

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Тернопільський національний економічний університет
Навчально-науковий інститут інноваційних освітніх технологій
Кафедра комп'ютерної інженерії

ТИМЧУК Юлія Сергіївна

**ARIMA-моделі прогнозування прибутків
комп'ютерної фірми / ARIMA-models for finance
forecasting of a computer firm**

спеціальність: 123 - Комп'ютерна інженерія
магістерська програма - Комп'ютерна інженерія

Магістерська робота

Виконала студентка групи КІзм-21
Ю.С. Тимчук

Науковий керівник:
к.т.н., доцент, К. М. Березька

Магістерську роботу допущено до захисту:

ТЕРНОПІЛЬ -2018

РЕЗЮМЕ

Магістерська робота на тему «ARIMA-моделі прогнозування прибутків комп'ютерної фірми» на здобуття освітньо-кваліфікаційного рівня «Магістр» зі спеціальності «Комп'ютерні системи та мережі» написана обсягом 95 сторінок і містить 34 ілюстрації, 4 таблиці, 4 додатки та 50 джерел за переліком посилань.

Метою роботи є розроблення алгоритмів побудови ARIMA-моделей та проектування програмної системи прогнозування прибутків комп'ютерної фірми.

Методи досліджень. Для розв'язання поставлених задач у магістерській роботі використано методи: математична статистика, статистика, кореляційний і регресійний аналіз, авторегресійні моделі та моделі ковзної середньої, метод найменших квадратів, метод максимальної правдоподібності.

Результати дослідження: алгоритми методології Бокса-Дженкінса підбору ARIMA моделі, програмна система комп'ютерного моделювання алгоритмів прогнозування за ARIMA-моделями.

Результати роботи можуть бути використані для наукових досліджень прогнозування стратегії розвитку фірм та в навчальному процесі.

Орієнтовні напрямки розвитку досліджень: продовження робіт по побудові ARIMA моделей: коінтеграція, уявна регресія, авторегресійні моделі при наявності автокореляції помилок, проектування програмних систем побудови ARIMA моделей.

КЛЮЧОВІ СЛОВА: АЛГОРИТМ, ARIMA-МОДЕЛЬ, ТРЕНД, МЕТОДОЛОГІЯ БОКСА-ДЖЕНКІНСА, ПРОГРАМНА СИСТЕМА.

RESUME

Diploma work: «ARIMA-models for finance forecasting of a computer firm» to education and qualification of «Master» specialty «Computer systems and networks» written 95 page volume and contains 34 illustrations, 4 tables, 4 applications and 50 sources for references.

The purpose of the work is to develop algorithms for building ARIMA-models and design a software system for forecasting profits of a computer firm.

Research methods. Methods for solving the tasks set in master's work include mathematical statistics, statistics, correlation and regression analysis, autoregressive models and moving average models, least squares method, maximum likelihood method.

Results of the research: algorithms of the Box-Jenkins methodology selection ARIMA model, software system of computer simulation of prediction algorithms by ARIMA-models.

The results of the work can be used for scientific research forecasting the strategy of development of firms and in the educational process.

Approximate directions of development of researches: continuation of works on construction of ARIMA models: cointegration, imaginary regression, autoregressive models in the presence of error autocorrelation, design of software systems for constructing ARIMA models.

KEY WORDS: ALGORITHMS, ARIMA-MODEL, TREND, BOX-JENKINS METHODOLOGY, SOFTWARE SYSTEM.

ЗМІСТ

Вступ	7
1. Аналіз методів, алгоритмів і програмних засобів прогнозування прибутку комп'ютерних фірм	10
1.1. Розвиток ІТ-бізнесу в Україні	10
1.2. Методи та алгоритми прогнозування	16
1.3. Програмні засоби прогнозування	24
1.4. Постановка задач дипломного проектування	33
2. Алгоритми побудови ARIMA-моделей	34
2.1. Алгоритми аналізу тренду	34
2.2. Алгоритми оцінювання моделі	39
2.3. Алгоритми методології Бокса-Дженкінса підбору ARIMA-моделі..	47
3. Комп'ютерне моделювання алгоритмів прогнозування за ARIMA-моделями	56
3.1. Програмно-апаратні засоби	56
3.2. Структура програмної системи.....	58
3.3. Модуль ідентифікації моделі	61
3.4. Модуль оцінювання моделі і перевірки адекватності моделі. Модуль прогнозування за моделлю	66
Висновки.....	73
Список використаних джерел	74
Додаток А Екранні форми програмної системи	79
Додаток Б Лістинг коду програми.....	82
Додаток В Світлокопія виданої публікації.....	87
Додаток Г Довідка про використання результатів дипломної роботи	95

ВСТУП

Актуальність теми. Розвиток інформаційних технологій набирає у сучасному світі усе більших обертів. ІТ-сфера тісно та швидко втілюється у всі сфери світової економіки, впливає на загальне зростання економічних та соціальних показників підприємств і фірм. Для України ІТ-сфера має особливо важливе значення, адже на фоні загальної економічної та політичної кризи, вона здатна згладити ці явища за допомогою стрімкого розвитку, а також забезпечення функціонування багатьох суміжних галузей [1]. Тому дослідження тенденцій розвитку українських комп'ютерних фірм, а також виявлення перспектив та основних напрямів їх руху є актуальними.

Підхід ARIMA – це модний напрям в прогнозуванні. Він дозволяє скромними засобами за допомогою невеликої кількості параметрів описати нескінченну різноманітність часових рядів, що спостерігаються в реальному житті.

Мета і завдання дослідження. Метою роботи є розробка алгоритмів для аналізу, побудови та прогнозу за ARIMA-моделями.

Для досягнення поставленої мети необхідно розв'язати такі завдання:

1. Проаналізувати розвиток ІТ-бізнесу в Україні.
2. Зробити аналіз методів та алгоритмів прогнозування.
3. Проаналізувати програмні засоби прогнозування.
4. Розробити алгоритми побудови ARIMA-моделей.
5. Розробити алгоритми обчислення показників якості побудованої моделі.
6. Розробити алгоритми вибору моделі з декількох існуючих.
7. Здійснити комп'ютерне моделювання програмної системи.
8. Розробити модуль ідентифікації моделі.
9. Розробити модуль оцінювання моделі і перевірки її адекватності.
10. Розробити модуль прогнозування за ARIMA-моделями.

Об'єкт дослідження – статистичні дані прибутків комп'ютерної фірми.

Предмет дослідження – алгоритми побудови ARIMA-моделей.

Методи досліджень. При розробці ARIMA-моделей, а також методів прогнозування в роботі застосовувалися математична статистика, статистика, кореляційний і регресійний аналіз, метод максимальної правдоподібності [2,3].

Наукова новизна

1. Алгоритм аналізу тренду.
2. Алгоритм стратегії вибору моделі за методом Бокса-Дженкінса.

Практичне значення отриманих результатів полягає в розробленні:

1. Модуля ідентифікації моделі.
2. Модуля оцінювання моделі і перевірки її адекватності.
3. Модуля прогнозування за ARIMA-моделями.

Публікації та апробація ДР. Основні результати досліджень опубліковано в тезах доповідей на всеукраїнській науково-практичній конференції «Прикладна геометрія та інформаційні технології в моделюванні об'єктів, явищ і процесів» (AGIT-2018), (м. Миколаїв) [2]:

Березька К.М. Прогнозування прибутків комп'ютерної фірми на основі ARIMA-моделей / К.М. Березька, В.М. Неміш, Ю.С. Тимчук // Матеріали III всеукраїнської науково-практичної конференції «Прикладна геометрія та інформаційні технології в моделюванні об'єктів, явищ і процесів» (AGIT-2018), м. Миколаїв, 17–19 жовтня 2018 р. – Миколаїв: МНУ імені В.О. Сухомлинського, 2018. – С. 120-121.

Дипломна робота складається із трьох розділів, висновків, списку використаної літератури та додатків [3].

У першому розділі здійснено аналіз методів, алгоритмів прогнозування (в тому числі й прибутку комп'ютерних фірм), точність і перевірку прогнозів, контроль прогнозу, метод екстраполяції тренду, факторний аналіз. Проаналізовано розвиток ІТ-бізнесу в Україні і програмні засоби, що використовуються для прогнозування.

У другому розділі розроблено алгоритми аналізу тренду, оцінювання моделі, перевірки її адекватності даним, алгоритми методології Бокса-Дженкінса підбору *ARIMA* моделі.

У третьому розділі розроблено програмну систему комп'ютерного моделювання алгоритмів підбору *ARIMA* моделі, розроблено модуль ідентифікації моделі, оцінювання моделі і перевірки її адекватності, прогнозування за *ARIMA*-моделями.

У додатках приведено екранні форми програмної системи, текст програми, довідку про використання результатів дипломної роботи, світлокопію виданої публікації.

1 АНАЛІЗ МЕТОДІВ, АЛГОРИТМІВ І ПРОГРАМНИХ ЗАСОБІВ ПРОГНОЗУВАННЯ ПРИБУТКУ КОМП'ЮТЕРНИХ ФІРМ

1.1 Розвиток ІТ-бізнесу в Україні

ІТ-галузь України була створена з нуля і стала є однією з головних експортних галузей економіки України, яка створила понад 100 000 робочих місць високооплачуваних фахівців.

Якщо порівнювати розвинені країни і Україну, то розвинені зараз країни в минулому були схожі на сучасну Україну, у якій сьогодні є проблеми з технологіями, які створюють найбільшу додану вартість. У країн, які змогли забезпечити добробут своїм громадянам, є спільна риса: вони досягли успіху в розвитку технологій. Уряди створили відповідні умови (і на цьому їх втручання закінчилося), інноватори за рахунок вкладених коштів бізнесу отримали можливість для застосування своїх розробок, які в результаті бізнес використовував для того, щоб стати ефективнішим.

Вартість технологій у світі падає, тому Україні потрібно поспішати з розвитком галузі. Наприклад, 10 років тому, щоб розшифрувати геном людини треба було \$10 тис., зараз – в десять разів менше, а через десяток років це буде коштувати всього \$100 [4]. Тому необхідно створювати такі умови, щоб українські вчені та програмісти працювали в Україні, а не виїжджали з країни, щоб, навпаки, сюди приїжджали іноземні працівники, а міжнародні компанії відкривали тут свої офіси.

Темпи зростання ІТ в Україні зростають, хоча в глобальному масштабі успіх української ІТ-індустрії непомітний.

Приведемо аналіз ІТ-індустрії України на світовому ринку.

1. За рівнем витрат на ІТ.

У рейтингу рівня витрат на ІТ та залежності від цього рівня економічного розвитку Україна (\$32,2 ВВП на особу) займає одне з останніх місць (за даними за 2015 рік). Нижче від нас стоїть тільки Індія (рисунок 1.1).

Через такий низький рівень споживання ІТ в Україні спостерігається низький рівень економічного добробуту. Основою економіки України є сировина, а не технології, що зумовило відставання у темпах розвитку від розвинутих країн.

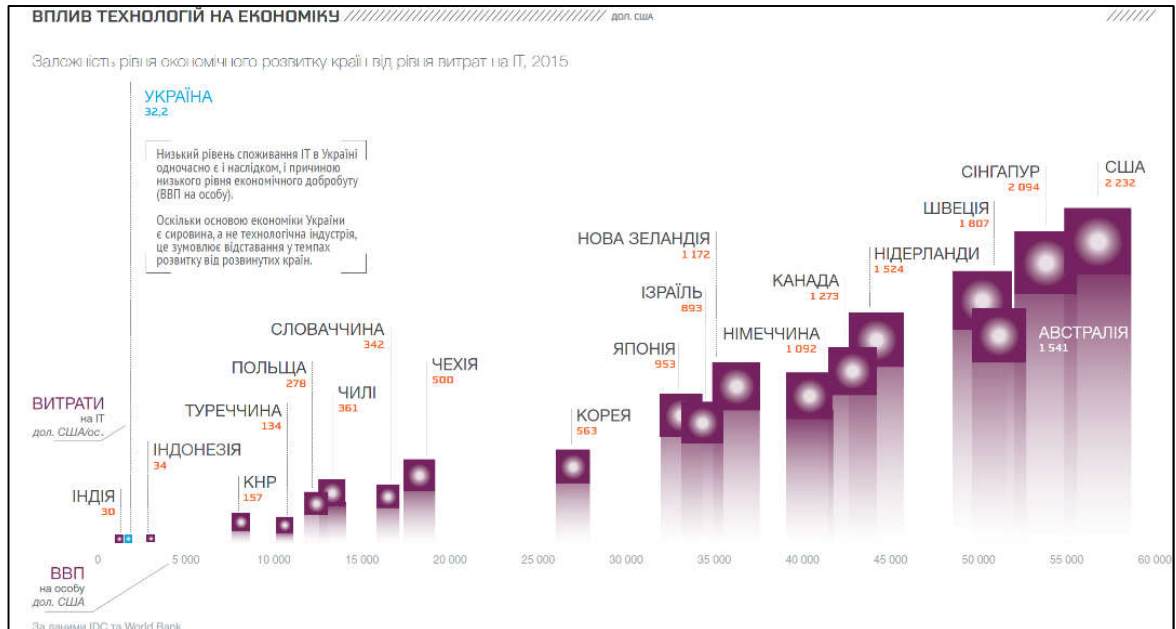


Рисунок 1.1 – Залежність рівня економічного розвитку країн від рівня витрат на ІТ

2. За обсягом внутрішнього ринку ІТ.

Обсяг внутрішнього ринку ІТ – це важливий економічний показник. Він показує, наскільки в країні розвинені інформаційні технології у порівнянні з іншими країнами.

Як бачимо (рисунок 1.2) обсяг ринку ІТ серед приведених країн є нарівні Ізраїлю і Сінгапуру та одним з найнижчих.

3. За розвитком аутсорсингу.

Україна займає перше місце на ринку ІТ-аутсорсингу розробки програмного забезпечення у Центральній та Східній Європі, а на світовому ринку входить у перші 50 країн за обсягом ІТ-аутсорсингу [5]. Погано розвиненими є створення готового ПЗ і "заліза". На глобальному ІТ-ринку наша країна нарівні з Німеччиною і Японією. Обсяг ринку аутсорсингу України становить \$1,1 млрд. (рисунок 1.3). При аутсорсингу виконуються певні

фрагменти кодів для програми, а замовник на їх основі створює готовий продукт. При цьому інвестор не наймає працівника в штат, а наймає за договором на виконання певної частини проекту. Ціни на працю українських спеціалістів є меншими, а якість виконання робіт не поступається провідним країнам світу. Тому закордонним компаніям вигідно наймати їх. Але прибуток від готового

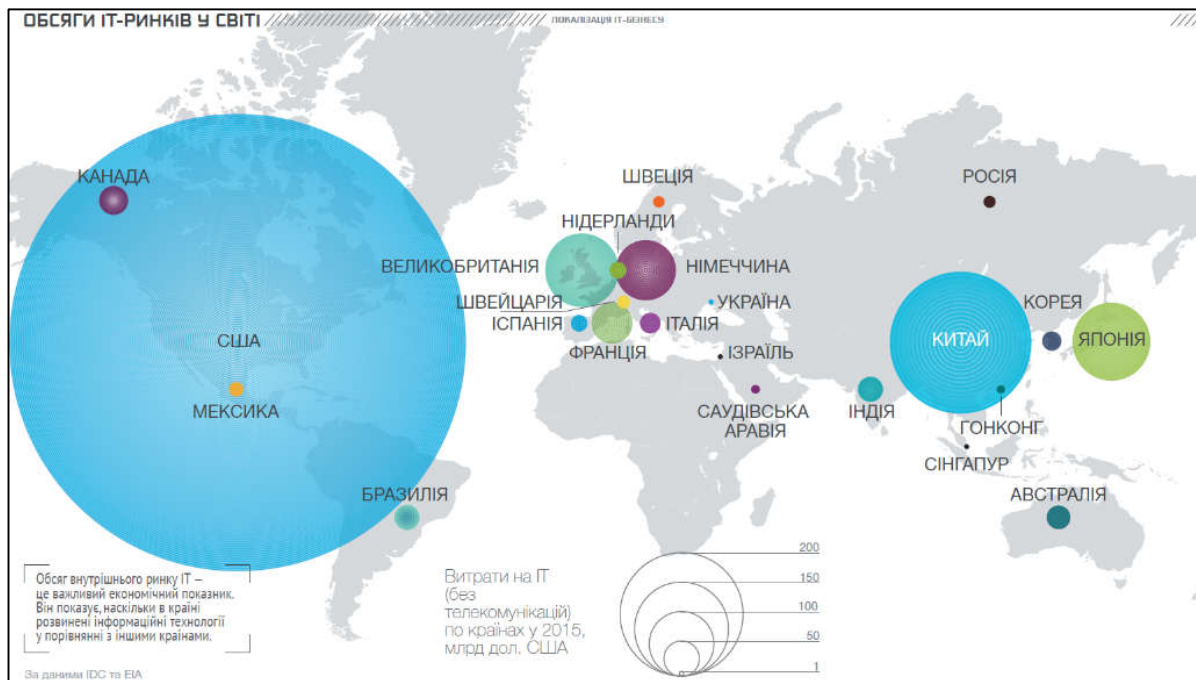


Рисунок 1.2 – Обсяги ІТ-ринків у світі

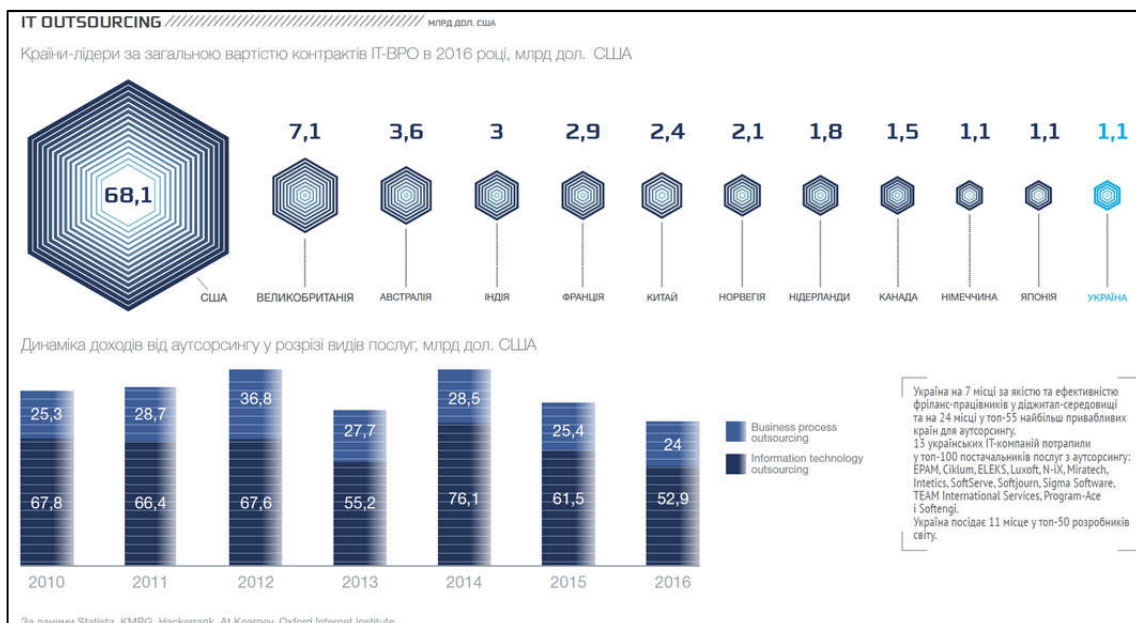


Рисунок 1.3 – Обсяг ринку аутсорсингу

продукту в Україні не залишається, лише українські фахівці набувають досвід роботи з іноземними проектами, набувають навичок у своїй спеціалізації, а також підіймають авторитет ІТ-працівників України на міжнародному рівні.

4. За ІСТ-експортом (ІТ+телеком).

ІСТ-експорт України є незначний (рисунок 1.4). Вона посідає 33 місце і наприклад, обсяг експорту ІСТ Польщі у 5 разів більший за український.

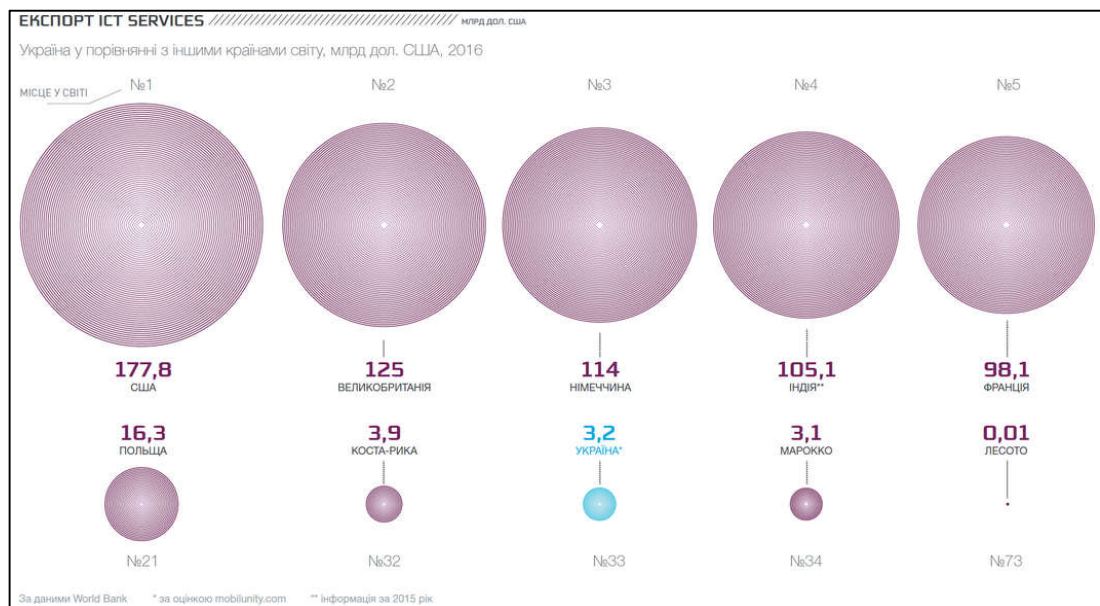


Рисунок 1.4 – ІСТ-експорт

На думку World Bank наша держава ніяк не підтримує ІТ галузь. Українське законодавство є недосконалим, існує слабкий правовий захист бізнесу, провідні спеціалісти ІТ галузі емігрують закордон, спостерігається низький рівень володіння англійською мовою, слабкість деяких секторів системи освіти тощо [5].

