Методи згладжування

К.е.н., доц. Ставицький А.В.

www.andriystav.cc.ua

План

1. Класичні підходи
2. Експоненціальне згладжування
3. Фільтр Ходріка-Прескотта
4. Виділення сезонності

Методи згладжування

Методи згладжування використовуються для зменшення впливу випадкового компонента (випадкових коливань) у часових рядах. Вони дають можливість отримувати більш “чисті” значення, які складаються лише з детермінованих компонентів. Одні з методів направлені на виділення деяких компонентів, наприклад, тренду.

1. КЛАСИЧНІ ПІДХОДИ

Метод усереднення - 1

Цей метод є одним з найпростіших, який дозволяє виділити тренд. Для застосування цього методу дослідник повинен мати доволі довгий ряд спостережень. Формально метод описується виразом:

$$\tilde{y}_t = \frac{1}{k} \sum_{j=-k_1}^{k_2} y_{t+j}, \quad k = k_1 + k_2 + 1.$$

Метод усереднення - 2

Для квартальних даних часового ряду формула набуває вигляду

$$\tilde{y}_t = \frac{1}{4}\sum_{j=-2}^{1} y_{t+j} = \frac{1}{4}\left(y_{t-2} + y_{t-1} + y_t + y_{t+1}\right).$$

Метод усереднення - 3

За допомогою цього методу можна не тільки більш чітко спостерігати трендовий компонент, але й сезонні та випадкові коливання:

$$\frac{y_t}{\tilde{y}_t} = \frac{tr_t \cdot c_t \cdot s_t \cdot r_t}{tr_t \cdot c_t} = s_t \cdot r_t, \ t = \overline{1, T-k}.$$

Приклад (умова)

Менеджер фірми, яка виробляє програмне забезпечення, отримує щомісяця скарги на продукцію фірми. Йому необхідно визначити, чи існує тенденція зміни у кількості скарг.

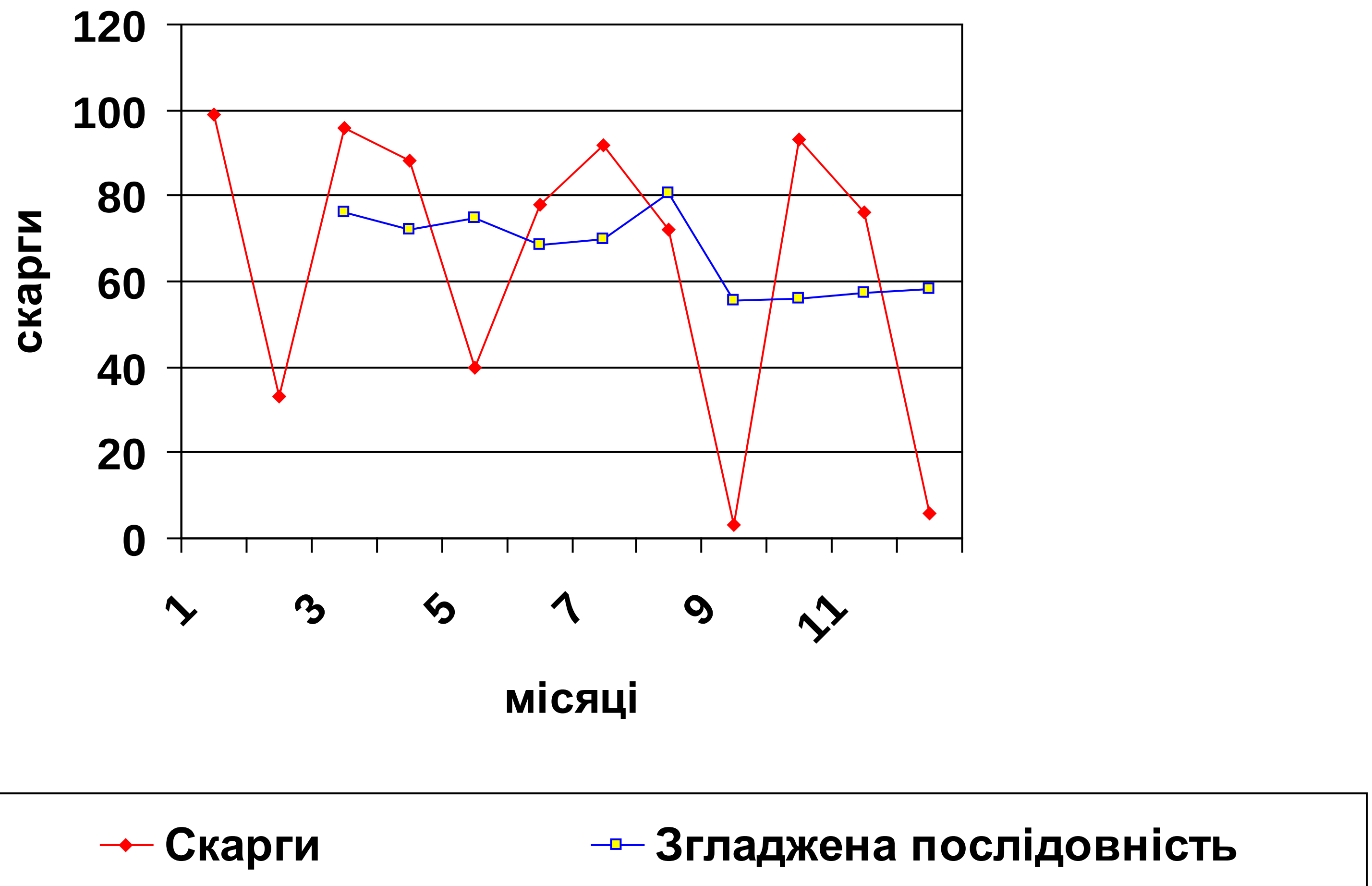
Приклад (дані)

Місяць	Скарги
1	99
2	33
3	96
4	88
5	40
6	78
7	92
8	72
9	3
10	93
11	76
12	6

Приклад (розрахунки)

Місяць	Скарги	Згладжена послідовність
1	99	#Н/Д
2	33	#Н/Д
3	96	76,0
4	88	72,3
5	40	74,7
6	78	68,7
7	92	70,0
8	72	80,7
9	3	55,7
10	93	56,0
11	76	57,3
12	6	58,3

Приклад (графік)



Подвійне усереднення

Цей метод двічі використовує усереднення часового ряду. При цьому кількість спостережень зменшується на два повних цикли сезонності, тому для використання методу необхідно мати часовий ряд, який складається щонайменше з 3-х повних циклів сезонності.

Взяття різниць

Нова послідовність будується за правилом

$$\tilde{y}_t = y_t - y_{t-k},$$

причому *k* визначається завдяки аналізу попередніх значень при порівнянні помилок прогнозування

Річна процентна різниця

Прогнозування здійснюється завдяки порівнянню даних минулого року з даними позаминулого року:

$$\tilde{y}_t = y_{t-1} \frac{y_{t-1-s}}{y_{t-1-2s}}, \quad t = \overline{2s+1, \ T}.$$

Цей метод також потребує знання даних за 2 роки до обчислень.

2. МЕТОДИ ЕКСПОНЕНЦІАЛЬНОГО ЗГЛАДЖУВАННЯ

Звичайне експоненціальне згладжування

Цей метод значно переважає усі попередні моделі. Найкраще цей метод зарекомендував себе, коли дані мають дуже гладкий, або навіть горизонтальний тренд. Нова послідовність будується за правилом:

$$S_t = \alpha y_t + (1-\alpha)S_{t-1}, \quad 0 < \alpha < 1.$$

Вибір початкового значення

Початкове значення послідовності можна вибрати як

$$S_1 = y_1, \qquad \text{або} \qquad S_1 = \bar{y}.$$

Вибір константи - 1

Єдина вага α може обиратися кількома шляхами.

По-перше, якщо обирається значення близьке до 1, то будуть більш важливими при прогнозуванні останні дані часового ряду, при виборі близьким до 0, більш впливовими будуть минулі значення.

