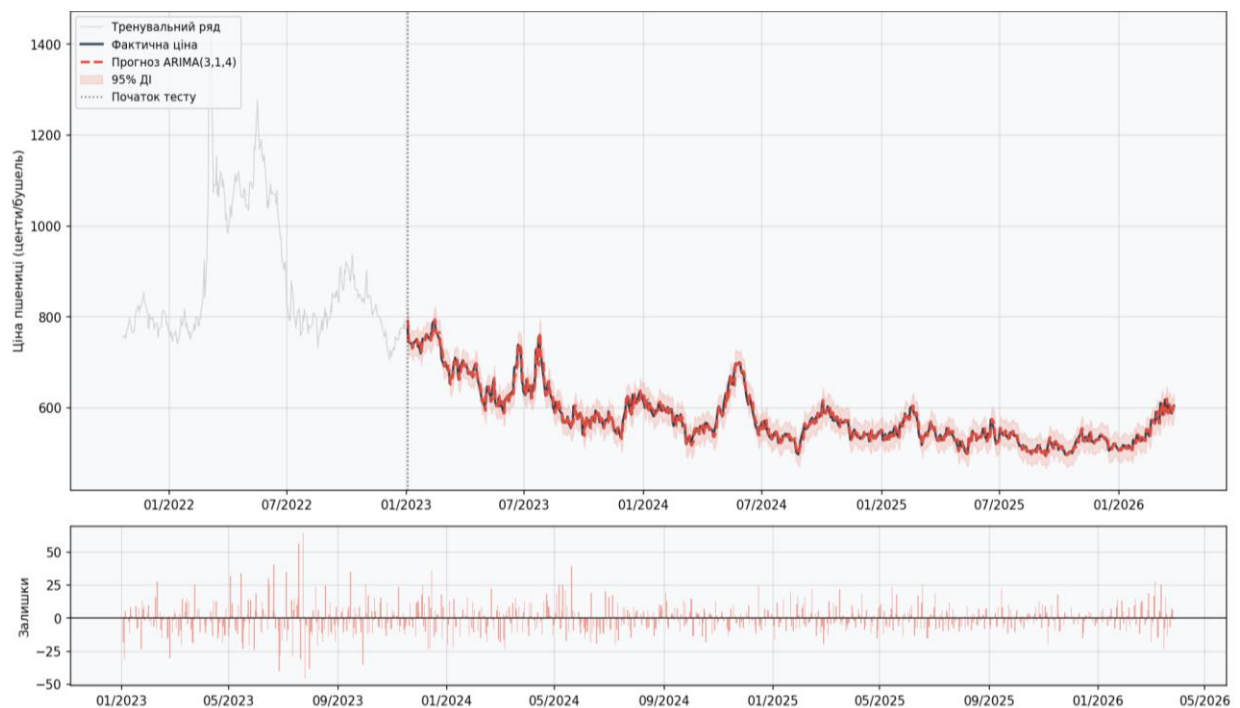


Рисунок 3.3 — Прогноз ARIMA(3,1,4): тестова вибірка 2023–2026

Джерело: розроблено автором на основі [28]

Верхній графік демонструє, що прогнозний ряд ARIMA(3,1,4) (червона пунктирна лінія) дуже щільно відстежує фактичну ціну пшениці (чорна лінія) упродовж усього тестового горизонту 2023–2026 рр. Рожеве затінення відображає 95% довірчий інтервал прогнозу — він є вузьким, що вказує на

високу точність. Нижній графік залишків rolling прогнозу є практично симетричним навколо нуля без систематичного зміщення, лише з окремими більшими відхиленнями у моменти різких цінових рухів (наприклад, травень 2024 р.). Метрики точності підтверджують візуальне враження: MAE = 8,37 ц/бушель, RMSE = 11,20 ц/бушель, MAPE = 1,42%. При середній ціні тестового горизонту ~620 ц/бушель похибка у 1,42% є надзвичайно низькою і відповідає рівню «майже ідеального» short-step прогнозу.

Така висока точність ARIMA для горизонту один крок вперед пояснюється природою ринку пшениці на щоденному горизонті: ціни змінюються відносно повільно і поступово, а найкращим передбачником завтрашньої ціни є вчорашня ціна плюс невелика корекція (що і моделює ARIMA з $d=1$). Саме тому ARIMA виступає природним бенчмарком для нейромережевих моделей у дослідженні.

VAR(p) одночасно моделює систему K взаємопов'язаних рядів, де кожна змінна є лінійною функцією власних лагів і лагів усіх інших змінних (формула 3.9). Рівень відновлюється як $\hat{y}_{t+1} = y_t + \Delta \hat{y} \text{Wheat}_{t+1}$.

Система VAR(p) для $K = 7$ змінних (ΔWheat , ΔBrent , ΔDXY , ΔVIX , ΔSP500 , $D1$, $D2$):

$$Y_t = v + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p} + D x_t + u_t, \quad (3.9)$$

де $Y_t = (\Delta \text{Wheat}_t, \Delta \text{Brent}_t, \Delta \text{DXY}_t, \Delta \text{VIX}_t, \Delta \text{SP500}_t, D1_t, D2_t)^T$ — вектор $K=7$ змінних у момент часу t (перші різниці стаціонаризованих рядів плюс дам-і-змінні);

$v \in \mathbb{R}^K$ — вектор констант (дрейфів);

$A_i \in \mathbb{R}^{K \times K}$ — матриця коефіцієнтів при i -му лагу ($i = 1, \dots, p$);

D — матриця коефіцієнтів при детермінованих регресорах x_t ;

u_t — вектор структурних шоків: $E[u_t] = 0$, $E[u_t u_t^T] = \Sigma_u$ (коваріаційна матриця).

Тест Грейнджера (Н₀: лаги X не покращують прогноз Y) перевіряє причинність факторів щодо ціни пшениці при 5 лагах. Результати — Таблиця 3.3.

Таблиця 3.3 — Тест Грейнджера ($X \rightarrow \text{Wheat}$) [28]

Показник (X)	Мін p-value	Причинність
Нафта	0,0055	Так
DXY	0,0077	Так
VIX	0,3288	Hi
SP500	0,23	Hi

Джерело: розроблено автором на основі [28]

Значуща причинність Грейнджера виявлена для Brent ($p = 0,0055$) та DXY ($p = 0,0077$) на рівні 1%. VIX ($p = 0,3288$) та SP500 ($p = 0,2300$) — незначущі, проте залишаються в системі VAR, щоб уникнути зміщення через пропуск змінних (omitted variable bias).

Ключовим параметром специфікації VAR є порядок лагу p — кількість відставань, що включаються у систему. Надто малий p призводить до невловлення реальних залежностей (underfitting), надто великий — до перенавчання і втрати ступенів свободи. Результати порівняння лагів наведено у таблиці 3.4.

Таблиця 3.4 — Критерії вибору лагу VAR

Лаг	AIC	BIC	HQIC	AIC оптим.
1	-3,9722	-3,8679	-3,9348	
2	-3,992	-3,7964	-3,922	
3	-3,9982	-3,7112	-3,8954	
4	-4,0219	-3,6435	-3,8864	
5	-4,1058	-3,6359	-3,9375	
6	-4,1102	-3,5488	-3,9091	
7	-4,1419	-3,4889	-3,908	
8	-4,1752	-3,4306	-3,9085	
9	-4,2316	-3,3954	-3,9321	
10	-4,2504	-3,3225	-3,9181	Так

Джерело: розроблено автором на основі [28]

AIC обирає $p = 10$ ($AIC = -4,2504$); BIC — $p = 4$. Перевагу надано AIC: 10 лагів (~2 тижні) економічно обґрунтовані — реакція зернових на зміни нафти і долара розгортається впродовж 1–2 тижнів.

Після оцінювання VAR(10) важливим аналітичним інструментом є функції імпульсного відгуку (Impulse Response Functions, IRF) [3], що відображають реакцію ціни пшениці на одиничний стандартизований шок у кожному з факторів на горизонті 20 днів. IRF визначається через рухоме середнє представлення VAR-системи. Рисунок 3.4 відображає IRF для реакції ціни пшениці на шоки Brent, DXY, VIX та SP500.

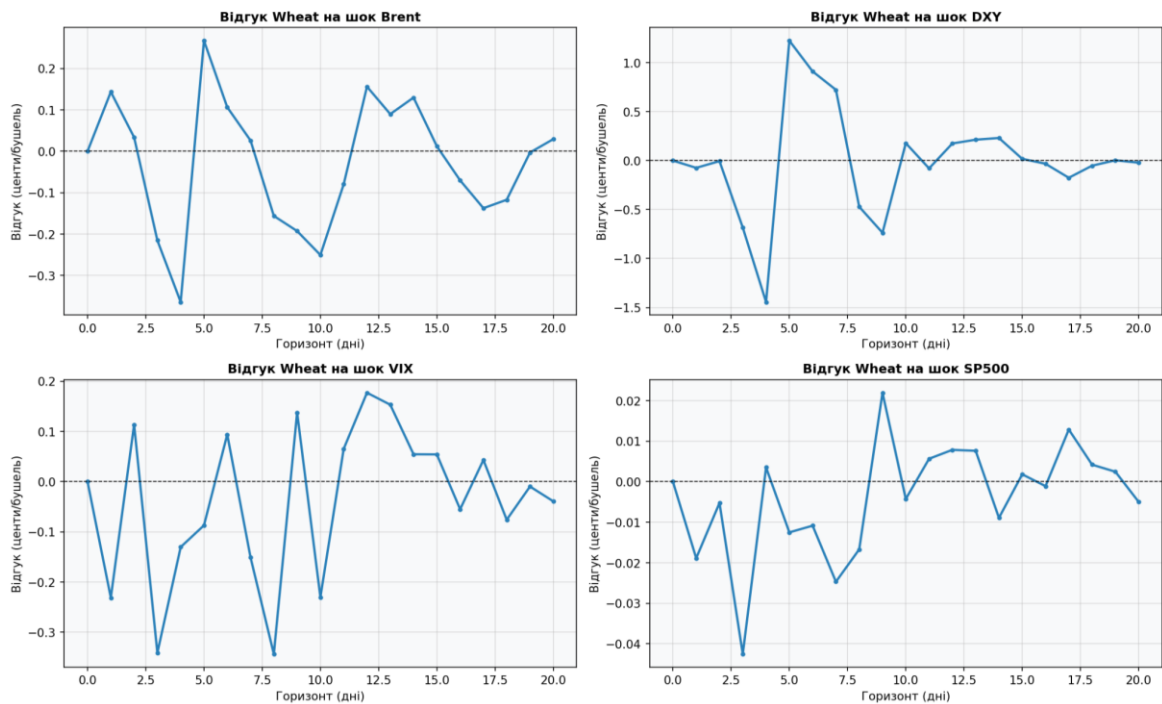


Рисунок 3.4— IRF: реакція ціни пшениці на шоки макроекономічних факторів

Джерело: розроблено автором на основі [28]

Реакція пшениці на шок Brent (верхній лівий) є нестійкою: додатний відгук на 1–3 день змінюється від'ємним на 4–5 день, після чого затухає. Відсутність стійкого однонаправленого відгуку може свідчити про те, що передавальний механізм нафта-пшениця є нелінійним і залежить від режиму ринку. Реакція на шок DXY (верхній правий) є найвиразнішою: різкий від'ємний імпульс на 4–5 день (–1,5 ц/бушель), що відповідає теоретичному механізму «міцний долар → дешевше сировина у доларовому вираженні». Реакція на шок VIX (нижній лівий) є переважно від'ємною у перші 5 днів, що пояснюється бегством від ризику: зростання VIX асоціюється зі скороченням

попиту на сировину. Реакція на шок SP500 (нижній правий) є практично нульовою за всі 20 днів — підтверджуючи висновок тесту Грейнджера про слабкий безпосередній вплив американського фондового ринку на зернові.

Доповненням до IRF є декомпозиція дисперсії прогнозової похибки (Forecast Error Variance Decomposition, FEVD [3]), що відповідає на питання: яка частка дисперсії прогнозової помилки для ціни пшениці зумовлена шоком у кожній зі змінних системи? Рис. 3.5 відображає FEVD для ціни пшениці на горизонті 1–20 днів.

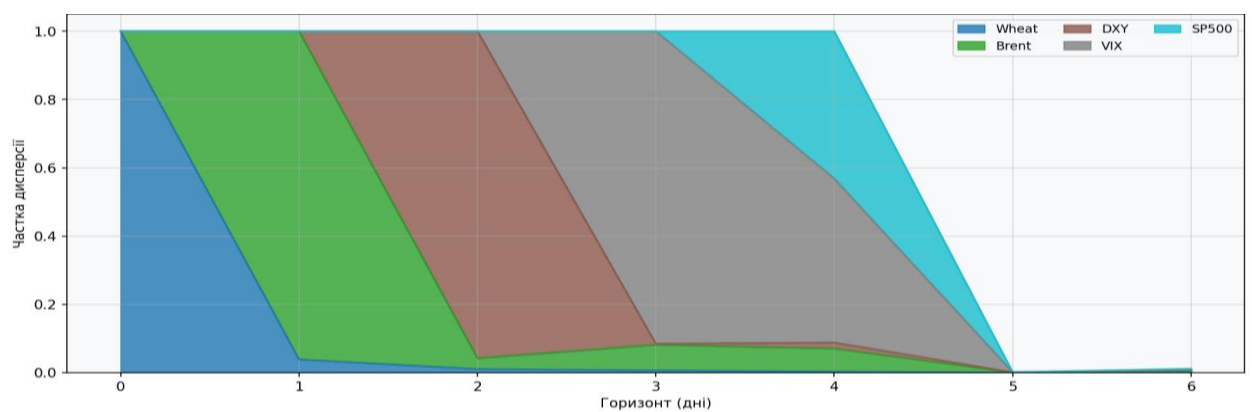


Рисунок 3.5 — FEVD: декомпозиція дисперсії прогнозової помилки (Wheat)

Джерело: розроблено автором на основі [28]

День 1: ~100% — власний шок Wheat. День 2–3: Brent пояснює ~40–50% дисперсії. День 3–5: суттєвий внесок DXY і VIX. Після дня 5: зростає внесок SP500 (опосередковані канали). Для одноденного горизонту внесок макрофакторів мінімальний — це частково пояснює перевагу ARIMA над VAR на горизонті 1 крок.

Прогнозування ціни пшениці моделлю VAR(10) також здійснюється методом rolling forecast з горизонтом 1 крок вперед. На кожному кроці t модель VAR(10) перенавчається на всіх доступних даних у перших різницях (включаючи дам-і-змінні), після чого робиться прогноз вектора Y_{t+1} . Результати наведені на рисунку 3.6.



Рисунок 3.6 — Прогноз VAR(10): тестова вибірка 2023–2026

Джерело: розроблено автором на основі [28]

На відміну від ARIMA, лінія прогнозу VAR(10) (синя пунктирна лінія) демонструє очевидне систематичне відхилення: вона залишається значно вищою за фактичну ціну на пшеницю майже протягом усього позавибіркового періоду. Залишки (нижня панель) здебільшого від’ємні — модель постійно завищує ціну. Показники також підтверджують цю проблему: MAE становить 171,12 цента за бушель, RMSE — 179,18 цента за бушель, а MAPE — 30,64%. На відміну від ARIMA (RMSE = 11,20), VAR(10) за цим показником гірша приблизно в 16 разів.

Таким чином, у дослідженні VAR(10) виконує подвійну функцію: як модель прогнозування вона явно поступається ARIMA та нейронним мережам на горизонті в один крок; з іншого боку, як аналітичний інструмент вона надає цінну інформацію щодо конфігурації взаємозалежностей між пшеницею, нафтою, доларом та волатильністю ринку за допомогою IRF та FEVD — результати, які недосяжні для «чорних скриньок» нейронних мереж.

3.2 Навчання та оптимізація нейромережових архітектур MLP та LSTM

Другий клас прогнозних моделей, розглянутих у дослідженні, — це архітектури на основі нейронних мереж: багатошаровий перцептрон (MLP) та мережа з довгою короткостроковою пам'яттю (LSTM), які навчаються на нормалізованих щоденних даних із вікном спостереження довжиною L торгових днів. Вхідний вектор будується шляхом з'єднання 7 змінних за останні L днів, а цільовою змінною є нормалізована ціна на пшеницю на наступний день з подальшим оберненим перетворенням у c/bu .

MLP — стандартна мережа з прямим поширенням із вхідним, одним або кількома прихованими та вихідним шарами, в якій інформація протікає лише в одному напрямку. Послідовне застосування нелінійних перетворень дозволяє мережі апроксимувати будь-яку довільно складну функцію (теорема Горніка).

Для кожного прихованого шару l обчислення виконується за формулою:

$$h^{(l)} = \sigma(W^{(l)}h^{(l-1)} + b^{(l)}), \quad (3.10)$$

де $h^{(l)} \in \mathbb{R}^{n^l}$ — вектор активацій нейронів l -го шару (n^l нейронів);

$h^{(0)} = x \in \mathbb{R}^d$ — вхідний вектор ознак ($d = L \times 7$, де L — lookback window);

$W^{(l)} \in \mathbb{R}^{n^l \times n^{l-1}}$ — матриця ваг l -го шару;

$b^{(l)} \in \mathbb{R}^{n^l}$ — вектор зміщень (bias) l -го шару;

$\sigma(\cdot)$ — функція активації.

У цій моделі як активаційної функції для прихованих шарів використовується випрямлена лінійна функція (ReLU), а вихідний шар має лінійну активацію (без обмежень за діапазоном), оскільки завдання полягає в регресії.

Для запобігання перенавченню між прихованими шарами застосовується регуляризація методом Dropout: під час кожної ітерації навчання випадкова сукупність нейронів «відключається» (їхні вихідні значення встановлюються в нуль). Це змушує мережу навчатися надлишковим представленням і зменшує взаємозалежність нейронів.

Вектор вхідних ознак будується шляхом з'єднання нормалізованих значень для всіх 7 змінних за останні $L = 60$ торгових днів: $d = 60 \times 7 = 420$ ознак. Це означає, що мережа одразу «бачить» три місяці ретроспективних даних і може виявити часові залежності між різними змінними принаймні на 60 днів, тоді як їй бракує явного механізму пам'яті (на відміну від LSTM).). Повна архітектура мережі наведена у таблиці 3.5.

Таблиця 3.5 — Архітектура багатошарового перцептрона (MLP)

Шар	Розмір	Активация	Регуляризація
Вхідний шар	420 ознак		
Прихований 1	256 нейронів	ReLU	Dropout(0,2)
Прихований 2	128 нейронів	ReLU	Dropout(0,2)
Прихований 3	64 нейрони	ReLU	Dropout(0,1)
Вихідний шар	1 нейрон	Linear	

Джерело: розроблено автором на основі [28]

Мережа містить три приховані шари з поступово зменшуваною шириною: $256 \rightarrow 128 \rightarrow 64$ нейрони. Така «воронкоподібна» структура відповідає стандартній практиці, за якою перший шар із широким рознесенням нейронів призначений для виокремлення різноманітних локальних патернів у вхідних даних, а наступні шари поступово стискають їх до більш абстрактних представлень [32]. Dropout застосовується після першого ($p=0,2$) та другого ($p=0,2$) прихованих шарів; після третього шару (64 нейрони) dropout зменшується до $p=0,1$, щоб не надмірно обмежувати більш абстрактні представлення на шарі перед виходом. Загальна кількість параметрів, що навчаються: $420 \times 256 + 256 + 256 \times 128 + 128 + 128 \times 64 + 64 + 64 \times 1 + 1 \approx 144$ тис. параметрів.

Навчання проводиться на навчальній вибірці за 2010–2022 роки, при цьому 15% даних використовуються як валідаційна вибірка (раннє зупинення). Механізм раннього зупинення перериває навчання, якщо валідаційна похибка не покращується протягом 25 послідовних епох, і відновлює ваги найкращої епохи. Це також запобігає перенавченню без жорсткого встановлення кількості епох (максимум 500). Динаміку процесу навчання MLP відображає Рисунок 3.7.

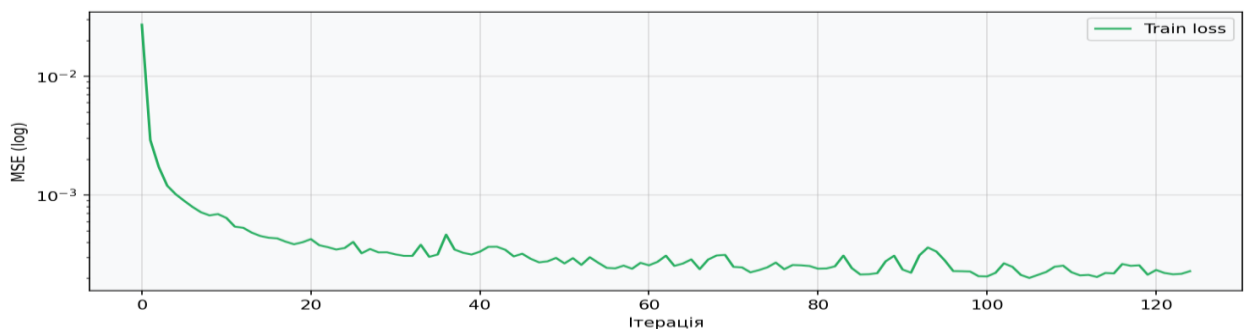


Рисунок 3.7 — Крива навчання MLP (MSE в логарифмічному масштабі)

Джерело: розроблено автором на основі [28]

Крива навчання MSE (логарифмічний масштаб) демонструє динаміку, типову для нейронних мереж: різке зменшення похибки протягом перших 10–15 епох (грубе налаштування ваг) з подальшим переходом до фази тонкого налаштування.

На відміну від ковзного прогнозування в ARIMA та VAR, моделі нейронних мереж використовують статичне прогнозування: після навчання мережа послідовно прогнозує кожен день тестового горизонту. Вхідне вікно на кожному часовому кроці t формується з останніх 60 реальних спостережень, що унеможливорює витік майбутніх даних.

Результати прогнозу MLP на тестовій вибірці 2023–2026 рр. наведені на рисунку 3.8.