Інвестиції в ІТ грають важливу роль в українській економіці. ІТ-індустрія створює 4% ВВП країни. Україна є лідером в Європі за обсягом експорту ІТ-послуг. У 2015 році Україна збільшила об'єм експорту ІТ-послуг на 17%, у грошовому еквіваленті це становить 2,1 млрд доларів [6].

У 2014 році в Україні (за даними Ukraine Digital News) працювала найбільша кількість ІТ професіоналів у Європі – 90 тис. осіб, більше тисячі компаній з надання ІТ-послуг. В Україні розмістили об'єкти своїх науково-дослідницьких робіт такі міжнародні гіганти як Aricent, Boeing, Ericsson, Huawei, Oracle, Siemens і Teleperformance [5]. У 2014 році в Україні було сформовано 6 ІТ-кластерів у містах Львові, Одесі, Києві, Харкові, Дніпропетровську та Луцьку. Ці спільноти провідних компаній у галузі інформаційних технологій спільно із університетами та міською владою покращують та розвивають ІТ у цих містах[6].

На рисунку 1.5 приведена динаміка ІСТ- експорту з України.

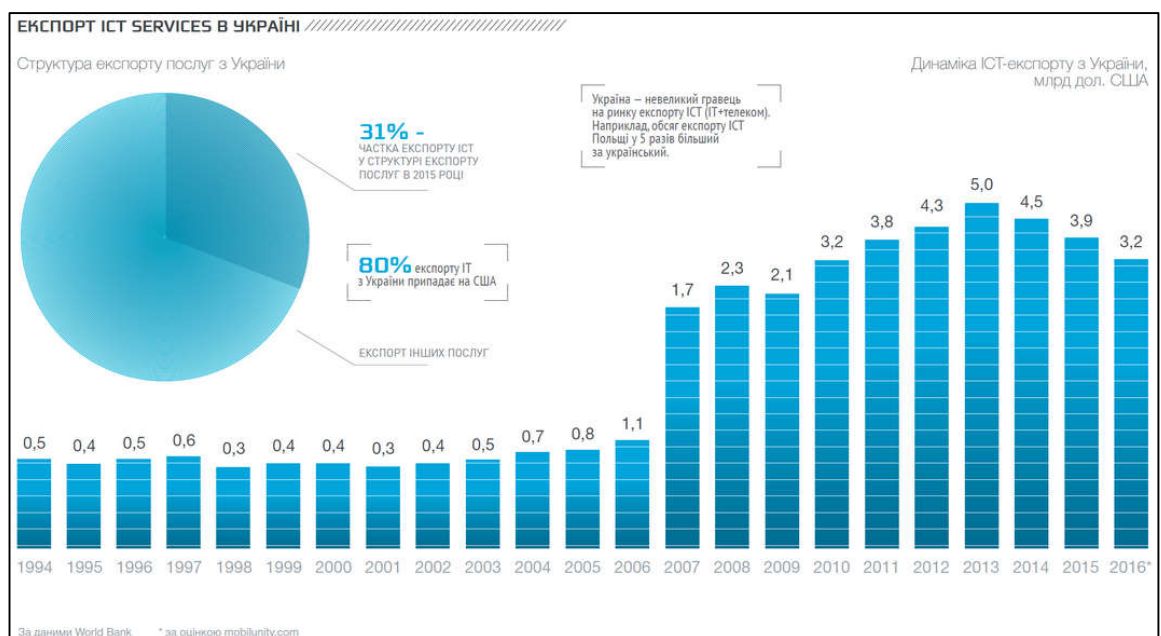


Рисунок 1.5 – Динаміка ІСТ- експорту з України

5. За ростом внутрішнього ринку.

Внутрішній ринок ІТ в 2016 році виріс в загальному на 22%. У той же час розвинені світові ринки ростуть не більше ніж на 5% в рік (рисунок 1.6).

Як бачимо (рисунок 1.7) три сегменти ринку ІТ: hardware – «залізо», packaged software – готове ПЗ і services – послуги в 2016 році зросли і за прогнозами [4], вони будуть рости ще більше. «Залізо» виросло з \$1,2 до \$1,5

млрд; ринок готового ПЗ виріс з \$0,11 до \$0,12 млрд; ринок послуг – з \$0,18 до \$0,19 млрд.

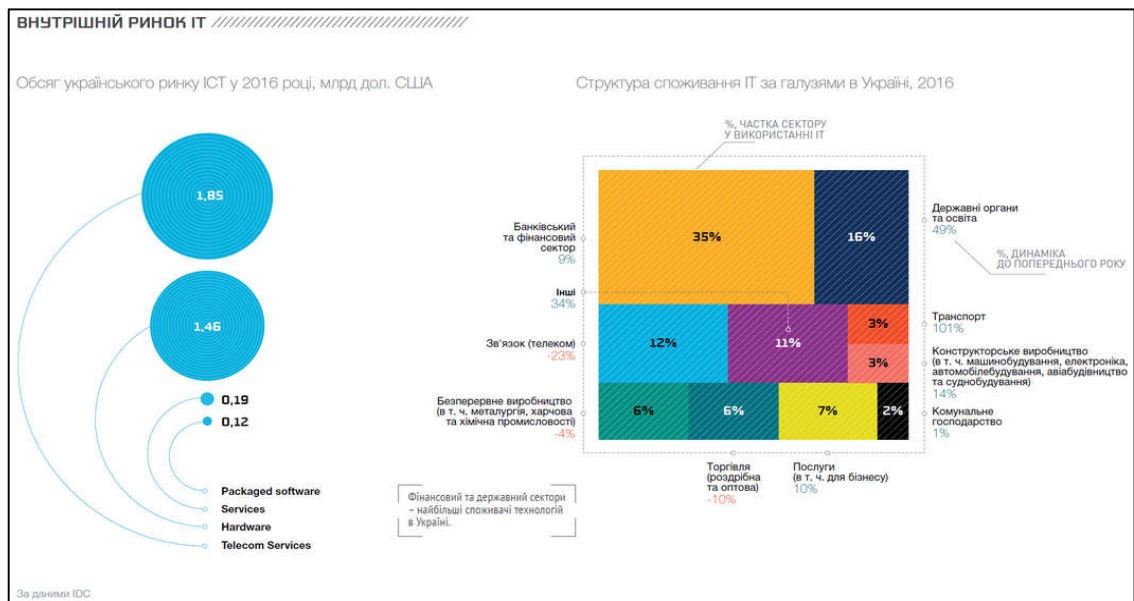


Рисунок 1.6 – Структура споживання ІТ за галузями в Україні

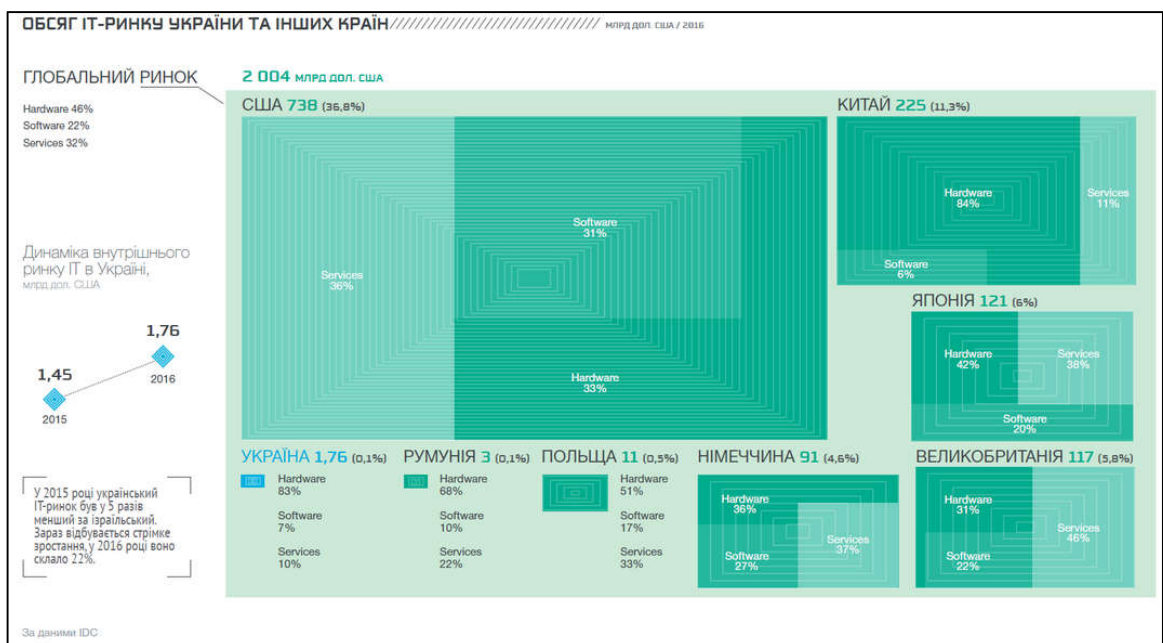


Рисунок 1.7 – Структура трьох сегментів ринку ІТ

На рисунку 1.8 приведено розподіл ІТ-ринку за регіонами.

Показники прогнозу (числові характеристики об'єкта, обсяги і терміни робіт і т. п.) мають ймовірнісну природу. На основі прогнозів здійснюється передбачення і приймаються управлінські рішення.

Мета прогнозування – отримати науково обґрунтовані варіанти тенденцій розвитку (зміни) керованого об'єкта (показників його стану) в часі і просторі. Джерелами інформації для прогнозів являються вербальні і письмові тексти, що отримуються в процесі комунікацій між людьми або у відкритому друці. Інформацію з відкритого друку отримують, використовуючи методи: структурно-морфологічний, визначення публічної активності, виявлення груп патентних документів, аналізу показників, термінологічного і лексичного аналізу.

Для прогнозування в практичній діяльності застосовуються різні кількісні і якісні методи.

Кількісні методи базуються на інформації, яку можна отримати, на основі вивчення тенденцій зміни параметрів або маючи статистично достовірні залежності, що характеризують виробничу діяльність об'єкта управління. Як приклади таких методів можуть виступати: аналіз тимчасових рядів, каузальне (причинно-наслідкове) моделювання.

Вирізняють такі типові методи кількісного прогнозування:

- аналіз часових рядів, що ґрунтується на дослідженні подій, які відбулися в минулому, є основою для планування. Його можна провести за допомогою таблиці або графіка шляхом нанесення на координатну сітку точок, що відповідають подіям минулого. Цей метод часто використовується з метою оцінювання попиту на товари й послуги, визначення потреб у запасах і прогнозування структури збуту;

- каузальне (причинно-наслідкове) моделювання – спроба спрогнозувати те, що відбудеться у подібних ситуаціях, шляхом дослідження статистичної залежності між фактором, що розглядається, й іншими змінними. Ця залежність називається кореляцією. Чим тісніша кореляція, тим більша придатність моделі для прогнозування.

У основу якісних методів покладені експертні оцінки фахівців в сфері прийняття рішень: наприклад, методи експертних оцінок, висновок, моделі очікування споживача (опитування клієнтів).

Складні об'єкти прогноуються з використанням різних кількісних і якісних методів. Наприклад, прогноз економічної кон'юнктури (сукупність ознак, що характеризують стан економіки в певний період) базується на прогнозах у сфері обмежень по захисту навколишнього середовища, міжнародної торгівлі, попиту на продукцію, пропозиції продукції та їх співвідношень. При цьому кожний з вказаних прогнозів, у свою чергу, ґрунтується на проміжних прогнозах різних процесів. Відпрацьованими методами прогнозування економічної кон'юнктури є: "мозкова атака", метод Дельфі, екстраполяції тенденцій, морфологічний аналіз, імітаційне динамічне моделювання, структурний аналіз і ін. Існують і інші класифікації методів прогнозування, які визначені особливостями прогнозів.

Цілі, час, умови прогнозу і специфіка його вироблення визначають комплекс методів і прийомів прогнозування. При цьому різні методи можуть використовуватися в розробці різних прогнозів.

Точність і перевірка прогнозів. Точність і перевірка прогнозів – важливий аспект прогнозування. Прогнозування майбутніх значень на регулярній основі ускладнюється важкою природою змінних. Саме тому необхідно включити в прогноз показник можливого відхилення значення змінної. Необхідність вибору еталону точності визначається:

- різними показниками прогнозів, що отримуються;
- потребою мати в прогнозі декілька показників;
- важливістю контролю помилок прогнозу (в порівнянні з еталоном).

Помилка (Π) в прогнозі означає різницю між фактичним і прогнозним значеннями:

$$\Pi = \Phi Z - \Pi Г ,$$

де ΦZ – фактичне значення; $\Pi \Gamma$ – прогноз.

Помилка позитивна, якщо прогноз занижений, і негативна, якщо він завищений. Помилки в прогнозі впливають на рішення при виборі різних варіантів прогнозу і на результат використання того або іншого методу прогнозування.

Обчислення точності прогнозу. Визначити помилку у вчисленому значенні прогнозу неможливо, оскільки невідомо його дійсне значення. Однак існує ймовірність того, що помилка обчисленого прогнозу не перевищить деяку величину або максимальну помилку прогнозу, яку можна очікувати із заданою ймовірністю:

- математичне очікування (МО):

$$MO = \frac{\sum |DZ - \Pi \Gamma|}{n},$$

де DZ – дійсне значення.

- середньоквадратичне відхилення (СКВ)

$$СКВ = \frac{\sum (DZ - \Pi \Gamma)^2}{n - 1}.$$

Контроль прогнозу. Контроль прогнозу здійснюється порівнянням помилок прогнозу із заздалегідь визначеними значеннями (межами). На практиці контроль здійснюється за допомогою розрахунку відношення сукупної помилки прогнозу до відповідного значення MO і використовується для спостереження за прогнозом:

$$\Pi_0 = \frac{\sum (DZ - \Pi \Gamma)}{MO},$$

де P_0 – показник відхилення.

Значення показників відхилення порівнюються із межами значень показників, що засновані на судженнях і досвіді.

Згідно теорії ймовірності з урахуванням нових законів розсіювання, відхилення випадкової величини від центру групування не перевищує трьох середніх квадратних відхилень. З урахуванням цього значення показника відхилення повинне знаходитися в межах ± 4 , що відповідає межам трьох стандартних відхилень.

Використання контрольних графіків. Графік передбачає установку верхніх і нижніх меж для окремих помилок прогнозу (замість сукупних помилок, як у випадку з показником відхилення). Ці межі кратні квадратному кореню із СКВ.

Графічне відображення помилок з використанням контрольної діаграми інформативно і наглядно представляє відхилення, що виходять за встановлені межі.

Вибір методу прогнозування. Вибір методу прогнозування значно покращує якість прогнозу, оскільки:

- забезпечує функціональну повноту, достовірність і точність прогнозу;
- зменшує тимчасові і матеріальні витрати на прогнозування. На вибір методу прогнозування впливають:
 - сутність практичної проблеми, що підлягає розв'язанню;
 - динамічні характеристики об'єкта прогнозування у ринковому середовищі;
 - вид і характер наявної інформації, типове уявлення про об'єкт прогнозування;
 - комбінація фаз життєвого, ринкового циклу товару або послуги;
 - період попередження і його співвідношення із передбачуваною тривалістю ринкового, життєвого циклу, циклу розробки або модифікації товару, послуги.

Всі перераховані чинники повинні розглядатися в системній єдності, лише несуттєві можуть виключатися з розгляду. Той, хто прогнозує внаслідок передпрогнозних досліджень повинен структурувати інформацію про об'єкт прогнозування, проаналізувати її і, вибравши, залишити ту, що відповідає умовам прогнозу. При підготовці рішення про вибір методу прогнозування рекомендується виділити ті, що можливо застосовувати до умов даного завдання. Ті, що застосувати не можна, виключаються. На практиці ж, вибираючи метод прогнозування, рекомендується враховувати два найбільш важливих чинники — затрати і точність. Слід з'ясувати, скільки грошей виділено на підготовку прогнозу, яка можлива ціна спрогнозованих помилок і вихід. Найкращий прогноз, як правило, є оптимальною комбінацією точності і вартості.

При виборі методу прогнозування потрібно враховувати:

- наявність статистичних даних за необхідний період;
- компетентність прогнозиста, наявність ВОТ;
- час, що необхідний для збору і аналізу інформації. У ряді випадків для отримання незалежних прогнозів використовуються одночасно декілька методів. Вимоги до прогнозів:

- своєчасність, певною мірою точності і визначеності інших показників;
- надійність, що виражено в знакових одиницях (доларах, одиницях продукції, обладнанні, кваліфікації персоналу і т. п.) і зафіксована на папері;
- простота методики прогнозування для розуміння і використання.

Проаналізуємо найбільш вживані методи прогнозування [7], а саме: метод екстраполяції тренду та метод головних компонент.

Метод екстраполяції тренду дозволяє тенденції минулого розвитку досліджуваного об'єкта перенестись в майбутній період. Метод екстраполяції тренду використовується в основному для короткострокового прогнозування, строком до року. Даний метод реалізується для кожного певного об'єкта окремо і поетапно на кожний наступний момент поточного часу [8].

До переваг цього методу зазвичай відносять:

- невисокі витрати на реалізацію;
- швидке отримання результатів прогнозу;
- якщо показники макросередовища стабільні, то цей метод один з найефективніших;
- у підприємства завжди є необхідна інформація для розрахунків [8].

Недоліками методу екстраполяції тренду є:

- швидкі зміни тренду;
- можливість розробки тільки короткострокового прогнозу.

Факторний аналіз у сучасному вигляді являє собою певний розділ математичної статистики. Є велика різноманітність методів факторного аналізу та їх модифікацій, які відомі в даний час.

Основною перевагою прогнозування розвитку факторів, а не окремих змінних є те, що в процесі прогнозування факторів автоматично вирішується задача синтезу та пов'язування прогнозів окремих показників. У цьому і полягає основна перевага і перспектива використання методів факторного аналізу в дослідженнях складних багатопараметричних стохастичних процесів.

За джерелами [9] факторний аналіз може використовуватися у поєднанні з моделями регресійного аналізу. У цьому випадку будують регресію деякого вихідного показника досліджуваного комплексу, але не на сукупності вхідних показників, як зазвичай, а на головних компонентах, отриманих у процесі факторного аналізу.

Виділення головних факторів в компонентній моделі відбувається за максимально падаючій на них дисперсії вибірки. При цьому, як правило, вже невелике число головних компонент забирає на себе 80–90% загальної дисперсії. У зв'язку з цим решта незначущих в рівнянні регресії компонент виключаються з розгляду [9,10].

Слід зазначити, що оцінку головних компонент доцільно проводити двічі: спочатку в просторі вхідних змінних X , а потім в об'єднаному просторі всіх змінних комплексу YX . У разі близькості одержуваних факторних структур можна говорити про достатнє подання вхідних показників. У разі значної

різниці одержуваних факторних структур в обох просторах слід уточнити і доповнити склад і вимірювання вхідних показників для більш повного відображення в ньому вихідних змінних комплексу.

Переваги методу головних компонент:

– по-перше, розмірність моделі істотно менша, ніж регресії на вихідних змінних;

– по-друге, відбір найбільш істотних компонент проводиться за чітким кількісним критерієм;

– по-третє, факторна модель виключає мультиколінеарність за рахунок ортогональності головних чинників.

При прийнятті оперативних рішень на макро- та мікрорівні часто необхідно спиратись на якісні прогностні значення певних показників, не заглиблюючись у детальний аналіз факторів, які впливають на їх зміну. Якщо збір даних та аналіз всіх факторів для побудови багатофакторної регресійної моделі може бути тривалим і потребувати великих витрат ресурсів, невідповідних отриманому результату. Саме в таких випадках доцільно використовувати методи прогнозування на основі часових рядів (категорія – метод екстраполяції тренду), зокрема побудову ARIMA моделі часового ряду. Зауважимо, що побудова ARIMA моделей є також ефективним допоміжним інструментарієм для обчислення прогностичних значень окремих факторів, що впливають на змінну залежного показника під час прогнозування на основі багатофакторних регресійних моделей, моделей симульативних рівнянь тощо. Крім того, на практиці можна будувати різновиди ARIMA моделей, зокрема ARMAX моделі, які одночасно з лаговими змінними досліджуваного показника можуть враховувати і додаткові екзогенні фактори в різних формах. Такі своєрідні «гібриди» лінійної багатофакторної регресії та ARIMA моделей за правильної специфікації доволі ефективні та можуть забезпечити зменшення похибки прогнозу порівняно з класичними багатофакторними регресійними моделями або навіть з чистими ARIMA моделями [11].

В даному підрозділі проаналізовані кількісні і якісні методи та алгоритми прогнозування. Наступні розділи будуть зосереджені на методах та алгоритмах прогнозування за ARIMA моделями.

