

Автоматизований вибір прогнозуючих моделей у системах підтримки прийняття рішень

Виконав:

студент групи КА-54 - Муравійник Єгор Андрійович

Керівник: проф., д. т. н.

Бідюк Петро Іванович

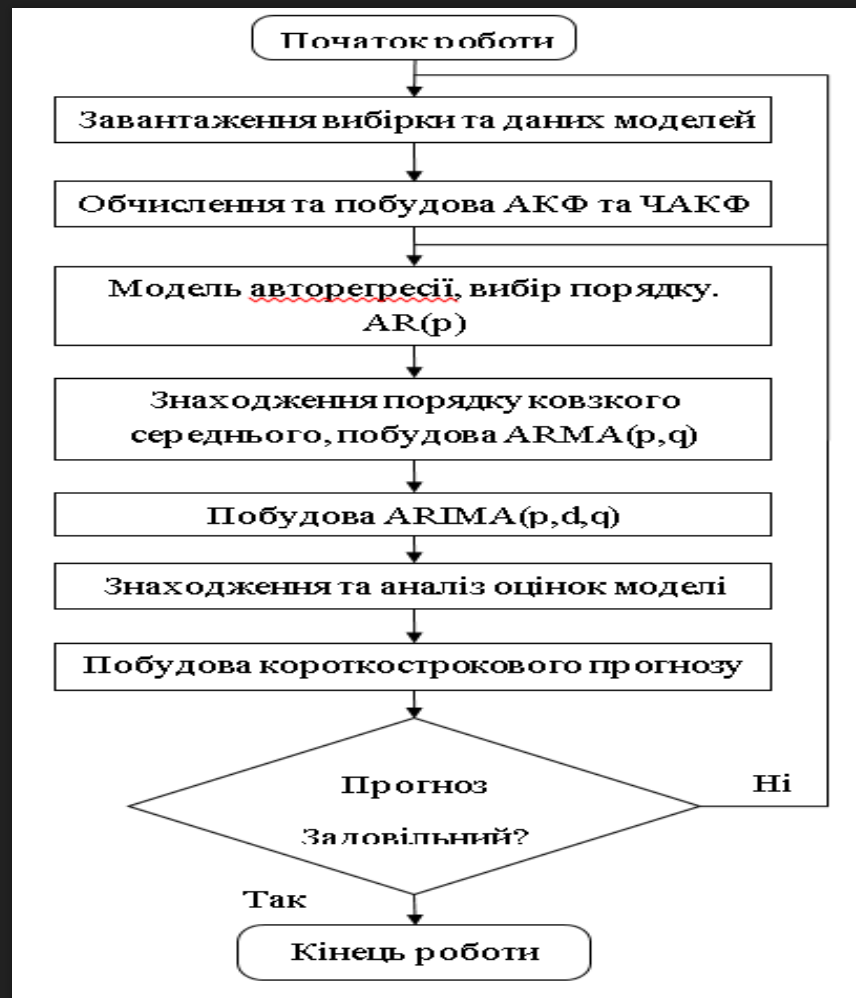
Об'єкт, предмет і мета дослідження:

- Об'єктом дослідження є вибрані фінансово-економічні процеси, представлені статистичними даними у формі часових рядів.
- Предмет дослідження: система підтримки прийняття рішень, методи статистичного і регресійного аналізу даних фінансових процесів.
- Мета даної роботи: створення інструментарію для побудови та автоматизованого вибору кращої моделі в процесі аналізу вибраних процесів.

Постановка задачі

1. Спроекувати і реалізувати СППР для моделювання і прогнозування вибраних фінансових процесів.
2. Застосувати СППР до аналізу вибраних процесів, побудувати прогнозуючі моделі.
3. Обчислити оцінки короткострокових прогнозів та статистичні параметри їх якості.
4. Виконати порівняльний аналіз використаних методів прогнозування з метою вибору кращого для даного процесу.
5. Виробити рекомендації стосовно можливостей розробленої системи до аналізу різних процесів.