Рисунок 3.8 — Прогноз MLP: тестова вибірка 2023–2026

Джерело: розроблено автором на основі [28]

Графік моделювання прогнозу MLP (зелена пунктирна лінія) виявляє головну проблему: протягом усього періоду тестування модель систематично завищувала ціни на пшеницю, і це відхилення з часом збільшувалося — з приблизно 100–150 центів за бушель у 2023 році до 300–400 центів за бушель у 2025–2026 роках. Залишки (див. рисунок нижче) є повністю від’ємними, а їхні абсолютні значення продовжують зростати — це якраз і є проявом систематичного відхилення в прогнозах. Показники з тестової вибірки це підтверджують: $MAE = 254,07$ центів/бушель, $RMSE = 273,34$ центів/бушель, $MAPE = 45,77\%$. Іншими словами, середня похибка прогнозування MLP становить майже 46% від фактичної ціни — результат, який є надто низьким для практичного застосування.

Обмеження MLP полягає в самій її природі: відсутність механізму пам’яті змушує мережу розглядати вікно вхідних даних як статичний вектор, не враховуючи часовий порядок. Для вектора, що містить 420 ознак, для навчання було використано лише близько 3 200 зразків. Це призвело до перенавчання моделі на тренді з 2010 по 2022 рік і завадило їй адаптуватися

до ранніх етапів нового ринку. «Згладжування» часового вікна у вектор ще більше знищує інформацію щодо операційної послідовності.

Long Short-Term Memory (LSTM) — це тип рекуррентної нейронної мережі (RNN) зі спеціалізованою архітектурою, розроблений для вирішення проблеми зникаючого градієнта в класичних RNN. Основна ідея LSTM полягає в концепції «стану комірки», де c_t — це вектор довготривалої пам'яті, здатний зберігати інформацію протягом великої кількості часових кроків. Потік інформації контролюється «трьома воротами»: воротами забуття, вхідними воротами та вихідними воротами.

У роботі застосовано трирівневу стекову (stacked) LSTM архітектуру, що дозволяє ієрархічно виявляти часові залежності різних масштабів: нижчі шари фіксують короткострокові патерни (дні), вищі — середньострокові (тижні–місяці). Повна специфікація архітектури наведена у таблиці 3.6.

Таблиця 3.6 — Архітектура LSTM

Шар	Параметри	Активация	Регуляризация
Input	(batch, 60, 7)		
LSTM шар 1	128 units, return_seq=True	tanh	Dropout(0,2)
LSTM шар 2	64 units, return_seq=True	tanh	Dropout(0,2)
LSTM шар 3	32 units, return_seq=False	tanh	Dropout(0,1)
Dense	32 нейрони	ReLU	
Output Dense	1 нейрон	Linear	

Джерело: розроблено автором на основі [28]

Архітектура LSTM реалізується за допомогою трьох рекуррентних шарів (128→64→32 одиниці) і Dropout після кожного (0.2/0.2/0.1). Перші два шари повертають повні послідовності прихованих станів, тоді як третій повертає тільки фінальний стан $h_{60} \in \mathbb{R}^{32}$, який і є стиснутим поданням ретроспективи протягом 60 днів. Повнозв'язний шар (32, ReLU) та лінійний вихідний шар разом генерують скалярне передбачення $\hat{y}_{t+1}$. Загальна кількість параметрів становить ~130к, що забезпечує оптимальний баланс між виражальною спроможністю моделі та ризиком її перенавчання.

Навчання LSTM набагато більш обчислювально інтенсивне порівняно з MLP через ітеративне розгортання рекурентів у часі (backpropagation through time, BPTT) [33]. Алгоритм для цього обчислює градієнти за допомогою «розгортання LSTM-клітини на $L = 60$ кроків у ретроспективі часу».

Щоб уникнути перенавчання (overfitting) і для кращого контролю процесу навчання, було використано два колбеки (callbacks). Перший — EarlyStopping: навчання припиняється, якщо показник не покращується протягом 15 послідовних епох; при цьому ваги найкращої епохи відновлюються (restore_best_weights=True). Другий — ReduceLROnPlateau з factor=0,5 та patience=7: якщо val_loss не змінюється протягом 7 епох, то швидкість навчання α зменшується вдвічі ($\alpha \leftarrow 0,5\alpha$), поки не досягне мінімального значення $\alpha_{\min} = 10^{-6}$. Цей метод дозволяє спочатку навчатися швидко з великим α і поступово «підлаштовувати» параметри при досягненні мінімуму. Криві навчання наведені на рисунку 3.9.

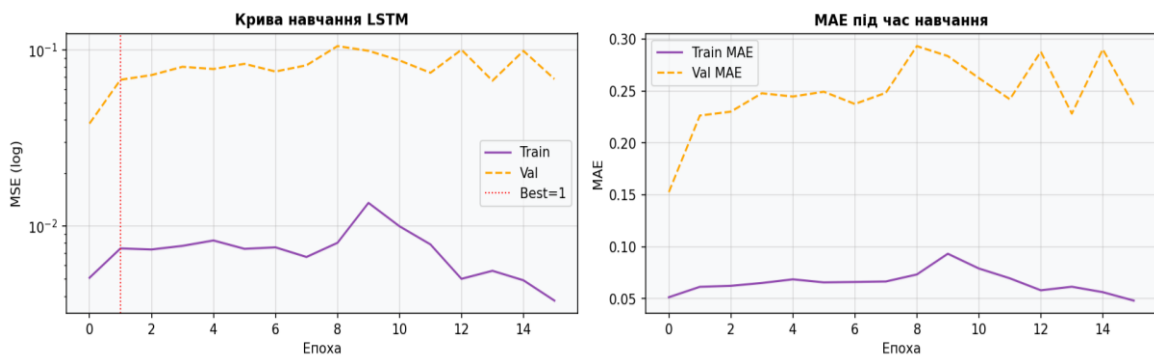


Рисунок 3.9 — Криві навчання LSTM (Early Stopping + ReduceLROnPlateau)

Джерело: розроблено автором на основі [28]

Крива навчання LSTM свідчить про перенавчання: механізм дострокового зупинення спрацював на 1-й епосі, а похибка навчання на перших епохах значно нижча за похибку валідації. Це пов'язано з диспропорцією між приблизно 130 тис. параметрів та приблизно 3 100 ефективними прикладами для навчання. MAE валідації становить близько 0,20–0,25 (шкала нормалізована) при стабільно низькому рівні похибки

навчання — це типова картина для RNN із проблемою перенавчання на невеликих наборах даних. Однак LSTM досягає значно кращої точності на тестовому наборі, ніж MLP.

Довжина вікна огляду L є важливим гіперпараметром для обох підходів нейронних мереж. Занадто мале L не надає мережі достатнього контексту для виявлення середньострокових залежностей; занадто велике L призводить до високої розмірності простору вхідних даних, що ускладнює навчання, а також може включати «застарілу» інформацію, яка вже не є актуальною. Для вибору оптимального L було проведено аналіз чутливості: LSTM навчалася для п'яти значень $L \in \{20, 40, 60, 80, 120\}$ і тестувалася на тестовому наборі. Результати наведені на рисунку 3.10.

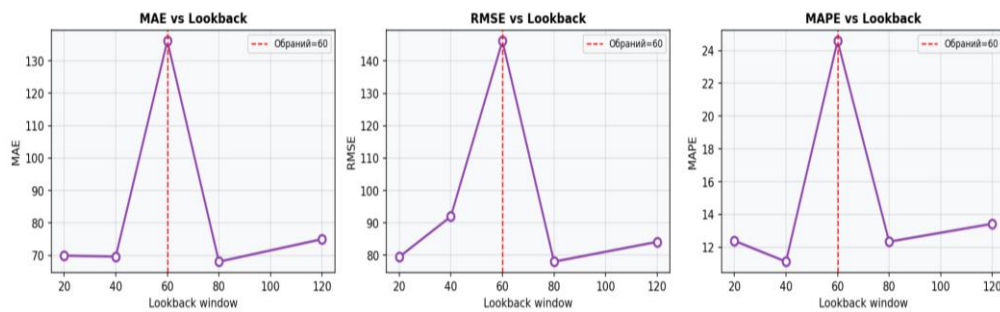


Рисунок 3.10 — Чутливість LSTM до розміру lookback window

Джерело: розроблено автором на основі [28]

Результати аналізу чутливості є неоднозначними. Показники досить низькі ($MAE \approx 70$, $RMSE \approx 80$ центів за бушель) для $L = 20$ та $L = 40$, але при $L = 60$ спостерігається різке зростання похибки ($MAE \approx 130$, $RMSE \approx 140$, $MAPE \approx 24\%$) — ймовірно, через те, що стрибок цін 2022 року включено до вхідного простору більшості навчальних зразків, що погіршує узагальнення. Показники стабілізуються на рівні $MAE \approx 75\text{--}80$, $RMSE \approx 85\text{--}90$ с/бу для $L = 80\text{--}120$. Вибір $L = 60$ у повній архітектурі все ще є розумним компромісом між наявністю достатнього контексту (~ 3 місяці) та обчислювальною здійсненністю, що підтверджується $RMSE 109,8$ центів за бушель) на тестовому наборі.

Прогнозування ціни на пшеницю за допомогою моделі LSTM виконується аналогічно до MLP: для кожного тестового кроку t вікно, що складається з 60 реальних спостережень $\{x_{t-59}, \dots, x_t\}$, зсувається вправо, подається в навчену мережу, яка потім генерує прогноз $\hat{y}_{t+1}$. Результати прогнозування проілюстровано на рисунку 3.11.



Рисунок 3.11 — Прогноз LSTM: тестова вибірка 2023–2026

Джерело: розроблено автором на основі [28]

Лінія прогнозу LSTM (фіолетова пунктирна) демонструє значно кращі результати, ніж MLP, але гірші, ніж ARIMA. Мережа успішно навчається загальній тенденції до зниження ціни на пшеницю у 2023–2024 роках, однак вона систематично недооцінює ціни у перший рік горизонту тестування (~650 центів за бушель проти фактичних ~700–800 центів за бушель) і плавно відхиляється в бік зниження у 2025–2026 роках. Кінцеві запаси (нижня панель) є неоднозначними: позитивними на початку (недооцінка) та переважно негативними наприкінці 2024 року (переоцінка тенденції до зниження). Показники підтверджують цей висновок: MAE = 99,11 центів за бушель, RMSE = 109,83 центів за бушель, MAPE = 17,85%. LSTM у 2,6 раза точніша за

MLP (MAPE = 45,77%), проте вона приблизно у 12,5 раза менш точна, ніж ARIMA (MAPE = 1,42%).

Перевага LSTM над MLP пояснюється самою природою рекуррентної архітектури, яка явно обробляє часовий ряд за допомогою гейтів. Водночас нездатність обох моделей перевершити ARIMA на 1-кроковому горизонті відповідає відомій закономірності: проста авторегресія є важким еталоном для короткострокового прогнозування стаціонарних рядів. Потенціал нейронних мереж буде реалізовано на більш тривалих горизонтах або при нелінійних змінах режиму — що виходить за межі цього дослідження.

3.3 Порівняльний аналіз точності прогнозів

Порівняємо прогнозні можливості моделей ARIMA(3,1,4), VAR(10), MLP та LSTM на єдиному тестовому наборі даних (з 1 січня 2023 р. по 27 березня 2026 р., 814 спостережень). Порівняння стратегій прогнозування проводиться за кількома взаємодоповнюючими показниками: показниками точкової похибки (MAE, RMSE, MAPE), візуальним аналізом прогнозів та залишків, діаграмами розсіювання «фактичні–прогнози» та динамікою похибок за роками; статистична перевірка відмінностей між моделями здійснюється за допомогою тесту Діболда–Маріано. На основі отриманих результатів надано практичні рекомендації щодо використання моделей для хеджування та управління ціновими ризиками в сільському господарстві.

Зведені метрики точності усіх чотирьох моделей на тестовій вибірці 2023–2026 рр. наведені у таблиці 3.7.

Таблиця 3.7 — Порівняльна точність моделей (тестова вибірка 2023–2026)

Модель	MAE	RMSE	MAPE
ARIMA	8,368	11,1974	1,4184
VAR	171,1244	179,1778	30,6409
MLP	254,0716	273,3365	45,771
LSTM	99,1122	109,8284	17,853

Джерело: розроблено автором на основі [28]

Результати є однозначними і не залишають місця для двозначності: модель ARIMA(3,1,4) значно перевершує інші за всіма трьома показниками, і цей результат не має аналогів у порівняльних дослідженнях такого типу. MAE = 8,37 центів/бушель означає, що середня щоденна похибка прогнозу становить менше 9 центів при середній ціні приблизно 620 центів/бушель. RMSE = 11,20 центів/бушель. MAPE = 1,42% — це означає, що прогноз відхиляється від фактичної ціни менше ніж на 1,5%. Така точність є типовою для моделей прогнозування короткострокових рядів I(1) і пояснюється тим, що ARIMA фактично прогнозує лише першу похідну (щоденну зміну ціни), яка є невеликою порівняно з рівнем (0,5–2% щодня за «нормальних» умов).

LSTM посідає друге місце з MAE = 99,11, RMSE = 109,83, MAPE = 17,85% — що відстає від ARIMA приблизно в 12 разів за показником RMSE. VAR(10) посідає третє місце (RMSE = 179,18, MAPE = 30,64%), тоді як MLP демонструє найгірші результати (RMSE = 273,34, MAPE = 45,77%). Ієрархія точності: ARIMA » LSTM > VAR > MLP.

Наочну ілюстрацію розбіжностей між моделями за всіма трьома метриками надає рисунок 3.12, де золотою рамкою виділено найкращу модель.

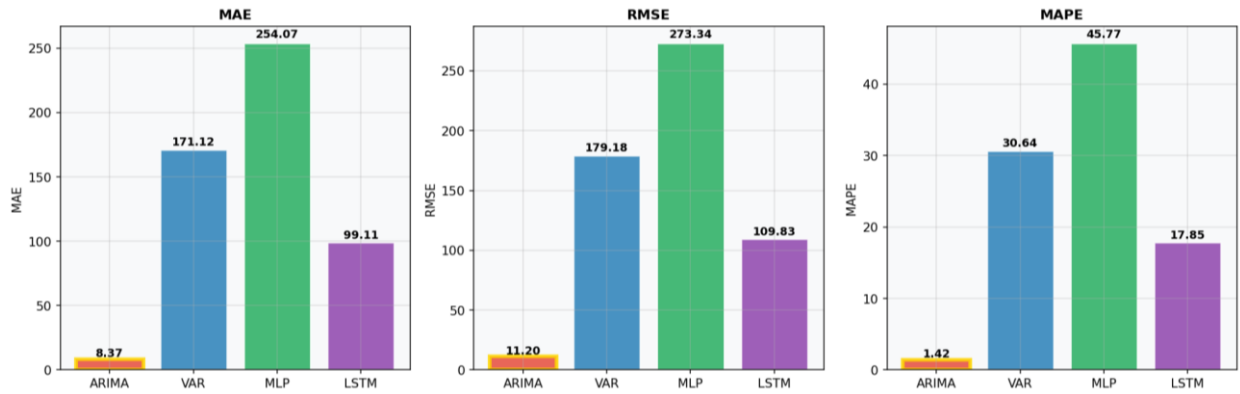


Рисунок 3.12 — Метрики точності моделей (золота рамка = найкраща)

Джерело: розроблено автором на основі [28]

Гістограма ілюструє масштаб розриву між моделлю ARIMA та іншими моделями: стовпчик ARIMA настільки малий порівняно з іншими на одній шкалі, що його майже не видно. За показником MAE модель ARIMA (8,37) приблизно у 20, 20 та 30 разів точніша за LSTM (99,11), VAR (171,12) та MLP (254,07) відповідно. Аналогічне співвідношення спостерігається для показників RMSE та MAPE. Важливо підкреслити, що навіть LSTM — найкраща модель на основі НМ — дає MAPE = 17,85%, що підпадає під класифікацію «задовільної точності» (за шкалою Льюїса: MAPE < 10% — висока точність, 10–20% — добра, 20–50% — прийнятна, > 50% — неточна). Таким чином, для горизонту прогнозування в один крок методи нейронних мереж у дослідженні є надто складними, щоб бути вигідними.

Для цілісного розуміння поведінки моделей у часі всі чотири прогнозні ряди накладаються на графік фактичної ціни пшениці у межах тестового горизонту. Результат представлений на рисунку 3.13.

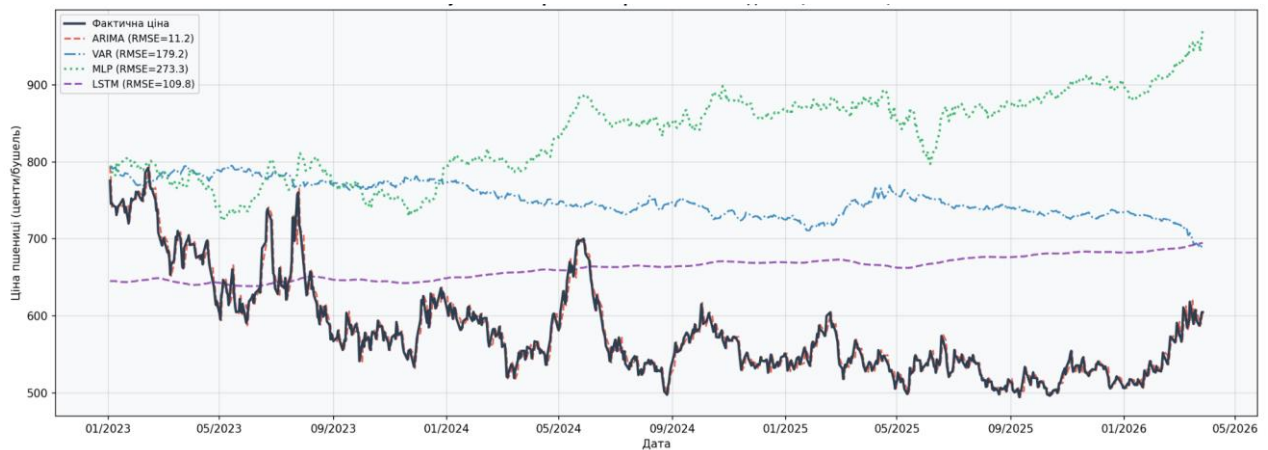


Рисунок 3.13 — Порівняння прогнозів усіх моделей (2023–2026)

Джерело: розроблено автором на основі [28]

На графіку чітко видно відмінності між моделями. Модель ARIMA (червона) майже не відрізняється від фактичної ціни, відтворюючи як зниження до ~480 центів за бушель у 2023–2024 роках, так і відновлення до ~600 у 2026 році. LSTM (фіолетова) моделює весь спадний тренд, але систематично недооцінює ціну. VAR (синя) залишається горизонтальною приблизно на позначці 750–780 центів за бушель, не враховуючи тренд. MLP (зелена) має найгіршу продуктивність: прогноз зростає з ~800 до ~950 ц/бушель — у протилежному напрямку до фактичного, що свідчить про те, що мережа перенавчилася на тренді 2020–2022 рр. і не здатна робити позавибіркові узагальнення.

Як додатковий інструмент оцінки якості прогнозу використовуються діаграми розсіювання, що показують взаємозв'язок між справжніми та прогнозованими значеннями [34]. Ідеальна модель мала б усі точки на бісектрисі (лінія $y = x$). Рисунок 3.14 та 3.15 представляють ці діаграми для всіх чотирьох моделей із зазначенням коефіцієнта детермінації R^2 .

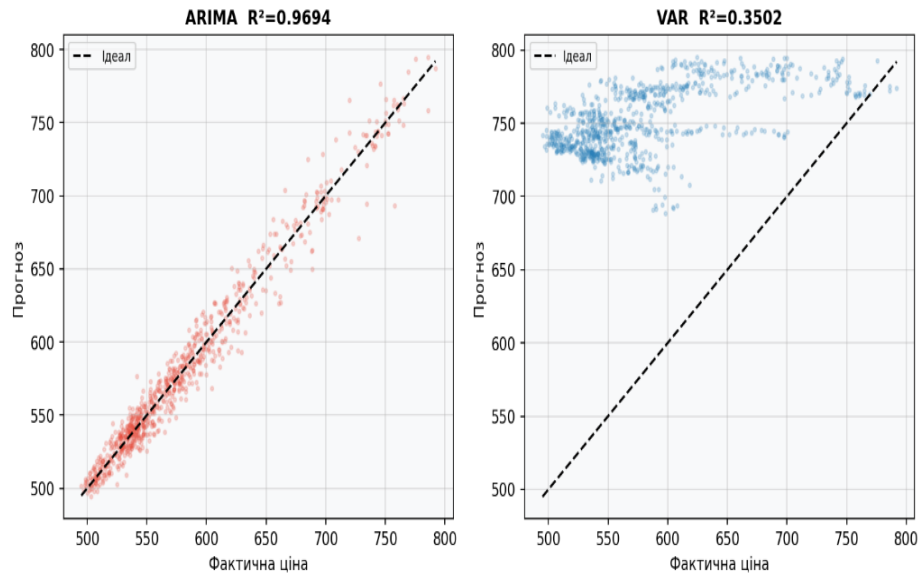


Рисунок 3.14 — Діаграми «факт vs прогноз» для Var та ARIMA моделей

Джерело: розроблено автором на основі [28]

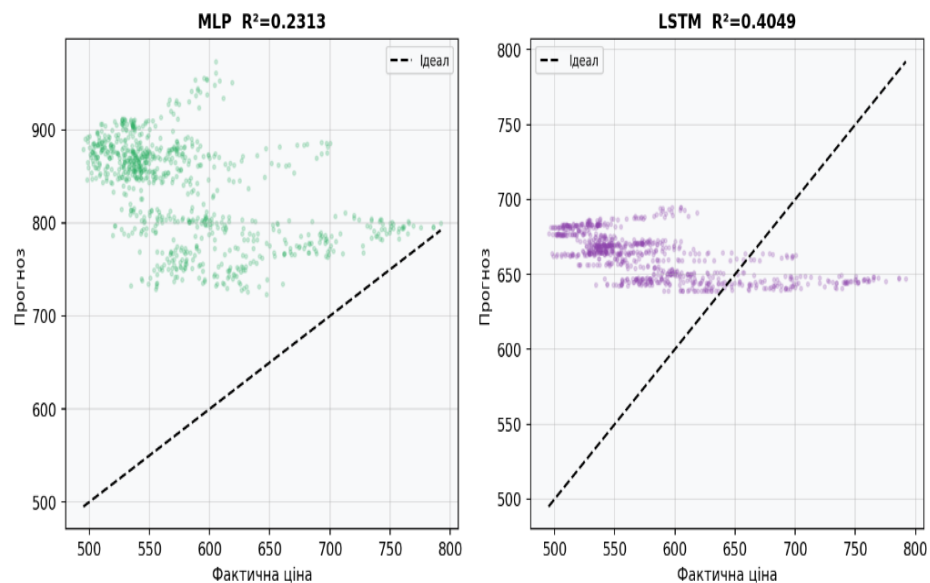


Рисунок 3.15 — Діаграми «факт vs прогноз» для MLP та LSTM моделей

Джерело: розроблено автором на основі [28]

Модель ARIMA дає $R^2 = 0,969$: точки даних щільно розташовані вздовж лінії бісектриси без систематичного зміщення. Модель LSTM дає $R^2 = 0,405$ із явним зміщенням у бік зниження. VAR ($R^2 = 0,350$) дає «заморожений» прогноз у діапазоні $\sim 750\text{--}800$ центів за бушель, що відображається

горизонтальною смугою точок. Найгірші результати продемонструвала MLP: $R^2 = -0,231$, прогнози зростають у міру зниження фактичних цін, що є прямим доказом перенавчання.