1.3 Програмні засоби прогнозування

Потреба інтелектуального аналізу даних зростає практично в усіх сферах діяльності. Ринок програмного забезпечення (ПЗ) для статистичної обробки даних нестримно розвивається. Більшість програмних пакетів для інтелектуального аналізу даних містять процедури прогнозування. Тому в підрозділі розглянемо основні присутні на ринку програмні пакети для інтелектуального аналізу даних, які можуть бути використані при прогнозуванні. ПЗ такого виду розробляється багатьма організаціями, як невеликими компаніями, так і великими корпораціями світового рівня.

Ринок пакетів програм для інтелектуального аналізу даних пропонує велику кількість різноманітного ПЗ [12-18].

При виборі важливо звернути увагу на наступні характеристики: відповідність комп'ютерного устаткування користувача системним вимогам ПЗ; відповідність можливостей ПЗ до параметрів поставлених завдань; об'єм даних для статистичного аналізу; кваліфікація (рівень знань) користувача в області статистики.

ПЗ, на думку авторів [19] повинне відповідати певним вимогам: модульність; можливість асистування при виборі способу обробки даних; використання простої проблемно-орієнтованої мови для формулювання завдання користувача; автоматична організація процесу обробки даних та зв'язків з модулями пакета; ведення банку даних користувача і складання звіту про результати зробленого аналізу; діалоговий режим роботи користувача з пакетом; сумісність з іншими програмами.

А також наявна інформаційна система повинна задовольняти відповідним системним вимогам, що показують, яка конфігурація системи здатна забезпечити максимально комфортні умови роботи ПЗ [20]. Ці вимоги включають, як апаратне забезпечення (hardware), так і програмне забезпечення (software - необхідні драйвери, операційна система тощо).

Ринок статистичних пакетів досить великий і досягає декількох тисяч [12, 13, 21]. За функціональністю програми для інтелектуального аналізу можна розділити на 3 основних групи: універсальні пакети, або пакети загального призначення; професійні пакети; спеціалізовані пакети (рисунок 1.9).

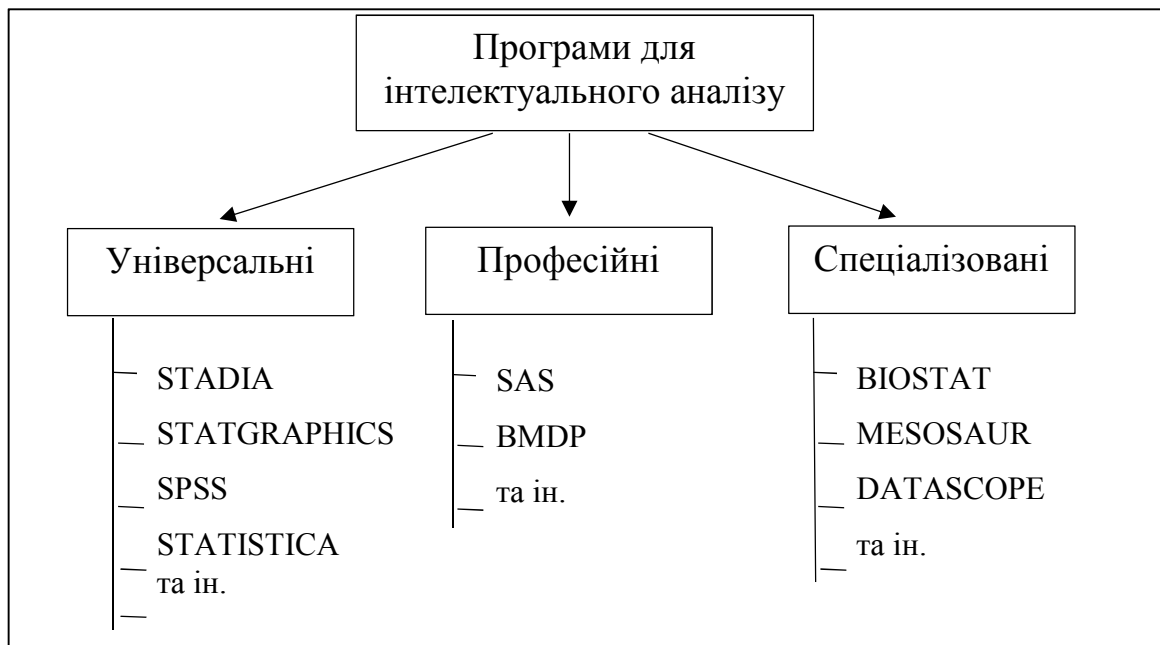


Рисунок 1.9 – Класифікація програм для інтелектуального аналізу

Охарактеризуємо найпопулярніші та функціонально повні програмні продукти з наявними засобами інтелектуального аналізу даних, а саме, які будують ARIMA-моделі і роблять прогнозування на деякий період.

EViews – Econometric Views (Quantitative Micro Software / IHS Markit Ltd.) [22]. Статистичний пакет для аналізу економетричних даних часових рядів, аналізу і моделювання даних, побудови регресійних моделей. EViews є найпотужнішим пакетом для побудови економетричних моделей. EViews може бути успішно використаний для вирішення наступних завдань: аналіз наукової інформації; фінансовий аналіз; макроекономічне прогнозування; моделювання економічних процесів; прогнозування станів ринків тощо. Широкі можливості пакет EViews надає для аналізу даних часових рядів. Функції цього ПЗ: проста лінійна регресія, багатофакторна регресія, нелінійна регресія, симулятивні моделі, ARIMA-моделі. Завдяки стрункій і логічній ідеології побудови

Windows- інтерфейса пакет EViews дуже простий в засвоєнні, містить розвинену підказку (help), що є, по суті, довідником по економетричним методам.

MATLAB (MathWorks, Inc.) – комп'ютерна оболонка для інтерактивних та командних обчислень і візуалізації [23]. Містить пакет прикладних програм для числового аналізу, а також мова програмування, що використовується в даному пакеті. Об'єднує в собі чисельний аналіз, операції з матрицями, сигнальні процеси та графіки в зручному для використання середовищі, де задачі та розв'язки подаються у математичному запису без використання традиційного програмування. Є зручним засобом для роботи з математичними матрицями, малюванням функцій, роботи з алгоритмами, створенням робочих оболонок (user interfaces) з програмами в інших мовах програмування. Спеціалізується на чисельному обчисленні, спеціальні інструментальні засоби працюють з програмним забезпеченням Maple, що робить його повноцінною системою для роботи з алгеброю.

MATLAB надає користувачеві велику кількість функцій для аналізу даних, які покривають майже всі області математики, зокрема: матриці та лінійна алгебра – алгебра матриць, лінійні рівняння, власні значення і вектори, сингулярності, факторизація матриць тощо; многочлени та інтерполяція – корені многочленів, операції над многочленами та їх диференціювання, інтерполяція та екстраполяція кривих; математична статистика та аналіз даних – статистичні функції, статистична регресія, цифрова фільтрація, швидке перетворення Фур'є; обробка даних – набір спеціальних функцій, включаючи побудову графіків, оптимізацію, пошук нулів, чисельне інтегрування; диференційні рівняння – вирішення диференційних і диференційно-алгебраїчних рівнянь, диференційних рівнянь із запізнюванням, рівнянь з обмеженнями, рівнянь в часткових похідних та інше; розріджені матриці – спеціальний клас даних пакету MATLAB, що використовується у спеціалізованих додатках; цілочисельна арифметика – виконання операцій цілочисельної арифметики в середовищі MATLAB. Це професіональний пакет

розв'язування математичних задач різної складності, моделювання, рішення рівнянь, побудова графіків тощо.

МЕЗОЗАВР [21]. Програма МЕЗОЗАВР призначена для аналізу часових рядів. У формі часових рядів можуть виступати різноманітні дані з економіки, демографії, техніки, медицини тощо. У системі МЕЗОЗАВР реалізовані функції: ділова графіка; параметричні тести (описова статистика); непараметричні тести (хі-квадрат, біноміальний, знаків, серій, кореляція Кенделла, кореляція Спірмана); регресійний аналіз (проста регресія, множинна регресія, покрокова регресія, робаста регресія, нелінійна регресія); аналіз часових рядів (згладжування, фільтрація, автокореляція, кроскореляція, спектральний, кросспектральний, модель Бокса-Дженкінса (ARIMA) з експертною підтримкою). Система МЕЗОЗАВР розроблена для персональних комп'ютерів IBM PC/XT/AT і призначена для аналізу часових рядів помірної (до декількох тисяч спостережень) довжини. Вона орієнтована на фахівця середньої кваліфікації, що має достатнє уявлення про методи аналізу часових рядів. Суттєвою відмінністю системи МЕЗОЗАВР від більшості інших систем є те, що в ній кожному значенню часового ряду автоматично зіставляється відповідний момент часу, що враховується при здійсненні будь-яких операцій над часовими рядами, побудові графіків, побудові моделей тощо. Пакет МЕЗОЗАВР має перевагами по введенню та зберіганню інформації: свій стандарт файлів, введення інформації, яке здійснюється через вбудований редактор даних типу «електронної таблиці»; надає можливість збереження в стандартних файлах будь-яких даних, отриманих в ході аналізу; допускає експорт та імпорт інформації з текстових файлів.

Гранична довжина одного аналізованого часового ряду дорівнює 16 тис. значень, однак можливості аналізу такого ряду будуть дуже обмежені і тому найбільш ефективно працювати з рядами до 2-3 тис. значень. Одночасно можна аналізувати до 256 різних змінних, проте їх сумарна кількість значень не повинна перевищувати 60 тис. чисел. Редагування даних здійснюється за допомогою вбудованого табличного редактора або ж у графічному режимі, а їх

перетворення – за допомогою додаткового меню перетворень спеціального характеру. Крім того, є можливість безпосередньо вказати попереднє перетворення даних, яке далі враховується в процесі аналізу. При всіх арифметичних операціях над рядами враховується сумісність часових шкал.

MicroTSP – Time Series Processor (TSP International, Inc.) [21,24]. Функції пакету: проста регресія, регресія часових рядів, прогнозування, авторегресивні моделі та моделі ковзної середньої, одночасні (симультаивні) та імітаційні моделі, використання електронних таблиць у поєднанні з МістоTSP, логістичні та ймовірнісні моделі.

Minitab (Minitab, Inc) [25]. Функції пакету: управління даними і файлами; електронні таблиці для поліпшення аналізу даних; регресійний аналіз – проста регресія, регресія часових рядів; потужність і розмір вибірки; таблиці і діаграми; багатовимірний аналіз, а саме – факторний аналіз, кластерний аналіз, аналіз відповідностей тощо; непараметричний аналіз – різноманітні тести, включаючи вокальний тест, критерій серій, тест Фрідмана тощо; часовий ряд і прогнозування – інструмент, який допомагає виявити тенденції в даних, а також прогнозування майбутніх значень, діаграми часових рядів, експоненціальне згладжування, аналіз тенденцій; статистичне управління процесами; системний аналіз вимірів; дисперсійний аналіз – визначення відмінностей між точками даних.

MS Excel. Найбільш поширений додаток з пакету офісних програм MS Office [26]. MS Excel – це електронна таблиця з досить потужними математичними можливостями, де деякі статистичні функції є просто додатковими вбудованими формулами. Розрахунки зроблені при її допомозі не визнаються авторитетними науковими журналами. MS Excel добре підходить для накопичення даних, проміжного перетворення, попередніх статистичних обчислень, для побудови деяких видів діаграм. Проте остаточний статистичний аналіз необхідно робити в програмах, які спеціально створені для цих цілей. Існують макроси-доповнення для MS Excel, що включають додаткові

статистичні функції, які в основних випадках є достатніми для звичайного застосування. Пробну версію макросів можна узяти на сайті виробника.

RATS – Regression Analysis of Time Series (Estima, Inc.) [27]. Програма для економетричного аналізу часових рядів, широко використовується в учбових цілях. Функції: метод найменших квадратів, ймовірності та логістичні моделі, двокроковий метод найменших квадратів, зважений метод найменших квадратів, інструментальні змінні, критерій Кочрена-Оркатта, поліноміальні розподілені лаги [27, 28].

SPSS – Statistical Package for the Social Sciences, PASW Statistics-Predictive Analytics SoftWare (SPSS: An IBM Company, Inc.) [29]. Потужний, але й дорогий статистичний пакет обробки даних. Один із лідерів ринку в області комерційних статистичних продуктів, призначених для проведення прикладних досліджень в соціальних науках. Найбільш відомий і найстаріший пакет статистичної обробки даних. Модульний, повністю інтегрований, має усі необхідні можливості програмного комплексу, охоплює всі етапи аналітичного процесу: планування, збір даних, доступ до даних і керування даними, аналіз, створення звітів та поширення результатів. Має зручний інтерфейс. Містить досить повний набір статистичних (усього понад 60) та графічних процедур, а також процедур для створення звітів. Має вбудований засіб, що виконує інтелектуальну функцію, наприклад, пояснює користувачеві, яку статистику краще застосувати в кожному конкретному випадку. Можливості: введення і зберігання даних; можливість використання змінних різних типів; частотність ознак, таблиці, графіки, таблиці зв'язаності, діаграми; первинна описова статистика; маркетингові дослідження; аналіз даних маркетингових досліджень. SPSS має якісні прогностичні моделі і різні методи аналізу. Включає широкий набір функцій для простого управління і аналізу великих об'ємів даних. Програма може аналізувати усі надані дані і створювати графіки та діаграми для їх представлення в зручному вигляді. Усі аналізовані дані і результати представляються в окремих діалогових вікнах. Редактор даних – це ще одна відмінна функція IBM SPSS Statistics Base. Редактор включає

багатофункціональне меню для редагування файлів, аналізу ризиків, ідентифікації клієнта, аналізу витрат і прибутку, а також аналіз трендів для кращого планування стратегії організації і майбутніх виробничих процесів. IBM SPSS Statistics Base також здатний виявляти шахрайські дані, щоб мінімізувати ризики. Програма дозволяє проаналізувати, які характеристики клієнти зв'язує з торговою маркою. Дані результатів, які використовуються в цьому інструменті, відображаються разом з графіками і діаграмами, які можна скопіювати і вставити в інші програми, а також експортувати в PDF або DOC формати, роздрукувати або зберегти для подальшого використання. Пакет відрізняється гнучкістю та потужністю, може застосовуватися для усіх видів статистичних розрахунків. Серед основних функцій цього пакету – можливість використання широкого набору математичних методів статистичної обробки даних, формування різних типів звітів, підтримка роботи з більшістю форматів даних, у тому числі підготовлених в електронних таблицях.

STADIA – це вичерпний набір найсучасніших і ефективніших методів аналізу. Надає повний комплект ділової і наукової графіки: функції, залежності, розподіли, діаграми розсіяння, багатовимірні діаграми, карти, поверхні, обертання, сплайни, прогнози, гістограми, стовпчикові, вежі і кругові діаграми, дендрограми, установка розмірів, написів по осях і під малюнком, графічний архів, різноманітні перетворення і обчислення, імпорт/експорт даних і результатів в стандартний міжнародний формат (ASCII і DBF); розвинена екранна допомога, зрозуміла інтерпретація результатів, великий сервіс, до 64 000 чисел в матриці даних. Функції: ділова графіка; параметричні тести (критерії Стьюдента, Фішера, гістограми, розподіли, узгодження частот, послідовний аналіз); непараметричні тести (хі-квадрат, знаків, Вілкоксона, Колмогорова-Смірнова, Ван-дер-Вардена, Клотца, Анасрі-Бредлі, кореляція Кенделла, кореляція Спірмана); категоріальний аналіз (кростабуляція, хі-квадрат оцінка, коефіцієнти узгодження); дисперсійний аналіз (одно-, дво-, багатофакторний, коваріаційний); регресійний аналіз (проста регресія, множинна регресія, покрокова регресія, нелінійна регресія); аналіз часових

рядів (згладжування, фільтрація, автокореляція, кроскореляція, спектральний, крос-спектральний, ARIMA-моделі); багатовимірні методи дискримінантний аналіз, кластерний аналіз, факторний аналіз). Має розвинену систему контекстної екранної допомоги. Включає усі необхідні статистичні функції. Зовнішній інтерфейс фактично не змінюється з 1996 року. Графіки і діаграми, побудовані за допомогою STADIA, виглядають в сучасних презентаціях архаїчно. Колірна гамма програми (червоний шрифт на зеленому) дуже стомлює в роботі. До позитивних якостей програми можна віднести російськомовний інтерфейс і наявність великої кількості книг опису роботи.

Stata (Stata Corp.) [30]. Потужний пакет для обробки статистичних і графічних даних. Пакет Stata позиціонується інструментом аналізу, що призначений для фахівців, які займаються науковими дослідженнями. Завдяки гнучкій модульній структурі пакет можливо застосовувати для аналізу даних із різних областей знань: суспільні науки (економіка, політологія), медицина (біостатистика, епідеміологія) тощо. Stata надає користувачам наступні можливості статистичного аналізу даних: динамічні факторні моделі, лінійні, узагальнені лінійні моделі і нелінійні моделі, багаторівневі змішані моделі, моделі вибуття; узагальнені оцінки рівнянь (GEE); непараметричні методи, методи повторних вибірок і статистичного експерименту; перевірка гіпотез і робота з оціненими моделями; максимізація функцій правдоподібності, заданих користувачем; матричні команди.

STATISTICA (StatSoft, Inc.) – добре збалансоване за співвідношенням «потужність/зручність» ПЗ [31]. Має широкий спектр функціональних алгоритмів і розвинену графіку, а також відповідні засоби для редагування графічних матеріалів. Містить більше 250 статистичних функцій. Користувач має знати статистичну термінологію, а об'ємна довідкова система дає змогу досить повно ознайомлюватися з алгоритмами, що використовуються. Широко розповсюджена. Вбудовані функції об'єднані спеціалізованими статистичними модулями: основні статистики і таблиці, непараметрична статистика, дисперсійний аналіз, множинна регресія, нелінійне оцінювання, аналіз часових

рядів і прогнозування, кластерний аналіз, факторний аналіз, функціональний аналіз, дискримінанта, аналіз тривалості життя, канонічна кореляція, багатовимірні шкали, моделювання структурними рівняннями тощо. Нескладний в засвоєнні, пакет можна рекомендувати для статистичних досліджень будь-якої складності [12, 16]. STATISTICA має суттєві переваги перед іншими статистичними пакетами: за допомогою реалізованих в системі STATISTICA мов програмування (SCL, STATISTICA BASIC), забезпечених спеціальними засобами підтримки, легко створюються закінчені рішення, що вбудовуються в різні інші застосування або обчислювальні середовища. Пакет перекладено російською мовою і можна придбати ліцензійну, русифіковану версію, видана велика кількість книг з детальним описом системи STATISTICA; можливе розширення користувачем бібліотеки функцій, що дозволять вирішувати більшість завдань по теорії вірогідності; реалізовано обмін даними між STATISTICA і Windows додатками; пакет має сенс використати при рішенні досить трудомістких, математично складних і громіздких в реалізації методів багатовимірного аналізу; будь-яка графічна і текстова інформація в STATISTICA може бути виведена у файл формату RTF [12, 21].

SYSTAT (Systat Software, Inc.) [32,33]. Програма для статистичної обробки. Функції: ділова графіка; параметричні тести (критерії Стьюдента, Фішера, гістограма); непараметричні тести (хі-квадрат, знаків, серій, Вілкоксона, Колмогорова-Смірнова, кореляція Спірмана); категоріальний аналіз (кростабуляція, хі-квадрат оцінка, коефіцієнти узгодження); дисперсійний аналіз (одно-, дво-, багатofакторний); регресійний аналіз (проста регресія, множинна регресія, покрокова регресія, нелінійна регресія); аналіз часових рядів (згладжування, фільтрація, автокореляція, кроскореляція, спектральний, ARIMA-моделі); багатовимірні методи (кореляція (коваріація), кластерний аналіз, факторний аналіз); побудова теоретичних моделей для емпіричних отриманих даних, відображення результатів у вигляді тривимірних

поверхонь чи двовимірних кривих; глибокий аналіз і обробка цифрових зображень.

ЕВРИСТА [21]. Функції: ділова графіка; параметричні тести (описова статистика, гістограма); непараметричні тести (хі-квадрат, знаків, Колмогорова-Смірнова); регресійний аналіз (проста регресія, множинна регресія); аналіз часових рядів (згладжування, фільтрація, автокореляція, кроскореляція, спектральний, крос-спектральний, моделі інтервенції, фазовий простір, гармонічні моделі, ARIMA-моделі); багатовимірні методи (факторний аналіз.

Інформаційні технології та їх складові частини (hardware, software) стрімко розвиваються. Висока вартість статистичних програм не дозволяє їх міняти. У своїй роботі для прогнозування прибутків комп'ютерної фірми я використала пакет STATISTICA.

1.4 Постановка задач дипломного проектування

Проаналізувавши розвиток ІТ-бізнесу в Україні, методи та алгоритми прогнозування, встановлено, що для ефективного прийняття рішень необхідний прогноз майбутніх значень часового ряду. Показано, що методами аналізу даних є перевірка статистичних гіпотез, дисперсійний аналіз, кореляційний аналіз, факторний аналіз, методи побудови й дослідження регресійних моделей та ін. Вибрано для аналізу бази даних комп'ютерних фірм кореляційний аналіз, методи і задачі регресійного аналізу, а саме для ARIMA-моделей.

Метою роботи є розробка алгоритмів аналізу прогнозування прибутку комп'ютерних фірм.

