

Міністерство освіти і науки України
Сумський державний університет
Навчально-науковий інститут бізнесу, економіки та менеджменту
Кафедра економічної кібернетики

КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА

на тему «МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ КІБЕРАТАК ДЛЯ
ЕКОНОМІЧНИХ АГЕНТІВ»

Виконала студентка 2 курсу, групи ЕКм.- 91н.а.

Спеціальності 051 «Економіка»

(«Економічна кібернетика»)

Дятленко А. О.

Керівниця: к.е.н., доцентка Яровенко А.Н.

Суми – 2021 рік

РЕФЕРАТ

Кваліфікаційної магістерської роботи на тему «МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ КІБЕРАТАК ДЛЯ ЕКОНОМІЧНИХ АГЕНТІВ»

студентки Дятленко Анни Олексіївни

Актуальність теми, обраної для дослідження, визначається тим, що кожного року суб'єкти господарювання стикаються з протиправними діями зі сторони кіберзлочинності. Саме тому при прогнозуванні збільшення кількості кібератак, суб'єкт зможе збільшити захист в потрібний час.

Метою даного дослідження є розробка математичної моделі прогнозування можливості та кількості кібератак на організації чи підприємства на основі моделей авто регресії та ковзного середнього.

Предметом дослідження виступають методи і моделі для прогнозування часових рядів.

Об'єктом дослідження є кількість кіберподій спрямованих на суб'єктів господарювання.

Для досягнення поставленої мети сформовано наступні задачі дослідження:

- розглянути основні види кібератак та способи захисту;
- проаналізувати сучасний стан кібербезпеки у світі;
- описати підходи до моделювання та прогнозування часових рядів;
- сформулювати вимоги до моделі;
- обрати програмне забезпечення для реалізації моделі;
- побудувати ARIMA модель;
- перевірити її на адекватність;
- спрогнозувати можливу кількість кібератак;

Під час виконання робіт були використані такі методи дослідження: аналіз часових рядів, прогнозування на основі моделі авто регресії та ковзного середнього.

Інформаційною базою кваліфікаційної магістерської роботи є Дані з виявлених аналітиками інцидентів від постачальника послуг комп'ютерної безпеки (CSSP) для Міністерства оборони США.

Основний науковий результат роботи полягає у розробці і перевірці на адекватність моделі прогнозування кількості кібератак, що дозволяє отримати інформацію про кількість кібератак на майбутній період.

Ключові слова: кібератака, прогнозування, моделювання.

Зміст кваліфікаційної магістерської роботи викладено на 41 сторінці. Список використаних джерел із 29 найменувань, розміщений на 3 сторінках. Робота містить 3 таблиці, 31 рисунок, а також 2 додатки, розміщених на 3 сторінках.

Рік виконання кваліфікаційної роботи – 2021 рік.

Рік захисту роботи – 2021 рік.

ЗМІСТ

| | |
|---|----|
| ВСТУП | 4 |
| РОЗДІЛ 1 ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ОСНОВИ МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ ВИПАДКІВ КІБЕРАТАК..... | 6 |
| 1.1 Поняття та основні види кібератак, основні засоби захисту..... | 6 |
| 1.2 Сучасний стан кібербезпеки у світі | 8 |
| 1.3 Підходи до моделювання та прогнозування кібератак..... | 10 |
| РОЗДІЛ 2 РОЗРОБКА МОДЕЛІ ПРОГНОЗУВАННЯ МОЖЛИВИХ ВИПАДКІВ КІБРАТАК..... | 12 |
| 2.1 Моделі прогнозування часових рядів | 12 |
| 2.2 Формулювання вимог до моделі та опис вхідних змінних | 15 |
| РОЗДІЛ 3 МОДЕЛЮВАННЯ МОЖЛИВИХ ОБСЯГІВ КІБЕРАТАК | 17 |
| 3.1 Аналіз часового ряду | 17 |
| 3.2 Побудова ARIMA моделі | 22 |
| 3.3 Прогнозування кількості кібератак на основі ARIMA моделей. | 35 |
| ВИСНОВКИ..... | 38 |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ..... | 39 |
| ДОДАТОК А..... | 43 |
| ДОДАТОК Б | 44 |

ВСТУП

З розвитком сучасних інформаційних технологій та повсюдному доступу до них майже в кожному куточку світу, зростає і кількість особистої чи корпоративної інформації, що зберігається на просторах Інтернету. Незважаючи на всі переваги новітніх технологій, вони також надали безмежне поле для розвитку кіберзлочинності, яка вдосконалюється з кожним роком. В час, коли більшість людей працювали, навчалися та спілкувалися онлайн, кібербезпека для кожної окремої особи чи для великої корпорації ще не була актуальніша.

Кіберзлочини ставлять під загрозу фізичну, економічну, соціальну та політичну безпеку. Наразі, коли протиправні дії можливо скоювати навіть не виходячи з власного дому, питання прогнозування та попередження кібератак є дуже актуальним. З подальшим розвитком технологій дане питання не тільки не втратить своєї значимості, а лише зростає.

Основне завдання моделювання та подальшого прогнозування кількості кібератак для суб'єктів господарювання є попередження атак шляхом збільшення захисту в необхідний момент.

Актуальність теми, обраної для дослідження, визначається тим, що кожного року суб'єкти господарювання стикаються з протиправними діями зі сторони кіберзлочинності. Саме тому при прогнозуванні збільшення кількості кібератак, суб'єкт зможе збільшити захист в потрібний час.

Метою даного дослідження є розробка математичної моделі прогнозування можливості та кількості кібератак на організації чи підприємства на основі моделей авто регресії та ковзного середнього.

Предметом дослідження виступають методи і моделі для прогнозування часових рядів.

Об'єктом дослідження є кількість кіберподій спрямованих на суб'єктів господарювання.

Для досягнення поставленої мети сформовано наступні задачі дослідження:

- розглянути основні види кібератак та способи захисту;
- проаналізувати сучасний стан кібербезпеки у світі;
- описати підходи до моделювання та прогнозування часових рядів;
- сформулювати вимоги до моделі;
- обрати програмне забезпечення для реалізації моделі;
- побудувати ARIMA модель;
- перевірити її на адекватність;
- спрогнозувати можливу кількість кібератак;

Результатами дослідження буде сформована модель часового ряду на основі методології ARIMA , яка зможе прогнозувати кількість кібератак для запобігання нанесення збитків підприємству.

РОЗДІЛ 1 ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ОСНОВИ МОДЕЛЮВАННЯ ТА ПРОГНОЗУВАННЯ ВИПАДКІВ КІБЕРАТАК

1.1 Поняття та основні види кібератак, основні засоби захисту

Кібератака - це будь-яка спрямована спроба отримати несанкціонований доступ до комп'ютера, обчислювальної системи або комп'ютерної мережі з метою заподіяння шкоди. Кібератаки мають на меті заподіяння шкоди певному фізичному суб'єкту . юридичній особі чи навіть державі. [1]

Більшість кібератак на сьогодні спрямована на комерційні організації для отримання фінансової вигоди. Простим прикладом є викрадення конфіденційних даних – номерів кредитних карток. Інші фінансово мотивовані атаки покликані вивести з ладу самі комп'ютерні системи, при цьому кіберзлочинці блокують комп'ютери, щоб їх власники та уповноважені користувачі не могли отримати доступ до необхідних їм програм або даних. Також мотивами для правопорушень такого роду може бути помста, радикальний активізм тощо.

Уряди усього світу також беруть участь у кібератаках, при цьому багато національних урядів визнають або підозрюють у розробці та здійсненні нападів на інші країни в рамках постійних політичних, економічних та соціальних суперечок. Ці типи атак класифікуються як кібервійни.

Найпоширеніші види кібератак:

– шкідливе ПЗ. Прикладами такого програмного забезпечення - програми-вимагачі, шпигунські програми та троянські програми. Залежно від типу шкідливого коду, зловмисне програмне забезпечення може використовуватися зловмисниками для викрадення або таємного копіювання конфіденційних даних, блокування доступу до файлів, порушення роботи

системи або приведення системи в непрацездатність;

- фішинг, в якому хакери соціально розробляють повідомлення електронної пошти, щоб спонукати одержувачів відкрити їх. Одержувачів обманюють завантажити шкідливе програмне забезпечення, що міститься в електронному листі, відкривши вкладений файл або вбудоване посилання;

- DDoS, в якому хакери бомбардують сервери організації великими обсягами одночасних запитів даних, тим самим роблячи сервери не в змозі обробляти будь-які законні запити;

- введення SQL, де хакери вставляють шкідливий код на сервери за допомогою мови програмування Structured Query Language, щоб сервер розкривав конфіденційні дані;

- тунелювання системи доменних імен (DNS), складна атака, при якій зловмисники встановлюють, а потім використовують постійно доступний доступ - або тунель - до систем своїх цілей. Завантаження відбувається, коли людина відвідує веб-сайт, який, у свою чергу, заражає комп'ютер нічого не підозрюючого шкідливим програмним забезпеченням.

Очевидно, що фізичним особам та організаціям кіберзлочини краще попереджувати та якомога мінімізувати, аніж намагатися скоріш позбутися всіх негативних наслідків. Саме тому важливо не нехтувати власною інформаційною безпекою, використовувати всі доступні та необхідні засоби захисту.

Сучасний захист від шкідливих програм може блокувати небезпечні електронні листи та завантаження шкідливих програм з різних веб-сайтів. Брандмауери та проксі-сервери заблокують небезпечні та непотрібні служби. Примусова політика паролів не дозволить користувачам вибирати легкі паролі і блокувати облікові записи після певної кількості невдалих спроб. Для доступу до особливо конфіденційної корпоративної або особливо важливої інформації мають бути передбачені додаткові заходи аутентифікації.

Критично важливим для фактичного виявлення порушення є

можливість відслідковувати всю мережеву активність і аналізувати її для виявлення будь-яких зловмисних або незвичайних дій, а також застосовувати всі доступні засоби захисту та комбінувати їх.

1.2 Сучасний стан кібербезпеки у світі

Cybersecurity Ventures (провідний світовий дослідник та видавець, що висвітлює світову кіберекономіку та надійне джерело фактів та даних) - очікує, що глобальні витрати на кіберзлочинність зростатимуть на 15 відсотків на рік протягом наступних п'яти років, досягаючи 10,5 трлн. Дол. США щорічно до 2025 р., порівняно з 3 трлн. Дол. США в 2015 році. Оцінка вартості збитків базується на історичних цифрах кіберзлочинності, включаючи нещодавні темпи зростання за минулий рік, різке збільшення обсягів зловмисних груп та організованої злочинності.

Витрати на кіберзлочинність включають пошкодження та знищення даних, викрадені гроші, втрату продуктивності, крадіжку інтелектуальної власності, крадіжку персональних та фінансових даних, розкрадання, шахрайство, відновлення та видалення зламаних даних та шкоду репутації. Організовані суб'єкти кіберзлочинності об'єднують зусилля, і ймовірність розкриття та переслідування за оцінками становить 0,05% у США, згідно з Глобальним звітом Світового економічного форуму за 2020 рік.

Характер і частота кібератак, а також масштаби збитків від них також залежить від сфери діяльності і масштабів структури об'єкта нападу[10]. Масштабні наслідки несуть за собою кібератаки на великі виробництва та промислові об'єкти, особливо державного значення, що може призвести навіть до техногенної катастрофи. Але, як правило, саме малий та середній бізнес стає основною ціллю кіберзлочинців.

Великий удар направлений на такий бізнес-сектор, адже хакери знають, що його власники часто майже не знаються на кібербезпеці або застосовують

спрощені заходи, яких недостатньо. Також хакери користуються тим фактом, що більшість атак залишаються непоміченими.

Для успішної боротьби зі зловмисними намірами надзвичайно важливо, щоб компанії зробили обізнаність, запобігання та кібербезпеку частиною своєї бізнес-культури. Також незахищений малий бізнес є шляхом доступу до зламу більших підприємств. Наприкінці 2013 року, хакери отримали номери кредитних карт 40 мільйонів клієнтів пункту продажів Target, вкравши облікові дані у провайдера контактів Target – HVAC.[11].

