

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КРИВОРІЗЬКИЙ ДЕРЖАВНИЙ ПЕДАГОГІЧНИЙ УНІВЕРСИТЕТ
Фізико-математичний факультет
Кафедра інформатики та прикладної математики

«Допущено до захисту»

В.о. завідувача кафедри

_____ Семеріков С.О.

«_____» _____ 2022 р.

Реєстраційний № _____

«_____» _____ 2022 р.

**МУЛЬТИВАРІАТИВНЕ АВТОРЕГРЕСІЙНЕ
МОДЕЛЮВАННЯ ФІНАНСОВИХ АКТИВІВ**

Кваліфікаційна робота студента
групи Ім-17

ступінь вищої освіти «магістр»

спеціальності 014 Середня освіта (Інформатика)

Потапенка Артура Олеговича

Керівник: проф., д. ф.-м. н.

Соловйов Володимир Миколайович

Оцінка:

Національна шкала _____

Шкала ECTS ___ Кількість балів ____

Голова ЕК _____

Члени ЕК _____

ЗАПЕВНЕННЯ

Я, Потапенко Артур Олегович, розумію і підтримую політику Криворізького державного педагогічного університету з академічної доброчесності. Запевняю, що ця кваліфікаційна робота виконана самостійно, не містить академічного плагіату, фабрикації, фальсифікації. Я не надавав і не одержував недозволену допомогу під час підготовки цієї роботи. Використання ідей, результатів і текстів інших авторів мають покликання на відповідне джерело.

Із чинним Положенням про запобігання та виявлення академічного плагіату в роботах здобувачів вищої освіти Криворізького державного педагогічного університету ознайомлений. Чітко усвідомлюю, що в разі виявлення у кваліфікаційній роботі порушення академічної доброчесності робота не допускається до захисту або оцінюється незадовільно.

ЗМІСТ

ВСТУП	4
РОЗДІЛ 1. ТЕОРЕТИЧНІ АСПЕКТИ МОДЕЛЮВАННЯ ФІНАНСОВИХ АКТИВІВ	
1.1. Визначення понять, досліджуваних у магістерській роботі.....	6
1.2. Аналіз дослідження проблеми мультиваріативного авторегресійного моделювання фінансових активів.....	9
Висновки до розділу 1	12
РОЗДІЛ 2. МЕТОДОЛОГІЇ МОДЕЛЮВАННЯ ФІНАНСОВИХ АКТИВІВ	
2.1. Визначення і компоненти часового ряду.....	14
2.2. Використання методології ARIMA (Бокса-Дженкінса) для моделювання фінансових активів.....	22
2.3. Метод експоненціального згладжування.....	26
2.4. Векторні моделі авторегресії Кевіна Коце.....	32
Висновки до розділу 2	36
РОЗДІЛ 3. ПРАКТИЧНЕ МОДЕЛЮВАННЯ ФІНАНСОВИХ АКТИВІВ ІЗ ЗАСТОСУВАННЯМ VAR МОДЕЛІ	
3.1. Дослідження даних та тренування VAR моделі.....	38
3.2. Побудова прогнозів поза вибірки навчання та тестування VAR моделі.....	52
Висновки до розділу 3	56
ВИСНОВКИ	57
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	60

ВСТУП

Методи та моделі соціально-економічного прогнозування засновані на теорії прогнозування або прогностиці, яка сформувалась порівняно недавно, хоча питаннями прогнозування людство цікавилось протягом усієї історії розвитку. Необхідність вивчати явище майбутнього є постійним прагненням людства.

Основними поняттями прогностики є поняття прогнозу та прогнозування. Під прогнозом розуміють науково-обґрунтоване судження про можливий стан об'єкта в майбутньому або альтернативні шляхи та терміни досягнення цих станів.

Одним із найбільш важливих напрямів суспільного розвитку є соціально-економічне прогнозування. Останнє розглядається як процес розробки прогнозів, заснований на наукових методах пізнання соціально-економічних явищ і використання сукупності методів і засобів соціально-економічної прогностики.

Етап прогнозування – частина процесу розробки прогнозів, що характеризується завданнями, методами та результатами. Розподіл на етапи пов'язаний зі специфікою побудови систематизованого опису об'єкта прогнозування, збору даних, з побудовою моделі, верифікацією прогнозу.

Модель прогнозування – модель об'єкта прогнозування, дослідження якої дозволяє отримати інформацію про можливі стани об'єкта прогнозування в майбутньому або шляхи та терміни їх здійснення. Особливо актуальним є моделювання фінансових активів, що дозволяє передбачити кризові явища на фінансовому ринку.

Верифікація прогнозу фінансових активів, тобто оцінювання їх вірогідності, точності й обґрунтованості досягається як правило шляхом використання авторегресійних моделей. Серед авторегресійних моделей виділяють VAR-моделі (модель Кевіна Коце), модель ковзного середнього ARIMA (модель Бокса-Дженкінса).

Проблематику мультиваріативного авторегресійного моделювання досліджено у працях відомих українських фахівців: В. Алімпієва [1], В. Маслія [14], В. Соловійова [5]. Також цю проблему було висвітлено в наукових доробках зарубіжних дослідників: Д. Бокса, Г. Дженкінса, К. Коце.

Виходячи з актуальності даної проблематики ми обрали темою магістерської роботи **«Мультиваріативне авторегресійне моделювання фінансових активів»**.

Метою дослідження є теоретичне вивчення особливостей практичного використання авторегресійних моделей для прогнозування фінансових активів.

Мета дослідження визначила постановку таких його **завдань**:

1. З'ясувати сутність та завдання авторегресійного моделювання.
2. Проаналізувати методологію моделювання фінансових активів.
3. Дослідити ефективності використання методології ARIMA (Бокса-Дженкінса) і векторної моделі авторегресії Кевіна Коце.

Об'єкт дослідження – методи авторегресійного моделювання.

Предмет дослідження: – використання методології ARIMA (Бокса-Дженкінса) і векторної моделі авторегресії Кевіна Коце та оцінка їх ефективності.

Методи дослідження:

- загальнонаукові (аналіз, синтез, систематизація, узагальнення, прогнозування);
- математичні (метод математичного аналізу, моделювання).

Практична значущість дослідження полягає в можливості практичного використання його результатів для написання курсових, дипломних і магістерських робіт.

Структура магістерської роботи. Магістерська робота складається зі вступу, трьох розділів, висновків до кожного розділу, загальних висновків, списку використаних джерел. Загальний обсяг роботи складає 64 сторінки. Список використаних джерел містить 44 позиції.

РОЗДІЛ 1

ТЕОРЕТИЧНІ АСПЕКТИ МОДЕЛЮВАННЯ ФІНАНСОВИХ АКТИВІВ

1.1. Визначення понять, досліджуваних у магістерській роботі

Модель – це схема, що імітує будову й дію будь-якого процесу. Слово «модель» з латинської означає «міра», зразок, примірник чого-небудь, схема для пояснення якогось явища або процесу. Відображаючи чи відтворюючи об'єкт дослідження модель здатна заміщати його таким чином, що її вивчення дає нам нову інформацію про цей об'єкт [14].

У науковій літературі знаходимо висвітлення методу моделювання як методу наукового пізнання, що має за основну функцію відтворення властивостей і відносин предметів і процесів у ході здійснення дослідження. Оригінальні предмети й процеси заміщаються моделлю, причому обираються ті відносини й взаємозв'язки об'єкту, які виступають безпосередньо предметом дослідження. У вузькому розумінні моделювання – це спосіб пізнання, за умов якого одна система (об'єкт дослідження) відтворюється в іншій (моделі).

Система є множиною елементів, які перебувають у відношеннях і зв'язках один з одним й утворюють певну цілісність, єдність. Система в теорії управління визначається як складне утворення, що складається з підсистем, функціонування яких підпорядковано єдиній меті [14].

VAR моделі – це мультіваріативні динамічні економетричні моделі, що використовуються для відображення лінійних взаємозв'язків між декількома часовими рядами. Популярність VAR моделей в економіці зросла після того, як її було запропоновано як альтернативну систему симультивних рівнянь, що часто включали екзогенні змінні. У векторній авторегресійній моделі всі змінні трактуються ендогенними, що дозволяє відобразити динамічні взаємозв'язки між досліджуваними економічними показниками. Такі моделі об'єднують прості авторегресійні рівняння в одну ендогенну систему [25].

VAR моделі найчастіше застосовуються для прогнозування, однак популярність даної методології сприяє розвитку іншого інструментарію, що дозволяє оцінити структурні зв'язки між ендогенними компонентами.

Сучасні методи прогнозування часових рядів базуються, переважно, на принципі історичного обумовлення майбутнього. При цьому постійного удосконалення набувають як методи тлумачення інформації, що представляє минулі події, так і способи їх екстраполяції на майбутнє. Широкого розповсюдження набуло параметричне моделювання, класичним прикладом якого є регресійний аналіз. Найбільш поширеними методами оцінки параметрів моделі залишаються метод найменших квадратів Гауса та метод максимальної правдоподібності. Підходи до специфікації параметричних моделей умовно можна поділити на структурні моделі, які базуються на системі рівнянь та обмежень на параметри, і спеціальні моделі «ad hoc», які не мають теоретичного обґрунтування [24].

Так, після масового використання структурного підходу у макроекономічному моделюванні в 1950-х роках, у 1960-х рр. науковці почали активно досліджувати сезонні та циклічні характеристики часових рядів. У 1970 р. Д. Бокс та Г. Дженкінс систематизували результати цих робіт та запропонували комплексний підхід до моделювання і прогнозування часових рядів на підставі авторегресійних моделей та моделей ковзного середнього (ARMA / ARIMA) [13].

Дана методика набула популярності, оскільки була простішою за складні структурні моделі та дозволяла отримувати не менш якісні прогнози. Її програмна реалізація представлена в таких програмних економетричних пакетах, як EViews, Stata, Statistica, SPSS. В рамках структурного підходу, починаючи з 1980-х років, успішно себе зарекомендували динамічні лінійні та нелінійні Баєсовські моделі із застосуванням фільтра Кальмана. Фільтр Кальмана використовується також і в моделях ARIMA для усунення проблеми недостовірного оцінювання параметрів методом максимальної правдоподібності та визначення неспостережуваних компонент ряду [25].

Мультиваріативне авторегресійне моделювання часто використовується для фінансових активів.

Фінансові активи – частина активів, яка являє собою фінансові ресурси: грошові кошти, цінні папери, боргові зобов'язання та право вимоги боргу. Фінансові активи включають в себе готівку каси, депозити в банках, вклади, чеки, страхові поліси, вклади в цінні папери, зобов'язання інших підприємств і організацій з виплати коштів за відвантажену продукцію (комерційний кредит), портфельні інвестиції в акції інших підприємств, пакети акцій інших підприємств, які надають право контролю, паї [26].

Відповідно до ст.4 П(С)БО 13 «Фінансові інструменти»:

Фінансовий актив – це:

1. Грошові кошти та їх еквіваленти.
2. Контракт, що надає право отримати грошові кошти або інший фінансовий актив від іншого підприємства.
3. Контракт, що надає право обмінятися фінансовими інструментами з іншим підприємством на потенційно вигідних умовах.
4. Інструмент власного капіталу іншого підприємства [28].

Фінансовий актив, призначений для перепродажу – фінансовий актив, придбаний з метою подальшого продажу для отримання прибутку від короткотермінових змін його ціни та/або винагороди посередника [27].

Фінансовий інструмент – контракт, який одночасно приводить до виникнення (збільшення) фінансового активу в одного підприємства і фінансового зобов'язання або інструмента власного капіталу в іншого [28].

Інструмент власного капіталу – це контракт, який підвищує право на частину в активах підприємства, що залишається після вирахування сум за всіма його зобов'язаннями.

Фінансовий актив – це будь-який актив, що є:

1. Грошовими коштами;
2. Інструментом власного капіталу іншого суб'єкта господарювання;
3. Контрактним правом:
 - отримувати грошові кошти або інший фінансовий актив від іншого суб'єкта господарювання;

- обмінювати фінансові інструменти з іншим суб'єктом господарювання за умов, які є потенційно сприятливими;
4. контрактом, розрахунки за яким здійснюватимуться або можуть здійснюватися власними інструментами капіталу суб'єкта господарювання та який є:
- непохідним інструментом, за який суб'єкт господарювання зобов'язаний отримати змінну кількість власних інструментів капіталу,
 - похідним інструментом, розрахунки за яким здійснюватимуться або можуть здійснюватися іншим чином, ніж обміном фіксованої суми грошових коштів або іншого фінансового активу на фіксовану кількість власних інструментів капіталу [18].

З цією метою до інструментів власного капіталу суб'єкта господарювання не включають: фінансові інструменти з правом дострокового погашення, класифіковані як інструменти власного капіталу; інструменти, які створюють для суб'єкта господарювання зобов'язання надати іншій стороні пропорційну частку чистих активів суб'єкта господарювання тільки при ліквідації і класифіковані як інструменти власного капіталу або інструменти, що є контрактами на майбутнє отримання або надання інструментів власного капіталу суб'єкта господарювання [8].

1.2. Аналіз дослідження проблеми мультиваріативного авторегресійного моделювання фінансових активів

На сьогодні, за оцінками фахівців, налічується більше ста п'ятдесяти різноманітних методів прогнозування, але на практиці в якості основних використовується близько двадцяти. Під методами прогнозування розуміється сукупність прийомів і способів мислення, які дозволяють на основі ретроспективних даних, екзогенних (зовнішніх) й ендогенних (внутрішніх) зв'язків об'єкта прогнозування та їх змін вивести судження певної вірогідності відносно його майбутнього розвитку [4].

До найважливіших класифікаційних ознак методів прогнозування належать: ступінь формалізації; загальні принципи дії; спосіб отримання й обробки інформації; напрям і призначення прогнозування; процедура отримання параметрів прогнозування та інші [3].

Вибір методу та моделі реалізації визначається наявністю ретроспективної інформації, терміном прогнозування, цілями та вимогами до прогнозування.

За ступенем формалізації методи соціально-економічного прогнозування можна розділити на інтуїтивні, формалізовані та комбіновані. У виборі методів прогнозування важливим показником є глибина попередження прогнозу. При цьому важливо не тільки знати дану величину, але й співвіднести її до тривалості циклу функціонування об'єкта прогнозування.