Для досягнення поставленої мети необхідно розв'язати такі задачі:

1. Проаналізувати стан розвитку ІТ-бізнесу в Україні.
2. Проаналізувати основні методи та алгоритми прогнозування.
3. Зробити аналіз програмних засобів прогнозування.

4. Розробити алгоритми побудови ARIMA-моделей.
5. Розробити алгоритми проведення кореляційного і регресійного аналізу.
6. Спроекувати і програмно реалізувати комп'ютерну систему моделювання алгоритмів прогнозування за ARIMA-моделями.
7. Розробити модуль ідентифікації моделі.
8. Розробити модуль оцінювання моделі і перевірки адекватності моделі.
9. Розробити модуль прогнозування за моделлю.

2 АЛГОРИТМИ ПОБУДОВИ ARIMA-МОДЕЛЕЙ

2.1 Алгоритми аналізу тренду

Розглянемо моделі часових рядів, що пояснюють поведінку часового ряду виходячи з його значень в попередні моменти часу.

Статистичні властивості стаціонарних і нестаціонарних часових рядів суттєво відрізняються і для їх моделювання повинні застосовуватися різні методи [34-38].

Часовий ряд – це сукупність послідовних вимірювань деякої змінної економічного процесу, проведених через рівні інтервали значень деякого параметру (часу чи просторової координати).

Аналіз часових рядів дає можливість розв'язати такі основні задачі:

- дослідження структури часового ряду, яка як правило відображається закономірною зміною середнього рівня (тренду) та наявністю періодичних коливань;
- побудова математичної моделі економічного процесу, представленого часовим рядом;
- прогнозування майбутнього розвитку економічного процесу;
- дослідження причинно-наслідкових взаємозв'язків між економічними процесами, які проявляються у вигляді різноманітних взаємних кореляцій між часовими рядами;
- перетворення часового ряду методами згладжування та фільтрації.

Значна частина методів аналізу зорієнтована на стаціонарні процеси, статистичні властивості яких не змінюються протягом часу (середнє та дисперсія постійні у випадку нормального розподілу). Однак багато часових рядів мають нестаціонарний характер. Оскільки, більшість часових рядів можна привести до стаціонарних після виділення тренду, сезонної компоненти чи взяття різниці, тому стаціонарним часовим рядам приділяється велика увага.

Розглянемо різні приклади нестаціонарних часових рядів.

Спочатку розглянемо тренд. Нехай часовий ряд можна представити

$$y_t = \alpha + \beta t + \varepsilon_t. \quad (2.1)$$

Тут ряд y_t представлено у вигляді композиції детермінованої складової $\alpha + \beta t$ (лінійний тренд) і випадкової складової ε_t , що є стаціонарним часовим рядом з нульовою середньою. Також зустрічаються інші приклади тренду: квадратичний $\alpha + \beta t + \gamma t^2$, експоненційний $\alpha e^{\beta t}$ і т. п.

Для того щоб виділити тренд в моделі (2.1) (і їй подібних), можна застосувати звичайну техніку оцінювання параметрів регресійних рівнянь, вважаючи t незалежною змінною. Після цього ми отримаємо ряд залишків, для опису якого можна буде застосувати моделі стаціонарних часових рядів.

Розглянемо сезонність. В економічних даних часто зустрічається сезонна компонента. Наприклад, в квартальних даних може спостерігатися сезонна компонента з періодом 4:

$$y_t = S(t) + \varepsilon_t, \quad S(t + 4) = S(t). \quad (2.2)$$

Тут ряд y_t представлений в вигляді композиції періодичної детермінованої складової $S(t)$ (сезонна компонента) і випадкової складової ε_t , що є стаціонарним часовим рядом з нульовою середньою. Сезонну компоненту $S(t)$ можна представити в вигляді $S(t) = \beta_1 d_{1t} + \beta_2 d_{2t} + \beta_3 d_{3t} + \beta_4 d_{4t}$, де d_i – фіктивні (бінарні) змінні для кварталів. Для виділення сезонної компоненти можна застосувати методи оцінювання параметрів регресій до рівняння:

$$y_t = \beta_1 d_{1t} + \beta_2 d_{2t} + \beta_3 d_{3t} + \beta_4 d_{4t} + \varepsilon_t. \quad (2.3)$$

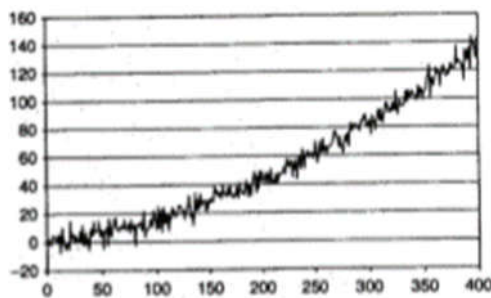
Часто модель (2.3) представляють в вигляді регресії з обмеженням, включаючи в неї константу:

$$y_t = \alpha + \beta_1 d_{1t} + \beta_2 d_{2t} + \beta_3 d_{3t} + \beta_4 d_{4t} + \varepsilon_t, \sum \beta_i = 0. \quad (2.4)$$

В (2.4) коефіцієнти β_i представляють відхилення від середнього за рік рівня в кварталі i .

Як і в випадку виділення тренду, методи моделювання стаціонарних часових рядів застосовуються далі до ряду залишків регресії.

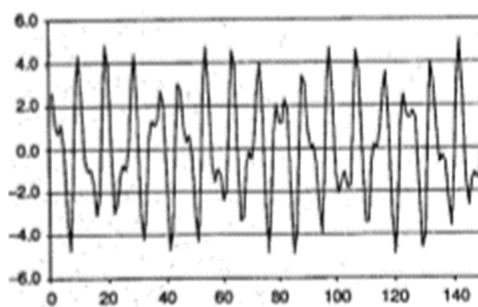
Типові приклади графіків нестационарних часових рядів приведені на рисунку 2.1 [35].



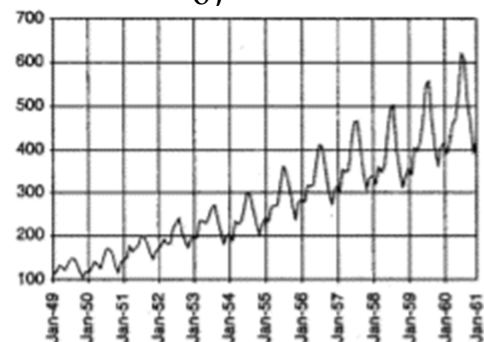
а)



б)



в)



г)

Рисунок 2.1. – Типові приклади графіків нестационарних часових рядів:

а) тренд, б) випадкове блукання, в) сезонність, г) тренд і сезонність

Розглянемо взяття послідовної різниці. Випадковий процес $y_t = y_{t-1} + \varepsilon_t$, $\varepsilon_t \sim iid(0, \sigma^2)$, $t = 1, \dots, n$ є прикладом нестационарного часового ряду. Проте, якщо до нього застосувати операцію взяття послідовної різниці, отримаємо стаціонарний часовий ряд:

$$z_t = \Delta y_t = (1 - L)y_t = y_t - y_{t-1}, z_t = \varepsilon_t$$

Для більш загального процесу $A(L)y_t = \varepsilon_t$, де $A(L)$ має один одиничний корінь, тобто $A(L) = B(L)(1 - L)$, а всі корені $B(L)$ лежать зовні одиничного кола, це ж перетворення $z_t = \Delta y_t$ приводить до стаціонарного процесу $B(L)z_t = \varepsilon_t$.

Взяття різниці також приводить до стаціонарного процесу ряд (2.1) з лінійним трендом:

$$\Delta y_t = \beta + u_t, u_t = \Delta \varepsilon_t = \varepsilon_t - \varepsilon_{t-1}. \quad (2.5)$$

В випадку квадратичного тренду $\alpha + \beta t + \gamma t^2$ взяття першої різниці не приводить до стаціонарного ряду, але якщо взяти другу різницю $\Delta^2 y_t = \Delta(\Delta y_t) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$, то

$$\Delta y_t = \beta + \gamma(2t - 1) + \Delta \varepsilon_t,$$

$$\Delta^2 y_t = 2\gamma + \Delta^2 \varepsilon_t,$$

і $\Delta^2 y_t$ вже є стаціонарним часовим рядом.

В випадку наявності сезонної компоненти (2.2) усунути останню можна з допомогою оператора взяття сезонної послідовної різниці $\Delta_4 y_t = (1 - L^4)y_t =$

$= y_t - y_{t-4}$. Якщо період сезонної компоненти дорівнює 12 (це може бути для місячних даних), то треба застосувати оператор Δ_{12} .

Слід відмітити, що застосування оператора послідовної різниці не обов'язково приводить нестационарний ряд до стаціонарного.

Таким чином, застосовуючи виділення тренду, сезонності і / або оператор послідовної (і сезонної) різниці, часто можна отримати з вихідного часового ряду стаціонарний.

Яким чином визначають за наявними спостереженнями, чи є ряд стаціонарним [39,40].

Перше – подивитись на графік отриманих спостережень. Можливо, він містить очевидний на око тренд або періодичну компоненту (сезонність). Також можливо, що розкид спостережень зростає чи спадає з часом. Це може служити вказівкою на залежність середньої або відповідно дисперсії від часу. В обох випадках ряд буде, найшвидше, нестационарний.

Друге – побудувати графік вибіркової автокореляційної функції (ACF) або корелограми (correlogram):

$$r_k = \hat{\rho}_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^n (y_t - \bar{y})^2}, \quad k = 1, 2, \dots \quad (2.6)$$

Корелограма стаціонарного часового ряду «швидко спадає» з ростом k після декількох перших значень. Якщо ж графік спадає дуже повільно, то є основи припустити нестационарність ряду. Крім ACF, можна також побудувати графік часткової автокореляційної функції PACF, яка теж повинна швидко спадати для стаціонарного процесу.

В загальному випадку, коефіцієнт часткової кореляції між y та x_1 при виключенні впливу x_2 визначається за формулою:

$$r(y, x_1 | x_2) = \frac{r(y, x_1) - r(y, x_2)r(x_1, x_2)}{\sqrt{1 - r^2(y, x_2)}\sqrt{1 - r^2(x_1, x_2)}}. \quad (2.7)$$

Значення його лежить в інтервалі $[-1; 1]$, як у звичайного коефіцієнта кореляції.

Часткова автокореляційна функція $PACF(k)$ є «чиста кореляція» між y_t і y_{t-k} при виключенні впливу проміжних значень $y_{t-1}, y_{t-2}, \dots, y_{t-k-1}$.

Якщо застосувати процедуру обчислення вибіркового часткового коефіцієнта кореляції, то виявляється, що у випадку стаціонарного ряду y_t значення вибіркової часткової автокореляційної функції $PACF(k)$ обчислюється як МНК-оцінка останнього коефіцієнта β_k в $AR(k)$ регресійному рівнянні:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_k y_{t-k} + \varepsilon_t. \quad (2.8)$$

Третє – можна використовувати формальні тести на наявність одиничного кореня (тест Дікі – Фуллера DF, розширений тест Дікі – Фуллера ADF, тест МакКінлі і ін.).

Таким чином, рисунок 2.2 підсумовує аналіз тренду.

2.2 Алгоритми оцінювання моделі за методологією Бокса-Дженкінса

Розглянемо спочатку частковий випадок моделі Бокса-Дженкінса – ARMA моделі для стаціонарних часових рядів.

Нехай є наступний клас моделей стаціонарних рядів:

$$y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (2.9)$$

$$\varepsilon_t \sim iid(0, \sigma^2),$$

або в короткій формі

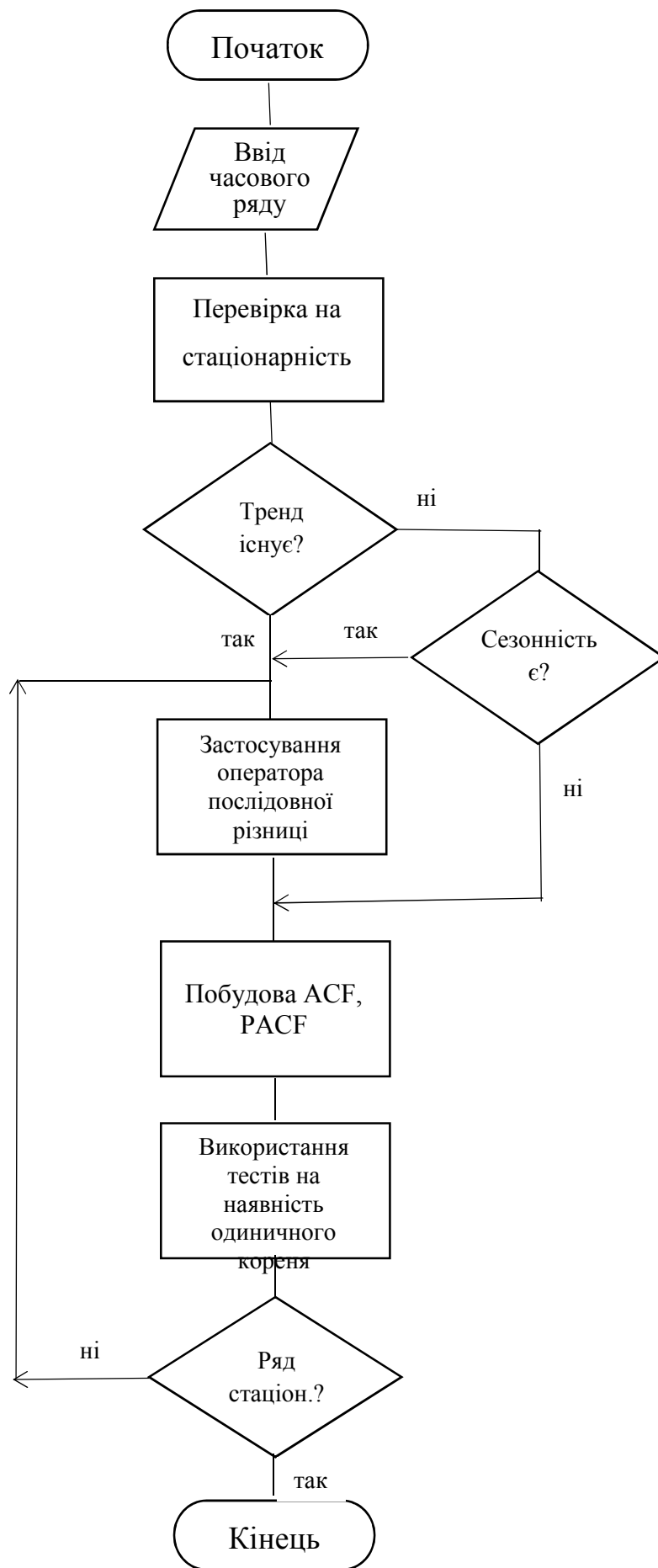


Рисунок 2.2 – Схема алгоритму аналізу тренду

$$\Phi(L)y_t = \delta + \Theta(L)\varepsilon_t, \varepsilon_t \square iid(0, \sigma^2), \quad (2.10)$$

де $\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ і $\Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$ – поліноми від оператора зсуву. Така модель називається моделлю авторегресії і ковзної середньої – ARMA(p, q).

Розглянемо прості ARMA моделі.

Процес AR(1) або ARMA(1,0) представляється формулою:

$$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t, \varepsilon_t \square iid(0, \sigma^2). \quad (2.11)$$

Його числові характеристики

$$E(y_t) = \frac{\delta}{1 - \phi_1}, V(y_t) = \gamma_0 = \frac{\sigma^2}{1 - \phi_1^2}, \gamma_k = \phi_1^k \gamma_0, \rho_k = \gamma_k / \gamma_0 = \phi_1^k \quad (2.12)$$

Необхідною умовою стаціонарності процесу y_t є умова $|\phi_1| < 1$.

Часткова автокореляційна функція AR(1) процесу дорівнює нулю для значень $k > 1$.

Розглянемо авторегресійний процес AR(2) або ARMA(2,0):

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \varepsilon_t \square iid(0, \sigma^2). \quad (2.13)$$

Нехай вільний член рівняння дорівнює нулю:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \varepsilon_t \square iid(0, \sigma^2). \quad (2.14)$$

Для $k > 0$ обчислимо коваріацію обох частин (2.14) з y_{t-k} :

$$\gamma_k = \text{Cov}(y_t, y_{t-k}) = \text{Cov}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, y_{t-k}) = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} \quad (2.15)$$

Розділимо на γ_0 і отримаємо

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k = 1, 2, \dots \quad (2.16)$$

Формула (2.16) при $k = 1, 2$ буде наступною системою рівнянь з невідомими ρ_1 і ρ_2 (врахувавши, що $\rho_0 = 1$, $\rho_{-1} = \rho_1$):

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1, \\ \rho_2 &= \phi_1 \rho_1 + \phi_2. \end{aligned} \quad (2.17)$$

(2.17) – це система рівнянь Юла-Уолкера для процесу AR(2). Розв'язуючи систему (2.17), отримаємо перші два значення автокореляційної функції:

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}, \quad \rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2. \quad (2.18)$$

Всі наступні значення вираховуються за (2.16). Формула дисперсії отримується наступним чином: множиться (2.14) на y_t і береться математичне сподівання, тоді отримуємо:

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2. \quad (2.19)$$

Розв'язавши це рівняння з (2.15) для $k = 1, 2$, отримаємо:

$$\gamma_0 = \frac{(1 - \phi_2)\sigma^2}{(1 + \phi_2)(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)}. \quad (2.20)$$

Тоді, умови стаціонарності AR(2) процесу:

$$|\phi_2| < 1, \phi_2 + \phi_1 < 1, \phi_2 - \phi_1 < 1. \quad (2.21)$$

Якщо корені характеристичного полінома $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2$ дійсні, то автокореляційна функція спадає експоненційно, а якщо корені – комплексні, то змінюється по синусоїді з експоненційно спадною амплітудою.

Схема обчислення часткової автокореляційної функції для процесу AR(2) наступна. Записуємо три рівняння Юла-Уолкера (див. (2.17)) для AR(3) процесу. Коефіцієнт ϕ_3 дорівнює коефіцієнту часткової кореляції між y_t і y_{t-3} . З (2.16) для AR(2) отримуємо $\rho_3 = \phi_1 \rho_2 + \phi_2 \rho_1$. Якщо підставити цей вираз в третє рівняння Юла-Уолкера, отримаємо $\phi_3 = 0$. Отже, $PACF(k) = 0$ для $k > 2$. Аналогічно, можна показати, що для AR(p) процесу часткова автокореляційна функція $PACF(k) = 0$, починаючи з $k = p + 1$. Необхідно наголосити, що цей результат вірний для теоретичної часткової автокореляційної функції і не завжди вірний для вибіркової часткової автокореляційної функції. Проте, на практиці відбувається різке спадання $PACF$ до значень близьких до нуля після k , що дорівнює порядку авторегресійного процесу.

Розглянемо процеси ковзної середньої. Моделлю ковзної середньої (МА) порядку q називається модель ARMA(0, q)

$$y_t = \delta + \Theta(L)\varepsilon_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \varepsilon_t \square iid(0, \sigma^2). \quad (2.22)$$

яка позначається MA(q). З (2.22) видно, що процес MA(q) стаціонарний при будь-якому q і будь-яких θ_i .

Розглянемо MA(1)

$$y_t = \delta + \Theta(L)\varepsilon_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1}, \varepsilon_t \square iid(0, \sigma^2). \quad (2.23)$$

МА(1) в вигляді авторегресійного процесу:

$$\Theta(L)^{-1} y_t = \Theta(L)^{-1} \delta + \varepsilon_t \quad (2.24)$$

або

$$y_t = \frac{\delta}{1 - \theta_1} - \theta_1 y_{t-1} - \theta_1^2 y_{t-2} - \dots + \varepsilon_t. \quad (2.25)$$

Таке $AR(\infty)$ представлення МА(1) процесу (2.23) можливе тільки у випадку оборотності оператора $\Theta(L) = 1 - \theta L$, тобто, коли виконується умова оборотності $|\theta| < 1$.

Обчислимо середню і дисперсію процесу МА(1):

$$E(y_t) = \delta, \quad V(y_t) = \sigma^2(1 + \theta_1^2). \quad (2.26)$$

Знайдемо автокореляційну функцію процесу МА(1):

$$\gamma_1 = Cov(y_t, y_{t-1}) = -\theta_1 \sigma^2. \quad (2.27)$$

Автокореляційна функція $ACF(k) = 0$ для $k > q$, тобто її вигляд аналогічний вигляду $PACF$ для $AR(q)$ процесу.

Часткова автокореляційна функція $PACF(k)$ для МА(q) процесу, аналогічно $ACF(k)$ для $AR(q)$ процесу, експоненційно спадає. Таким чином, має місце деяка симетрія: пара графіків ($ACF, PACF$) для МА(q) процесу мають такий же вигляд, як пара графіків ($PACF, ACF$) для $AR(q)$.

Слід відмітити, що подібно $AR(\infty)$ представленню (2.25) для МА(1) процесу (2.23) існує МА(∞) представлення для $AR(1)$ процесу (2.11):

$$y_t = (1 - \phi_1 L)^{-1}(\delta + \varepsilon_t) = \frac{\delta}{1 - \phi_1} + \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \dots \quad (2.28)$$

Розглянемо найпростіший змішаний $ARMA(1,1)$ процес (2.10) з $\Phi(L) = 1 - \phi_1 L$, і $\Theta(L) = 1 - \theta_1 L$:

$$y_t - \phi_1 y_{t-1} = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1}, \quad \varepsilon_t \square iid(0, \sigma^2). \quad (2.29)$$

Будем вважати, що $|\phi_1| < 1$ і $|\theta_1| < 1$. Як і у випадку $AR(1)$ і $MA(1)$ процесів, можна показати, що тоді процес $ARMA(1,1)$ є стаціонарним і зворотним.