По-друге, можна покласти $$\alpha = \frac{2}{T+1}.$$

Вибір константи - 2

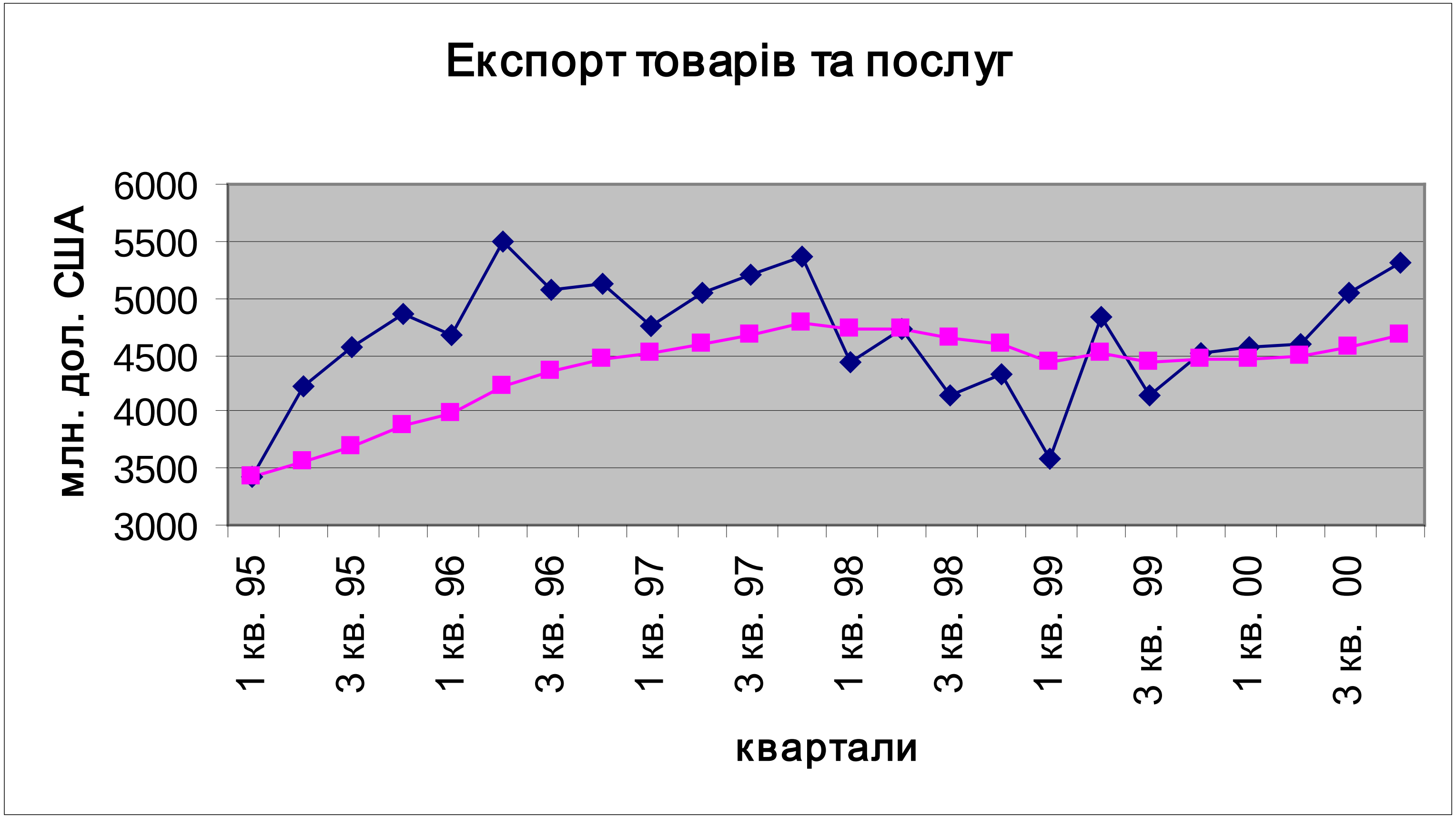
По-третє – вибір такого значення, при якому мінімізується один з критеріїв точності прогнозів.

Прогноз

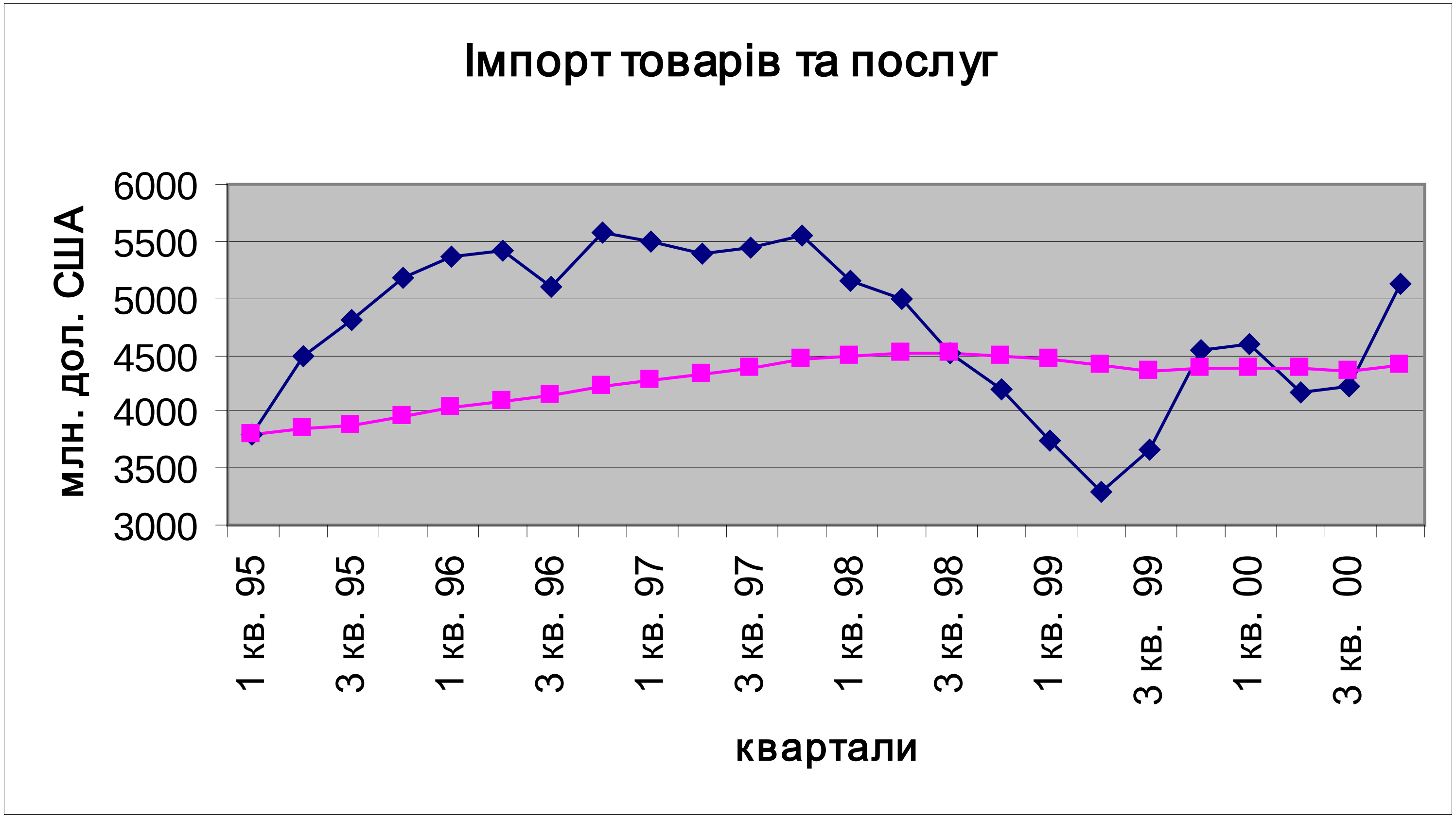
Прогноз значень часового ряду дорівнює останньому члену послідовності

$$\hat{y}_{T+p} = S_T, \qquad p = 1,\ 2,\ \ldots.$$

Приклад - 1


Експорт товарів та послуг
6000
5500
5000
4500
4000
3500
3000
млн. дол. США
1 кв. 95
3 кв. 95
1 кв. 96
3 кв. 96
1 кв. 97
3 кв. 97
1 кв. 98
3 кв. 98
1 кв. 99
3 кв. 99
1 кв. 00
3 кв. 00
квартали

Приклад - 2


Імпорт товарів та послуг
млн. дол. США
6000
5500
5000
4500
4000
3500
3000
1 кв. 95
3 кв. 95
1 кв. 96
3 кв. 96
1 кв. 97
3 кв. 97
1 кв. 98
3 кв. 98
1 кв. 99
3 кв. 99
1 кв. 00
3 кв. 00
квартали

Подвійне експоненціальне згладжування Брауна

Цей метод будується аналогічно попередньому, тільки процес згладжування робиться двічі:

$$S_t^{/} = \alpha y_t + (1-\alpha)S_{t-1}^{/},$$

$$S_t^{//} = \alpha S_t^{/} + (1-\alpha)S_{t-1}^{//}, \quad 0 < \alpha < 1.$$

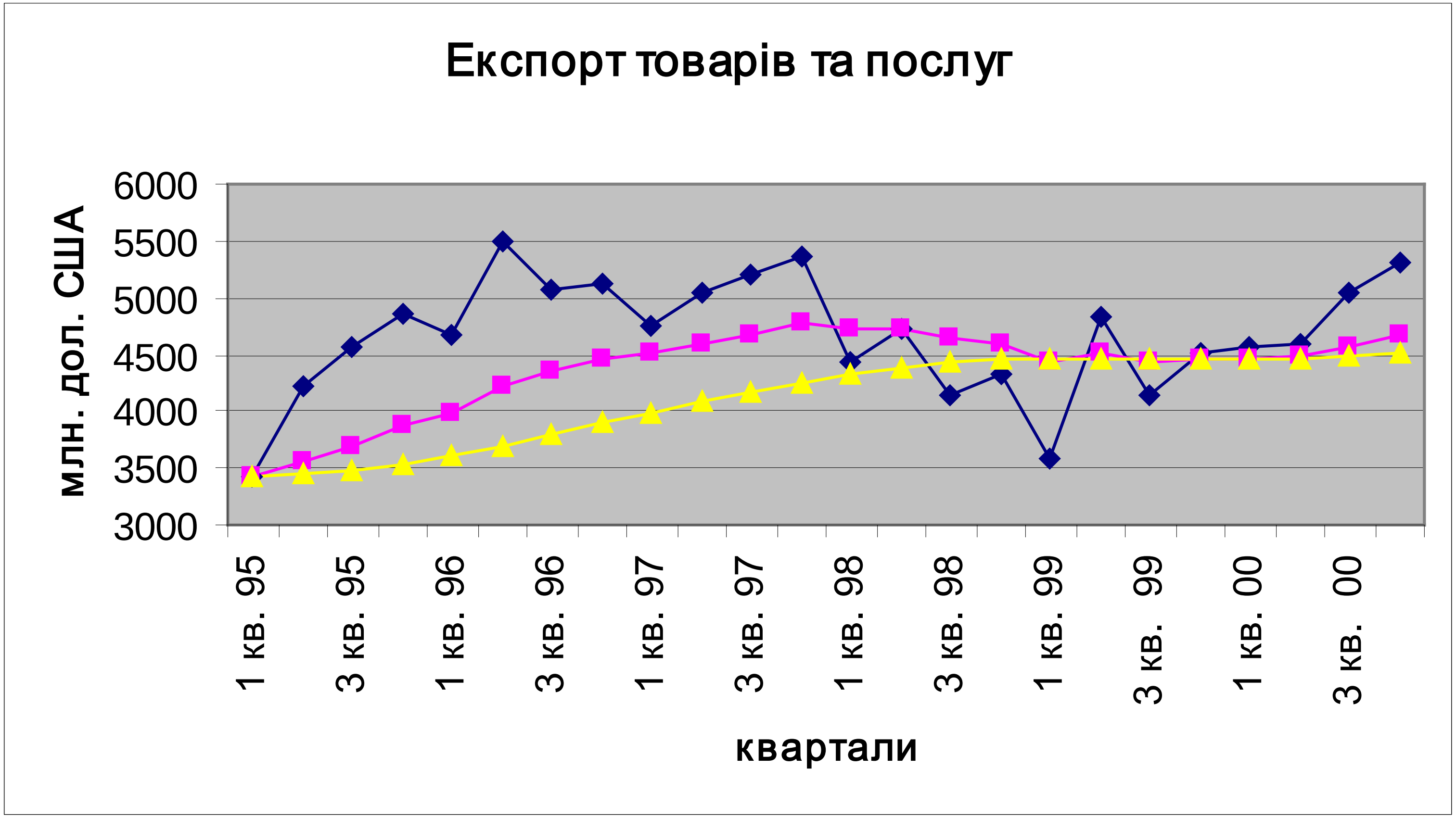
Метод використовується, коли дані часового ряду мають тренд