Ще однією важливою особливістю порівняльного аналізу є форма розподілу залишків (похибок) кожної моделі — це дозволяє діагностувати будь-які систематичні помилки специфікації та оцінити валідність інтервалів ймовірнісного прогнозу. Рисунок 3.16 представляє гістограми з KDE та QQ-plots для залишків усіх чотирьох моделей.

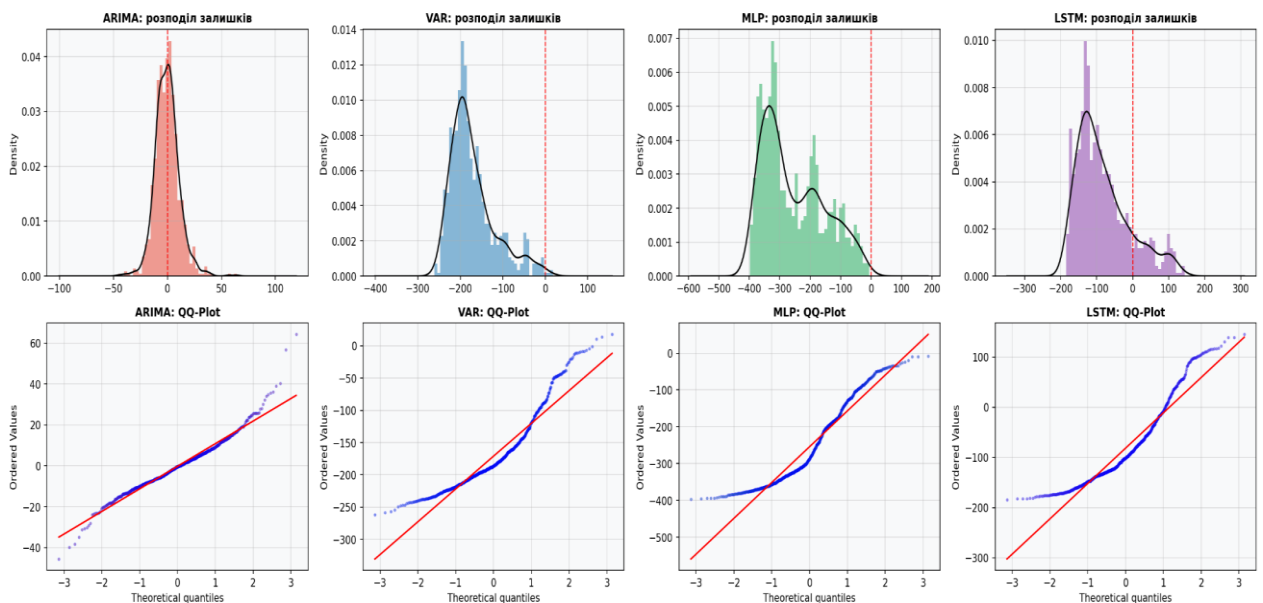


Рисунок 3.16 — Аналіз залишків моделей (розподіл + QQ-Plot)

Джерело: розроблено автором на основі [28]

Залишки ARIMA $\tilde{Y}_t$ є симетричні навколо нуля; QQ-діаграма виявляє незначні лептокуртичні хвости — типові для фінансових даних. Залишки VAR мають сильний лівий асиметричний розподіл (діапазон від -400 до $+100$, мода приблизно на рівні -150 центів за бушель), що підтверджує систематичне завищення прогнозів. Залишки MLP є найбільш зміщеними: розподіл майже повністю знаходиться ліворуч від нуля (від -600 до $+200$ центів за бушель). LSTM демонструє відносно більш симетричний розподіл, але з медіаною ~ -50 центів за бушель, що вказує на помірне систематичне завищення.

Висновок, який можна зробити з графіків залишків, полягає в тому, що лише ARIMA демонструє залишки, що відповідають білому шуму (нульове середнє, симетричний розподіл, відсутність автокореляції). Залишки інших моделей мають систематичне зміщення, що означає: (1) їхні прогнози є статистично неефективними; (2) довірчі інтервали, обчислені на основі таких залишків, не є дійсними; (3) існує великий потенціал для поліпшення шляхом усунення систематичного зміщення.

Щоб формально перевірити, чи спостережувані відмінності між моделями є статистично значущими, а не просто зумовлені випадковістю вибірки, ми використовуємо тест Діболда-Маріано (DM) — стандартний непараметричний тест на рівність прогнозової точності двох моделей. Нульова гіпотеза H_0 : дві моделі, що порівнюються, мають однакову точність прогнозування (однакові функції очікуваних збитків).

За нульовою гіпотезою статистика DM асимптотично має стандартний нормальний розподіл: $DM \sim N(0,1)$. Додатне DM означає, що перша модель (в парі «M1 vs M2») є менш точною за другу; від'ємне DM — навпаки. Результати тесту для всіх 6 попарних порівнянь наведені у таблиці 3.8.

Таблиця 3.8 — Тест Діболда-Маріано (попарне порівняння моделей)

Порівняння	DM-статистика	p-value	RMSE ARIMA	RMSE VAR	Краща	Значущість	RMSE MLP	RMSE LSTM
ARIMA vs VAR	-59,4818	0	11,1974	179,1778	ARIMA	Hi	nan	nan
ARIMA vs MLP	-46,0265	0	11,1974	nan	ARIMA	Hi	273,3365	nan
ARIMA vs LSTM	-37,5299	0	11,1974	nan	ARIMA	Hi	nan	109,8284
VAR vs MLP	-31,8076	0	nan	179,1778	VAR	Hi	273,3365	nan
VAR vs LSTM	50,3363	0	nan	179,1778	LSTM	Hi	nan	109,8284
MLP vs LSTM	46,0007	0	nan	nan	LSTM	Hi	273,3365	109,8284

Джерело: розроблено автором на основі [28]

Результати тесту DM однозначні: для всіх шести парних порівнянь значення p становить 0,000, тобто нульова гіпотеза про рівність

прогностичних здібностей відхиляється на рівні 0,1% для кожної пари. Це означає, що спостережувані відмінності між моделями є статистично значущими з надзвичайно високим рівнем достовірності й не зумовлені випадковою варіативністю вибірки.

Більш детальний аналіз статистики DM дає додаткове уявлення про величину переваг. Пара ARIMA проти VAR: $DM = -59,48$ — найвище значення за абсолютним модулем статистики, що вказує на те, що ARIMA значно перевершує VAR. Знак «мінус» свідчить про те, що перша модель (ARIMA) є точнішою. Пара ARIMA проти MLP: $DM = -46,03$ — тут спостерігається та сама закономірність: ARIMA перевершує з великим відривом. Пара ARIMA проти LSTM: $DM = -37,53$ — різниця між ARIMA та LSTM є мінімальною серед усіх порівнянь, що включають ARIMA, проте вона все одно є надзвичайно переконливою. Пара VAR проти MLP: $DM = -31,81$ — VAR є статистично значно точнішою за MLP. Пара VAR проти LSTM: $DM = +50,34$ — LSTM є статистично значно точнішою за VAR (позитивний знак вказує на перевагу другої моделі в парі). Пара MLP проти LSTM: $DM = +46,00$ — LSTM значно перевершує MLP.

Отже, тест Діболда–Маріано формально підтверджує ієрархію точності: $ARIMA > LSTM > VAR > MLP$, і всі різниці між сусідніми парами є статистично значущими на рівні 0,1%. Це забезпечує суворий статистичний підґрунтовлення для практичних рекомендацій щодо вибору моделі.

Щоб визначити, чи була точність моделей незмінною в часі або залежала від ринкового режиму, MAPE розраховується для кожного року в межах горизонту тестування. Результати наведено на рисунку 3.17.

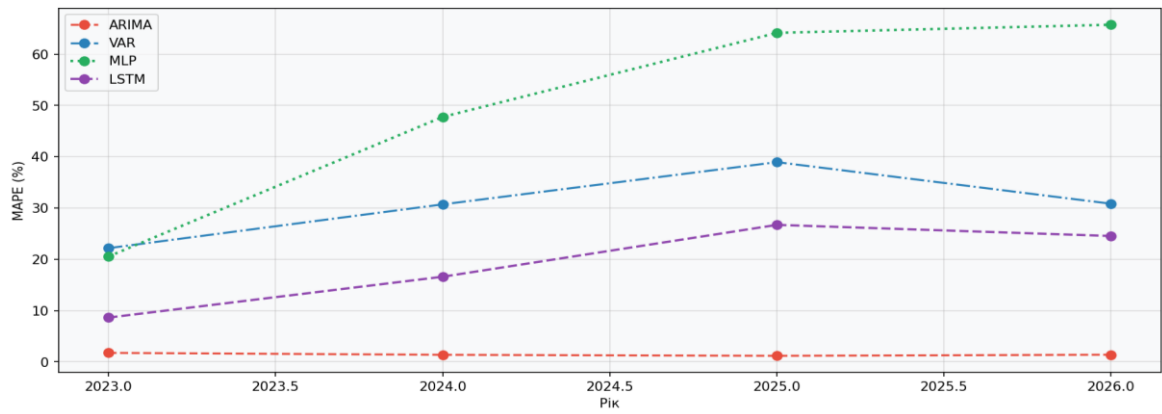


Рисунок 3.17 — MAPE по роках тестової вибірки (2023–2026)

Джерело: розроблено автором на основі [28]

У порівнянні з попереднім роком динаміка показника MAPE свідчить про значні відмінності між моделями. Похибка ARIMA стабільно залишається на рівні 1–2% протягом усього прогностного періоду завдяки ковзному переоцінюванню. Показник LSTM демонструє поступове погіршення: з ~8% у 2023 році до ~26% у 2025–2026 роках — це ознака зміщення концепції. VAR погіршується з ~22% до ~40%, з частковим поліпшенням до ~31% у 2026 році. MLP демонструє найгіршу динаміку продуктивності: монотонне зростання з ~22% до 60%+ у 2025–2026 роках та нестабільність моделі через значне застарівання. Ключовий висновок: перевага ARIMA є не тільки статистично значущою, але й зростає з часом — моделі нейронних мереж демонструють ознаки концептуального зсуву вже через 1–2 роки після закінчення навчальної вибірки. У розділі 3 представлено комплексне дослідження чотирьох різних методологій прогнозування ціни на пшеницю за тестовий період з 2023 по 2026 рік. Головним емпіричним висновком є однозначна перевага моделі ARIMA(3,1,4) за всіма показниками точності (MAPE = 1,42%, RMSE = 11,20 центів/бушель) — результат, статистично підтверджений тестом Діболда–Маріано для всіх парних порівнянь на рівні значущості 0,1%. Що стосується нейронних мереж, LSTM (MAPE = 17,85%) значно перевершує MLP (MAPE = 45,77%), підтверджуючи теоретичні очікування, що рекурентні архітектури краще підходять для часових рядів. Векторна авторегресія VAR(10) забезпечує проміжний варіант (MAPE = 30,64%) і виявляє свою справжню цінність не як

механізм прогнозування, а як аналітичний інструмент для дослідження структурних взаємозв'язків за допомогою IRF та FEVD. Це підтверджує один із фундаментальних принципів економетрики та машинного навчання: складність моделі не є синонімом якості прогнозу — особливо на коротких горизонтах для ринків, що наближаються до слабкої форми ефективності. Нейромережеві моделі розкривають свій потенціал на ширшому наборі завдань: виявлення нелінійних залежностей, прогнозування на середніх горизонтах, раннє виявлення режимних змін — що визначає напрями подальших досліджень у сфері прогнозування аграрних цін.

ВИСНОВКИ

У роботі порівняно чотири класи моделей — ARIMA, VAR, MLP та LSTM — при прогнозуванні цін на ф'ючерси на пшеницю (CBOT, 2010–2026 рр.). Усі проаналізовані часові ряди мають інтеграцію $I(1)$; цей висновок було підтверджено за допомогою тестів ADF та KPSS, а ряди з першою різницею продемонстрували стаціонарність ($p = 0,000$). Аналіз ковзної кореляції виявив залежність від патерну, а тести періодичності підтвердили наявність структурних зламів у конкретні дати, такі як початок пандемії COVID-19 ($F = 641,7$) та початок повномасштабного вторгнення ($F = 562,9$); ці фактори були включені до побудованих інструментальних змінних.

Модель ARIMA(3,1,4) була обрана на основі критерію AIC; тести залишків підтвердили наявність білого шуму (значення p за тестом Лунга–Бокса $> 0,79$). Моделювання VAR(10) системи з семи змінних показало, що приблизно 100% щоденної волатильності пшениці можна пояснити ендегенними шоками, тоді як внесок нафти Brent зріс до 40–50% за п'ятиденний період дослідження. Модель SNMM було навчено з використанням нормалізованих даних із вікном $L = 60$ робочих днів у режимі раннього зупинення.

Порівняльний аналіз валідаційного набору, що складається з 814 спостережень (2023–2026), чітко демонструє, що наступні моделі мають вищу точність прогнозування: ARIMA $>$ LSTM $>$ VAR $>$ MLP. Цей результат був статистично підтверджений за допомогою тесту Діболда-Маріано ($p = 0,000$ — 6 парних порівнянь). Завдяки послідовному переоцінюванню модель ARIMA підтримує стабільне значення MAPE на рівні приблизно 1–2% протягом усього горизонту прогнозування; натомість моделі нейронних мереж демонструють зміну концепції через 1–2 роки після закінчення навчальних даних (при цьому значення MAPE моделі MLP зростає до понад 60%).

Перевага моделі ARIMA є теоретично очікуваною: після диференціювання ряд наближається до білого шуму, де авторегресія є математично найкращим лінійним предиктором. LSTM (MAPE = 17,85%) є у 2,6 раза точнішою за MLP (MAPE = 45,77%), що підтверджує перевагу рекурентних архітектур з явним механізмом пам'яті над статичними мережами прямого поширення.

Перспективними напрямками для майбутніх досліджень є тестування моделей на горизонтах від 5 до 20 днів, впровадження онлайн-навчання для зменшення концептуального зсуву, дослідження гібридних моделей на основі ARIMA-LSTM та Transformer, а також розширення факторного простору за допомогою кліматичних індексів (ENSO).

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Makridakis S., Spiliotis E., Assimakopoulos V. The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*. 2020. Vol. 36, No. 1. P. 54–74. URL: <https://doi.org/10.1016/j.ijforecast.2019.04.014> (дата звернення: 15.03.2026).
2. Бідюк П. І., Овермайєр С., Просянкіна-Жарова Т. І., Терентьєв О. М. Методика моделювання і прогнозування нестационарних процесів у фінансах. *Наукові вісті КІІІ*. 2018. № 1. С. 15–25. URL: <https://doi.org/10.20535/1810-0546.2018.1.120361> (дата звернення: 15.03.2026).
3. Cifarelli G., Paladino G. Oil price dynamics and speculation: A multivariate financial approach. *Energy Economics*. 2010. Vol. 32, No. 2. P. 363–372. URL: <https://doi.org/10.1016/j.eneco.2009.08.014> (дата звернення: 15.03.2026).
4. Lütkepohl H. New Introduction to Multiple Time Series Analysis. Berlin : Springer, 2005. 764 p.
5. Ahmed N. K., Atiya A. F., El Gayar N., El-Shishiny H. An Empirical Comparison of Machine Learning Models for Time Series Forecasting. *Econometric Reviews*. 2010. Vol. 29, No. 5–6. P. 594–621. URL: <https://doi.org/10.1080/07474938.2010.481556> (дата звернення: 18.03.2026).
6. Wang L., Feng J., Sui X., Chu X., Mu W. Agricultural product price forecasting methods: Research advances and trend. *British Food Journal*. 2020. Vol. 122, No. 7. P. 2121–2138. URL: <https://doi.org/10.1108/BFJ-09-2019-0683> (дата звернення: 18.03.2026).
7. Fischer T., Krauss C. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*. 2018. Vol. 270, No. 2. P. 654–669. URL: <https://doi.org/10.1016/j.ejor.2017.11.054> (дата звернення: 18.03.2026).

8. Brignoli P. L., Varacca A., Gardebroek C., Sckokai P. Machine learning to predict grains futures prices. *Agricultural Economics*. 2024. Vol. 55, No. 3. P. 479–497. URL: <https://doi.org/10.1111/agec.12828> (дата звернення: 20.03.2026).

9. Niu Z., Zhong G., Yu H. A review on the attention mechanism of deep learning. *Neurocomputing*. 2021. Vol. 452. P. 48–62. URL: <https://doi.org/10.1016/j.neucom.2021.03.091> (дата звернення: 20.03.2026).

10. Боднар О. В. Вплив війни на розвиток ринку зерна в Україні. *Інноваційна економіка*. 2023. № 4. URL: <https://inneco.org/index.php/innecoua/article/view/1151> (дата звернення: 01.05.2026).

11. Hyndman R. J., Koehler A. B. Another look at measures of forecast accuracy. *International Journal of Forecasting*. 2006. Vol. 22, No. 4. P. 679–688. URL: <https://doi.org/10.1016/j.ijforecast.2006.03.001> (дата звернення: 20.03.2026).

12. Chai T., Draxler R. R. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*. 2014. Vol. 7, No. 3. P. 1247–1250. URL: <https://doi.org/10.5194/gmd-7-1247-2014> (дата звернення: 20.03.2026).

13. Kim S., Kim H. A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*. 2016. Vol. 32, No. 3. P. 669–679. URL: <https://doi.org/10.1016/j.ijforecast.2015.12.003> (дата звернення: 21.03.2026).

14. Diebold F. X. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of Diebold–Mariano tests. *Journal of Business and Economic Statistics*. 2015. Vol. 33, No. 1. P. 1–9. URL: <https://doi.org/10.1080/07350015.2014.983236> (дата звернення: 21.03.2026).

15. Harvey D., Leybourne S., Newbold P. Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting*. 1997. Vol. 13, No. 2.

futures and spot markets. *Agriculture*. 2025. Vol. 15, No. 9. P. 923. URL: <https://doi.org/10.3390/agriculture15090923> (дата звернення: 27.03.2026).

24. Mensi W., Hammoudeh S., Reboredo J., Nguyen D. Do Global Factors Impact BRICS Stock Markets? A Quantile Regression Approach. *Emerging Markets Review*. 2014. Vol. 19. P. 1–17. URL: <https://doi.org/10.1016/j.ememar.2014.04.002> (дата звернення: 26.03.2026).

25. Bouteska A., Hajek P., Fisher B., Abedin M. Z. Nonlinearity in forecasting energy commodity prices: Evidence from a focused time-delayed neural network. *Research in International Business and Finance*. 2023. Vol. 64. P. 101863. URL: <https://doi.org/10.1016/j.ribaf.2022.101863> (дата звернення: 27.03.2026).

26. Tang K., Xiong W. Index Investment and the Financialization of Commodities. *Financial Analysts Journal*. 2012. Vol. 68, No. 6. P. 54–74. URL: <https://doi.org/10.2469/faj.v68.n6.5> (дата звернення: 27.03.2026).

27. Devadoss S., Ridley W. Impacts of the Russian invasion of Ukraine on the global wheat market. *World Development*. 2024. Vol. 173. P. 106396. URL: <https://doi.org/10.1016/j.worlddev.2023.106396> (дата звернення: 27.03.2026).

28. Yahoo Finance. URL: <https://finance.yahoo.com> (дата звернення: 27.03.2026).

29. Ray S., Lama A., Mishra P., Biswas T., Das S. S., Gurung B. An ARIMA-LSTM model for predicting volatile agricultural price series with random forest technique. *Applied Soft Computing*. 2023. Vol. 149. P. 110939. URL: <https://doi.org/10.1016/j.asoc.2023.110939> (дата звернення: 27.03.2026).

30. Gujarati D. N., Porter D. C. Basic Econometrics. 5th ed. New York : McGraw-Hill, 2009. 922 p.

31. Fat Tails in Financial Return Distributions Revisited: Evidence from the Korean Stock Market. arXiv. 2019. URL: <https://arxiv.org/abs/1904.02567> (дата звернення: 28.03.2026).

32. LeCun Y., Bengio Y., Hinton G. Deep learning. *Nature*. 2015. Vol. 521, No. 7553. P. 436–444. URL: <https://doi.org/10.1038/nature14539> (дата звернення: 28.03.2026).

33. Sherstinsky A. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network. *Physica D: Nonlinear Phenomena*. 2020. Vol. 404. P. 132306. URL: <https://doi.org/10.1016/j.physd.2019.132306> (дата звернення: 28.03.2026).

34. Hyndman R. J., Athanasopoulos G. Forecasting: Principles and Practice. 3rd ed. Melbourne : OTexts, 2021. URL: <https://otexts.com/fpp3> (дата звернення: 15.04.2026).