Формалізовані методи прогнозування використовуються, коли величина глибини попередження укладається в рамки циклу функціонування (еволюційного циклу). У разі виникнення в рамках прогнозного періоду «стрибка», у розвитку об'єкта необхідно використовувати інтуїтивні методи як для визначення «чинності» стрибка так і для оцінки часу його здійснення. У проміжках між стрибками можуть бути використані формалізовані методи [4].

Розрізняють індивідуальні та колективні експертні оцінки. Індивідуальні експертні оцінки включають:

1. метод «інтерв'ю», при якому здійснюється безпосередній контакт експерта з фахівцем за схемою «питання – відповідь»;
2. аналітичний метод, при якому здійснюється логічний аналіз будь-якої прогнозованої ситуації;
3. метод написання сценарію, заснований на визначенні логіки процесу або явища в часі за різних умов [4].

До складу колективних експертних оцінок входять:

1. метод «комісій»;
2. метод Дельфі;
3. метод «колективної генерації ідей» (мозкова атака);

4. матричний метод (метод сценаріїв).

Застосування цих методів, засноване на колективному мисленні, дозволяє знизити ступінь суб'єктивізму думок експертів і, як наслідок, підвищити точність результатів.

У групі формалізованих методів виділяють дві підгрупи: екстраполяції та моделювання. До першої підгрупи належать методи найменших квадратів, експонентного згладжування та інші. До другої підгрупи належать структурне, мережне, матричне та імітаційне моделювання [11].

Значне місце серед методів соціально-економічного прогнозування займають так звані комбіновані методи. До них належать методи зі змішаною інформаційною основою, в яких як первинна інформація використовується фактографічна й експертна інформація. Наприклад, при проведенні експертного опитування може бути використана фактографічна інформація й, навпаки, при екстраполяції тенденції поряд з фактичними даними – експертні оцінки.

Комбіновані методи використовують під час побудови досить складних соціально-економічних прогнозів, де дуже важлива обробка якісної та кількісної інформації.

Моделі авторегресії широко застосовуються для опису стаціонарних процесів в економічних дослідженнях. Побудова авторегресійних моделей в економіці заснована на такому важливому властивості рядів економічних явищ і процесів, як взаємозалежність рівнів одного і того ж ряду один від одного. Умова нормальності розподілу ряду при побудові їх економетричних моделей не є обов'язковою [5].

Застосування авторегресійних моделей широко поширене для опису стаціонарних випадкових процесів. Визначною характеристикою стаціонарних часових рядів є те, що їх розвиток відбувається в незмінних стабільних умовах без вираженої тенденції, тому ймовірнісні властивості рядів залишаються незмінними від часу. Функції розподілу стаціонарних динамічних рядів також зберігають сталість при зсуві часу.

Висновки до розділу 1

У науковій літературі термін «модель» трактується як схема, що імітує будову й дію будь-якого процесу. Відображаючи чи відтворюючи об'єкт дослідження модель здатна заміщати його таким чином, що її вивчення дає нам нову інформацію про цей об'єкт. Метод моделювання – це метод наукового пізнання, що має за основну функцію відтворення властивостей і відносин предметів і процесів у ході здійснення дослідження.

VAR моделі – це мультиваріативні динамічні економетричні моделі, що використовуються для відображення лінійних взаємозв'язків між декількома часовими рядами. У векторній авторегресійній моделі всі змінні трактуються ендогенними, що дозволяє відобразити динамічні взаємозв'язки між досліджуваними економічними показниками.

У 1960-х рр. науковці почали активно досліджувати сезонні та циклічні характеристики часових рядів. Д. Бокс та Г. Дженкінс систематизували результати цих робіт та запропонували комплексний підхід до моделювання і прогнозування часових рядів на підставі авторегресійних моделей та моделей ковзного середнього (ARMA / ARIMA).

Мультиваріативне авторегресійне моделювання часто використовується для фінансових активів. Фінансові активи – частина активів, яка являє собою фінансові ресурси: грошові кошти, цінні папери, боргові зобов'язання та право вимоги боргу. Під фінансовими активами розуміють: грошові кошти та їх еквіваленти; контракт, що надає право отримати грошові кошти або інший фінансовий актив від іншого підприємства; контракт, що надає право обмінятися фінансовими інструментами з іншим підприємством на потенційно вигідних умовах; інструмент власного капіталу.

За ступенем формалізації методи соціально-економічного прогнозування можна розділити на інтуїтивні, формалізовані та комбіновані. У виборі методів прогнозування важливим показником є глибина попередження прогнозу.

Моделі авторегресії широко застосовуються для опису стаціонарних процесів в економічних дослідженнях. Побудова авторегресійних моделей в

економіці заснована на такому важливому властивості рядів економічних явищ і процесів, як взаємозалежність рівнів одного і того ж ряду один від одного.

Застосування авторегресійних моделей широко поширене для опису стаціонарних випадкових процесів. Визначною характеристикою стаціонарних часових рядів є те, що їх розвиток відбувається в незмінних стабільних умовах без вираженої тенденції, тому ймовірнісні властивості рядів залишаються незмінними від часу.

РОЗДІЛ 2

МЕТОДОЛОГІЇ МОДЕЛЮВАННЯ ФІНАНСОВИХ АКТИВІВ

2.1. Визначення і компоненти часового ряду

Часовим рядом називають послідовність значень статистичного показника (ознаки), впорядковану у хронологічному порядку. Застосовують також терміни «ряд динаміки» та «динамічний ряд» [3].

В англomовній літературі використовують термін «time series». Окремі спостереження часового ряду називають його рівнями, або елементами. Кожний рівень ряду відповідає певному моменту часу. Рівні ряду можуть набувати як детермінованих, так і випадкових значень. Порядок розташування рівнів є істотною характеристикою ряду і не може змінюватися довільно. Іноді кожному моменту часу приводять у відповідність декілька значень різних показників досліджуваного об'єкта. Тоді отримують багатовимірний часовий ряд [19].

Основними завданнями дослідження часових рядів є:

- відокремлення та опис основних характерних особливостей ряду; підбір статистичної моделі, що найкращим у певному розумінні способом відображає ряд;
- прогнозування майбутніх значень показників, що утворюють ряд, за попередніми спостереженнями;
- підготовка рекомендацій з управління процесом, що породжує досліджуваний часовий ряд [13].

Аналіз часових рядів, як правило, передбачає проведення таких основних етапів:

- графічне подання й попередній аналіз поведінки часового ряду;
- відокремлення і видалення закономірних складових ряду (тренду, сезонних та циклічних компонентів);
- відокремлення та видалення низько та високочастотних складових (фільтрація);

- дослідження випадкової складової часового ряду, що залишилася після видалення вищезазначених компонентів;
- побудова і перевірка адекватності моделі випадкової складової;
- побудова загальної моделі досліджуваного ряду;
- дослідження отриманої моделі і прогнозування майбутньої поведінки об'єкта, що вивчається;
- вивчення взаємодії між різними часовими рядами, що характеризують певну систему або процес [37].

При дослідженні часових рядів застосовують основні методи:

1. Кореляційний аналіз, який дає змогу виявляти істотні періодичні залежності та їх лаги (затримки) всередині певного процесу (автокореляція) або між декількома процесами (крос-кореляція).
2. Спектральний аналіз, що застосовують для визначення періодичних та квазіперіодичних компонент часового ряду.
3. Методи згладжування та фільтрації, призначені для перетворення часових рядів з метою видалення з них високочастотних та сезонних коливань.
4. Методи авторегресії та ковзних середніх, які використовують для опису і прогнозування процесів, що здійснюють випадкові коливання навколо певного середнього значення.
5. Методи прогнозування, що дають можливість на основі обраної моделі часового ряду оцінювати його найбільш імовірні значення в майбутньому [9].

Залежно від характеру часового параметра розрізняють моментні та інтервальні часові ряди. У моментних рядах рівні характеризують значення показника станом на певний момент часу. Їм відповідають змінні типу запасу (stock variables). Зокрема, моментними є ряди цін на різноманітні товари станом на певні моменти часу, курсів валют, кількості населення, маси об'єкта, електричного напруження у мережі, температури хворого [21].

В інтервальних рядах рівні характеризують значення показника за певні періоди (інтервали) часу. Їм відповідають змінні типу потоку (flow variables).

Прикладами є ряди динаміки виробництва продукції (річного, квартального, місячного, добового), міграції населення, прибутку підприємства, валового національного доходу, розсіюваної потужності, витрат реагентів у хімічному реакторі [9].

Часовий ряд, сформований із перших різниць рівнів моментного ряду, є інтервальним. Часовий ряд, рівні якого отримані підсумовуванням наростаючим підсумком всіх рівнів інтервального ряду, починаючи з певного визначеного рівня, є моментним. Співвідношення між моментними та інтервальними часовими рядами аналогічно співвідношенню між функціями та їх похідними.

Розрізняють повні й неповні часові ряди. У неповних часових рядів відсутні рівні, що відповідають певним моментам часу.

Рівні рядів динаміки можуть бути абсолютними, відносними або середніми значеннями певних показників. Якщо вони не є величинами, вимірюваними безпосередньо, а похідними від них – середніми, відносними й т.д., то відповідні ряди називають похідними. Зокрема, похідними є ряди середньодобового або середньомісячного виробництва певної продукції [2].

Особливістю інтервальних рядів є можливість підсумовування їх рівнів. Результатом цього є так звані нагромаджені підсумки, які є значеннями певного показника за відповідний період часу. Наприклад, підсумовуючи добові обсяги виробництва продукції протягом місяця, отримують загальний обсяг її виробництва за відповідний місяць.

Суми рівнів моментного ряду не мають змісту. Зокрема, якщо ці рівні є значеннями середньодобової температури, то підсумовуючи їх протягом певного періоду, не можна отримати величини, яка б мала фізичний зміст. Для цього потрібно поділити результат на кількість днів у досліджуваному періоді. Тоді отримаємо величину, що є середнім значенням температури за відповідний період часу. Різниця рівнів моментного ряду є зміною показника за відповідний період [9].

Часовим рядам притаманні такі властивості:

- рівні часового ряду не є статистично незалежними;
- члени часового ряду не є однаково розподіленими [1].

Успішність аналізу часового ряду істотно залежить від правильності вибору інтервалу між його сусідніми рівнями. Зручніше вибирати рівновіддалені рівні ряду, але це не завжди є можливим. Зокрема, офіційні курси валют і фондові індекси не визначаються у вихідні та святкові дні, тому рівні відповідних рядів не є рівновіддаленими. Вибір надмірно великого інтервалу може призвести до втрати інформації про суттєві особливості ряду. Крім того, кількість членів ряду при цьому може виявитися недостатньою для застосування деяких методів аналізу. З іншого боку, надмірно малі інтервали збільшують обсяг розрахунків і можуть затінювати особливості динаміки показника випадковими флуктуаціями [37].

Важливою умовою аналізу є порівнянність рівнів ряду. Їх непорівнянність може бути зумовлена змінами законодавства та нормативної бази, методик розрахунку показників, меж адміністративно-територіальних одиниць і держав, технологій виробництва, термінологій, природними, екологічними та техногенними катастрофами й іншими факторами. Інформація про ряд має бути достатньо повною. Зокрема, для рядів, що містять циклічну складову, залежно від застосовуваних методів математичної обробки, потрібна інформація щодо проміжку, який перевищує 3–6 повних циклів. При побудові регресійних моделей необхідно мати ряди, довжина яких у кілька разів перевищує кількість параметрів, що визначаються [19].

Рівні часових рядів можуть включати аномальні значення. Причинами їх появи найчастіше є помилки під час збору та обробки інформації. Але викиди також можуть бути зумовлені реальними стрибками в динаміці досліджуваних показників.

В часовому ряді взагалі виділяють чотири основні компоненти: тренд, циклічна, сезонна та випадкова компоненти [11].

Основна тенденція – тренд. Тренд – це загальна систематична лінійна або нелінійна компонента, яка може змінюватися в часі.

Циклічні коливання – цикли кон’юнктури. Циклічний компонент пояснює відхилення від тренда з періодичністю від двох до десяти років. Зазвичай цей компонент може змінюватися за довжиною періоду та інтенсивністю та добре корелює з циклом ділової активності. На підйомі ділової активності значення відгуку вищі за суто трендові, а в періоди спаду та стагнації – помітно нижчі від очікуваних за трендом [7].

Сезонні коливання – коливання, що періодично повторюються в певний час кожного року, місяця, дня та години. Сезонний компонент – це періодично повторювана компонента. Вона визначає короткоперіодичні коливання, пов’язані саме зі змінами внутрішньорічної активності, що повторюються через більш-менш фіксовані моменти часу [14]. Природно зв’язати сезонний компонент із впливом традицій (сезонні та різдвяні розпродажі), соціальних звичок (висока активність у курортному бізнесі в літню пору та «мертві сезони» в інші періоди), релігійних чинників (різдвяні поїздки до родичів або друзів у християнських країнах, продажі харчових продуктів і громадське харчування під час мусульманського поста) і навіть періоди важко передбачуваної погоди (продаж морозива та прохолодних напоїв, діяльність гірськолижних курортів).

Тренд, циклічні та сезонні коливання є регулярною складовою рядів динаміки [5].

Нерегулярний компонент викликає відхилення від ходу відгуку, обумовленого трендовою, циклічною та сезонною складовими. Вона може бути розглянута як випадкова складова і тому не прогнозована. У термінах статистики випадкову складову можна вважати помилкою спостереження й обробляти аналогічно до випадкових помилок вимірювань у статистиці. Випадкову складову пов’язують з випадковими явищами зовнішнього світу.

Нерегулярні складові можна розділити на дві групи: спорадичні зміни, викликані війною або катастрофою; випадкові зміни, що є результатом дії великої кількості відносно незначних другорядних чинників [5].

Ще один компонент – інтервенція, під якою розуміють короткочасний вплив на часовий ряд. Часовий ряд може бути поданий такими математичними моделями:

$$\text{Мультиплікативна: } Y(t) = T(t) \times S(t) \times C(t) \times I(t);$$

$$\text{Адитивна: } Y(t) = T(t) + S(t) + C(t) + I(t);$$

Тут $Y(t)$ є спостереженням, а $T(t)$, $S(t)$, $C(t)$ і $I(t)$ є відповідно трендом, сезонною, циклічною і випадковою компонентами в момент часу t .