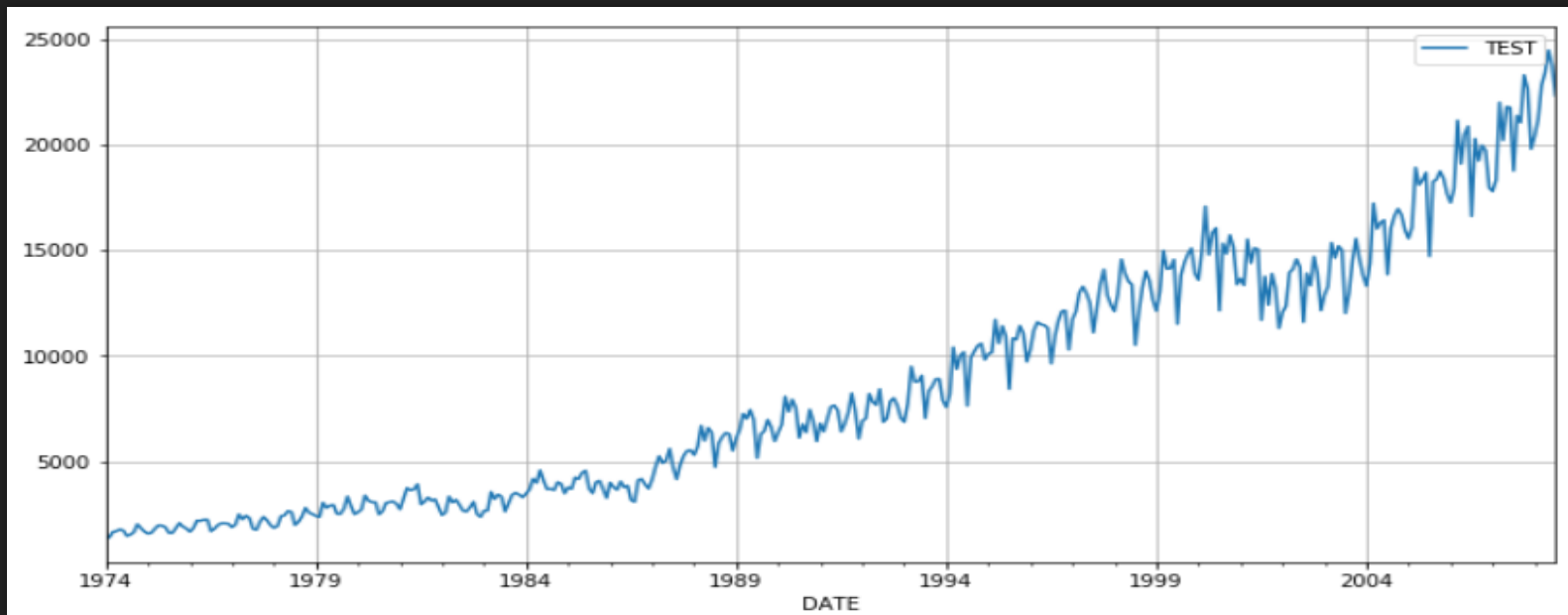
Блок-схема розробленої програми



Початок роботи програми

Для початку роботи програми треба завантажити файл з вибіркою даних у форматі .csv. Програма автоматично порахує усі допоміжні статистики.

Для прикладу було взято Розмір експорту зі США до Канади (мільйони доларів): 1974.1 – 2007.7.

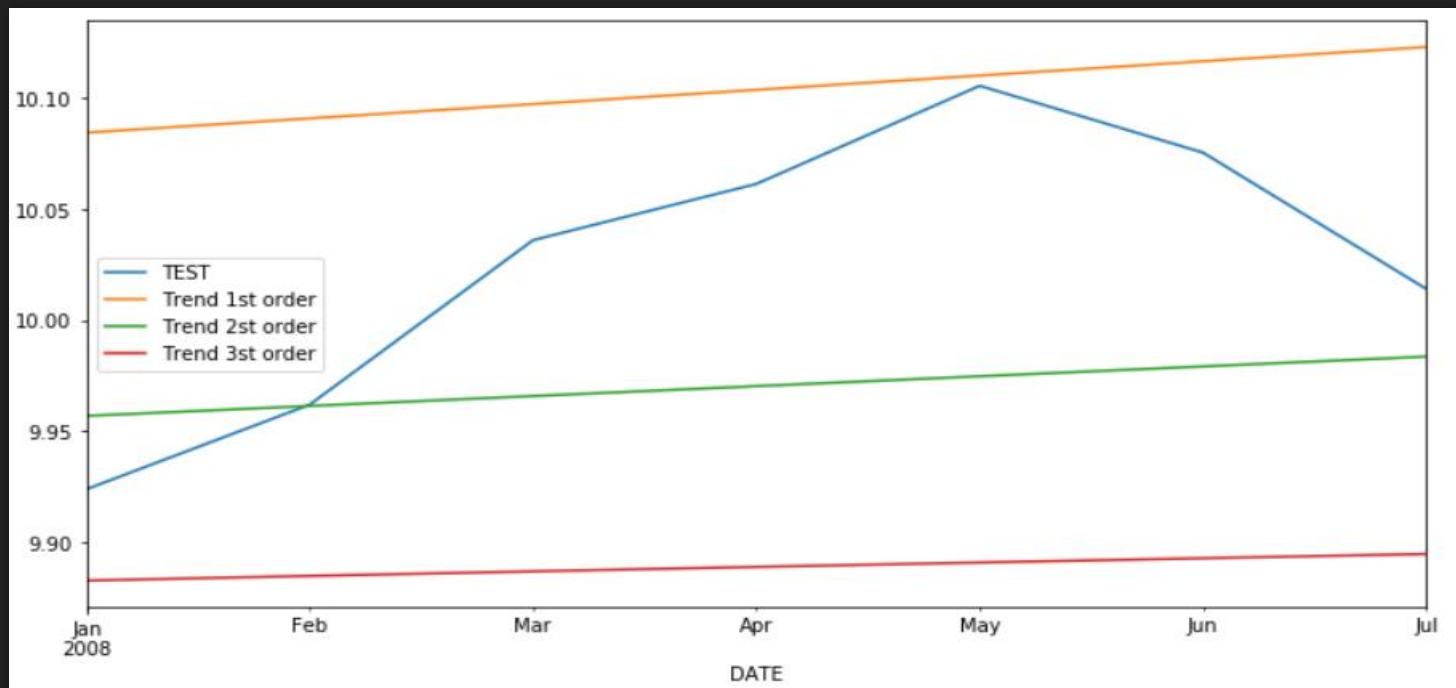


Побудова трендів перших трьох порядків:

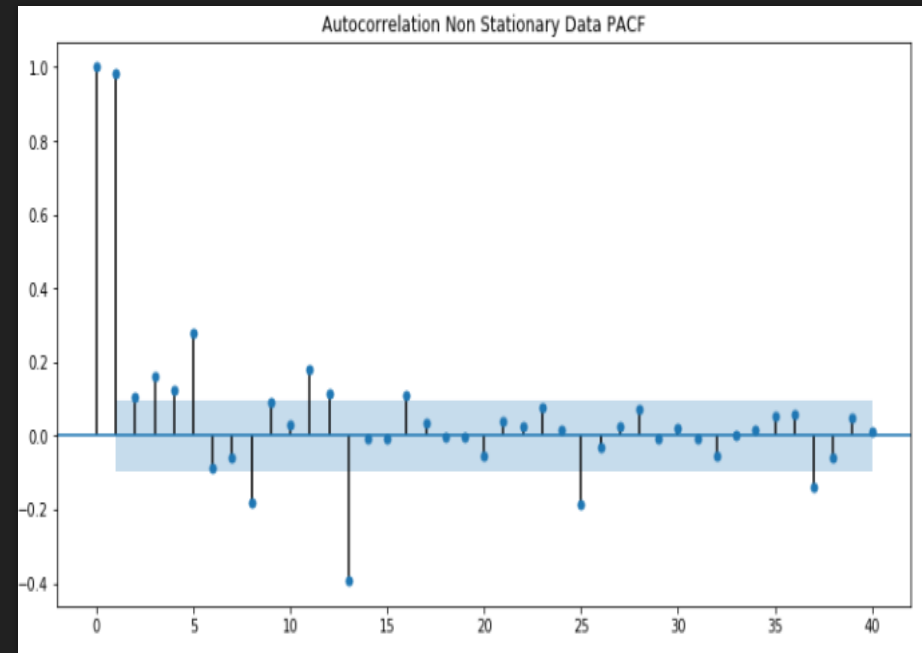
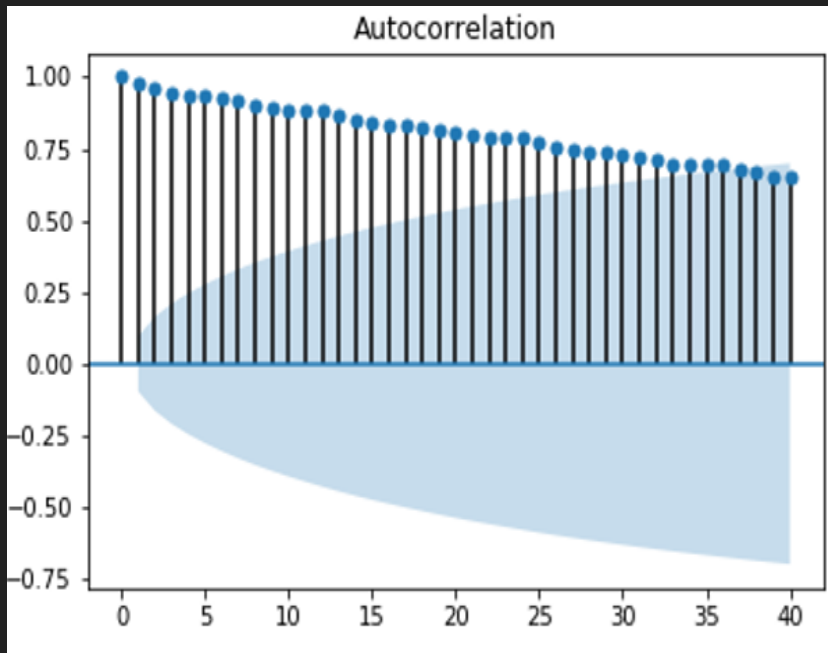
Тренд 1-го порядку: $y(k) = 7.443681 + 0.006453 * k$

Тренд 2-го порядку: $y(k) = 7.303271 + 0.008508 * k - 5.02E-06 * k * k$

Тренд 3-го порядку: $y(k) = 7.406829 + 0.005488 * k + 1.34E-05 * k * k - 3.01E-08 * k * k * k$