ДОДАТКИ

Додаток А

Сформована база показників за період 2010-2026

Date	Wheat	Brent	DXY	VIX	SP500
04.01.2010	557,75	80,12	77,53	20,04	1 132,99
05.01.2010	553	80,59	77,62	19,35	1 136,52
06.01.2010	567,25	81,89	77,49	19,16	1 137,14
07.01.2010	557,75	81,51	77,91	19,06	1 141,69
08.01.2010	568,5	81,37	77,47	18,13	1 144,98
11.01.2010	572,5	80,97	77,00	17,55	1 146,98
12.01.2010	535,75	79,30	76,95	18,25	1 136,22
13.01.2010	537	78,31	76,85	17,85	1 145,68
14.01.2010	527,75	77,82	76,73	17,63	1 148,46
15.01.2010	510	77,11	77,32	17,91	1 136,03
19.01.2010	500,5	77,63	77,50	17,58	1 150,23
20.01.2010	497,5	76,32	78,34	18,68	1 138,04
21.01.2010	499,5	74,58	78,32	22,27	1 116,48
22.01.2010	498,5	72,83	78,28	27,31	1 091,76
25.01.2010	498,25	73,69	78,19	25,41	1 096,78
26.01.2010	494	73,29	78,43	24,55	1 092,17
27.01.2010	483,75	72,24	78,68	23,14	1 097,50
28.01.2010	487	72,13	78,90	23,73	1 084,53
29.01.2010	474	71,46	79,46	24,62	1 073,87
01.02.2010	474,75	73,11	79,24	22,59	1 089,19
02.02.2010	487,25	76,06	79,01	21,48	1 103,32
03.02.2010	469	75,92	79,37	21,60	1 097,28

04.02.2010	475,75	72,13	79,92	26,08	1 063,11
05.02.2010	473,25	69,59	80,44	26,11	1 066,19
08.02.2010	484	70,11	80,30	26,51	1 056,74
09.02.2010	482,25	72,13	79,86	26,00	1 070,52
10.02.2010	496,75	72,54	80,03	25,40	1 068,13
11.02.2010	493,5	73,05	79,99	23,96	1 078,47
12.02.2010	486,5	72,90	80,32	22,73	1 075,51
16.02.2010	505	75,68	79,70	22,25	1 094,87
17.02.2010	494,75	76,27	80,38	21,72	1 099,51
18.02.2010	485	77,78	80,40	20,63	1 106,75
19.02.2010	489,75	78,19	80,64	20,02	1 109,17
22.02.2010	501,25	78,61	80,51	19,94	1 108,01
23.02.2010	491,75	77,25	80,85	21,37	1 094,60
24.02.2010	500,25	78,09	80,85	20,27	1 105,24
25.02.2010	489,75	76,29	80,79	20,10	1 102,94
26.02.2010	506,5	77,59	80,36	19,05,2026	1 104,49
01.03.2010	492,75	76,89	80,66	19,26	1 115,71
02.03.2010	492	78,18	80,52	19,06	1 118,31
03.03.2010	503,5	79,25	79,98	18,83	1 118,79
04.03.2010	490,5	78,54	80,56	18,72	1 122,97
05.03.2010	482,25	79,89	80,43	17,42	1 138,70
08.03.2010	484,5	80,47	80,43	17,79	1 138,50
09.03.2010	478,5	79,91	80,59	17,92	1 140,45
10.03.2010	470,75	80,48	80,45	18,57	1 145,61
11.03.2010	468,5	80,28	80,32	18,06	1 150,24

12.03.2010	472,25	79,39	79,83	17,58	1 149,99
15.03.2010	479,25	77,89	80,25	18,00	1 150,51
16.03.2010	487	79,02	79,75	17,69	1 159,46
17.03.2010	496	81,96	79,64	16,91	1 166,21
18.03.2010	489,25	81,48	80,22	16,62	1 165,83
19.03.2010	483,75	79,88	80,72	16,97	1 159,90
22.03.2010	486,5	80,54	80,65	16,87	1 165,81
23.03.2010	476,75	80,70	80,90	16,35	1 174,17
24.03.2010	476	79,62	81,84	17,55	1 167,72
25.03.2010	466,5	79,61	82,12	18,40	1 165,73
26.03.2010	464,75	79,29	81,68	17,77	1 166,59
29.03.2010	464,75	81,17	81,36	17,59	1 173,22
30.03.2010	472	81,28	81,47	17,13	1 173,27
31.03.2010	450,5	82,70	81,07	17,59	1 169,43
01.04.2010	454,75	84,01	80,78	17,47	1 178,10
05.04.2010	453,5	85,88	81,09	17,02	1 187,44
06.04.2010	463,5	86,15	81,39	16,23	1 189,44
07.04.2010	475,25	85,59	81,44	16,62	1 182,45
08.04.2010	469,25	84,81	81,53	16,48	1 186,44
09.04.2010	465,75	84,83	81,09	16,14	1 194,37
12.04.2010	467,75	84,77	80,56	15,58	1 196,48
13.04.2010	476	84,72	80,51	16,20	1 197,30
14.04.2010	474,75	86,15	80,19	15,59	1 210,65
15.04.2010	480,25	87,17	80,48	15,89	1 211,67
16.04.2010	490,5	85,99	80,82	18,36	1 192,13

19.04.2010	467,75	84,23	81,00	17,34	1 197,52
20.04.2010	486	84,80	81,05	15,73	1 207,17
21.04.2010	487,25	85,70	81,16	16,32	1 205,94
22.04.2010	498,5	85,67	81,57	16,47	1 208,67
23.04.2010	493,25	87,25	81,35	16,62	1 217,28
26.04.2010	476	86,83	81,48	17,47	1 212,05
27.04.2010	478,25	85,78	82,15	22,81	1 183,71
28.04.2010	476,25	86,16	82,38	21,08	1 191,36
29.04.2010	484,5	86,90	82,00	18,44	1 206,78
30.04.2010	491,75	87,44	81,87	22,05	1 186,69
03.05.2010	490	88,94	82,27	20,19	1 202,26
04.05.2010	499,75	88,94	83,30	23,84	1 173,60
05.05.2010	501	82,61	84,08	24,91	1 165,87
06.05.2010	497,25	79,83	84,89	32,80	1 128,15
07.05.2010	500,25	78,27	84,45	40,95	1 110,88
10.05.2010	482,5	80,12	84,16	28,84	1 159,73
11.05.2010	483,75	80,49	84,47	28,32	1 155,79
12.05.2010	482,75	81,20	84,83	25,52	1 171,67
13.05.2010	471	80,11	85,21	26,68	1 157,44
14.05.2010	463,5	77,18	86,10	31,24	1 135,68
17.05.2010	469	75,10	86,21	30,84	1 136,94
18.05.2010	467,75	74,43	87,16	33,55	1 120,80
19.05.2010	469,25	73,69	86,39	35,32	1 115,05
20.05.2010	469,75	71,84	85,57	45,79	1 071,59
21.05.2010	472	71,68	85,37	40,10	1 087,69

24.05.2010	467,5	71,17	86,21	38,32	1 073,65
25.05.2010	460,5	69,55	86,78	34,61	1 074,03
26.05.2010	461,75	71,74	87,12	35,02	1 067,95
27.05.2010	467,75	74,66	86,18	29,68	1 103,06
28.05.2010	457,75	74,02	86,48	32,07	1 089,41
01.06.2010	450,75	72,71	86,64	35,54	1 070,71
02.06.2010	442,5	73,75	86,79	30,17	1 098,38
03.06.2010	441,75	75,41	87,14	29,46	1 102,83
04.06.2010	435,75	72,09	88,23	35,48	1 064,88
07.06.2010	432,25	72,09	88,40	36,57	1 050,47
08.06.2010	432,25	72,09	88,39	33,70	1 062,00
09.06.2010	428	72,09	87,90	33,73	1 055,69
10.06.2010	433,25	75,29	87,17	30,57	1 086,84
11.06.2010	440,75	74,35	87,51	28,79	1 091,60
14.06.2010	451,5	75,20	86,51	28,58	1 089,63
15.06.2010	451,75	76,20	85,98	25,87	1 115,23
16.06.2010	461,25	78,14	86,08	25,92	1 114,61
17.06.2010	462,75	78,68	85,69	25,05	1 116,04
18.06.2010	461,75	78,22	85,70	23,95	1 117,51
21.06.2010	462	78,82	85,93	24,88	1 113,20
22.06.2010	460,75	78,04	86,11	27,05	1 095,31
23.06.2010	462,25	76,27	85,74	26,91	1 092,04
24.06.2010	463	76,47	85,73	29,74	1 073,69
25.06.2010	456,25	78,12	85,31	28,53	1 076,76
28.06.2010	449,5	77,59	85,65	29,00	1 074,57

29.06.2010	442	75,44	86,03	34,13	1 041,24
30.06.2010	464,75	75,01	86,02	34,54	1 030,71
01.07.2010	484	72,34	84,72	32,86	1 027,37
02.07.2010	487,75	71,65	84,43	30,12	1 022,58
06.07.2010	492,75	71,45	84,08	29,65	1 028,06
07.07.2010	515,75	73,51	83,82	26,84	1 060,27
08.07.2010	534	74,71	83,82	25,71	1 070,25
09.07.2010	523,5	75,42	83,95	24,98	1 077,96
12.07.2010	521,25	74,37	84,20	24,43	1 078,75
13.07.2010	535,75	76,65	83,64	24,56	1 095,34
14.07.2010	548,25	76,77	83,43	24,89	1 095,17
15.07.2010	596,25	76,77	82,56	25,14	1 096,48
16.07.2010	587,25	75,37	82,49	26,25	1 064,88
19.07.2010	582,25	75,62	82,51	25,97	1 071,25
20.07.2010	577	76,22	82,75	23,93	1 083,48
21.07.2010	588,25	75,37	83,39	25,64	1 069,59
22.07.2010	596,5	77,82	82,60	24,63	1 093,67
23.07.2010	596,25	77,45	82,46	23,47	1 102,66
26.07.2010	589,5	77,50	82,09	22,73	1 115,01
27.07.2010	595	76,13	82,18	23,19	1 113,84
28.07.2010	615,5	76,06	82,18	24,25	1 106,13
29.07.2010	627,5	77,59	81,64	24,13	1 101,53
30.07.2010	661,5	78,18	81,54	23,05,2026	1 101,60
02.08.2010	693,25	80,82	80,94	22,01	1 125,86
03.08.2010	680	82,68	80,60	22,63	1 120,46

04.08.2010	725,75	82,20	80,89	22,21	1 127,24
05.08.2010	785,75	81,61	80,82	22,10	1 125,81
06.08.2010	725,75	80,16	80,41	21,74	1 121,64
09.08.2010	712,5	80,99	80,71	22,14	1 127,79
10.08.2010	694,75	79,60	80,80	22,37	1 121,06
11.08.2010	694,75	77,64	82,29	25,39	1 089,47
12.08.2010	713	75,52	82,64	25,73	1 083,61
13.08.2010	702,5	75,11	82,95	26,24	1 079,25
16.08.2010	663,75	74,85	82,53	26,10	1 079,38
17.08.2010	651	76,93	82,23	24,33	1 092,54
18.08.2010	656	76,47	82,22	24,59	1 094,16
19.08.2010	681,25	75,30	82,45	26,44	1 075,63
20.08.2010	679	74,26	83,06	25,49	1 071,69
23.08.2010	692,25	73,62	83,12	25,66	1 067,36
24.08.2010	674,75	72,38	83,15	27,46	1 051,87
25.08.2010	647,75	73,48	83,26	26,70	1 055,33
26.08.2010	656,5	75,02	82,93	27,37	1 047,22
27.08.2010	662,5	76,65	82,92	24,45	1 064,59
30.08.2010	671,25	76,60	83,16	27,21	1 048,92
31.08.2010	652,5	74,64	83,20	26,05	1 049,33
01.09.2010	675,5	76,35	82,52	23,89	1 080,29
02.09.2010	680,75	76,93	82,46	23,19	1 090,10
03.09.2010	708,25	76,67	83,10	21,31	1 104,51
07.09.2010	702,5	77,74	82,82	23,80	1 091,84
08.09.2010	678,75	78,17	82,59	23,25	1 098,87

09.09.2010	706,75	77,47	82,67	22,81	1 104,18
10.09.2010	704,75	78,16	82,70	21,99	1 109,55
13.09.2010	713,75	79,03	81,91	21,21	1 121,90
14.09.2010	704,5	79,16	81,08	21,56	1 121,10
15.09.2010	726,75	78,91	81,49	22,10	1 125,07
16.09.2010	719,25	78,48	81,24	21,72	1 124,66
17.09.2010	739,25	78,21	81,40	22,01	1 125,59
20.09.2010	731,75	79,32	81,34	21,05,2026	1 142,71
21.09.2010	718	78,42	80,44	22,35	1 139,78
22.09.2010	719,75	77,95	79,83	22,51	1 134,28
23.09.2010	697,25	78,11	80,01	23,87	1 124,83
24.09.2010	720	78,87	79,39	21,71	1 148,67
27.09.2010	706,5	78,57	79,34	22,54	1 142,16
28.09.2010	684,75	78,71	79,01	22,60	1 147,70
29.09.2010	683,5	80,77	78,69	23,25	1 144,73
30.09.2010	674	82,31	78,72	23,70	1 141,20
01.10.2010	655	83,75	78,09	22,05,2026	1 146,24
04.10.2010	647,25	83,28	78,45	23,53	1 137,03
05.10.2010	663,5	84,84	77,75	21,76	1 160,75
06.10.2010	658,25	85,06	77,39	21,49	1 159,97
07.10.2010	659,25	83,43	77,39	21,56	1 158,06
08.10.2010	719,25	84,03	77,32	20,71	1 165,15
11.10.2010	709,25	83,72	77,44	18,96	1 165,32
12.10.2010	710	83,50	77,36	18,93	1 169,77
13.10.2010	702,75	84,64	77,07	19,07	1 178,10

14.10.2010	700,75	84,53	76,65	19,88	1 173,81
15.10.2010	704,5	82,45	77,04	19,03	1 176,19
18.10.2010	690	84,37	76,93	19,09	1 184,71
19.10.2010	671,5	81,10	78,18	20,63	1 165,90
20.10.2010	683	83,60	77,17	19,79	1 178,17
21.10.2010	668,75	81,83	77,42	19,27	1 180,26
22.10.2010	670,75	82,96	77,47	18,78	1 183,08
25.10.2010	674	83,54	77,10	19,85	1 185,62
26.10.2010	692	83,66	77,71	20,22	1 185,64
27.10.2010	702,75	83,23	78,15	20,71	1 182,45
28.10.2010	718,25	83,59	77,31	20,88	1 183,78
29.10.2010	717,25	83,15	77,27	21,20	1 183,26
01.11.2010	702,5	84,62	77,30	21,83	1 184,38
02.11.2010	694,25	85,41	76,72	21,57	1 193,57
03.11.2010	690,25	86,38	76,48	19,56	1 197,96
04.11.2010	713,75	88,00	75,88	18,52	1 221,06
05.11.2010	728,75	88,11	76,55	18,26	1 225,85
08.11.2010	736,25	88,46	77,03	18,29	1 223,25
09.11.2010	721,75	88,33	77,44	19,08	1 213,40
10.11.2010	710	88,96	77,63	18,47	1 218,71
11.11.2010	704	88,81	78,22	18,64	1 213,54
12.11.2010	669,25	86,34	78,08	20,61	1 199,21
15.11.2010	672,75	86,70	78,52	20,20	1 197,75
16.11.2010	626,25	84,73	79,21	22,58	1 178,34
17.11.2010	632,5	83,28	79,08	21,76	1 178,59

18.11.2010	645,25	85,05	78,62	18,75	1 196,69
19.11.2010	644,5	84,34	78,50	18,04	1 199,73
22.11.2010	645,5	83,96	78,68	18,37	1 197,84
23.11.2010	642	83,25	79,68	20,63	1 180,73
24.11.2010	647,5	85,84	79,87	19,56	1 198,35
26.11.2010	648,25	85,58	80,36	22,22	1 189,40
29.11.2010	650,25	87,34	80,84	21,53	1 187,76
30.11.2010	650,25	85,92	81,20	23,54	1 180,55
01.12.2010	697,5	88,87	80,71	21,36	1 206,07
02.12.2010	706,5	90,69	80,30	19,39	1 221,53
03.12.2010	738	91,42	79,38	18,01	1 224,71
06.12.2010	752	91,45	79,57	18,02	1 223,12
07.12.2010	743	91,39	79,86	17,99	1 223,75
08.12.2010	742	90,77	80,00	17,74	1 228,28
09.12.2010	748,25	90,99	80,07	17,25	1 233,00
10.12.2010	735,5	90,48	80,07	17,61	1 240,40
13.12.2010	740,25	91,19	79,28	17,55	1 240,46
14.12.2010	721	91,21	79,37	17,61	1 241,59
15.12.2010	764,75	92,20	80,26	17,94	1 235,23
16.12.2010	749,75	91,71	80,18	17,39	1 242,87
17.12.2010	756,75	91,67	80,37	16,11	1 243,91
20.12.2010	769,5	92,74	80,63	16,41	1 247,08
21.12.2010	765	93,20	80,72	16,49	1 254,60
22.12.2010	783,5	93,65	80,73	15,45	1 258,84
23.12.2010	783	94,25	80,50	16,47	1 256,77

27.12.2010	780,25	93,85	80,37	17,67	1 257,54
28.12.2010	798,25	94,38	80,36	17,52	1 258,51
29.12.2010	799,25	94,14	79,80	17,28	1 259,78
30.12.2010	784,75	93,09	79,52	17,52	1 257,88
31.12.2010	794,25	94,75	79,03	17,75	1 257,64
03.01.2011	805,5	94,84	79,13	17,61	1 271,87
04.01.2011	789,25	93,53	79,44	17,38	1 270,20
05.01.2011	808,25	95,50	80,26	17,02	1 276,56
06.01.2011	789	94,52	80,79	17,40	1 273,85
07.01.2011	774	93,33	81,01	17,14	1 271,50
10.01.2011	767,25	95,70	80,88	17,54	1 269,75
11.01.2011	759,5	97,61	80,85	16,89	1 274,48
12.01.2011	770,5	98,12	80,03	16,24	1 285,96
13.01.2011	783,5	98,06	79,19	16,39	1 283,76
14.01.2011	773,25	98,06	79,16	15,46	1 293,24
18.01.2011	793,25	97,80	78,96	15,87	1 295,02
19.01.2011	797,25	98,16	78,64	17,31	1 281,92
20.01.2011	803,5	96,58	78,82	17,99	1 280,26
21.01.2011	824,5	97,60	78,21	18,47	1 283,35
24.01.2011	835,25	96,61	78,05	17,65	1 290,84
25.01.2011	838,25	95,25	78,00	17,59	1 291,18
26.01.2011	856,5	97,91	77,90	16,64	1 296,63
27.01.2011	846,25	97,39	77,73	16,15	1 299,54
28.01.2011	825,75	99,42	78,13	20,04	1 276,34
31.01.2011	840,75	101,01	77,74	19,53	1 286,12

01.02.2011	835,75	101,74	77,07	17,63	1 307,59
02.02.2011	863	102,34	77,16	17,30	1 304,03
03.02.2011	859	101,76	77,75	16,69	1 307,10
04.02.2011	853,75	99,83	78,04	15,93	1 310,87
07.02.2011	858,75	99,25	78,03	16,28	1 319,05
08.02.2011	874,25	99,92	78,00	15,81	1 324,57
09.02.2011	886	101,82	77,64	15,87	1 320,88
10.02.2011	862,75	100,87	78,25	16,09	1 321,87
11.02.2011	867	101,43	78,46	15,69	1 329,15
14.02.2011	872	103,08	78,61	15,95	1 332,32
15.02.2011	840,25	101,64	78,57	16,37	1 328,01
16.02.2011	837	103,78	78,22	16,72	1 336,32
17.02.2011	850,75	102,59	78,00	16,59	1 340,43
18.02.2011	822,25	102,52	77,66	16,43	1 343,01
22.02.2011	762,25	105,78	77,77	20,80	1 315,44
23.02.2011	763	111,25	77,41	22,13	1 307,40
24.02.2011	747,25	111,36	77,06	21,32	1 306,10
25.02.2011	776,5	112,14	77,28	19,22	1 319,88
28.02.2011	782,5	111,80	76,89	18,35	1 327,22
01.03.2011	775,75	115,42	77,05	21,01	1 306,33
02.03.2011	776,75	116,35	76,67	20,70	1 308,44
03.03.2011	790,5	114,79	76,48	18,60	1 330,97
04.03.2011	800,5	115,97	76,40	19,06	1 321,15
07.03.2011	769	115,04	76,50	20,66	1 310,13
08.03.2011	751,25	113,06	76,80	19,82	1 321,82

09.03.2011	732,25	115,94	76,72	20,22	1 320,02
10.03.2011	714	115,43	77,28	21,88	1 295,11
11.03.2011	695	113,84	76,78	20,08	1 304,28
14.03.2011	700,5	113,67	76,35	21,13	1 296,39
15.03.2011	667,75	108,52	76,33	24,32	1 281,87
16.03.2011	662	110,62	76,68	29,40	1 256,88
17.03.2011	710,25	114,90	76,04	26,37	1 273,72
18.03.2011	723	113,93	75,72	24,44	1 279,21
21.03.2011	721	114,96	75,40	20,61	1 298,38
22.03.2011	722,25	115,70	75,43	20,21	1 293,77
23.03.2011	714,25	115,55	75,79	19,17	1 297,54
24.03.2011	739,5	115,72	75,66	18,00	1 309,66
25.03.2011	733,25	115,59	76,22	17,91	1 313,80
28.03.2011	725,25	114,80	76,13	19,44	1 310,19
29.03.2011	737,25	114,80	76,25	18,16	1 319,44
30.03.2011	727,25	115,13	76,12	17,71	1 328,26
31.03.2011	763,25	117,36	75,86	17,74	1 325,83
01.04.2011	759,5	118,70	75,83	17,40	1 332,41
04.04.2011	790	121,06	75,91	17,05,2026	1 332,87
05.04.2011	786,25	122,22	75,88	17,25	1 332,63
06.04.2011	782,25	122,30	75,51	16,90	1 335,54
07.04.2011	773,25	122,67	75,59	17,11	1 333,51
08.04.2011	797,5	126,65	75,07	17,87	1 328,17
11.04.2011	798,25	123,98	75,06	16,59	1 324,46
12.04.2011	759,5	120,92	74,85	17,09	1 314,16

