KPI

Міністерство освіти та науки

Національний технічний університет України

«Київський політехнічний інститут»

Фізико-технічний інститут

**Лабороторна робота №2**

на тему:

## «Виділення тренда та прогнозування часового ряду. Статистика Дарбіна-Уотсона»

з предмету

«**Аналіз даних та статистична обробка сигналів**»

**Виконала:**

студентка 6 курсу ФТІ

групи ФІ-31м

Донченко Валентина

**Перевірив:**

Архипов О.Є.

**Київ 2014**

1. **Теоретичні відомості**

Поліноміальна трендова модель (41) фактично є рівнянням множинної регресії, тому для її ідентифікації цілком застосовні розглянуті в першій частині цього видання методи та процедури регресійного аналізу. Зокрема розширена матриця даних для моделі (41) матиме вигляд:

, (42)

а вектор МНК-оцінок параметрів визначатиметься співвідношенням:

. (43)

Проте на етапах вибору варіанту ММ та її верифікації виникає, завдяки специфічним особливостям часових рядів, можливість використання додаткових способів оцінювання якості варіантів ММ. Найчастіше це стосується аналізу нев’язок (залишків, похибок) ,  трендової моделі, де  – оцінка значення *t*-ого рівня часового ряду, розрахована за поліноміальною трендовою моделлю (41).

Зокрема, зважаючи на характер обмежень, що накладаються на властивості випадкової складової *et* рівня часового ряду, а саме, на припущення щодо некорельованості її значень:



видається доцільним перевірити, як виконується це обмеження для нев’язок , що за певних обставин сприймаються у якості оцінок відповідних значень випадкової складової *Е*. Найбільш відомим і поширеним тестом перевірки моделі на наявність автокореляції між нев’язками є *критерій Дарбіна-Уотсона*. Критерій застосовується у тих випадках, коли тренд визначається для всього ряду.

Змістовну суть критерію можна пояснити наступним чином. У разі правильної специфікації трендової моделі за даними всього часового ряду, тобто коли структура функції тренду адекватно відображає динаміку змін значень рівнів часового ряду на інтервалі його спостереження, послідовність нев’язок  дійсно слід вважати оцінками значень випадкової складової *Е*. Очевидно, що в цьому випадку автокорельованість нев’язок має бути мінімальною. У протилежному випадку, коли отримана трендова модель є недостатньо точною, в нев’язках присутня, крім випадкової складової , частка детермінованої складової  часового ряду, яка має високий рівень автокореляції. Тому, наприклад, серія позитивних або негативних знаків у послідовності нев’язок є своєрідним індикатором того, що побудована модель неадекватна вихідним даним. Більш вагомим статистичним підтвердженням цього факту є існування автокореляцій у серії нев’язок, одна з основних причин появи якої – наявність в них залишкового тренду.

Значення статистики Дарбіна-Уотсона розраховується за формулою:

. (44)

З аналізу формули (44) витікає, що якщо між значеннями  присутня сильна позитивна автокореляція , де  – вибірковий коефіцієнт автокореляції між двома послідовними нев’язками часового ряду, то величина *d≈*0, якщо маємо сильну негативну автокореляцію , то *d≈*4. При відсутності автокореляції , статистика *d≈*2. Таким чином, значення *d*-статистики Дарбіна-Уотсона знаходяться у межах від 0 до 4 (). На рис. 3 зображено схему розташування граничних значень *d*-статистики:

позитивна зона автокореляція зона негативна

кореляція невизнач. відсутня невизнач. кореляція

0 *d*1 *d*2 2 4– *d*2 4– *d*1 4

**Рис.3. Зони розташування граничних значень *d*-статистики**

Крім формули (44) для приблизного розрахунку статистики Дарбіна-Уотсона можна використати простішу формулу:

. (45)

Співвідношення (45) дає найбільш прозору інтерпретацію змісту критерію Дарбіна-Уотсона та його кількісним характеристикам.

Практичне використання критерію Дарбіна-Уотсона засновано на порівнянні величини *d*, розрахованої за формулою (44), з теоретичними значеннями *d*1та  *d*2, які показані на рис.3:

1)при  – приймається гіпотеза про відсутність автокореляції між значеннями випадкової величини;

2)при  – виявляється позитивна автокореляція;

3)при  – виявляється негативна автокореляція;

4)при  або  – немає достатньої інформації для прийняття рішень, тобто отримуємо зону невизначеності.

