

Реалізація методики побудови моделей часових рядів

Виконав:

студент групи КА-24

Дудка Богдан Романович

Науковий керівник:

д.т.н., проф. Бідюк П.І.

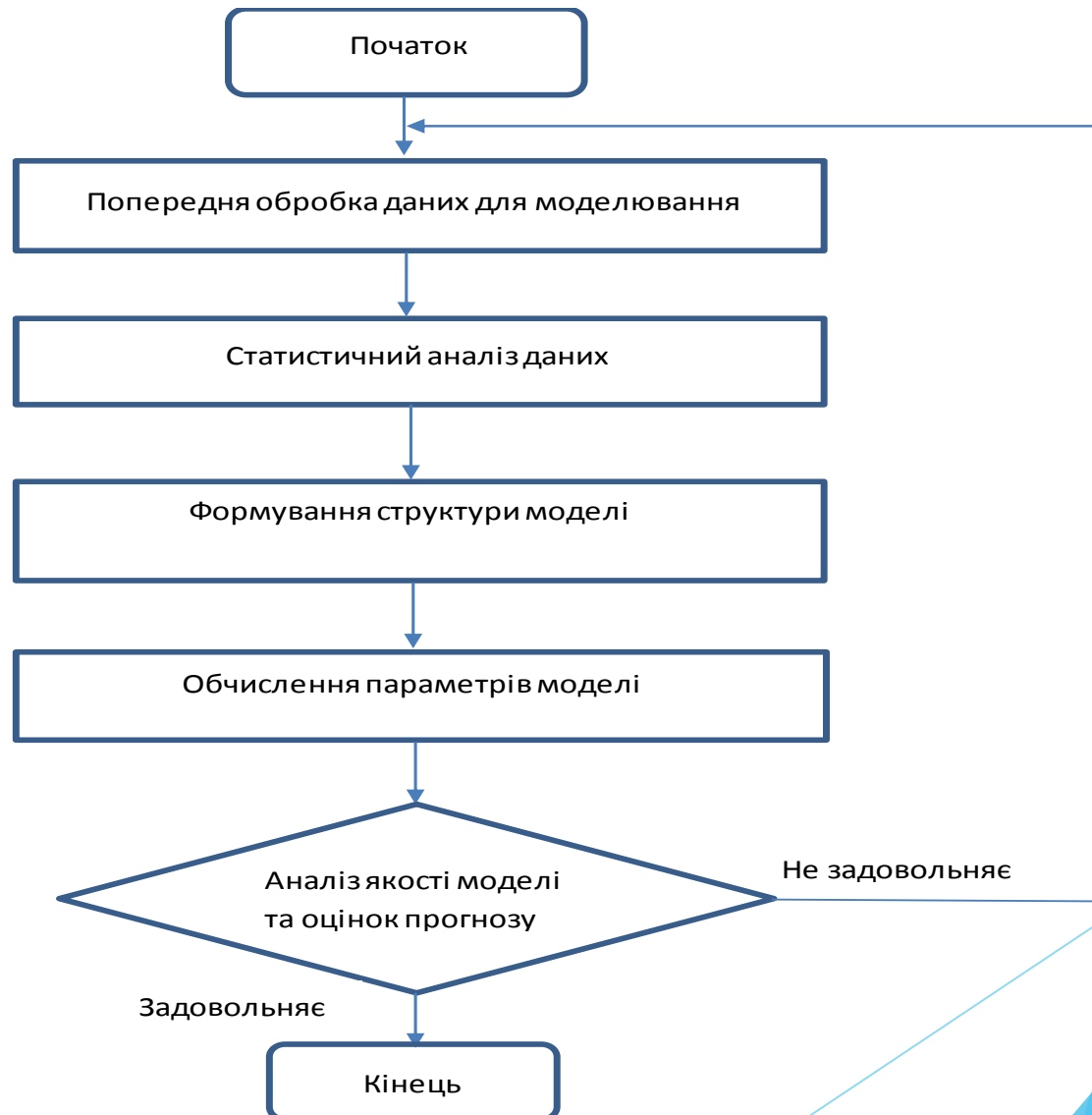
Об'єкт, предмет і мета дослідження

- ▶ **Об'єкт** дослідження – статистичні дані стосовно розвитку вибраних фінансово-економічних процесів
- ▶ **Предмет** дослідження – методика побудови моделей часових рядів, методи дослідження пропусків даних, статистичні характеристики адекватності моделей і оцінок прогнозів, методи згладжування часових рядів.
- ▶ **Методи** дослідження – статистичний аналіз даних, методи заповнення пропусків даних, регресійний аналіз.
- ▶ **Мета** дослідження – реалізація та модифікація методики побудови моделей часових рядів, дослідження впливу пропусків даних та застосування методів згладжування часових рядів на статистичні характеристики моделей процесів з динамікою, розробка програмного продукту для побудови моделей часових рядів на основі представленої методики.

Постановка задачі

- ▶ Зібрати статистичні дані для виконання обчислювальних експериментів.
- ▶ Модифікувати та реалізувати методика побудови моделей часових рядів.
- ▶ Вибрати та розробити методи заповнення пропусків і згладжування даних.
- ▶ Виконати обчислювальні експерименти щодо заповнення пропусків даних і побудові моделей на основі цих даних.
- ▶ Виконати аналіз впливу застосування методів згладжування на статистичні характеристики моделей процесів.
- ▶ Виконати порівняльний аналіз отриманих результатів і надати рекомендації стосовно подальшої модифікації розглянутої методики побудови моделей часових рядів та вибору методів заповнення пропусків у статистичних даних.

