­­МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ “ЛЬВІВСЬКА ПОЛІТЕХНІКА”

­­­

Контрольна робота

з курсу «Інтелектуальна обробка даних»

для студентів базового напрямку 6.08.04 "Комп’ютерні науки"

(заочна форма навчання)

Варіант 10

Виконав студент гр. КНз-3

Чалий Михайло

­­

Львів 2015

## Мета роботи

Мета даної лабораторної роботи полягає у вивченні аналізу і прогнозуванню часових рядів.

## Прогнозування на основі часових рядів

Часовий ряд (time series) – це сукупність вимірювань деякої змінної величини, які проводяться у часі. Характерною особливістю часових рядів є те, що спостереження за деяким об'єктом проводяться послідовно в часі. Наприклад, температура повітря в середині кожної години доби, щорічна врожайність зернових, щоденний об'єм продажів якого-небудь товару, вартість акції підприємства, рівень інфляції, обмінний курс валют – все це часові ряди.

Для аналізу часового ряду порядок у послідовності є суттєвим, тобто час виступає одним із визначальних чинників. Це відрізняє часовий ряд від звичайної випадкової вибірки, де індекси вводять лише для зручності ідентифікації. Принциповою відмінністю часового ряду від простих статистичних сукупностей є:

* по-перше, рівні часового ряду не є незалежними. Інакше кажучи, якщо майбутні значення змінної можна визначити, то вони є функцією від минулих значень цієї змінної;
* по-друге, рівні часового ряду неоднаково розподілені. Закон розподілу ймовірностей цих випадкових величин і, зокрема, їхні математичні сподівання та дисперсії можуть залежати від часу.

Прогнозування на основі часових рядів – один із самих популярних підходів до прогнозування розвитку економічних процесів, об’ємів торгових операцій, об’ємів виробництва та накопичення продукції на складах, оцінювання альтернативних економічних стратегій, формування бюджетів підприємств та держави, прогнозування та менеджмент економічних і фінансових ризиків та інше. Загалом методи прогнозування можна розділити на три широкі класи:

1. Прогнозування на основі суджень, тобто, прогнозування, що ґрунтується на суб’єктивних судженнях (оцінках), інтуїції, поглиблених знаннях конкретної області та іншій інформації, що має відношення до прогнозованого процесу – так зване передбачення;

2. Методи прогнозування на основі використання часового ряду однієї змінної, тобто, на основі авторегресії, авторегресії з ковзним середнім (АРКС) та АРКС плюс модель тренду;

3. Методи прогнозування на основі використання часових рядів декількох змінних.

В останньому випадку ендогенна змінна, що прогнозується, залежить від декількох регресорів або екзогенних змінних у правій частині рівняння. Очевидно, що в загальному випадку метод прогнозування може поєднувати у собі 2-3 наведених вище методи. На сьогоднішній день в спеціальній літературі описано багато методів прогнозування на основі використання часових рядів. Найбільш поширеними серед них є метод групового врахування аргументів (МГВА), регресійний аналіз (авторегресія (АР), авторегресія з ковзним середнім (АРКС), авторегресія з інтегрованим ковзним середнім (АРІКС), лінійна та нелінійна множинна регресія, квантильна регресія, регресійні дерева), нейромережі, байєсівські мережі, нечіткі множини, нечіткі нейромережі та інші.

В загальному випадку прогноз може бути представлений одним (точковим) значенням змінної, інтервалом, в який попадає випадкова змінна, а також ймовірністю прийняття змінною (чи подією) деякого значення у вибраному інтервалі. Якщо для опису процесу застосовують лінгвістичні змінні, то прогнозом буде нечітке лінгвістичне значення, але його також можна перетворити в чітке число.