Теоретичні граничні значення *d*1та *d*2 для рівня значущості 5% і різних степенів поліному *p* представлено у табл. 2 (більш повні дані наведено у додатках в [3,4]).

**Таблиця 2. Граничні значення *d*1та  *d*2 ( *р* – степінь поліному, n – об’єм вибірки)**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *n* | *p*=1 | | *p*=2 | | *p*=3 | | *p*=4 | |
|  | *d*1 | *d*2 | *d*1 | *d*2 | *d*1 | *d*2 | *d*1 | *d*2 |
| 30 | 1,35 | 1,49 | 1,28 | 1,57 | 1,21 | 1,65 | 1,14 | 1,74 |
| 40 | 1,44 | 1,54 | 1,39 | 1,6 | 1,34 | 1,66 | 1,29 | 1,72 |

Для визначення за відомою (вже ідентифікованою) поліноміальною моделлю виду (41) прогнозних значень тренду достатньо ввести до цієї моделі замість змінної *t* її кількісне значення *l*, в якому враховано інтервал упередження *L*. Так, якщо цей інтервал дорівнює одному, двом, трьом,... відлікам (крокам), тобто *L*=1,2,3,..., то значення змінної *t* кількісно дорівнюватиме *l= n + L*.

В матричній формі оцінка прогнозу тренду визначається виразом

, (46)

де  – вектор значень регресорів моделі (41) у точці прогнозу.

Помилка отриманої оцінки трендового прогнозу  матиме (за умови адекватності структури моделі вихідним даним та обчисленню коефіцієнтів поліноміальної моделі методом найменших квадратів) дисперсію

. (47)

Слід зазначити, що процедура виділення тренду суттєво ускладнюється в разі наявності в даних часового ряду циклічного компоненту (особливо в коротких чи середніх за обсягом рядах). Зокрема, помилкове включення рівнів, у складі яких присутній циклічний компонент, до сукупності даних, що використовується для ідентифікації трендової моделі, призводить до створення трендової залежності і робить неможливим її використання для прогнозу тренду часового ряду. В цьому випадку рекомендується проаналізувати одночасну сукупну зміну у часі трендового та циклічного компонентів, відділивши від них шумову складову ,  шляхом згладжування даних часового ряду.

1. **Завдання:**
2. Побудувати за допомогою *Майстра діаграм* графік вихідних даних.
3. У поле діаграми викликати контекстне меню для елемента *Ряд даних*, обрати команду *Додати лінію тренду*.
4. У вікні *Лінія тренду* вибрати *Лінійний* вид тренду.
5. На вкладці *Параметри* виправити назву лінії тренду та відмітити:

🗹 Показати рівняння тренду;

🗹 Додати коефіцієнт детермінації .

1. Використовуючи параметри рівняння тренду (на діаграмі), розрахувати модельні значення для лінійного тренду.
2. На цю ж діаграму додати *Логарифмічний*, *Степеневий* та *Експоненціальний*  види тренду.
3. Побудувати новий графік вихідного часового ряду й додати на одну діаграму три поліноміальні тренди для степені поліному 
4. Для трьох поліномів на вкладці *Параметри* відмітити *Прогноз наперед* на 3 періоди.
5. Побудувати новий графік вихідного часового ряду й додати на одну діаграму три тренди *Лінійна фільтрація* на 2, 4 и 10 точок.
6. Розрахувати для кожного виду тренду ряд модельних значень.
7. Розрахувати для кожного виду тренду суму квадратів нев’язок.
8. За допомогою функції ЛИНЕЙН оцінити коефіцієнти  для тренду 4-го степеня:

 .

1. Розрахувати суму квадратів нев’язок для цієї моделі.
2. Вибрати кращий тип тренду для апроксимації вихідних даних  за коефіцієнтом детермінації  і за сумою квадратів нев’язок.
3. Застосувати статистику Дарбіна-Уотсона для виявлення автокореляції в значеннях нев’язки.
4. Вибрати кращий вид тренду за статисткою Дарбіна-Уотсона.