Методика побудови моделей часових рядів



Показники якості моделей і прогнозів

Коефіцієнт детермінації	$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)}$
Сума квадратів похибок моделі	$SSE = \sum_{k=1}^N [\hat{y}(k) - y(k)]^2$
Статистика Дарбіна-Уотсона	$DW = \frac{\sum_{k=2}^N [e(k) - e(k-1)]^2}{\sum_{k=1}^N e^2(k)}$
Коефіцієнт Тейла	$U = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_i)^2 + \frac{1}{N} \sum_{i=1}^N (\hat{y}_i)^2}}$
Середня похибка в процентах	$СПП = \frac{1}{S} \sum_{i=1}^S \frac{y(k+s) - \hat{y}(k+s, k)}{y(k+s)} \times 100\%$
Абсолютна середня похибка в процентах	$АСПП = \frac{1}{S} \sum_{i=1}^S \frac{ y(k+s) - \hat{y}(k+s, k) }{ y(k+s) } \times 100\%$

Приклади прогнозуючих функцій

- Прогнозуюча функція для процесу AP(1) без розв'язку:

$$E_k[y(k + s)] = \hat{y}(k + s, k) = a_0 \sum_{i=0}^{s-1} a_1^i + a_1^s y(k);$$

- Прогнозуюча функція для процесу ARКC(1,1) на основі повного розв'язку:

$$E_k[y(k + s)] = \frac{a_0}{1-a_1} (1 - a_1^s) + \beta_1 a_1^{s-1} \varepsilon(k) + a_1^s y(k);$$

- Прогнозуюча функція для рівняння AP(1) на основі повного розв'язку:

$$E_k[y(k + s)] = \frac{a_0}{1-a_1} (1 - a_1^s) + a_1^s y(k);$$

Типи моделей, що будуються за даною методикою

- ▶ Авторегресія (АР(p)):

$$y(k) = a_0 + \sum_{i=0}^p a_i y(k-i) + \varepsilon(k)$$

- ▶ Авторегресія з ковзним середнім (АРКС(p, q)):

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=0}^q b_j \varepsilon(k-j)$$

- ▶ Авторегресія з трендом (АР+тренд(p, q)):

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j t^j(k) + \varepsilon(k)$$

- ▶ Узагальнена авторегресійна умовно гетероскедастична модель (УАРУГ):

$$h(k) = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i) + \sum_{i=1}^p \beta_i h(k-i) + v(k)$$

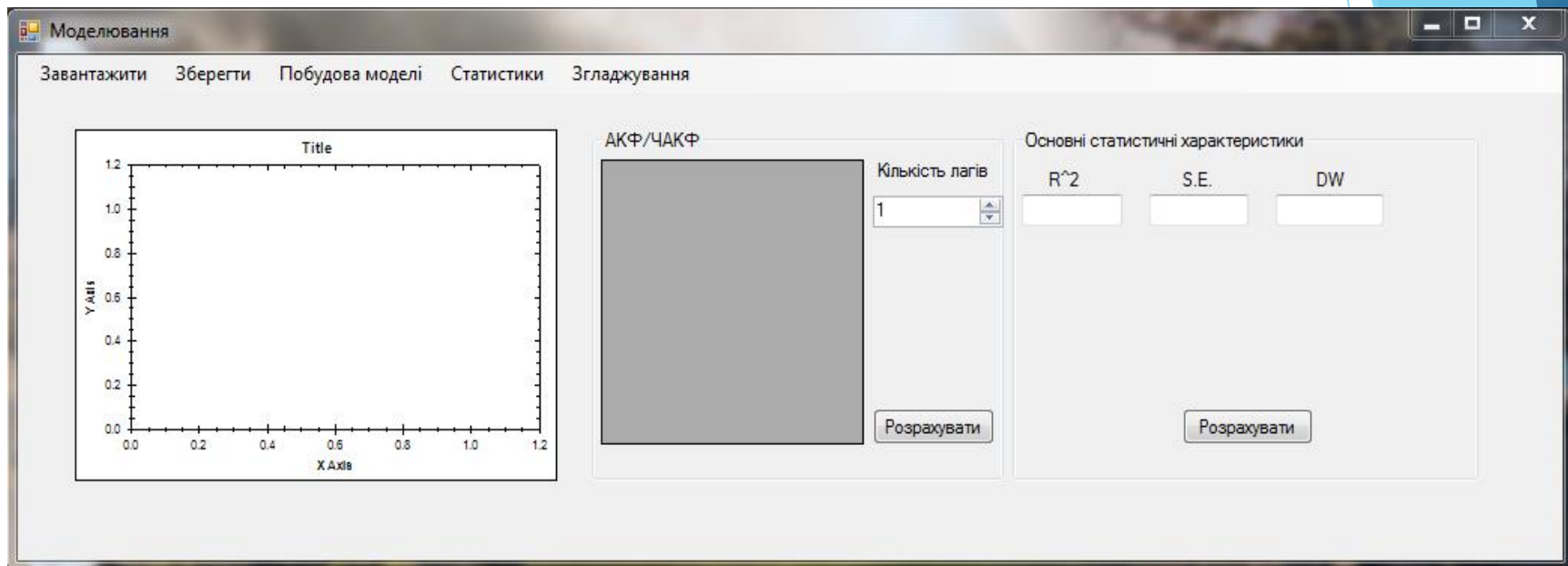
- ▶ Австорегресійна умовно гетероскедастична модель (АРУГ):

$$\varepsilon^2(k) = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon^2(k-i)$$