Мультиплікативна модель ґрунтується на припущенні, що чотири компоненти часового ряду не обов'язково є незалежними і можуть впливати один на одного, тоді як в адитивній моделі передбачається, що чотири компоненти є незалежними один від одного [19].

Поняття стаціонарності стохастичного процесу можна уявити як форму статистичної рівноваги. Статистичні властивості, такі як середнє і дисперсія стаціонарного процесу, не залежать від часу. Існують два типи стаціонарних процесів.

Процес $\{x(t), t = 0, 1, 2, \dots\}$ є сильно стаціонарним або строго стаціонарним, якщо функція розподілу випадкових величин $\{x_{t-s}, x_{t-s+1}, \dots, x_t, \dots, x_{t+s-1}, x_{t+s}\}$ не залежить від t для всіх s .

Таким чином, для сильно стаціонарного процесу спільний розподіл будь-якого можливого набору випадкових величин з процесу не залежить від часу.

Стохастичний процес називається слабо стаціонарним порядку k або інтегрованим порядку k ($I(k)$), якщо випадкові моменти процесу до цього порядку залежать тільки від часових різниць, а не від часу входження даних, що використовуються для оцінки моментів.

Важливо відзначити, що ні сильна, ні слабка стаціонарність не передбачає іншого. Однак слабо стаціонарний процес, що слідує за нормальним розподілом, також сильно стаціонарний [25].

Статистика та економетрика використовують одномірні або багатомірні регресійні моделі часових рядів для моделювання змінних та їх взаємозв'язків. Ці моделі ґрунтуються на методології Бокса-Дженкінса і фундаментальним

припущенням для їх використання є стаціонарність часових рядів. Однак на практиці ця умова не завжди виконується, оскільки більшість економічних часових рядів є нестаціонарними. Ці часові ряди позначаються як інтегровані порядку d , коли вони є стаціонарними після диференціювання d разів.

Тому побудова аналітичних моделей вимагає ідентифікації порядку інтегрованості або порядку диференціації d [12].

Економетристи використовують кілька підходів для визначення цього порядку. Найпростішим методом є оцінка графіків часових рядів. Оригінальні часові ряди порівнюються з рядами перших та других різниць. Хоча цей метод має суб'єктивний характер, він є дуже ефективним у багатьох випадках, особливо для досвідчених аналітиків.

Тест Дікі-Фуллера.

У статистиці тест Дікі-Фуллера використовують для перевірки нульової гіпотези про наявність одиничного кореня в авторегресійній моделі. Альтернативна гіпотеза варіюється в залежності від версії тесту, але зазвичай означає наявність стаціонарності або тренд-стаціонарності. Тест названий на честь статистиків Девіда Дікі та Вейна Фулера, які розробили його в 1979 році [31].

Тест Дікі-Фуллера є одним з найбільш відомих і найбільш широко використовуваних тестів на одиничний корінь. Він базується на моделі авторегресійного процесу першого порядку:

$$y_t = \phi_1 y_{t-1} + \varepsilon_t, t = 1, \dots, T,$$

де ϕ_1 - параметр авторегресії, ε_t - випадкова складова моделі, що відповідає характеристикам процесу білого шуму.

Нульова гіпотеза $H_0: \phi_1 = 1$, тобто процес містить одиничний корінь і тому нестаціонарний, і позначається як $I(1)$.

Альтернативна гіпотеза $H_1: |\phi_1| < 1$, тобто процес не містить одиничного кореня і є стаціонарним, $I(0)$.

Для обчислення тестової статистики для тесту DF використовується рівняння, яке отримується, якщо y_{t-1} вираховується з обох сторін рівняння:

$$\Delta y_t = \beta y_{t-1} + \varepsilon_t,$$

де $\beta = \phi_1 - 1$.

Тому перевірка гіпотези про одиничний корінь тут означає перевірку нульової гіпотези про рівність нулю коефіцієнта β . Оскільки випадок «вибухових» процесів виключається, то тест є одностороннім, тобто альтернативною гіпотезою є гіпотеза про те, що коефіцієнт β менше нуля. Статистика тесту (DF-статистика) – це звичайна t-статистика для перевірки значущості коефіцієнтів лінійної регресії. Однак розподіл цієї статистики відрізняється від класичного розподілу t-статистики, тобто розподілу Стьюдента або асимптотично нормального розподілу [34].

Розподіл DF-статистики виражається через винерівський процес і називається розподілом Дікі-Фуллера. Гіпотеза про рівність нулю коефіцієнта β відхиляється, якщо $t_{\text{факт}} < t_{\text{критич}}$, $t_{\text{факт}}$ – це обчислена за емпіричними даними t-статистика Стьюдента, а $t_{\text{критич}}$ – це критичне значення статистики Дікі-Фуллера. Існує три версії тесту (тестових регресій):

1. без константи й тренду:

$$\Delta y_t = \beta y_{t-1} + \varepsilon_t$$

2. з константою але без тренду:

$$y_t = \beta_0 + \phi_1 y_{t-1} + \varepsilon_t$$

3. з константою та лінійним трендом:

$$y_t = \beta_0 + \beta_1 t + \phi_1 y_{t-1} + \varepsilon_t$$

Для кожної з трьох тестових регресій існують свої критичні значення DF-статистики. Якщо значення статистики лежить лівіше критичного значення (критичні значення є негативними) за цього рівня значущості, то нульова гіпотеза при одиничному корені відхиляється, а процес визнається стаціонарним. В іншому разі гіпотеза не відкидається, а процес може містити

одиночні коріння, тобто бути нестационарним часовим рядом. Тест Дікі-Фуллера, як і багато інших тестів, перевіряє наявність лише одного одиночного кореня. Однак процес може мати теоретично декілька одиночних коренів. Тоді тест може бути некоректним [31].

Оскільки зазвичай передбачається, що більше трьох одиночних коренів навряд чи можуть зустрічатися в реальних економічних часових рядах, теоретично обґрунтованим є тестування перш за все другої різниці ряду. Якщо гіпотеза одиночного кореня для цього ряду відкидається, то тестується одиночний корінь в перших різницях. Якщо на цьому етапі гіпотеза не відкидається, то вихідний ряд має два одиночні кореня. Якщо відкидається, то перевіряється одиночний корінь в самому часовому ряді.

2.2. Використання методології ARIMA (Бокса-Дженкінса) для моделювання фінансових активів

У статистиці та економетриці, зокрема в аналізі часових рядів, модель авторегресійної інтегрованої ковзної середньої (ARIMA) є узагальненням моделі авторегресійної ковзної середньої (ARMA). Обидві ці моделі адаптуються до даних часових рядів або для кращого розуміння даних, або для прогнозування майбутніх точок у ряді [29].

Моделі ARIMA застосовуються в деяких випадках, коли дані показують докази нестационарності в сенсі середнього (але не дисперсії/автоковаріації), де початковий крок розрізнення може бути застосований один або більше разів, щоб усунути нестационарність середньої функції. Коли сезонність відображається в часовому ряді, сезонна різниця може бути застосована для усунення сезонної складової.

Необхідно звернути увагу, що якщо часовий ряд містить передбачуваний підпроцес (чистий синус або комплексно-значний експоненціальний процес), передбачуваний компонент розглядається як ненульовий середній, але періодичний компонент в ARIMA таким чином, щоб він був усунутий сезонною різницею [17].

Частина AR ARIMA вказує на те, що змінна, яка цікавить, регресує на її власні відсталі (тобто попередні) значення. Частина MA вказує на те, що помилка регресії є лінійною комбінацією членів помилки, значення яких мали місце одночасно та в різний час у минулому. I (для «інтегрованого») вказує на те, що значення даних було замінено на різницю між їхніми значеннями та попередніми значеннями [16].

Мета кожної з цих функцій полягає в тому, щоб модель якомога краще відповідала даним. Несезонні моделі ARIMA зазвичай позначаються як ARIMA (p, d, q), де параметри p, d і q є невід'ємними цілими числами, p – порядок (кількість часових лагів) авторегресійної моделі, d – ступінь різниці (кількість разів, коли дані мали віднімання минулих значень), і q є порядком моделі ковзного середнього [24].

Сезонні моделі ARIMA зазвичай позначаються як ARIMA (p, d, q)(P, D, Q)_m, де m означає кількість періодів у кожному сезоні, а великі літери P, D, Q позначають авторегресію, різницю, і умови ковзного середнього для сезонної частини моделі ARIMA [42].

Моделі ARIMA базуються на припущенні, що минулі значення мають деякий залишковий вплив на поточні або майбутні значення. Наприклад, інвестор, який використовує модель ARIMA для прогнозування цін на акції, припустив би, що на нових покупців і продавців цих акцій впливають нещодавні ринкові транзакції, коли вони вирішують, скільки запропонувати або прийняти за цінний папір.

Хоча це припущення буде справедливим за багатьох обставин, це не завжди так [16]. Наприклад, за роки до фінансової кризи 2008 року більшість інвесторів не знали про ризики, пов'язані з великими портфелями іпотечних цінних паперів (MBS), якими володіють багато фінансових компаній.

У ті часи інвестор, який використовує модель авторегресії для прогнозування показників американських фінансових акцій, мав би вагомні підстави передбачити триваючу тенденцію стабільних або зростаючих цін на акції в цьому секторі. Однак, як тільки стало відомо, що багатьом фінансовим

установам загрожує неминучий крах, інвестори раптово стали менше турбуватися про останні ціни на ці акції, а набагато більше – про їх основний ризик. Таким чином, ринок швидко переоцінив фінансові акції до набагато нижчого рівня, що повністю збентежило б авторегресійну модель.

ARIMA поєднує функції авторегресії з функціями ковзних середніх. Наприклад, авторегресійний процес AR – це процес, у якому поточне значення базується на безпосередньо попередньому значенні, тоді як процес AR – це процес, у якому поточне значення базується на двох попередніх значеннях [5].

Ковзне середнє – це обчислення, яке використовується для аналізу точок даних шляхом створення серії середніх різних підмножин повного набору даних, щоб згладити вплив викидів. Завдяки цій комбінації методів моделі ARIMA можуть враховувати тренди, цикли, сезонність та інші нестатичні типи даних під час прогнозування.

Прогнозування ARIMA досягається шляхом підключення даних часових рядів для цікавої змінної. Статистичне програмне забезпечення визначить відповідну кількість затримок або величину різниці, яка буде застосована до даних, і перевірить їх на стаціонарність. Потім воно виведе результати, які часто інтерпретуються подібно до моделі множинної лінійної регресії [5].

Деякі сильні сторони моделей ARIMA:

- вони забезпечують дуже гнучке сімейство моделей, яке є закритим для додавання моделей;
- властивості моделей добре зрозумілі і прості у використанні;
- їх короткострокові прогнози важко перевершити більш складними моделями;
- вони можуть бути легко узагальнені (наприклад, вони можуть генерувати незворотній час).

Деякі недоліки моделей ARIMA:

- вони мало допомагають для довгострокового прогнозування;

- експоненціальне згладжування легше використовувати і часто може забезпечувати зручне наближення до реальності та прогнози майже такі ж хороші, як і у повністю обладнаної ARIMA апроксимації;
- структура ARIMA моделі може бути незручними у використанні (наприклад, більшість моделі не є інваріантними до змін інтервалу вибірки, що важливо для управління процесом або до тимчасової агрегації) [24].

Процесом з авторегресією AR(p) називається послідовність рівнів ряду $\{y_i\}$, яка задовольняє рівняння:

$$y_{n+1} = a_0 + a_1 y_n + a_2 y_{n-1} + \dots + a_{k-1} y_{n-k+1} + a_k y_{n-k} + e,$$

де $\{a_i\}$ – параметри авторегресії, k – порядок авторегресійної моделі, y_i – рівні ряду [5].

Проілюструємо даний метод на прикладі авторегресійної моделі другого порядку:

$$y(k) = 2,1 + 0,8y(k-1) - 0,15y(k-2).$$

Сформуємо однорідне рівняння $\lambda^2 - 0,8\lambda + 0,15 = 0$.

Воно має два дійсних розв'язки $\lambda_1 = 0,5$ та $\lambda_2 = 0,3$, тому часткові розв'язки авторегресійного рівняння мають вигляд:

$$y_1^h(k) = A_1 \times (0,5)^k \text{ та } y_2^h(k) = A_2 \times (0,3)^k.$$

Знаходимо частинний розв'язок для однорідних збурень: $y^p = 2,1 + 0,8y^p - 0,15y^p \Rightarrow y^p = 6$.

Об'єднуємо однорідний розв'язок з частинними: $y(k) = 6,0 + A_1 \times (0,5)^k + A_2 \times (0,3)^k$, де A_1, A_2 – довільні константи, які визначаються при заданих початкових умовах [5].

Нехай задані початкові умови $y(0) = 1$; $y(1) = 2$, тоді $\{1,0 = 6,0 + A_1 + A_2 \times 2,0 = 6,0 + 0,5 \times A_1 + 0,3 \times A_2\}$, звідки $A_1 = -12,5$; $A_2 = 7,5$.

Остаточний розв'язок матиме вигляд: $y(k) = 6,0 - 12,5 \times (0,5)^k + 7,5 \times (0,3)^k$.

Для рівнянь вищого порядку можна виконати аналогічні перетворення, але виникають труднощі, пов'язані з розв'язанням алгебраїчного рівняння високого ступеня. Даний метод не покращує прогностичні можливості авторегресійних моделей у порівнянні з ітераційним обчисленням прогнозованих значень ряду, але розв'язок авторегресійного рівняння відносно невідомої функції $y(t)$ дозволяє провести класифікацію можливих режимів авторегресійних процесів.

Процесом ковзного середнього $MA(q)$ називається послідовність рівнів ряду $\{y_i\}$, яка задовольняє рівняння:

$$y_{n+1} = b_0 + b_1 e_n + b_2 e_{n-1} + \dots + b_{k-1} e_{k-2} + b_k e_{k-1},$$
 де $\{b_i\}$ – параметри авторегресії, k – порядок авторегресійної моделі, e_i – залишки.

Для моделювання та прогнозування частіше використовуються комбіновані $ARMA(p, q)$ -моделі.