Прогноз

Прогноз будується як останнє значення другої послідовності:

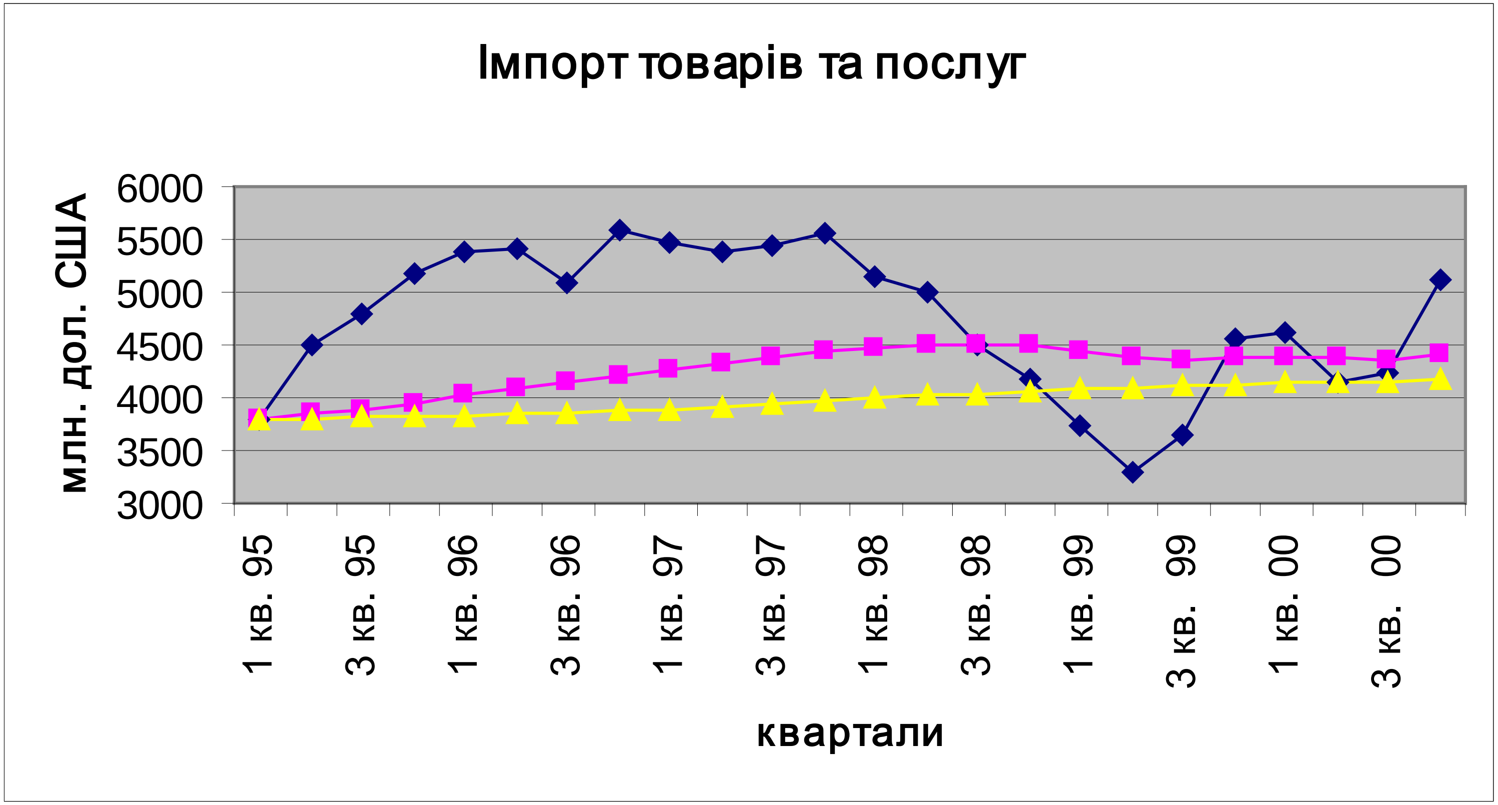
$$\hat{y}_{T+p} = S_T^{//}, \qquad p = 1, \quad 2, \quad \ldots .$$

Приклад - 1



Експорт товарів та послуг
млн. дол. США
6000
5500
5000
4500
4000
3500
3000
1 кв. 95
3 кв. 95
1 кв. 96
3 кв. 96
1 кв. 97
3 кв. 97
1 кв. 98
3 кв. 98
1 кв. 99
3 кв. 99
1 кв. 00
3 кв. 00
квартали

Приклад - 2



Імпорт товарів та послуг
млн. дол. США
6000
5500
5000
4500
4000
3500
3000
1 кв. 95
3 кв. 95
1 кв. 96
3 кв. 96
1 кв. 97
3 кв. 97
1 кв. 98
3 кв. 98
1 кв. 99
3 кв. 99
1 кв. 00
3 кв. 00
квартали

Потрійне експоненціальне згладжування Брауна

Цей метод аналогічний двом попереднім, тільки згладжування проводиться тричі. Це дозволяє прогнозувати нестаціонарні часові ряди з великими перепадами мінімального та максимального значень. Нові послідовності будуються за правилом:

$$S_t^{/} = \alpha y_t + (1-\alpha) S_{t-1}^{/},$$

$$S_t^{//} = \alpha S_t^{/} + (1-\alpha) S_{t-1}^{//},$$

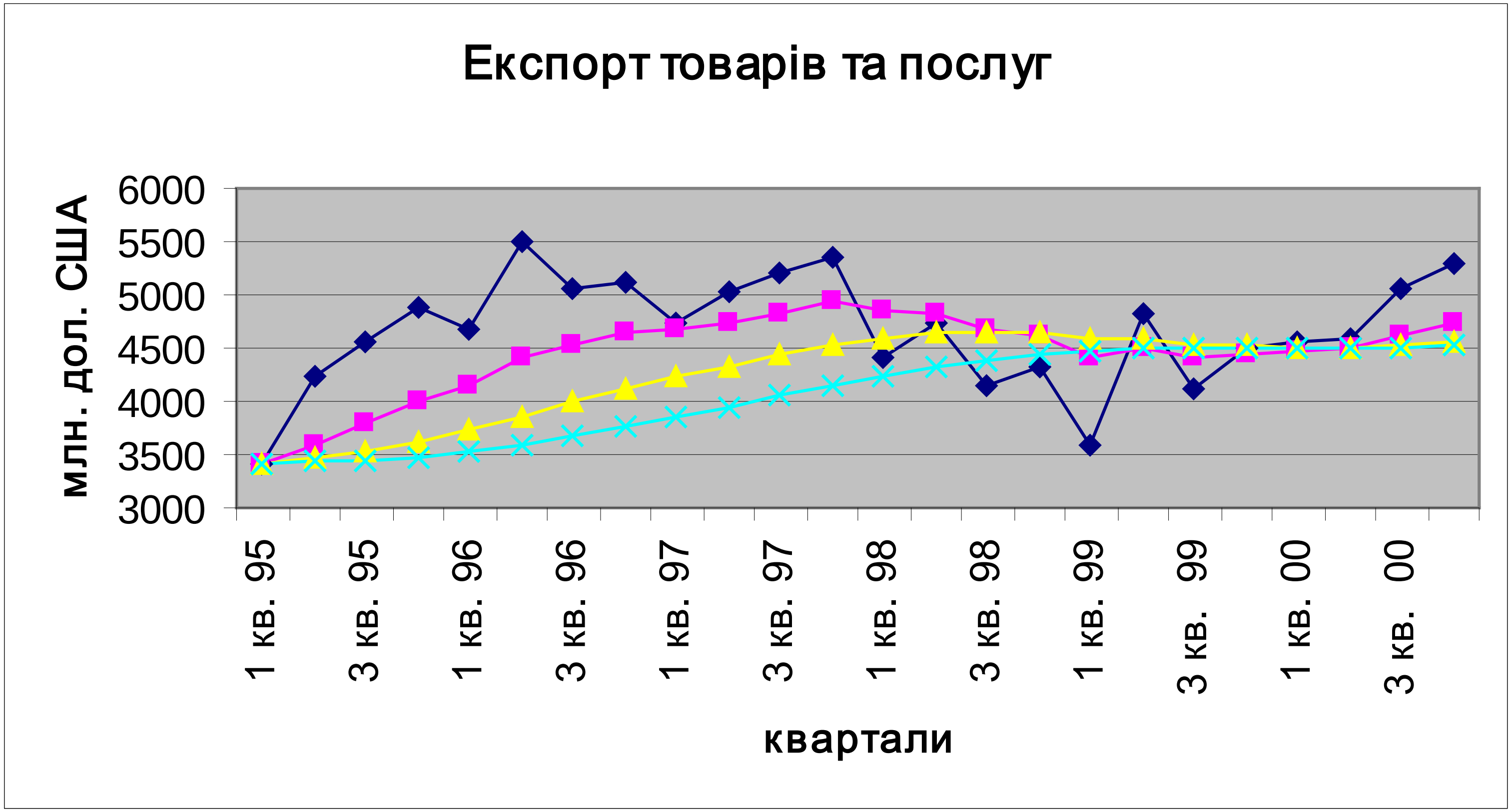
$$S_t^{///} = \alpha S_t^{//} + (1-\alpha) S_{t-1}^{///}, \quad 0 < \alpha < 1.$$