Важливо відмітити, що для багатьох представників малого та середнього бізнесу будь-який злом та викрадення и псування даних може обернутися банкрутством, адже для відновлення цих комерційних даних необхідна буде велика кількість фінансових та часових ресурсів.

За даними PurpelSec, які постійно доповнюють актуальну світову статистичну інформацію щодо кібербезпеки:

- 43% кібератак націлені на малий бізнес;
- 70% малих підприємств не готові до кібератаки;
- 51% малих підприємств заявляють, що не виділяють жодного бюджету на кібербезпеку;
- 58% жертв шкідливого програмного забезпечення класифікуються як малий бізнес.

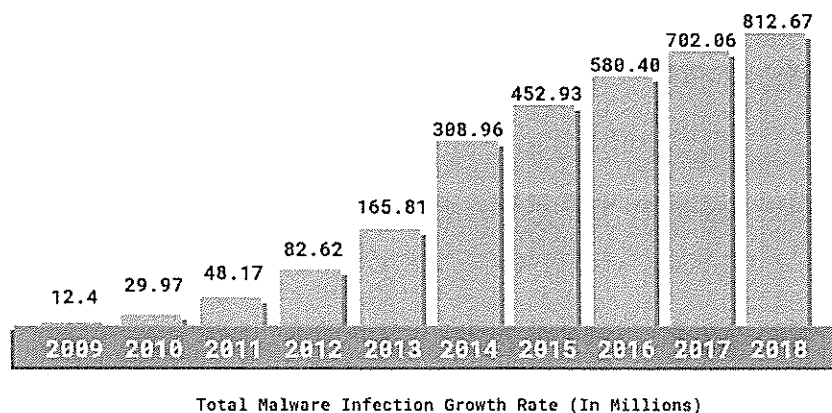


Рисунок 1.1 – Зростання загальна кількість заражень шкідливим програмним забезпеченням зростає протягом десяти років

На рисунку 1.1 зображене постійне зростання кіберзлочинів, а саме зараження шкідливим ПЗ. До типів шкідливих програм належать комп'ютерні віруси, троянські програми, шпигунські програми, програми-вимагателі, рекламні програми, хробаки, шкідливе програмне забезпечення без файлів або гібридні атаки.

Отже, можемо зробити висновки про розвиток протилежних сил. З одного боку – рівня кіберзлочинності, кіберугруповувань, вдосконалення кіберзлочинів загалом, а з іншого розуміння важливості захисту, прививання культури та систем кібербезпеки. [12]

1.3 Підходи до моделювання та прогнозування кібератак

Метод прогнозування складається з послідовності дій, які потрібно зробити для отримання моделі прогнозування. Модель прогнозування – це формалізоване співвідношення, що адекватно описує досліджуваний процес і є базою для отримання прогнозованих значень. Поєднання методу і моделі формує процес моделювання.

Велика кількість систем кібербезпеки ідентифікують загрози через вже відомі прототипи процесів та дій, що вже колись використовувались хакерами. Процес створення нових підходів для зламу інформаційної системи невпинний, а комп'ютерні віруси видозмінюються та розвиваються майже, як біологічні. Саме тому, перевірка однотипних загроз не завжди даватиме очікуваний результат і підходить лише у випадку скоєння кібератаки не досить досвідченим шахраєм. В такому випадку корисно буде прогнозувати саме кількість кібератак, для прийняття додаткових дій для забезпечення надійного захисту.

Всі сучасні методи прогнозування можна поділити на дві групи: імовірнісні і точкові. В останній час все більше дослідників та науковців звертають свою увагу на імовірнісне прогнозування кібератак. Це можна пояснити тим, що імовірнісні прогнози дозволяють отримати результат не

тільки саме майбутніх подій, а й оцінки їх здійснення. Різновидом імовірнісного прогнозування являється інтервальне прогнозування [13]. Головна ідея такого прогнозування полягає в тому, що прогнозується інтервал (із двох уже заданих інтервалів), у якому може знаходитись майбутнє значення показника на основі оцінок ймовірностей цих подій. Розмежувальна межа інтервалів задається розрахунковим способом, ґрунтуючись на статистичних характеристиках цього показника.

Наприклад, прогнозування може не починатися з уже спостереженої зловмисної події, а скоріше з імовірністю, що дана атака відбудеться, ґрунтуючись на вже відомих статистичних даних. Прикладом підходу, заснованого на безперервній моделі, є часові ряди, що представляють собою кількість атак на певну систему або мережу за певний проміжок часу. Потім часовий ряд може бути використаний для прогнозування того, чи станеться атака чи ні, або для прогнозування кількості можливих атак. Удосконалені методи можуть врахувати типи нападів та характеристики нападників та жертв щоб фахівці з кібербезпеки могли оцінити, який тип атаки буде найбільш ймовірним, хто стане нападником та хто має стати жертвою [16].

Таблиця 1.1 – Порівняльна характеристика методів прогнозування

| Метод | Переваги | Недоліки |
|---------------------------|--|--|
| Дослідження часових рядів | Великий вибір способу побудови прогнозу, простота реалізації | Складний вибір моделі |
| Теорія ігор | Результат високої точності | Кожен гравець повинен знати стратегію супротивника |

Отже, для побудови моделі прогнозування можливості кібератак було обрано дослідження часових рядів, через його простоту і способи побудови моделі.

РОЗДІЛ 2 РОЗРОБКА МОДЕЛІ ПРОГНОЗУВАННЯ МОЖЛИВИХ ВИПАДКІВ КІБРАТАК

2.1 Моделі прогнозування часових рядів

Часовий ряд - це послідовність значень досліджуваної ознаки (статистичного показника), впорядкована у хронологічному порядку. Використовують також терміни: ряд динаміки, динамічний ряд. Окремі спостереження часового ряду називають його рівнями, або елементами. [18]

Прикладами часових рядів є щомісячна кількість опадів, щорічний прибуток Google, щоденні ціни на акції тощо. На рисунку зображене квартальне виробництво пива в Австралії з 1992 до 2010 року [19]

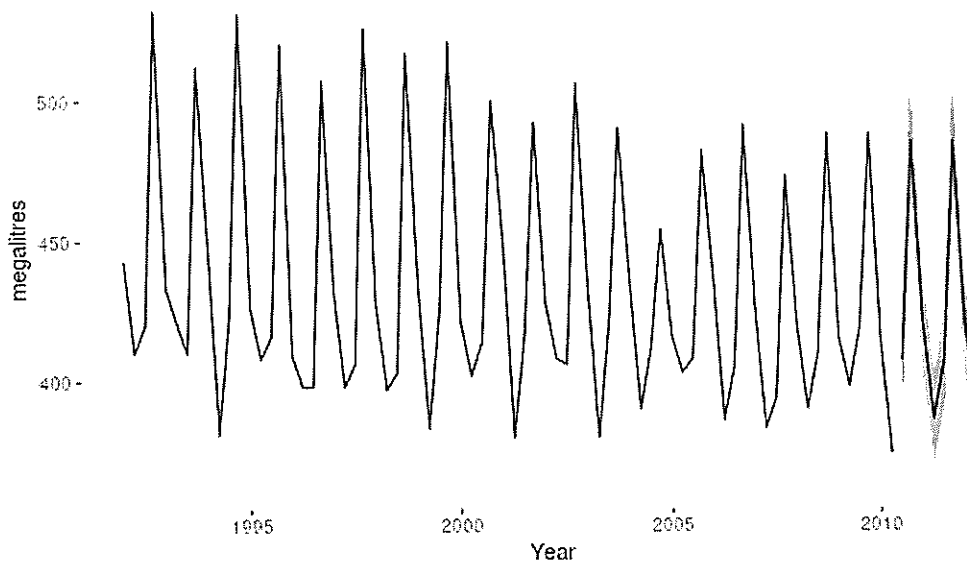


Рисунок 2.1 – Квартальне виробництво пива в Австралії з 1992 до 2010 року

Часовий ряд можна записати у короткому вигляді:

$$y_t, t = 1, 2, 3, \dots, n, \quad (2.1)$$

де t – рівновіддалені моменти спостережень (година, доба, місяць, рік);

Часовий ряд доцільно використовувати при розрахунках та будувати

математичну модель на його основі, якщо рівні ряду відповідають деяким умовам. Вони мають бути порівняні та однорідні.

Якщо часовий ряд має однакові одиниці вимірювання, однакову періодичність обліку окремих спостережень, однаковий ступінь агрегування, обчислюються за однією методикою, такий ряд можна назвати порівняним. Непорівняність часового ряду може бути через безліч причин і її неможливо усунути лише формальними методами.

Якщо у часовому ряді відсутні аномальні, нетипові спостереження та викривлення тенденції, його називають однорідним. Під аномальним рівнем розуміють таке значення часового ряду, що не відповідає потенційним можливостям даної економічної системи. Такий аномальний стрибок чи спад може чинити суттєвий вплив на основні характеристики ряду. Причинами аномальних значень можуть бути посилки першого та другого роду. Помилки першого роду, тобто технічні, зустрічаються частіше та потребують виправлення.

Для знаходження таких аномальних рядів використовують різні статистичні методи. Найпопулярнішим з них є модифікований метод Ірвіна. Він базується на порівнянні двох сусідніх значень та розрахунку значення λ , яке знаходиться за формулою 2.2:

$$\lambda_t = \frac{|y_t - y_{t-1}|}{\sigma_{\hat{y}}}; t = 2, 3, \dots, n; \quad (2.2)$$

де $\sigma_{\hat{y}}$ оцінка середньоквадратичного відхилення вибіркового ряду y_t , яка розраховується з використанням формули (2.3):

$$\sigma_{\hat{y}} = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y})^2 + (y_{t+1} - \hat{y})^2}{2}}; \hat{y} = \frac{y_{t-1} + y_{t+1}}{2}. \quad (2.3)$$

Якщо розрахункові дані λ не перевищують критичні значення, то такі

рівні часового ряду не містять аномалій. Нижче в таблиці 2.1 наведені критичні значення для рівня значущості $\alpha = 0,05$.

Таблиця 2.1 – Значення критерію Ірвіна

| | | | | | | | | | |
|------------------|---|-----|-----|-----|-----|-----|-----|-----|------|
| n | 1 | 2 | 3 | 10 | 30 | 50 | 100 | 400 | 1000 |
| λ_α | 2 | 2,8 | 2,2 | 1,5 | 1,2 | 1,1 | 1 | 0,9 | 0,8 |

Прогнозування є важливою частиною при управлінні будь-яким бізнесом. Практично все, що робить підприємство, вимагає прогнозування майбутніх потреб, щоб вони могли бути передбачені або заплановані, включаючи продажі, сировину / напівпродукти та робочу силу. Зараз добре розроблені науковцями різні прийоми дослідження часових рядів. Далі зупинимося більш детально на класичних статистично-обґрунтованих підходах – моделях авторегресії.

ARIMA (p, d, q) - методика прогнозування, яка проектує майбутні значення серії на основі власної інерції або відсталих значень. Його основне застосування в області короткострокового прогнозування і вимагає щонайменше 40 історичних даних. Дана модель найкраще працює, коли дані демонструють послідовну закономірність у часі з мінімальною кількістю відхилень. Іноді ARIMA, яку іноді називають Бокс-Дженкінсом (за оригінальними авторами), перевершує експоненціальні методи згладжування, коли дані досить довгі та кореляція між минулими спостереженнями стабільна. Якщо дані короткі або дуже мінливі, тоді якийсь метод згладжування може працювати ефективніше.

Частина AR ARIMA вказує на те, що еволюційна змінна, що цікавить, регресується на власних відсталих (тобто попередніх) значеннях:

p - порядок (кількість часових лагів) авторегресивної моделі I (інтегрована диференціація);

d - ступінь різниці (скільки разів дані віднімали минулі значення);

Частина MA вказує на те, що помилка регресії насправді є лінійною комбінацією термінів помилок, значення яких мали місце одночасно та в

різний час у минулому;

q - порядок моделі ковзного середнього.