Використані методи заповнення пропусків

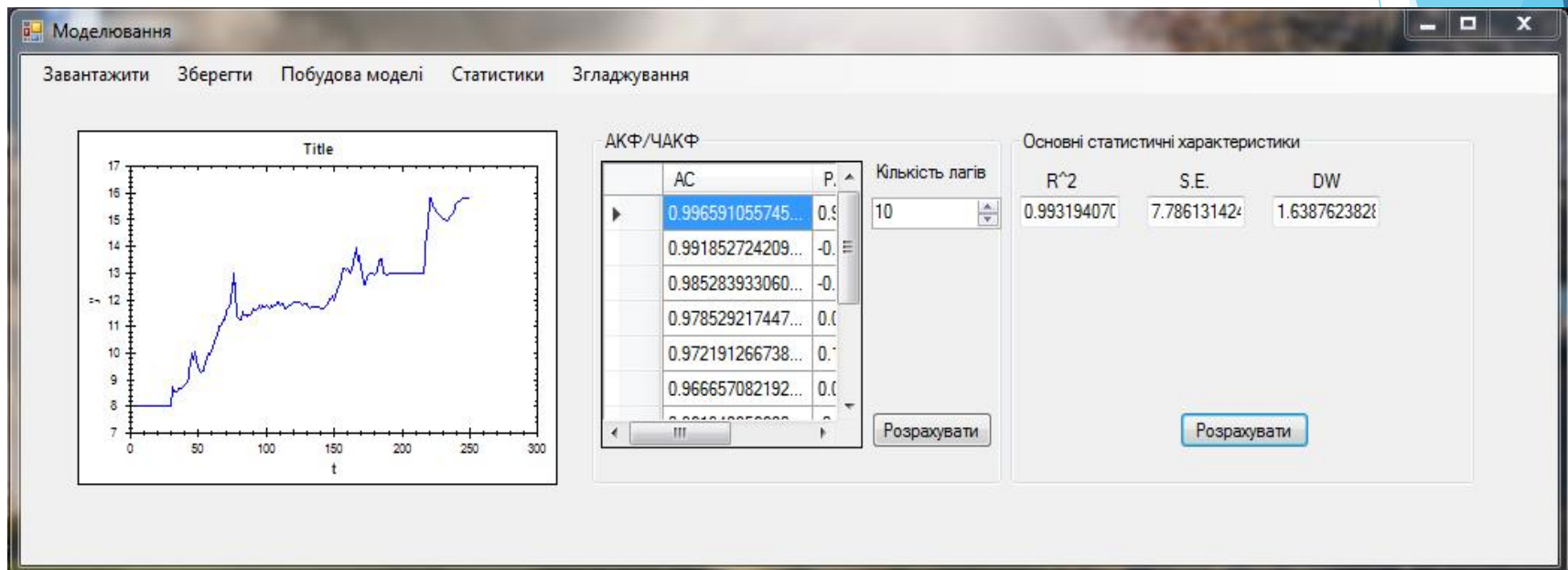
- Середнім значенням
- Оцінками прогнозів
- Випадковими величинами з нормального розподілу
- Випадковими величинами з рівномірного розподілу

Головне вікно програми



Вхідні дані

- Курс валют: долар/гривня 2014 рік.
- Потужність часового ряду становить 249 значень.



АКФ і ЧАКФ

АКФ/ЧАКФ

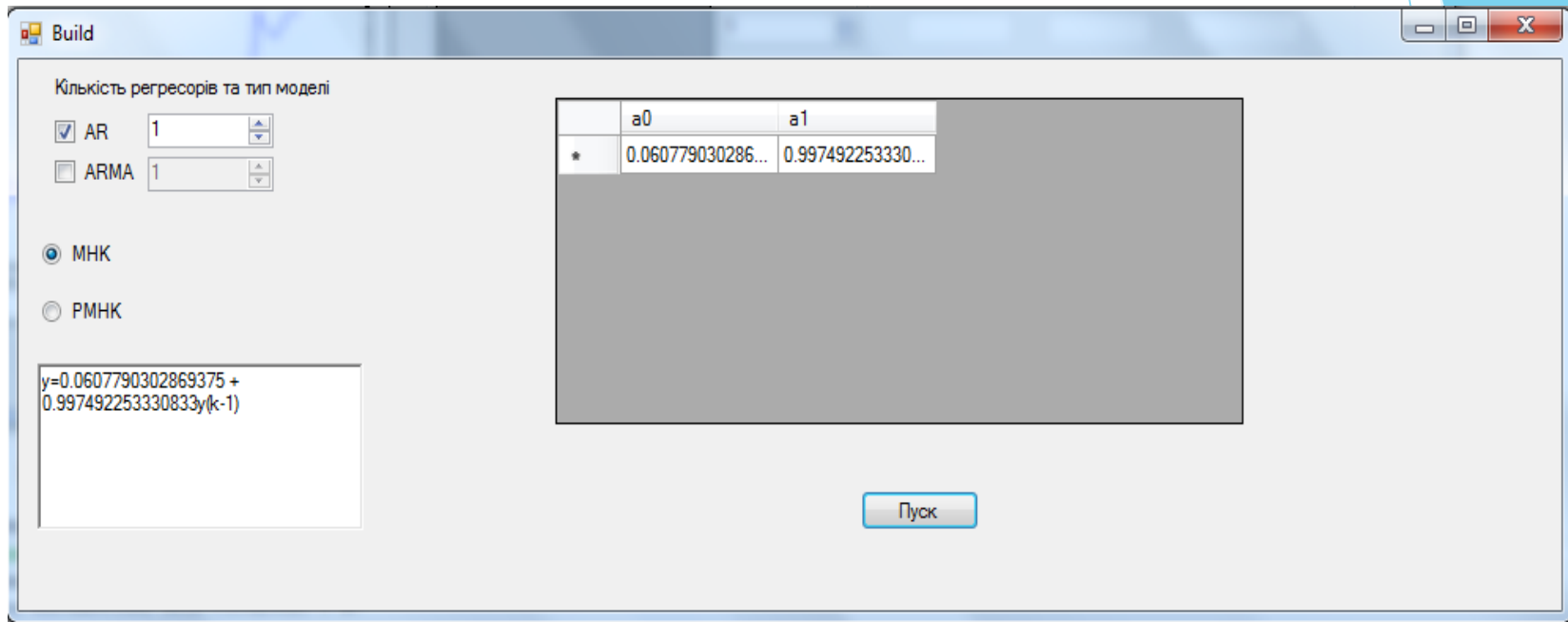
	AC	P
▶	0.996591055745...	0.9
	0.991852724209...	-0.
	0.985283933060...	-0.
	0.978529217447...	0.0
	0.972191266738...	0.1
	0.966657082192...	0.0
	0.961118858888...	0.0