13.04.2011	752,75	122,88	74,98	16,92	1 314,41
14.04.2011	740,5	122,36	74,68	16,27	1 314,52
15.04.2011	744,25	123,45	74,83	15,32	1 319,68
18.04.2011	775	121,61	75,50	16,96	1 305,14
19.04.2011	785,75	121,33	75,03	15,83	1 312,62
20.04.2011	785	121,33	74,37	15,07	1 330,36
21.04.2011	799,5	123,99	74,00	14,69	1 337,38
25.04.2011	826	123,66	73,99	15,77	1 335,25
26.04.2011	811,25	124,14	73,84	15,62	1 347,24
27.04.2011	777	125,13	73,52	15,35	1 355,66
28.04.2011	743	125,02	73,12	14,62	1 360,48
29.04.2011	769,25	125,89	72,93	14,75	1 363,61
02.05.2011	759,75	125,12	72,95	15,99	1 361,22
03.05.2011	760,75	122,45	73,14	16,70	1 356,62
04.05.2011	741	121,19	73,03	17,08	1 347,32
05.05.2011	722,5	110,80	74,19	18,20	1 335,10
06.05.2011	724,5	109,13	74,84	18,40	1 340,20
09.05.2011	759,25	115,90	74,74	17,16	1 346,29
10.05.2011	769	117,63	74,61	15,91	1 357,16
11.05.2011	727,75	112,57	75,33	16,95	1 342,08
12.05.2011	700,5	112,98	75,24	16,03	1 348,65
13.05.2011	696	113,83	75,76	17,07	1 337,77
16.05.2011	736,5	112,73	75,47	18,24	1 329,47
17.05.2011	764	109,99	75,40	17,55	1 328,98
18.05.2011	817	112,30	75,47	16,23	1 340,68

19.05.2011	812	111,42	75,12	15,52	1 343,60
20.05.2011	806,5	112,39	75,43	17,43	1 333,27
23.05.2011	803	110,10	76,10	18,27	1 317,37
24.05.2011	779,75	112,53	75,88	17,82	1 316,28
25.05.2011	796,5	114,93	75,93	17,07	1 320,47
26.05.2011	814,5	115,05	75,55	16,09	1 325,69
27.05.2011	819,75	115,03	74,95	15,98	1 331,10
31.05.2011	782,25	116,73	74,64	15,45	1 345,20
01.06.2011	759,25	114,53	74,68	18,30	1 314,55
02.06.2011	769,75	115,54	74,33	18,09	1 312,94
03.06.2011	773,75	115,84	73,78	17,95	1 300,16
06.06.2011	744	114,48	73,95	18,49	1 286,17
07.06.2011	733,75	116,78	73,52	18,07	1 284,94
08.06.2011	748	117,85	73,93	18,79	1 279,56
09.06.2011	745	119,57	74,19	17,77	1 289,00
10.06.2011	759,25	118,78	74,79	18,86	1 270,98
13.06.2011	743	119,10	74,51	19,61	1 271,83
14.06.2011	731,25	120,16	74,33	18,26	1 287,87
15.06.2011	708,5	117,10	75,60	21,32	1 265,42
16.06.2011	673,25	114,02	75,75	22,73	1 267,64
17.06.2011	672,25	113,21	74,99	21,85	1 271,50
20.06.2011	659,25	111,69	75,03	19,99	1 278,36
21.06.2011	674,25	110,95	74,54	18,86	1 295,52
22.06.2011	638,25	114,21	74,78	18,52	1 287,14
23.06.2011	649	107,26	75,45	19,29	1 283,50

24.06.2011	635,75	105,12	75,67	21,10	1 268,45
27.06.2011	622,75	105,99	75,39	20,56	1 280,10
28.06.2011	640,25	108,78	75,06	19,17	1 296,67
29.06.2011	641,25	112,40	74,69	17,27	1 307,41
30.06.2011	584,75	112,48	74,30	16,52	1 320,64
01.07.2011	584,5	111,77	74,35	15,87	1 339,67
05.07.2011	613,75	111,77	74,69	16,06	1 337,88
06.07.2011	611,25	113,62	75,12	16,34	1 339,22
07.07.2011	624,5	118,59	74,96	15,95	1 353,22
08.07.2011	650,5	118,33	75,18	15,95	1 343,80
11.07.2011	636	117,24	75,99	18,39	1 319,49
12.07.2011	662,75	117,75	75,81	19,87	1 313,64
13.07.2011	702	117,75	75,21	19,91	1 317,72
14.07.2011	691,25	118,32	75,25	20,80	1 308,87
15.07.2011	694,75	117,26	75,13	19,53	1 316,14
18.07.2011	689,5	116,05	75,48	20,95	1 305,44
19.07.2011	693,5	116,05	75,22	19,21	1 326,73
20.07.2011	697	118,15	74,79	19,09	1 325,84
21.07.2011	677,25	117,51	74,02	17,56	1 343,80
22.07.2011	692,25	118,67	74,20	17,52	1 345,02
25.07.2011	688,5	117,94	74,06	19,35	1 337,43
26.07.2011	694	118,28	73,47	20,23	1 331,94
27.07.2011	704,75	117,43	74,09	22,98	1 304,89
28.07.2011	693,25	117,36	74,22	23,74	1 300,67
29.07.2011	672,5	116,74	73,90	25,25	1 292,28

01.08.2011	676,5	116,81	74,26	23,66	1 286,94
02.08.2011	718	116,46	74,51	24,79	1 254,05
03.08.2011	710,5	113,23	74,04	23,38	1 260,34
04.08.2011	681,75	107,25	75,12	31,66	1 200,07
05.08.2011	679	109,37	74,60	32,00	1 199,38
08.08.2011	656,5	103,74	74,79	48,00	1 119,46
09.08.2011	671,75	102,57	74,61	35,06	1 172,53
10.08.2011	685	106,68	74,64	42,99	1 120,76
11.08.2011	701,25	108,02	74,68	39,00	1 172,64
12.08.2011	702,5	108,03	74,61	36,36	1 178,81
15.08.2011	712,5	109,91	73,84	31,87	1 204,49
16.08.2011	724,75	109,91	74,01	32,85	1 192,76
17.08.2011	727,5	110,60	73,67	31,58	1 193,89
18.08.2011	707,75	106,99	74,24	42,67	1 140,65
19.08.2011	730,75	108,62	73,99	43,05	1 123,53
22.08.2011	735,5	108,36	74,07	42,44	1 123,82
23.08.2011	757,25	109,31	73,87	36,27	1 162,35
24.08.2011	749,25	110,15	74,01	35,90	1 177,60
25.08.2011	757,25	110,15	74,28	39,76	1 159,27
26.08.2011	762,25	111,36	73,81	35,59	1 176,80
29.08.2011	757,25	111,88	73,72	32,28	1 210,08
30.08.2011	750,25	114,02	73,94	32,89	1 212,92
31.08.2011	745,25	114,85	74,13	31,62	1 218,89
01.09.2011	715,5	114,29	74,48	31,82	1 204,42
02.09.2011	730	112,33	74,75	33,92	1 173,97

06.09.2011	716,25	112,89	75,92	37,00	1 165,24
07.09.2011	714,75	115,80	75,46	33,38	1 198,62
08.09.2011	709,25	114,55	76,24	34,32	1 185,90
09.09.2011	701	112,77	77,19	38,52	1 154,23
12.09.2011	699,75	112,25	77,57	38,59	1 162,27
13.09.2011	688,75	111,89	76,91	36,91	1 172,87
14.09.2011	701,75	112,40	76,83	34,60	1 188,68
15.09.2011	696	115,34	76,24	31,97	1 209,11
16.09.2011	688,25	112,22	76,58	30,98	1 216,01
19.09.2011	673	109,14	77,39	32,73	1 204,09
20.09.2011	674,75	110,54	77,03	32,86	1 202,09
21.09.2011	666,75	110,36	77,34	37,32	1 166,76
22.09.2011	633,75	105,49	78,38	41,35	1 129,56
23.09.2011	640,75	103,97	78,30	41,25	1 136,43
26.09.2011	648,25	103,97	78,36	39,02	1 162,95
27.09.2011	658,25	107,14	77,51	37,71	1 175,38
28.09.2011	638,75	103,81	77,85	41,08	1 151,06
29.09.2011	654,25	103,95	77,92	38,84	1 160,40
30.09.2011	609,25	102,76	78,55	42,96	1 131,42
03.10.2011	619,5	101,71	79,45	45,45	1 099,23
04.10.2011	604	99,79	79,60	40,82	1 123,95
05.10.2011	625,25	99,79	78,91	37,81	1 144,03
06.10.2011	616	105,73	78,53	36,27	1 164,97
07.10.2011	607,5	105,88	78,72	36,20	1 155,46
10.10.2011	611,5	108,95	77,60	33,02	1 194,89

11.10.2011	660,75	110,73	77,58	32,86	1 195,54
12.10.2011	626,75	111,36	76,99	31,26	1 207,25
13.10.2011	618	111,11	76,99	30,70	1 203,66
14.10.2011	622,75	114,68	76,62	28,24	1 224,58
17.10.2011	624,25	110,16	77,14	33,39	1 200,86
18.10.2011	625,25	111,15	77,13	31,56	1 225,38
19.10.2011	619,5	108,39	77,11	34,44	1 209,88
20.10.2011	630,75	109,76	76,97	34,78	1 215,39
21.10.2011	632	109,56	76,30	31,32	1 238,25
24.10.2011	642,5	111,45	76,02	29,26	1 254,19
25.10.2011	636,25	110,92	76,25	32,22	1 229,05
26.10.2011	619,5	108,91	76,20	29,86	1 242,00
27.10.2011	644	112,08	74,87	25,46	1 284,59
28.10.2011	644,5	109,91	75,09	24,53	1 285,09
31.10.2011	628,25	109,56	76,17	29,96	1 253,30
01.11.2011	630	109,54	77,26	34,77	1 218,28
02.11.2011	623,5	109,34	77,01	32,74	1 237,90
03.11.2011	636	110,83	76,67	30,05,2026	1 261,15
04.11.2011	636,75	111,97	76,91	30,16	1 253,23
07.11.2011	638,75	114,56	76,96	29,85	1 261,12
08.11.2011	657	115,00	76,67	27,48	1 275,92
09.11.2011	643	112,31	77,93	36,16	1 229,10
10.11.2011	620	113,71	77,64	32,81	1 239,70
11.11.2011	616,75	114,16	76,94	30,04	1 263,85
14.11.2011	615,75	111,89	77,56	31,13	1 251,78

Програмний код для розрахунку у середовищі Visual Code Studio

```
import warnings
warnings.filterwarnings('ignore')

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
import seaborn as sns
from scipy import stats
from scipy.stats import jarque_bera

import yfinance as yf
from statsmodels.tsa.stattools import adfuller, kpss, coint
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.regression.linear_model import OLS
from statsmodels.tools.tools import add_constant
from statsmodels.stats.diagnostic import acorr_ljungbox
from sklearn.preprocessing import MinMaxScaler, StandardScaler

#


---


# СТИЛЬ ГРАФІКІВ
#


---


plt.rcParams.update({
    'figure.dpi': 150,
    'figure.facecolor': 'white',
    'axes.facecolor': '#f8f9fa',
    'axes.grid': True,
    'grid.alpha': 0.4,
    'grid.color': '#cccccc',
    'font.size': 10,
    'axes.titlesize': 12,
    'axes.titleweight': 'bold',
    'axes.labelsize': 10,
    'xtick.labelsize': 9,
    'ytick.labelsize': 9,
    'legend.fontsize': 9,
    'lines.linewidth': 1.5,
})

COLORS = {
    'wheat': '#f5a623',
    'brent': '#e74c3c',
    'dxy': '#2980b9',
    'vix': '#8e44ad',
    'sp500': '#27ae60',
    'accent': '#2c3e50',
}
```

#

1. ЗАВАНТАЖЕННЯ ДАНИХ

#

START_DATE = "2010-01-01"

END_DATE = "2026-03-30"

TICKERS = {

'Wheat': 'ZW=F',

'Brent': 'BZ=F',

'DXY': 'DX-Y.NYB',

'VIX': '^VIX',

'SP500': '^GSPC',

}

print("=" * 70)

print("ЗАВАНТАЖЕННЯ ДАНИХ З YAHOO FINANCE")

print("=" * 70)

raw = {}

for name, ticker in TICKERS.items():

print(f" Завантажую {name} ({ticker})...", end=" ")

try:

 df_t = yf.download(ticker, start=START_DATE, end=END_DATE,
 progress=False, auto_adjust=True)

raw[name] = df_t['Close'].squeeze()

print(f"ОК ({len(df_t)} рядків)")

except Exception as e:

print(f"ПОМИЛКА: {e}")

df_daily = pd.DataFrame(raw).dropna(how='all')

df_daily.index = pd.to_datetime(df_daily.index)

df_daily.index.name = 'Date'

df_daily = df_daily.ffill().dropna()

print(f"\n Діапазон даних: {df_daily.index.min().date()} → {df_daily.index.max().date()}")

print(f" Всього рядків: {len(df_daily)}")

#

ДОПОМІЖНА ФУНКЦІЯ — таблиця як рисунок

#

def table_figure(df, title, filename):

fig, ax = plt.subplots(figsize=(14, max(2.5, len(df) * 0.6 + 1.2)))

ax.axis('off')

 tbl = ax.table(
 cellText=df.values,
 colLabels=df.columns,
 rowLabels=df.index,
 cellLoc='center',
 loc='center'
)

cellText=df.values,

colLabels=df.columns,

rowLabels=df.index,

cellLoc='center',

loc='center'

)

tbl.auto_set_font_size(False)

tbl.set_fontsize(9)

tbl.scale(1, 1.55)

for (r, c), cell in tbl.get_celld().items():

```

if r == 0 or c == -1:
    cell.set_facecolor('#2c3e50')
    cell.set_text_props(color='white', fontweight='bold')
else:
    val = cell.get_text().get_text()
    if 'Так' in val:
        cell.set_facecolor('#d5f5d5')
    elif 'Hi' in val:
        cell.set_facecolor('#fdd')
    else:
        cell.set_facecolor('#f8f9fa' if r % 2 == 0 else 'white')
ax.set_title(title, fontsize=11, fontweight='bold', pad=15)
plt.tight_layout()
plt.savefig(filename, bbox_inches='tight')
plt.show()
print(f" Збережено: {filename}")

#

```

```

# 2.1 СТАТИСТИЧНИЙ АНАЛІЗ ЧАСОВИХ РЯДІВ
#

```

```

print("\n" + "=" * 70)
print("РОЗДІЛ 2.1 — СТАТИСТИЧНИЙ АНАЛІЗ ЧАСОВИХ РЯДІВ")
print("=" * 70)

# — Описова статистика

```

```

desc = df_daily.describe().T
desc['CV%'] = (desc['std'] / desc['mean'] * 100).round(2)
desc['skewness'] = df_daily.skew().round(4)
desc['kurtosis'] = df_daily.kurtosis().round(4)

jb_rows = {}
for col in df_daily.columns:
    stat, p = jarque_bera(df_daily[col].dropna())
    jb_rows[col] = {
        'JB_stat': round(stat, 2),
        'JB_p': round(p, 4),
        'Normal?': 'Так' if p > 0.05 else 'Hi'
    }
jb_df = pd.DataFrame(jb_rows).T
desc_full = pd.concat(
    [desc[['count', 'mean', 'std', 'min', '25%', '50%', '75%', 'max',
        'CV%', 'skewness', 'kurtosis']], jb_df], axis=1
)
print("\nТаблиця 2.1 — Описова статистика:")
print(desc_full.to_string())

# Рис. 1 — таблиця описової статистики
cols_show = ['mean', 'std', 'min', 'max', 'CV%', 'skewness', 'kurtosis', 'Normal?']
fig, ax = plt.subplots(figsize=(16, 3.5))
ax.axis('off')
tbl = ax.table(
    cellText=desc_full[cols_show].round(2).values,
    rowLabels=desc_full.index,
    colLabels=['Середнє', 'Ст.відх.', 'Мін', 'Макс',
        'CV%', 'Асиметрія', 'Експес', 'Нормальність'],
    cellLoc='center', rowLoc='center', loc='center'
)

```

```
tbl.auto_set_font_size(False)
tbl.set_fontsize(9)
tbl.scale(1, 1.6)
for (r, c), cell in tbl.get_celld().items():
    if r == 0 or c == -1:
        cell.set_facecolor('#2c3e50')
        cell.set_text_props(color='white', fontweight='bold')
    else:
        cell.set_facecolor('#f8f9fa' if r % 2 == 0 else 'white')
ax.set_title('Таблиця 2.1: Описова статистика часових рядів (2010–2026)',
            fontsize=12, fontweight='bold', pad=15)
plt.tight_layout()
plt.savefig('fig_01_descriptive_stats.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_01_descriptive_stats.png")
```

— Рис. 2 — Динаміка всіх рядів —

```
vars_plot = [
    ('Wheat', 'Ціна пшениці (центи/бушель)', COLORS['wheat']),
    ('Brent', 'Нафта Brent (USD/барель)', COLORS['brent']),
    ('DXY', 'Індекс долара (DXY)', COLORS['dxy']),
    ('VIX', 'VIX (індекс волатильності)', COLORS['vix']),
    ('SP500', 'S&P 500 (індекс)', COLORS['sp500']),
]

# Два шоки (без ГФК — датасет з 2010)
shocks = {
    'COVID\n2020': ('2020-02-01', '2020-09-30', '#e67e22'),
    'Вторгнення\n2022': ('2022-02-24', '2022-12-31', '#c0392b'),
}

fig, axes = plt.subplots(5, 1, figsize=(14, 16), sharex=True)
for ax, (col, ylabel, color) in zip(axes, vars_plot):
    ax.plot(df_daily.index, df_daily[col],
            color=color, linewidth=0.9, alpha=0.9)
    ax.set_ylabel(ylabel, fontsize=9)
    for label, (s, e, sc) in shocks.items():
        ax.axvspan(pd.Timestamp(s), pd.Timestamp(e),
                  alpha=0.13, color=sc)
    ax.fill_between(df_daily.index, df_daily[col],
                   df_daily[col].min(), alpha=0.08, color=color)

from matplotlib.patches import Patch
legend_els = [
    Patch(facecolor='#e67e22', alpha=0.4, label='COVID 2020'),
    Patch(facecolor='#c0392b', alpha=0.4, label='Вторгнення 2022'),
]
axes[0].legend(handles=legend_els, loc='upper left', fontsize=8)
axes[-1].xaxis.set_major_locator(mdates.YearLocator(2))
axes[-1].xaxis.set_major_formatter(mdates.DateFormatter('%Y'))
fig.suptitle('Рисунок 2.1: Динаміка часових рядів (2010–2026)',
            fontsize=13, fontweight='bold', y=1.01)
plt.tight_layout()
plt.savefig('fig_02_time_series_dynamics.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_02_time_series_dynamics.png")
```

— Рис. 3 — Пшениця zoom по двох шоках —

```
fig, axes = plt.subplots(1, 2, figsize=(13, 4))
periods = {
    'COVID 2020': ('2019-10-01', '2021-06-01'),
    'Вторгнення 2022': ('2021-10-01', '2023-06-01'),
```

```

}
for ax, (title, (s, e)) in zip(axes, periods.items()):
    mask = (df_daily.index >= s) & (df_daily.index <= e)
    sub = df_daily.loc[mask, 'Wheat']
    ax.plot(sub.index, sub, color=COLORS['wheat'], linewidth=1.5)
    ax.fill_between(sub.index, sub, sub.min(),
                    alpha=0.15, color=COLORS['wheat'])
    ax.set_title(f'Пшениця: {title}', fontweight='bold')
    ax.set_ylabel('Центи/бушель')
    ax.xaxis.set_major_formatter(mdates.DateFormatter('%m/%Y'))
    plt.setp(ax.xaxis.get_majorticklabels(), rotation=30, ha='right')
fig.suptitle('Рисунок 2.2: Поведінка цін на пшеницю під час структурних шоків',
            fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig_03_wheat_shocks.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_03_wheat_shocks.png")

```

```

# — Рис. 4 — Гістограми + KDE
fig, axes = plt.subplots(1, 5, figsize=(18, 4))
for ax, (col, _, color) in zip(axes, vars_plot):
    data = df_daily[col].dropna()
    ax.hist(data, bins=60, color=color, alpha=0.5,
            density=True, edgecolor='none')
    data.plot.kde(ax=ax, color=color, linewidth=2)
    ax.axvline(data.mean(), color='black', linestyle='--',
               linewidth=1.2, label=f'Середнє: {data.mean():.1f}')
    ax.axvline(data.median(), color='gray', linestyle=':',
               linewidth=1.2, label=f'Медіана: {data.median():.1f}')
    ax.set_title(col, fontweight='bold')
    ax.legend(fontsize=7)
    ax.set_xlabel('Значення')
    ax.set_ylabel('Щільність')
fig.suptitle('Рисунок 2.3: Розподіл значень змінних (гістограма + KDE)',
            fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig_04_distributions.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_04_distributions.png")

```

```

# — Рис. 5 — Волатильність (30-денне ковзне ст. відхилення)
wheat_vol = (df_daily['Wheat'].pct_change()
             .rolling(30).std() * np.sqrt(252) * 100)
fig, ax = plt.subplots(figsize=(14, 5))
ax.fill_between(wheat_vol.index, wheat_vol,
                alpha=0.4, color=COLORS['wheat'])
ax.plot(wheat_vol.index, wheat_vol,
        color=COLORS['wheat'], linewidth=1)
for label, (s, e, sc) in shocks.items():
    ax.axvspan(pd.Timestamp(s), pd.Timestamp(e),
               alpha=0.13, color=sc,
               label=label.replace("\n", ' '))
ax.set_ylabel('Річна волатильність (%)')
ax.set_title('Рисунок 2.4: Волатильність цін на пшеницю '
            '(30-денне ковзне стандартне відхилення)',
            fontweight='bold')
ax.legend()
plt.tight_layout()
plt.savefig('fig_05_volatility.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_05_volatility.png")