Процесом авторегресії – ковзного середнього $ARMA(p, q)$ називається послідовність рівнів ряду $\{y_i\}$, яка задовольняє рівняння:

$$y_{n+1} = a_0 + a_1 y_n + a_2 y_{n-1} + \dots + a_{k-1} y_{k-2} + a_k y_{k-1} + b_0 + b_1 e_n + b_2 e_{n-1} + \dots + b_{k-1} e_{k-2} + b_k e_{k-1},$$
 де $\{a_i\}$, $\{b_i\}$ – параметри авторегресії, p – порядок авторегресійної складової моделі, q – порядок складової ковзного середнього, e_i – залишки [24].

2.3. Метод експоненціального згладжування

Досить ефективним і надійним методом прогнозування є метод експоненціального згладжування. Основні переваги цього методу полягають у можливості оцінювання вагів початкової інформації, у простоті обчислювальних операцій, в гнучкості опису різних динамічних процесів.

Метод експоненціального згладжування дає можливість отримати оцінку параметрів тренду, що характеризує не середній рівень процесу, а тенденцію, що склалася до моменту останнього спостереження. Після появи робіт Р. Брауна найчастіше даний метод застосовується для реалізації короткострокових і середньострокових прогнозів. Для методу експоненціального згладжування

основним і найбільш важким моментом є вибір параметра згладжування α , початкових умов і ступеня прогнозуючого полінома [35].

Метод експоненціального згладжування, що є узагальненням методу ковзного середнього, дозволяє побудувати такий опис процесу, при якому новим спостереженням додаються більші ваги в порівнянні з старими, причому ваги спостережень спадають за експонентою.

Важливу роль в методі експоненціального згладжування грає вибір оптимального параметра згладжування α , оскільки саме він визначає оцінки коефіцієнтів моделі і результати прогнозу. Вибір параметра α доцільно пов'язувати з точністю прогнозу, тому для більш обгрунтованого вибору α можна використовувати процедуру узагальненого згладжування, яка дозволяє отримати співвідношення, що пов'язують дисперсію прогнозу і параметр згладжування. Істотним для практичного використання є питання про вибір порядку прогнозуючого полінома, що багато в чому визначає якість прогнозу [5].

Даний метод є одним з найбільш популярних і широко застосовується в практиці прогнозування. Один з найбільш перспективних напрямів розвитку даного методу є метод різницевого прогнозування, в якому саме експоненціальне згладжування розглядається як окремий випадок.

Головний недолік цього методу полягає в тому, що часовий ряд розглядається ізольовано від інших явищ. Крім того, точність прогнозу помітно падає при довгостроковому прогнозуванні.

Слабким місцем усіх адаптивних методів і методів експоненціального згладжування є підбір відповідного до даного конкретного завдання параметра згладжування (адаптації) α . Навіть при оптимальному підборі параметра модель Брауна поступається в точності прогнозу ARIMA (0,1,1) моделі [15].

Недоліками методу найменших квадратів є використання процедури оцінки, що передбачає обов'язкове виконання цілого ряду передумов, порушення яких може привести до значних помилок: 1) випадкові похибки повинні мати нульову середню та скінченні дисперсії і коваріації; 2) кожен

вимір випадкової похибки характеризується нульовим середнім, не залежним від значень спостережуваних змінних; 3) дисперсії кожної випадкової похибки є постійними, їх величини незалежні від значень спостережуваних змінних (гомоскедастичність); 4) відсутність автокореляції похибок, тобто значення похибок різних спостережень мають бути незалежні один від одного; 5) випадкові похибки повинні відповідати нормальному закону розподілу; 6) значення ендогенної змінної x не містять похибок вимірювання і мають скінченні середні значення та дисперсії [34].

Вибір моделі у кожному конкретному випадку здійснюється за цілим рядом статистичних критеріїв, наприклад за дисперсією або кореляційним відношенням. Класичний метод найменших квадратів передбачає рівноцінність початкової інформації в моделі. У реальній практиці майбутня поведінка процесу значно більшою мірою визначається більш новими спостереженнями, ніж минулими. Ця властивість вказує на зменшення цінності ранньої (більш віддаленої у часі) інформації.

Важливим моментом отримання прогнозу за допомогою методу найменших квадратів є оцінка адекватності отриманого результату. Для цієї мети використовується цілий ряд статистичних характеристик:

- оцінка стандартної помилки;
- середня відносна помилка оцінки;
- середнє лінійне відхилення;
- кореляційне відношення для оцінки надійності моделі.

Недоліки методу найменших квадратів обумовлені жорсткою фіксацією тренду. Надійний прогноз при цьому можна отримати тільки на невеликий період упередження [35].

Недоліком методу адаптивного згладжування є те, що тільки при дуже довгих рядах можна отримати надійний прогноз на інтервал більший, ніж при звичайному експоненціальному згладжуванні. Для даного методу немає строгої процедури оцінки необхідної або достатньої довжини початкової інформації.

Проблеми з методом Бокса-Дженкінса пов'язані з неоднорідністю часових рядів і складністю практичної реалізації методу. В цілому, результати застосування традиційних технологій оцінювання і прогнозування фінансового стану компаній можна назвати обмеженими. Обмеженість цих методів полягає у їх залежності від початкових умов і відсутності гнучкості [1]. Жорсткі статистичні припущення про властивості часових рядів обмежують можливості методів математичної статистики, теорії розпізнавання образів, теорії випадкових процесів, нечітких методів. Вони не здатні враховувати те, що відносна значущість окремих показників фінансової звітності та чинників, що їх визначають, на практиці змінюються з часом, іноді дуже різко і непередбачувано.

Окрім цього, традиційні підходи характеризуються обмеженою інформативністю, оскільки призначені для опису якісних чинників або закономірностей в кількісних термінах. Таким чином, на зміну традиційним технологіям повинні прийти нові підходи, ефективніші в умовах структурної нестабільності світової економіки [39].

Лабораторна робота «Моделі авторегресії-ковзного середнього»

Для побудови прогнозів на основі авторегресійних моделей реалізовані процедури у багатьох статистичних пакетах (Statistica, Matlab та багато інших).

В даній лабораторній роботі обрано Excel як найбільш наочний засіб, який дозволяє студентам, не знайомих з особливостями авторегресійних моделей, більш глибоко усвідомити їх суть і можливості при побудові прогнозів [5].

Лист Excel розробляється наступним чином. В секції «Дано» необхідно задавати наступні параметри:

1. значення початкового моменту навчальної вибірки t_0 . При зміні значення цього параметра буде змінюватись фрагмент вибірки для побудови прогнозної моделі;

2. значення кроку дискретизації Δt . Даний параметр містить ціле число – крок дискретизації в одиницях інтервалу дискретизації вхідного ряду. Задає розрідження вхідного ряду.

Секцію «Дано» треба доповнити наступними розрахунковими параметрами:

1. параметри моделі a_0, a_1, \dots, a_m ;
2. середнє лінійне відхилення (MAE) на навчальній вибірці;
3. середньоквадратичне відхилення (MSE) на навчальній вибірці;
4. дисперсія процесу;
5. коефіцієнт детермінації;
6. середнє лінійне відхилення (MAE) на перевірочній вибірці;
7. середньоквадратичне відхилення (MSE) на перевірочній вибірці.

Секцію «Таблиця даних» пропонується розділити на 2 частини. Перша буде містити вихідні дані (дати та значення показника, що прогнозується), а друга – формується з першої у вигляді векторів лагових змінних із заданими початковим значенням t та кроком Δt :

1. час t ,
2. лаговий вектор $x_t = (y_{t-(m-1)\Delta t}, \dots, y_{t-\Delta t})$, що відповідає моменту часу t .

Значення y_t є останньою координатою лагового вектора,

3. наступне значення, що відповідає моменту часу $t + \Delta t$.

Використовується як фактичні значення відгуку авторегресійної моделі при заданому вхідному векторі,

4. прогнозний ряд $y(x_t)$,
5. модулі відхилень,
6. квадрати відхилень.

Розглянемо приклад листа Excel для побудови авторегресійної моделі. Стовець часу t задає послідовність індексів, що відповідають часу поточного лагового вектору. Тому значення t_0 в секції «Дано» повинне бути задане з відступом, щоб даних вистачило для першого вектору, враховуючи заданий

крок Δt . Перше значення стовця часу необхідно взяти рівним t_0 з секції «Дано» [5].

Наступні комірки, починаючи з D5, заповнюються значеннями, що збільшуються з кроком 1, це дозволить розширити об'єм навчальної вибірки, в порівнянні з розрідженням. Для комірки D5 формула матиме вигляд $=D4+1$.

Таким чином, компоненти лагового вектора будуть формуватись із значень вихідного ряду через інтервал Δt , а між лаговими векторами інтервал буде рівний одному кроку вихідної дискретизації. У комірку I2, яка знаходиться зліва від J2, записується формула $=J2+Q2$, при якій минула величина лагу збільшується на Δt . Усі формули зліва від розглянутих отримуються копіюванням формули в комірці I2 у комірки E2... H2. Для значення відгуку системи (комірка K2) формула матиме вигляд $=J2-Q2$. Таким чином можна побудувати лагову модель не тільки для розмірності лагового вектору 6, але і для будь-якого. При цьому необхідно правильно спланувати розміщення лагових векторів на листі. При зміні значення Δt , усі компоненти перерахуються [5].

Далі необхідно сформулювати значення лагових векторів у таблиці. Наприклад, для комірки J4, яка містить останню координату першого лагового вектору, формула матиме вигляд $=ИНДЕКС(\$C\$2:\$C\$13510;\$D4-J2;1)$. Формули для попередніх координат лагових векторів можна отримати, копіюванням комірки J4 в усі комірки E4... K4. Формула для майбутнього значення (відгуку системи, наприклад, для комірки K4) також може мати аналогічний вигляд, так як абсолютні та відносні посилання налаштовані на можливість копіювання як у вертикальному, так і в горизонтальному напрямі. Усі інші формули таблиці заповнені аналогічно. Після заповнення першого рядку, формули інших рядків можуть бути скопійовані з першого [5].

Після настання моменту прогнозу, подальше використання минулої формули для формування лагового вектору стає неможливим, так як недоступні нові дані для побудови нових лагових векторів. Тому лагові вектори будуємо, використовуючи обчислені прогнози.

Оцінювання параметрів ARIMA-моделі та прогнозування часового ряду можна виконати за допомогою Garch Toolbox, яка включена в Econometrics Toolbox системи Matlab.

2.4. Векторні моделі авторегресії Кевіна Коце

Моделі векторної авторегресії (VAR) широко використовуються в дослідженнях часових рядів для вивчення динамічних зв'язків, які існують між змінними. Крім того, вони також є важливими інструментами прогнозування, які використовуються більшістю макроекономічних установ або установ, що формують політику. Розробка цих моделей є предметом багатьох поточних досліджень, які слідують за основоположним внеском Сімса (1972 р.), який встановив структуру для моделювання ендогенних змінних у багатофакторних умовах [40].

Після розробки цієї основи було розроблено низку статистичних тестів для визначення взаємозалежностей і характеру динамічних зв'язків між змінними. Крім того, новітні розробки включають використання структурних розкладів, знакових обмежень, включення змінних у часі параметрів, структурних розривів та стохастичної волатильності [40].

Структура моделей VAR дозволяє пояснити значення ендогенних змінних на основі їх минулих спостережуваних значень. Ці моделі дещо відрізняються від моделей структурної векторної авторегресії (SVAR), які дозволяють явно моделювати сучасну взаємозалежність між лівими змінними. Крім того, важливі внески Енгла та Грейнджера (1987) надали економетрикам потужні інструменти для моделювання коінтегрованих зв'язків між змінними.

Модель VAR є багатоваріантною версією моделі AR. Модель VAR можна записати як модель VAR першого порядку за допомогою матриці супутньої форми. Якщо всі власні значення матриці супутньої форми менші за одиницю за абсолютним значенням, тоді модель VAR є стабільною (коваріація стаціонарна). Стабільний VAR модель може бути обернена і записана як векторна модель ковзного середнього нескінченного порядку. Параметри в

моделі VAR можна оцінити за допомогою OLS, використовуючи метод рівняння за рівнянням. Відповідно до стандартного припущення, оцінка МНК буде подібна до оцінки максимальної правдоподібності [43].

Прогнозування зі стабільною моделлю VAR буде зближуватися до безумовного середнього значення змінних у моделі, а матриця MSE помилок прогнозу зближуватиметься до безумовної дисперсії змінних у моделі. Прогнози щільності та інтервали прогнозів можуть бути побудовані на основі припущень про нормальність помилок. Можна використати тест причинності Грейнджера, який передбачає перевірку того, чи допомагають запізнілі значення певної змінної в системі VAR передбачити одну з інших ендогенних змінних у системі. Такий тест можна просто провести за допомогою F – процедури перевірки, де нульова гіпотеза не є причинністю Грейнджера.

Векторні авторегресійні моделі ковзного середнього (VARMA) та моделі VAR використовуються в економетриці, зокрема в аналізі часових рядів, для виявлення перехресних кореляцій між рядами, що виходять за рамки ізольованого аналізу рядів даних [41].

На тлі успіхів одновимірних ARMA-моделей у розробці прогнозів було здійснено перехід до VARMA-моделей для багатовимірного контексту. Мета їх запровадження відповідає визначенню Грейнджера і пов'язана з підвищенням точності прогнозування за рахунок використання моделі із взаємопов'язаними змінними. Вперше ці моделі були використані у 1957 році. З того часу вони були предметом досліджень багатьох економістів. В усіх цих випадках кількість змінних була невеликою, не більше 3 змінних. Іншою проблемою була неможливість ідентифікувати VARMA репрезентації [41].

Ці проблеми були проаналізовані Е. Hannan (1970, 1976, 1979, 1981) та W. Dunsmuir. Е. Hannan та М. Deistler (1988) були першими, хто надав теоретичне представлення репрезентативності моделей VARMA. Lautkepohl (1991, 2002) та Reinsel (1993) проаналізували прогнози на основі цих моделей. ARIMA-моделі широко використовуються на практиці, але вони не можуть описати динаміку всіх взаємозв'язків між обраними змінними [41].