Кількість лагів

10

Розрахувати

Побудова моделі



Build

Кількість регресорів та тип моделі

AR 1

ARMA 1

МНК

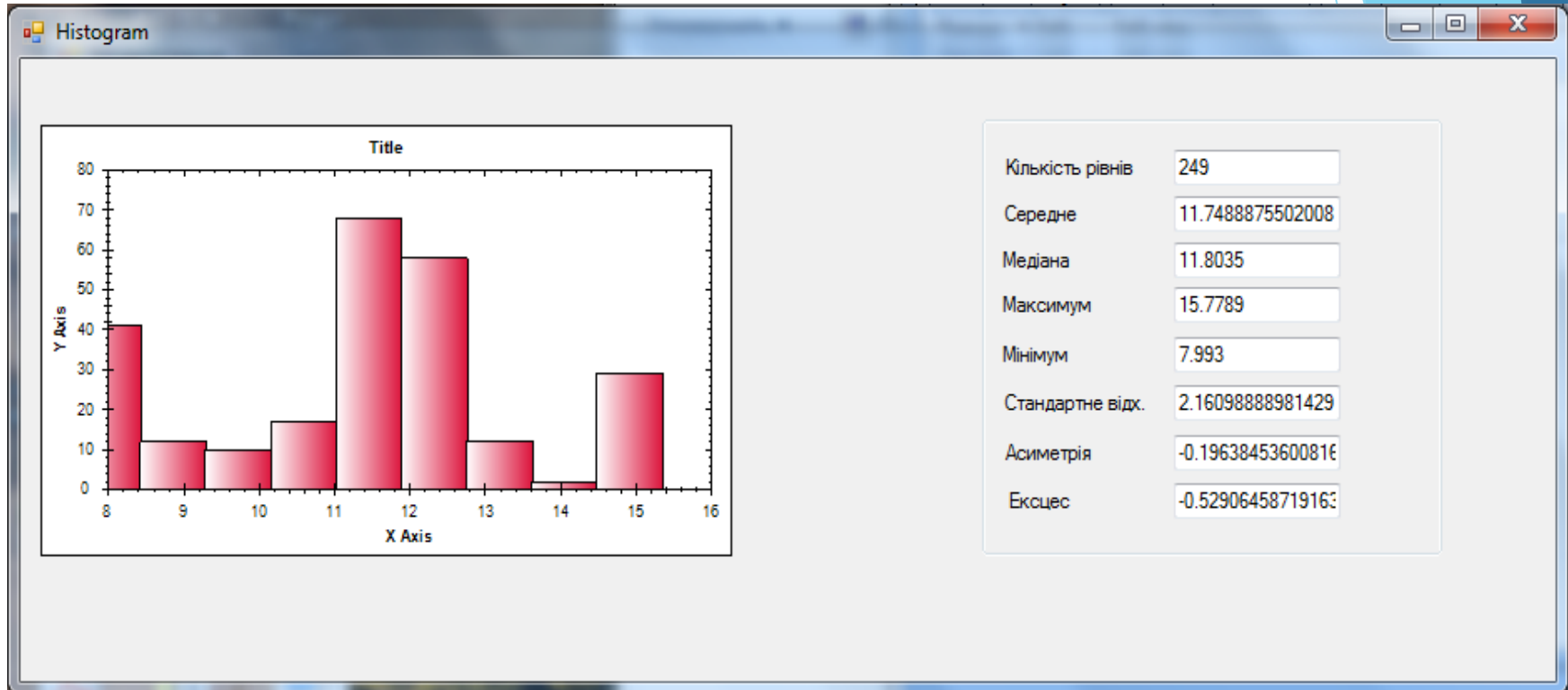
РМНК

	a0	a1
*	0.060779030286...	0.997492253330...

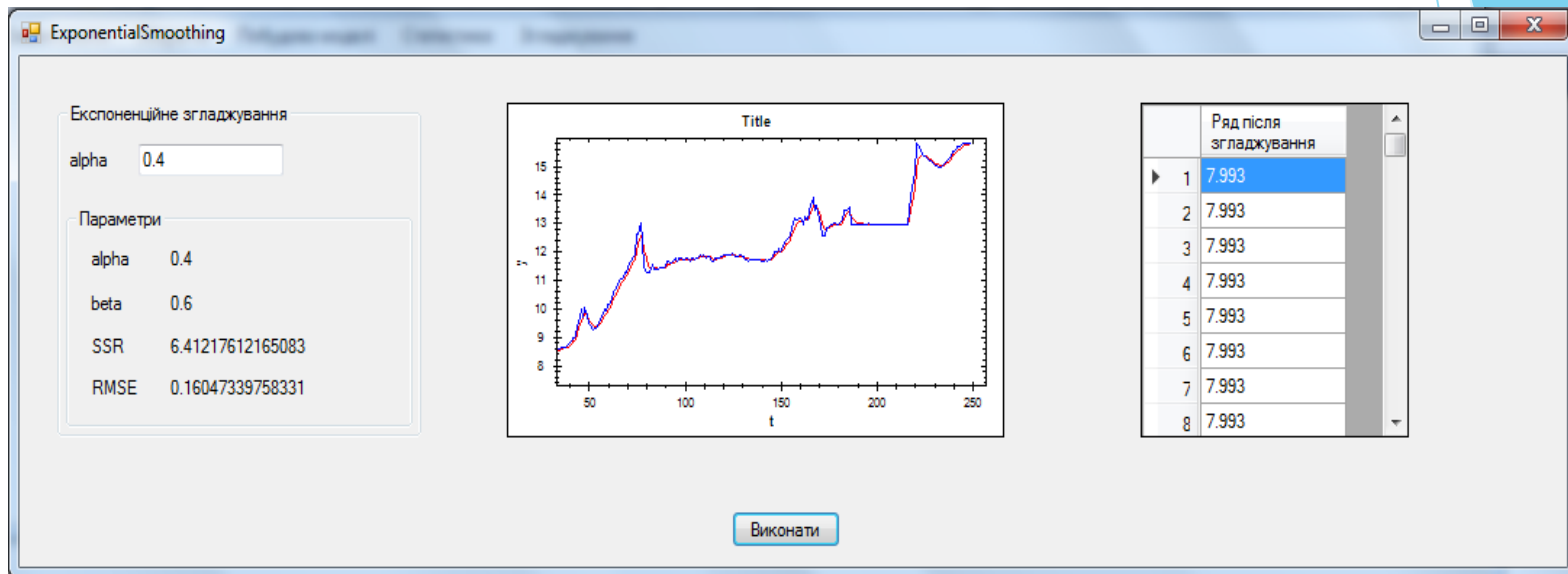
$y=0.0607790302869375 + 0.997492253330833y_{(k-1)}$