Застосовуючи ті ж методи, що і раніше, отримаємо наступні вирази для середньої, дисперсії і коваріації $ARMA(1,1)$ процесу:

$$E(y_t) = \frac{\delta}{1 - \phi_1} \quad (2.30)$$

$$\gamma_0 = V(y_t) = \sigma^2 \frac{1 + \theta_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2} \quad (2.31)$$

$$\gamma_1 = Cov(y_t, y_{t-1}) = \phi_1 \gamma_0 - \theta_1 \sigma^2 \quad (2.32)$$

Для автокореляцій порядку більше 1 отримуємо рекурентне співвідношення

$$\gamma_k = Cov(y_t, y_{t-k}) = \phi_1 \gamma_{k-1}, \quad \rho_k = \phi_1^{k-1} \rho_1, \quad k > 1, \quad \rho_1 = \frac{(1 - \phi_1 \theta_1)(\phi_1 - \theta_1)}{1 + \theta_1^2 - 2\phi_1 \theta_1}. \quad (2.33)$$

З (2.33) видно, що ACF для $ARMA(1,1)$ процесу веде себе так як ACF для $AR(1)$ процесу. Хоча значення ρ_1 інше, але співвідношення між ρ_1 і наступними значеннями ACF точно такі ж.

Цей висновок можна узагальнити на випадок $ARMA(p, q)$ процесу. Перші q значень ACF визначаються взаємодією AR і MA компонент, а подальша поведінка автокореляційної функції така ж, як в $AR(p)$ процесі.

Аналогічний висновок справедливий для часткової автокореляційної функції $ARMA(p, q)$ процесу. Вона спадає подібно $PACF$ для $MA(q)$ процесу.

Узагальнимо вище приведені результати у вигляді таблиць 2.1.

Таблиця 2.1 – Поведінка автокореляційної і часткової автокореляційної функцій для різних процесів $ARMA$

Процеси	Функції	
	Автокореляційна функція (корелограма)	Часткова автокореляційна функція
$AR(1)$	Експоненційно спадає	$PACF(k) = 0$ при $k > 1$
$AR(2)$	Експоненційно спадає	$PACF(k) = 0$ при $k > 2$
$AR(p)$	Експоненційно спадає	$PACF(k) = 0$ при $k > p$
$MA(1)$	$ACF(k) = 0$ при $k > 1$	Експоненційно спадає
$MA(2)$	$ACF(k) = 0$ при $k > 2$	Експоненційно спадає
$MA(q)$	$ACF(k) = 0$ при $k > q$	Експоненційно спадає
$ARMA(p, q)$	Експоненційно спадає при $k > q$	Експоненційно спадає при $k > p$

2.3 Алгоритми методології Бокса-Дженкінса підбору *ARIMA* моделі

В попередніх підрозділах розглянуто алгоритми, за якими аналізуються часові ряди, проходить перевірка їх стаціонарності. Для вибору найкращої *ARIMA*-моделі переважно застосовують алгоритм за методом Бокса-Дженкінса [41-43] (рисунок 2.3). Він передбачає побудову *ARIMA*-моделі на основі наявного динамічного ряду. В деяких випадках навіть акцентують увагу на тому, що чим більша довжина динамічного ряду, тим буде кращою модель.



Рисунок 2.3 – Алгоритм стратегії вибору моделі за методом Бокса-Дженкінса

Деякі нестационарні часові ряди можуть бути приведені до стаціонарних з допомогою оператора послідовної різниці. Нехай, часовий ряд y_t після застосування d раз оператора послідовної різниці, став стаціонарним рядом $\Delta^d y_t$, що задовольняє *ARMA*(p, q) моделі (2.10). Тоді процес y_t називається інтегрованим процесом авторегресії і ковзної середньої *ARIMA*(p, d, q). Слід

відмітити, що з моделі, наприклад, для ряду Δy_t легко отримати модель для вихідного ряду y_t , використовуючи співвідношення $y_t = y_{t-1} + \Delta y_t$.

Методологія Бокса-Дженкінса підбору *ARIMA* моделі для даного ряду спостережень складається з трьох етапів.

1. Ідентифікація моделі

1.1. Перший крок – отримання стаціонарного ряду. Тестується ряд на стаціонарність, використовуючи описані вище методи: візуальний аналіз графіку, візуальний аналіз *ACF* і *PACF*, тести на одиничні корені. Якщо отримується стаціонарний ряд, то переходимо до наступного пункту, якщо ні – застосовуємо оператор взяття послідовної різниці і повторяємо тестування. На практиці послідовна різниця береться, як правило, не більше двох разів.

1.2. Після того як отримано стаціонарний часовий ряд, будуються його вибіркві *ACF* і *PACF*, які дозволяють сформулювати декілька гіпотез про можливі порядки авторегресії p і ковзної середньої q . Зазвичай, використовуються моделі більш низьких порядків, як правило, з $p + q \leq 3$ (якщо немає сезонної компоненти).

Вибіркові *ACF* і *PACF*, зазвичай, не співпадають з теоретичними аналогами, але достатньо близькі до них.

2. Оцінювання моделі і перевірка адекватності моделі

2.1. Для кожної з вибраних на першому етапі моделей оцінюються їх параметри і обчислюються залишки.

2.2. Кожна з моделей перевіряється, наскільки вона відповідає даним. З моделей, адекватних даним, вибирається найпростіша (з меншою кількістю параметрів).

3. Прогнозування

Після того, як на другому етапі вибрана модель, можна будувати прогноз на один або декілька кроків за часом і оцінювати довірчі межі прогнозних значень.

Розглянемо оцінювання $ARMA$ моделей. В сучасних комп'ютерних пакетах є різні методи оцінювання $ARMA$ моделей, такі, як лінійний або нелінійний МНК, повний або умовний метод максимальної правдоподібності.

Розглянемо приклад $ARMA(1,1)$ моделі (2.29). Напишемо її в вигляді:

$$\Theta(L)^{-1}y_t = \Theta(L)^{-1}(\delta + \phi_1 y_{t-1}) + \varepsilon_t, \quad (2.34)$$

де $\Theta(L) = 1 - \theta_1 L$ і $\Theta(L)^{-1} = 1 + \theta_1 L + \theta_1^2 L^2 + \dots$.

Для оцінювання рівняння (2.34) застосуємо умовний метод максимальної правдоподібності, коли y_1 є заданим, вважаючи, що помилки $\varepsilon_t \square iidN(0, \sigma^2)$.

Умовна функція правдоподібності дорівнює:

$$\begin{aligned} L^* &= p(y_2^*, y_3^*, \dots, y_n^* | y_1^*) = \prod_{t=2}^n p(y_t^* | y_{t-1}^*) = \\ &= \prod_{t=2}^n \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2} (y_t^* - \delta^* - \phi_1 y_{t-1}^*)^2\right). \end{aligned} \quad (2.35)$$

Логарифм умовної функції максимальної правдоподібності дорівнює

$$l^* = \ln L^* = const - \frac{n-1}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=2}^n (y_t^* - \delta^* - \phi_1 y_{t-1}^*)^2. \quad (2.36)$$

З вигляду l^* видно, що оцінка коефіцієнтів δ , ϕ_1 за умовним методом максимальної правдоподібності співпадає з оцінкою нелінійного методу найменших квадратів. Сума в правій частині (2.36) є нелінійною функцією параметрів δ , ϕ_1 .

Повний метод максимальної правдоподібності полягає в максимізації функції правдоподібності

$$L = p(y_1)L^*.$$

Відомо, що при гіпотезі нормальності помилок $y_1^* \square N(\delta^* / (1 - \phi_1), \sigma^2 / (1 - \phi_1^2))$. Тому логарифм функції правдоподібності дорівнює:

$$\begin{aligned} l(\delta, \phi_1, \theta_1, \sigma^2) &= \ln p(y_1) + l^* = \\ &= \text{const} - \frac{n}{2} \ln \sigma^2 + \frac{1}{2} \ln(1 - \phi_1^2) - \frac{1 - \phi_1^2}{2\sigma^2} \left(y_1^* - \frac{\delta^*}{1 - \phi_1} \right)^2 - \\ &\quad - \frac{1}{2\sigma^2} \sum_{t=2}^n (y_t^* - \delta^* - \phi_1 y_{t-1}^*)^2. \end{aligned} \quad (2.37)$$

Є декілька критеріїв оцінки того, наскільки *ARMA* модель, яку ми оцінюємо, відповідає нашим даним, тобто перевірки адекватності (рисунок 2.4).

По-перше, оцінки коефіцієнтів моделі повинні статистично достовірно відрізнятися від нуля, тобто відповідні *P*-значення *t*-статистик мають бути меншими вибраного порогового значення.

По-друге, згідно моделі помилки ε_t є білим шумом. Відповідно їх оцінки, тобто залишки регресії e_t , мають бути подібними на білий шум. Тому залишки повинні мати нульову автокореляцію.

В моделі, що містить константу, середня залишків дорівнює нулю. Тому вибіркова автокореляційна функція залишків обчислюється за формулою:

$$r_k = \frac{\sum_{t=k+1}^n e_t e_{t-k}}{\sum_{t=1}^n e_t^2}, \quad k = 1, 2, \dots \quad (2.38)$$

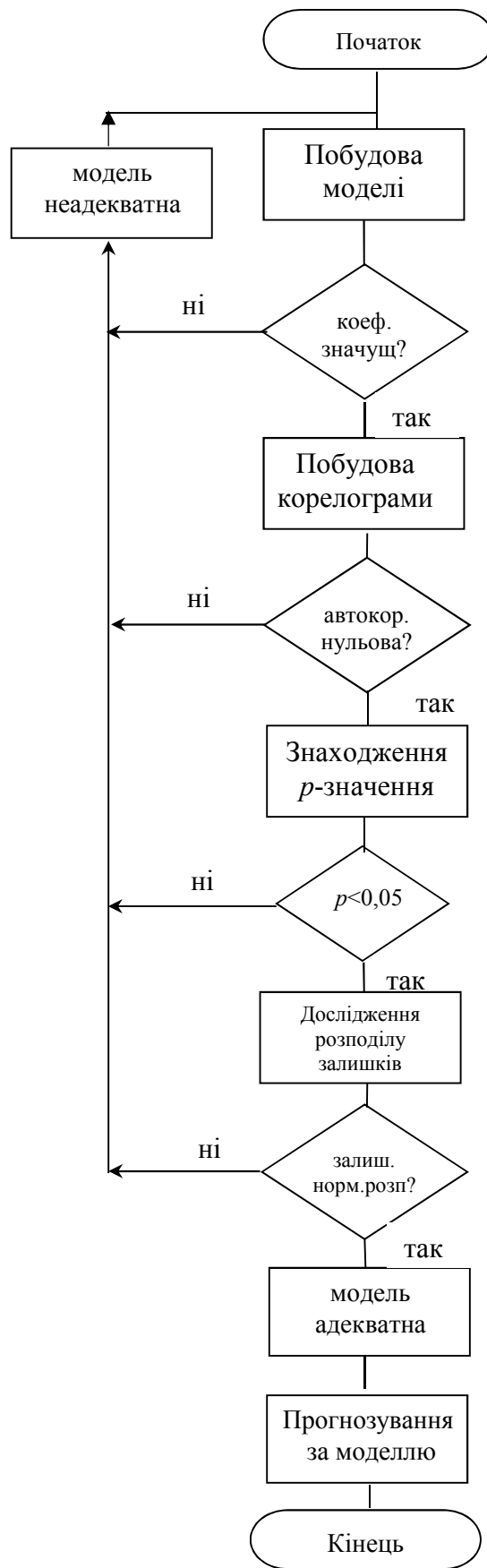


Рисунок 2.4 – Блок-схема алгоритму перевірки моделі на адекватність

Якщо модель адекватна даним, тоді помилки є білим шумом, і при великих значеннях n і k величина r_k має розподіл, близький до нормального $N(0, \frac{1}{n})$.

Причому на практиці хороша апроксимація починається з $k = 5 \div 6$. Тому значення r_k зовні інтервалу $0 \pm \frac{2}{\sqrt{n}}$ дозволяє на 5%-му рівні значущості відхилити гіпотезу рівності нулю коефіцієнта кореляції ρ_k .

Інші тести перевіряють гіпотезу рівності нулю зразу K перших значень автокореляційної функції залишків.

Q -статистика Бокса-Пірса визначається як

$$Q = n \sum_{k=1}^K r_k^2. \quad (2.39)$$

При нульовій гіпотезі відсутності автокореляції Q має розподіл $\chi^2(K - p - q)$, де p, q – параметри $ARMA$ моделі. Нульова гіпотеза відхиляється, якщо отримане значення Q є більшим відповідного критичного значення.

Тест Льюнга-Бокса є модифікацією тесту Бокса-Пірса. Визначається як

$$\tilde{Q} = n(n+2) \sum_{k=1}^K \frac{r_k^2}{n-k} \quad (2.40)$$

і має такий асимптотичний розподіл, як і Q . Проте розподіл статистики ближчий до χ^2 для скінчених вибірок.

Якщо тести показують наявність автокореляції залишків, це означає, що розглядувана $ARMA$ модель не підходить, і її треба модифікувати. Якщо, наприклад, в автокореляційній функції відмінні від нуля значення з номерами, кратними 4, то треба поспробувати ввести сезонну авторегресію четвертого

порядку. Якщо єдине відмінне від нуля значення відповідає лагу, що дорівнює 4, можна поспробувати ввести сезонний MA -член порядку 4.

Якщо ми маємо ситуацію, коли декілька $ARMA$ моделей є адекватними даним, то треба вибрати модель з найменшою кількістю параметрів.

В комп'ютерних пакетах серед результатів оцінювання приводиться інформаційний критерій Акайке (AIC), що визначається формулою:

$$AIC = 2 \frac{p + q}{n} + \ln \left(\frac{\sum_{t=1}^n e_t^2}{n} \right), \quad (2.41)$$

Критерій Акайке є евристичною спробою звести в один показник дві вимоги: зменшення числа параметрів моделі і якість підгонки моделі. Згідно цього критерію, з двох моделей треба вибрати модель з меншим значенням AIC.

Зазвичай також приводиться значення критерію Шварца (BIC):

$$BIC = \frac{(p + q) \ln n}{n} + \ln \left(\frac{\sum_{t=1}^n e_t^2}{n} \right), \quad (2.42)$$

відмінність якого від AIC полягає в великому штрафі за кількість параметрів.

За своєю ідеологією критерії Акайке і Шварца близькі до скорегованого R^2 .

Головна мета використання $ARIMA$ моделей – побудова прогнозу за межами вибірки. Є два джерела неточності прогнозу: перший – ігнорування майбутніх помилок ε_t , другий – відхилення оцінок коефіцієнтів моделі від їх істинних значень. Розглянемо перше джерело або, іншими словами, прогнозування в рамках теоретичних моделей.

Розглянемо проблему прогнозування на прикладі $ARMA(1,1)$ і $ARIMA(1,1,0)$ моделей.

З (2.29) отримуємо значення y в момент $n + 1$:

$$y_{n+1} = \delta + \phi_1 y_n + \varepsilon_{n+1} - \theta_1 \varepsilon_n.$$

Використовуючи позначення $\mu = E(y_t) = \delta / (1 - \phi_1)$, отримуємо:

$$(y_{n+1} - \mu) = \phi_1 (y_n - \mu) + \varepsilon_{n+1} - \theta_1 \varepsilon_n \quad (2.43)$$

Прогноз на один крок, що мінімізує середньоквадратичне відхилення, дорівнює $\hat{y}_{n+1} = E(y_{n+1} | I_n)$, де I_n – інформація, що доступна в момент n :

$$(\hat{y}_{n+1} - \mu) = \phi_1 (y_n - \mu) - \theta_1 \varepsilon_n \quad (2.44)$$

Помилка прогнозу і її дисперсія рівні

$$e_{n+1} = y_{n+1} - \hat{y}_{n+1} = \varepsilon_{n+1}, \quad V(e_{n+1}) = \sigma^2. \quad (2.45)$$

Використовуючи дві ітерації рівняння (2.43), аналогічно отримуємо прогноз на два кроки:

$$\begin{aligned} (\hat{y}_{n+2} - \mu) &= \phi_1^2 (y_n - \mu) - \phi_1 \theta_1 \varepsilon_n, \\ V(e_{n+2}) &= \sigma^2 (1 + (\phi_1 - \theta_1)^2). \end{aligned} \quad (2.46)$$

Продовжуючи ітерації, можна отримати прогноз на s кроків і його дисперсію:

$$(\hat{y}_{n+s} - \mu) = \phi_1^s (y_n - \mu) - \phi_1^{s-1} \theta_1 \varepsilon_n, \quad (2.47)$$

$$\lim_{s \rightarrow \infty} V(e_{n+s}) = \sigma^2 \left(\frac{1 - 2\phi_1 \theta_1 + \theta_1^2}{1 - \phi_1^2} \right). \quad (2.48)$$

Прогноз прямує до середнього μ , при зростанні s .

Прогноз нестационарного часового ряду дещо відрізняється. Розглянемо часовий ряд y_t , перші різниці якого $z_t \in AR(1)$ процесом:

$$z_t = y_t - y_{t-1}, \quad z_t - \mu = \phi_1(z_{t-1} - \mu) + \varepsilon_t. \quad (2.49)$$

Багатократне застосування (2.49) дасть

$$y_{n+s} = y_n + s\mu + \frac{\phi_1(1-\phi_1^s)}{1-\phi_1}(y_n - y_{n-1} - \mu) + e_{n+s}, \quad (2.50)$$

де

$$e_{n+s} = \varepsilon_{n+s} + (1 + \phi_1)\varepsilon_{n+s-1} + \dots + (1 + \phi_1 + \phi_1^2 + \dots + \phi_1^{s-1})\varepsilon_{n+1}. \quad (2.51)$$

Як бачимо, що у випадку нестационарного часового ряду дисперсія помилки прогнозу монотонно росте з ростом s .

Всі обчислення були проведені для теоретичної моделі, тобто в припущенні, що коефіцієнти моделі відомі точно. Зазвичай на практиці ми маємо справу з оцінками коефіцієнтів, що додає додаткову невизначеність в прогноз.

Таким чином, було проаналізовано алгоритм вибору моделі за методом Бокса-Дженкінса. В цьому підрозділі приведено алгоритм перевірки моделі на адекватність та створення прогнозів.

3 КОМП'ЮТЕРНЕ МОДЕЛЮВАННЯ АЛГОРИТМІВ ПРОГНОЗУВАННЯ ЗА ARIMA-МОДЕЛЯМИ

3.1 Програмно-апаратні засоби

Для побудови алгоритмів прогнозування за ARIMA-моделями необхідно правильно підібрати програмно-апаратний комплекс для швидкої та безперебійної роботи системи. Характеристики апаратного забезпечення наведено в таблиці 3.1.

Таблиця 3.1 – Характеристики апаратного забезпечення

Система	Характеристики
Виробник	Dell Inc.
Модель	Latitude
Процесор	Intel(R) Core(TM) i5-2520M CPU 2,50GHz
Оперативна пам'ять	4,00 ГБ
Тип системи	64-розрядна ОС на базі процесора x64
Операційна система	Windows 10 Pro

Програмним засобом для проведення експерименту є Matlab. Система Matlab – математична лабораторія, представляє собою інтегроване програмне середовище для виконання числових розрахунків, комп'ютерного моделювання і обчислювальних експериментів, які в тій чи іншій мірі охоплюють області класичної та сучасної математики, спектр інженерних додатків. Це комп'ютерна програма для автоматичного аналізу великих наборів даних, пошуку в них закономірностей, для автоматичного складання прогнозів.

В таблиці 3.2 наведено основну інформацію про використовуваний програмний засіб Matlab.

Таблиця 3.2 – Характеристики програмного засобу Matlab

Програмне забезпечення	Характеристики
Розробник	The MathWorks
Версія	R2015a (8.5.0.197613)
Тип ПЗ	64-bit (win64)
Зчитувані формати файлів	Файл MATLAB-M
Створювані формати файлів	Файл MATLAB-M

Пакет STATISTICA – це програмний комплекс, призначений для проведення статистичного аналізу і володіє широким набором функцій. Серед них – тисячі типів графіків, аналіз даних, кореляційний аналіз, використання Т-критеріїв (і не тільки їх), непараметричні статистики, коваріаційний і дисперсійний аналіз, застосування множинної регресії, таблиць частот, і ін. [14,16].

Переваги пакету:

- широкий вибір різноманітних інструментів для статистичного аналізу даних;
- дозволяє обробити масивні обсяги даних;
- можливість проведення кластерного, дискримінантного, канонічного, дисперсійного і факторного аналізу;
- наявність функції угруповання даних;
- можливість побудови рядів, лінійних і нелінійних моделей;
- великий набір інструментів для дослідження кореляції між різними змінними;
- підтримка нейронних мереж;
- можливість імпортування даних з Excel-документів;
- наявність блоку інструментів data mining;
- можливість побудови діаграм і 2D / 3D графіків (більше 10000 типів);
- можливість категоризації по змінним (ієрархічна угруповання);

- якісна технічна підтримка;
- інтеграція з Visual Basic (можна використовувати DLL-бібліотеки) і ін.

3.2 Структура програмної системи

В даному підрозділі на базі алгоритмів аналізу тренду, вибору і побудови ARIMA-моделей, прогнозування за моделями розроблена програмна система прогнозування прибутків комп'ютерної фірми (рисунок 3.1).