Можна по-різному ставити задачу прогнозування в залежності від рівня прийняття рішення та конкретної поставленої задачі управління чи контролю. Прогнозування може стосуватись таких складових процесу:

– детермінованого тренду, як індикатора довгострокових змін процесу;

– випадкового (нерегулярного) тренду, як показника коротко- та середньострокових змін;

– короткострокових змін, тобто, прогнозування коливань (відхилень), що накладаються на тренд;

– сезонних ефектів;

– приростів (швидкості) зміни процесу, які визначаються першими різницями;

– дисперсії або стандартного відхилення, як міри розсіювання процесу (наприклад, волатильність, яку часто використовують за міру ризику у інвестуванні або міру якості на виробництві);

– якісних змінних (за допомогою нечітких множин, мереж Байєса і т. ін.);

– комбінацій вказаних елементів процесів.

Відповідно до того, які складові процесу необхідно прогнозувати, ставиться задача побудови математичної, ймовірнісної (ймовірнісностатистичної) або логічної моделі, що має меті забезпечити високу якість прогнозу на заданому горизонті.

Тренд, сезонна і циклічна компоненти не є випадковими і називаються систематичними компонентами часового ряду. Складова частина часового ряду, що залишається пі­сля вилучення з нього систематичних компонент, являє собою випадкову компоненту (залишки, помилки) *εt*. Оскільки випадкові відхилення неминуче супроводжують будь-яке макроекономічне явище, випадкова компонента є обов’язковою складовою часового ряду і визначає стохастичний характер його елементів *уt*. Якщо побудована «якісна» модель прогнозування, то *εt* є близькою до нуля, випадковою, незалежною, нормально розподіленою компонентою, інакше модель вважається поганою.

Аналіз динаміки часового ряду містить такі послідовні зав­дання:

1. коригування рівнів динамічного ряду, якщо цього вимагають умови порівняльності;
2. визначення систематичних компонент динамічного ряду (функції *ft*, *st*, *ct*), які присутні у його розкладенні;
3. розрахунок оцінок тих функцій, які входять у розкладення часового ряду;
4. підбір моделі, яка адекватно описує поведінку випадкової компоненти *εt*, і статистичне оцінювання параметрів цієї моделі.

Тренд - загальна тенденція при різнонаправленому русі, визначена загальною спрямованістю змін показників часового ряду. Графіки можуть описуватись різними рівняннями - лінійними, логарифмічними, степеневими і т.д.. Фактичний тип графіка встановлюють за графічним зображенням даних часового ряду, шляхом усереднення показників часового ряду, на основі статистичної перевірки гіпотези про сталість параметрів графіка.Тренд — це тривала зміна рівня середнього випадкового процесу.

## Систематична складова і випадковий шум

Як і більшість інших видів аналізу, аналіз часових рядів передбачає, що дані містять систематичну складову (зазвичай включає кілька компонент) і випадковий шум (помилку), який ускладнює виявлення регулярних компонент. Більшість методів дослідження часових рядів включає різні способи фільтрації шуму, що дозволяють побачити регулярну складову більш чітко.

Часовий ряд — це послідовність впорядкованих у часі числових показників, що характеризують рівень стану і зміни досліджуваного явища.

## Класифікація часових рядів

Всякий часовий ряд включає два обов'язкових елемента: по-перше, час і, по-друге, конкретне значення показника, або рівень ряду.

Часові ряди розрізняються за такими ознаками: 1) за часом — моментні та інтервальні. Інтервальний ряд — послідовність, в якій рівень явища відносять до результату, накопиченому або знову виробленому за певний інтервал часу. Такі, наприклад, такі ряди показників обсягу продукції підприємства по місяцях року, кількості відпрацьованих людиною днів по окремих періодах (місяцях, кварталах, півріччях, роках, тощо) і т. д. Якщо ж рівень ряду характеризує досліджуване явище в конкретний момент часу, то сукупність рівнів формує моментний ряд.