Пуск

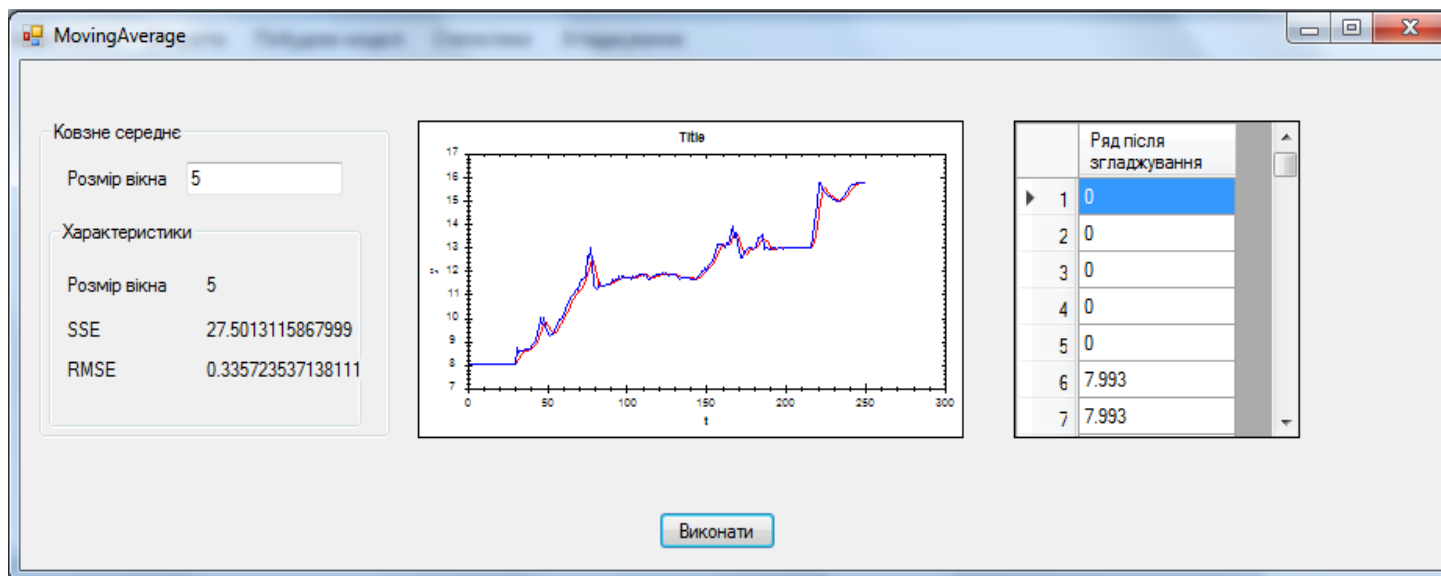
Статистичний аналіз



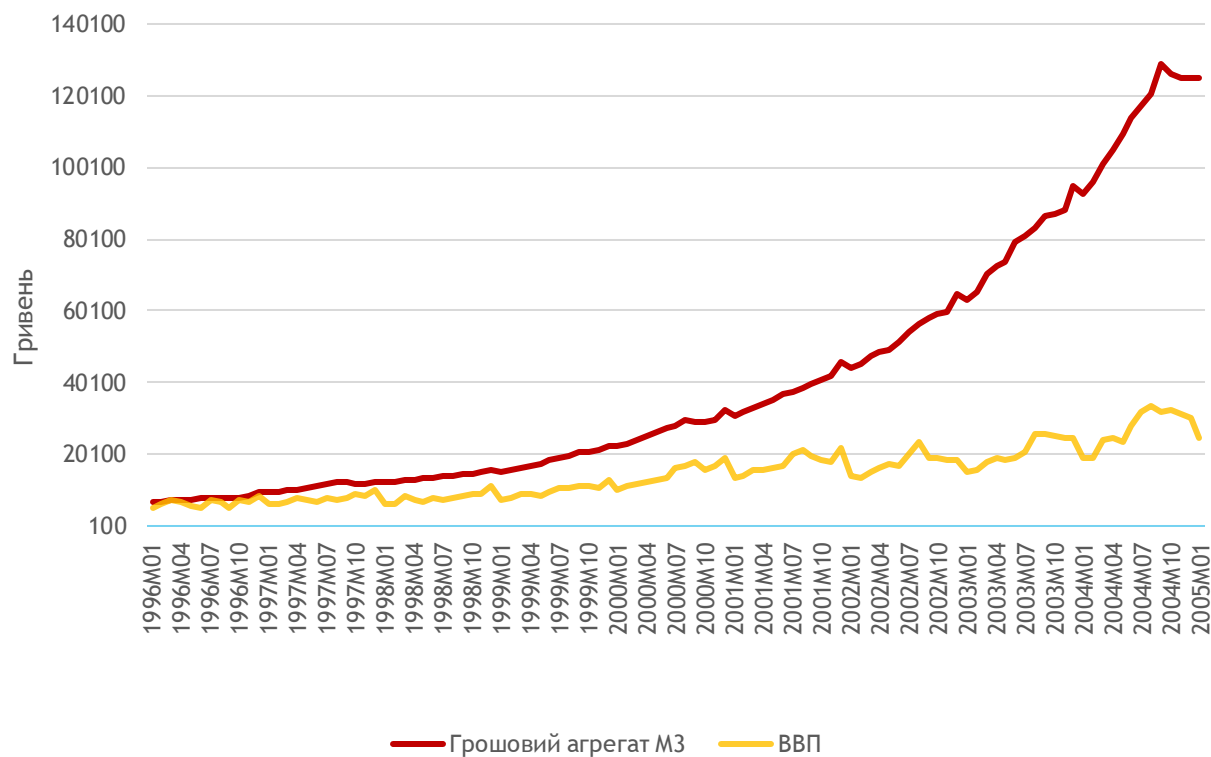
Експоненційне згладжування

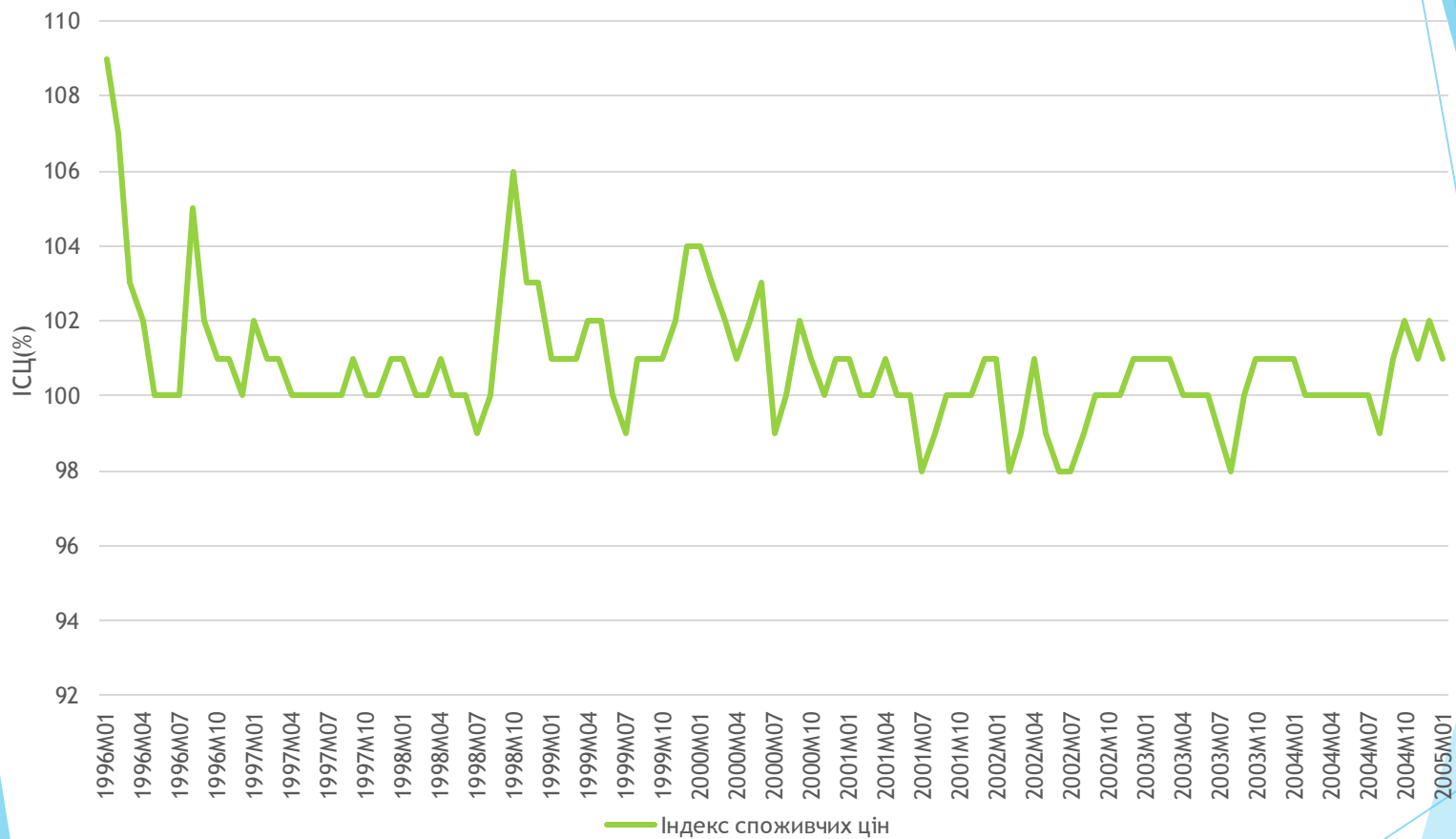


Згладжування за допомогою ковзного середнього



Реалізація методики побудови моделей часових рядів

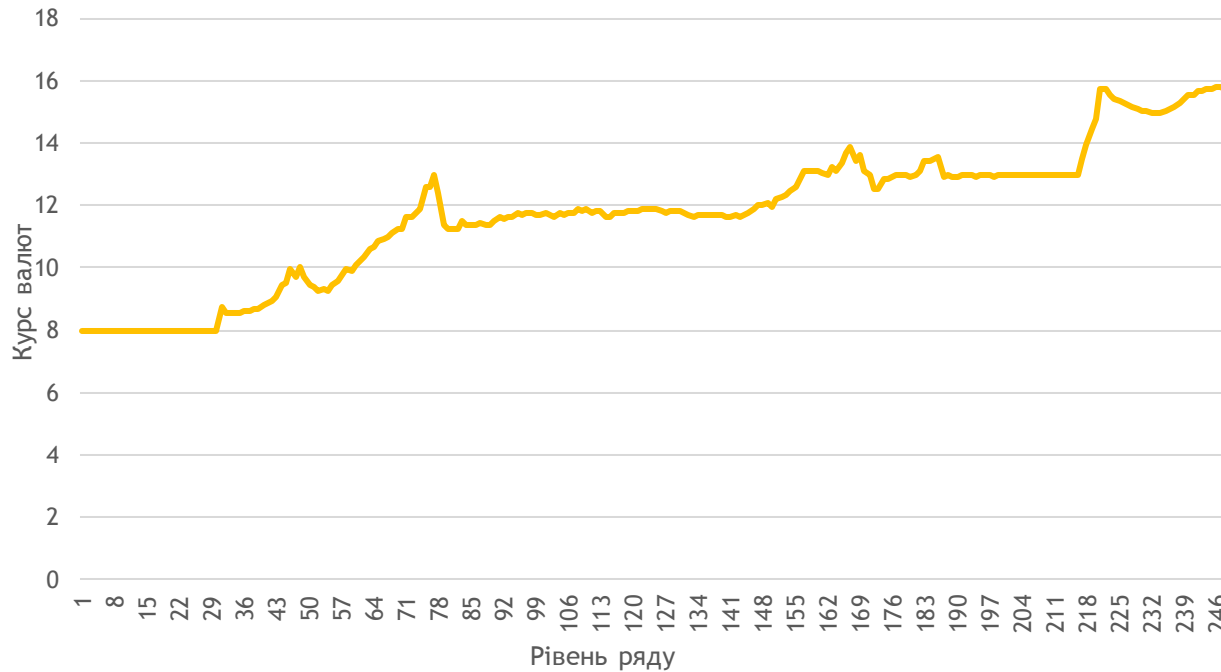




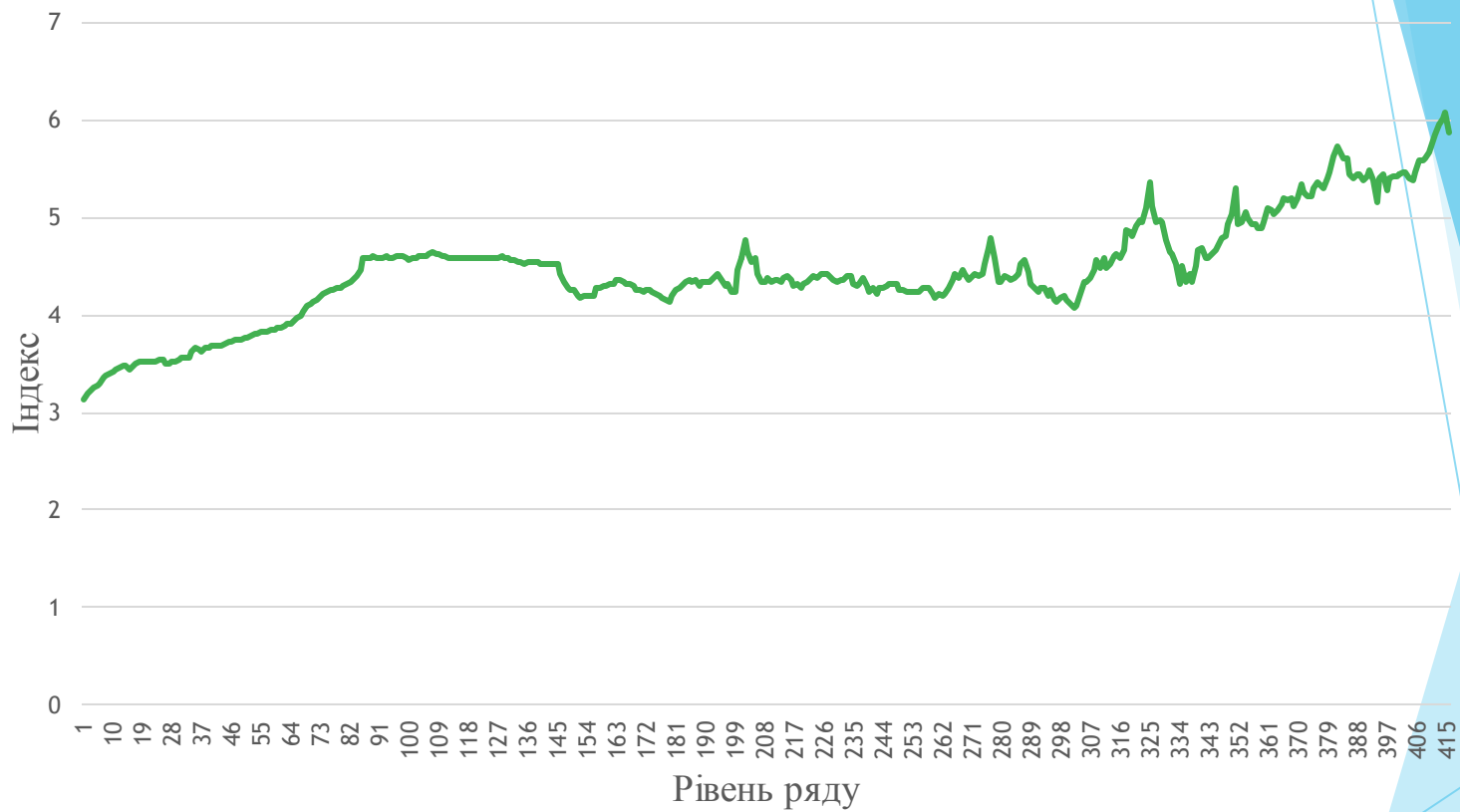
Результати моделювання та короткострокового прогнозування

Тип моделі	Характеристики моделі			Характеристики прогнозу			
	R ²	$\sum e^2(k)$	DW	СеКП	САП	САПП	U
АР(1)	0,415	141,99	1,931	1,360	1,020	1,008	0,0067
АР(3)	0,317	135,148	1,992	1,360	1,020	1,011	0,0068
АР(7)	0,346	127,244	1,811	1,360	1,012	1,002	0,0067
АР(12)	0,435	97,80	1,941	1,337	1,020	1,013	0,0066
АРКС(1,1)	0,416	141,78	1,996	1,362	1,016	1,005	0,0067
АР(1)+МЗ	0,419	141,007	1,919	1,340	1,004	0,994	0,0066
АР(1)+ВВП	0,419	141,054	1,916	1,335	1,004	0,993	0,0066

Імплементація методів заповнення пропусків



Курс валют: долар/гривня(2014 рік)



Індекс цін товаровиробників – первинні енергетичні матеріали(1974.1-2008.8)