Прогноз

Прогноз на наступні періоди має вигляд:

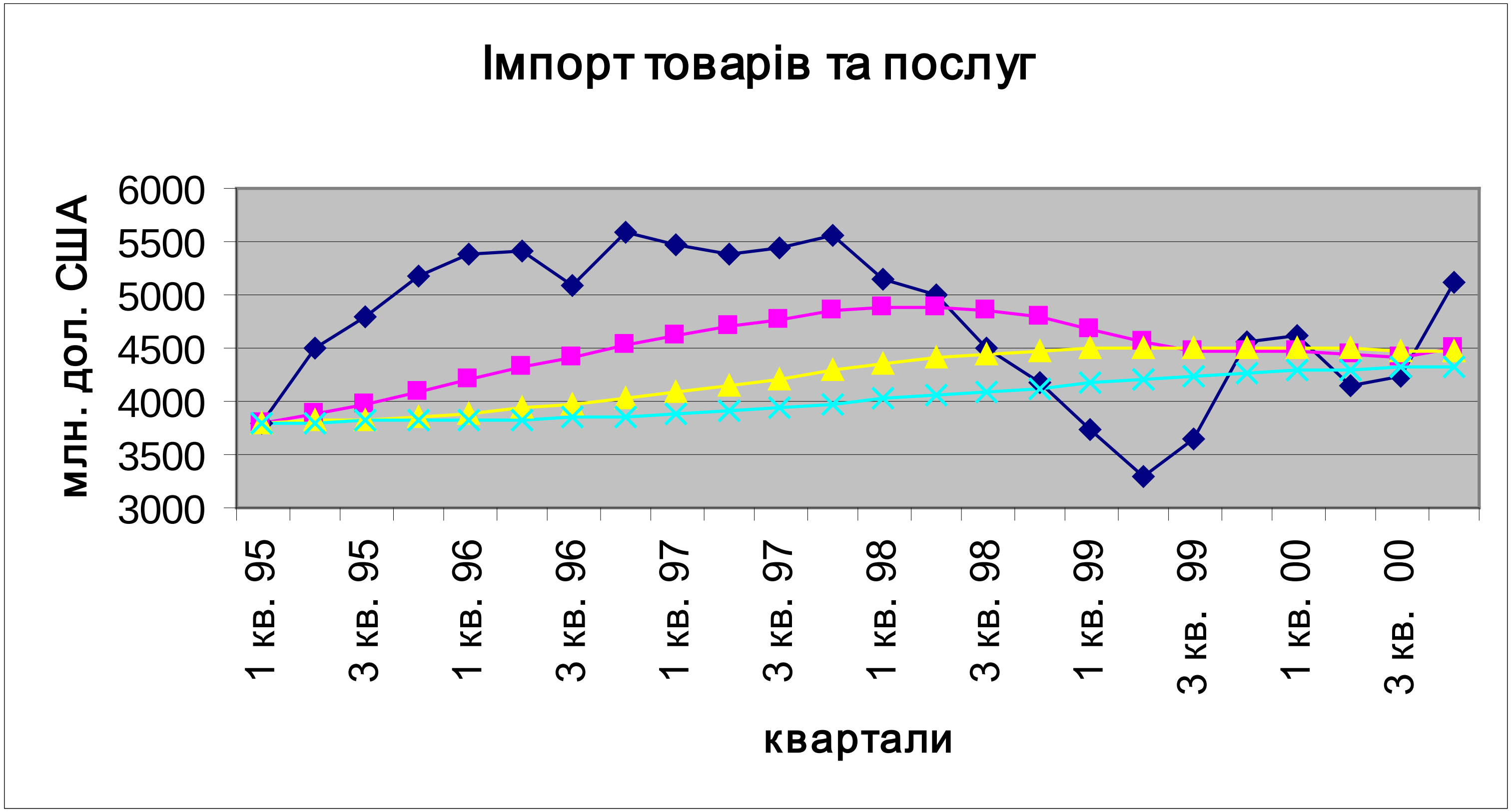
$$\hat{y}_{T+p} = S_T^{///}, \qquad p = 1, \quad 2, \quad \ldots.$$

Приклад - 1



Експорт товарів та послуг
млн. дол. США
6000
5500
5000
4500
4000
3500
3000
1 кв. 95
3 кв. 95
1 кв. 96
3 кв. 96
1 кв. 97
3 кв. 97
1 кв. 98
3 кв. 98
1 кв. 99
3 кв. 99
1 кв. 00
3 кв. 00
квартали

Приклад - 2



Імпорт товарів та послуг
6000
5500
5000
4500
4000
3500
3000
млн. дол. США
1 кв. 95
3 кв. 95
1 кв. 96
3 кв. 96
1 кв. 97
3 кв. 97
1 кв. 98
3 кв. 98
1 кв. 99
3 кв. 99
1 кв. 00
3 кв. 00
квартали

Адаптивне згладжування

Цей метод дозволяє автоматично змінювати константу згладжування в процесі обрахунку. Нова послідовність будується за правилом

,

$$S_{t+1} = \alpha_t y_t + (1 - \alpha_t) S_t$$

.

де α змінюється з часом залежно від похибки прогнозування

Вибір константи

$$\alpha_t = \left| \frac{E_t}{M_t} \right|$$

$$E_t = \beta\left(y_t - \hat{y}_t\right) + \left(1 - \beta\right)E_{t-1}$$

$$M_t = \beta\left|y_t - \hat{y}_t\right| + \left(1 - \beta\right)M_{t-1}$$

Параметр β знаходиться у межах від 0 до 1. Для випадкових похибок коефіцієнт α буде близьким до 0.5.

Несезонна модель Холта-Вінтерса

Ця модель схожа на подвійне експоненціальне згладжування, але дозволяє виділяти трендовий компонент за допомогою другої послідовності:

$$S_2^{/} = y_2, \quad S_2^{//} = y_2 - y_1,$$

$$S_t^{/} = \alpha y_t + (1-\alpha)\left(S_{t-1}^{/} + S_{t-1}^{//}\right), \quad 0 < \alpha < 1,$$

$$S_t^{//} = \beta\left(S_t^{/} - S_{t-1}^{/}\right) + (1-\beta) S_{t-1}^{//}, \quad 0 < \beta < 1.$$

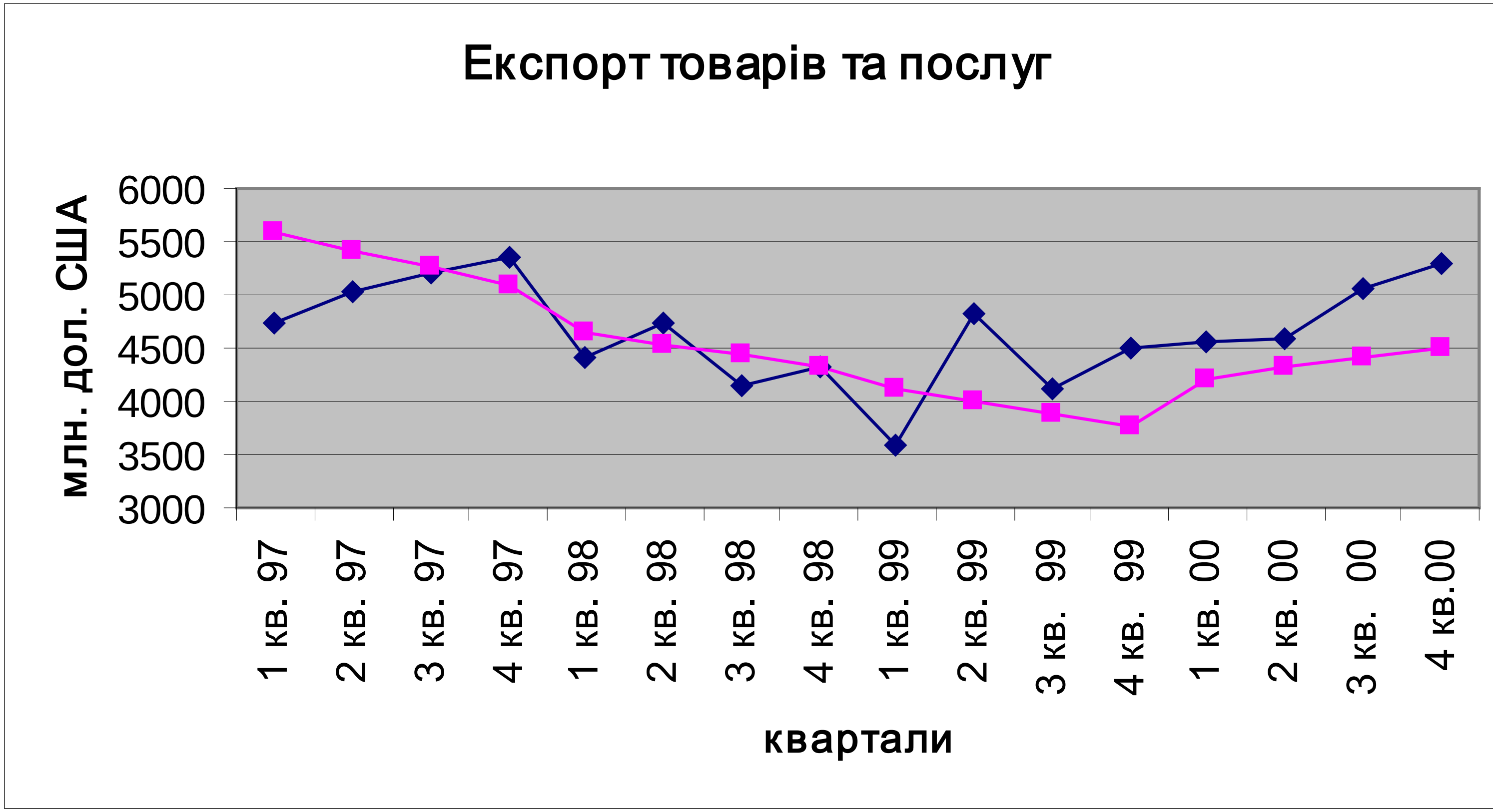
Прогноз

Прогноз на наступні періоди:

$$\hat{y}_{T+p} = S_T^{I} + pS_T^{II}, \qquad p = 1, \quad 2, \quad \ldots.$$

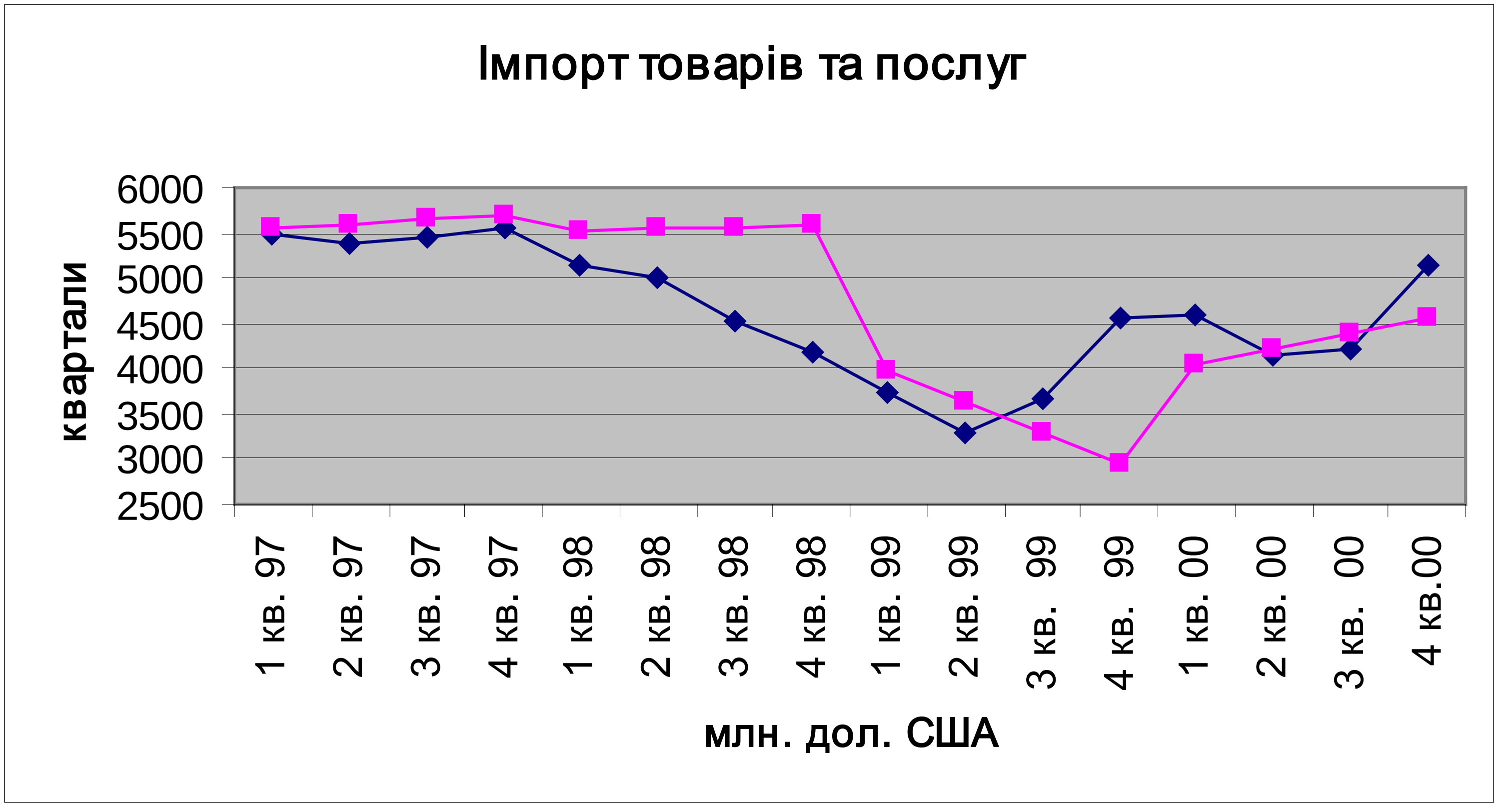
Слід відмітити, що прогнози, зроблені за цим методом, як правило, або сильно завищені, або занижені внаслідок того, що додається постійний трендовий компонент, який на практиці змінюється протягом року.

Приклад - 1



Експорт товарів та послуг
млн. дол. США
6000
5500
5000
4500
4000
3500
3000
1 кв. 97
2 кв. 97
3 кв. 97
4 кв. 97
1 кв. 98
2 кв. 98
3 кв. 98
4 кв. 98
1 кв. 99
2 кв. 99
3 кв. 99
4 кв. 99
1 кв. 00
2 кв. 00
3 кв. 00
4 кв.00
квартали

Приклад - 2



Імпорт товарів та послуг
6000
5500
5000
4500
4000
3500
3000
2500
квартали
1 кв. 97
2 кв. 97
3 кв. 97
4 кв. 97
1 кв. 98
2 кв. 98
3 кв. 98
4 кв. 98
1 кв. 99
2 кв. 99
3 кв. 99
4 кв. 99
1 кв. 00
2 кв. 00
3 кв. 00
4 кв.00
млн. дол. США

3. ФІЛЬТР ХОДРІКА-ПРЕСКОТТА

Опис методу – 1

При наявності багатьох обстежень деякого процесу, можна будувати регресійні моделі, які більш-менш точно відображають тренд. При недовгих спостереженнях за процесом, а особливо коли він має сезонні коливання, побудована таким чином регресія не тільки не дає задовільного результату з прийнятною похибкою, але й може взагалі невірно визначити напрям руху. Тому при дослідженні процесів з малою кількістю даних, варто застосовувати іншу методику.

Опис методу – 2

Нехай

$$y_t = f(t) + \varepsilon_t,$$

де y_t – значення змінної, що досліджується, t – час; – похибка; $f(t)$ – деяка функція, яка може залежати від декількох параметрів, а також обов'язково залежить від часу.

Опис методу – 3

Параметри у функції f підбираються таким чином, щоб мінімізувати вираз:

$$S = \sum_{t=1}^{T} \left(y_t - f(t) \right)^2 +$$

$$+ \lambda \sum_{t=2}^{T-2} \left(\left(f(t+1) - f(t) \right) - \left(f(t) - f(t-1) \right) \right)^2 \to \min$$

Опис методу – 4

Взявши частинні похідні, та поклавши їх рівними нулю, ми отримаємо умови першого порядку для знаходження мінімуму функції *S*:

$$t = 1: \quad y_1 = f(1) + \lambda f(1) - 2\lambda f(2) + \lambda f(3);$$

$$t = 2: \quad y_2 = f(2) - 2\lambda f(1) + 5\lambda f(2) - 4\lambda f(3) + \lambda f(4);$$

$$\ldots$$

$$t = T: \quad y_T = f(T) + \lambda f(T-2) - 2\lambda f(T-1) + \lambda f(T).$$

Опис методу – 5

Таким чином

$$\bar{y} = X\bar{f},$$

$$\bar{y} = (y_1, y_2, \ldots, y_T)^T, \ \bar{f} = (f(1), f(2), \ldots, f(T))^T$$

Опис методу – 6

$$X = \begin{Bmatrix}
1+\lambda & -2\lambda & \lambda & 0 & 0 & \dots & & & & & & & \\
-2\lambda & 1+5\lambda & -4\lambda & \lambda & 0 & 0 & \dots & & & & & & \\
\lambda & -4\lambda & 1+6\lambda & -4\lambda & \lambda & 0 & 0 & \dots & & & & & \\
0 & \lambda & -4\lambda & 1+6\lambda & -4\lambda & \lambda & 0 & \dots & & & & & \\
 & & & & & & & \dots & & & & & \\
0 & 0 & \dots & & & & & 0 & \lambda & -4\lambda & 1+6\lambda & -4\lambda & \lambda & 0 \\
0 & 0 & \dots & & & & & 0 & 0 & \lambda & -4\lambda & 1+6\lambda & -4\lambda & \lambda \\
0 & 0 & \dots & & & & & & 0 & 0 & \lambda & -4\lambda & 1+5\lambda & -2\lambda \\
0 & 0 & \dots & & & & & & & 0 & 0 & \lambda & -2\lambda & 1+\lambda
\end{Bmatrix}$$