Відповідно до цієї схеми часові ряди прибутків комп'ютерної фірми подаються для отримання інформації по часових даних прибутків. Дані перевіряються на стаціонарність і при потребі трансформуються в стаціонарний застосуванням оператора послідовної різниці в блоці перетворення стаціонарності. В результаті цього на виході блоку буде сформований стаціонарний ряд. Далі вибірка оброблених статистичних даних подається на вхід блоку вибору моделі в результаті чого маємо набір моделей (2.9).

Для того, щоб переконатися, яку модель вартніше використати для прогнозування, переходимо до блоку перевірки адекватності моделі або перевірки гіпотези, що оцінки коефіцієнтів моделі статистично достовірно відрізняються від нуля або гіпотези відсутності автокореляції залишків. На виході цього блоку отримуємо рішення про адекватність моделі. Якщо нульова гіпотеза відхиляється, а була прийнята альтернативна гіпотеза з певним рівнем значущості, то переходимо до блоку прогнозування за моделлю. В результаті отримуємо прогнозні дані.

Особливості розробленої ІС алгоритмів прогнозування за ARIMA-моделями наступні:

- 1) блок перетворення в стаціонарний ряд включає запропонований в підрозділі 2.1 алгоритм аналізу тренду (рисунок 2.2), перевірку наявності тренду,

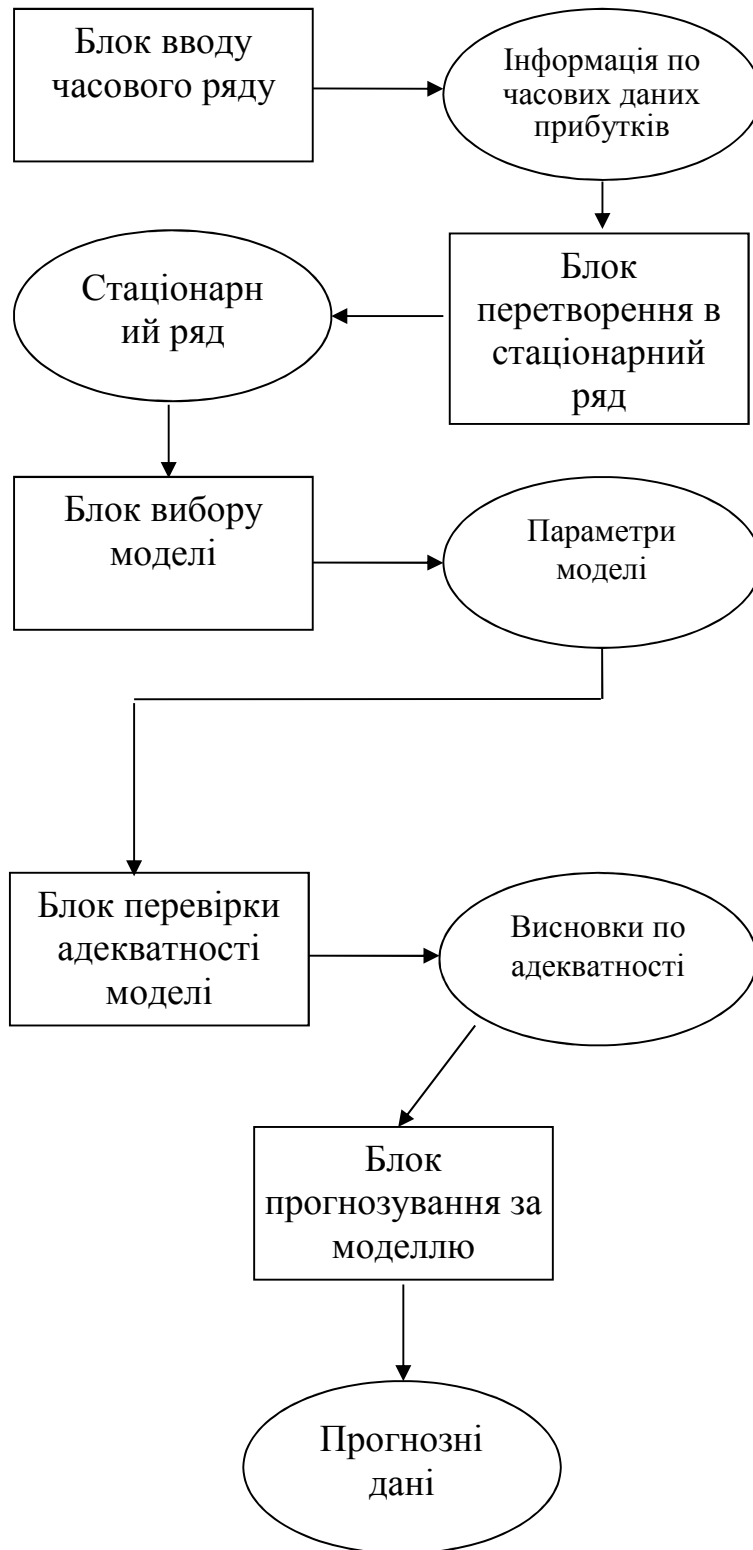


Рисунок 3.1 – Програмна система прогнозування

наявності сезонності, застосування оператора послідовної різниці, побудову ACF, PACF, таблицю 2.1, за допомогою якої можна визначити порядок MA і AR , використання тестів на наявність одиничного кореня. Кожен з цих алгоритмів є підготовчим до побудови моделі;

2) в блоці вибору моделі та перевірки адекватності моделі застосовується запропонований в підрозділі 2.2, алгоритм стратегії вибору моделі за методом Бокса-Дженкінса (рисунок 2.3). Згідно нього оцінюються параметри моделей, обчислюються їх залишки, які використовуються для перевірки адекватності;

2) в блоці прогнозування за моделлю використовується алгоритм створення прогнозів (формули (2.43-2.47), (2.50)), що включає прогноз на один крок, прогноз на s кроків, обчислення помилки прогнозу і її дисперсії, приведений в підрозділі 2.3.

Розроблена програмна система була застосована для прогнозу прибутків комп'ютерної фірми.

В пакеті STATISTICA аналіз і прогноз за моделлю здійснюється в модулі *Часові ряди і прогнозування (Time Series / Forecasting)* (рисунок А.1). В процесі роботи генерується велика кількість діалогових і допоміжних вікон, таблиць, графіків. Приведемо схему переходів між чотирма основними діалоговими вікнами першого рівня (рисунок 3.2).

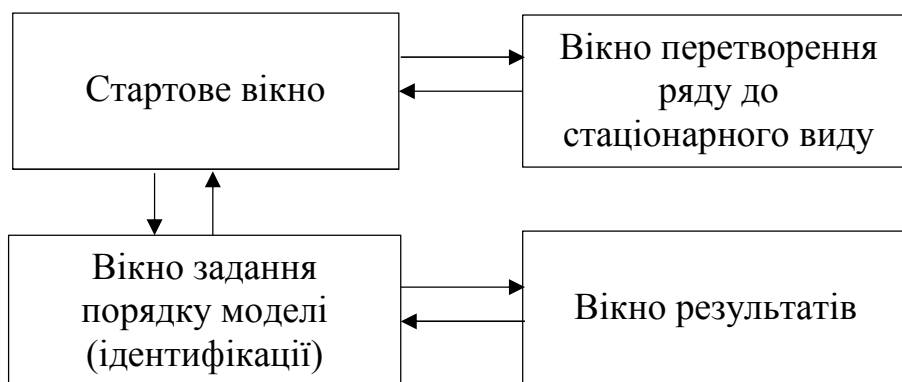


Рисунок 3.2 – Схема переходів між діалоговими вікнами

Розроблену програмну систему прогнозування можна розбити на три модулі: модуль ідентифікації моделі, модуль оцінювання моделі і перевірки адекватності моделі, модуль прогнозування за моделлю. Розглянемо реалізацію кожного з цих модулів.

3.3 Модуль ідентифікації моделі

Для реалізації цього модуля використано пакет STATISTICA. У ньому передбачено такі методи для аналізу часових рядів [45-47]:

- ARIMA – метод авторегресії та проінтегрованого ковзної середньої;
- Interrupted time series analysis – аналіз дискретного часового ряду (моделі інтервенції для ARIMA);
- Exponential smoothing & forecasting – експоненційне згладжування та прогнозування;
- Seasonal decomposition (1, 2) – сезонна декомпозиція 1, 2 (квартальна, за місяцями);
- Distributed lags analysis – аналіз розподілених лагів (регресійна модель для двох часових рядів);
- Spectral (Fourier) analysis – спектральний (Фур'є-) аналіз.

В вхідному вікні (рисунок А.1) приведені часові ряди прибутків комп'ютерної фірми за 2 роки помісячно. Перед тим як підігнати до часового ряду модель, треба зробити його стаціонарним. Для цього будуємо графік за початковим рядом даних. В діалоговому вікні *Time Series Analysis* натискаємо кнопку *ARIMA & autocorrelation function* потім вибираємо *Review series* (рисунок 3.3) \Rightarrow *Review and plot variables* \Rightarrow *Review highlighted variable* \Rightarrow *Plot*.

На рисунку 3.4 побудовано графік ряду для попередньої візуальної оцінки даних.

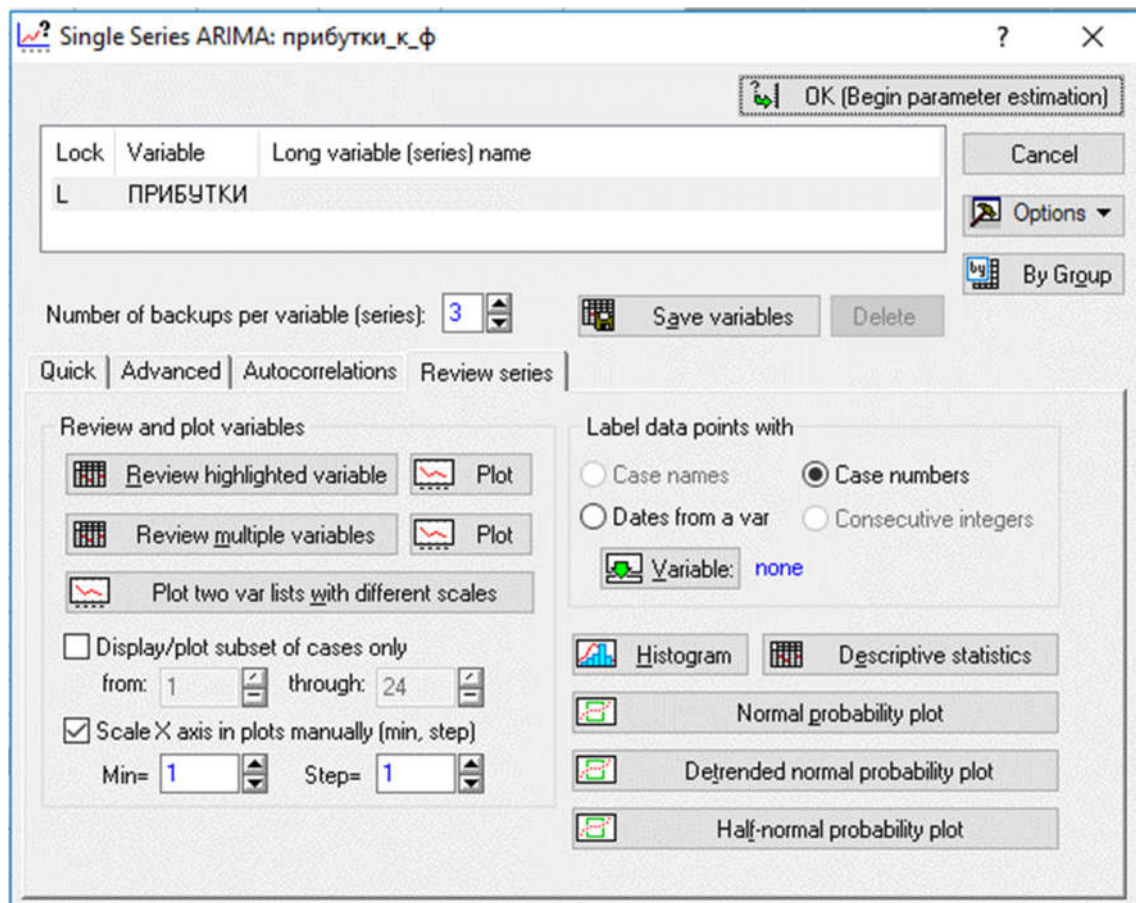


Рисунок 3.3 – Діалогове вікно *Single Series ARIMA*

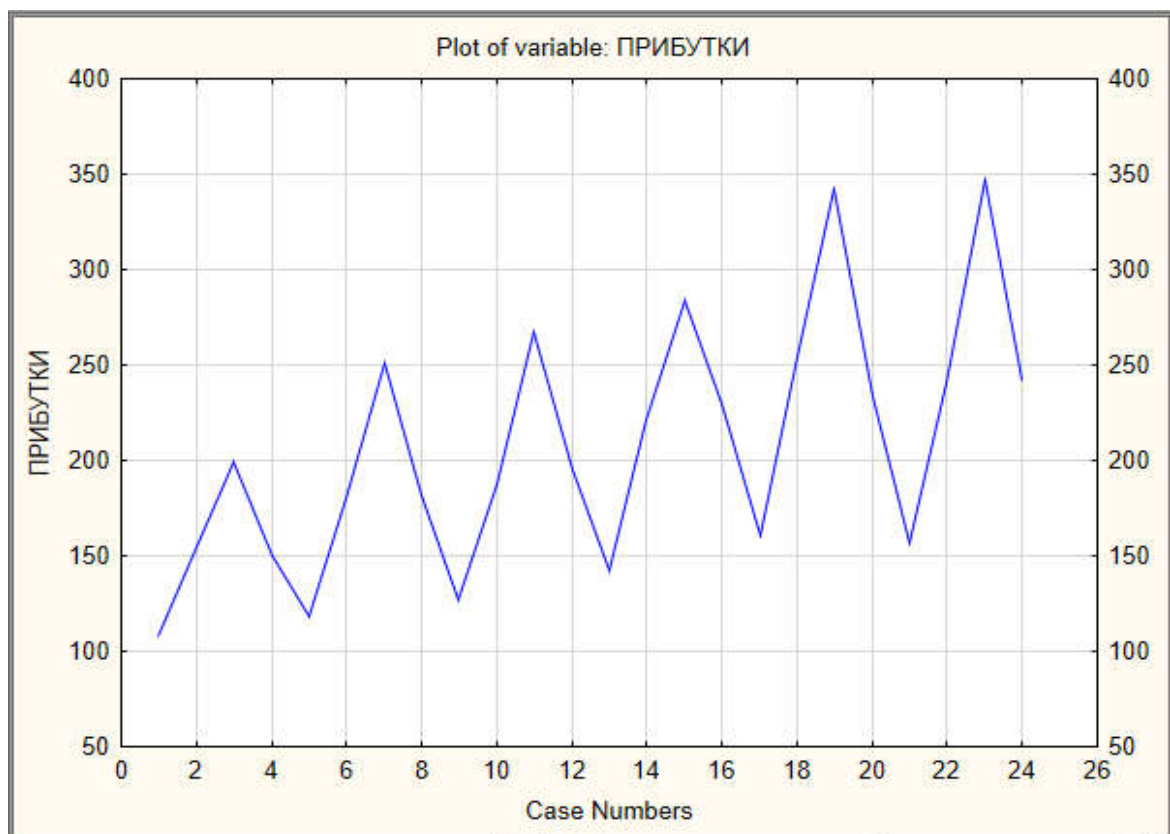


Рисунок 3.4 – Графік вихідних даних часового ряду

Даний ряд не є стаціонарним, спостерігається тренд ряду (плавне збільшення значень ряду), деяка сезонність.

Обрати спосіб перетворення ряду можна за допомогою кнопки *Other transformations & Plots*, розташованої на вкладці *Advanced* діалогового вікна *Single Series ARIMA* (рисунок А.3) [48,49].

Для автоматичного відображення графіку перетворених даних після кожного перетворення ряду встановлюється опція *Plot variables (Series) after each transformation* на вкладці *Review & plot* вікна *Transformation of Variables* (рисунок А.4).

Для відображення коефіцієнтів автокореляції і часткової автокореляції використовуємо вкладку *Autocorr*s вікна *Transformation of Variables*.

На перших шести вкладках вікна *Transformation of Variables* можна вибрати різні способи перетворення часового ряду, встановивши відповідну опцію і натиснувши кнопку *OK (Transform selected series)*.

Рекомендуються наступні перетворення:

- логарифмування для стабілізації дисперсії і зменшення амплітуд коливання часового ряду (вкладка $x = f(x)$);
- видалення тренду (вкладка $x = f(x)$);
- взяття різниці першого або вищого порядків на вкладці *Різниця (Difference)* для видалення лінійного тренда;
- згладжування (вкладка *Smoothing*);
- віднімання середнього, стандартизація ряду (вкладка $x = f(x)$).

Щоб зробити ряд даних (рисунок А.1) стаціонарним, а потім підібрати авторегресійну модель, спочатку потрібно послідовно застосувати кілька перетворень до часового ряду прибутків комп'ютерної фірми за місяцями. Застосуємо логарифмічне перетворення *Natural log* на вкладці $x = f(x)$ (рисунок 3.5).

Графік прологарифмованого ряду (рисунок 3.6) має меншу дисперсію, його коливання зменшилися.

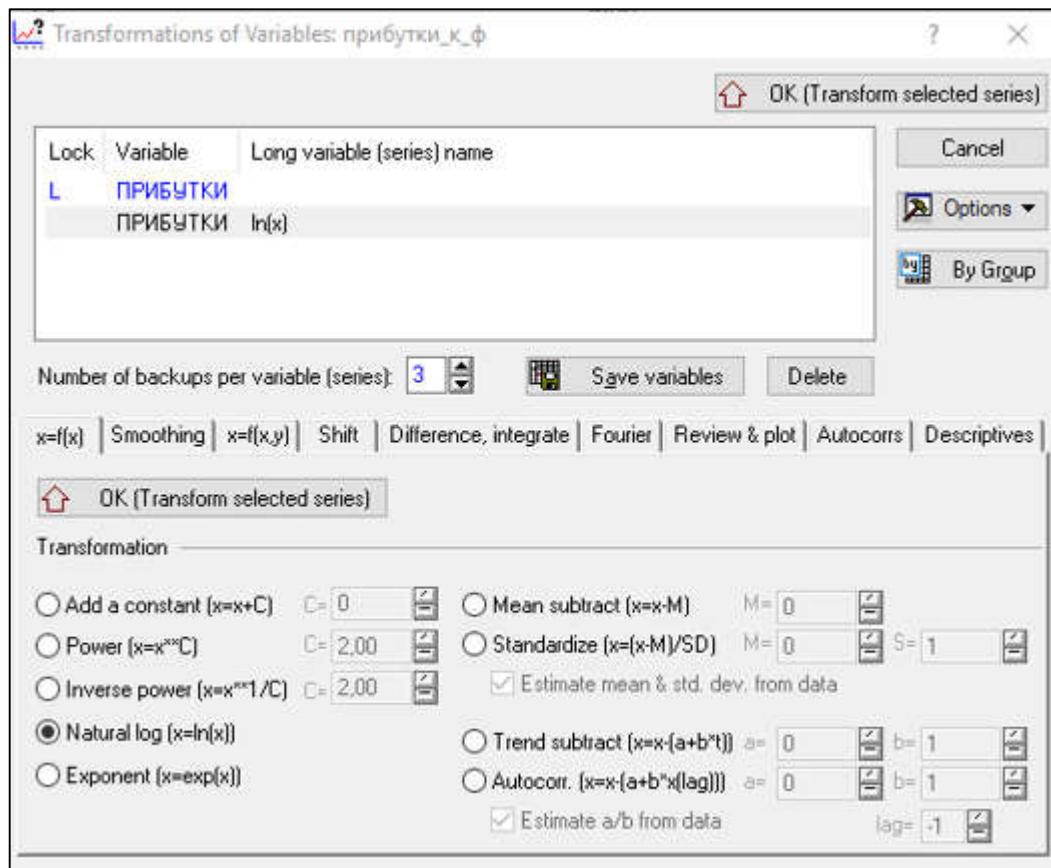


Рисунок 3.5 – Вкладка $x = f(x)$ діалогового вікна

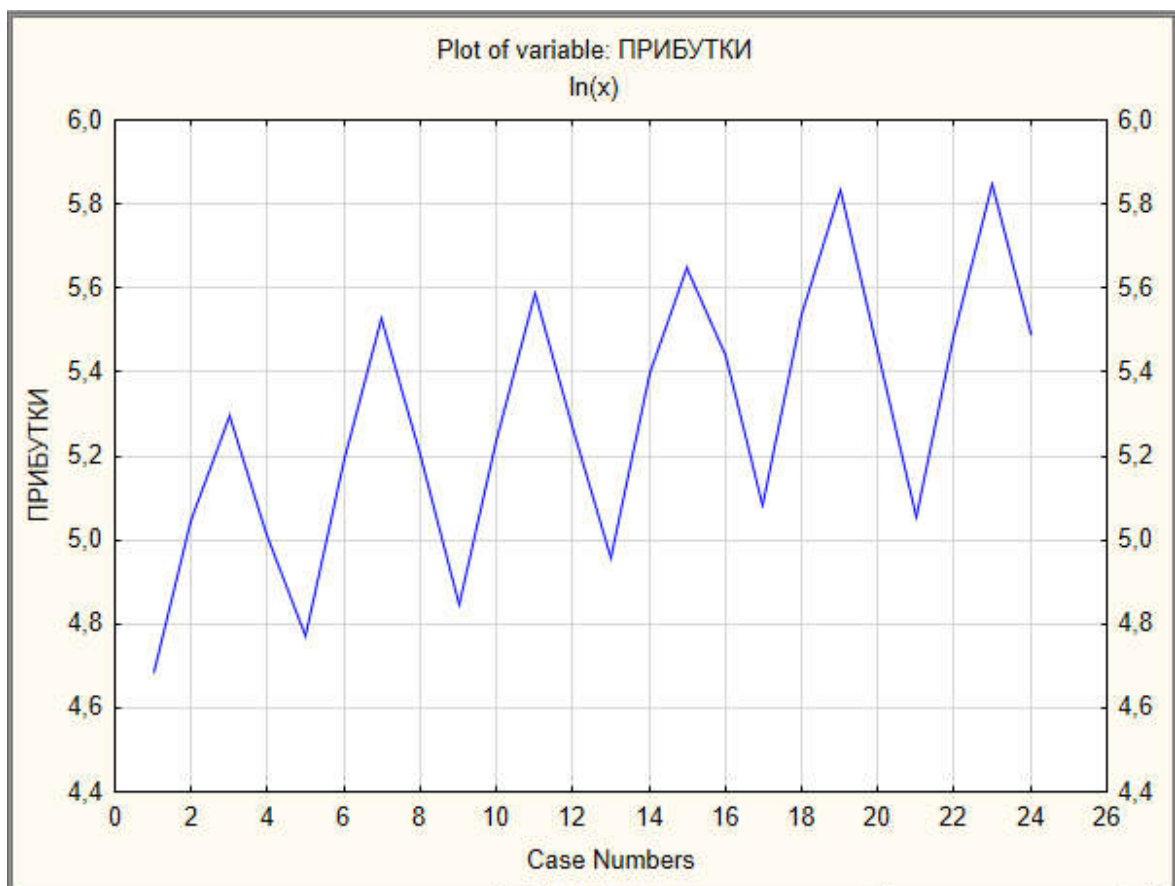


Рисунок 3.6 – Графік прологарифмованих даних часового ряду

Після зменшення розкиду ідентифікуємо параметри моделі для подальшого їх оцінювання.