Рівняння прогнозування ARIMA для стаціонарного часового ряду є рівнянням лінійної регресії, в якому предиктори складаються з лагів залежної змінної та / або лагів помилок прогнозу.

Часовий ряд може включати в себе тренд, сезонну складову, циклічність та випадкову компоненту. Якщо в часовому ряді спостерігається сезонність, використовуємо сезонну декомпозицію. В методі сезонної декомпозиції тренд і циклічну компоненту поєднують. Виділяємо два основних види взаємодії компонент : адитивний та мультиплікативний.

Метод експоненціального згладжування передбачує згладжування кожного рівня часового ряду за допомогою зваженої ковзної середньої. Для розрахунку такої експоненціальної середньої використовується наведена формула 2.6:

$$E_n = W \cdot Y_n + (1 - W) \cdot E_{n-1}. \quad (2.6)$$

де n – параметр згладжування;

E_n – експоненціальна середня в точці n .

2.2 Формулювання вимог до моделі та опис вхідних змінних

Модель прогнозування кількості кібератак має відповідати таким основним вимогам:

1) адекватність – модель часового ряду має відповідати реальності та враховувати найважливіші ознаки.

2) точність – побудований прогноз на основі моделі має бути наближеним до фактичних значень показника. Точність моделі ми перевіримо на останніх кроках, порівнюючи прогнозовані дані з фактичними, відомими.

3) модель має бути проста для розуміння та подальшого її використання. Чим менше менші обчислювальні ресурсні затрати (час та оперативна пам'ять) на побудову моделі, тим вона краще.

4) актуальність – модель має надавати характеристику досліджуваного об'єкта, який на чинить вагомий вплив на економіку та суспільство.

5) своєчасність – дані часового ряду мають бути актуальними.

Вхідна змінна майбутньої моделі – кількість кібер-подій, що являють собою підтвержені аналітиками прояви шкідливого ПЗ. Дані для розрахунків складаються з виявлених аналітиками інцидентів від великого постачальника послуг комп'ютерної безпеки (CSSP) для Міністерства оборони США. Набір даних включає в себе щотижневі підрахунки кібер-атак приблизно за 7 років.

На рисунку 2.2 зображена концептуальна модель прогнозування кількості кібератак, її вхідні та вихідні дані.



Рисунок 2.2 – Концептуальна модель прогнозування кількості кібератак способом ARIMA

РОЗДІЛ 3 МОДЕЛЮВАННЯ МОЖЛИВИХ ОБСЯГІВ КІБЕРАТАК

3.1 Аналіз часового ряду

Першим кроком необхідно перевірити ряд на однорідність та стаціонарність.

Найпоширенішим тестом перевірки часового ряду на однорідність є модифікований метод Ірвіна, результати якого наведені на рисунку 3.1. Для розрахунків була обрана програма MS EXCEL.

| Модифікований метод Ірвіна | | | | | | |
|----------------------------|-------|-------|---------------------|---------------------|-----------|-----------|
| week | y_i | y_t | $(y_{(i-1)}-y_t)^2$ | $(y_{(i+1)}-y_t)^2$ | \bar{y} | λ |
| 1 | 8 | - | - | - | - | - |
| 2 | 11 | 5 | 9 | 9 | 18 | 0,0189 |
| 3 | 2 | 6,5 | 20,25 | 20,25 | 40,5 | 0,0567 |
| 4 | 2 | 1 | 1 | 1 | 2 | 0,0000 |
| 5 | 0 | 2,5 | 0,25 | 0,25 | 0,5 | 0,0126 |
| 6 | 3 | 2,5 | 6,25 | 6,25 | 12,5 | 0,0189 |
| 7 | 5 | 3 | 0 | 0 | 0 | 0,0126 |
| 8 | 3 | 5 | 0 | 0 | 0 | 0,0126 |
| 9 | 5 | 3 | 0 | 0 | 0 | 0,0126 |
| 10 | 3 | 6 | 1 | 1 | 2 | 0,0126 |
| 11 | 7 | 2,5 | 0,25 | 0,25 | 0,5 | 0,0252 |
| 12 | 2 | 7,5 | 0,25 | 0,25 | 0,5 | 0,0315 |
| 13 | 8 | 3 | 1 | 1 | 2 | 0,0378 |
| 14 | 4 | 6 | 4 | 4 | 8 | 0,0252 |
| 15 | 4 | 4 | 0 | 0 | 0 | 0,0000 |
| 16 | 4 | 10,5 | 42,25 | 42,25 | 84,5 | 0,0000 |
| 17 | 17 | 6 | 4 | 4 | 8 | 0,0819 |
| 18 | 8 | 12 | 25 | 25 | 50 | 0,0567 |
| 19 | 7 | 7,5 | 0,25 | 0,25 | 0,5 | 0,0063 |
| 20 | 7 | 7 | 0 | 0 | 0 | 0,0000 |
| 21 | 7 | 6 | 1 | 1 | 2 | 0,0000 |
| 22 | 5 | 6 | 1 | 1 | 2 | 0,0126 |

Рисунок 3.1 – Фрагмент перевірки часового ряду на однорідність

Після перевірки було отримано розрахункове значення λ для кожного рівня ряду, яке порівнюємо з табличним значенням (табл. 3.1). Часовий ряд – однорідний та підходить для подальшого аналізу.

Перевірку на стаціонарність проводимо за допомогою статистичного пакету EViews. Дане програмне забезпечення спеціалізується на аналізі економічних даних, моделюванні та прогнозуванні, та має все необхідне для побудови ARIMA-моделі.

Першим кроком завантажуюмо початкові дані (рисунок 3.2). Графічна візуалізація часового ряду представлена на рисунку 3.3.

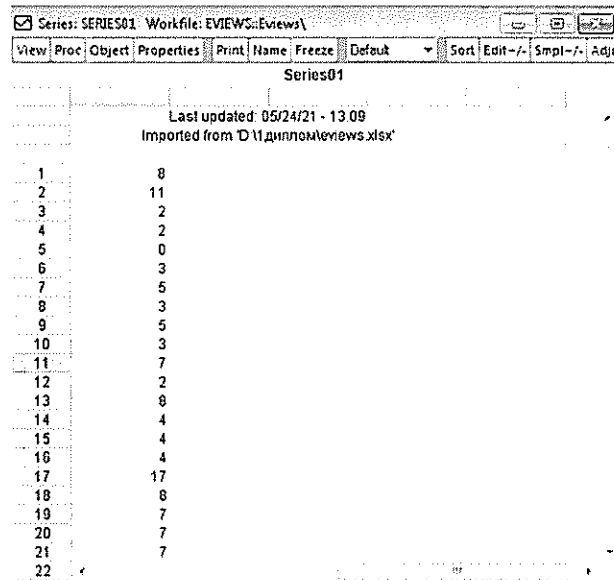


Рисунок 3.2 – Початкові данні

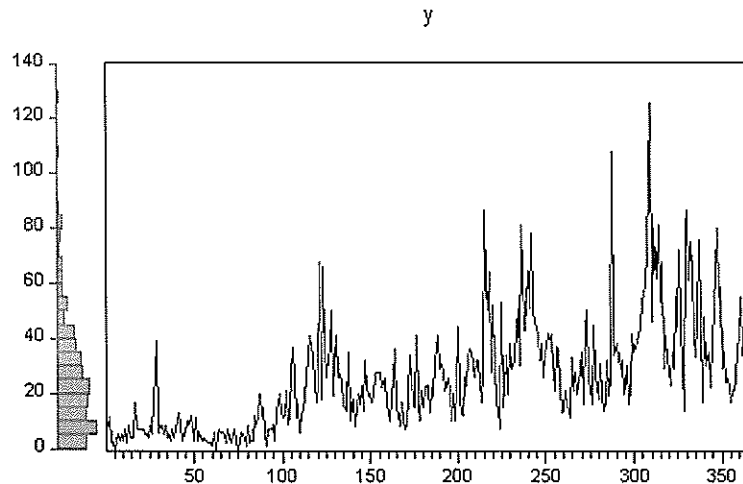


Рисунок 3.3 – Графік часового ряду кількості кібератак

Перш ніж перейти до основної частини моделювання часового ряду, необхідно проаналізувати його на стаціонарність. Під стаціонарними часовими рядами розуміють такі часові ряди, елементи яких є випадковими величинами з постійним математичним очікуванням і постійної дисперсією. Найпоширенішим тестом для перевірки ряду на стаціонарність є розширений

тест Дікі-Фулера. Даний тест проводимо за допомогою вбудованої функції EViews (Unit root test), використаємо 3 тести – з константою, з трендом та константою, та без жодної складової. Отримані дані на рисунку 3.4-3.6.

Null Hypothesis: Y has a unit root
Exogenous: None
Lag Length: 15 (Automatic - based on AIC, maxlag=16)

| | t-Statistic | Prob.* |
|--|-------------|--------|
| Augmented Dickey-Fuller test statistic | -0.596735 | 0.4584 |
| Test critical values: | | |
| 1% level | -2.571586 | |
| 5% level | -1.941732 | |
| 10% level | -1.616093 | |

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(Y)
Method: Least Squares
Date: 05/26/21 Time: 21:41
Sample (adjusted): 17 366
Included observations: 350 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|--------|
| Y(-1) | -0.014369 | 0.024080 | -0.596735 | 0.5511 |
| D(Y(-1)) | -0.661984 | 0.058745 | -11.26886 | 0.0000 |
| D(Y(-2)) | -0.426097 | 0.067033 | -6.356545 | 0.0000 |

Рисунок 3.4 – Результати розширеного тесту Дікі-Фулера для початкового ряду

Null Hypothesis: Y has a unit root
Exogenous: Constant
Lag Length: 15 (Automatic - based on AIC, maxlag=16)

| | t-Statistic | Prob.* |
|--|-------------|--------|
| Augmented Dickey-Fuller test statistic | -1.969873 | 0.3002 |
| Test critical values: | | |
| 1% level | -3.448835 | |
| 5% level | -2.869581 | |
| 10% level | -2.571122 | |

*Mackinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(Y)
Method: Least Squares
Date: 05/26/21 Time: 21:43
Sample (adjusted): 17 366
Included observations: 350 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|--------|
| Y(-1) | -0.094561 | 0.048004 | -1.969873 | 0.0497 |
| D(Y(-1)) | -0.593776 | 0.068367 | -8.685127 | 0.0000 |
| D(Y(-2)) | -0.367021 | 0.073455 | -4.996567 | 0.0000 |

Рисунок 3.5 – Результати розширеного тесту Дікі-Фулера з константою для початкового ряду

Null Hypothesis: Y has a unit root
 Exogenous: Constant, Linear Trend
 Lag Length: 2 (Automatic - based on AIC, maxlag=16)

| | t-Statistic | Prob.* |
|--|-------------|--------|
| Augmented Dickey-Fuller test statistic | -5.845028 | 0.0000 |
| Test critical values: | | |
| 1% level | -3.983541 | |
| 5% level | -3.422252 | |
| 10% level | -3.133975 | |

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(Y)
 Method: Least Squares
 Date: 05/26/21 Time: 21:44
 Sample (adjusted): 4 366
 Included observations: 363 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|--------|
| Y(-1) | -0.335818 | 0.057454 | -5.845028 | 0.0000 |
| D(Y(-1)) | -0.359711 | 0.061571 | -5.842254 | 0.0000 |
| D(Y(-2)) | -0.162484 | 0.052575 | -3.090512 | 0.0022 |

Рисунок 3.6 – Результати розширеного тесту Дікі-Фулера з константою та трендом для початкового ряду

За результатом тесту Дікі-Фулера було отримано розрахункові значення та критичні значення t-статистики МакКінона . Порівнявши ці значення отримуємо: за першим та другим тестом ряд не стаціонарний, абсолютна величина розрахункового значення t-статистики більша за абсолютні величини критичного значення при 1 %, 5 %, 10 % рівнях значущості.