```

```
# — Рис. 6 — Матриця кореляцій
returns = df_daily.pct_change().dropna()
fig, axes = plt.subplots(1, 2, figsize=(14, 5))
mask_upper = np.triu(np.ones_like(df_daily.corr(), dtype=bool))
sns.heatmap(df_daily.corr(), annot=True, fmt='.2f', cmap='RdYlGn',
            center=0, mask=mask_upper, ax=axes[0],
            linewidths=0.5, cbar_kws={'label': 'Pearson r'})
axes[0].set_title('Кореляція: півні ціні', fontweight='bold')
sns.heatmap(returns.corr(), annot=True, fmt='.2f', cmap='RdYlGn',
            center=0, mask=mask_upper, ax=axes[1],
            linewidths=0.5, cbar_kws={'label': 'Pearson r'})
axes[1].set_title('Кореляція: денні прибутковості', fontweight='bold')
fig.suptitle('Рисунок 2.5: Матриця кореляцій між змінними',
            fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig_06_correlation_matrix.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_06_correlation_matrix.png")
```

```
# — Рис. 7 — Scatter Wheat vs фактори
x_vars = [('Brent', COLORS['brent']), ('DXY', COLORS['dxy']),
          ('VIX', COLORS['vix']), ('SP500', COLORS['sp500'])]
fig, axes = plt.subplots(1, 4, figsize=(16, 4))
for ax, (xvar, color) in zip(axes, x_vars):
    ax.scatter(df_daily[xvar], df_daily['Wheat'],
              alpha=0.15, s=4, color=color)
    z = np.polyfit(df_daily[xvar].dropna(),
                  df_daily['Wheat'].dropna(), 1)
    xline = np.linspace(df_daily[xvar].min(),
                       df_daily[xvar].max(), 200)
    ax.plot(xline, np.poly1d(z)(xline), 'k--', linewidth=1.5)
    r, pval = stats.pearsonr(df_daily[xvar].dropna(),
                           df_daily['Wheat'].dropna())
    ax.set_xlabel(xvar)
    ax.set_ylabel('Wheat (центи/бушель)')
    ax.set_title(fr = {r:.3f} p = {pval:.3f}', fontweight='bold')
fig.suptitle('Рисунок 2.6: Діаграми розсіювання Пшениця vs фактори',
            fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig_07_scatter_plots.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_07_scatter_plots.png")
```

```
#
```

2.2 ДІАГНОСТИКА ТА СТАЦІОНАРНІСТЬ

```
#
```

```
print("\n" + "=" * 70)
print("РОЗДІЛ 2.2 — ДІАГНОСТИКА ТА СТАЦІОНАРНІСТЬ")
print("=" * 70)

def run_adf(series, name, max_lags=20):
    result = adfuller(series.dropna(), maxlag=max_lags, autolag='AIC')
    return {
        'Змінна': name,
        'ADF stat': round(result[0], 4),
        'p-value': round(result[1], 4),
        'Крит. 1%': round(result[4][1%], 4),
        'Крит. 5%': round(result[4][5%], 4),
```

```

        'Крит. 10%': round(result[4]['10%'], 4),
        'Стационарний?': 'Так ✓' if result[1] < 0.05 else 'Hi X'
    }

def run_kpss(series, name):
    try:
        stat, p, _, crit = kpss(series.dropna(),
                                regression='c', nlags='auto')

        return {
            'Змінна': name,
            'KPSS stat': round(stat, 4),
            'p-value': round(p, 4),
            'Крит. 5%': round(crit['5%'], 4),
            'Стационарний?': 'Так ✓' if p > 0.05 else 'Hi X'
        }
    except Exception:
        return {'Змінна': name, 'KPSS stat': 'N/A',
                'p-value': 'N/A', 'Крит. 5%': 'N/A',
                'Стационарний?': 'N/A'}

cols = list(df_daily.columns)
adf_levels, kpss_levels = [], []
adf_diff1, kpss_diff1 = [], []

for col in cols:
    series = df_daily[col].dropna()
    adf_levels.append(run_adf(series, col))
    kpss_levels.append(run_kpss(series, col))
    diff1 = series.diff().dropna()
    adf_diff1.append(run_adf(diff1, f' $\Delta$ {col}'))
    kpss_diff1.append(run_kpss(diff1, f' $\Delta$ {col}'))

adf_lev_df = pd.DataFrame(adf_levels).set_index('Змінна')
kpss_lev_df = pd.DataFrame(kpss_levels).set_index('Змінна')
adf_d1_df = pd.DataFrame(adf_diff1).set_index('Змінна')
kpss_d1_df = pd.DataFrame(kpss_diff1).set_index('Змінна')

print("\nADF рівні:"); print(adf_lev_df.to_string())
print("\nKPSS рівні:"); print(kpss_lev_df.to_string())
print("\nADF різниці:"); print(adf_d1_df.to_string())
print("\nKPSS різниці:"); print(kpss_d1_df.to_string())

table_figure(adf_lev_df,
              'Таблиця 2.2а: ADF тест — рівні (2010–2026)',
              'fig_08a_adf_levels.png')
table_figure(kpss_lev_df,
              'Таблиця 2.2b: KPSS тест — рівні (2010–2026)',
              'fig_08b_kpss_levels.png')
table_figure(adf_d1_df,
              'Таблиця 2.2с: ADF тест — перші різниці',
              'fig_08c_adf_diff.png')
table_figure(kpss_d1_df,
              'Таблиця 2.2d: KPSS тест — перші різниці',
              'fig_08d_kpss_diff.png')

# — Рис. 9 — Рівень vs Перша різниця (пшениця) —————
wheat = df_daily['Wheat']
wheat_d1 = wheat.diff().dropna()

fig, axes = plt.subplots(2, 2, figsize=(14, 8))
axes[0,0].plot(wheat.index, wheat,

```

```

        color=COLORS['wheat'], linewidth=0.8)
axes[0,0].set_title('Рівень ціни (нестационарний)', fontweight='bold')
axes[0,0].set_ylabel('Центи/бушель')

axes[0,1].plot(wheat_d1.index, wheat_d1,
               color=COLORS['wheat'], linewidth=0.8, alpha=0.8)
axes[0,1].axhline(0, color='black', linewidth=0.8, linestyle='--')
axes[0,1].set_title('Перша різниця (стаціонарна)', fontweight='bold')
axes[0,1].set_ylabel('Δ центи/бушель')

axes[1,0].hist(wheat, bins=50, color=COLORS['wheat'],
               alpha=0.6, density=True)
wheat.plot.kde(ax=axes[1,0], color='darkred', linewidth=2)
axes[1,0].set_title('Розподіл рівня', fontweight='bold')

axes[1,1].hist(wheat_d1, bins=50, color=COLORS['wheat'],
               alpha=0.6, density=True)
wheat_d1.plot.kde(ax=axes[1,1], color='darkred', linewidth=2)
axes[1,1].set_title('Розподіл першої різниці', fontweight='bold')

fig.suptitle('Рисунок 2.7: Рівень та перша різниця ціни пшениці',
             fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig_09_levels_vs_diff.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_09_levels_vs_diff.png")

```

```

# — Рис. 10 — ACF / PACF пшениця —
fig, axes = plt.subplots(2, 2, figsize=(14, 8))
plot_acf(wheat.dropna(), lags=60, ax=axes[0,0],
         color=COLORS['wheat'], title='ACF: Рівень ціни пшениці')
plot_pacf(wheat.dropna(), lags=60, ax=axes[0,1],
          color=COLORS['wheat'], title='PACF: Рівень ціни пшениці')
plot_acf(wheat_d1.dropna(), lags=60, ax=axes[1,0],
         color=COLORS['accent'], title='ACF: Перша різниця')
plot_pacf(wheat_d1.dropna(), lags=60, ax=axes[1,1],
          color=COLORS['accent'], title='PACF: Перша різниця')
fig.suptitle('Рисунок 2.8: Корелограми ACF та PACF (Пшениця)',
             fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig_10_acf_pacf.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_10_acf_pacf.png")

```

```

# — Рис. 11 — ACF/PACF факторів —
x_cols = ['Brent', 'DXY', 'VIX', 'SP500']
fig, axes = plt.subplots(len(x_cols), 2, figsize=(14, 14))
for i, xcol in enumerate(x_cols):
    s = df_daily[xcol].diff().dropna()
    plot_acf(s, lags=40, ax=axes[i,0], title=f'ACF: Δ{xcol}')
    plot_pacf(s, lags=40, ax=axes[i,1], title=f'PACF: Δ{xcol}')
fig.suptitle('Рисунок 2.9: ACF/PACF для перших різниць факторів',
             fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig_11_acf_pacf_factors.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_11_acf_pacf_factors.png")

```

```

# — Тест коінтеграції Енгла-Грейнджера —
print("\nТаблиця 2.3 — Тест на коінтеграцію (Wheat ~ Xi):")
coint_rows = []
for xcol in x_cols:

```

```

try:
    score, p, crits = coint(df_daily['Wheat'].dropna(),
                           df_daily[xcol].dropna())
    coint_rows.append({
        'Папа':      f'Wheat ~ {xcol}',
        't-stat':     round(score, 4),
        'p-value':    round(p, 4),
        'Крит. 1%':   round(crits[0], 4),
        'Крит. 5%':   round(crits[1], 4),
        'Коінтегровані?': 'Так ✓' if p < 0.05 else 'Ні ✗'
    })
except Exception as e:
    coint_rows.append({'Папа': f'Wheat ~ {xcol}',
                      'Помилка': str(e)})

coint_df = pd.DataFrame(coint_rows).set_index('Папа')
print(coint_df.to_string())
table_figure(coint_df,
             'Таблиця 2.3: Тест на коінтеграцію Енгла–Грейнджера',
             'fig_12_cointegration.png')

# — Рис. 12 — Ковзна кореляція (252 дні)
colors_x = [COLORS['brent'], COLORS['dxy'],
            COLORS['vix'],  COLORS['sp500']]
fig, axes = plt.subplots(2, 2, figsize=(14, 8))
for ax, xcol, color in zip(axes.flatten(), x_cols, colors_x):
    rc = (df_daily['Wheat'].pct_change()
          .rolling(252)
          .corr(df_daily[xcol].pct_change()))
    ax.plot(rc.index, rc, color=color, linewidth=1.2)
    ax.axhline(0, color='black', linewidth=0.8, linestyle='--')
    ax.fill_between(rc.index, rc, 0,
                    where=(rc > 0), alpha=0.2, color='green')
    ax.fill_between(rc.index, rc, 0,
                    where=(rc < 0), alpha=0.2, color='red')
    ax.set_title(f'Пшениця ~ {xcol}', fontweight='bold')
    ax.set_ylabel('Кореляція (252-денне вікно)')
    ax.set_ylim(-1, 1)
fig.suptitle('Рисунок 2.10: Ковзна кореляція (252 дні) Пшениця з факторами',
             fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig_13_rolling_correlation.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_13_rolling_correlation.png")

# — Рис. 13 — QQ-plot
fig, axes = plt.subplots(1, 2, figsize=(12, 5))
for ax, (series, title, color) in zip(axes, [
    (wheat, 'Рівень ціни',  COLORS['wheat']),
    (wheat_d1, 'Перша різниця',  COLORS['accent'])]):
    stats.probplot(series.dropna(), dist='norm', plot=ax)
    ax.get_lines()[0].set(markerfacecolor=color,
                          markersize=2, alpha=0.5)
    ax.get_lines()[1].set(color='red', linewidth=2)
    ax.set_title(f'QQ-Plot: Пшениця ({title})', fontweight='bold')
fig.suptitle('Рисунок 2.11: QQ-Plot — перевірка нормальності',
             fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig_14_qqplot.png', bbox_inches='tight')
plt.show()

```

```

print(" Збережено: fig_14_qqplot.png")

#

```

```

# 2.3 СТРУКТУРНІ ЗЛАМКИ, ДАМІ-ЗМІННІ, НОРМАЛІЗАЦІЯ
#

```

```

print("\n" + "=" * 70)
print("РОЗДІЛ 2.3 — СТРУКТУРНІ ЗЛАМКИ ТА ПІДГОТОВКА ДАНИХ")
print("=" * 70)

# — Тест Чоу

```

```

def chow_test(y, x, break_date):
    mask1 = y.index < break_date
    mask2 = y.index >= break_date
    if mask1.sum() < 10 or mask2.sum() < 10:
        return None, None
    X = add_constant(x)
    rss_full = OLS(y, X).fit().ssr
    rss1 = OLS(y[mask1], X[mask1]).fit().ssr
    rss2 = OLS(y[mask2], X[mask2]).fit().ssr
    k = X.shape[1]
    n = len(y)
    chow_f = ((rss_full - (rss1 + rss2)) / k) / \
        ((rss1 + rss2) / (n - 2 * k))
    p_val = 1 - stats.f.cdf(chow_f, k, n - 2 * k)
    return round(chow_f, 4), round(p_val, 4)

# Тестуємо тільки COVID та вторгнення (датасет з 2010, ГФК не охоплює)
break_dates = {
    'COVID (2020-03-11)': pd.Timestamp('2020-03-11'),
    'Вторгнення (2022-02-24)': pd.Timestamp('2022-02-24'),
}

wheat_clean = df_daily['Wheat'].dropna()
brent_clean = df_daily['Brent'].reindex(wheat_clean.index).ffill().dropna()
common_idx = wheat_clean.index.intersection(brent_clean.index)

chow_rows = []
for label, bd in break_dates.items():
    f_stat, p_val = chow_test(
        wheat_clean.loc[common_idx],
        brent_clean.loc[common_idx],
        bd
    )
    chow_rows.append({
        'Точка зламу': label,
        'F-статистика': f_stat,
        'p-value': p_val,
        'Структурний злам?': 'Так ✓' if (p_val and p_val < 0.05)
        else 'Ні ✗'
    })

chow_df = pd.DataFrame(chow_rows).set_index('Точка зламу')
print("\nТаблиця 2.4 — Тест Чоу:")
print(chow_df.to_string())
table_figure(chow_df,
    'Таблиця 2.4: Тест Чоу на структурні зламки',

```

```

'fig_15_chow_test.png')

# — Рис. 14 — Візуалізація зламів —————
fig, ax = plt.subplots(figsize=(14, 5))
ax.plot(df_daily.index, df_daily['Wheat'],
        color=COLORS['wheat'], linewidth=1, label='Ціна пшениці')
shock_lines = [
    ('2020-03-11', 'COVID\n11.03.2020', '#e67e22'),
    ('2022-02-24', 'Вторгнення\n24.02.2022', '#c0392b'),
]
for dt, label, color in shock_lines:
    ax.axvline(pd.Timestamp(dt), color=color,
               linewidth=2, linestyle='--', label=label)
ax.set_ylabel('Ціна пшениці (центи/бушель)')
ax.set_title('Рисунок 2.12: Точки структурних зламів '
             'на ряді цін пшениці (2010–2026)',
             fontweight='bold')
ax.legend(loc='upper left')
ax.xaxis.set_major_locator(mdates.YearLocator(2))
ax.xaxis.set_major_formatter(mdates.DateFormatter('%Y'))
plt.tight_layout()
plt.savefig('fig_16_structural_breaks.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_16_structural_breaks.png")

# — Дамі-змінні (тільки COVID та вторгнення) —————
df_model = df_daily.copy()

df_model['d_covid'] = (
    (df_model.index >= '2020-02-01') &
    (df_model.index <= '2020-09-30')
).astype(int)

df_model['d_war'] = (
    df_model.index >= '2022-02-24'
).astype(int)

dummy_counts = df_model[['d_covid', 'd_war']].sum()
print("\nТаблиця 2.5 — Кількість спостережень = 1 по дамі-змінних:")
print(dummy_counts.to_frame('Кількість').to_string())

# — Рис. 15 — Дамі-змінні —————
fig, axes = plt.subplots(3, 1, figsize=(14, 8), sharex=True)

axes[0].plot(df_model.index, df_model['Wheat'],
             color=COLORS['wheat'], linewidth=0.9)
axes[0].set_ylabel('Центи/бушель')
axes[0].set_title('Ціна пшениці (2010–2026)', fontweight='bold')

for ax, (dcol, dname, dcolor) in zip(axes[1:], [
    ('d_covid', 'D1: COVID 2020', '#e67e22'),
    ('d_war', 'D2: Вторгнення 2022', '#c0392b'),
]):
    ax.fill_between(df_model.index, df_model[dcol],
                   step='mid', alpha=0.75, color=dcolor)
    ax.set_ylabel(dname, fontsize=9)
    ax.set_ylim(-0.05, 1.2)
    ax.set_yticks([0, 1])
    ax.set_yticklabels(['0', '1'])

axes[-1].xaxis.set_major_locator(mdates.YearLocator(2))

```

```

axes[-1].xaxis.set_major_formatter(mdates.DateFormatter('%Y'))
fig.suptitle('Рисунок 2.13: Дамі-змінні структурних шоків (2010–2026)',
            fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig_17_dummy_variables.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_17_dummy_variables.png")

# — Нормалізація

```

```

feature_cols = ['Wheat', 'Brent', 'DXY', 'VIX', 'SP500']
df_features = df_model[feature_cols].copy()

scaler_mm = MinMaxScaler(feature_range=(0, 1))
df_minmax = pd.DataFrame(
    scaler_mm.fit_transform(df_features),
    index=df_features.index, columns=df_features.columns
)

scaler_std = StandardScaler()
df_standard = pd.DataFrame(
    scaler_std.fit_transform(df_features),
    index=df_features.index, columns=df_features.columns
)

print("\nТаблиця 2.6 — Параметри Min-Max нормалізації:")
scaling_params = pd.DataFrame({
    'Мін (оригінал)': scaler_mm.data_min_,
    'Макс (оригінал)': scaler_mm.data_max_,
    'Діапазон': scaler_mm.data_max_ - scaler_mm.data_min_,
}, index=feature_cols)
print(scaling_params.to_string())

# — Рис. 16 — Порівняння нормалізацій

```

```

fig, axes = plt.subplots(3, 1, figsize=(14, 10))

for col, color in zip(feature_cols, list(COLORS.values())[:5]):
    norm = ((df_features[col] - df_features[col].min()) /
            (df_features[col].max() - df_features[col].min()))
    axes[0].plot(df_features.index, norm,
                 linewidth=0.7, label=col, alpha=0.85)
axes[0].set_title('Вихідні дані (ручна нормалізація для відображення)',
                  fontweight='bold')
axes[0].legend(ncol=5, fontsize=8)

for col in feature_cols:
    axes[1].plot(df_minmax.index, df_minmax[col],
                 linewidth=0.7, label=col, alpha=0.85)
axes[1].set_title('Після Min-Max нормалізації [0, 1]', fontweight='bold')
axes[1].legend(ncol=5, fontsize=8)

for col in feature_cols:
    axes[2].plot(df_standard.index, df_standard[col],
                 linewidth=0.7, label=col, alpha=0.85)
axes[2].axhline(0, color='black', linewidth=0.5, linestyle='--')
axes[2].set_title('Після Z-score стандартизації ( $\mu=0$ ,  $\sigma=1$ )', fontweight='bold')
axes[2].legend(ncol=5, fontsize=8)

for ax in axes:
    ax.xaxis.set_major_locator(mdates.YearLocator(2))
    ax.xaxis.set_major_formatter(mdates.DateFormatter('%Y'))

```

```

fig.suptitle('Рисунок 2.14: Порівняння методів нормалізації даних',
             fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig_18_normalization.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig_18_normalization.png")

# — Збереження фінального датасету —————
df_final_diff = df_features.diff()
df_final_diff.columns = [f'd_{c}' for c in feature_cols]
df_export = pd.concat([
    df_model,
    df_final_diff,
    df_minmax.add_suffix('_mm'),
    df_standard.add_suffix('_std')
], axis=1)
df_export.to_csv('dataset_wheat_analysis.csv')
print(f"\n Збережено: dataset_wheat_analysis.csv")
print(f" Розмір: {df_export.shape[0]} рядків × {df_export.shape[1]} колонок")

#
—————

for fname, desc in files:
    print(f" {fname:<42} → {desc}")
print(f"\n Всього: {len(files)} файлів")
print("=" * 70)
print("РОЗДІЛ 2 ЗАВЕРШЕНО УСПІШНО")
print("=" * 70)
#
—————

# ІМПОРТИ
#
—————

import warnings
warnings.filterwarnings('ignore')

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.dates as mdates
import seaborn as sns
from scipy import stats
from scipy.stats import t as t_dist
from itertools import product

import yfinance as yf
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.tsa.api import VAR
from statsmodels.tsa.stattools import adfuller, grangercausalitytests
from statsmodels.stats.diagnostic import acorr_ljungbox
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

from sklearn.preprocessing import MinMaxScaler
from sklearn.neural_network import MLPRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error

try:
    import tensorflow as tf
    from tensorflow.keras.models import Sequential

```

```

from tensorflow.keras.layers import LSTM, Dense, Dropout
from tensorflow.keras.callbacks import EarlyStopping, ReduceLROnPlateau
from tensorflow.keras.optimizers import Adam
tf.random.set_seed(42)
TF_OK = True
print(f" TensorFlow {tf.__version__} знайдено")
except ImportError:
    TF_OK = False
    print(" TensorFlow не знайдено. LSTM буде пропущено.")
    print(" pip install tensorflow")

np.random.seed(42)

#


---


# СТИЛЬ
#


---


plt.rcParams.update({
    'figure.dpi': 150, 'figure.facecolor': 'white',
    'axes.facecolor': '#f8f9fa', 'axes.grid': True,
    'grid.alpha': 0.4, 'font.size': 10,
    'axes.titlesize': 11, 'axes.titleweight': 'bold',
    'lines.linewidth': 1.5,
})
C = {
    'actual': '#2c3e50', 'arima': '#e74c3c',
    'var': '#2980b9', 'mlp': '#27ae60',
    'lstm': '#8e44ad', 'train': '#bdc3c7',
}

TRAIN_END = '2022-12-31'
TEST_START = '2023-01-01'
LOOKBACK = 60 # вікно LSTM/MLP (торгових днів)
FEAT_COLS = ['Wheat', 'Brent', 'DXY', 'VIX', 'SP500']

#


---


# ЗАВАНТАЖЕННЯ ДАНИХ
#


---


print("=" * 65)
print("ЗАВАНТАЖЕННЯ ДАНИХ")
print("=" * 65)

try:
    df_raw = pd.read_csv('dataset_wheat_analysis.csv', index_col=0, parse_dates=True)
    print(" dataset_wheat_final.pkl — завантажено")
except FileNotFoundError:
    print(" pkl не знайдено — завантажуюмо з Yahoo Finance...")
    TICKERS = {'Wheat': 'ZW=F', 'Brent': 'BZ=F',
               'DXY': 'DX-Y.NYB', 'VIX': '^VIX', 'SP500': '^GSPC'}
    raw = {}
    for name, tkr in TICKERS.items():
        tmp = yf.download(tkr, start="2010-01-01",
                          end="2026-03-30", progress=False, auto_adjust=True)
        raw[name] = tmp['Close'].squeeze()
    df_raw = pd.DataFrame(raw).ffill().dropna()