VARMA-моделі є результатом теореми Вольда про декомпозицію для багатовимірних стаціонарних рядів. С. Kascha та С. Trenkler стверджують, що існує дуже мало досліджень, присвячених оцінці ефективності прогнозів на основі VARMA або коінтегрованих VARMA моделей. D. Poskitt, G. Athanasopoulos та F. Vahid оцінювали точність прогнозів, зроблених за допомогою VARMA моделей і вони отримали результати, що перевищують результати VAR моделей [36].

Моделі VARMA були використані багатьма дослідниками, такими як M. Quenouille (1957), E. Hannan (1969) та G. Wilson (1973). Однак перевага надається VAR моделям скінченного порядку, а не VARMA, оскільки в літературі не ставилося питання про їх альтернативність.

T. Cooley та M. Dwyer (1998) стверджували, що моделювання макроекономічних часових рядів з використанням VAR моделей не узгоджується з економічною теорією. Але складність методології VARMA зумовила вибір VAR моделей, результати яких є досить успішними. Функція правдоподібності базується на припущенні про нормальність і може бути визначена рекурсивно, коли задаються перші p спостережень, а наступні дорівнюють нулю. Починаючи з форми простору станів моделі, ця функція правдоподібності може бути визначена рекурентно. Визначення порядків моделі VARMA (p, q) є досить складним завданням, враховуючи той факт, що параметри повинні відповідати певним обмеженням. С. Kascha та К. Mertens (2009) здійснили порівняльний аналіз ідентифікації структурної форми для VAR та VARMA моделей [36].

В літературі існує декілька методів визначення VARMA. Athanasopoulos та Vahid (2008) визначили дві методології, які можуть бути застосовані для отримання унікальної ідентифікації VARMA. Вони провели порівняння результатів прогнозів, зроблених на основі моделей VARMA.

Перша методологія є розширенням методології G. Tiao та R. Tsay (1989 р.). Друга методика передбачає оцінку індексів Кронекера, що

розраховуються як максимальний ранг кожного рядка з кожного рівняння моделі.

Індекси Кронекера оцінюються за допомогою методу найменших квадратів, який застосовується до регресій. Вони визначаються на другому етапі з використанням критерію відбору моделі [44].

G. Tiao та R. Tsay (1989) запропонували перший метод, який зазнав критики, але потім Athanasopoulos та Vahid (2006) удосконалили його. Їх методика складається з трьох етапів:

1. Ідентифікація скалярних компонентів моделі (СКМ) шляхом застосування канонічних кореляційних тестів між різними наборами змінних;
2. Ідентифікація структурної форми моделі за допомогою тих же тестів та певних логічних висновків;
3. Оцінка моделі з використанням методу максимальної правдоподібності за повною інформацією (FIML).

Методологія G. Tiao та R. Tsay (1989) оцінює параметри у два етапи. На першому етапі отримують відхилення стандартних помилок, які потім коригуються у триетапній версії, запропонованій G. Athanasopoulos та F. Vahid [36].

Всі ці проблеми були вирішені в роботах G. Athanasopoulos та F. Vahid (2008), які запропонували формулу для визначення кількості надлишкових параметрів матриці A . Вони описали процедуру, за допомогою якої певні параметри нормалізуються до 1, але вони відрізняються від тих, які мають бути прирівняні до нуля.

Методологія скалярних компонент, що використовується при побудові VARMA моделей, є досить складною на практиці, але на невеликих часових горизонтах прогнози на основі цих моделей є кращими.

Висновки до розділу 2

Статистика та економетрика використовують одномірні або багатомірні регресійні моделі часових рядів для моделювання змінних та їх взаємозв'язків. Ці моделі ґрунтуються на методології Бокса-Дженкінса і фундаментальним припущенням для їх використання є стаціонарність часових рядів.

На практиці ця умова не завжди виконується, оскільки більшість економічних часових рядів є нестаціонарними. Ці часові ряди позначаються як інтегровані порядку d , коли вони є стаціонарними після диференціювання d разів. Тому побудова аналітичних моделей вимагає ідентифікації порядку інтегрованості або порядку диференціації d .

Економетристи використовують кілька підходів для визначення цього порядку. Найпростішим методом є оцінка графіків часових рядів. Оригінальні часові ряди порівнюються з рядами перших та других різниць. Хоча цей метод має суб'єктивний характер, він є дуже ефективним у багатьох випадках, особливо для досвідчених аналітиків.

В аналізі часових рядів модель авторегресійної інтегрованої ковзної середньої (ARIMA) є узагальненням моделі авторегресійної ковзної середньої (ARMA). Обидві ці моделі адаптуються до даних часових рядів або для кращого розуміння даних, або для прогнозування майбутніх точок у ряді.

Моделі ARIMA застосовуються в деяких випадках, коли дані показують докази нестаціонарності в сенсі середнього, де початковий крок розрізнення може бути застосований один або більше разів, щоб усунути нестаціонарність середньої функції. Коли сезонність відображається в часовому ряді, сезонна різниця може бути застосована для усунення сезонної складової.

ARIMA поєднує функції авторегресії з функціями ковзних середніх. Наприклад, авторегресійний процес AR – це процес, у якому поточне значення базується на безпосередньо попередньому значенні, тоді як процес AR – це процес, у якому поточне значення базується на двох попередніх значеннях.

Ковзне середнє – це обчислення, яке використовується для аналізу точок даних шляхом створення серії середніх різних підмножин повного набору

даних, щоб згладити вплив викидів. Завдяки цій комбінації методів моделі ARIMA можуть враховувати тренди, цикли, сезонність та інші нестатичні типи даних під час прогнозування.

Прогнозування ARIMA досягається шляхом підключення даних часових рядів для цікавої змінної. Статистичне програмне забезпечення визначить відповідну кількість затримок або величину різниці, яка буде застосована до даних, і перевірить їх на стаціонарність. Потім воно виведе результати, які часто інтерпретуються подібно до моделі множинної лінійної регресії.

Моделі векторної авторегресії (VAR) широко використовуються в дослідженнях часових рядів для вивчення динамічних зв'язків, які існують між змінними. Крім того, вони також є важливими інструментами прогнозування, які використовуються більшістю макроекономічних установ або установ, що формують політику. Розробка цих моделей є предметом багатьох поточних досліджень, які слідують за основоположним внеском Сімса (1972 р.), який встановив структуру для моделювання ендогенних змінних у багатофакторних умовах.

Структура моделей VAR дозволяє пояснити значення ендогенних змінних на основі їх минулих спостережуваних значень. Ці моделі дещо відрізняються від моделей структурної векторної авторегресії (SVAR), які дозволяють явно моделювати сучасну взаємозалежність між лівими змінними.

РОЗДІЛ 3

ПРАКТИЧНЕ МОДЕЛЮВАННЯ ФІНАНСОВИХ АКТИВІВ ІЗ ЗАСТОСУВАННЯМ VAR МОДЕЛІ

3.1. Дослідження даних та тренування VAR моделі

Модель векторної авторегресії постає дуже корисною при прогнозуванні масиву часових рядів, а також при дослідженні взаємної інформації між ними. Дана модель збирає інформацію зі своїх власних атрибутів (в даному випадку наших акцій), одночасно навчаючись з атрибутів інших часових рядів, в даному випадку дані, які впливають на ціну акцій.

Ми визначимо такі змінні / часові ряди:

- Ціна ХОМ: ціна акцій компанії Exxon.
- S&P 500: індекс фондового ринку S&P 500.
- Ціна WTI: ціна бареля сирої нафти West Texas Intermediate (США).
- Ціна на нафту марки Brent: ціна бареля сирої нафти марки Brent (ЄС).
- Видобуток (змінна Production): видобуток сирої нафти в США.
- Споживання (змінна Consumption): споживання нафтопродуктів у США.
- Запаси (Inventory (Crude)): запаси сирої нафти в США.
- Витрати на нафтопереробні заводи (Input to Refineries): споживання сирої нафти на нафтопереробних заводах США.
- Запаси (Inventory (Gas)): запаси бензину в США.

Представлені показники вилучатимуться з сайту Yahoo Finance за допомогою інструментарію мови програмування Python. Значення цін акцій компанії Exxon, індексу S&P 500, цін на нафту марок Brent та WTI вилучимо, використовуючи API бібліотеки *yfinance*. Вилучення зайвих стовпців та об'єднання в матрицю показників виконаємо з використанням бібліотеки *pandas*.

```
xom = yf.Ticker("XOM")
sp500 = yf.Ticker("^GSPC")
bre = yf.Ticker("BZ=F")
wt = yf.Ticker("CL=F")
```

(a)

```
histxom = xom.history(period="max")
histsp = sp500.history(period="max")
histbre = bre.history(period="max")
histwt = wt.history(period="max")
exxon = histxom.drop(columns = ['Open', 'High', 'Low', 'Volume', 'Dividends', 'Stock Splits'])
exxon = exxon.rename(columns={'Close': 'XOM Price'})
market = histsp.drop(columns = ['Open', 'High', 'Low', 'Volume', 'Dividends', 'Stock Splits'])
market = market.rename(columns={'Close': 'S&P 500'})
brent = histbre.drop(columns = ['Open', 'High', 'Low', 'Volume', 'Dividends', 'Stock Splits'])
brent = brent.rename(columns={'Close': 'Brent Price'})
wti = histwt.drop(columns = ['Open', 'High', 'Low', 'Volume', 'Dividends', 'Stock Splits'])
wti = wti.rename(columns={'Close': 'WTI Price'})
fin = pd.concat([exxon,market,brent,wti], axis = 1)
```

(б)

Рис. 3.1. Вилучення та фільтрація значень індексів XOM, S&P 500, Brent та WTI (а) та їх фільтрація (б).

Для отримання показників пов'язаних із видобуванням, споживанням та зберіганням нафти, скористаємось ресурсом «Управління енергетичної інформації США». У коді визначимо назви стовпців додаткових змінних та встановимо їх кодові позначення для парсингу.

```

titles = ['Production',
          'Consumption',
          'Inventory (Crude)',
          'Input to Refineries',
          'Inventory (Gas)']

codes = ['PET.MCRFPUS2.M',
         'PET.WRPUPUS2.W',
         'PET.WTTSTUS1.W',
         'PET.WCRRIUS2.W',
         'PET.WGTSTUS1.W']

eia_dat = []

```

Далі здійснюється парсинг відповідних рядів та формування таблиці типу *DataFrame* бібліотеки *pandas*.

```

for i in range(len(codes)):
    url = 'https://api.eia.gov/series/?api_key=' + api_key + '&series_id=' + codes[i]
    r = requests.get(url)
    json_data = r.json()

    if r.status_code == 200:
        print('Success!')
    else:
        print('Error')

    df = pd.DataFrame(json_data.get('series')[0].get('data'),
                      columns = ['Date', titles[i]])
    df.set_index('Date', drop=True, inplace=True)
    eia_dat.append(df)

```

Рис. 3.3. Парсинг рядів за посиланням та формування таблиці.

Відповідні ряди обробляються за допомогою ковзного середнього різного періодів та зрівнюються за довжиною найменшого ряду. У нашому випадку найменше значень мають ряд із значеннями цін на сиру нафту Brent.


```

data['XOM Price'] = data['XOM Price'].rolling(5).mean().round(decimals = 2)
data['S&P 500'] = data['S&P 500'].rolling(5).mean().round(decimals = 2)
data['WTI Price'] = data['WTI Price'].rolling(5).mean().round(decimals = 2)
data['Brent Price'] = data['Brent Price'].rolling(5).mean().round(decimals = 2)
data['Production'] = data['Production'].rolling(60).mean().round(decimals = 2)
data['Consumption'] = data['Consumption'].rolling(15).mean().round(decimals = 2)
data['Inventory (Crude)'] = data['Inventory (Crude)'].rolling(15).mean().round(decimals = 2)
data['Input to Refineries'] = data['Input to Refineries'].rolling(15).mean().round(decimals = 2)
data['Inventory (Gas)'] = data['Inventory (Gas)'].rolling(15).mean().round(decimals = 2)

```

(a)

```

startDate = data['Brent Price'].first_valid_index()
mask = (data.index >= startDate) & (data.index <= endDate)
data = data.loc[mask]

```

(б)

Рис. 3.4. Фільтрація часових рядів шляхом усереднення значень (а) та співставлення всіх за першим ненульовим значенням.

Унаслідок коронавірусної пандемії значення сирової нафти індексу WTI зазнали шокової події, що призвела до негативних значень цін на сирову нафту. У цьому випадку необхідна додаткова фільтрація, оскільки ціни не можуть бути від'ємними.

```

data['WTI Price'] = data['WTI Price'].mask(data['WTI Price'].lt(0)).ffill().fillna(0).convert_dtypes()

```

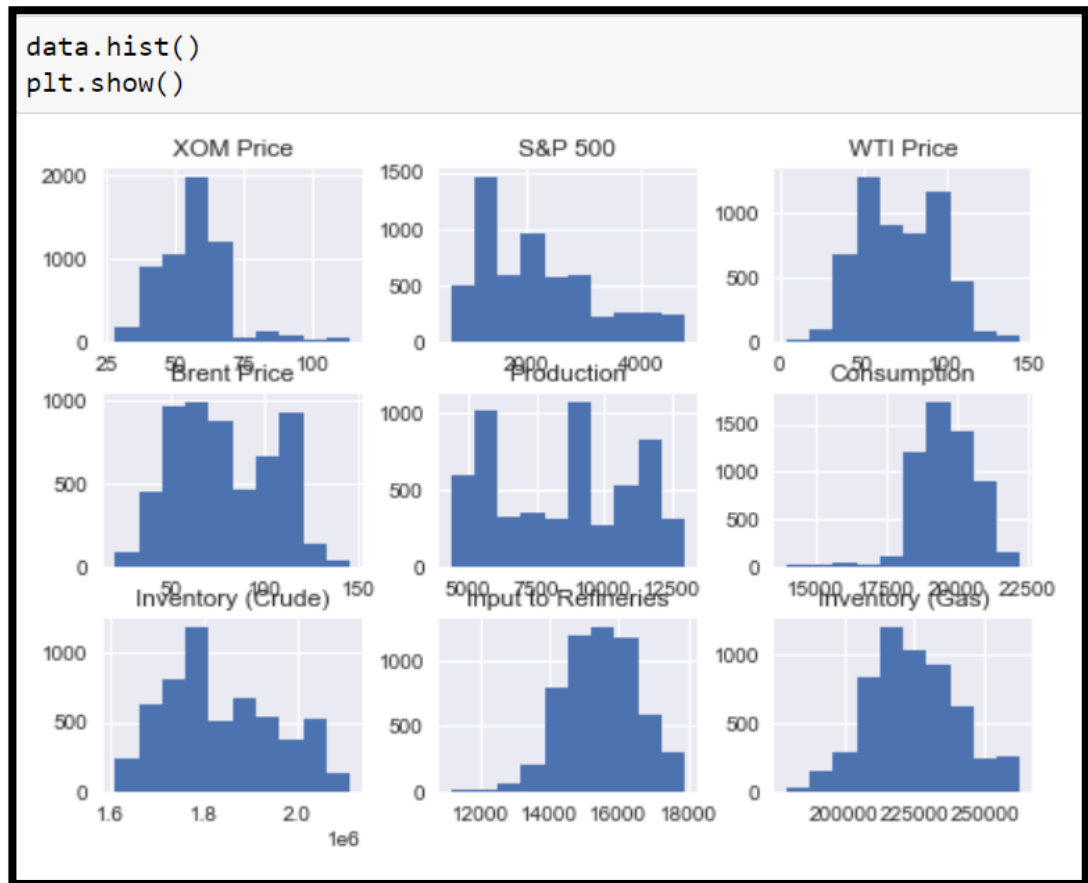
Рис. 3.5. Приклад коду, що замінює від'ємну ціну нульовим значенням.