Ми не відкидаємо нульову гіпотезу про те, що часовий ряд має одиничний корінь і є нестаціонарним за двома тестами Дікі-Фулера, а також через наявність тренду . Наступним кроком проводимо повторний розширений тест Дікі-Фулера для перших різниць (рис. 3.7), аби перевірити, чи стане ряд стаціонарним.

Null Hypothesis: $D(Y)$ has a unit root
 Exogenous: None
 Lag Length: 3 (Automatic - based on AIC, maxlag=4)

| | t-Statistic | Prob.* |
|--|-------------|--------|
| Augmented Dickey-Fuller test statistic | -13.28461 | 0.0000 |
| Test critical values: | | |
| 1% level | -2.571383 | |
| 5% level | -1.941704 | |
| 10% level | -1.616111 | |

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: $D(Y,2)$
 Method: Least Squares
 Date: 05/25/21 Time: 22:17
 Sample (adjusted): 6 366
 Included observations: 361 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------|-------------|------------|-------------|--------|
| $D(Y(-1))$ | -2.234338 | 0.168190 | -13.28461 | 0.0000 |
| $D(Y(-1),2)$ | 0.622476 | 0.140992 | 4.414966 | 0.0000 |

Рисунок 3.7 – Результати розширеного тесту Дікі-Фулера для перших різниць

За результатом тесту Дікі-Фулера для перших різниць абсолютна величина розрахункового значення t-статистики (-13,285) менша за абсолютні величини критичного значення при 1 %, 5 %, 10 % рівнях значущості. Отже, приймаємо нульову гіпотезу про те, що ряд стаціонарний.

Для попереднього визначення загального вигляду майбутньої ARIMA моделі і кількості лагів для кожної складової, скористаємося графіками автокореляційної і часткової автокореляційної функції показника кількості кабератак. Аналіз рисунку 3.8 показує, що сезонність в даному ряді відсутня, а значущими є перший лаг.

Sample: 1 366
Included observations: 365

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob | |
|-----------------|---------------------|----|--------|--------|--------|-------|
| | | 1 | -0.451 | -0.451 | 75.001 | 0.000 |
| | | 2 | -0.016 | -0.276 | 75.098 | 0.000 |
| | | 3 | 0.081 | -0.072 | 77.520 | 0.000 |
| | | 4 | -0.091 | -0.110 | 80.584 | 0.000 |
| | | 5 | 0.069 | -0.012 | 82.369 | 0.000 |
| | | 6 | -0.090 | -0.106 | 85.415 | 0.000 |
| | | 7 | 0.055 | -0.035 | 86.567 | 0.000 |
| | | 8 | -0.000 | -0.019 | 86.567 | 0.000 |
| | | 9 | -0.082 | -0.100 | 89.094 | 0.000 |
| | | 10 | 0.015 | -0.123 | 89.181 | 0.000 |
| | | 11 | 0.081 | 0.018 | 91.645 | 0.000 |
| | | 12 | -0.059 | -0.017 | 92.982 | 0.000 |
| | | 13 | -0.033 | -0.081 | 93.390 | 0.000 |
| | | 14 | -0.051 | -0.183 | 94.395 | 0.000 |
| | | 15 | 0.042 | -0.135 | 95.066 | 0.000 |
| | | 16 | 0.068 | 0.001 | 96.858 | 0.000 |
| | | 17 | -0.044 | 0.009 | 97.608 | 0.000 |
| | | 18 | 0.076 | 0.071 | 99.825 | 0.000 |
| | | 19 | -0.074 | -0.027 | 101.94 | 0.000 |
| | | 20 | -0.105 | -0.208 | 106.22 | 0.000 |
| | | 21 | 0.251 | 0.104 | 130.80 | 0.000 |
| | | 22 | -0.164 | -0.005 | 141.25 | 0.000 |
| | | 23 | 0.040 | -0.014 | 141.87 | 0.000 |
| | | 24 | 0.027 | 0.006 | 142.16 | 0.000 |
| | | 25 | -0.041 | 0.024 | 142.81 | 0.000 |

Рисунок 3.8 – Графік автокореляційної та часткової автокореляційної функції перших різниць часового ряду

3.2 Побудова ARIMA моделі

Першим кроком оцінюємо AR складову за допомогою звичайного методу найменших квадратів. Статистично значущими є перший та другий лаг. Оптимальними лагами для включення в модель є ті, за яких досягається мінімальне значення інформаційного критерію Акайк (AIC).

Включаючи перший лаг, отримуємо результат, вказані на рисунку 3.9. Критерій Акайк = 8,156.

Dependent Variable: D(Y)
Method: Least Squares
Date: 05/26/21 Time: 22:18
Sample (adjusted): 3 366
Included observations: 364 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|--------|
| C | 0.034964 | 0.746579 | 0.046832 | 0.9627 |
| D(Y(-1)) | -0.454526 | 0.047014 | -9.667859 | 0.0000 |

| | | | |
|--------------------|-----------|-----------------------|----------|
| R-squared | 0.205212 | Mean dependent var | 0.000000 |
| Adjusted R-squared | 0.203017 | S.D. dependent var | 15.95500 |
| S.E. of regression | 14.24365 | Akaike info criterion | 8.155979 |
| Sum squared resid | 73443.16 | Schwarz criterion | 8.177392 |
| Log likelihood | -1482.388 | Hannan-Quinn criter. | 8.164490 |
| F-statistic | 93.46750 | Durbin-Watson stat | 2.244618 |
| Prob(F-statistic) | 0.000000 | | |

Рисунок 3.9 – Результати оцінювання AR-складової при включеному першому лагу

Наступний кроком додаємо другий лаг (рис. 3.10). Критерій Акайк = 8,082. Оптимальним вважається таке значення лагу, при якому досягається найменше значення критерію АІС, а це означає, що оптимальним порядком авторегресії нашої моделі є 2.

Dependent Variable: D(Y)
Method: Least Squares
Date: 05/26/21 Time: 22:25
Sample (adjusted): 4 366
Included observations: 363 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------|-------------|------------|-------------|--------|
| C | 0.079369 | 0.719819 | 0.110263 | 0.9123 |
| D(Y(-1)) | -0.580417 | 0.050757 | -11.43524 | 0.0000 |
| D(Y(-2)) | -0.278979 | 0.050777 | -5.494248 | 0.0000 |

| | | | |
|--------------------|-----------|-----------------------|----------|
| R-squared | 0.266646 | Mean dependent var | 0.024793 |
| Adjusted R-squared | 0.262571 | S.D. dependent var | 15.97000 |
| S.E. of regression | 13.71402 | Akaike info criterion | 8.082945 |
| Sum squared resid | 67706.78 | Schwarz criterion | 8.115130 |
| Log likelihood | -1464.054 | Hannan-Quinn criter. | 8.095738 |
| F-statistic | 65.44749 | Durbin-Watson stat | 2.034181 |
| Prob(F-statistic) | 0.000000 | | |

Рисунок 3.10 - Результати оцінювання AR-складової першого порядку

Мінімальне значення критерію Акайк було досягнуто для специфікації такого виду:

$$D(Y) = m + \alpha_1 D(Y(-1)) + \alpha_2 D(Y(-2)); \quad (3.1)$$

Процедура пошуку оптимального порядку MA-складової моделі ARIMA аналогічна процедурі визначення оптимального порядку AR - складової. На цьому етапі до вже визначеної AR-складової поступово додають MA(q)-складові та розраховують значення Шварц-критерію. Модель, яка має найменше значення критерію – претендент для подальшого аналізу. Мінімальне значення критерію Шварца – 8,127 (рисунок 3.11), досягається для загальної специфікації моделі в EViews :

$$D(Y) C D(Y(-1)) D(Y(-2)) RESIDES(-1); \quad (3.2)$$

Dependent Variable: D(Y)
Method: Least Squares
Date: 05/26/21 Time: 22:38
Sample (adjusted): 5 366
Included observations: 362 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| C | 0.081843 | 0.720588 | 0.113578 | 0.9096 |
| D(Y(-1)) | -0.323934 | 0.182228 | -1.777633 | 0.0763 |
| D(Y(-2)) | -0.162160 | 0.094419 | -1.717447 | 0.0868 |
| RESID01(-1) | -0.279011 | 0.190036 | -1.468198 | 0.1429 |
| R-squared | 0.271250 | Mean dependent var | | 0.024862 |
| Adjusted R-squared | 0.265144 | S.D. dependent var | | 15.99210 |
| S.E. of regression | 13.70903 | Akaike info criterion | | 8.084975 |
| Sum squared resid | 67281.64 | Schwarz criterion | | 8.127977 |
| _log likelihood | -1459.380 | Hannan-Quinn criter. | | 8.102070 |
| F-statistic | 44.41747 | Durbin-Watson stat | | 2.008643 |
| Prob(F-statistic) | 0.000000 | | | |

Рисунок 3.11 – Результати оцінювання MA-складової першого порядку

Dependent Variable: D(Y)
 Method: Least Squares
 Date: 05/26/21 Time: 22:38
 Sample (adjusted): 5 366
 Included observations: 362 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| C | 0.081843 | 0.720588 | 0.113578 | 0.9096 |
| D(Y(-1)) | -0.323934 | 0.182228 | -1.777633 | 0.0763 |
| D(Y(-2)) | -0.162160 | 0.094419 | -1.717447 | 0.0868 |
| RESID01(-1) | -0.279011 | 0.190036 | -1.468198 | 0.1429 |
| R-squared | 0.271250 | Mean dependent var | | 0.024862 |
| Adjusted R-squared | 0.265144 | S.D. dependent var | | 15.99210 |
| S.E. of regression | 13.70903 | Akaike info criterion | | 8.084975 |
| Sum squared resid | 67281.64 | Schwarz criterion | | 8.127977 |
| _log likelihood | -1459.380 | Hannan-Quinn criter. | | 8.102070 |
| F-statistic | 44.41747 | Durbin-Watson stat | | 2.008643 |
| Prob(F-statistic) | 0.000000 | | | |

Рисунок 3.12 – Результати оцінювання MA-складової другого порядку

Оцінивши всі варіанти, найоптимальнішим буде AR-складова 2-го порядку, а MA-складова першого порядку. Константу виключаємо з моделі, адже вона не значуща. На рисунку 3.13 відображені результати оцінювання ARIMA-моделі, де $p = 2$, $d = 1$, $q = 1$.

Dependent Variable: D(Y)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 05/26/21 Time: 22:58
 Sample: 2 366
 Included observations: 365
 Convergence achieved after 35 iterations
 Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| AR(1) | 0.282655 | 0.048309 | 5.851014 | 0.0000 |
| AR(2) | 0.198866 | 0.049480 | 4.019089 | 0.0001 |
| MA(1) | -0.922956 | 0.027651 | -33.37874 | 0.0000 |
| SIGMASQ | 177.5636 | 5.979614 | 29.69483 | 0.0000 |
| R-squared | 0.298699 | Mean dependent var | | 0.008219 |
| Adjusted R-squared | 0.292871 | S.D. dependent var | | 15.93384 |
| S.E. of regression | 13.39892 | Akaike info criterion | | 8.041858 |
| Sum squared resid | 64810.72 | Schwarz criterion | | 8.084596 |
| Log likelihood | -1463.639 | Hannan-Quinn criter. | | 8.058843 |
| Durbin-Watson stat | 2.047142 | | | |
| Inverted AR Roots | .61 | -.33 | | |
| Inverted MA Roots | .92 | | | |

Рисунок 3.12 – Результати оцінювання ARIMA моделі (2,1,1)

Основна специфікація ARIMA моделі (2,1,1):

$$D(Y) AR(Y(-1)) AR(Y(-2)) MA(-1); \quad (3.3)$$

Наступним кроком необхідно перевірити якість побудованої моделі. Для того, щоб модель вважалась адекватною, необхідно, щоб залишки оціненого фінального рівняння були білим шумом. Для перевірки скористаємося корелограмами та тестом Дікі-Фулера (рис 3.13-3.14).

Як можна побачити з графіку автокореляції (рис 3.13) всі її коефіцієнти статистично не значущі, отже, попередньо можна зробити висновок про стаціонарність залишків.