Таблиці з результатами моделювання після заповнення пропусків даних

Тип моделі	Адекватність			Якість прогнозу		
	R ²	Σe ²	DW	RMSE	MAPE	V
Ідеальний ряд AP(1)(без пропусків)	0,993194	7,786131	1,638762	0,796331	5,579418	0,032757
5 пропусків AP(1)	0,582792	801,6419	2,623437	2,719252	16,13595	0,115064
10 пропусків AP(1)	0,474081	1404,563	2,673507	3,246950	19,56524	0,138383
15 пропусків AP(1)	0,378006	2107,858	2,568085	3,677957	22,82081	0,157754
Заповнення пропусків середнім значенням (5,10,15 значень)						
AP(1)(5 пропусків)	0,966663	35,71873	1,906365	1,274126	9,932036	0,052084
AP(1)(10 пропусків)	0,964183	35,67618	1,905186	1,369852	10,56082	0,055778
AP(1)(15 пропусків)	0,961187	35,62485	1,903806	1,435057	10,94749	0,058208
Заповнення пропусків значеннями з прогнозу (5,10,15 значень)						
AP(1)(5 пропусків)	0,993124	7,835027	1,641374	0,796766	5,563862	0,032768
AP(1)(10 пропусків)	0,992940	7,962562	1,646944	0,797219	5,523656	0,032767
AP(1)(15 пропусків)	0,992641	8,164928	1,655420	0,796761	5,459403	0,032718
Заповнення пропусків нормально розподіленими величинами(5,10,15 значень)						
AP(1)(5 пропусків)	0,992377	8.589816	1.713261	0.809218	5.570993	0.033236
AP(1)(10 пропусків)	0.992112	8.761841	1.708382	0.812561	5.679683	0.033347
AP(1)(15 пропусків)	0.992111	8.683712	1.676252	0.811381	5.627265	0.033289
Заповнення пропусків рівномірно розподіленими величинами (5,10,15 значень)						
AP(1)(5 пропусків)	0,989885	11,54448	2,222202	0,854502	6,232994	0,035032
AP(1)(10 пропусків)	0,979245	22,35595	1,665215	1,063332	8,534481	0,043373
AP(1)(15 пропусків)	0,969518	31,45628	1,737932	1,199706	9,401084	0,048955
Експоненційне згладжування(Exponential, alpha = 0,6)						
AP(1)	0,966438	4,052998	0,829825	0,752145	5,141824	0,031166
Згладжування за допомогою ковзного середнього(MA, T=2)						
AP(1)	0,995866	4,592002	0,475521	0,770693	5,258617	0,031892

Статистичні характеристики моделі процесу (курс валют)

Тип моделі	Адекватність			Якість прогнозу		
	R ²	Σe^2	DW	RMSE	MAPE	V
Ідеальний ряд AP(1)(без пропусків)	0.984131	1.903643	1.836867	0.441836	8.040758	0.048126
5 пропусків AP(1)	0.584089	86.03976	2.595776	0.679893	9.855768	0.076287
10 пропусків AP(1)	0.349443	193.0927	2.490091	0.837722	10.16546	0.094328
15 пропусків AP(1)	0.089041	401.2624	2.233811	1.028839	10.68965	0.115551
	Заповнення пропусків середнім значенням (5,10,15 значень)					
AP(1)(5 пропусків)	0.962612	4.426393	1.921697	0.427444	7.949919	0.046897
AP(1)(10 пропусків)	0.961391	4.204673	1.922654	0.444045	8.323843	0.048633
AP(1)(15 пропусків)	0.960591	4.104981	1.90977	0.456435	8.627456	0.049901
	Заповнення пропусків значеннями з прогнозу (5,10,15 значень)					
AP(1)(5 пропусків)	0.984161	1.905217	1.836562	0.441782	8.031592	0.048122
AP(1)(10 пропусків)	0.984311	1.899629	1.836921	0.441979	8.015475	0.048147
AP(1)(15 пропусків)	0.984373	1.899204	1.838121	0.441989	8.003015	0.048151
	Заповнення пропусків нормально розподіленими величинами(5,10,15 значень)					
AP(1)(5 пропусків)	0.961227	5.013048	2.095854	0.412973	7.762677	0.045489
AP(1)(10 пропусків)	0.964579	4.995774	2.168566	0.42034	8.030101	0.046316
AP(1)(15 пропусків)	0.970291	4.527164	2.195121	0.427064	8.280872	0.047021
	Заповнення пропусків рівномірно розподіленими величинами (5,10,15 значень)					
AP(1)(5 пропусків)	0.925631	8.856709	1.639623	0.467016	8.105878	0.051599
AP(1)(5 пропусків)	0.925256	8.616504	1.526491	0.478304	8.339294	0.052729
AP(1)(5 пропусків)	0.896539	11.70821	1.782207	0.491936	8.377905	0.054297
	Експоненційне згладжування(Exponential, alpha = 0,6)					
AP(1)	0.990344	0.01063	1.499641	0.058445	5.631185	0.032945
	Згладжування за допомогою ковзного середнього(MA, T=2)					
AP(1)	0.989192	0.011906	1.1274	0.058444	5.690629	0.032872

Статистичні характеристики моделі процесу (індекс цін товаровиробників)

Рекомендації для подальших досліджень

- ▶ Розглянуту методику моделювання доцільно доповнити тестами на аналіз нелінійностей та типу нестационарності досліджуваних процесів.
- ▶ Доцільно увести комбінований критерій аналізу адекватності моделей для реалізації автоматизованого режиму вибору кращої моделі.
- ▶ Доповнити наведену методику моделювання альтернативними методами заповнення пропусків даних і методами оцінювання параметрів нелінійних моделей.

Висновки по роботі

- ▶ Реалізована методика моделювання забезпечує отримання адекватних моделей за умови відповідності даних вимогам інформативності та повноти.
- ▶ Якщо за міру придатності моделі для подальшого використання взяти критерій якості прогнозу $MAPE$ (середня абсолютна похибка у відсотках), то найкращим методом для заповнення пропусків у даних та побудови моделі є метод заповнення пропусків оцінками прогнозів.
- ▶ Також показано, що ефективним методом є заповнення пропусків випадковими величинами, які розподілені за законом Гауса.
- ▶ Застосування методів згладжування дає можливість покращити якість оцінок прогнозів та статистичних характеристик моделей досліджуваних процесів.
- ▶ Попередня обробка даних дає можливість підвищити якість остаточного результату – оцінок прогнозів та статистичних характеристик моделей процесів.

Публікація за темою роботи

- ▶ Дудка Б.Р. Реалізація методики побудови моделей часових рядів // Київ: ІПСА НТУУ «КПІ», «Системні науки та кібернетика», 2016.

Дякую за увагу!