Опис методу – 7

Тоді

$$X^T X = XX \Rightarrow (XX)^{-1} = X^{-1} X^{-1} \Rightarrow$$

$$\Rightarrow (X^T X)^{-1} X^T = (XX)^{-1} X^T = X^{-1};$$

X є квадратною та симетричною матрицею:

$$\bar{y} = X\bar{f},$$

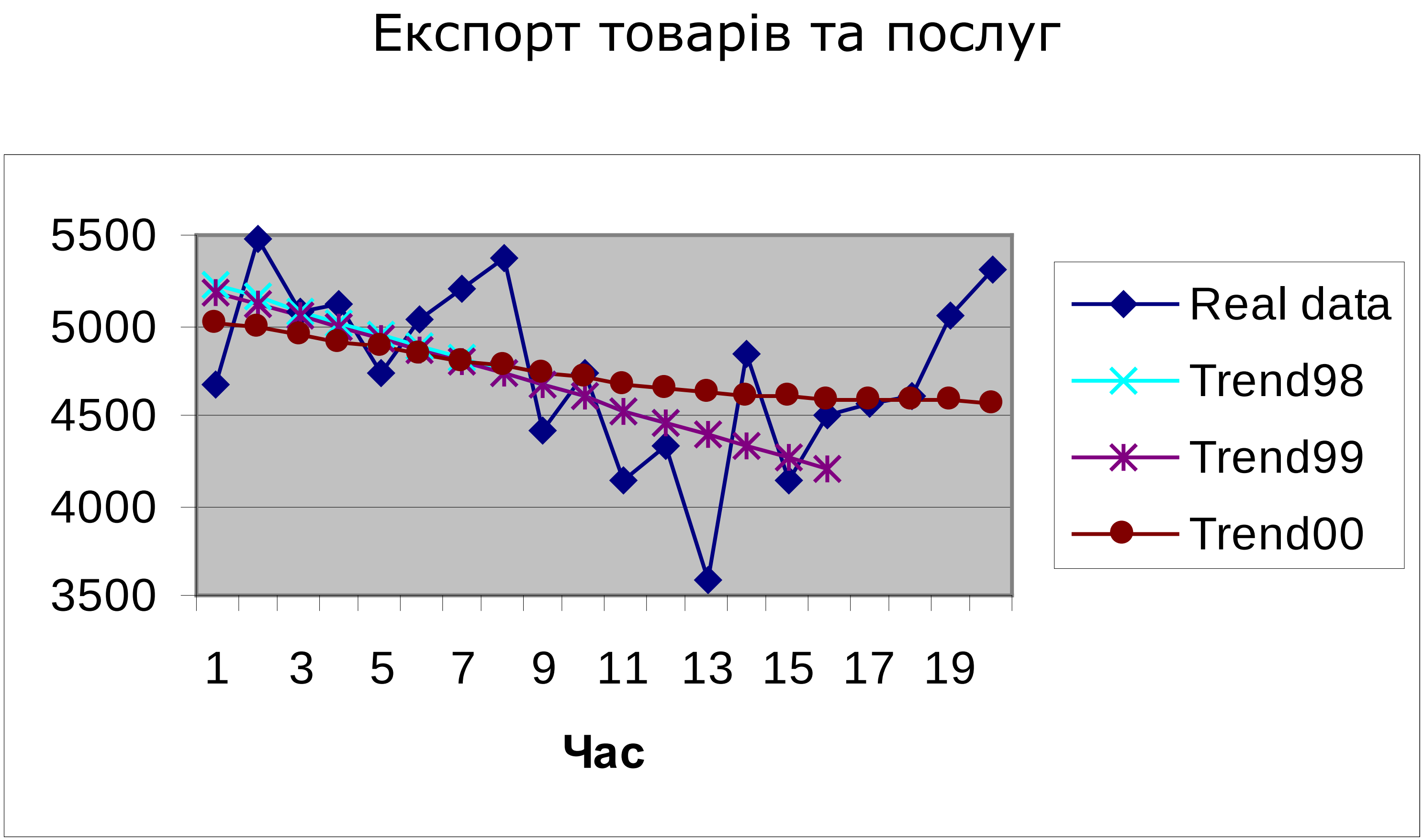
$$\bar{f} = X^{-1}\bar{y} = (X^T X)^{-1} X^T \bar{y}.$$

Опис методу – 8

Таким чином, можна знайти значення функції f для кожного значення t. Після цього встановлюється вигляд функції f чи знаходяться її параметри, прогнозується її значення на наступний період.

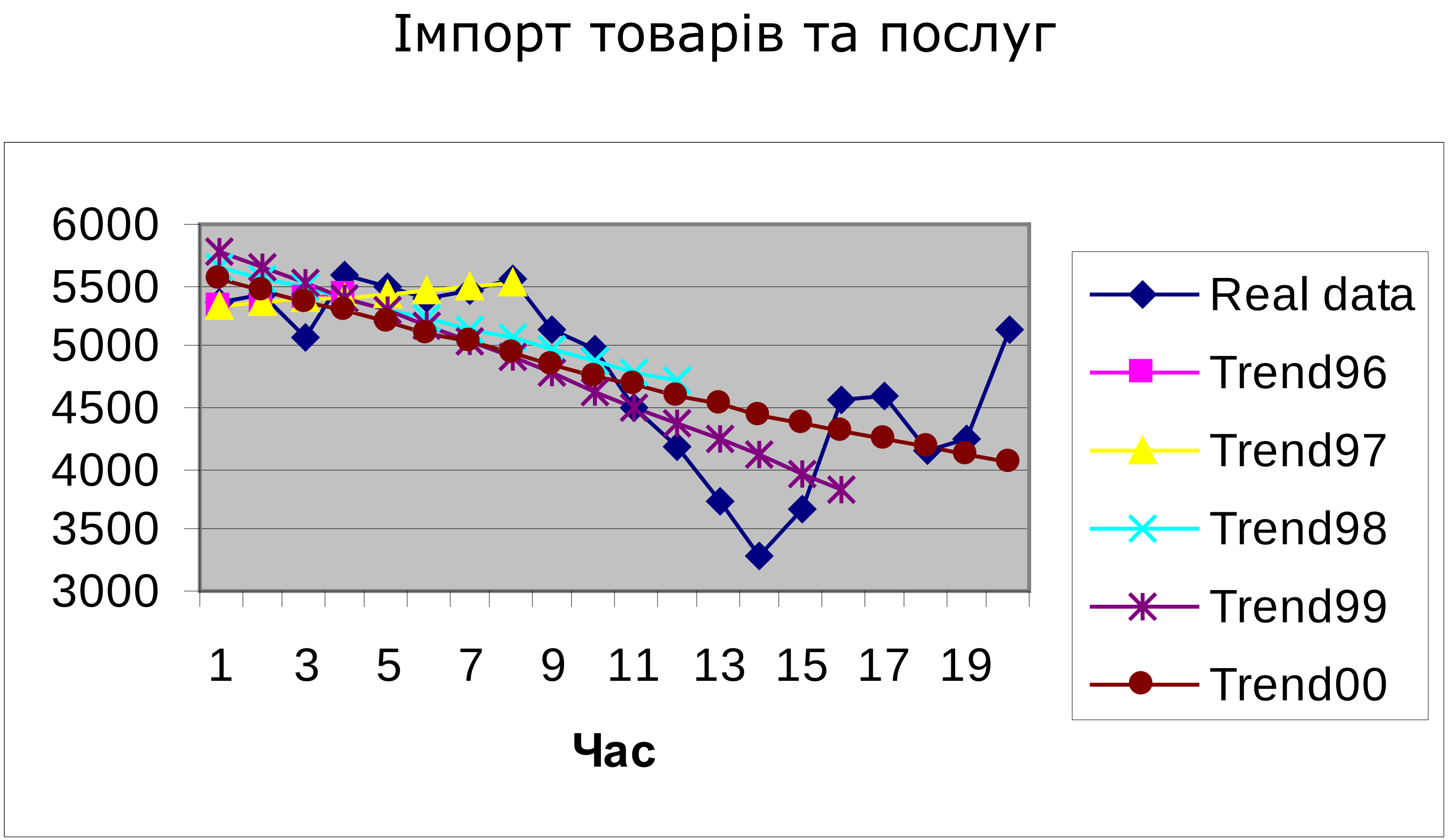
Практично було винайдено, що для річних даних прийнятним значенням вагового коефіцієнта є $\lambda=100$, для квартальних - $\lambda=1600$.

Приклад – 1



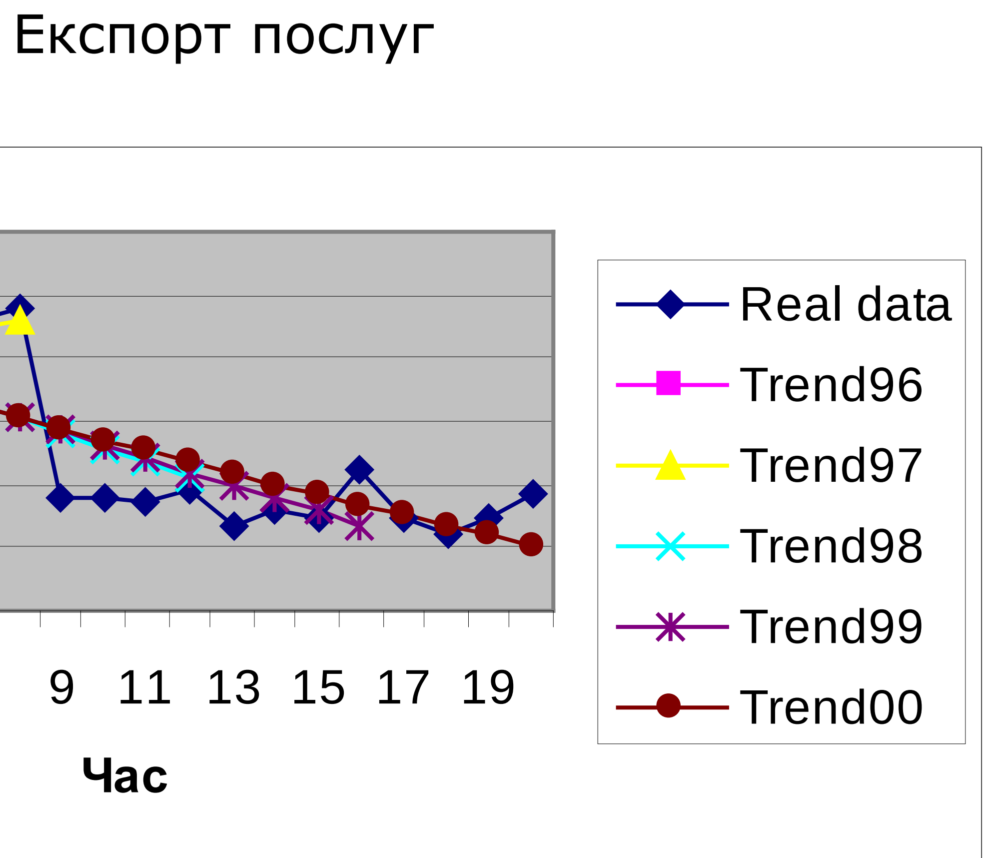
Експорт товарів та послуг
5500
5000
4500
4000
3500
1
3
5
7
9
11
13
15
17
19
Час
Real data
Trend98
Trend99
Trend00