Перевірка залишків на білий шум на основі тесту Дікі-Фулера (рис. 3.14) показує відсутність одиничного кореня, тобто залишки стаціонарні. Таким чином, залишки оціненої моделі можна вважати білим шумом, а модель адекватною.

Після перевірки залишків на стаціонарність, побудуємо та проаналізуємо гістограму та показники нормальності розподілу залишків, отримаємо такі основні значення:

- Probability менше за 0,05, у свою чергу це свідчить про те, що гіпотеза про нормальність розподілу відкидається.

- Коефіцієнт асиметрії (Skewness) більш нуля, відслідковується правостороння асиметрія, тобто позитивні залишки переважають над негативними. Звідси можемо зробити висновок, що в кількості кібератак частіше спостерігались різкі підйоми, ніж різкі аналогічні спади.

- Коефіцієнт ексцесу (Kurtosis) більше 3 – розподіл гостроверхий. Це означає, що в даному розподілі є виражене ядро щільності розподілу, всередині якого діапазон коливань залишків не значний і розсіяне «гало», де розкид коливань величини залишків досить значний. З точки зору передбачення кількості кібератак, таких характер розподілу означає, що 99% прогнозу буде лише при дуже великому рині надійності.

Date: 05/25/21 Time: 23.48

Sample (adjusted): 2 366

Included observations: 365 after adjustments

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 -0.029 | -0.029 | 0.3193 | 0.572 |
| | | 2 -0.040 | -0.040 | 0.8974 | 0.638 |
| | | 3 0.123 | 0.121 | 6.4663 | 0.091 |
| | | 4 -0.003 | 0.002 | 6.4705 | 0.167 |
| | | 5 0.042 | 0.052 | 7.1314 | 0.211 |
| | | 6 -0.045 | -0.058 | 7.8841 | 0.247 |
| | | 7 0.034 | 0.037 | 8.3274 | 0.305 |
| | | 8 -0.017 | -0.032 | 8.4325 | 0.392 |
| | | 9 -0.103 | -0.090 | 12.432 | 0.190 |
| | | 10 -0.001 | -0.019 | 12.432 | 0.257 |
| | | 11 0.056 | 0.061 | 13.633 | 0.254 |
| | | 12 -0.078 | -0.060 | 15.946 | 0.194 |
| | | 13 -0.103 | -0.097 | 19.976 | 0.096 |
| | | 14 -0.073 | -0.096 | 22.005 | 0.079 |
| | | 15 0.051 | 0.053 | 22.985 | 0.084 |
| | | 16 0.087 | 0.114 | 25.871 | 0.056 |
| | | 17 0.007 | 0.047 | 25.890 | 0.076 |
| | | 18 0.053 | 0.039 | 26.978 | 0.079 |
| | | 19 -0.088 | -0.112 | 29.976 | 0.052 |
| | | 20 -0.038 | -0.051 | 30.522 | 0.062 |
| | | 21 0.220 | 0.202 | 49.351 | 0.000 |
| | | 22 -0.061 | -0.049 | 50.812 | 0.000 |
| | | 23 0.013 | 0.016 | 50.881 | 0.001 |
| | | 24 0.042 | 0.012 | 51.566 | 0.001 |
| | | 25 -0.020 | 0.002 | 51.723 | 0.001 |
| | | 26 0.008 | -0.031 | 51.748 | 0.002 |
| | | 27 -0.009 | -0.010 | 51.783 | 0.003 |
| | | 28 0.049 | 0.030 | 52.738 | 0.003 |
| | | 29 -0.090 | -0.057 | 55.937 | 0.002 |
| | | 30 -0.078 | -0.012 | 58.385 | 0.001 |
| | | 31 0.074 | 0.053 | 60.595 | 0.001 |
| | | 32 -0.023 | -0.067 | 60.802 | 0.002 |
| | | 33 -0.039 | -0.029 | 61.400 | 0.002 |
| | | 34 -0.030 | -0.006 | 61.753 | 0.002 |
| | | 35 -0.060 | -0.021 | 63.227 | 0.002 |
| | | 36 0.120 | 0.127 | 69.042 | 0.001 |

Рисунок 3.13 – Корелограма залишків моделі

Null Hypothesis: RESID02 has a unit root
 Exogenous: None
 Lag Length: 2 (Automatic - based on AIC, maxlag=4)

| | t-Statistic | Prob.* |
|--|-------------|--------|
| Augmented Dickey-Fuller test statistic | -9.937325 | 0.0000 |
| Test critical values: | | |
| 1% level | -2.571366 | |
| 5% level | -1.941701 | |
| 10% level | -1.616113 | |

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(RESID02)
 Method: Least Squares
 Date: 05/25/21 Time: 23:54
 Sample (adjusted): 5 366
 Included observations: 362 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|-----------|
| RESID02(-1) | -0.935764 | 0.094167 | -9.937325 | 0.0000 |
| D(RESID02(-1)) | -0.088246 | 0.075475 | -1.169213 | 0.2431 |
| D(RESID02(-2)) | -0.123577 | 0.052650 | -2.347138 | 0.0195 |
| R-squared | 0.520177 | Mean dependent var | | -0.053813 |
| Adjusted R-squared | 0.517504 | S.D. dependent var | | 19.16341 |
| S.E. of regression | 13.31128 | Akaike info criterion | | 8.023353 |
| Sum squared resid | 63611.27 | Schwarz criterion | | 8.055605 |
| Log likelihood | -1449.227 | Hannan-Quinn criter. | | 8.036174 |
| Durbin-Watson stat | 1.991455 | | | |

Рисунок 3.14 – Результати перевірки залишків моделі на білий шум

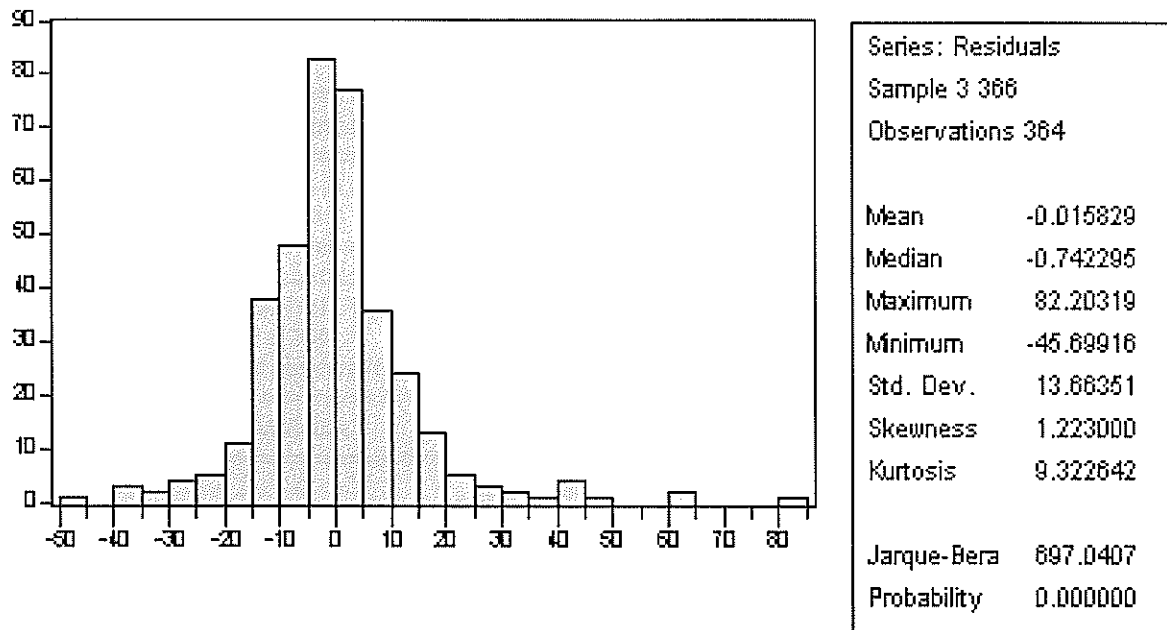


Рисунок 3.15 – Результати тестування на нормальній розподіл залишків

Перевіряємо нульову гіпотезу про те, що не існує зв'язку між дисперсією залишків вихідного рівняння і незалежними змінними, тобто залишки гомоскедастичні. Однією з важливих передумов МНК є передумова про гомоскедастичність залишків, тобто про сталість дисперсії залишкових членів рівняння регресії. Виконання цієї передумови (поряд з іншими) забезпечує хороші якості обраних оцінок моделі методом найменших квадратів.

В нашому випадку гіпотеза про гомоскедастичність залишків приймається, адже показник $Obs \cdot R^2$ менше прийнятого рівня значущості. Такі дані отримали за допомогою тесту Уайта в пакеті EViews (рис. 3.16).

Тест Дікі-Фулера (рис.3.18) для других різниць часового ряду зазначає, що ряд стаціонарний, адже розрахункове значення менше, ніж табличне. Отже, наступним кроком побудуємо модель з такими ж складовими AR та MA для других різниць, адже є вірогідність, що така модель краще описуватиме у.

Heteroskedasticity Test: White

| | | | |
|---------------------|----------|----------------------|--------|
| F-statistic | 253.3695 | Prob. F(10,353) | 0.0000 |
| Obs*R-squared | 319.4882 | Prob. Chi-Square(10) | 0.0000 |
| Scaled explained SS | 1299.545 | Prob. Chi-Square(10) | 0.0000 |

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 05/27/21 Time: 15:42

Sample: 3 366

Included observations: 364

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|-------------------|-------------|------------|-------------|--------|
| C | 75.14853 | 10.51813 | 7.144664 | 0.0000 |
| GRADF_01^2 | 25.26117 | 2.319053 | 10.89288 | 0.0000 |
| GRADF_01*GRADF_02 | 24.68379 | 3.188909 | 7.740511 | 0.0000 |
| GRADF_01*GRADF_03 | 56805.09 | 5837.969 | 9.730283 | 0.0000 |
| GRADF_01*GRADF_04 | 280386.7 | 28653.41 | 9.785459 | 0.0000 |
| GRADF_02^2 | 6.142150 | 1.909769 | 3.216173 | 0.0014 |
| GRADF_02*GRADF_03 | 24492.15 | 4479.353 | 5.467788 | 0.0000 |
| GRADF_02*GRADF_04 | 120701.0 | 21502.48 | 5.613354 | 0.0000 |
| GRADF_03^2 | 34489202 | 3409470. | 10.11571 | 0.0000 |
| GRADF_03*GRADF_04 | 3.40E+08 | 33590833 | 10.11122 | 0.0000 |

Рисунок 3.16 – Результати тесту Уайта на гетероскедатичність залишків моделі

Null Hypothesis: D(Y,2) has a unit root

Exogenous: None

Lag Length: 16 (Automatic - based on AIC, maxlag=16)

| | t-Statistic | Prob.* |
|--|-------------|--------|
| Augmented Dickey-Fuller test statistic | -10.05182 | 0.0000 |
| Test critical values: | | |
| 1% level | -2.571643 | |
| 5% level | -1.941740 | |
| 10% level | -1.616087 | |

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(Y,3)

Method: Least Squares

Date: 05/26/21 Time: 23:30

Sample (adjusted): 20 366

Included observations: 347 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|------------|-------------|------------|-------------|--------|
| D(Y(-1),2) | -21.46783 | 2.135715 | -10.05182 | 0.0000 |
| D(Y(-1),3) | 18.92107 | 2.111808 | 8.959653 | 0.0000 |
| D(Y(-2),3) | 17.15005 | 2.056997 | 8.337422 | 0.0000 |
| D(Y(-3),3) | 15.42138 | 1.974985 | 7.808350 | 0.0000 |
| D(Y(-4),3) | 13.73784 | 1.871882 | 7.339049 | 0.0000 |
| D(Y(-5),3) | 12.14347 | 1.748804 | 6.943874 | 0.0000 |
| D(Y(-6),3) | 10.57879 | 1.607293 | 6.581744 | 0.0000 |