## Забезпечення порівнянності рівнів часових рядів

Найважливішою умовою правильного формування часових рядів є порівнянність рівнів, що утворюють ряд. Рівні ряду, що підлягають вивченню, повинні бути однорідні за економічним змістом і враховувати сутність досліджуванного явища і його мету. Статистичні дані, представлені у вигляді часових рядів, повинні бути порівняні по території, колу охоплюваних об'єктів, одиницям вимірювання, моменту реєстрації, методикою розрахунку, цінами, достовірності. Неспівмірність за територією виникає в результаті змін кордонів країн, регіонів, господарств тощо Для приведення даних до порівнянного вигляду проводиться перерахунок колишніх даних з урахуванням нових кордонів. Цілісність охоплення різних частин явища — найважливіша умова порівнянності рівнів ряду. Вимога однакової цілісності охоплення різних частин досліджуваного об'єкта означає, що рівні ряду за окремі періоди повинні характеризувати різноманітність заходів того чи іншого явища по одному і тому ж колу, які входять до складу його частин. Наприклад, при характеристиці динаміки врожайності овочевих культур у регіоні по роках не можна в одні роки враховувати тільки сільськогосподарські підприємства, а в інші — всі категорії господарств, у тому числі присадибні ділянки сільських жителів і сади, городи городян. При визначенні порівнюваних рівнів ряду необхідно використовувати єдину методику їх розрахунку. Особливо часто ця проблема виникає при міжнародних порівняннях.

Неспівмірність показників, що виникає в силу неоднакового застосовуваних одиниць виміру, сама по собі очевидна. З різницею застосовуваних одиниць вимірювання доводиться зустрічатися при вивченні динаміки: виробничих ресурсів, коли вони представляються то у вартісному, то в трудовому обчисленні; енергетичних потужностей (кВт-год, л.с.); атмосферного тиску і т. д. Труднощі при порівнянні даних по моменту реєстрації виникають через сезонні явища. Чисельність худоби в домашніх господарствах через економічну доцільность відмінна взимку і влітку, тому рівні при порівнянні повинні ставитися до певної дати щорічно. При аналізі показників у вартісному вираженні слід враховувати, що з плином часу відбувається безперервна зміна цін. Причин цього процесу безліч — інфляція, зростання витрат, ринкові умови (попит і пропозиція) і т. д. У цьому зв'язку при характеристиці вартісних показників обсягів продукції в часі має бути усунуто вплив трансформаційних змін цін. Для вирішення цього завдання кількість продукції, виробленої в різні періоди, оцінюють в цінах одного періоду, які називають фіксованими або в визначеними в статистичних органах — порівнянними цінами.

Широке використання в статистичних дослідженнях вибіркового методу вимагає враховувати достовірність кількісних і якісних характеристик досліджуваних явищ в динаміці. Різна репрезентативність вибірки за періодами внесе суттєві похибки в величини рівнів ряду. Однією з умов порівнянності рівнів інтервального ряду, крім рівності періодів, за які наводять дані, є однорідність етапів, у межах яких показник підпорядковується одному закону розвитку. У цих випадках проводять періодизацію часових рядів, типологічну угруповання в часі. Всі вищеназвані обставини слід враховувати при підготовці інформації для аналізу змін явищ у часі (динаміці).

Зазвичай, метою прикладного статистичного аналізу часових рядів є побудова моделі ряду, за допомогою якої можна пояснити поведінку ряду і здійснити прогноз на майбутні періоди.

## Побудова і вивчення графіка

Аналіз часового ряду починається з побудови і вивчення його графіка. Якщо нестаціонарність часового ряду очевидна, то спочатку необхідно виокремити його нестаціонарну складову. Процес виокремлення [тренду](http://uk.wikipedia.org/wiki/%D0%A2%D1%80%D0%B5%D0%BD%D0%B4_(%D1%81%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D0%BA%D0%B0)) та інших компонент ряду, що призводять до порушення стаціонарності, може проходити в декілька етапів. На кожному з них розглядається ряд залишків, отриманий у результаті вирахування з вихідного ряду підібраної моделі тренду, або результат різницевих і інших перетворень ряду. Крім графіків, ознаками нестаціонарності часового ряду можуть служити автокореляційна функція, що прямує не до нуля (за винятком дуже великих значень [лагів](http://uk.wikipedia.org/wiki/%D0%9B%D0%B0%D0%B3_%D1%87%D0%B0%D1%81%D0%BE%D0%B2%D0%B8%D0%B9)) і наявність яскраво виражених піків на низьких частотах у періодограмі. За допомогою автокореляційної функції досліджують також внутрішні зв'язки між елементами часових рядів.