Візуально розглянемо наші показники. На рис. 3.6 представлено таблицю із останніми значеннями індексів, що використовуються та частотні розподіли.

```
data.tail()
```

Date	XOM Price	S&P 500	WTI Price	Brent Price	Production	Consumption	Inventory (Crude)	Input to Refineries	Inventory (Gas)
2022-11-25	113.12	4006.83	78.57	86.05	11975.0	20442.2	1609793.0	16289.6	209570.93
2022-11-26	113.56	4022.07	77.88	85.29	11975.0	20361.6	1609909.0	16306.8	209774.80
2022-11-27	113.37	4026.58	76.94	84.34	11975.0	20281.0	1610025.0	16324.0	209978.67
2022-11-28	112.61	4013.91	76.8	83.90	11975.0	20200.4	1610141.0	16341.2	210182.53
2022-11-29	112.00	3999.99	77.01	83.53	11975.0	20119.8	1610257.0	16358.4	210386.40

(a)



(б)

Рис. 3.6. Таблиця останніх значень рядів (а) та їх частотні розподіли (б).

Згідно розподілів можна припустити, що майже всі дані є нестационарними. Це проблема багатьох часових рядів, оскільки

нестационарність проявляється в різні середні значення та відхилення в різні моменти часу.

Із даною проблемою можна розібратись за допомогою диференціації $\Delta x = x_i - x_{i-1}$. Але спершу перевіримо нашу гіпотезу про те, що наведені вище дані часових рядів впливають один на одного. Це можна зробити за допомогою тесту Грейнджера на причинність.

Причинність Грейнджера – це статистична концепція причинності, заснована на прогнозуванні. Відповідно до причинно-наслідкового зв'язку згідно Грейнджера, якщо сигнал X_1 «Грейнджер-викликає» (або «G-викликає») сигнал X_2 , то минулі значення X_1 повинні містити інформацію, яка допомагає передбачити X_2 , і дані X_1 повинні в такому разі мати більшу прогнозну силу ніж минулі значення X_2 . Його математичне формулювання базується на лінійному регресійному моделюванні випадкових процесів. Існують більш складні доповнення для нелінійних випадків, однак ці покращення часто важче застосовувати на практиці. У цьому тесті використовуються наступні нульові та альтернативні гіпотези:

Нульова гіпотеза (H_0): часовий ряд X не має прямого зв'язку із часовим рядом Y .

Альтернативна гіпотеза (H_A): часовий ряд X Грейнджер-викликає часовий ряд Y .

Перевіримо причинно-наслідковий зв'язок Грейнджера всіх можливих комбінацій часових рядів. Рядки – це змінні, що прогнозуються, стовпці – предиктори. Значення в таблиці є p -значеннями. P -значення, менші за рівень значущості (0,05), передбачають альтернативну гіпотезу про те, що X має вплив на Y . Тобто, нульова гіпотеза може бути відхилена.

На наступному рисунку представлено код для виконання тесту Грейнджера та відповідні p -значення.

```

maxlag=12
test = 'ssr_chi2test'
def grangers_causation_matrix(data, variables, test='ssr_chi2test', verbose=False):

    df = pd.DataFrame(np.zeros((len(variables), len(variables))), columns=variables, index=variables)
    for c in df.columns:
        for r in df.index:
            test_result = grangercausalitytests(data[[r, c]], maxlag=maxlag, verbose=False)
            p_values = [round(test_result[i+1][0][test][1],4) for i in range(maxlag)]
            if verbose: print(f'Y = {r}, X = {c}, P Values = {p_values}')
            min_p_value = np.min(p_values)
            df.loc[r, c] = min_p_value
    df.columns = [var + '_x' for var in variables]
    df.index = [var + '_y' for var in variables]
    return df

grangers_causation_matrix(data, variables = df.columns)

```

(a)

	XOM Price_x	S&P 500_x	WTI Price_x	Brent Price_x	Production_x	Consumption_x	Inventory (Crude)_x	Input to Refineries_x	Inventory (Gas)_x
XOM Price_y	1.0000	0.0000	0.0000	0.0000	0.0338	0.0020	0.0692	0.0000	0.0081
S&P 500_y	0.0000	1.0000	0.0001	0.0001	0.0010	0.0000	0.0000	0.0207	0.0020
WTI Price_y	0.0000	0.0000	1.0000	0.0000	0.0646	0.0000	0.0000	0.0000	0.0000
Brent Price_y	0.0000	0.0000	0.0000	1.0000	0.0950	0.1342	0.0000	0.0003	0.0000
Production_y	0.0000	0.0053	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
Consumption_y	0.0002	0.0006	0.4100	0.2268	0.0004	1.0000	0.0000	0.0000	0.0000
Inventory (Crude)_y	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000	0.2211	0.0000
Input to Refineries_y	0.0000	0.0163	0.0094	0.0236	0.0000	0.0274	0.1352	1.0000	0.0000
Inventory (Gas)_y	0.0120	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

(б)

Рис. 3.7. Код для виконання тесту Грейджера (а) та побудови матриці p -значень і відповідна матриця отримана з тесту Грейджера (б).

Як ми бачимо вище, більшість значень у матриці менше 0,05, що відкидає нульову гіпотезу про те, що будь-який даний X не є причинною Y . Це означає, що ми можемо сміливо виходити з нашого припущення, що змінні впливають одна на одну.

Тепер розділимо досліджувані дані на навчальний та тестувальний набір. У якості тесту скористаємось останніми 10 значеннями кожного з рядів.

```

nobs = 10
df_train, df_test = data[0:-nobs], data[-nobs:]

print(df_train.shape)
print(df_test.shape)

(5588, 9)
(10, 9)

```

Рис. 3.8. Розбиття вибірки на навчальну і тестувальну та вивід розмірності даних.

Можемо бачити, що в якості навчального набору отримали матрицю розмірністю 5588 днів на 9 змінних, в той час як для тестувального набору отримали останні 10 днів на 9 тих самих параметрів.

Перевіримо значення деяких основних рядів на стаціонарність із використанням тесту Дікі-Фуллера. На рис. 3.9 зображено результати тестів для вихідних часових рядів.

Augmented Dickey-Fuller Test on «XOM Price»

Null Hypothesis: Data has unit root. Non-Stationary.

Significance Level = 0.05

Test Statistic = 0.4298

No. Lags Chosen = 33

Critical value 1% = -3.432

Critical value 5% = -2.862

Critical value 10% = -2.567

=> P-Value = 0.9826. Weak evidence to reject the Null Hypothesis.

=> Series is Non-Stationary.

(a)

Augmented Dickey-Fuller Test on «WTI Price»

Null Hypothesis: Data has unit root. Non-Stationary.

Significance Level = 0.05

Test Statistic = -2.522

No. Lags Chosen = 32

Critical value 1% = -3.432

Critical value 5% = -2.862

Critical value 10% = -2.567

=> P-Value = 0.1102. Weak evidence to reject the Null Hypothesis.

=> Series is Non-Stationary.

(б)

Augmented Dickey-Fuller Test on «Brent Price»

Null Hypothesis: Data has unit root. Non-Stationary.

Significance Level = 0.05

Test Statistic = -2.345

No. Lags Chosen = 32

Critical value 1% = -3.432

Critical value 5% = -2.862

Critical value 10% = -2.567

=> P-Value = 0.1579. Weak evidence to reject the Null Hypothesis.

=> Series is Non-Stationary.

(в)

Рис. 3.9. Результати тесту Дікі-Фуллера для вихідних часових рядів індексу ХОМ (а), WTI (б) та Brent (с).

Можемо бачити, що основні нафтові індекси є нестационарними. Для використання моделі VAR у подальшому потрібно виконати диференціювання першого порядку. Виконаємо та повторимо тест.

```
df_differenced = df_train.diff().dropna()
```

(a)

Augmented Dickey-Fuller Test on «XOM Price»

Null Hypothesis: Data has unit root. Non-Stationary.

Significance Level = 0.05

Test Statistic = -12.8689

No. Lags Chosen = 33

Critical value 1% = -3.432

Critical value 5% = -2.862

Critical value 10% = -2.567

=> P-Value = 0.0. Rejecting Null Hypothesis.

=> Series is Stationary.

(б)

Augmented Dickey-Fuller Test on «WTI Price»

Null Hypothesis: Data has unit root. Non-Stationary.

Significance Level = 0.05

Test Statistic = -11.1401

No. Lags Chosen = 33

Critical value 1% = -3.432

Critical value 5% = -2.862

Critical value 10% = -2.567

=> P-Value = 0.0. Rejecting Null Hypothesis.

=> Series is Stationary.

(в)

Augmented Dickey-Fuller Test on «Brent Price»

Null Hypothesis: Data has unit root. Non-Stationary.

Significance Level = 0.05

Test Statistic = -10.8719

No. Lags Chosen = 33

Critical value 1% = -3.432

Critical value 5% = -2.862

Critical value 10% = -2.567

=> P-Value = 0.0. Rejecting Null Hypothesis.

=> Series is Stationary.

(г)

Рис. 3.10. Виконання процедури диференціювання першого порядку (а) та проведення повторного тесту Дікі-Фуллера для індексів ХОМ (б), WTI (в) та Brent (г).

Тепер можемо бачити, що дані стаціонарні.

Приступимо до підбору оптимального порядку моделі VAR з опорою на такі критерії як інформаційний критерій Баєса (BIC), критерій Акаїке (AIC), фінальна прогнозна похибка Акаїке (FPE) та інформаційний критерій Ханнана-Куїнна (HQIC). Максимальний порядок моделі дорівнюватиме 12.

```
model = VAR(df_differenced)
x = model.select_order(maxlags=12)
x.summary()
```

(а)

	AIC	BIC	FPE	HQIC
0	42.82	42.83	3.955e+18	42.83
1	28.15	28.26	1.683e+12	28.19
2	28.06	28.26	1.538e+12	28.13
3	28.01	28.31	1.460e+12	28.11
4	27.90	28.29	1.306e+12	28.04
5	27.62	28.11	9.870e+11	27.79
6	26.93	27.51	4.937e+11	27.13
7	26.76	27.44	4.185e+11	27.00
8	25.96	26.74	1.883e+11	26.23
9	25.76	26.64	1.547e+11	26.07
10	25.65	26.63	1.382e+11	25.99
11	25.34	26.41*	1.015e+11	25.72*
12	25.33*	26.50	1.002e+11*	25.74

(б)

Рис. 3.11. Код для навчання моделі згідно максимального порядку та виводу результату (а). Значення різних інформаційних критеріїв у відповідності до порядку моделі.

Згідно результатам приведені вище порядок моделі дорівнюватиме 12, оскільки і AIC, і FPE є найнижчими при даному значенні. Ми могли б продовжувати рухатися далі, але оскільки BIC має найнижчий показник на рівні 11, як і HQIC, ми можемо зупинитися на 12 порядку.

Отримаємо порядок моделі, сформуємо базу даних із попередніх 12 значень та виконаємо прогноз на 10 днів уперед.

```
lag_order = model_fitted.k_ar

forecast_input = df_differenced.values[-lag_order:]
```

(a)

```
fc = model_fitted.forecast(y=forecast_input, steps=nobs)
df_forecast = pd.DataFrame(fc, index=df.index[-nobs:], columns=df.columns + '_1d')
df_forecast.tail()
```

(б)

Рис. 3.12. Формування бази даних із 12 попередніх значень кожного з показників (а). Прогнозування показників на 10 днів уперед та формування бази даних (б).

Оскільки над даними проводилась диференціація першого порядку, прогнози мають таке саме представлення і тому потрібно зробити зворотне перетворення прогнозних значень.

```
def invert_transformation(df_train, df_forecast, second_diff=False):
    df_fc = df_forecast.copy()
    columns = df_train.columns
    for col in columns:
        if second_diff:
            df_fc[str(col)+'_1d'] = (df_train[col].iloc[-1]-df_train[col].iloc[-2]) + df_fc[str(col)+'_2d'].cumsum()
            df_fc[str(col)+'_forecast'] = df_train[col].iloc[-1] + df_fc[str(col)+'_1d'].cumsum()
    return df_fc
```

(a)

```
df_results = invert_transformation(df_train, df_forecast, second_diff=False)
```

(б)

Рис. 3.13. Функція для зворотного перетворення (а) та відповідне приведення даних до оригінального масштабу (б).

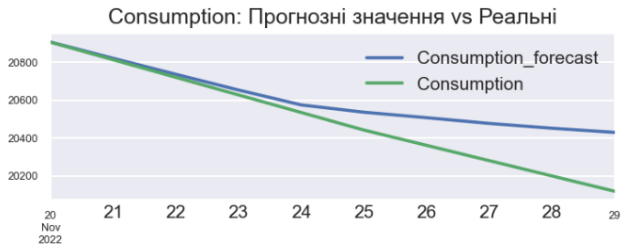
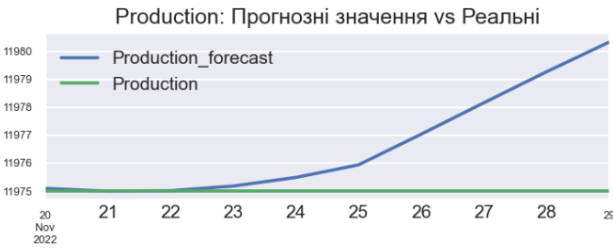
Виведемо результати прогнозів по кожному показнику (Рис. 3.14).