Рисунок 3.17 - Результати розширеного тесту Дікі-Фулера для других різниць

Date: 05/26/21 Time: 23:36
 Sample: 1 366
 Included observations: 364

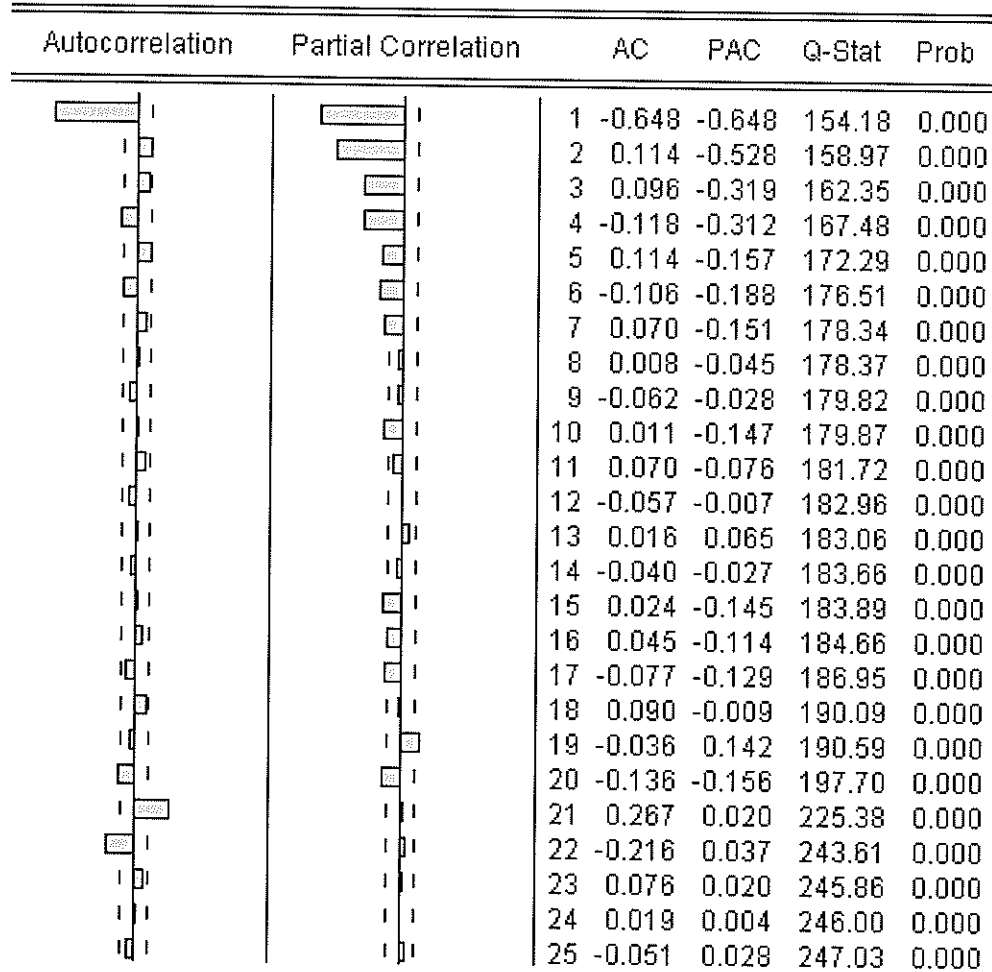


Рисунок 3.19 - Графік автокореляційної та часткової автокореляційної функції других різниць часового ряду

Основна специфікація ARIMA моделі (2,2,1):

$$D(Y, 2) AR(Y(-1)) AR(Y(-2)) MA(-1); \quad (3.3)$$

Якщо звернути увагу на значення R^2 , очевидно, що його значення більше для моделі ARIMA (2,2,1). Така модель краще описує фактичні дані, але MA-складова не значуща, тому побудуємо модель, виключаючи MA-складову, ARIMA (2,2,0).

Dependent Variable: D(Y,2)
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 05/26/21 Time: 23:39
 Sample: 3 366
 Included observations: 364
 Failure to improve objective (non-zero gradients) after 24 iterations
 Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|-----------|
| AR(1) | -0.577150 | 0.040391 | -14.28904 | 0.0000 |
| AR(2) | -0.275710 | 0.042737 | -6.451310 | 0.0000 |
| MA(1) | -0.999999 | 42.78601 | -0.023372 | 0.9814 |
| SIGMASQ | 186.1789 | 208.5105 | 0.892900 | 0.3725 |
| R-squared | 0.746784 | Mean dependent var | | -0.076923 |
| Adjusted R-squared | 0.744674 | S.D. dependent var | | 27.15296 |
| S.E. of regression | 13.72033 | Akaike info criterion | | 8.107214 |
| Sum squared resid | 67769.13 | Schwarz criterion | | 8.150040 |
| Log likelihood | -1471.513 | Hannan-Quinn criter. | | 8.124235 |
| Durbin-Watson stat | 2.038820 | | | |
| Inverted AR Roots | -.29+.44i | -.29-.44i | | |
| Inverted MA Roots | 1.00 | | | |

Рисунок 3.20 – Результати оцінювання ARIMA моделі (2,2,1)

Dependent Variable: Y
 Method: ARMA Maximum Likelihood (OPG - BHHH)
 Date: 05/28/21 Time: 04:44
 Sample: 1 366
 Included observations: 366
 Convergence achieved after 23 iterations
 Coefficient covariance computed using outer product of gradients

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|----------|
| D | -0.453785 | 0.034366 | -13.20449 | 0.0000 |
| AR(1) | 0.795524 | 0.000175 | 4558.843 | 0.0000 |
| AR(2) | 0.204451 | 0.001133 | 180.4002 | 0.0000 |
| SIGMASQ | 175.8300 | 7.153254 | 24.58042 | 0.0000 |
| R-squared | 0.552555 | Mean dependent var | | 25.41530 |
| Adjusted R-squared | 0.548846 | S.D. dependent var | | 19.85046 |
| S.E. of regression | 13.33315 | Akaike info criterion | | 8.036655 |
| Sum squared resid | 64353.78 | Schwarz criterion | | 8.079307 |
| Log likelihood | -1466.708 | Hannan-Quinn criter. | | 8.053604 |
| Durbin-Watson stat | 2.004225 | | | |
| Inverted AR Roots | 1.00 | -.20 | | |

Рисунок 3.21 – Результати оцінювання ARIMA моделі (2,2,0)

Основна специфікація ARIMA моделі (2,2,1):

$$D(Y, 2) AR(Y(-1)) AR(Y(-2)); \quad (3.4)$$

Перевірка залишків на білий шум на основі тесту Дікі-Фулера (рис. 3.22) показує відсутність одиничного кореня, тобто залишки стаціонарні. Таким чином, залишки оціненої моделі можна вважати білим шумом, а модель адекватною.

Date: 05/28/21 Time: 04:57

Sample: 1 366

Included observations: 347

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob |
|-----------------|---------------------|-----------|--------|--------|-------|
| | | 1 -0.026 | -0.026 | 0.2298 | 0.632 |
| | | 2 -0.054 | -0.055 | 1.2648 | 0.531 |
| | | 3 0.094 | 0.091 | 4.3623 | 0.225 |
| | | 4 -0.034 | -0.033 | 4.7757 | 0.311 |
| | | 5 0.040 | 0.049 | 5.3397 | 0.376 |
| | | 6 -0.060 | -0.072 | 6.6244 | 0.357 |
| | | 7 0.027 | 0.037 | 6.8761 | 0.442 |
| | | 8 -0.028 | -0.045 | 7.1493 | 0.521 |
| | | 9 -0.112 | -0.096 | 11.665 | 0.233 |
| | | 10 -0.007 | -0.028 | 11.682 | 0.307 |
| | | 11 0.027 | 0.030 | 11.954 | 0.367 |
| | | 12 -0.122 | -0.117 | 17.367 | 0.136 |
| | | 13 -0.129 | -0.132 | 23.365 | 0.038 |
| | | 14 -0.113 | -0.145 | 28.046 | 0.014 |
| | | 15 0.046 | 0.038 | 28.811 | 0.017 |
| | | 16 0.073 | 0.079 | 30.772 | 0.014 |
| | | 17 0.024 | 0.058 | 30.978 | 0.020 |
| | | 18 0.057 | 0.036 | 32.174 | 0.021 |
| | | 19 -0.070 | -0.081 | 34.009 | 0.018 |
| | | 20 -0.018 | -0.041 | 34.131 | 0.025 |
| | | 21 0.255 | 0.236 | 58.225 | 0.000 |
| | | 22 -0.052 | -0.059 | 59.235 | 0.000 |
| | | 23 0.018 | 0.020 | 59.354 | 0.000 |

Рисунок 3.22 – Корелограма залишків моделі

Null Hypothesis: RESID02 has a unit root
 Exogenous: None
 Lag Length: 16 (Automatic - based on AIC, maxlag=16)

| | t-Statistic | Prob.* |
|--|-------------|--------|
| Augmented Dickey-Fuller test statistic | -4.213897 | 0.0000 |
| Test critical values: | | |
| 1% level | -2.571988 | |
| 5% level | -1.941787 | |
| 10% level | -1.616056 | |

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
 Dependent Variable: D(RESID02)
 Method: Least Squares
 Date: 05/28/21 Time: 04:57
 Sample (adjusted): 18 347
 Included observations: 330 after adjustments

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|----------------|-------------|------------|-------------|--------|
| RESID02(-1) | -1.004127 | 0.238289 | -4.213897 | 0.0000 |
| D(RESID02(-1)) | -0.019700 | 0.230141 | -0.085600 | 0.9318 |
| D(RESID02(-2)) | -0.056289 | 0.219779 | -0.256118 | 0.7980 |
| D(RESID02(-3)) | 0.083719 | 0.207650 | 0.403174 | 0.6871 |
| D(RESID02(-4)) | 0.057860 | 0.197324 | 0.293223 | 0.7695 |

Рисунок 3.23 – Результати перевірки залишків моделі на білий шум

Після перевірки залишків на стаціонарність, побудували гістограму та показники нормальності розподілу залишків. Показники свідчать, що аналогічно до моделі (2,1,1), про нормальність розподілу відкидається, відслідковується правостороння асиметрія, розподіл гостроверхий.

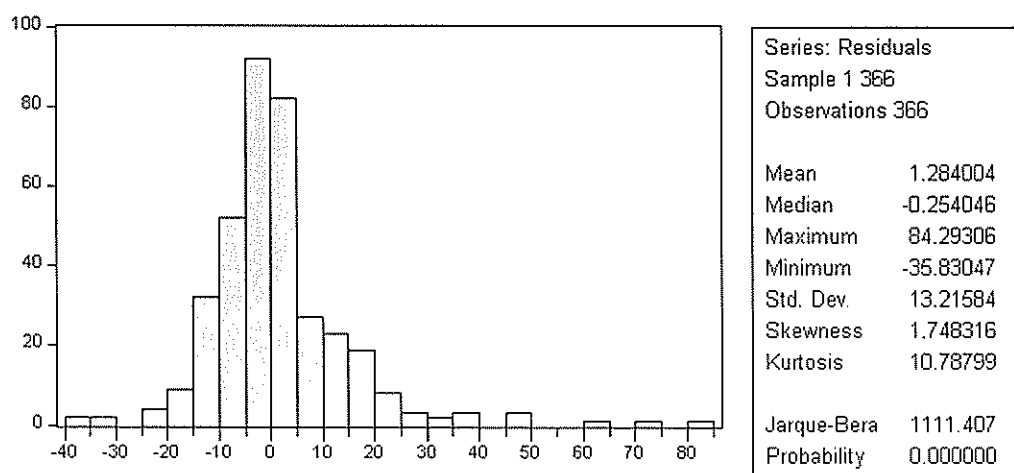


Рисунок 3.23– Результати тестування на нормальній розподіл залишків

В випадку моделі (2,2,0) гіпотеза про гомоскедастичність залишків приймається, адже показник $\text{Obs} \cdot R^2$ менше прийнятого рівня значущості. Такі дані отримали за допомогою тесту Уайта в пакеті EViews (рис. 3.24).