Приклад – 2



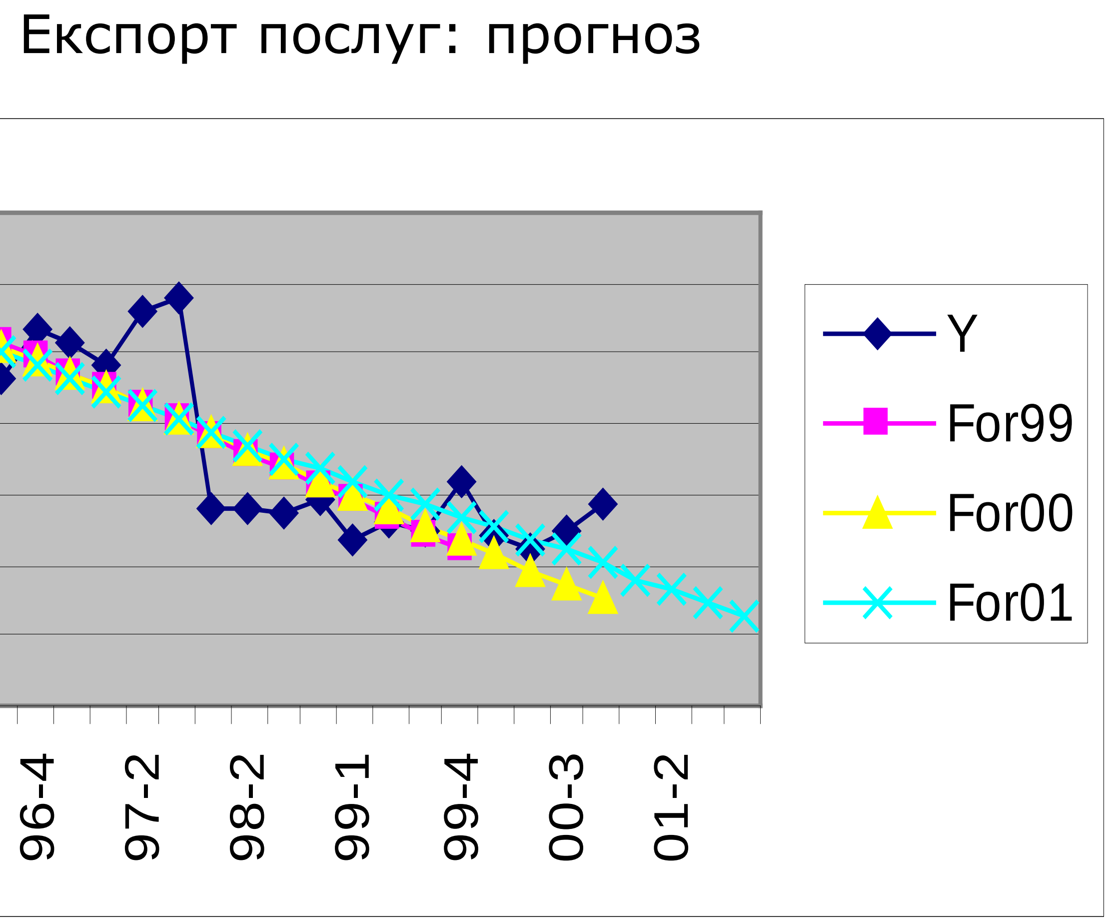
Імпорт товарів та послуг
6000
5500
5000
4500
4000
3500
3000
1
3
5
7
9
11
13
15
17
19
Час
Real data
Trend96
Trend97
Trend98
Trend99
Trend00

Приклад – 3



Експорт послуг
1400
1300
1200
1100
1000
900
800
1
3
5
7
9
11
13
15
17
19
Час
Real data
Trend96
Trend97
Trend98
Trend99
Trend00

Приклад – 4


Експорт послуг: прогноз
млн. дол. США
1400
1300
1200
1100
1000
900
800
700
96-1
96-4
97-2
98-2
99-1
99-4
00-3
01-2
Y
For99
For00
For01

4. ВИДІЛЕННЯ СЕЗОННОСТІ

Адитивна модель із визначенням сезонних коливань - 1

В цьому методі визначаються трендовий та сезонний компоненти часового ряду. Нехай p – цикл сезонності, тобто $s_t = s_{t+p}$ для будь-якого t. Наша задача – оцінити значення s_t за спостереженнями y_t. Для квартальних даних логічним є припущення про річний період сезонних коливань, тобто $p = 4$, для місячних – $p = 12$.

Адитивна модель із визначенням сезонних коливань - 2

Тепер, за допомогою методу усереднення, оцінимо трендовий компонент. В результаті отримуємо десезонолізовану (вільну від сезонних впливів) величину:

$$\hat{tr}_t = \frac{\left(\dfrac{y_{t-2} + y_{t-1} + y_t + y_{t+1}}{4} + \dfrac{y_{t-1} + y_t + y_{t+1} + y_{t+2}}{4}\right)}{2} =$$

$$= \frac{y_{t-2} + 2y_{t-1} + 2y_t + 2y_{t+1} + y_{t+2}}{8}, \qquad t = \overline{3, T-2}$$

Адитивна модель із визначенням сезонних коливань - 3

Тоді оцінка сезонної компоненти буде мати вигляд:

$$\hat{s}_t = y_t - \hat{tr}_t = y_t - \frac{y_{t-2} + 2y_{t-1} + 2y_t + 2y_{t+1} + y_{t+2}}{8}$$

$$= \frac{6y_t - (y_{t-2} + 2y_{t-1} + 2y_{t+1} + y_{t+2})}{8},$$

$$t = \overline{3, T-2}$$

Адитивна модель із визначенням сезонних коливань - 4

Підрахуємо тепер середні значення сезонної компоненти для кожного періоду циклу сезонності $\overline{s_t}$

Вважаючи, що $\overline{s_{t+4}} = \overline{s_t}$ можна отримати скоригований сезонний компонент за формулою

$$\overline{s_t}* = \overline{s_t} - \frac{\sum_{i=1}^{p} \overline{s_i}}{p}$$

Отримані значення задовольняють властивості:

$$\sum_{t=1}^{p} \overline{s_t}* = 0$$

Адитивна модель із визначенням сезонних коливань - 5

,
,
.

Тепер вилучимо з даних сезонний компонент і отримаємо ряд:

$$tr_t = y_t - \bar{s}_t \ *$$

Якщо припустити наявність лінійного тренду в моделі, то

$$y_t = a_0 + a_1 t + \bar{s}_t \ * \qquad t = \overline{1, T}$$

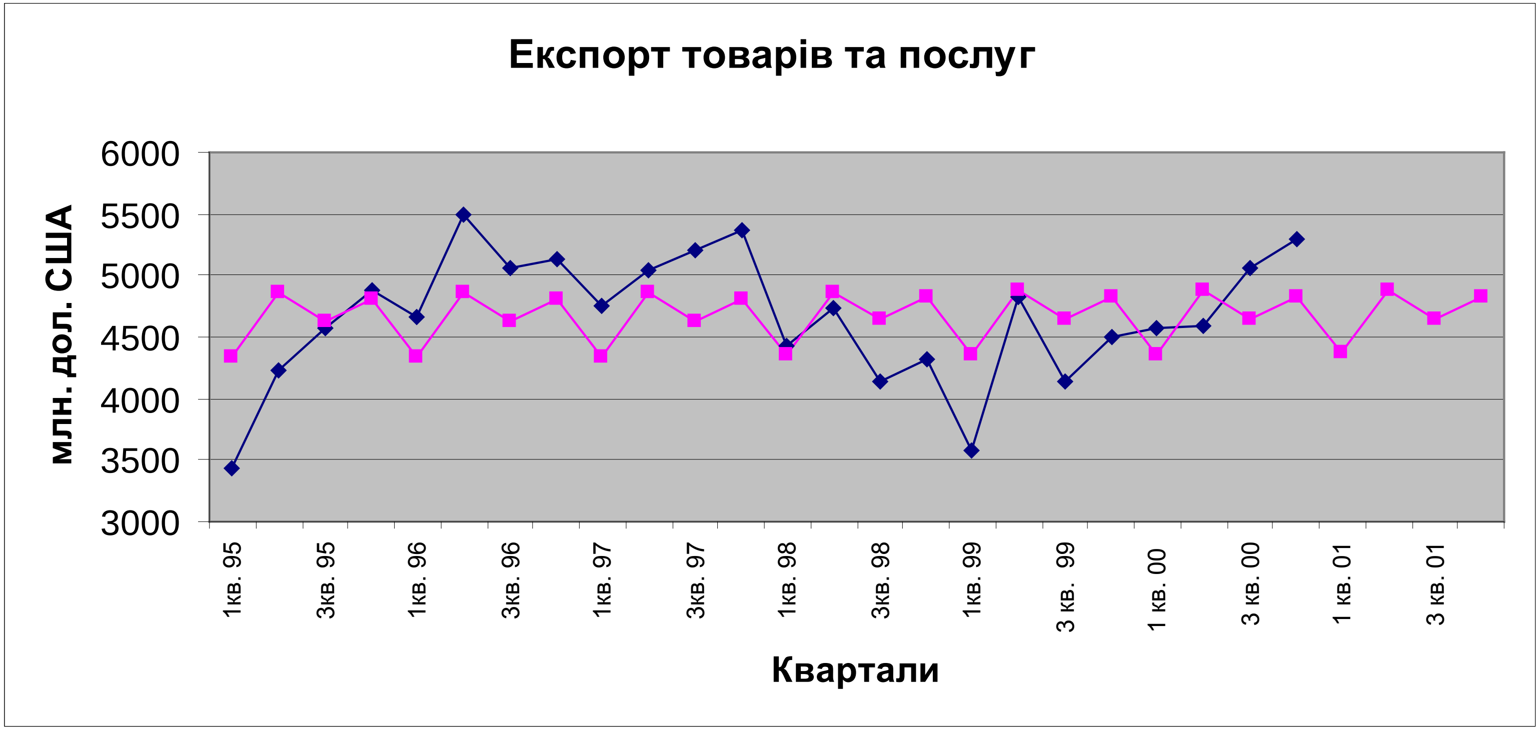
Прогноз

Прогнози на наступні періоди будуються за формулою:

$$\hat{y}_t = a_0 + a_1 t + \bar{s}_t \ *$$

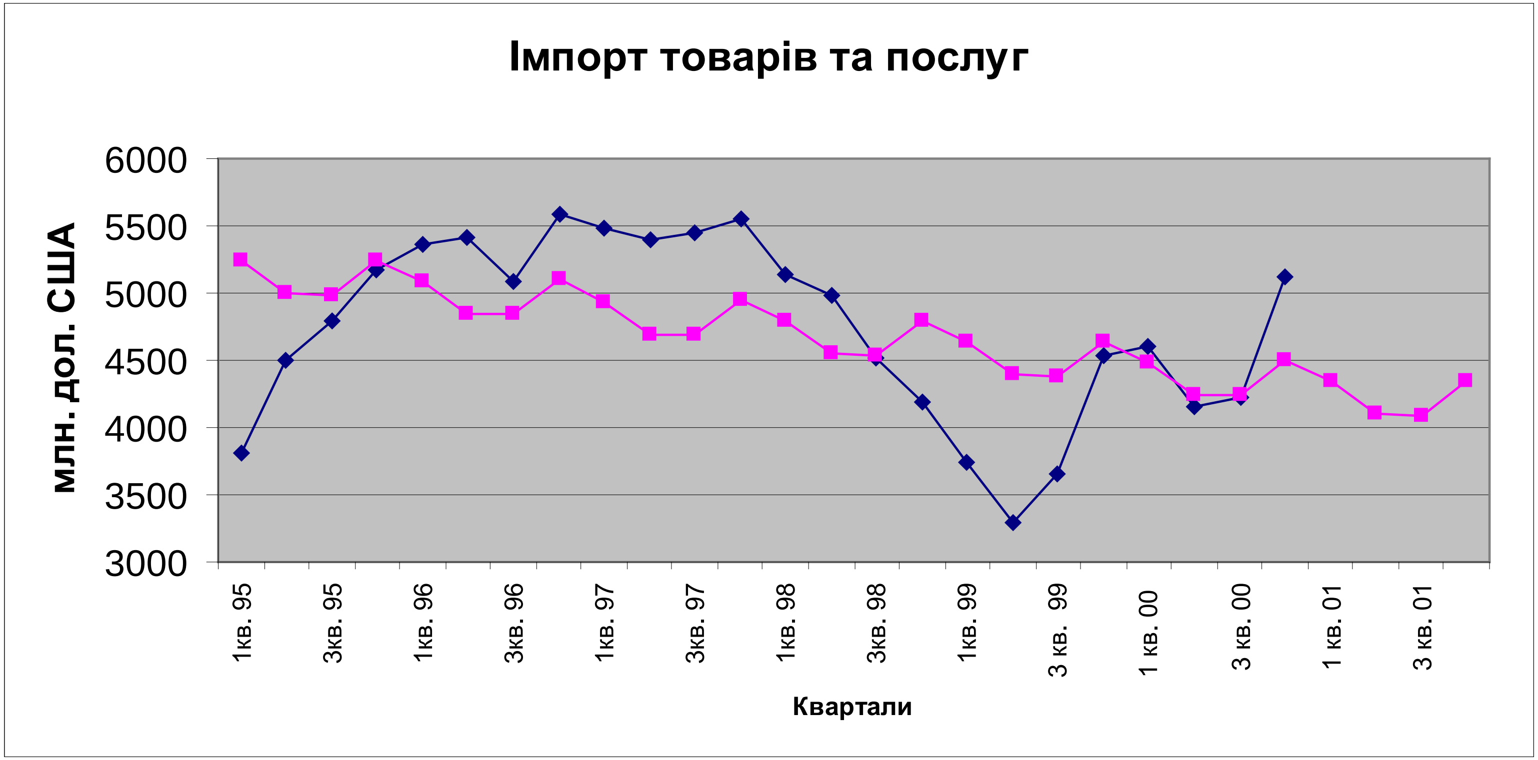
$$t = T + 1, T + 2, \ldots.$$

Приклад - 1



Експорт товарів та послуг
6000
5500
5000
4500
4000
3500
3000
млн. дол. США
1кв. 95
3кв. 95
1кв. 96
3кв. 96
1кв. 97
3кв. 97
1кв. 98
3кв. 98
1кв. 99
3 кв. 99
1 кв. 00
3 кв. 00
1 кв. 01
3 кв. 01
Квартали

Приклад - 2



Імпорт товарів та послуг
млн. дол. США
6000
5500
5000
4500
4000
3500
3000
1кв. 95
3кв. 95
1кв. 96
3кв. 96
1кв. 97
3кв. 97
1кв. 98
3кв. 98
1кв. 99
3 кв. 99
1 кв. 00
3 кв. 00
1 кв. 01
3 кв. 01
Квартали

Адитивна модель Вінтерса

Адитивна модель Вінтерса є розвиненням експоненціального згладжування. Аналізуються три статистично залежні ряди, які використовуються для побудови дійсного прогнозу:

- a_t - згладжені дані,
- b_t - трендовий компонент,
- c_t - індекс сезонності

Модель

$$a_t = \alpha\left(\frac{y_t}{c_{t-s}}\right) + (1-\alpha)(a_{t-1} + b_{t-1}),$$

$$b_t = \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1},$$

$$c_t = \gamma\left(\frac{y_t}{a_t}\right) + (1-\gamma)c_{t-s},$$

$$t = \overline{2s+1,\, T},$$

де *s* - кількість циклів сезонності.

Прогноз

Прогноз на період $T+p$ будується наступним чином:

$$\hat{y}_{T+p} = \left(a_T + pb_T\right)c_{T-s+p},$$

$$p = 1,\quad 2,\quad \ldots,\quad s,$$

$$\hat{y}_{T+p} = \left(a_T + pb_T\right)c_{T-2s+p},$$

$$p = s+1,\quad s+2,\quad \ldots,\quad 2s.$$

Сумуюче згладжування

Розглянемо модель сумування елементів часового ряду:

$$\tilde{y}_t = \sum_{i=1}^{t} y_t$$

та побудуємо лінійну регресію виду

$$\tilde{y}_t = a_0 + a_1 t + \varepsilon_t$$

Найпростіше застосування регресійного аналізу дозволяє отримати, як саме змінюються коефіцієнти моделі, що дозволяє будувати прогнози часового ряду.

Похибка прогнозування, % (2001)

Методи	Експорт товарів та послуг			Імпорт товарів та послуг		
	I кв.	II кв.	Перше півріччя	I кв.	II кв.	Перше півріччя
Реальне значення	-5,28	-12,84	-9,22	-7,32	-13,43	-10,48
Експоненціальне згладжування	-8,79	-16,07	-12,58	-12,13	-17,92	-15,12
Подвійне експоненціальне згладжування	-8,56	-15,86	-12,36	-8,68	-14,70	-11,79
Потрійне експоненціальне згладжування	9,15	6,42	7,73	7,23	5,38	6,27
Несезонна модель Холта-Вінтерса	-18,56	-15,04	-16,73	-18,58	-28,26	-23,59
Сумуюче згладжування	-11,41	-14,50	-13,01	19,50	-20,12	-0,99
Сезонна адитивна модель	-11,72	-9,14	-10,37	-8,68	-19,45	-14,25
Наївна модель	-7,56	-14,44	-11,14	-3,05	-18,27	-10,92

ОГЛЯД

Методи згладжування

Методи згладжування використовуються для зменшення впливу випадкового компонента (випадкових коливань) у часових рядах. Вони дають можливість отримувати більш “чисті” значення, які складаються лише з детермінованих компонентів. Одні з методів направлені на виділення деяких компонентів, наприклад, тренду.

Згладжування

- Експоненціальне
- Холта-Вінтерса
- Адитивна модель Вінтерса
- Фільтр Ходріка-Прескотта

ПИТАННЯ?

ДЯКУЮ ЗА УВАГУ!