Так як з рисунку 3.6 видно, що тренд ряду виражений і вибіркова автокореляційна функція (рисунок 3.7) (вкладка *Autocorr*s вікна *Transformation of Variables*), не має тенденції до згасання, то це вказує на нестационарність ряду. Тому на наступному етапі (рисунок А.5) застосуємо до часового ряду різницеве перетворення (вкладка *Differencing, Integrate* вікна *Transformation of Variables, lag=1*).

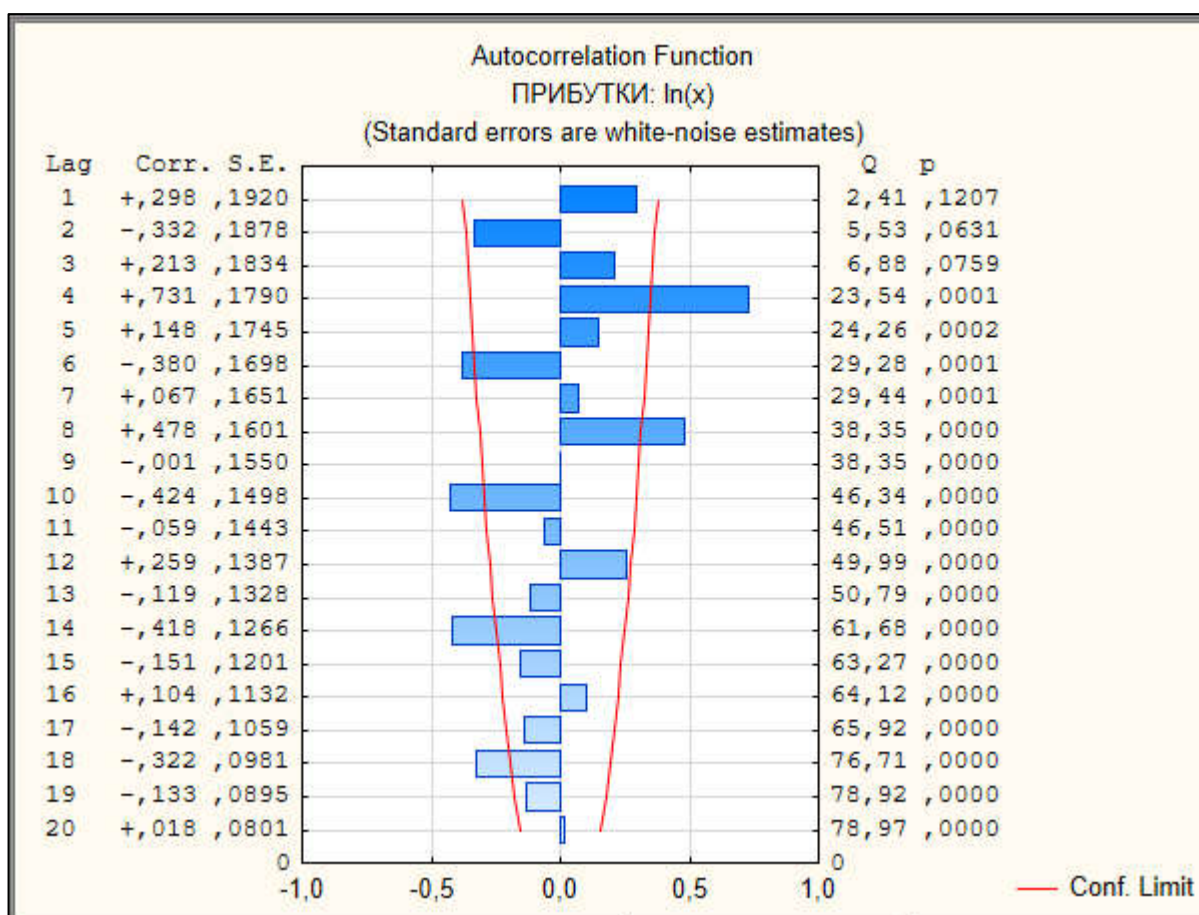


Рисунок 3.7 – Графік автокореляційної функції

Графік перетвореного ряду показаний на рисунку 3.8. Очевидно, що ряд став стаціонарний.

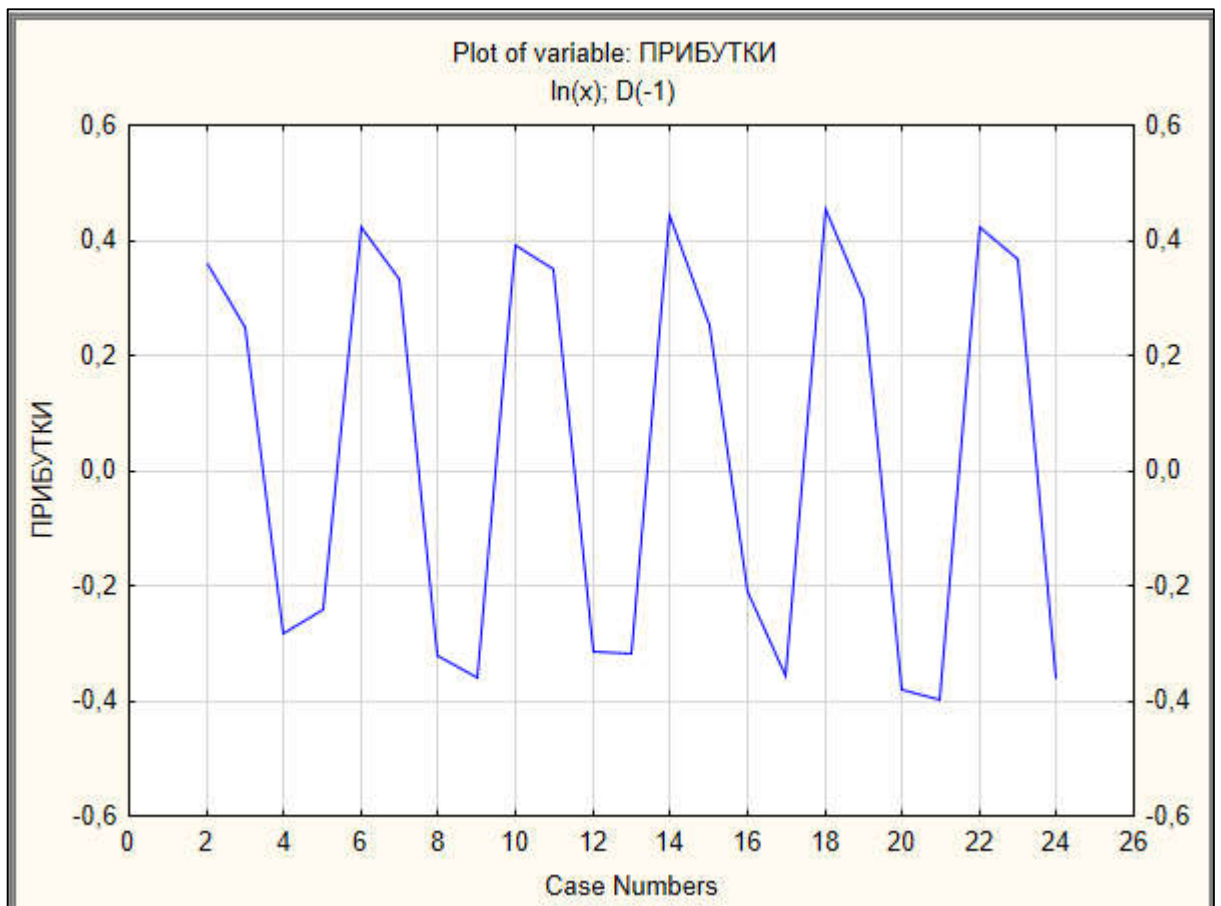


Рисунок 3.8 – Графік прибутків після застосування двох перетворень: Ln , $Differencing$ з лагом 1

3.4 Модуль оцінювання моделі і перевірки адекватності моделі. Модуль прогнозування за моделлю

Модель – це рівняння, за яким будуть обчислюватися прогнозовані прибутки. Оцінювання моделі – це вибір конкретного виду рівняння. В додатку Б міститься лістинг коду програми. Ми повинні вказати програмі, скільки доданків буде в рівнянні авторегресії (p), а скільки – в рівнянні ковзної середньої (q). Діалогове вікно, призначене для оцінювання моделі приведено на рисунку 3.9.

Зробимо побудову моделей ARIMA з різними параметрами і порівняємо їх. Для цього побудуємо моделі ARIMA(1,1,1), ARIMA(0,1,2), ARIMA(1,1,2), ARIMA(2,1,2), ARIMA(2,1,0), ARIMA(2,1,1). В таблиці 3.3 приведені значення оцінок, їх стандартні помилки. У випадку, якщо оцінка є значуща, то вона позначена жирним шрифтом.

Таблиця 3.3 – Порівняльна таблиця моделей ARIMA

Параметри	Модель ARIMA											
	(0,1,1)		(1,1,1)		(1,1,2)		(2,1,2)		(2,1,0)		(2,1,1)	
	оцінка	станд. помилка	оцінка	станд. помилка	оцінка	станд. помилка	оцінка	станд. помилка	оцінка	станд. помилка	оцінка	станд. помилка
Const	,03686	,01162	,04175	,02419	,04467	,00704			,03012	,01471	,02995	,01372
p(1)			,24676	,37984	,01369	,24242	,01539	не обч.	,01286	,08309	,02483	,08819
p(2)							-1,000	не обч.	-,9851	,08490	-,9856	,08715
q(1)	,87580	,12873	,96468	,26209	,19216	,00000	-,1405	не обч.			,09441	,19291
q(2)					,80783	,00000	-,0341	не обч.				

Як бачимо з таблиці 1 єдиною моделлю, в якій всі оцінки є значущими є модель ARIMA(0,1,1).

На рисунках 3.10 – 3.16 приведено аналіз моделі ARIMA(0,1,1).

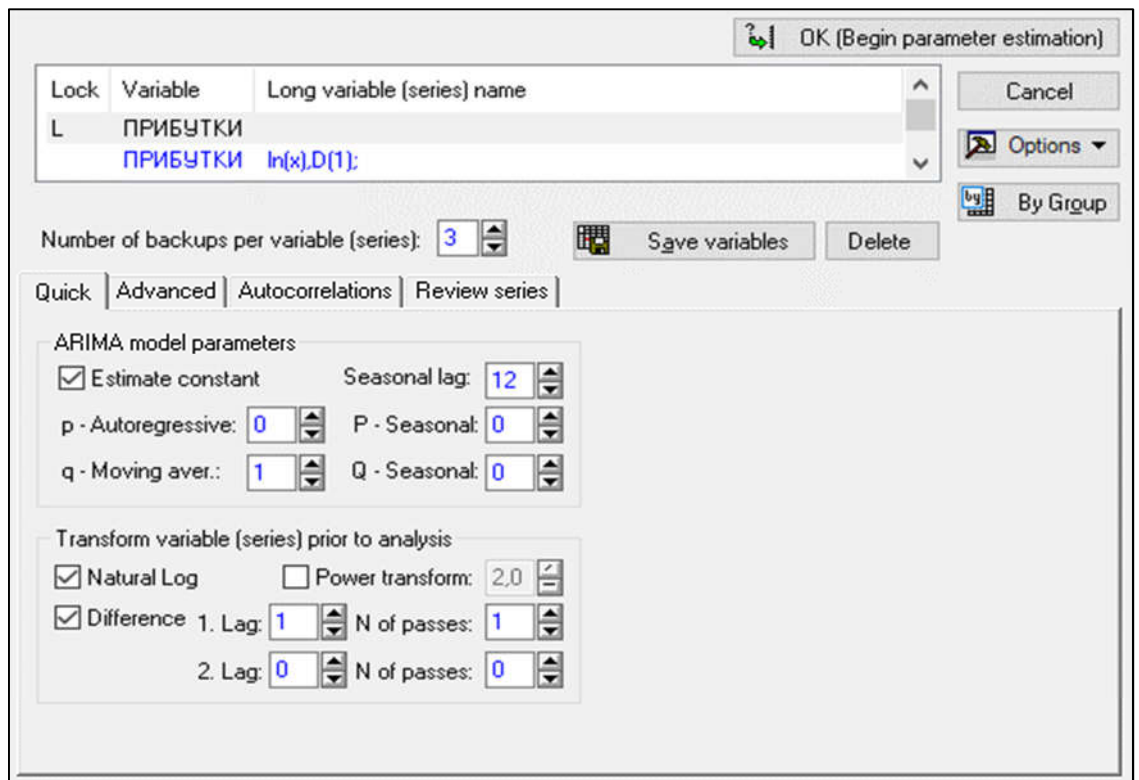


Рисунок 3.9 – Діалогове вікно для оцінювання моделі

На рисунку 3.10 – вікно з результатами обчислень. Коефіцієнти рівняння моделі стійкі, і значущі (на це вказує позначення червоним кольором). Вони більш, ніж в два рази перевищують свої стандартні помилки. Параметри прогнозу можна задати на вкладці Advanced в полі Forecasting.

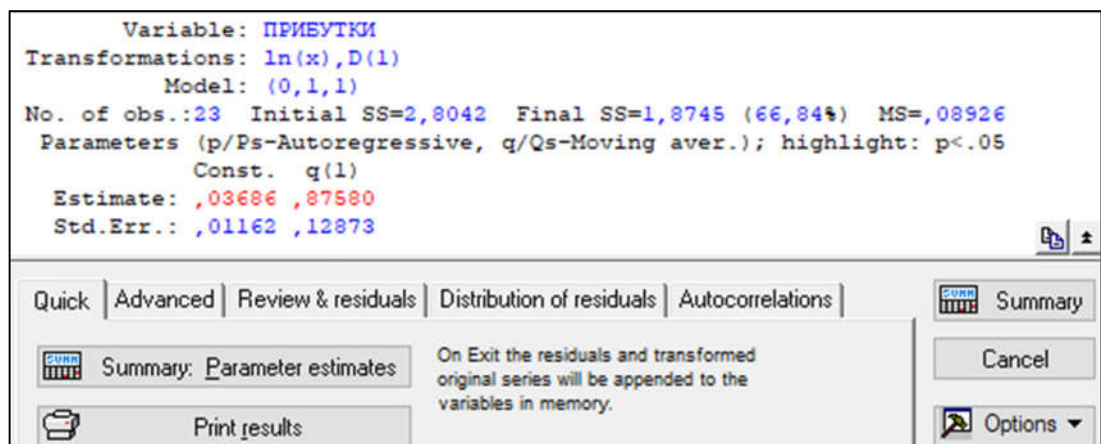


Рисунок 3.10 – Вікно з результатами обчислень

Дальше проходить перевірка адекватності моделі.

На рисунку 3.11 приведено значення оцінок, їх стандартні помилки, t -статистики, P -значення, довірчі інтервали (95%) оцінок. Оцінки коефіцієнтів моделі є статистично значущими (позначені червоним кольором).

Input: ПРИБУТКИ (прибутки_к_ф) Transformations: ln(x),D(1) Model:(0,1,1) MS Residual= ,08926						
Paramet.	Param.	Asympt. Std.Err.	Asympt. t(21)	p	Lower 95% Conf	Upper 95% Conf
Constant	0,036860	0,011624	3,171164	0,004601	0,012688	0,061033
q(1)	0,875802	0,128731	6,803371	0,000001	0,608092	1,143512

Рисунок 3.11 – Аналіз оцінок параметрів моделі ARIMA(0,1,1)

Перевіряємо, чи помилки моделі є білим шумом. Для цього перевіримо, чи залишки регресії, подібні на білий шум, тобто чи залишки мають нульову автокореляцію. На рисунку 3.12 зображена корелограма залишків.

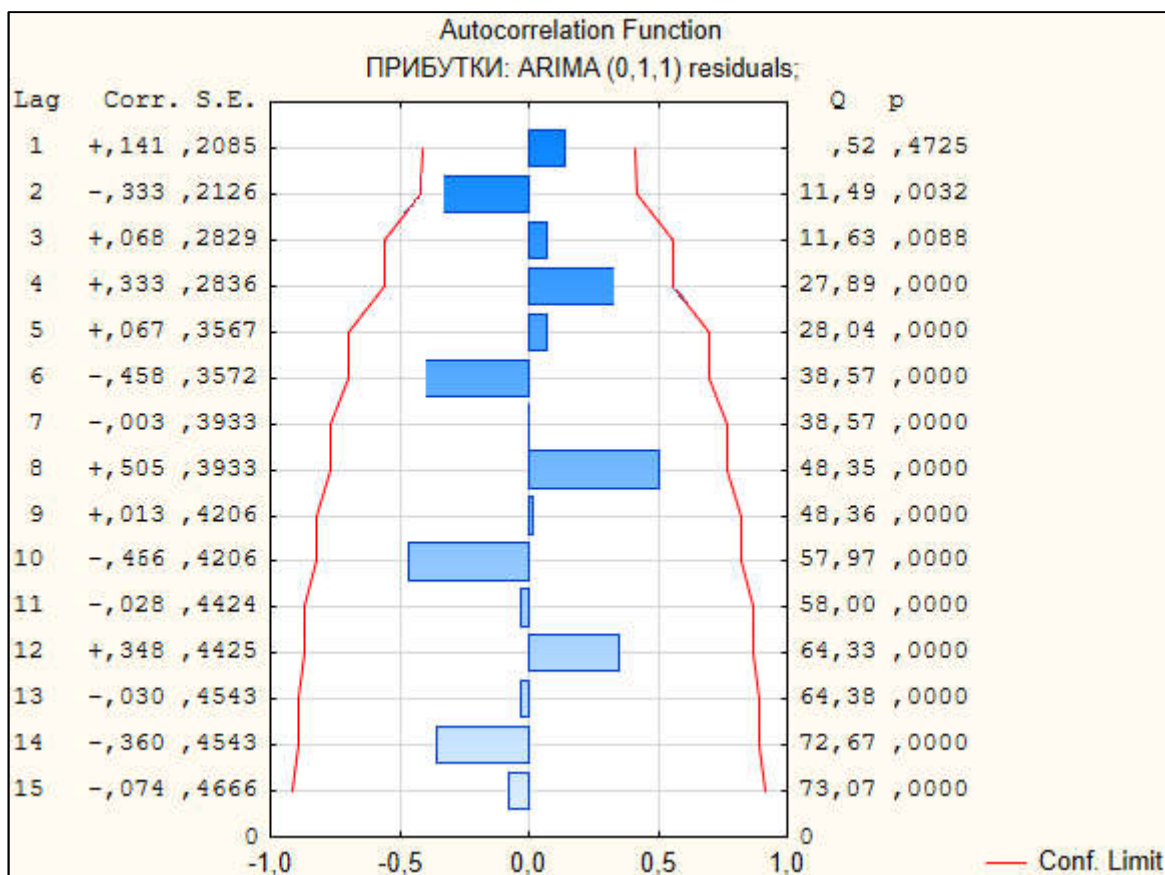


Рисунок 3.12 – Автокореляційна функція залишків моделі ARIMA(0,1,1)

З рисунку 3.12 бачимо, що коефіцієнти автокореляції незначущі (не виходять за межі довірчого коридору) і розташовані хаотично, тобто залишки незалежні, подібні на білий шум.

Для перевірки нормальності розподілу залишків моделі використовуємо гістограму залишків (рисунок 3.13) і графік нормального розподілу залишків (рисунок 3.14). Як бачимо, гістограма майже знаходиться в межах нормальної кривої, тому розподіл близький до нормального. Це підтверджується й з рисунку 3.14 – всі точки групуються близько прямої лінії на графіку.