```

```

df_raw.index = pd.to_datetime(df_raw.index)
df_raw['d_covid'] = ((df_raw.index>='2020-02-01') &
                    (df_raw.index<='2020-09-30')).astype(int)
df_raw['d_war'] = (df_raw.index >= '2022-02-24').astype(int)

# Обираємо тільки потрібні колонки
ALL_COLS = FEAT_COLS + ['d_covid', 'd_war']
df = df_raw[ALL_COLS].copy()

train = df[df.index <= TRAIN_END].copy()
test = df[df.index >= TEST_START].copy()

wheat_train = train['Wheat']
wheat_test = test['Wheat']

print(f" Train: {train.index.min().date()} → {train.index.max().date()} ({len(train)} спост.)")
print(f" Test: {test.index.min().date()} → {test.index.max().date()} ({len(test)} спост.)")

#

```

```

# ДОПОМІЖНІ ФУНКЦІЇ
#

```

```

def calc_metrics(actual, predicted):
    """MAE, RMSE, MAPE — NaN-safe."""
    a = np.array(actual).flatten()
    p = np.array(predicted).flatten()
    n = min(len(a), len(p))
    a, p = a[:n], p[:n]
    # Прибираємо NaN з обох векторів одночасно
    mask = ~(np.isnan(a) | np.isnan(p)) & (a != 0)
    if mask.sum() == 0:
        return {'MAE': np.nan, 'RMSE': np.nan, 'MAPE': np.nan}
    a_m, p_m = a[mask], p[mask]
    mae = mean_absolute_error(a_m, p_m)
    rmse = np.sqrt(mean_squared_error(a_m, p_m))
    mape = np.mean(np.abs((a_m - p_m) / a_m)) * 100
    return {'MAE': round(mae, 4),
            'RMSE': round(rmse, 4),
            'MAPE': round(mape, 4)}

def diebold_mariano(e1, e2):
    """Тест Диболда–Маріано. H0: однакова точність."""
    e1 = np.array(e1).flatten()
    e2 = np.array(e2).flatten()
    n = min(len(e1), len(e2))
    e1, e2 = e1[:n], e2[:n]
    mask = ~(np.isnan(e1) | np.isnan(e2))
    e1, e2 = e1[mask], e2[mask]
    if len(e1) < 5:
        return np.nan, np.nan
    d = e1**2 - e2**2
    d_mean = np.mean(d)
    var_d = np.var(d, ddof=1) / len(d)
    if var_d <= 0:
        return np.nan, np.nan
    dm = d_mean / np.sqrt(var_d)
    p = 2 * (1 - t_dist.cdf(abs(dm), df=len(d)-1))
    return round(dm, 4), round(p, 4)

```

```

def table_fig(df_in, title, filename):
    """DataFrame → красивий рисунок-таблиця."""
    fig, ax = plt.subplots(figsize=(14, max(2.2, len(df_in)*0.6+1.2)))
    ax.axis('off')
    tbl = ax.table(
        cellText=df_in.values, colLabels=df_in.columns,
        rowLabels=df_in.index, cellLoc='center', loc='center'
    )
    tbl.auto_set_font_size(False)
    tbl.set_fontsize(9)
    tbl.scale(1, 1.55)
    for (r, c), cell in tbl.get_celld().items():
        if r == 0 or c == -1:
            cell.set_facecolor('#2c3e50')
            cell.set_text_props(color='white', fontweight='bold')
        else:
            val = cell.get_text().get_text()
            if 'Так' in val:
                cell.set_facecolor('#d5f5d5')
            elif 'Hi' in val:
                cell.set_facecolor('#fde8e8')
            else:
                cell.set_facecolor('#f8f9fa' if r % 2 == 0 else 'white')
    ax.set_title(title, fontsize=11, fontweight='bold', pad=14)
    plt.tight_layout()
    plt.savefig(filename, bbox_inches='tight')
    plt.show()
    print(f" Збережено: {filename}")

def forecast_plot(train_s, test_s, preds, model_name, color,
                  filename, ci_lo=None, ci_hi=None):
    """Стандартний графік прогнозу з залишками."""
    n = min(len(test_s), len(preds))
    fig, axes = plt.subplots(2, 1, figsize=(14, 8),
                             gridspec_kw={'height_ratios': [3, 1]})
    ax = axes[0]
    ax.plot(train_s.index[-300:], train_s.values[-300:],
            color=C['train'], lw=0.8, alpha=0.7, label='Тренувальний ряд')
    ax.plot(test_s.index, test_s.values,
            color=C['actual'], lw=1.8, label='Фактична ціна')
    ax.plot(test_s.index[:n], np.array(preds[:n]),
            color=color, lw=2, ls='--',
            label=f'Прогноз {model_name}')
    if ci_lo is not None:
        ax.fill_between(test_s.index[:n],
                        ci_lo[:n], ci_hi[:n],
                        alpha=0.15, color=color, label='95% ДІ')
    ax.axvline(test_s.index[0], color='black', lw=1.2,
               ls=':', alpha=0.6, label='Початок тесту')
    ax.set_ylabel('Ціна пшениці (центи/бушель)')
    ax.set_title(f'Прогноз {model_name} — тестова вибірка 2023–2026',
                 fontweight='bold')
    ax.legend(loc='upper left', fontsize=9)
    ax.xaxis.set_major_formatter(mdates.DateFormatter('%m/%Y'))

    ax2 = axes[1]
    resid = test_s.values[:n] - np.array(preds[:n])
    ax2.bar(test_s.index[:n], resid, color=color, alpha=0.5, width=1)
    ax2.axhline(0, color='black', lw=0.8)
    ax2.set_ylabel('Залишки')
    ax2.xaxis.set_major_formatter(mdates.DateFormatter('%m/%Y'))

```

```

plt.tight_layout()
plt.savefig(filename, bbox_inches='tight')
plt.show()
print(f" Збережено: {filename}")

#

```

```

# 3.1a — ARIMA
#

```

```

print("\n" + "=" * 65)
print("3.1a ARIMA")
print("=" * 65)

wheat_d1 = wheat_train.diff().dropna()

# — Рис. ACF/PACF для визначення p, q
fig, axes = plt.subplots(2, 2, figsize=(14, 8))
plot_acf(wheat_train, lags=40, ax=axes[0,0],
         title='ACF: Рівень ціни пшениці')
plot_pacf(wheat_train, lags=40, ax=axes[0,1],
         title='PACF: Рівень ціни пшениці')
plot_acf(wheat_d1, lags=40, ax=axes[1,0],
         title='ACF: Перша різниця (стаціонарна)')
plot_pacf(wheat_d1, lags=40, ax=axes[1,1],
         title='PACF: Перша різниця')
fig.suptitle('Рисунок 3.1: ACF/PACF для ідентифікації ARIMA',
            fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig3_01_acf_pacf.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig3_01_acf_pacf.png")

# — Підбір (p, d=1, q) за AIC
print("\n Підбір ARIMA за AIC (p ∈ 0..4, q ∈ 0..4) ...")
arima_grid = []
for p, q in product(range(5), range(5)):
    try:
        m = ARIMA(wheat_train, order=(p, 1, q)).fit(
            method='innovations_mle',
            cov_type='none'
        )
        arima_grid.append({
            'p': p, 'd': 1, 'q': q,
            'AIC': round(m.aic, 2),
            'BIC': round(m.bic, 2),
        })
    except Exception:
        pass

aic_df = (pd.DataFrame(arima_grid)
         .sort_values('AIC')
         .reset_index(drop=True))
print(aic_df.head(10).to_string(index=False))

best_p = int(aic_df.iloc[0]['p'])
best_q = int(aic_df.iloc[0]['q'])
print(f"\n Найкраща модель: ARIMA({best_p},1,{best_q}) "
      f"AIC={aic_df.iloc[0]['AIC']}")

```

```

# Таблиця топ-10 AIC
aic_show = aic_df.head(10).copy()
aic_show.insert(0, 'Модель',
    aic_show.apply(lambda r: f"ARIMA({int(r.p)},1,{int(r.q)})", axis=1))
aic_show = aic_show[['Модель','AIC','BIC']].set_index('Модель')
table_fig(aic_show,
    'Таблиця 3.1: Порівняння специфікацій ARIMA за AIC',
    'fig3_02_arima_aic.png')

# — Фінальна модель


---


print(f'\n Навчання ARIMA({best_p},1,{best_q}) ...')
arima_final = ARIMA(wheat_train, order=(best_p, 1, best_q)).fit(
    method='innovations_mle', cov_type='robust'
)
print(arima_final.summary())

# Параметри як таблиця
coef_df = pd.DataFrame({
    'Коефіцієнт': arima_final.params.round(4),
    'Ст. похибка': arima_final.bse.round(4),
    't-stat': arima_final.tvalues.round(4),
    'p-value': arima_final.pvalues.round(4),
    'Значущий?': ['Так ✓' if p < 0.05 else 'Ні ✗'
        for p in arima_final.pvalues]
}, index=arima_final.param_names)
table_fig(coef_df,
    f'Таблиця 3.2: Параметри ARIMA({best_p},1,{best_q})',
    'fig3_03_arima_params.png')

# — Діагностика залишків ARIMA


---


resid_a = arima_final.resid.dropna()
lb_a = acorr_ljungbox(resid_a, lags=range(1, 21), return_df=True)

fig, axes = plt.subplots(2, 3, figsize=(16, 8))

axes[0,0].plot(resid_a.index, resid_a.values,
    color=C['arima'], lw=0.7, alpha=0.8)
axes[0,0].axhline(0, color='black', lw=0.8, ls='--')
axes[0,0].set_title('Залишки ARIMA в часі')

axes[0,1].hist(resid_a, bins=50, color=C['arima'],
    alpha=0.55, density=True)
resid_a.plot.kde(ax=axes[0,1], color='darkred', lw=2)
axes[0,1].set_title('Розподіл залишків')

stats.probplot(resid_a, dist='norm', plot=axes[0,2])
axes[0,2].get_lines()[0].set(markerfacecolor=C['arima'],
    markersize=2, alpha=0.4)
axes[0,2].set_title('QQ-Plot залишків')

plot_acf(resid_a, lags=40, ax=axes[1,0],
    title='ACF залишків (має бути "білий шум")')
plot_pacf(resid_a, lags=40, ax=axes[1,1],
    title='PACF залишків')

axes[1,2].bar(lb_a.index, lb_a['lb_pvalue'],
    color=[C['arima'] if p > 0.05 else 'red'
        for p in lb_a['lb_pvalue']], alpha=0.75)
axes[1,2].axhline(0.05, color='red', lw=1.5, ls='--', label='α=0.05')
axes[1,2].set_title('Ljung-Box: p-value по лагах')

```

```

axes[1,2].set_xlabel('Лар')
axes[1,2].legend()

fig.suptitle(f'Рисунок 3.2: Діагностика залишків ARIMA({best_p},1,{best_q})',
             fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig3_04_arima_diag.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig3_04_arima_diag.png")

# Ljung-Box таблиця
lb_show = lb_a.round(4).copy()
lb_show.index.name = 'Лар'
lb_show.columns = ['LB-статистика', 'p-value']
lb_show['Білий шум?'] = lb_show['p-value'].apply(
    lambda p: 'Так ✓' if p > 0.05 else 'Ні ✗')
table_fig(lb_show,
           'Таблиця 3.3: Тест Льюнга–Бокса (ARIMA залишки)',
           'fig3_05_ljungbox.png')
# ДІАГНОСТИКА rolling forecast
print("\n Тест rolling на 1 кроці:")
try:
    m_test = ARIMA(list(wheat_train.values), order=(best_p, 1, best_q)).fit(
        method='innovations_mle'
    )
    fc_test = m_test.get_forecast(steps=1)
    print(f" predicted_mean: {fc_test.predicted_mean}")
    print(f" conf_int: {fc_test.conf_int(alpha=0.05)}")
    print(f" iloc[0]: {float(fc_test.predicted_mean.iloc[0])}")
except Exception as e:
    import traceback
    traceback.print_exc()

# — Rolling forecast ARIMA — NaN-safe —————
print(f"\n Rolling forecast ARIMA({best_p},1,{best_q}) ...")
print(" (1 крок вперед, перенавчання на кожному кроці)")

arima_preds = []
arima_lo = []
arima_hi = []
history_a = list(wheat_train.values)

print(f"\n Rolling forecast ARIMA({best_p},1,{best_q}) ...")
print(" (1 крок вперед, перенавчання на кожному кроці)")

arima_preds = []
arima_lo = []
arima_hi = []
history_a = list(wheat_train.values)

for i in range(len(wheat_test)):
    pred_val = np.nan
    lo_val = np.nan
    hi_val = np.nan
    try:
        m_roll = ARIMA(history_a, order=(best_p, 1, best_q)).fit(
            method='innovations_mle'
        )
        fc = m_roll.get_forecast(steps=1)
        mn = fc.predicted_mean

```

```

ci = fc.conf_int(alpha=0.05)
pred_val = float(mn[0])
lo_val = float(ci[0, 0])
hi_val = float(ci[0, 1])
if not (0 < pred_val < 10_000):
    pred_val = lo_val = hi_val = np.nan
except Exception:
    pass

arima_preds.append(pred_val)
arima_lo.append(lo_val)
arima_hi.append(hi_val)
history_a.append(float(wheat_test.values[i]))

if (i + 1) % 200 == 0:
    n_ok = sum(1 for v in arima_preds if not np.isnan(v))
    print(f" [{i+1}/{len(wheat_test)}] "
          f"прогнозів ОК: {n_ok} NaN: {i+1-n_ok}")

n_nan_a = sum(1 for v in arima_preds if np.isnan(v))
print(f" Готово. NaN у прогнозі: {n_nan_a}/{len(arima_preds)}")

arima_metrics = calc_metrics(wheat_test.values, arima_preds)
print(f" ARIMA: MAE={arima_metrics['MAE']} "
      f"RMSE={arima_metrics['RMSE']} "
      f"MAPE={arima_metrics['MAPE']}%")

forecast_plot(wheat_train, wheat_test, arima_preds,
              f"ARIMA({best_p},1,{best_q})", C['arima'],
              'fig3_06_arima_forecast.png',
              ci_lo=arima_lo, ci_hi=arima_hi)

np.save('arima_preds.npy', np.array(arima_preds))
print("\n ДІАГНОСТИКА витоку даних:")
print(f" Перше фактичне значення тесту: {wheat_test.values[0]:.2f}")
print(f" Перший прогноз ARIMA: {arima_preds[0]:.2f}")
print(f" Різниця: {abs(wheat_test.values[0] - arima_preds[0]):.2f}")
print(f" Останнє тренувальне значення: {wheat_train.values[-1]:.2f}")
print(f"\n Останні 3 дати train: {wheat_train.index[-3:]}"
      f"Перші 3 дати test: {wheat_test.index[:3]}")
print(f" Перетин дат: {len(wheat_train.index.intersection(wheat_test.index))}")
#

```

```

# 3.16 — VAR
#

```

```

print("\n" + "=" * 65)
print("3.16 VAR")
print("=" * 65)

# VAR — перші різниці (стаціонарні ряди)
var_cols = FEAT_COLS + ['d_covid', 'd_war']
var_train = train[var_cols].diff().dropna()
var_test = test[var_cols].diff().dropna()

# — Тест Грейнджера

```

```

print("\n Тест Грейнджера (X → Wheat, 5 лагів):")
gr_rows = []
for xcol in ['Brent', 'DXY', 'VIX', 'SP500']:

```

```

try:
    res = grangercausalitytests(
        var_train[['Wheat', xcol]], maxlag=5, verbose=False)
    best_p_gr = min(res[l][0]['ssr_ftest'][1] for l in range(1, 6))
    gr_rows.append({
        'X → Wheat': xcol,
        'Мін p-value': round(best_p_gr, 4),
        'Причинність?': 'Так ✓' if best_p_gr < 0.05 else 'Ні X'
    })
except Exception as e:
    gr_rows.append({'X → Wheat': xcol,
        'Мін p-value': 'N/A',
        'Причинність?': str(e)})

gr_df = pd.DataFrame(gr_rows).set_index('X → Wheat')
print(gr_df.to_string())
table_fig(gr_df,
    'Таблиця 3.4: Тест Грейнджера (X → Wheat)',
    'fig3_07_granger.png')

```

— Вибір лагу VAR за AIC

```

print("\n Підбір лагу VAR ...")
lag_rows = []
for lag in range(1, 11):
    try:
        m_lag = VAR(var_train).fit(lag)
        lag_rows.append({
            'Лар': lag,
            'AIC': round(m_lag.aic, 4),
            'BIC': round(m_lag.bic, 4),
            'HQIC': round(m_lag.hqic, 4),
        })
    except Exception:
        pass

lag_df = pd.DataFrame(lag_rows).set_index('Лар')
best_lag = int(lag_df['AIC'].idxmin())
lag_df['AIC оптим.?'] = lag_df.index.map(
    lambda x: 'Так ✓' if x == best_lag else '')
print(lag_df.to_string())
print(f"\n Оптимальний лаг VAR: {best_lag}")
table_fig(lag_df,
    'Таблиця 3.5: Критерії вибору лагу VAR',
    'fig3_08_var_lag.png')

```

— Навчання VAR

```

print(f"\n Навчання VAR({best_lag}) ...")
var_fitted = VAR(var_train).fit(best_lag)
print(var_fitted.summary())

```

— IRF

```

print("\n IRF — функції імпульсного відгуку ...")
try:
    irf = var_fitted.irf(periods=20)
    wheat_idx = var_cols.index('Wheat')
    imp_names = ['Brent', 'DXY', 'VIX', 'SP500']

```

```

fig, axes = plt.subplots(2, 2, figsize=(14, 9))
for ax, imp in zip(axes.flatten(), imp_names):
    imp_idx = var_cols.index(imp)
    vals = irf.irfs[:, wheat_idx, imp_idx]
    ax.plot(range(21), vals, color=C['var'], lw=2,
            marker='o', markersize=3)
    ax.axhline(0, color='black', lw=0.8, ls='--')
    ax.set_title(f'Відгук Wheat на шок {imp}', fontweight='bold')
    ax.set_xlabel('Горизонт (дні)')
    ax.set_ylabel('Відгук (центи/бушель)')
fig.suptitle('Рисунок 3.3: IRF — реакція ціни пшениці на шоки факторів',
            fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig3_09_var_irf.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig3_09_var_irf.png")
except Exception as e:
    print(f" IRF: {e}")

```

— FEVD

```

try:
    fevd = var_fitted.fevd(periods=20)
    fevd_df = pd.DataFrame(
        fevd.decomp[:, wheat_idx, :],
        columns=var_cols
    )
    fig, ax = plt.subplots(figsize=(12, 5))
    fevd_df[FEAT_COLS].plot.area(ax=ax, colormap='tab10', alpha=0.8)
    ax.set_xlabel('Горизонт (дні)')
    ax.set_ylabel('Частка дисперсії')
    ax.set_title('Рисунок 3.4: FEVD — декомпозиція дисперсії (Wheat)',
                fontweight='bold')
    ax.legend(loc='upper right', ncol=3, fontsize=9)
    plt.tight_layout()
    plt.savefig('fig3_10_var_fevd.png', bbox_inches='tight')
    plt.show()
    print(" Збережено: fig3_10_var_fevd.png")
except Exception as e:
    print(f" FEVD: {e}")

```

— Rolling forecast VAR

```

print(f"\n Rolling forecast VAR({best_lag}) ...")
var_preds_diff = []
hist_var = var_train.values.tolist()

for i in range(len(var_test)):
    try:
        m_roll = VAR(np.array(hist_var)).fit(best_lag)
        fc_arr = m_roll.forecast(
            np.array(hist_var[-best_lag:]), steps=1)
        var_preds_diff.append(fc_arr[0, 0])
    except Exception:
        var_preds_diff.append(np.nan)
    hist_var.append(var_test.values[i].tolist())
    if (i + 1) % 200 == 0:
        print(f" [{i+1}/{len(var_test)}]")

```

```

# Різниця → рівні
last_known = float(wheat_train.values[-1])
var_preds_level = []

```

```

prev = last_known
for d in var_preds_diff:
    if np.isnan(d):
        var_preds_level.append(np.nan)
        # prev не змінюємо — зберігаємо ланцюжок
    else:
        val = prev + d
        if 0 < val < 10_000:
            var_preds_level.append(val)
            prev = val
        else:
            var_preds_level.append(np.nan)

n_var = min(len(var_preds_level), len(wheat_test))
var_metrics = calc_metrics(wheat_test.values[:n_var],
                           var_preds_level[:n_var])
print(f" VAR: MAE={var_metrics['MAE']} "
      f"RMSE={var_metrics['RMSE']} "
      f"MAPE={var_metrics['MAPE']}%")

forecast_plot(wheat_train, wheat_test.iloc[:n_var],
              var_preds_level[:n_var],
              f'VAR({best_lag})', C['var'],
              'fig3_11_var_forecast.png')

np.save('var_preds.npy', np.array(var_preds_level))

#


---


# 3.2a — MLP
#


---



print("\n" + "=" * 65)
print("3.2a MLP")
print("=" * 65)

ML_COLS = FEAT_COLS + ['d_covid', 'd_war']
w_idx = ML_COLS.index('Wheat')

# — Нормалізація


---



scaler_mlp = MinMaxScaler(feature_range=(0, 1))
tr_sc = scaler_mlp.fit_transform(train[ML_COLS])
te_sc = scaler_mlp.transform(test[ML_COLS])

def make_windows(data, lb, y_idx):
    X, y = [], []
    for i in range(lb, len(data)):
        X.append(data[i-lb:i, :].flatten())
        y.append(data[i, y_idx])
    return np.array(X), np.array(y)

X_tr_m, y_tr_m = make_windows(tr_sc, LOOKBACK, w_idx)
X_te_m, y_te_m = make_windows(
    np.vstack([tr_sc[-LOOKBACK:], te_sc]), LOOKBACK, w_idx)

print(f" MLP вхід: train={X_tr_m.shape} test={X_te_m.shape}")

# Архітектура
mlp_arch = pd.DataFrame([