У вибіркових дослідженнях найпростіші числові характеристики описової статистики (середнє, медіана, дисперсія, стандартне відхилення, коефіцієнти асиметрії й ексцесу) звичайно дають достатньо інформативне уявлення про вибірку. Графічні методи зображення й аналізу вибірок при цьому грають лише допоміжну роль, дозволяючи краще зрозуміти локалізацію і концентрацію даних, їхній закон розподілу.

Роль графічних методів при аналізі часових рядів цілком інша. Табличне представлення часового ряду й описових статистик частіше за все не дозволяє зрозуміти характер процесу, у той час як за графіком часового ряду можна зробити досить багато висновків. Надалі вони можуть бути перевірені й уточнені за допомогою розрахунків.

Людське око досить упевнено визначає за графіком часового ряду:

1. наявність тренду і його характер;
2. наявність сезонних і циклічних компонент;
3. ступінь повільності або переривчастості змін послідовних значень ряду після усунення тренду (по цьому показнику можна судити про характер і розмір кореляції між сусідніми елементами ряду).

Так графічний аналіз ряду звичайно задає напрямок його подальшого аналізу.

## Вибір моделі для часового ряду

Після того, як вихідний процес максимально наближений до стаціонарного, можна приступити до вибору різноманітних моделей отриманого процесу. Мета цього етапу — опис і урахування надалі аналізу кореляційної структури аналізованого процесу. Модель може вважатися підібраною, якщо залишкова компонента ряду є процесом типу, як правило, «білого шуму». Після підбору залишки аналізуються для перевірки адекватності моделі та побудови надійних інтервалів.

## Прогнозування або інтерполяція

Останнім етапом аналізу часового ряду може бути прогнозування його майбутніх (екстраполяція) або відновлення пропущених (інтерполяція) значень і визначення точності цього прогнозу на базі підібраної моделі. Добре підібрати математичну модель вдається не для всякого часового ряду. Нерідко буває і так, що для опису підходять відразу декілька моделей. Неоднозначність вибору моделі може спостерігатися як на етапі виділення детермінованого компонента ряду, так і при виборі структури ряду залишків. Тому досить часто розробляють декілька прогнозів, зроблених за допомогою різних моделей.

## Методи згладжування часових рядів

виокремлюються у дві основні групи:

1) механічне згладжування окремих рівнів часового ряду, яке не потребує знань про аналітичний вид згладженої функції;

2) аналітичне згладжування з використанням кривої, проведеної між певними рівнями ряду так, щоб вона відбивала тенденцію, притаманну ряду, і одночасно позбавляла його незначних коливань.

*Механічні* *методи* *згладжування* *часових* *рядів* використовують фактичні значення сусідніх рівнів ряду і не досліджують аналітичний вид згладженої функції. До механічних методів належать: згладжування по двох точках, метод простої ковзкої середньої, метод зваженої ковзкої середньої, метод експоненційного згладжування.

*Аналітичні* *методи* *згладжування* *часових* *рядів* ґрунтуються на припущенні, що відомий загальний вигляд невипадкової складової часового ряду. Вони реалізуються за допомогою регресійних та адаптивних методів.

Регресійні методи є основою побудови кривих зростання*.* Щоб правильно підібрати найкращу криву для моделювання і прогнозування економічного явища, необхідно знати особливості кожного виду кривих (табл. 2.2.1 [11]). Універсальним методом попереднього вибору кривих зростання, який дає можливість вибрати криву із широкого класу, є метод характеристик приросту. Він заснований на використанні окремих характерних властивостей кривих. При цьому методі вхідний часовий ряд поперед­ньо згладжується методом простої ковзкої середньої.