Heteroskedasticity Test: White

| | | | |
|---------------------|----------|----------------------|--------|
| F-statistic | 3357.372 | Prob. F(10,355) | 0.0000 |
| Obs*R-squared | 362.1705 | Prob. Chi-Square(10) | 0.0000 |
| Scaled explained SS | 1826.424 | Prob. Chi-Square(10) | 0.0000 |

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 05/28/21 Time: 05:26

Sample: 1 366

Included observations: 366

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|-------------------|-------------|--------------------|-------------|--------|
| C | 20.69130 | 3.428748 | 6.034652 | 0.0000 |
| GRADF_01^2 | -0.340602 | 0.420796 | -0.809421 | 0.4188 |
| GRADF_01*GRADF_02 | 1.009815 | 0.209933 | 4.810174 | 0.0000 |
| GRADF_01*GRADF_03 | 2.001881 | 0.423154 | 4.730881 | 0.0000 |
| GRADF_01*GRADF_04 | -422.6285 | 141.6083 | -2.984490 | 0.0030 |
| GRADF_02^2 | -0.001300 | 0.000585 | -2.221174 | 0.0270 |
| GRADF_02*GRADF_03 | -0.004990 | 0.007380 | -0.676181 | 0.4994 |
| GRADF_02*GRADF_04 | 691.0429 | 40.27479 | 17.15820 | 0.0000 |
| GRADF_03^2 | 0.011156 | 0.012434 | 0.897215 | 0.3702 |
| GRADF_03*GRADF_04 | 1357.436 | 81.36404 | 16.68349 | 0.0000 |
| GRADF_04^2 | -47698.33 | 23660.49 | -2.015948 | 0.0446 |
| R-squared | 0.989537 | Mean dependent var | 175.8300 | |

Рисунок 3.24 – Результати тесту Уайта на гетероскедастичність залишків моделі

3.3 Прогнозування кількості кібератак на основі ARIMA моделей

В цьому розділі побудуємо прогнози на основі двох моделей різного порядку різниць, які були побудовані та проаналізовані в попередньому розділі. Для початкової виборки обираємо 95% від початкових даних.

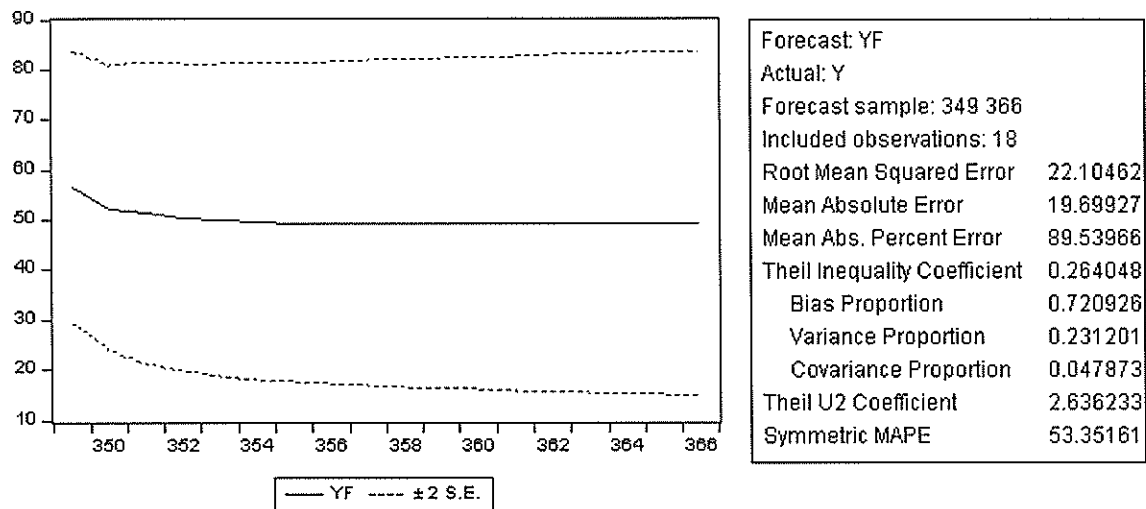


Рисунок 3.25 – Графічне зображення прогнозу моделі (2,1,1)

| | |
|-----|----------|
| 346 | 50.00000 |
| 347 | 80.00000 |
| 348 | 58.03503 |
| 349 | 57.48124 |
| 350 | 53.32073 |
| 351 | 52.07775 |
| 352 | 50.97732 |
| 353 | 50.44845 |
| 354 | 50.10247 |
| 355 | 49.91099 |
| 356 | 49.79530 |
| 357 | 49.72860 |
| 358 | 49.68918 |
| 359 | 49.66619 |
| 360 | 49.65268 |
| 361 | 49.64478 |
| 362 | 49.64015 |
| 363 | 49.63743 |
| 364 | 49.63584 |
| 365 | 49.63491 |
| 366 | 49.63436 |

Рисунок 3.26 - Прогнозні значення моделі (2,1,1)

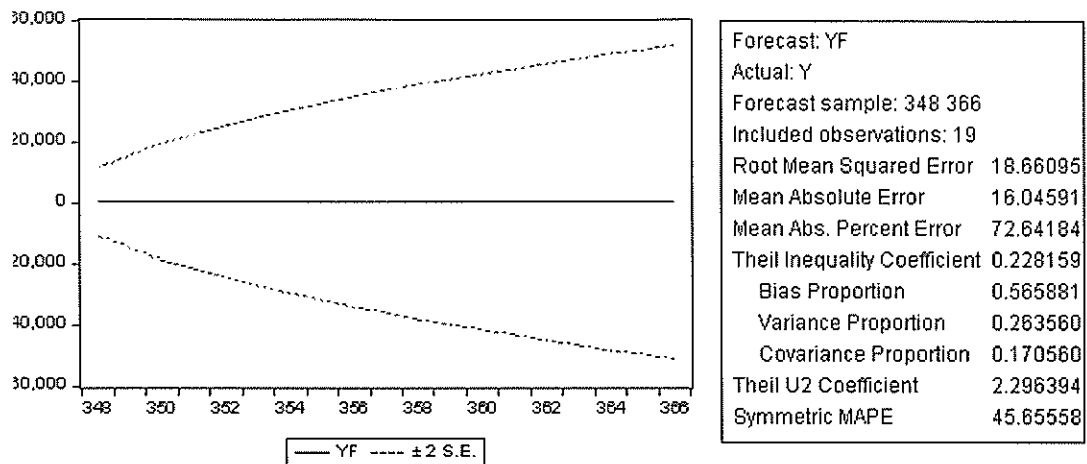


Рисунок 3.28 – Графічне зображення прогнозу моделі (2,2,0)

| | |
|-----|----------|
| 346 | 50.00000 |
| 347 | 80.00000 |
| 348 | 54.50287 |
| 349 | 54.17475 |
| 350 | 51.26023 |
| 351 | 49.90900 |
| 352 | 48.76009 |
| 353 | 47.87956 |
| 354 | 47.14460 |
| 355 | 46.51880 |
| 356 | 45.97181 |
| 357 | 45.48528 |
| 358 | 45.04621 |
| 359 | 44.64541 |
| 360 | 44.27609 |
| 361 | 43.93312 |
| 362 | 43.61256 |
| 363 | 43.31130 |
| 364 | 43.02685 |
| 365 | 42.75719 |
| 366 | 42.50066 |

Рисунок 3.26 - Прогнозні значення моделі (2,1,1)

Результати прогноза двох моделей дуже схожі, перевіряємо їх з фактичними даними, тобто з 5% виборки

Можна зробити висновки, що отримана нами ARIMA модель не

досить точно прогнозує значення в межах одного періоду. Вона не є придатною для довгострокового прогнозування, оскільки не враховує значних коливань кількості кібератак, які можуть відбуватися без впливу очевидних чинників, проте її можна успішно застосовувати для короткострокових прогнозів

ВИСНОВКИ

Дане дослідження базувалось на побудові ARIMA моделі для прогнозування кількості кібератак для суб'єктів господарювання. Під час виконання дипломної роботи були виконані такі основні завдання:

- розглянуті основні види кібератак та засоби захисту від них;
- проаналізовані методологія побудови ARIMA моделей часових рядів;
- визначені основні вимоги до моделі;
- проведений аналіз часового ряду;
- побудовані та проаналізовані ARIMA моделі різних порядків різниць та обрана найоптимальніша;

Для побудови моделі використовувалась статистична програма EViews, адже містить весь необхідний інструментарій.

Для створення прогнозної моделі було обрано ARIMA моделі порядку (2,1,1) та (2, 2, 0). Для підвищення точності та якості прогнозу було додатково було обрано методи згладження для обробки початкових даних.

Другий розділ дипломної роботи містить детальний аналіз процесу побудови моделей. Зокрема, в даному розділі проведений математичний опис моделей, використаних при моделюванні прогнозів. Другий розділ було завершено описом підготовкою часового ряду для побудови моделей.

В третьому розділі роботи була описана реалізація моделей прогнозування кількості кібератак на економічних агентів. Після побудови моделей була проаналізована їх якість на даних, що не брали участь у прогнозуванні. Візуальний огляд показав, що прогнозні значення досить не досить точно та якісно описують тенденцію зміни кількості кібератак. Отже, необхідно у подальших роботах розглядати додаткові методи прогнозування часових рядів.

24.05.2024

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Віруси і шкідливі програми / TREND MICRO 2017. URL: <http://docs.trendmicro.com/ru-ru/smb/worry-free-business-security-90-spl-agent-help/about/understanding-threat/viruses-and-malware.aspx>
2. Козьменко О. В. Економко-математичні методи та моделі (економетрика): навчальний посібник / О. В. Козьменко, О. В. Кузьменко. – Суми : Університетська книга, 2014. – 406 с.
3. Присенко Г. В. Прогнозування соціально-економічних процесів: навчальний посібник / Г. В. Приенко, Є. І Равікович — К.:КНЕУ, 2005. — 378 с.
4. Экономико-математические методы и прикладные модели: Учеб. пособие для вузов / Под ред. В. В. Федосеева. — М.: ЮНИТИ, 2000.
5. Кобелев Н. Б. Практика применения экономико-математических методов и моделей: Учебно-практическое пособие. — М.: ЗАО Финстатинформ, 2000.
6. Чумаченко Д. І. Математичні моделі та методи прогнозування епідемічних процесів: монографія / Д. І. Чумаченко, Т. О. Чумаченко. – Харків: ТОВ "Планета-Прінт", 2020. – 180 с
7. Крюков Ю. А. ARIMA – модель прогнозирования значений трафика / Ю. А. Крюков, Д. В. Чернягин. // ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ. – 2011. – №2. – С. 41–49.
8. Про рішення Ради національної безпеки і оборони України від 27 січня 2016 р. "Про Стратегію кібербезпеки України": Указ Президента України. Урядовий кур'єр, № 52, 2016.
9. Функції захисту персональних даних покладено на уповноваженого. [Електронний ресурс]. Режим доступу: <http://www.ombudsman.gov.ua/ua/page/zpd/>.
10. Шинкаренко А. Ю. Кібербезпека як один з механізмів забезпечення

стабільного розвитку економіки в Україні [Електронний ресурс] / А. Ю. Шинкаренко, О. В. Ставицький. – 2019. – Режим доступу до ресурсу: 102862-Текст статті-217650-1-10-20170530.pdf

11. Дубов Д. В. Стратегічні аспекти кібербезпеки України / Дмитро Володимирович Дубов. // Стратегічні пріоритети. – 2013. – №4. – С. 125–126.

12. 2021 Cyber Security Statistics The Ultimate List Of Stats, Data & Trends [Електронний ресурс] // purplesec.us. – 2021. – Режим доступу до ресурсу: <https://purplesec.us/resources/cyber-security-statistics/>

13. A Report on the Current State of Cybersecurity [Електронний ресурс] // theseoviews.com. – 2020. – Режим доступу до ресурсу: <https://theseoviews.com/a-report-on-the-current-state-of-cybersecurity/>.