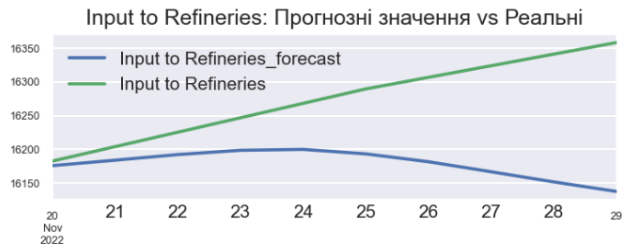
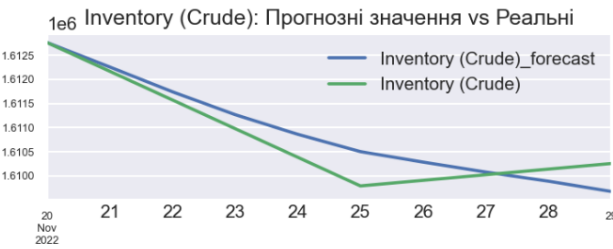
(a)



(б)



(B)



(Г)

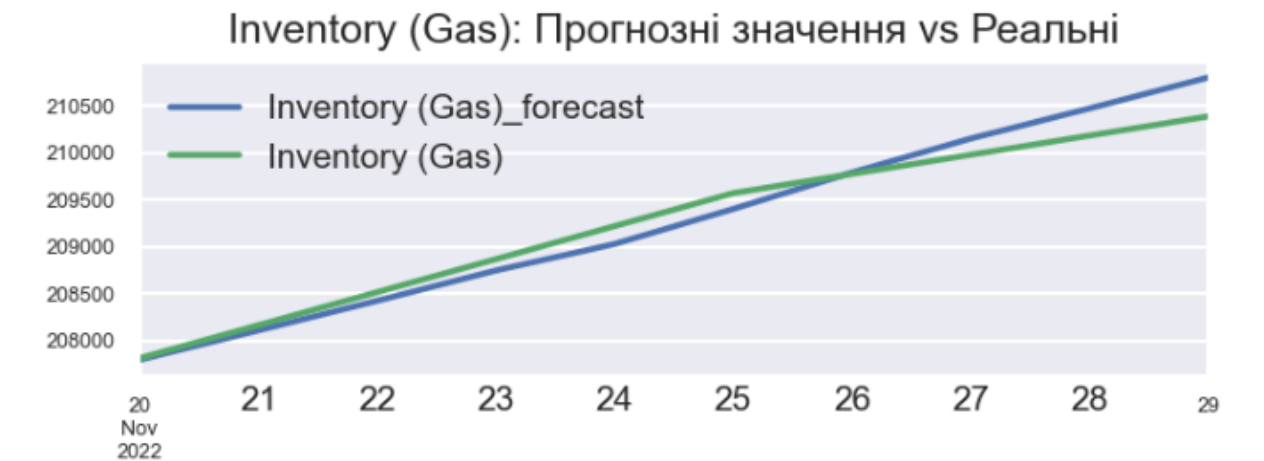


Рис. 3.14. Співставлення прогнозів VAR моделі та реальних значень за останні 10 днів.

Побудувавши графіки, ми бачимо, що більшість наших прогнозів точно відповідають тестувальному набору даних, що вказує на сильну стабільність нашої моделі.

3.2. Побудова прогнозів поза вибірки навчання та тестування VAR моделі

При побудові прогнозів поза вибірки на котрій ми навчались та перевіряли нашу модель, можна взяти увесь набір даних та відштовхуючись від тих параметрів, що ми використовували попередньо, робити прогнози в майбутнє.

Як можна було помітити, прогнозування в рамках VAR моделі, що в нашому випадку є частиною бібліотеки statsmodels, здійснюється за рахунок однієї команди – *model.forecast*. Код даної роботи виглядатиме наступним чином:

```
steps = 10

model = VAR(data_diff)
model_fit = model.fit(12)

pred = model_fit.forecast(data_diff.values[-model_fit.k_ar:], steps=steps)
```

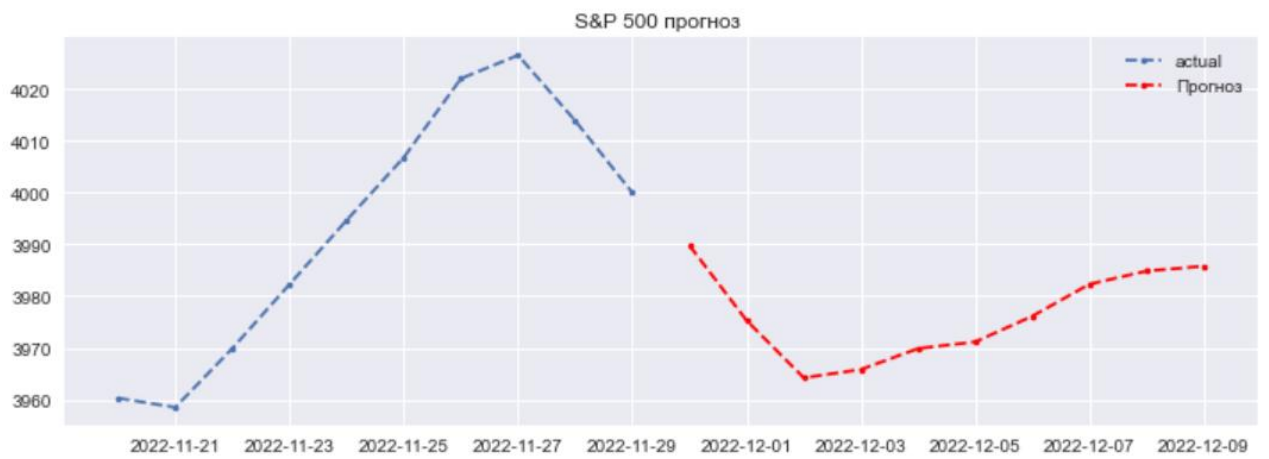
Рис. 3.15. Код ініціалізації, навчання та прогнозування для VAR моделі.

Ми спостерігаємо, що кожен із прогнозів має унікальний тренд. Подібно до того, що ми спостерігали на рис. 3.14, модель вилучає достатньо інформації з кожного часового ряду, щоб розрахувати для кожної змінної унікальний та максимально точний прогноз, що дана модель спроможна надати.

Представимо прогнози досліджуваних рядів, починаючи з 29 листопада по 9 грудня.



(a)



(б)



(B)



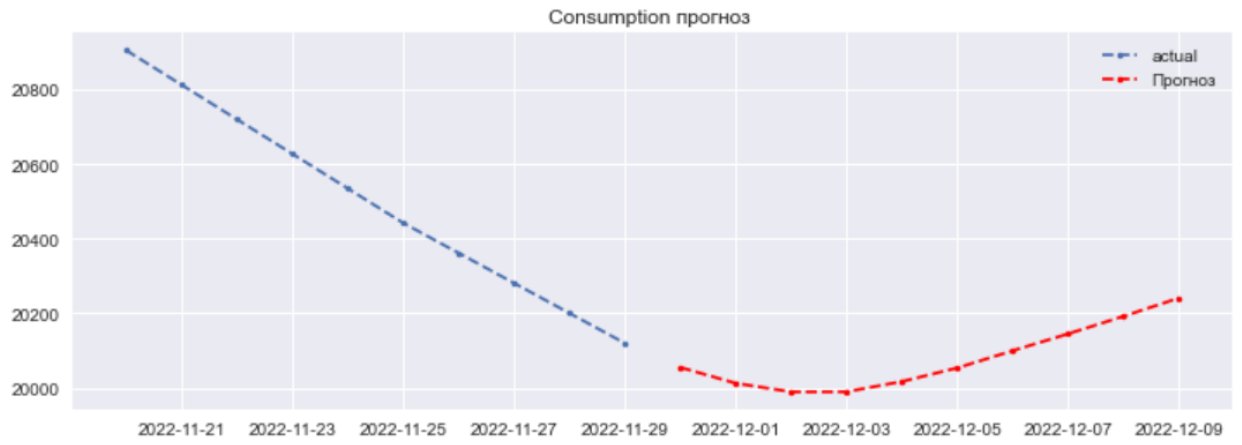
(г)

Рис. 3.16. Прогнозні значення для індексів ХОМ (а), S&P 500 (б), WTI (в), та Brent (г).

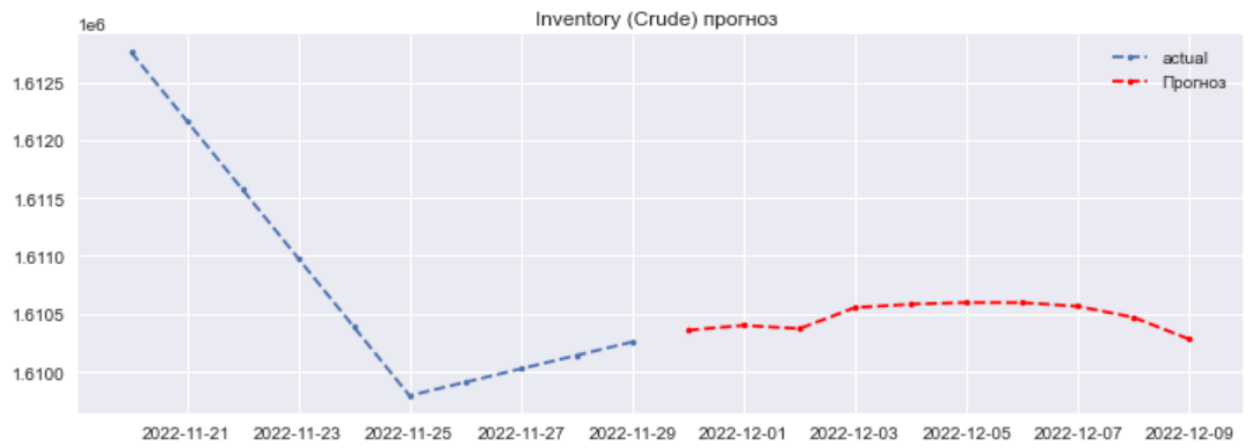
Схоже, що вибір моделі VAR, оптимального порядку моделі та репрезентативних змінних дають високу ефективність при прогнозуванні ціни акцій компанії Еххон у короткостроковій перспективі. Суперечливими поки залишаються результати WTI та Brent. Розглянемо прогнози для інших показників.



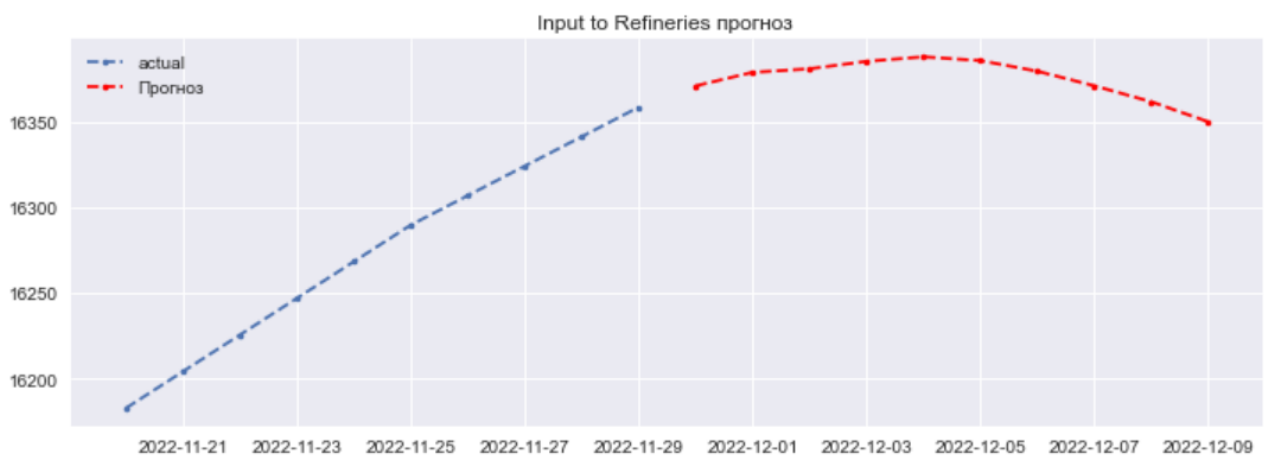
(а)



(б)



(в)



(г)



(д)

Рис. 3.17. Прогнозні значення на 10 днів уперед для видобутку сирої нафти в США (а), споживання нафтопродуктів (б), запасів сирої нафти (в), витрат на нафтопереробні заводи (г) та запасів бензину (д).

Висновки до розділу 3

У даному розділі було проаналізовано часові ряди цін акцій компанії Еххон, індекс фондового ринку S&P 500, ціна бареля сирої нафти West Texas Intermediate (США), ціна бареля сирої нафти марки Brent (ЄС), видобуток сирої нафти в США, споживання нафтопродуктів у США, запаси сирої нафти в США, споживання сирої нафти на нафтопереробних заводах США та запаси бензину в США.

Було розглянуто модель VAR. Даний підхід пройшов тест на причинність по Грейнджеру, було знайдено кореляційні відношення між рядами, проведено тест на стаціонарність, здійснено фільтрацію та нормалізацію даних та досліджено модель за різними інформаційними критеріями. Результати прогнозу на основі тестувального набору даних показують, що даний вид моделі може бути застосований для прогнозування акцій нафтопереробних компаній, цін на барелі сирої нафти різних марок, видобутку енергоресурсів.

ВИСНОВКИ

Під час виконання магістерської роботи ми зробили такі висновки.

Наукове поняття «модель» розуміється як схема, що імітує будову й дію будь-якого процесу. Будь-яка модель здатна заміщати певний об'єкт таким чином, що її вивчення дає нам нову інформацію про нього. Метод моделювання – це метод наукового пізнання, що містить функцію відтворення властивостей і відносин предметів і процесів у ході здійснення дослідження.