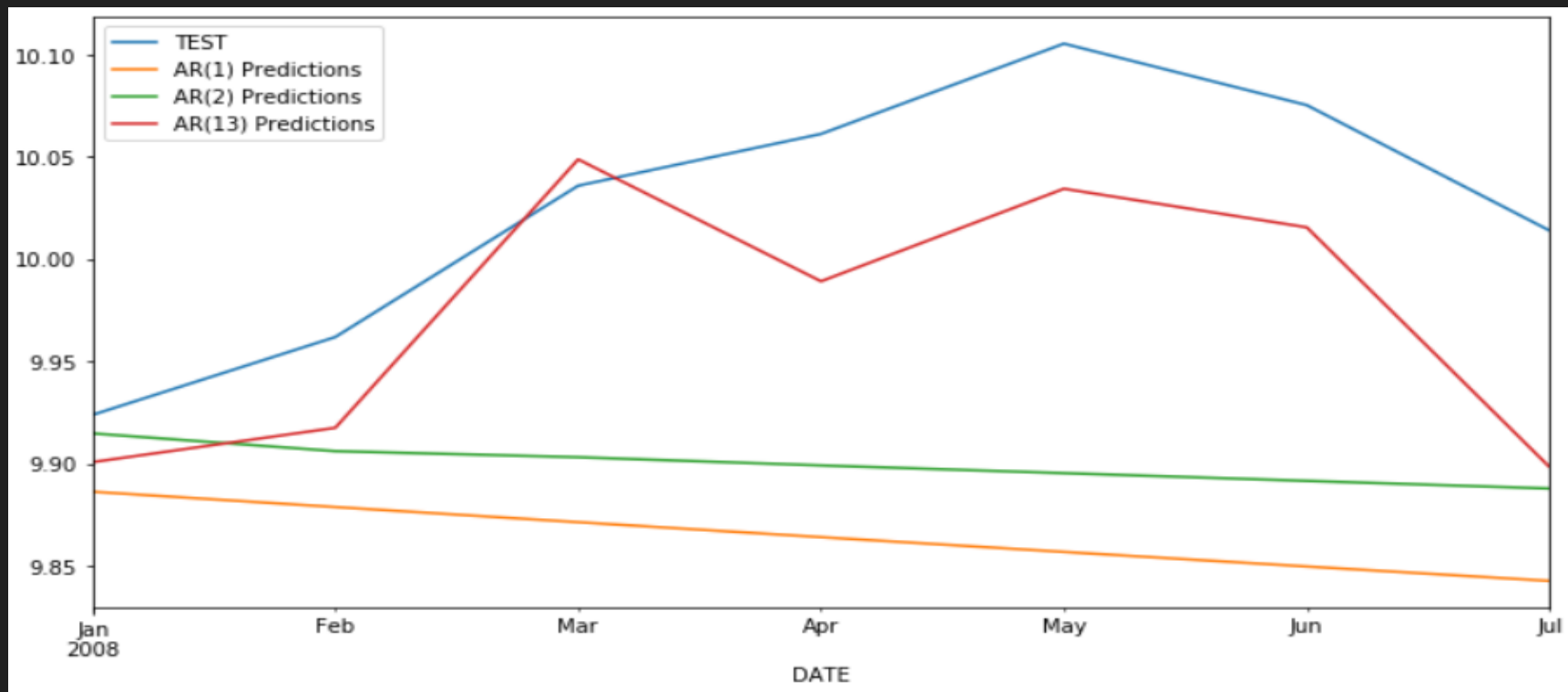


ACF та PACF для моделювання авторегресійних моделей :

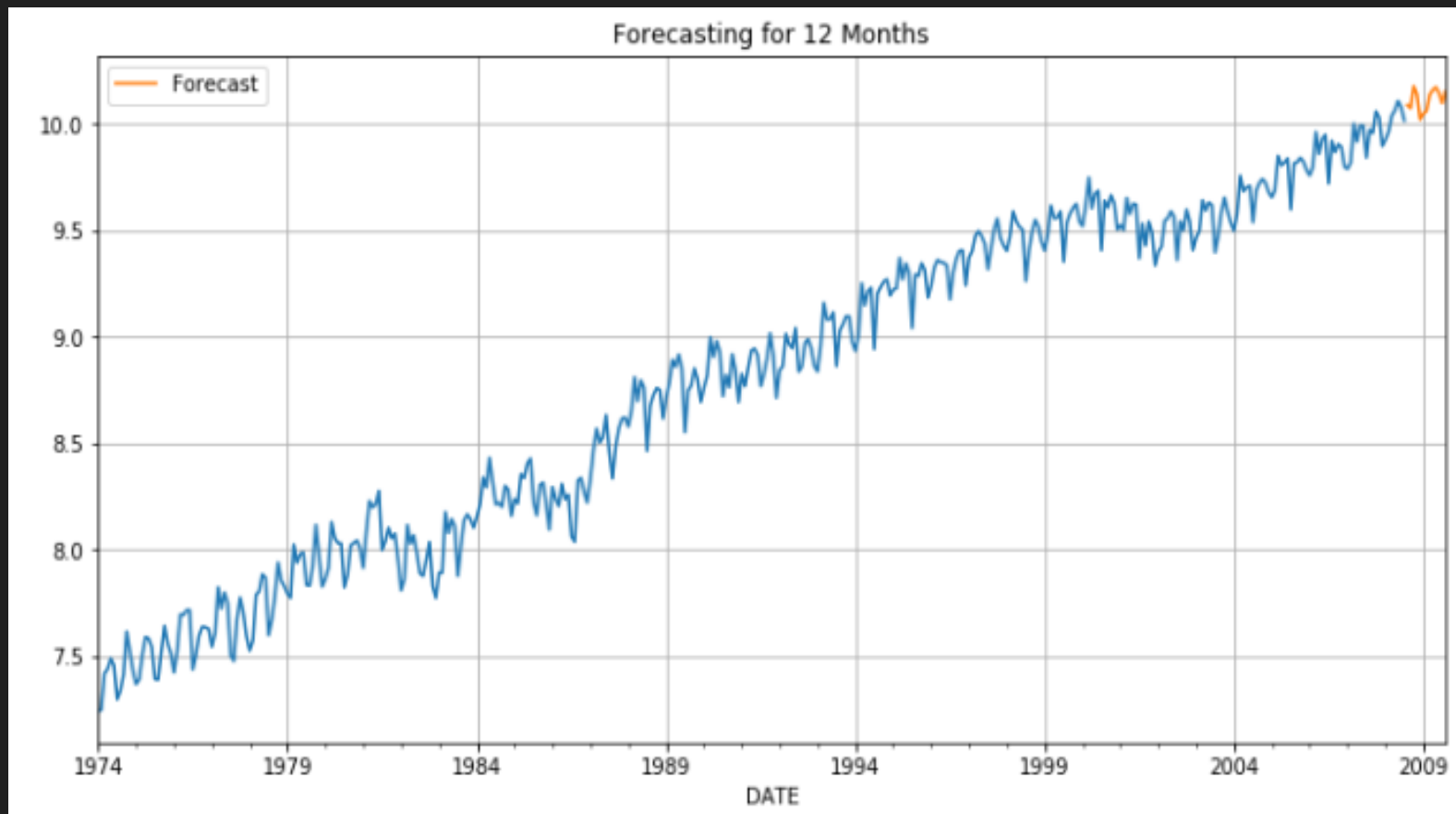


Вибір кращої авторегресійної моделі:

$$\text{AR}(13): y(k) = 12.14466 + 0.760891 \cdot \log y(k-1) + 0.811839 \cdot \log y(k-12) - 0.578414 \cdot \log y(k-13)$$