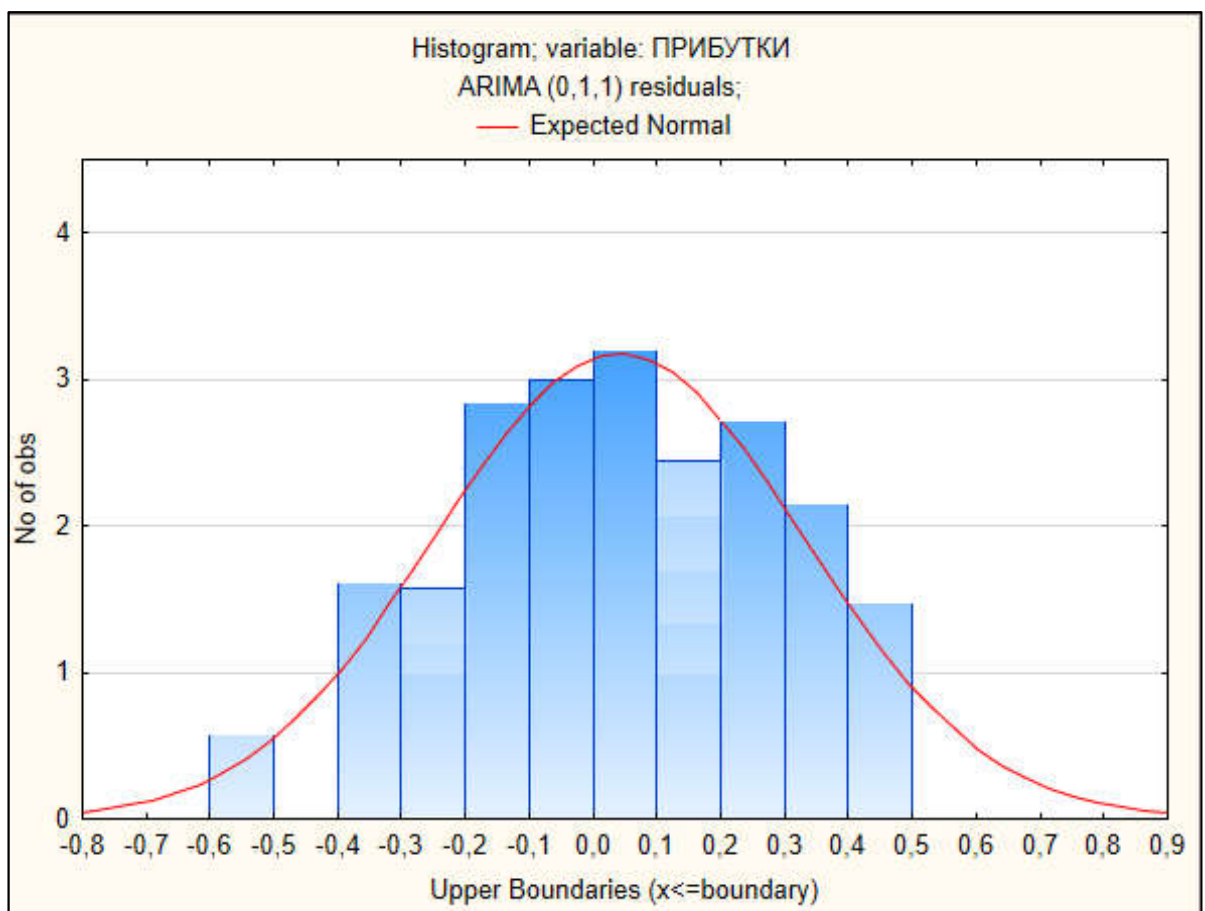


Рисунок 3.13 – Гістограма залишків моделі ARIMA(0,1,1)

Графік ряду з доданими прогнозними значеннями прибутків і довірчими інтервалами для них на три місяці приведено на рисунку 3.15, числові значення їх на рисунку 3.16.

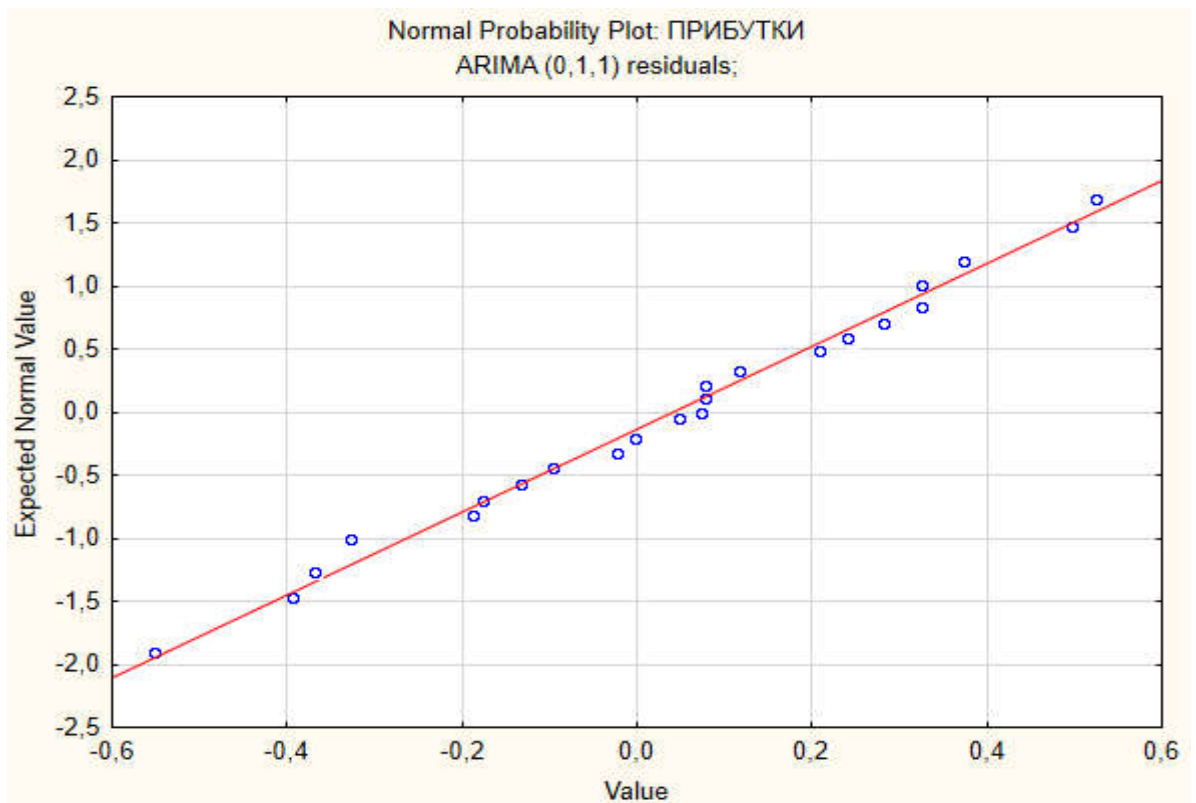


Рисунок 3.14 – Графік нормального розподілу залишків моделі ARIMA(0,1,1)

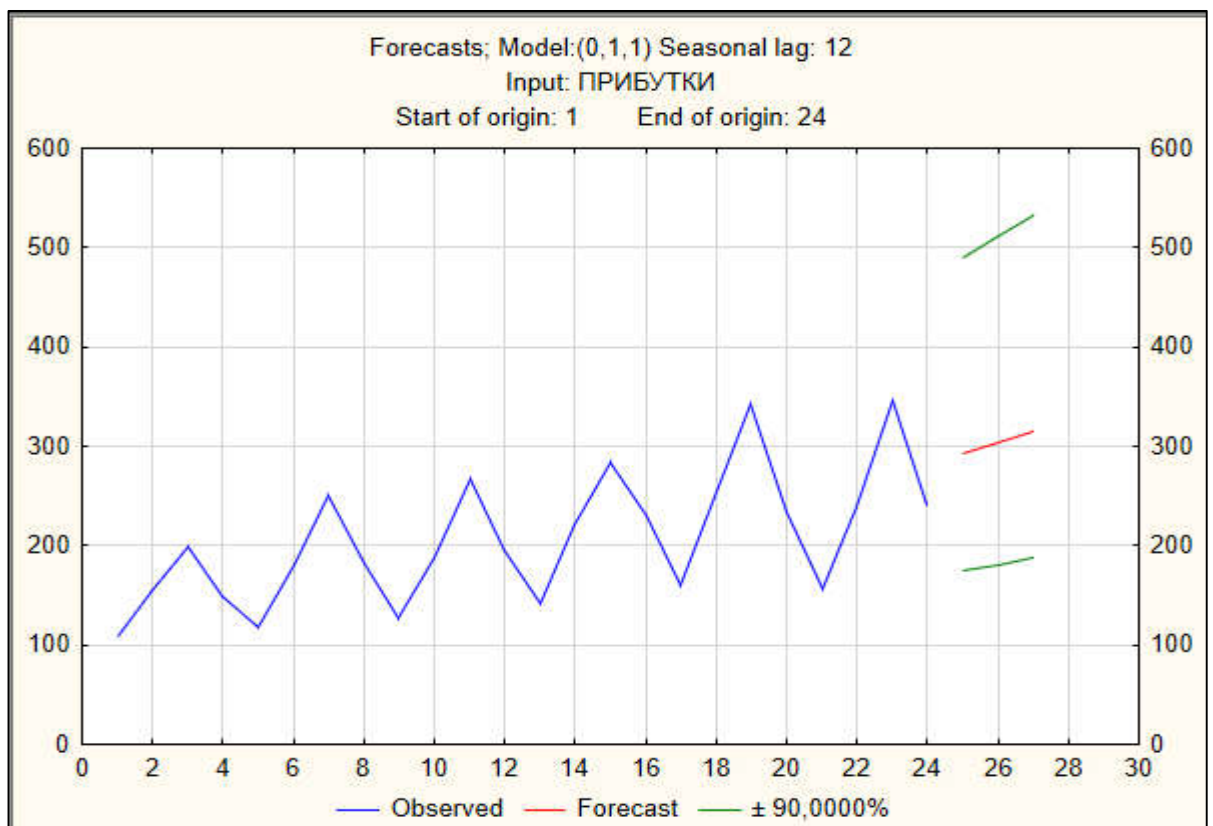


Рисунок 3.15 – Графік ряду з прогнозними значеннями і довірчими інтервалами

Forecasts; Model:(0,1,1) Seasonal lag: 12 (прибутки_к_ф) Input: ПРИБУТКИ Start of origin: 1 End of origin: 24			
CaseNo.	Forecast	Lower 90,0000%	Upper 90,0000%
25	293,3820	175,4544	490,5719
26	304,3979	181,3248	511,0064
27	315,8275	187,3971	532,2761

Рисунок 3.16 – Прогнозні значення і довірчі інтервали

Отже, в даному розділі на основі запропонованих алгоритмів розроблено модуль ідентифікації моделі, модуль оцінювання моделі і перевірки її адекватності, модуль прогнозування за моделлю, що дало можливість покращити прогнозування прибутків для фірми.

ВИСНОВКИ

1. Здійснено аналіз стану розвитку ІТ-бізнесу в Україні.
2. Проаналізовано методи та алгоритми прогнозування, встановлено, що для ефективного прийняття рішень на фірмі необхідний прогноз майбутніх значень часового ряду.
3. Розроблено алгоритми побудови ARIMA-моделей.
4. Розроблено алгоритми перевірки адекватності моделі і вибору з декількох адекватних кращої.
5. На основі розроблених алгоритмів побудови ARIMA-моделей здійснено комп'ютерне моделювання.
6. На основі запропонованих алгоритмів розроблено модуль ідентифікації моделі, модуль оцінювання моделі і перевірки адекватності моделі, модуль прогнозування за моделлю, що дало можливість покращити прогнозування прибутків для фірми.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Винничук Р. О. Особливості розвитку ІТ-ринку в Україні: стан та тенденції / Р. О. Винничук, Т. В. Склярчук // Вісник Національного університету "Львівська політехніка". Логістика. – 2015. – № 833. – С. 3-8.
2. Березька К.М. Прогнозування прибутків комп'ютерної фірми на основі ARIMA-моделей / К.М. Березька, В.М. Неміш, Ю.С. Тимчук // Матеріали III всеукраїнської науково-практичної конференції «Прикладна геометрія та інформаційні технології в моделюванні об'єктів, явищ і процесів» (AGIT-2018), м. Миколаїв, 17–19 жовтня 2018 р. – Миколаїв: МНУ імені В.О. Сухомлинського, 2018. – С. 120-121.
3. Методичні рекомендації до виконання дипломної роботи з освітньо-кваліфікаційного рівня «Магістр». Спеціальність «Комп'ютерні системи та мережі» / О.М. Березький, Л.О. Дубчак, Г.М. Мельник / Під ред. О.М. Березького – Тернопіль: ТНЕУ, 2016.– 47 с.
4. <https://biz.nv.ua/ukr/economics/ukrajina-na-svitovomu-rinku-it-top-10-najtsikavishih-faktiv-infografika-2353974.html>.
5. Грома Я. В. Перспективи розвитку ІТ-бізнесу в Україні [Електронний ресурс]. – Режим доступу: http://ela.kpi.ua/bitstream/123456789/20047/1/MNTS2017_57_Hroma.pdf
6. Інфляційний звіт НБУ за 2016 рік [Електронний ресурс] – Режим доступу: <http://www.bank.gov.ua/doccatalog/document?id=30350020>.
7. Щука В. Г. Дослідження методів прогнозування та обґрунтування вибору кращого з них для прогнозування показників фінансово - господарської діяльності підприємства / В. Г. Щука, Д. І. Мандрик // Вісник Хмельницького національного університету. – 2015. – № 1. – С. 102-104.
8. Грабовецький Б. Є. Методи експертних оцінок: теорія, методологія, напрямки використання: моногр. / Б. Є. Грабовецький. – Вінниця: ВНТУ, 2010. – С. 16–59.

9. Бестужев-Лада И. В. Рабочая книга по прогнозированию / И. В. Бестужев-Лада. – М.: Мысль, 1982. – С. 10–21.
10. https://pidruchniki.com/16400221/menedzhment/osnovni_metodi_prognozuvannya_klasifikatsiya_prognoziv
11. Лук'яненко І. Г. Аналіз часових рядів. Побудова ARIMA, ARCH/GARCH моделей з використанням пакета EViews 6.0: Практичний посібник для роботи в комп'ютерному класі / І. Г. Лук'яненко І. Г., В. М. Жук. – К.: НаУКМА; Аграр Медіа Груп, 2013. – 187 с.
12. Айвазян С.А. Программное обеспечение по статистическому анализу данных: методология сравнительного анализа и выборочный обзор рынка/ С.А.Айвазян, В.С.Степанов [Электронный ресурс]. – Режим доступа: <http://pubhealth.spb.ru/SAS/STatProg.htm>.
13. Василенко Ж.В. Программное обеспечение по статистическому анализу данных. Методология сравнительного анализа [Электронный ресурс]. – Режим доступа: http://www.giac.unibel.by/sm_full.aspx?guid=8313.
14. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учеб.пособ.– 2-е изд., испр. и доп. – М.: Форум, 2008.– 464 с.
15. Красильников Д.Е. Программное обеспечение эконометрического исследования [Электронный ресурс]. – Режим доступа: [http://www.unn.ru/pages/issues/vestnik/99999999_West_2011_3\(2\)/37.pdf](http://www.unn.ru/pages/issues/vestnik/99999999_West_2011_3(2)/37.pdf).
16. Майборода Р.Є., Сугакова О.В. Статистичний аналіз даних за допомогою пакету STATISTICA [Електронний ресурс]. – Режим доступу: <http://matphys.rpd.univ.kiev.ua/downloads/courses/mmatstat/StatAn.doc>.
17. Шарапов О. Д. Системний аналіз: Навч.-метод. посібник для самост. вивч. дисц./ О. Д. Шарапов, В. Д. Дербенцев, Д. Є. Семьонов. – К.: КНЕУ, 2003. – 154 с.
18. Електронний учебник по статистике [Електронний ресурс]. – Режим доступу: <http://statsoft.ru/home/textbook/default.htm>.

19. Роїк, М. В. Огляд програмних засобів статистичного аналізу даних / М. В. Роїк, О. І. Присяжнюк, В. О. Денисюк // Ефективна економіка. – 2017. – №7, [Електронний ресурс]. – Режим доступу: <http://www.economy.nauka.com.ua/?op=1&z=5676>
20. System requirements [Electronic resource]. – Mode of access: https://en.wikipedia.org/wiki/System_requirements.
21. Статистические ППП [Електронний ресурс]. – Режим доступу: <http://helpiks.org/6-54552.html>.
22. EViews Econometric Modeling Software [Electronic resource]. – Mode of access: <https://www.ihs.com/products/eviews-econometric-modeling-analysis-software.html>.
23. The Language of Technical Computing [Electronic resource]. – Mode of access: <https://www.mathworks.com/products/matlab.html>.
24. TSP International [Electronic resource]. – Mode of access: <http://www.tspintl.com/>.
25. Minitab 18 [Electronic resource]. – Mode of access: <https://www.minitab.com/en-us/products/minitab/>.
26. Функції Excel (за категоріями) [Електронний ресурс]. – Режим доступу: <https://support.office.com/uk-ua/article>.
27. Estima [Electronic resource]. – Mode of access: <https://estima.com/>.
28. The RATS Software Forum [Electronic resource]. – Mode of access: <https://estima.com/forum/viewforum.php?f=2>.
29. IBM SPSS Software [Electronic resource]. – Mode of access: <https://www.ibm.com/analytics/us/en/technology/spss>.
30. STATA [Electronic resource]. – Mode of access: <http://www.stata.com/products/>.
31. StatSoft's Electronic Statistics Textbook [Electronic resource]. – Mode of access: <http://www.statsoft.com/textbook/stathome.html>.
32. Systat Software Inc.(SSI) [Electronic resource]. – Mode of access: <http://www.sigmaplot.co.uk>.

33. 830-1998 - IEEE Recommended Practice for Software Requirements Specifications [Electronic resource]. – Mode of access: <http://standards.ieee.org/findstds/standard/830-1998.html>.

34. Економіко-математичне моделювання: Навч. посібник / За ред. О.Т. Іващука. – Тернопіль: ТНЕУ, Економічна думка, 2008. – 704 с.

35. Магнус Я. Р. Эконометрика. Начальный курс: Учебник. – 6-е изд., перераб. и доп. / Я. Р. Магнус, П. К. Катыхов, А. А. Пересецкий. – М.: Дело, 2004. – 576 с.

36. Грубер Й. Эконометрия. – К.: 1996. – Т.1. Введение в эконометрию. – 400 с.

37. Маленво Э. Статистические методы эконометрии. – М.: Статистика, 1975. – 423 с.

38. Джонстон Дж. Эконометрические методы. – М.: Статистика, 1980. – 444 с.

39. Толбатов Ю.А. Эконометрика. – К.: Четверта хвиля, 1997. – 320 с.

40. Кейн Э. Экономическая статистика и эконометрия. Введение в количественный экономический анализ. – М.: Статистика, 1977. – 254 с.

41. Березька К. М. Побудова ARIMA моделей часових рядів для прогнозування інвестицій / К. М. Березька, В.В. Маслій // Матеріали Міжнародної наукової конференції «Інтелектуальні системи прийняття рішень та проблеми обчислювального інтелекту» (ISDMCI'2015), м. Залізний Порт, 25–28 травня 2015 р. – Херсон: ХНТУ, 2015. – С. 28–32.

42. Березька К. М. Модифіковані ARIMA-моделі обсягів прямих іноземних інвестицій / К. М. Березька, В.В. Маслій // Матеріали ІХ-ої міжнародної конференції «Актуальні проблеми економіки 2015-2016», м. Київ, 19 лютого 2016 р. – К.: Національна академія управління, 2016. – С. 15 – 20.

43. Маслій В. В. Вибір та оцінка ARIMA-моделі для прогнозування обсягів прямих іноземних інвестицій / В. В. Маслій, К.М. Березька // Науковий вісник Міжнародного гуманітарного університету. – 2017. – Вип. 24, Ч. 2. – С. 115-119.

44. Олійник А. О. Інтелектуальний аналіз даних: навчальний посібник / А. О. Олійник, С. О. Субботін, О. О. Олійник. – Запоріжжя: ЗНТУ, 2012. – 271 с.
45. Ковальчук О. Я. Математичне моделювання та прогнозування в міжнародних відносинах (Ч. 1): навчальний посібник / О. Я. Ковальчук. – Тернопіль: ТНЕУ, 2016. – 423 с.
46. Бахрушин В. Є. Методи аналізу даних: навчальний посібник / В. Є. Бахрушин. – Запоріжжя: КПУ, 2011. – 268 с.
47. Василенко О. А. Математико-статистичні методи аналізу у прикладних дослідженнях: навчальний посібник / О. А. Василенко, І. А. Сенча. – Одеса: ОНАЗ ім. О. С. Попова, 2011. 166 с.
48. Мамчич Т. І. Статистичний аналіз даних з пакетом Statistica: навчально-методичний посібник / Т. І. Мамчич, А. Я. Оленко, М. М. Осипчук, В. Г. Шпортюк; Нац. ун-т «Києво-Могилянська академія». – Дрогобич: Вид. фірма «Відродження», 2006. – 207 с.
49. Боровиков В. Statistica. Искусство анализа данных на компьютере : Для профессионалов / В. Боровиков – [2-е изд.] – СПб.: Изд-во Питер, 2003 – 688 с.
50. «Моделирование ARIMA - временного ряда средствами MatLab 6.0» [Електронний ресурс] – Режим доступу: <http://www.excelprognoz.narod.ru/arima.html>.