```

```

['Вхідний шар', f'{X_tr_m.shape[1]} ознак', '—', '—'],
['Прихований 1', '256 нейронів', 'ReLU', 'Dropout(0.2)'],
['Прихований 2', '128 нейронів', 'ReLU', 'Dropout(0.2)'],
['Прихований 3', '64 нейрони', 'ReLU', 'Dropout(0.1)'],
['Вихідний шар', '1 нейрон', 'Linear', '—'],
], columns=['Шар', 'Розмір', 'Активация', 'Регуляризация'])
mlp_arch = mlp_arch.set_index('Шар')
table_fig(mlp_arch, 'Таблиця 3.6: Архітектура MLP',
          'fig3_12_mlp_arch.png')

# — Навчання

mlp_model = MLPRegressor(
    hidden_layer_sizes=(256, 128, 64),
    activation='relu', solver='adam',
    max_iter=500, random_state=42,
    early_stopping=True,
    validation_fraction=0.15,
    n_iter_no_change=25,
    learning_rate_init=0.001,
    alpha=0.0001,
    verbose=False
)
print(" Навчання MLP ...")
try:
    mlp_model.fit(X_tr_m, y_tr_m)
    print(f" MLP навчання ОК")
except Exception as e:
    import traceback
    traceback.print_exc()
n_iter = mlp_model.n_iter_ if mlp_model.n_iter_ is not None else 0
best_loss = mlp_model.best_loss_ if mlp_model.best_loss_ is not None else float('nan')
print(f" MLP: {n_iter} ітерацій best_loss={best_loss:.6f}")

# Крива навчання
if hasattr(mlp_model, 'loss_curve_'):
    fig, ax = plt.subplots(figsize=(10, 4))
    ax.plot(mlp_model.loss_curve_, color=C['mlp'], lw=1.5, label="Train loss")
    ax.set_yscale('log')
    ax.set_xlabel('Ітерація')
    ax.set_ylabel('MSE (log)')
    ax.set_title('Рисунок 3.5: Крива навчання MLP', fontweight='bold')
    ax.legend()
    plt.tight_layout()
    plt.savefig('fig3_13_mlp_loss.png', bbox_inches='tight')
    plt.show()
    print(" Збережено: fig3_13_mlp_loss.png")

# — Прогноз MLP → зворотна нормалізація
mlp_sc = mlp_model.predict(X_te_m)
inv = np.zeros((len(mlp_sc), len(ML_COLS)))
inv[:, w_idx] = mlp_sc
mlp_preds = scaler_mlp.inverse_transform(inv[:, w_idx])

n_mlp = min(len(mlp_preds), len(wheat_test))
mlp_metrics = calc_metrics(wheat_test.values[:n_mlp], mlp_preds[:n_mlp])
print(f" MLP: MAE={mlp_metrics['MAE']} "
      f"RMSE={mlp_metrics['RMSE']} "
      f"MAPE={mlp_metrics['MAPE']}%")

forecast_plot(wheat_train, wheat_test, mlp_preds,
              'MLP', C['mlp'], 'fig3_14_mlp_forecast.png')

```

```

np.save('mlp_preds.npy', mlp_preds)

#

```

```

# 3.26 — LSTM
#

```

```

print("\n" + "=" * 65)
print("3.26 LSTM")
print("=" * 65)

if not TF_OK:
    print(" TensorFlow не встановлено — LSTM пропущено")
    lstm_preds = np.full(len(wheat_test), np.nan)
    lstm_metrics = {'MAE': np.nan, 'RMSE': np.nan, 'MAPE': np.nan}
else:
    scaler_lstm = MinMaxScaler(feature_range=(0, 1))
    tr_sc_l = scaler_lstm.fit_transform(train[ML_COLS])
    te_sc_l = scaler_lstm.transform(test[ML_COLS])

    def make_windows_3d(data, lb, y_idx):
        X, y = [], []
        for i in range(lb, len(data)):
            X.append(data[i-lb:i, :])
            y.append(data[i, y_idx])
        return np.array(X), np.array(y)

    X_tr_l, y_tr_l = make_windows_3d(tr_sc_l, LOOKBACK, w_idx)
    X_te_l, y_te_l = make_windows_3d(
        np.vstack([tr_sc_l[-LOOKBACK:], te_sc_l]), LOOKBACK, w_idx)

    print(f" LSTM вхід: train={X_tr_l.shape} test={X_te_l.shape}")

# Архітектура
lstm_arch = pd.DataFrame([
    ['Input', f(batch, {LOOKBACK}, {len(ML_COLS)})], ['—', '—'],
    ['LSTM шар 1', '128 units, return_seq=True', 'tanh', 'Dropout(0.2)'],
    ['LSTM шар 2', '64 units, return_seq=True', 'tanh', 'Dropout(0.2)'],
    ['LSTM шар 3', '32 units, return_seq=False', 'tanh', 'Dropout(0.1)'],
    ['Dense', '32 нейрони', 'ReLU', '—'],
    ['Output Dense', '1 нейрон', 'Linear', '—'],
], columns=['Шар', 'Параметри', 'Активация', 'Регуляризация'])
lstm_arch = lstm_arch.set_index('Шар')
table_fig(lstm_arch, 'Таблиця 3.7: Архітектура LSTM',
    'fig3_15_lstm_arch.png')

# Модель
lstm_model = Sequential([
    LSTM(128, return_sequences=True,
        input_shape=(LOOKBACK, len(ML_COLS))),
    Dropout(0.2),
    LSTM(64, return_sequences=True),
    Dropout(0.2),
    LSTM(32, return_sequences=False),
    Dropout(0.1),
    Dense(32, activation='relu'),
    Dense(1)
])
lstm_model.compile(
    optimizer=Adam(learning_rate=0.001),

```

```

    loss='mse', metrics=['mae']
)
lstm_model.summary()

cbs = [
    EarlyStopping(monitor='val_loss', patience=15,
                  restore_best_weights=True, verbose=1),
    ReduceLROnPlateau(monitor='val_loss', factor=0.5,
                     patience=7, min_lr=1e-6, verbose=0)
]

print("\n Навчання LSTM ...")
hist_lstm = lstm_model.fit(
    X_tr_l, y_tr_l,
    epochs=100, batch_size=32,
    validation_split=0.15,
    callbacks=cbs, verbose=1,
    shuffle=False
)

# Крива навчання
best_ep = int(np.argmax(hist_lstm.history['val_loss'])) + 1
fig, axes = plt.subplots(1, 2, figsize=(13, 4))
axes[0].plot(hist_lstm.history['loss'],
             color=C['lstm'], lw=1.5, label='Train')
axes[0].plot(hist_lstm.history['val_loss'],
             color='orange', lw=1.5, ls='--', label='Val')
axes[0].axvline(best_ep, color='red', lw=1, ls=':',
               label=f'Best={best_ep}')
axes[0].set_yscale('log')
axes[0].set_xlabel('Епоха')
axes[0].set_ylabel('MSE (log)')
axes[0].set_title('Крива навчання LSTM', fontweight='bold')
axes[0].legend()

axes[1].plot(hist_lstm.history['mae'],
             color=C['lstm'], lw=1.5, label='Train MAE')
axes[1].plot(hist_lstm.history['val_mae'],
             color='orange', lw=1.5, ls='--', label='Val MAE')
axes[1].set_xlabel('Епоха')
axes[1].set_ylabel('MAE')
axes[1].set_title('MAE під час навчання', fontweight='bold')
axes[1].legend()

fig.suptitle('Рисунок 3.6: Криві навчання LSTM '
            '(Early Stopping + ReduceLROnPlateau)',
            fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig3_16_lstm_training.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig3_16_lstm_training.png")

# Прогноз
lstm_sc = lstm_model.predict(X_te_l, verbose=0).flatten()
inv_l = np.zeros((len(lstm_sc), len(ML_COLS)))
inv_l[:, w_idx] = lstm_sc
lstm_preds = scaler_lstm.inverse_transform(inv_l[:, w_idx])

n_lstm = min(len(lstm_preds), len(wheat_test))
lstm_metrics = calc_metrics(wheat_test.values[:n_lstm],
                          lstm_preds[:n_lstm])
print(f" LSTM: MAE={lstm_metrics['MAE']} ")

```

```

f"RMSE={lstm_metrics['RMSE']} "
f"MAPE={lstm_metrics['MAPE']}%")

forecast_plot(wheat_train, wheat_test, lstm_preds,
              'LSTM', C['lstm'], 'fig3_17_lstm_forecast.png')
np.save('lstm_preds.npy', lstm_preds)

# Аналіз lookback window
print("\n Аналіз чутливості до lookback ...")
lb_sens = []
for lb in [20, 40, 60, 80, 120]:
    try:
        X_t, y_t = make_windows_3d(tr_sc_l, lb, w_idx)
        X_e, _ = make_windows_3d(
            np.vstack([tr_sc_l[-lb:], te_sc_l]), lb, w_idx)
        m_lb = Sequential([
            LSTM(64, return_sequences=True,
                input_shape=(lb, len(ML_COLS))),
            Dropout(0.2),
            LSTM(32, return_sequences=False),
            Dropout(0.1),
            Dense(1)
        ])
        m_lb.compile(optimizer=Adam(0.001), loss='mse')
        m_lb.fit(X_t, y_t, epochs=40, batch_size=32,
            validation_split=0.15,
            callbacks=[EarlyStopping(patience=8,
                restore_best_weights=True,
                verbose=0)],
            verbose=0, shuffle=False)
        p_sc = m_lb.predict(X_e, verbose=0).flatten()
        inv_ = np.zeros((len(p_sc), len(ML_COLS)))
        inv[:, w_idx] = p_sc
        p_lv = scaler_lstm.inverse_transform(inv_[:, w_idx])
        n_ = min(len(p_lv), len(wheat_test))
        m_ = calc_metrics(wheat_test.values[:n_], p_lv[:n_])
        lb_sens.append({'Lookback': lb, **m_})
        print(f" LB={lb}: RMSE={m_['RMSE']}")
    except Exception as e:
        print(f" LB={lb}: {e}")

if lb_sens:
    lb_df = pd.DataFrame(lb_sens)
    fig, axes = plt.subplots(1, 3, figsize=(14, 4))
    for ax, metric in zip(axes, ['MAE', 'RMSE', 'MAPE']):
        ax.plot(lb_df['Lookback'], lb_df[metric],
            marker='o', color=C['lstm'], lw=2,
            markersize=7, markerfacecolor='white',
            markeredgewidth=2)
        ax.axvline(LOOKBACK, color='red', lw=1.2, ls='--',
            label=f'Обраний={LOOKBACK}')
        ax.set_xlabel('Lookback window')
        ax.set_ylabel(metric)
        ax.set_title(f'{metric} vs Lookback', fontweight='bold')
        ax.legend(fontsize=8)
    fig.suptitle('Рисунок 3.7: Чутливість LSTM до lookback window',
        fontsize=12, fontweight='bold')
    plt.tight_layout()
    plt.savefig('fig3_18_lstm_lookback.png', bbox_inches='tight')
    plt.show()
    print(" Збережено: fig3_18_lstm_lookback.png")

```

```

#


---


# 3.3 — КОМПАРАТИВНИЙ АНАЛІЗ
#


---



print("\n" + "=" * 65)
print("3.3 КОМПАРАТИВНИЙ АНАЛІЗ")
print("=" * 65)

preds_all = {
    'ARIMA': np.array(arima_preds),
    'VAR': np.array(var_preds_level),
    'MLP': mlp_preds,
}
if TF_OK:
    preds_all['LSTM'] = lstm_preds

metrics_rows = []
for mname, preds in preds_all.items():
    n = min(len(preds), len(wheat_test))
    m = calc_metrics(wheat_test.values[:n], preds[:n])
    m['Модель'] = mname
    metrics_rows.append(m)

metrics_df = pd.DataFrame(metrics_rows).set_index('Модель')
print("\nЗведена таблиця метрик:")
print(metrics_df.to_string())

best_m = metrics_df['RMSE'].idxmin()
print(f"\n Найкраща модель за RMSE: {best_m}")

# — Таблиця метрик


---


table_fig(metrics_df.round(4),
           'Таблиця 3.9: Порівняльна точність моделей (2023–2026)',
           'fig3_19_metrics_table.png')

# — Bar chart метрик


---


model_colors = [C.get(m.lower(), '#95a5a6')
                 for m in metrics_df.index]
fig, axes = plt.subplots(1, 3, figsize=(14, 5))
for ax, metric in zip(axes, ['MAE', 'RMSE', 'MAPE']):
    bars = ax.bar(metrics_df.index, metrics_df[metric],
                  color=model_colors, alpha=0.85,
                  edgecolor='white', lw=1.5)
    for bar in bars:
        h = bar.get_height()
        if not np.isnan(h):
            ax.text(bar.get_x() + bar.get_width()/2., h*1.01,
                    f'{h:.2f}', ha='center', va='bottom',
                    fontsize=9, fontweight='bold')
    # Золота рамка для кращої моделі
    best_idx = list(metrics_df.index).index(best_m)
    if best_idx < len(bars):
        bars[best_idx].set_edgecolor('gold')
        bars[best_idx].set_linewidth(3)
    ax.set_title(metric, fontweight='bold')
    ax.set_ylabel(metric)
fig.suptitle('Рисунок 3.8: Метрики точності моделей ')

```

```

        '(золота рамка = найкраща)',
        fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig3_20_metrics_bar.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig3_20_metrics_bar.png")

# — Всі прогнози на одному графіку —————
cfg = {
    'ARIMA': (C['arima'], '--', 1.6),
    'VAR': (C['var'], '-.', 1.6),
    'MLP': (C['mlp'], ':', 2.0),
    'LSTM': (C['lstm'], '--', 1.8),
}
fig, ax = plt.subplots(figsize=(16, 6))
ax.plot(wheat_test.index, wheat_test.values,
        color=C['actual'], lw=2.2, label='Фактична ціна', zorder=5)
for mname, preds in preds_all.items():
    color, ls, lw = cfg.get(mname, ('#999', '--', 1.5))
    n = min(len(preds), len(wheat_test))
    rmse = metrics_df.loc[mname, 'RMSE'] if mname in metrics_df.index else np.nan
    ax.plot(wheat_test.index[:n], preds[:n],
            color=color, lw=lw, ls=ls, alpha=0.9,
            label=f'{mname} (RMSE={rmse:.1f})')
ax.set_ylabel('Ціна пшениці (центи/бушель)')
ax.set_xlabel('Дата')
ax.set_title('Рисунок 3.9: Порівняння прогнозів всіх моделей (2023–2026)',
            fontweight='bold', fontsize=12)
ax.legend(loc='upper left', fontsize=9, framealpha=0.9)
ax.xaxis.set_major_formatter(mdates.DateFormatter('%m/%Y'))
plt.tight_layout()
plt.savefig('fig3_21_all_forecasts.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig3_21_all_forecasts.png")

# — Scatter факт vs прогноз —————
n_plots = len(preds_all)
fig, axes = plt.subplots(1, n_plots, figsize=(5*n_plots, 5))
if n_plots == 1:
    axes = [axes]
for ax, (mname, preds) in zip(axes, preds_all.items()):
    color, _ = cfg.get(mname, ('#999', '--', 1.5))
    n = min(len(preds), len(wheat_test))
    act = wheat_test.values[:n]
    pred = np.array(preds[:n])
    mask = ~np.isnan(pred)
    ax.scatter(act[mask], pred[mask],
               alpha=0.2, s=5, color=color)
    mn, mx = act[mask].min(), act[mask].max()
    ax.plot([mn, mx], [mn, mx], 'k--', lw=1.5, label='Ідеал')
    r2 = np.corrcoef(act[mask], pred[mask])[0,1]**2
    ax.set_xlabel('Фактична ціна')
    ax.set_ylabel('Прогноз')
    ax.set_title(f'{mname} R²={r2:.4f}', fontweight='bold')
    ax.legend(fontsize=8)
fig.suptitle('Рисунок 3.10: Діаграми Факт vs Прогноз',
            fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig3_22_scatter.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig3_22_scatter.png")

```

— Розподіл залишків

```
fig, axes = plt.subplots(2, n_plots, figsize=(5*n_plots, 8))
if n_plots == 1:
    axes = axes.reshape(2,1)
for col, (mname, preds) in enumerate(preds_all.items()):
    color, _ = cfg.get(mname, ('#999', '--', 1.5))
    n = min(len(preds), len(wheat_test))
    resid = wheat_test.values[:n] - np.array(preds[:n])
    resid = resid[~np.isnan(resid)]

    axes[0, col].hist(resid, bins=40, color=color,
                      alpha=0.55, density=True)
    pd.Series(resid).plot.kde(ax=axes[0, col],
                             color='black', lw=1.5)
    axes[0, col].axvline(0, color='red', ls='--', lw=1)
    axes[0, col].set_title(f'{mname}: розподіл залишків',
                           fontweight='bold')

    stats.probplot(resid, dist='norm', plot=axes[1, col])
    axes[1, col].get_lines()[0].set(markerfacecolor=color,
                                     markersize=2, alpha=0.4)
    axes[1, col].set_title(f'{mname}: QQ-Plot', fontweight='bold')

fig.suptitle('Рисунок 3.11: Аналіз залишків моделей',
             fontsize=12, fontweight='bold')
plt.tight_layout()
plt.savefig('fig3_23_residuals.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig3_23_residuals.png")
```

— Тест Диболда–Маріано

```
print("\n Тест Диболда–Маріано ...")
model_list = list(preds_all.keys())
dm_rows = []
for i, m1 in enumerate(model_list):
    for m2 in model_list[i+1:]:
        n = min(len(preds_all[m1]), len(preds_all[m2]),
                len(wheat_test))
        e1 = wheat_test.values[:n] - np.array(preds_all[m1][:n])
        e2 = wheat_test.values[:n] - np.array(preds_all[m2][:n])
        dm_stat, p_val = diebold_mariano(e1, e2)
        r1 = metrics_df.loc[m1, 'RMSE'] if m1 in metrics_df.index else np.nan
        r2 = metrics_df.loc[m2, 'RMSE'] if m2 in metrics_df.index else np.nan
        dm_rows.append({
            'Папа': f'{m1} vs {m2}',
            'DM-статистика': dm_stat,
            'p-value': p_val,
            f'RMSE {m1}': r1,
            f'RMSE {m2}': r2,
            'Краща': m1 if (not np.isnan(r1) and
                           not np.isnan(r2) and r1 < r2) else m2,
            'Значуща?': 'Так ✓' if (p_val and p_val < 0.05)
                           else 'Hi X'
        })

if dm_rows:
    dm_df = pd.DataFrame(dm_rows).set_index('Папа')
    print(dm_df.to_string())
    table_fig(dm_df,
```

'Таблиця 3.10: Тест Диболда–Маріано',
'fig3_24_dm_test.png')

— MAPE по роках

```
fig, ax = plt.subplots(figsize=(12, 5))
years = sorted(wheat_test.index.year.unique())
for mname, preds in preds_all.items():
    color, ls, lw = cfg.get(mname, ('#999', '--', 1.5))
    y_mape = []
    preds_arr = np.array(preds)
    for yr in years:
        mask_yr = wheat_test.index.year == yr
        idx_yr = np.where(mask_yr)[0]
        valid_idx = idx_yr[idx_yr < len(preds_arr)]
        if len(valid_idx) == 0:
            y_mape.append(np.nan); continue
        act_yr = wheat_test.values[valid_idx]
        pred_yr = preds_arr[valid_idx]
        valid = ~np.isnan(pred_yr) & (act_yr != 0)
        if valid.sum() == 0:
            y_mape.append(np.nan); continue
        y_mape.append(np.mean(
            np.abs((act_yr[valid]-pred_yr[valid])/act_yr[valid]))*100)
    ax.plot(years[:len(y_mape)], y_mape,
            marker='o', color=color, lw=lw, ls=ls,
            label=mname, markersize=7)
ax.set_xlabel('Рік')
ax.set_ylabel('MAPE (%)')
ax.set_title('Рисунок 3.12: MAPE по роках тестової вибірки',
            fontweight='bold')
ax.legend()
plt.tight_layout()
plt.savefig('fig3_25_mape_by_year.png', bbox_inches='tight')
plt.show()
print(" Збережено: fig3_25_mape_by_year.png")
```

#

ПІДСУМОК

#

```
print("\n" + "=" * 65)
print("ПІДСУМОК РОЗДІЛУ 3")
print("=" * 65)
print(f"\n Метрики на тестовій вибірці (2023–2026):")
print(metrics_df.to_string())
print(f"\n Найкраща модель (RMSE): {best_m}")

if dm_rows:
    sig = sum(1 for r in dm_rows if r.get("Значуща?") == 'Так ✓')
    print(f" Статистично значущих пар (тест ДМ): {sig}/{len(dm_rows)}")
]
for f in files:
    print(f" {f}")
print(f"\n Всього: {len(files)} файлів")
print("=" * 65)
print("РОЗДІЛ 3 ЗАВЕРШЕНО")
print("=" * 65)
```

Метадані

ДОКУМЕНТ

Заголовок

Турченко_Галіцина_дист_плагіат

Автор

Турченко

Науковий керівник / Експерт

Галіцина

ІД документа

334088755

ОРГАНІЗАЦІЯ

Назва організації

Kyiv National Economic University named after Vadym Hetman
KNEU

підрозділ

кафедра штучного інтелекту, моделювання та статистики

ЗВІТ

Дата звіту

5/29/2026

Дата редагування

Обсяг знайдених подібностей

Коефіцієнт подібності визначає, який відсоток тексту по відношенню до загального обсягу тексту було знайдено в різних джерелах. Зверніть увагу, що високі значення коефіцієнта не автоматично означають плагіат. Звіт має аналізувати компетентна / уповноважена особа.



25
Довжина фрази для коефіцієнта подібності 2



11488
Кількість слів



82671
Кількість символів