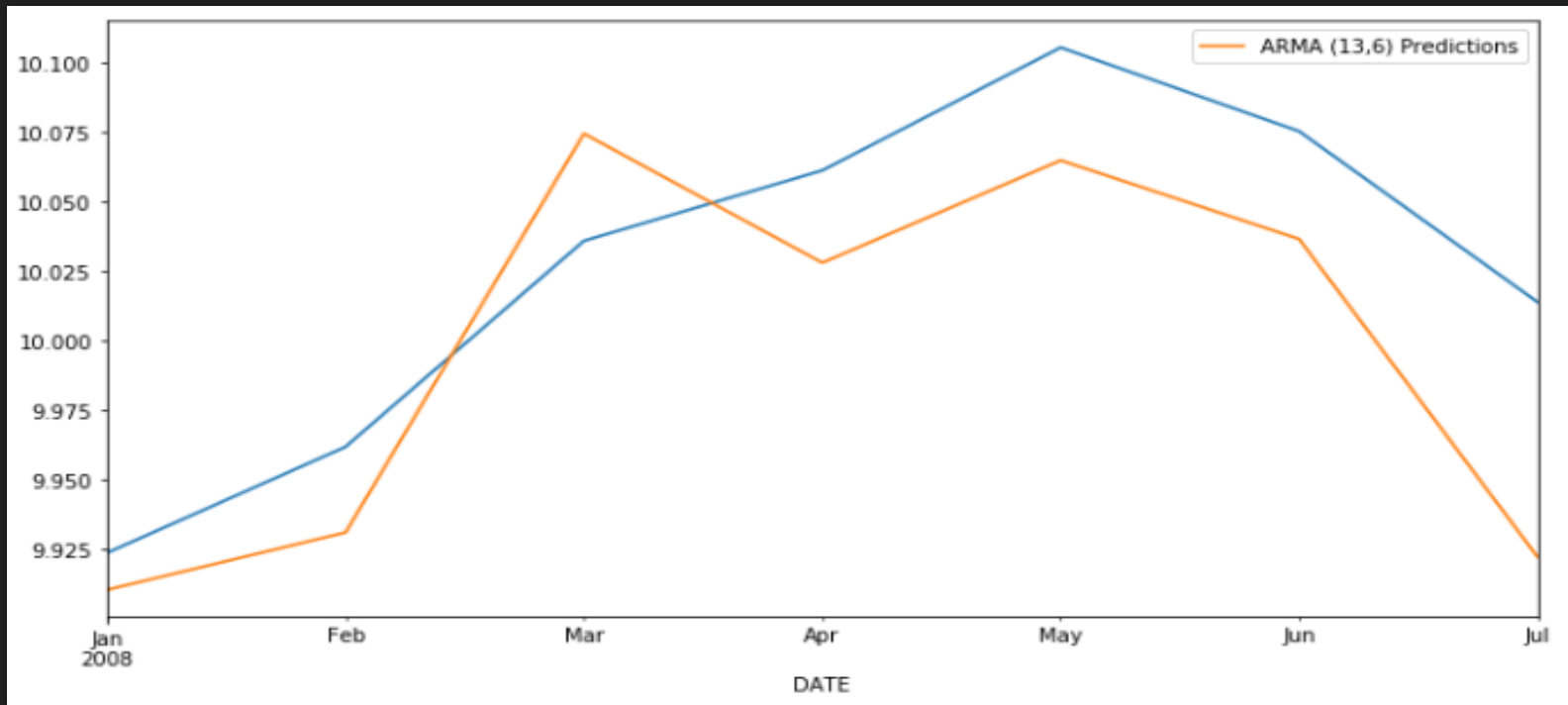


Прогноз моделі AR на 12 місяців:



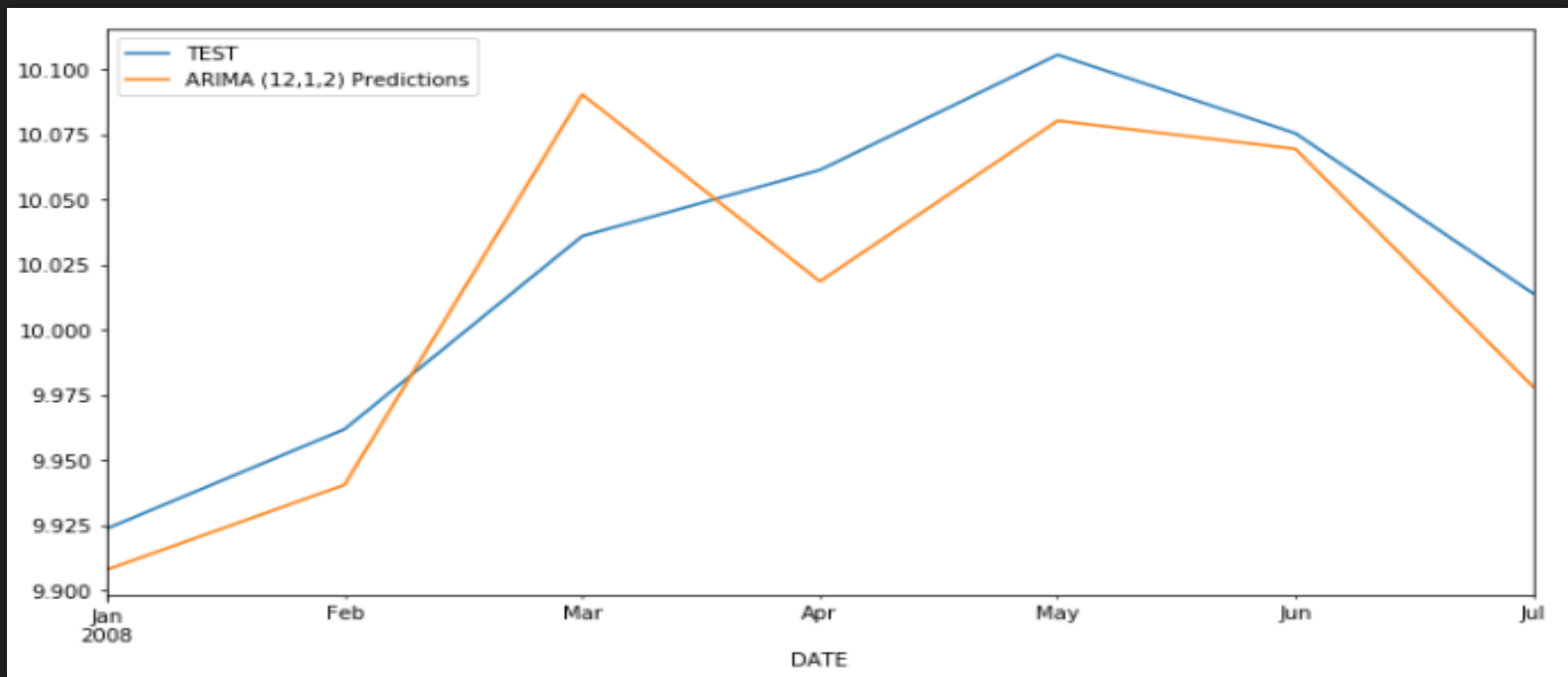
Побудова моделі ARMA(13,6):

ARMA(13,6): $y(k) = 11.77469 + 0.646552*y(k-1) + 0.825077*y(k-12) - 0.480886*y(k-13) + 0.261384*ma(k-5) + 0.306416*ma(k-6)$

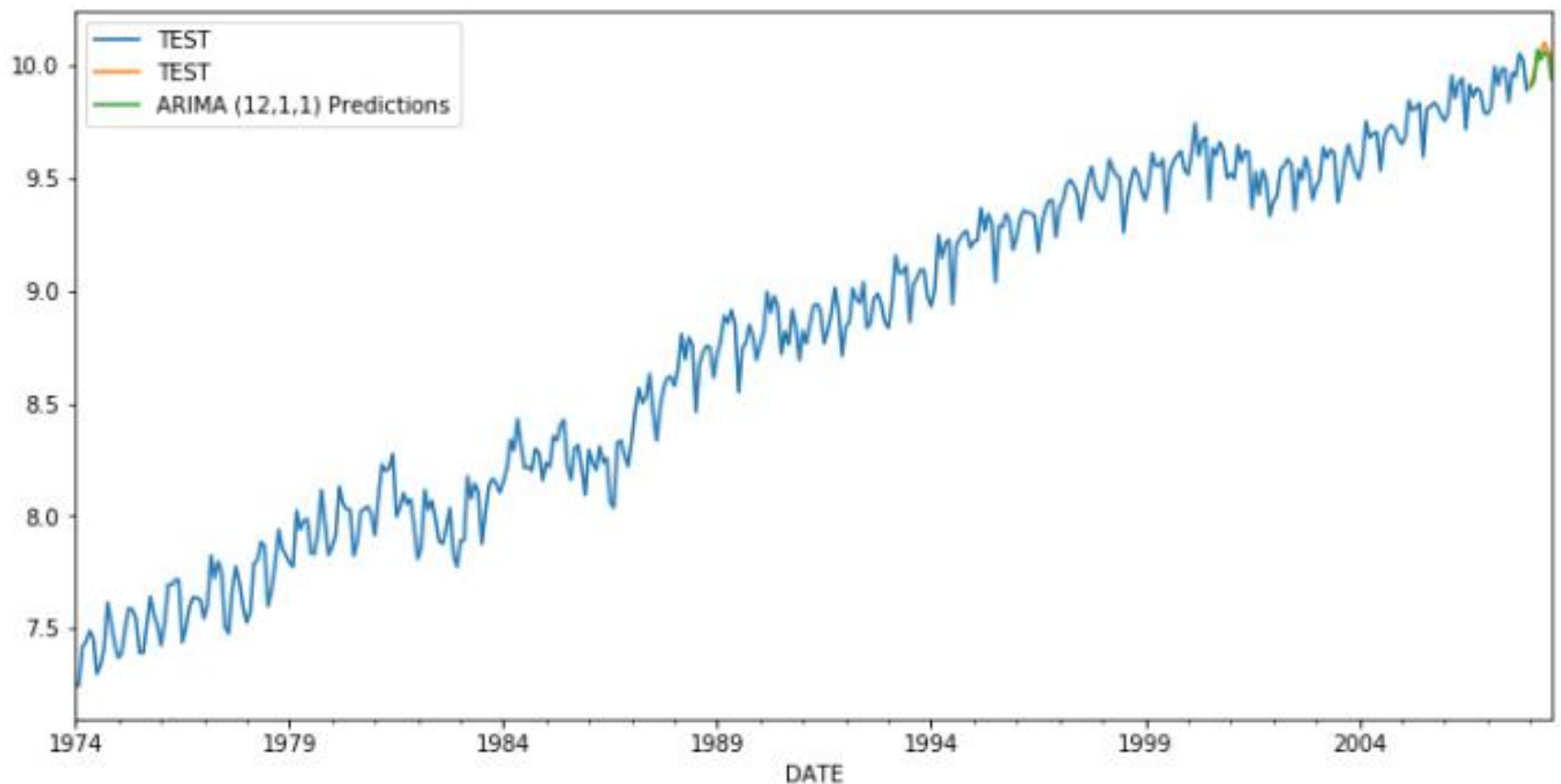


Побудова моделі ARIMA(12,1,2):