VAR моделі – це мультиваріативні динамічні економетричні моделі, що використовуються для відображення лінійних взаємозв'язків між декількома часовими рядами. У векторній авторегресійній моделі всі змінні трактуються ендогенними, що дозволяє відобразити динамічні взаємозв'язки між досліджуваними економічними показниками.

Мультиваріативне авторегресійне моделювання часто використовується для фінансових активів. Фінансові активи – частина активів, яка являє собою фінансові ресурси: грошові кошти, цінні папери, боргові зобов'язання та право вимоги боргу. Під фінансовими активами розуміють: грошові кошти та їх еквіваленти.

Можна зробити висновок, що моделі авторегресії широко використовуються для опису стаціонарних процесів в економічних дослідженнях. Побудова авторегресійних моделей в економіці заснована на такому важливому властивості рядів економічних явищ і процесів, як взаємозалежність рівнів одного і того ж ряду один від одного.

Застосування авторегресійних моделей поширене для опису стаціонарних випадкових процесів. Визначною характеристикою стаціонарних часових рядів є те, що їх розвиток відбувається в незмінних стабільних умовах без вираженої тенденції, тому ймовірнісні властивості рядів залишаються незмінними від часу.

Статистика та економетрика використовують одномірні або багатомірні регресійні моделі часових рядів для моделювання змінних та їх взаємозв'язків.

Ці моделі ґрунтуються на методології Бокса-Дженкінса і фундаментальним припущенням для їх використання є стаціонарність часових рядів.

Економетристи використовують кілька підходів для визначення цього порядку. Найпростішим методом є оцінка графіків часових рядів. Оригінальні часові ряди порівнюються з рядами перших та других різниць. Хоча цей метод має суб'єктивний характер, він є дуже ефективним у багатьох випадках, особливо для досвідчених аналітиків.

В аналізі часових рядів модель авторегресійної інтегрованої ковзної середньої (ARIMA) є узагальненням моделі авторегресійної ковзної середньої (ARMA). Обидві ці моделі адаптуються до даних часових рядів або для кращого розуміння даних, або для прогнозування майбутніх точок у ряді.

Моделі ARIMA застосовуються в деяких випадках, коли дані показують докази нестационарності в сенсі середнього, де початковий крок розрізнення може бути застосований один або більше разів, щоб усунути нестационарність середньої функції. Коли сезонність відображається в часовому ряді, сезонна різниця може бути застосована для усунення сезонної складової.

ARIMA поєднує функції авторегресії з функціями ковзних середніх. Наприклад, авторегресійний процес AR – це процес, у якому поточне значення базується на безпосередньо попередньому значенні, тоді як процес AR – це процес, у якому поточне значення базується на двох попередніх значеннях.

Ковзне середнє – це обчислення, яке використовується для аналізу точок даних шляхом створення серії середніх різних підмножин повного набору даних, щоб згладити вплив викидів. Завдяки цій комбінації методів моделі ARIMA можуть враховувати тренди, цикли, сезонність та інші нестатичні типи даних під час прогнозування.

Прогнозування ARIMA досягається шляхом підключення даних часових рядів для цікавої змінної. Статистичне програмне забезпечення визначить відповідну кількість затримок або величину різниці, яка буде застосована до даних, і перевірить їх на стаціонарність. Потім воно виведе результати, які часто інтерпретуються подібно до моделі множинної лінійної регресії.

Моделі векторної авторегресії (VAR) широко використовуються в дослідженнях часових рядів для вивчення динамічних зв'язків, які існують між змінними. Крім того, вони також є важливими інструментами прогнозування, які використовуються більшістю макроекономічних установ або установ, що формують політику. Розробка цих моделей є предметом багатьох поточних досліджень, які слідує за основоположним внеском Сімса (1972 р.), який встановив структуру для моделювання ендогенних змінних у багатofакторних умовах.

Структура моделей VAR дозволяє пояснити значення ендогенних змінних на основі їх минулих спостережуваних значень. Ці моделі дещо відрізняються від моделей структурної векторної авторегресії (SVAR), які дозволяють явно моделювати сучасну взаємозалежність між лівими змінними.

Нами було проаналізовано часові ряди цін акцій компанії Exxon, індекс фондового ринку S&P 500, ціна бареля сирої нафти West Texas Intermediate (США), ціна бареля сирої нафти марки Brent (ЄС), видобуток сирої нафти в США, споживання нафтопродуктів у США, запаси сирої нафти в США, споживання сирої нафти на нафтопереробних заводах США та запаси бензину в США.

Було розглянуто модель VAR. Даний підхід пройшов тест на причинність по Грейнджеру, було знайдено кореляційні відношення між рядами, проведено тест на стаціонарність, здійснено фільтрацію та нормалізацію даних та досліджено модель за різними інформаційними критеріями. Результати прогнозу на основі тестувального набору даних показують, що даний вид моделі може бути застосований для прогнозування акцій нафтопереробних компаній, цін на барелі сирої нафти різних марок, видобутку енергоресурсів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Алімпієв Є.В. Методика передмодельного аналізу у var-моделях фінансово-монетарної трансмісії. URL: <http://www.economy.nayka.com.ua/?op=1&z=3297>.
2. Андрієнко В.М., Семенов А.С., Методика статистичного аналізу економічних часових рядів. URL: http://www.visnyk-econom.uzhnu.uz.ua/archive/21_1_2018ua/3.pdf.
3. Антохонова І.В. Методи прогнозування соціально-економічних процесів, 2004. URL: <http://elbib.in.ua/metodyi-prognozirovaniya-sotsialno.html>.
4. Галушак М.П., Галушак О.Я., Кужда Т.І. Прогнозування соціально-економічних процесів: навчальний посібник для економічних спеціальностей. Тернопіль: ФОП Паляниця. 2021. 160 с.
5. Ганчук А.А., Соловійов В.М., Чабаненко Д.М. Методи прогнозування. Навч. посібник. Черкаси. Брама Україна. 2012. 140 с. URL: <http://prognoz.ck.ua/ForecastingMethodsBook-main.pdf>.
6. Гуржій А.М., Поворознюк М.І. Інформатика та інформаційні технології: підручник для учнів професійно-технічних навчальних закладів. Харків. 2007. 352 с.
7. Економетрія: навчальний посібник / Доля В.Т. Харківська національна академія міського господарства. ХНАМГ. 2010. 171 с.
8. Економічний аналіз: навчальний посібник / Грабовецький Б.Є., Вінниця: ВНТУ. 2013. 85 с. URL: <http://ктеп.kiev.ua/wp-content/uploads/2019/12/Ekonomichnij-analiz-V.-.-Graboveczkij.pdf>.
9. Єріна А.М. Статистичне моделювання та прогнозування: Навч. посібник. КНЕУ. 2001. 170 с.
10. Завадський Й.С., Осовська Т.В., Юшкевич О.О. Економічний словник. URL: https://library.nlu.edu.ua/POLN_TEXT/KNIGI/KONDOR/EKONOMIC_SL_2006.pdf.
11. Клебанова Т.С., Курзенев В.А., Наумов В.М. Прогнозування соціально-економічних процесів: навчальний посібник для студентів напряму

- підготовки 6.030502 «Економічна кібернетика». ХНЕУ ім. С. Кузнеця. 2015. 656 с.
12. Лук'яненко І.Г., Городніченко Ю.О. Сучасні економетричні методи у фінансах. Навчальний посібник. Літера ЛТД. 2002. 352 с.
13. Маринич Т.О., Назаренко Л.Д. Моделювання нестационарних процесів із структурними розривами. URL: file:///D:/User/Downloads/vestnik_KhPI_2016_6_Marynych_Modeliuvannia.pdf.
14. Маслій В.В. Авторегресійні моделі як інструмент для аналізу процесу іноземного інвестування. *Актуальні проблеми економіки*. 2012. Вип. 6. С. 311-317.
15. Метод експоненціального згладжування. URL: <https://studfile.net/preview/8988510/page:5/>.
16. Моделі ковзної середньої. URL: https://stud.com.ua/72679/ekonomika/modeli_kovznoyi_serednoyi.
17. Модель авторегресії скользящего среднего (ARIMA). URL: <https://wiki.loginom.ru/articles/arima.html>.
18. Облік, аудит і оподаткування: теорія та практика. Навчальний посібник для студентів спеціальностей 071 «Облік і оподаткування» / колектив авторів за заг. ред. проф. Давидюк Т.В. Харків. НТУ «ХПІ». 2021. 607 с. URL: <http://repository.kpi.kharkov.ua/handle/KhPI-Press/52695>.
19. Прогнозування та аналіз часових рядів. Методичні вказівки до практичних занять та самостійної роботи студентів спеціальності 051 «Економіка» освітня програма «Економічна кібернетика», «Економічна аналітика» / Укладач: Юрченко М.Є. Чернігів. ЧНТУ. 2018. 88 с.
20. Рзаєв Д.О., Шарапов О.Д., Ігнатенко В.М., Дибкова Л.М. Інформатика та комп'ютерна техніка: Навчально-методичний посібник для самостійного вивчення дисципліни. КНЕУ. 2002. 486 с.
21. Руська Р.В. Економетрика: навчальний посібник. Тернопіль. Тайп. 2012. 224 с. URL: <http://elcat.pnpu.edu.ua/docs/Економетрика.pdf>.

- 22.Словник термінів з інформатики. URL: <http://www.mktek.com.ua/wp-content/uploads/Slovník-z-informatiki.pdf>.
- 23.Тлумачний словник з інформатики / Півняк Г.Г., Бусигін Б.С., Дівізінюк М.М. Національний гірничий університет. 2010. 600 с. URL: <http://www.programmer.dp.ua/download/tlumachniy-slovník-z-informatiki.pdf>.
- 24.Трегуб А.В., Трегуб И.В. Методика построения модели арима для прогнозирования динамики временных рядов. URL: <https://cyberleninka.ru/article/n/metodika-postroeniya-modeli-arima-dlya-prognozirovaniya-dinamiki-vremennyh-ryadov/viewer>.
- 25.Фарина О.І. Динамічні моделі оцінювання стабільності фінансової системи України. Дисертація на здобуття наукового ступеня кандидата економічних наук. URL:http://ekmair.ukma.edu.ua/bitstream/handle/123456789/9075/Dysertatsiia_Farina.pdf?sequence=1&isAllowed=y.
- 26.Фінансовий менеджмент: підручник / за ред. д-ра екон. наук, проф. М. І. Крупки. Крупка М.І., Ковалюк О.М., Коваленко В.М. Львів. ЛНУ імені Івана Франка. 2019. 440 с. URL: <https://econom.lnu.edu.ua/wp-content/uploads/2021/02/Finansovyy-menedzhment-1.pdf>.
- 27.Фінансовий ринок: Навчальний посібник за заг. ред. Арутюнян С.С. / Арутюнян С.С., Добриніна Л.В. та інші – К: Гуляєва В.М. 2018. 484 с. URL: <http://dcmaup.com.ua/assets/files/finansovij-rinok.pdf>.
- 28.Фінансово-економічний словник: словник-довідник / кол. авт.: Юнін О.С., Круглова О.О., Савельєва М.О. Дніпро. Видавець Біла К.О. 2018. 164 с. URL: http://er.dduvs.in.ua/bitstream/123456789/1373/1/ФІН.-ЕКОН.%20СЛОВНИК__макет%20в%20печать.pdf.
29. Autoregressive Integrated Moving Average (ARIMA). URL: <https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp#:~:text=An%20autoregressive%20integrated%20moving%20average%2C%20or%20ARIMA%2C%20is%20a%20statistical,value%20based%20on%20past%20values.>

30. Autoregressive Integrated Moving Average. URL: <https://www.sciencedirect.com/topics/mathematics/autoregressive-integrated-moving-average>.
31. Dickey-Fuller Test. URL: <https://www.real-statistics.com/time-series-analysis/stochastic-processes/dickey-fuller-test/>.
32. Dickey-Fuller Test. URL: <https://www.sciencedirect.com/topics/economics-econometrics-and-finance/dickey-fuller-test>.
33. Duy T.A., Thoma M.A., Modeling and Forecasting Cointegrated Variables: Some Practical Experience. URL: <https://pages.uoregon.edu/mthoma/Papers/Modeling%20and%20Forecasting%20Cointegrated%20Variables.pdf>.
34. Eric Ghysels, Massimiliano Marcellino. Applied Economic Forecasting using Time Series Methods. URL: <https://file-lianxh.oss-cn-shenzhen.aliyuncs.com/Refs/Books/Ghysels-2018-Applied-Economic-Forecasting-using-Time-Series-Methods.pdf>.
35. Exponential Smoothing Forecasting. URL: <https://byjus.com/maths/exponential-smoothing/>.
36. George Athanasopoulos, Farshid Vahid. VARMA versus VAR for Macroeconomic Forecasting. URL: <https://www.jstor.org/stable/27638977>.
37. Hamilton J.D., Time Series Analysis. URL: <http://mayoral.iae-csic.org/timeseries2021/hamilton.pdf>.
38. Interpret all statistics and graphs for Augmented Dickey-Fuller Test. URL: <https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistical-modeling/time-series/how-to/augmented-dickey-fuller-test/interpret-the-results/all-statistics-and-graphs/>.
39. Interrupted time series analysis using autoregressive integrated moving average (ARIMA) models: a guide for evaluating large-scale health interventions. URL: <https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/s12874-021-01235-8>.

40. Kevin Kotze. Vector autoregression models. URL: <https://kevinkotze.github.io/ts-7-var/>.
41. Mihaela Simionescu. The use of VARMA models in forecasting macroeconomic indicators. URL: https://www.economics-sociology.eu/files/12_Simionescu_1_7.pdf.
42. Seasonal ARIMA processes. URL: <http://halweb.uc3m.es/esp/Personal/personas/amalonso/esp/TSAtema6.pdf>.
43. Vector Autoregressive Models for Multivariate Time Series. URL: <https://faculty.washington.edu/ezivot/econ584/notes/varModels.pdf>.
44. Vijay Kotu, Bala Deshpande. Data Science: Concepts and Practice, Second Edition. URL: <https://asolanki.co.in/wp-content/uploads/2019/04/Data-Science-Concepts-and-Practice-2nd-Edition-3.pdf>.