Адаптивні методи прогнозування застосовуються в ситуації зміни зовнішніх умов, коли найбільш важливими стають останні реалізації досліджуваного процесу. Загальна схема побудови адаптивних методів може бути подана так:

1) за кількома першими рівнями ряду будується модель і оцінюються її параметри;

2) на основі побудованої моделі розраховується прогноз на один крок вперед, причому його відхилення від фактичного рівня ряду розцінюється як помилка прогнозування, яка враховується відповідно до прийнятої схеми коригування моделі;

3) за моделлю з відкоригованими параметрами розраховується прогнозна оцінка на наступний момент часу тощо.

Таким чином, модель постійно вбирає в себе нову інформацію і до кінця періоду навчання відбиває тенденцію розвитку процесу, що існує на даний момент. Прогноз отримується як екстраполяція останньої тенденції. Численні адаптивні методи відрізняються один від одного лише способами числової оцінки параметрів моделі і визначення параметрів адаптації. Базовими адаптивними методами вважаються методи Хольта, Брауна і Хольта—Уїнтерса.

*Методи* *фільтрації* *сезонної* *компоненти* *st* . Проблема аналізу сезонності (та/або циклічності) полягає у дослідженні сезонних коливань і у вивченні зовнішнього циклічного механізму, що їх породжує. Для дослідження суто сезонних коливань необхідно виокремити з часового ряду *уt* сезонну компоненту *st* і потім аналізувати її динаміку. Більшість методів фільтрації побудовано таким чином, що попередньо виокремлюється тренд, а потім сезонна компонента. Тренд у чистому вигляді необхідний і для аналізу динаміки сезонної хвилі. Оскільки індекси сезонності сезонної хвилі величини безрозмірні і не змінюються з року в рік, то їх можна використовувати для визначення рівня сезонності у часовому ряду. У разі використання квартальних даних їх буде чотири, а місячних спостережень — 12.

Для повного дослідження тренд-сезонного часового ряду потрібно розв’язати сукупність завдань у такій послідовності:

1) визначення наявності тренду і визначення ступеня його гладкості;

2) виявлення наявності у часовому ряду сезонних коливань;

3) здійснення фільтрації сезонної компоненти у разі підтвердження сезонного процесу;

4) проведення аналізу динаміки (еволюції) сезонної хвилі;

5) дослідження чинників, які визначають сезонні коливання;

6) розроблення прогнозу тренд-сезонного процесу.

Найбільш поширеними методами фільтрації є ітераційні та гар­монічного аналізу.

### Методи прогнозування випадкових компонент.

На відміну від прогнозів, які, наприклад, послуговуються класичною регресійною моделлю, у прогнозі часових рядів суттєво використовуються взаємозалежність і прогноз самих випадкових залишків. Отже, мова йде про моделювання не самих часових рядів, а лише їх випадкових залишків. Для описання поведінки випадкових залишків *εt* і прогнозування їх значень використовується клас стаціонарних часових рядів, для яких розроблені спеціальні лінійні параметричні моделі, такі як авторегресійні (AR), ковзної середньої (MA) та ARMA. Однак реальні часові ряди, що зустрічаються в макроекономіці, є у багатьох випадках нестаціонарними. Їх нестаціонарність частіше за все виявляється в наявності невипадкової складової *ft*. У таких випадках йдеться про нестаціонарні однорідні часові ряди. Отже, нестаціонарний однорідний часовий ряд *yt* може бути перетворений у стаціонарний часовий ряд процедурою віднімання від ряду *yt* його невипадкової складової *ft*. Для моделювання нестаціонарних часових рядів з означеними властивостями використовується ARIMA-модель (авторегресійна інтегрована модель ковзної середньої), або модель Бокса—Дженкінса. При цьому AR, MA та ARMA моделі являють собою окремі випадки ARIMA-моделі.

## Висновки

Ознайомився з аналізом і прогнозуванням часових рядів.