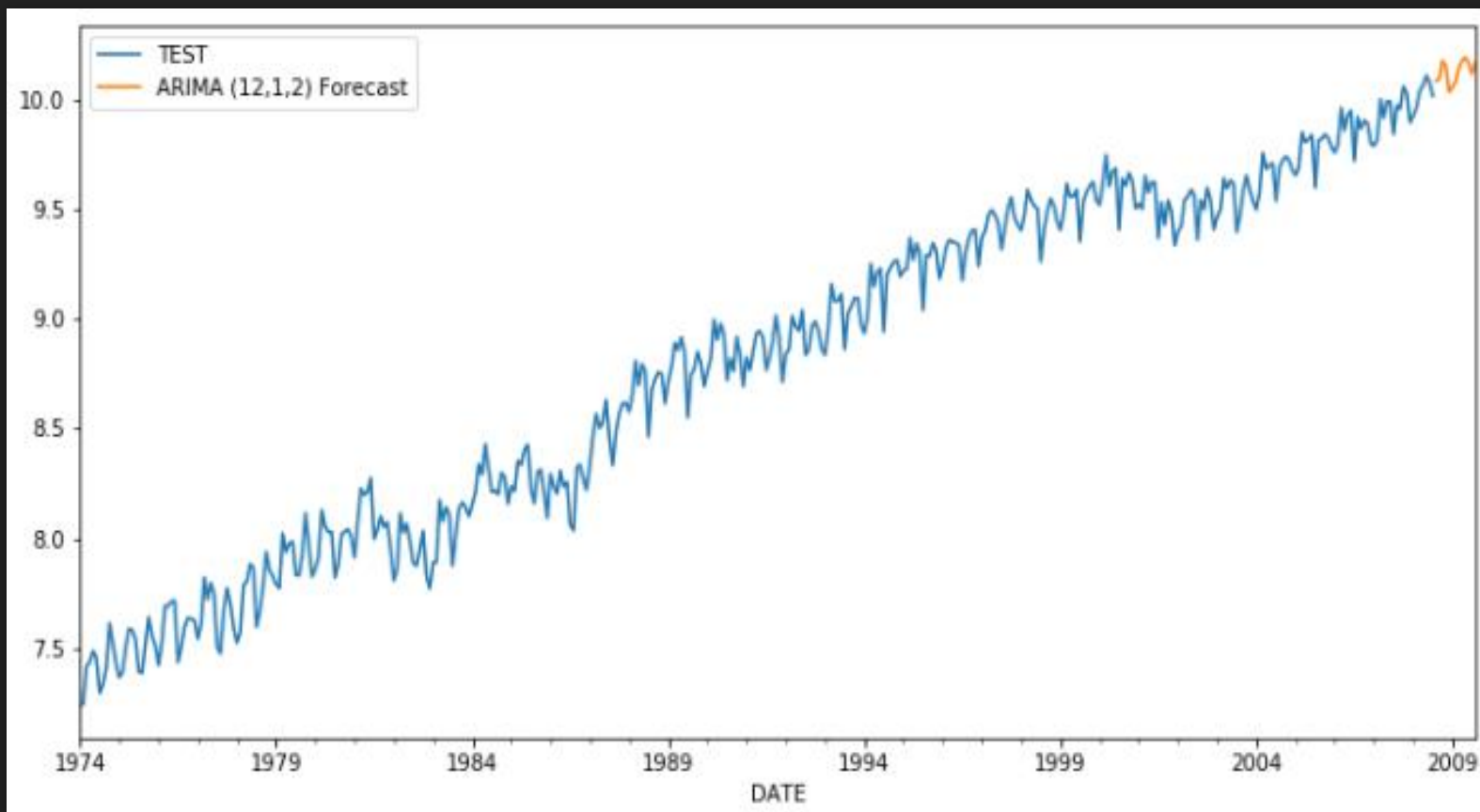
ARIMA(12,1,2): $y(k) = 0.006664 - 0.156679*y(k-3) - 0.022058*y(k-10) + 0.076470*y(k-11) + 0.748793*y(k-12) - 0.242140*ma(k-2)$



Модель ARIMA(12,1,2) в порівнянні з тестовою вибіркою:



Прогноз моделі ARIMA(12,1,2) на 12 місяців



Оцінки та порівняння різних моделей та прогнозу:

	R ²	Sum squared resid	DW	RMSE	MAE	MAPE	Theil
Тренд 1-го порядку	0,979665	8,429046571	0,611371	0.1420	0.1192	1.3594	0.0080
Тренд 2-го порядку	0,964948	6,673903179	0,739				
Тренд 3-го порядку	0,974241	6,114447888	0,81816				
ARIMA(12,1,2)	0,650623	1,722698851	2,735833	0.0651	0.0509	0.5917	0.0036
AR(13)	0,995521	1,693835131	2,397491	0.0617	0.0479	0.5579	0.0034
ARMA(13,6)	0,990497	1,512544612	2,24239	0.0659	0.0524	0.6144	0.0046