**3.Виконання роботи:**

Графік вихідного часового ряду з трьома поліноміальними трендами для степені поліному 2, 4, 6 (із застосуванням прогнозу наперед на 3 періоди):

Графік вихідного часового ряду з трендами *Лінійна фільтрація* на 2, 4 и 10 точок:

Модельні значення для трендів 2, 4 та 6 степенів:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | a2 | a1 | a0 |  |  |  |  |
| a | -0,003961 | 0,45821 | -0,046213 |  |  |  |  |
| Sa | 0,002022 | 0,08550 | 0,760141 |  |  |  |  |
| R^2, Sz | 0,847017 | 1,52305 |  |  |  |  |  |
| F, k | 102,428 | 37 |  |  |  |  |  |
| Sreg, Sost | 475,2099 | 85,8293 |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  | a4 | a3 | a2 | a1 | a0 |  |  |
| a | 1,81578E-06 | 0,00066 | -0,050045 | 1,251938 | -2,97416 |  |  |
| Sa | 1,5238E-05 | 0,00125 | 0,034550 | 0,354459 | 1,077259 |  |  |
| R^2, Sz | 0,915499 | 1,16383 |  |  |  |  |  |
| F, k | 94,79970 | 35 |  |  |  |  |  |
| Sreg, Sost | 513,6311 | 47,4080 |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  | a6 | a5 | a4 | a3 | a2 | a1 | a0 |
| a | 2,46268E-07 | -3,6339E-05 | 0,00201 | -0,053565 | 0,63996 | -2,4177 | 2,342045 |
| Sa | 1,09831E-07 | 1,3553E-05 | 0,00064 | 0,0146563 | 0,16388 | 0,80801 | 1,278509 |
| R^2, Sz | 0,959489 | 0,82989 |  |  |  |  |  |
| F, k | 130,2684 | 33 |  |  |  |  |  |
| Sreg, Sost | 538,3114 | 22,7277 |  |  |  |  |  |

Застосовуємо статистику Дарбіна-Уотсона для виявлення автокореляції в значеннях нев’язки:

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **y2** | **Ei = z-y2** | **(z-y2)^2** | **(Ei-Ei-1)^2** |  | **y4** | **Ei=z-y4** | **(z-y4)^2** | **(Ei-Ei-1)^2** |  | **y6** | **Ei=z-y6** | **(z-y6)^2** | **(Ei-Ei-1)^2** |
| 0,41 | -0,86 | 0,741 | 0,000 |  | -1,77 | 1,32 | 1,739 | 0,000 |  | 0,51 | -0,97 | 0,932 | 0,000 |
| 0,85 | 0,03 | 0,001 | 0,789 |  | -0,67 | 1,55 | 2,392 | 0,052 |  | -0,33 | 1,21 | 1,469 | 4,742 |
| 1,29 | -1,81 | 3,284 | 3,382 |  | 0,35 | -0,87 | 0,755 | 5,833 |  | -0,44 | -0,08 | 0,006 | 1,665 |
| … | … | … | … |  | … | … | … | … |  | … | … | … | … |
| … | … | … | … |  | … | … | … | … |  | … | … | … | … |
| 11,65 | 0,87 | 0,751 | 0,421 |  | 12,60 | -0,09 | 0,007 | 1,379 |  | 12,65 | -0,14 | 0,020 | 0,667 |
| 11,80 | 1,08 | 1,173 | 0,047 |  | 13,37 | -0,49 | 0,237 | 0,160 |  | 12,94 | -0,06 | 0,003 | 0,007 |
| 11,94 | 1,31 | 1,709 | 0,050 |  | 14,23 | -0,97 | 0,950 | 0,239 |  | 13,19 | 0,06 | 0,003 | 0,013 |
|  | **Sum:** | **85,829** | **52,540** |  |  | **Sum:** | **47,408** | **52,612** |  |  | **Sum:** | **22,728** | **50,909** |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  |  | **DW:** | **1,6336** |  |  |  | **DW:** | **0,9011** |  |  |  | **DW:** | **0,4464** |

**Висновки**

За результатами виконаної роботи було отримано, що за статисткою Дарбіна-Уотсона кращим є тренд для степені поліному 2.

У ході виконання роботи ми навчились виділенню тренду та прогнозуванню часового ряду з використанням можливостей Excel.