14. Krakovsky Y. M Applied aspects of application of interval forecasting of dynamic indicators in system analysis / Y. M. Krakovsky , A. N. Luzgin –Modern technology. System analysis. Modeling, 2017, no. 2/

16. Husak M. Survey of Attack Projection, Prediction, and Forecasting in Cyber Security URL: https://www.researchgate.net/publication/327449459_Survey_of_Attack_Projection_Prediction_and_Forecasting_in_Cyber_Security

17. Методы прогнозирования. [URL: <http://statsoft.ru/solutions/tasks/forecast/>

18 Прогнозування та аналіз часових рядів. Методичні вказівки до практичних занять та самостійної роботи студентів спеціальності 051«Економіка» освітня програма «Економічна кібернетика», «Економічнааналітика» / Укл.: Юрченко М. Є. – Чернігів: ЧНТУ, 2018. – 88 с.

19. Лук'яненко І. Г. Економетрика. Теорія та практика / І. Г. Лук'яненко, Л. І. Краснікова. — К. : Знання, 1998 — 493 с.

20. Dickey–Fuller test. [Електронний ресурс]. – Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Dickey%E2%80%93Fuller_test

21. ADF — Augmented Dickey Fuller Test. [Електронний ресурс]. – Режим доступу до ресурсу:

<https://www.statisticshowto.datasciencecentral.com/adfaugmented-dickey-fuller-test/>

22. Unit Root & Augmented Dickey-Fuller (ADF) Test. [Электронный ресурс]. – Режим доступа до ресурсу: http://www.ams.sunysb.edu/~zhu/ams586/UnitRoot_ADF.pdf

23. Bollerslev T. Generalized Autoregressive Conditional Heteroscedasticity / T. Bollerslev // *Journal of Econometrics*. – 1986. – Vol. 31. – P. 307–327.

24. Caner M. Treshold Autoregression with a Unit Root / M. Caner, B. Hansen // *Econometrica*. – 2001. – Vol. 69. – P. 1555–1596.

25. Charemza W. W. New Direction in Econometric Practice / W. W. Charemza, D. F. Deadman. – Brookfield : Edward Eglar, 1992. – 343 p.

26. Davidson R. Problems with the Estimation of Moving Average Processes / R. Davidson // *Journal of Econometrics*. – 1981. – № 16. – P. 295–310.

27. Davidson R. Several Tests for Model specification in the Presence of Alternative Hypothesis / R. Davidson, J. MacKinnon // *Econometrica*. – 1981. – Vol. 49. – P. 781–793.

28. Choudhry, T., Hassan, S.S. and Shabi, S. (2015) Relationship between Gold and Stock Markets during the Global Financial Crisis: Evidence from Nonlinear Causality Tests. *International Review of Financial Analysis*, 41, 247-256.

29. Adhikari R. An Introductory Study on Time Series Modeling and Forecasting [Текст] / Adhikari R. – Riga: LAP Lambert Academic Publishing, 2013. – 76 p.

ДОДАТКИ

ДОДАТОК А

SUMMARY

Diatlenko A.O. Modeling and forecasting possibility the number of cyberattacks for business entities. – Masters-level Qualification Thesis. Sumy State University, Sumy, 2021.

Cyber attacks (types and their essence), methods and models of forecasting the number of cyberattacks were studied. The main purpose of the study is to build a model for predicting the potential volume of cyberattacks on businesses.

Keywords: cyberattack, forecasting, modeling, time series.

АНОТАЦІЯ

Дятленко А.О. Моделювання та прогнозування можливості кібератак на суб'єкти господарювання. – Кваліфікаційна магістерська робота. Сумський державний університет, Суми, 2021р.

В даній роботі було досліджені кібератаки (види та їх сутність), методи і моделі прогнозування кількості кібератак. Основною метою дослідження є побудова моделі прогнозування кількості кібератак на суб'єкти господарювання.

Ключові слова: кібератака, моделювання, прогнозування, часові ряди.

ДОДАТОК Б

Таблиця Б-1 – Початкові

| Тиждень | Кількість атак | Тиждень | Кількість атак | Тиждень | Кількість атак | Тиждень | Кількість атак | Тиждень | Кількість атак | Тиждень | Кількість атак | Тиждень | Кількість атак | Тиждень | Кількість атак |
|---------|----------------|---------|----------------|---------|----------------|---------|----------------|---------|----------------|---------|----------------|---------|----------------|---------|----------------|
| 1 | 8 | 51 | 11 | 101 | 12 | 151 | 17 | 201 | 18 | 251 | 42 | 301 | 37 | 351 | 35 |
| 2 | 1 | 52 | 3 | 102 | 21 | 152 | 19 | 202 | 13 | 252 | 40 | 302 | 38 | 352 | 24 |
| 3 | 2 | 53 | 6 | 103 | 9 | 153 | 27 | 203 | 12 | 253 | 42 | 303 | 42 | 353 | 25 |
| 4 | 2 | 54 | 4 | 104 | 12 | 154 | 28 | 204 | 26 | 254 | 28 | 304 | 43 | 354 | 22 |
| 5 | 0 | 55 | 3 | 105 | 29 | 155 | 28 | 205 | 23 | 255 | 21 | 305 | 54 | 355 | 17 |
| 6 | 3 | 56 | 4 | 106 | 37 | 156 | 22 | 206 | 33 | 256 | 37 | 306 | 55 | 356 | 21 |
| 7 | 5 | 57 | 3 | 107 | 26 | 157 | 24 | 207 | 36 | 257 | 36 | 307 | 60 | 357 | 21 |
| 8 | 3 | 58 | 2 | 108 | 11 | 158 | 25 | 208 | 35 | 258 | 23 | 308 | 71 | 358 | 25 |
| 9 | 5 | 59 | 2 | 109 | 16 | 159 | 16 | 209 | 31 | 259 | 13 | 309 | 126 | 359 | 36 |
| 10 | 3 | 60 | 1 | 110 | 6 | 160 | 13 | 210 | 26 | 260 | 14 | 310 | 46 | 360 | 40 |
| 11 | 7 | 61 | 6 | 111 | 13 | 161 | 10 | 211 | 32 | 261 | 21 | 311 | 80 | 361 | 55 |
| 12 | 2 | 62 | 0 | 112 | 14 | 162 | 18 | 212 | 32 | 262 | 16 | 312 | 68 | 362 | 34 |
| 13 | 8 | 63 | 5 | 113 | 19 | 163 | 25 | 213 | 17 | 263 | 13 | 313 | 62 | 363 | 46 |
| 14 | 4 | 64 | 7 | 114 | 26 | 164 | 36 | 214 | 28 | 264 | 11 | 314 | 81 | 364 | 27 |
| 15 | 4 | 65 | 6 | 115 | 34 | 165 | 14 | 215 | 87 | 265 | 33 | 315 | 53 | 365 | 36 |
| 16 | 4 | 66 | 6 | 116 | 41 | 166 | 13 | 216 | 58 | 266 | 21 | 316 | 68 | 366 | 11 |
| 17 | 1 | 67 | 5 | 117 | 33 | 167 | 8 | 217 | 50 | 267 | 28 | 317 | 29 | | |
| 18 | 8 | 68 | 2 | 118 | 24 | 168 | 17 | 218 | 64 | 268 | 20 | 318 | 41 | | |
| 19 | 7 | 69 | 7 | 119 | 17 | 169 | 10 | 219 | 28 | 269 | 23 | 319 | 32 | | |
| 20 | 7 | 70 | 4 | 120 | 23 | 170 | 7 | 220 | 52 | 270 | 35 | 320 | 29 | | |
| 21 | 7 | 71 | 2 | 121 | 68 | 171 | 9 | 221 | 31 | 271 | 16 | 321 | 23 | | |
| 22 | 5 | 72 | 5 | 122 | 18 | 172 | 19 | 222 | 16 | 272 | 43 | 322 | 35 | | |
| 23 | 5 | 73 | 7 | 123 | 66 | 173 | 34 | 223 | 15 | 273 | 50 | 323 | 29 | | |
| 24 | 4 | 74 | 3 | 124 | 37 | 174 | 16 | 224 | 7 | 274 | 24 | 324 | 55 | | |
| 25 | 8 | 75 | 0 | 125 | 26 | 175 | 15 | 225 | 53 | 275 | 23 | 325 | 44 | | |
| 26 | 5 | 76 | 3 | 126 | 33 | 176 | 41 | 226 | 15 | 276 | 16 | 326 | 72 | | |
| 27 | 1 | 77 | 6 | 127 | 36 | 177 | 16 | 227 | 22 | 277 | 45 | 327 | 45 | | |
| 28 | 2 | 78 | 4 | 128 | 50 | 178 | 10 | 228 | 34 | 278 | 27 | 328 | 14 | | |
| 29 | 3 | 79 | 1 | 129 | 19 | 179 | 21 | 229 | 20 | 279 | 22 | 329 | 81 | | |
| 30 | 6 | 80 | 8 | 130 | 33 | 180 | 15 | 230 | 38 | 280 | 18 | 330 | 87 | | |
| 31 | 8 | 81 | 2 | 131 | 41 | 181 | 22 | 231 | 29 | 281 | 31 | 331 | 61 | | |
| 32 | 8 | 82 | 3 | 132 | 26 | 182 | 23 | 232 | 32 | 282 | 19 | 332 | 75 | | |

Продовження табл. Б1

| | | | | | | | | | | | | | |
|----|--------|-----|----|-----|----|-----|----|-----|----|-----|-----|-----|----|
| 33 | 5 | 83 | 3 | 133 | 27 | 183 | 23 | 233 | 44 | 283 | 14 | 333 | 66 |
| 34 | 8 | 84 | 12 | 134 | 24 | 184 | 13 | 234 | 51 | 284 | 16 | 334 | 42 |
| 35 | 4 | 85 | 9 | 135 | 16 | 185 | 22 | 235 | 30 | 285 | 32 | 335 | 33 |
| 36 | 5 | 86 | 12 | 136 | 14 | 186 | 19 | 236 | 81 | 286 | 19 | 336 | 46 |
| 37 | 3 | 87 | 20 | 137 | 35 | 187 | 32 | 237 | 64 | 287 | 26 | 337 | 76 |
| 38 | 7 | 88 | 12 | 138 | 10 | 188 | 36 | 238 | 43 | 288 | 108 | 338 | 47 |
| 39 | 4 | 89 | 15 | 139 | 16 | 189 | 41 | 239 | 53 | 289 | 34 | 339 | 17 |
| 40 | 1 0 | 90 | 6 | 140 | 18 | 190 | 29 | 240 | 65 | 290 | 38 | 340 | 48 |
| 41 | 1 3 | 91 | 1 | 141 | 8 | 191 | 31 | 241 | 51 | 291 | 33 | 341 | 30 |
| 42 | 9 | 92 | 7 | 142 | 15 | 192 | 28 | 242 | 78 | 292 | 29 | 342 | 33 |
| 43 | 3 | 93 | 7 | 143 | 20 | 193 | 21 | 243 | 51 | 293 | 32 | 343 | 35 |
| 44 | 6 | 94 | 7 | 144 | 16 | 194 | 25 | 244 | 46 | 294 | 20 | 344 | 22 |
| 45 | 6 | 95 | 9 | 145 | 23 | 195 | 21 | 245 | 44 | 295 | 23 | 345 | 40 |
| 46 | 1 0 | 96 | 3 | 146 | 14 | 196 | 10 | 246 | 42 | 296 | 30 | 346 | 50 |
| 47 | 9 | 97 | 14 | 147 | 32 | 197 | 20 | 247 | 32 | 297 | 16 | 347 | 80 |
| 48 | 1 2 | 98 | 20 | 148 | 22 | 198 | 10 | 248 | 38 | 298 | 21 | 348 | 55 |
| 49 | 8 | 99 | 14 | 149 | 19 | 199 | 22 | 249 | 26 | 299 | 42 | 349 | 59 |
| 50 | 3 | 100 | 11 | 150 | 20 | 200 | 44 | 250 | 34 | 300 | 35 | 350 | 29 |