Час	Реальне значення <u>logy</u>	Математична модель			
		тренд1-го порядку	АРІКС(12,1,2)	АР(13)	АРКС(13,6)
2008M01	9.92373125901317	10.08302507815833	9.896053589564989	9.87591570015485	9.877501985581905
2008M02	9.961879094649684	10.08947824231132	9.955348208683675	9.928329990617381	9.916879716327719
2008M03	10.03601031166086	10.09593140646431	10.09052555636867	10.08857532804753	10.10477485378197
2008M04	10.06135844663414	10.1023845706173	10.00016177022341	9.970627802217427	9.994565288015391
2008M05	10.10558722652197	10.10883773477029	10.13148192065793	10.09971372276784	10.07997723551675
2008M06	10.0754599072717	10.11529089892328	10.07209469053909	10.08910348524205	10.08013412993134
2008M07	10.01399529844661	10.12174406307627	9.97189185092043	9.946652342399974	9.96491893678675

Порівняння результатів з програмою Eviews:

Оцінки програми:

	R^2	Sum squared resid	DW
Тренд 1-го порядку	0,979665	8,429046571	0,611371
Тренд 2-го порядку	0,964948	6,673903179	0,739
Тренд 3-го порядку	0,974241	6,114447888	0,81816
ARIMA(12,1,2)	0,650623	1,722698851	2,735833
AR(13)	0,995521	1,693835131	2,397491
ARMA(13,6)	0,990497	1,512544612	2,24239

Оцінки Eviews:

Equation: TREND1 Workfile: EXPCA:Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOGY
 Method: Least Squares (Marquardt - EViews legacy)
 Date: 10/28/16 Time: 18:03
 Sample: 1974M01 2007M12
 Included observations: 408
 LOGY=C(1)+C(2)*K

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	7.443681	0.014193	524.4774	0.0000
C(2)	0.008453	6.01E-05	107.3022	0.0000

R-squared	0.965939	Mean dependent var	8.763353
Adjusted R-squared	0.965885	S.D. dependent var	0.774282
S.E. of regression	0.143075	Akaike info criterion	-1.046013
Sum squared resid	8.310948	Schwarz criterion	-1.026349
Log likelihood	215.3866	Hannan-Quinn criter.	-1.038232
F-statistic	11513.77	Durbin-Watson stat	0.602805
Prob(F-statistic)	0.000000		

Equation: AR13 Workfile: EXPCA:Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOGY
 Method: ARMA Conditional Least Squares (Marquardt - EViews legacy)
 Date: 10/29/16 Time: 18:11
 Sample (adjusted): 1975M02 2007M12
 Included observations: 395 after adjustments
 Convergence achieved after 4 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12.14466	2.702624	4.493655	0.0000
AR(1)	0.760891	0.033547	22.68119	0.0000
AR(12)	0.811839	0.030171	26.90815	0.0000
AR(13)	-0.578414	0.041054	-14.08910	0.0000

R-squared	0.992296	Mean dependent var	8.808010
Adjusted R-squared	0.992236	S.D. dependent var	0.745784
S.E. of regression	0.065712	Akaike info criterion	-2.597005
Sum squared resid	1.688348	Schwarz criterion	-2.556713
Log likelihood	516.9085	Hannan-Quinn criter.	-2.581041
F-statistic	16786.36	Durbin-Watson stat	2.389724
Prob(F-statistic)	0.000000		

Equation: AR13MA6 Workfile: EXPCA:Untitled\

View Proc Object Print Name Freeze Estimate Forecast Stats Resids

Dependent Variable: LOGY
 Method: ARMA Conditional Least Squares (Marquardt - EViews legacy)
 Date: 10/29/16 Time: 18:12
 Sample (adjusted): 1975M02 2007M12
 Included observations: 395 after adjustments
 Convergence achieved after 16 iterations
 MA Backcast: 1974M08 1975M01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.77469	2.177217	5.408139	0.0000
AR(1)	0.646552	0.040034	16.15024	0.0000
AR(12)	0.825077	0.030157	27.35922	0.0000
AR(13)	-0.480886	0.045014	-10.68304	0.0000
MA(5)	0.261384	0.047497	5.503146	0.0000
MA(6)	0.306416	0.046396	6.331387	0.0000

R-squared	0.993080	Mean dependent var	8.808010
Adjusted R-squared	0.992991	S.D. dependent var	0.745784
S.E. of regression	0.062437	Akaike info criterion	-2.694231
Sum squared resid	1.516489	Schwarz criterion	-2.633793
Log likelihood	538.1107	Hannan-Quinn criter.	-2.670285
F-statistic	11164.89	Durbin-Watson stat	2.248238
Prob(F-statistic)	0.000000		

Висновки

1.

- Розроблено програмний продукт для моделювання та аналізу фінансово-економічних процесів (інструментальна система Python). Отримано високі результати прогнозування вибраних процесів.

2.

- Створена комп'ютерна система для моделювання і прогнозування на платформі Python (jupyter). Вона може бути використана при моделюванні і прогнозуванні реальних процесів, а також для навчального процесу.

3.

- Побудовані нові математичні моделі для вибраних процесів, які використані для оцінювання коротко- і середньострокових прогнозів.

Перспективи подальших досліджень:

- Реалізувати інші методи прогнозування, такі як: експоненційне згладжування, МГВА, нейромережі та ін.
- Реалізувати нові методи оцінювання параметрів моделі – метод максимальної правдоподібності і Монте-Карло для марковських ланцюгів (МКМЛ).
- Створити інтегральний критерій для автоматизованого вибору кращої моделі.

Дякую за увагу!