УДК 336.72

Пугачов Олександр Миколайович
студент Научально-наукового комплексу
«Інститут прикладного системного аналізу»
НТУУ «КПІ», м. Київ, Україна

АНАЛІЗ ДЕЯКИХ МОДЕЛЕЙ ДЛЯ ПРОГНОЗУВАННЯ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ

Анотація. *У роботі проведено порівняльний аналіз деяких моделей авторегресії на прикладі нестаціонарних процесів у фінансово-економічній сфері.*

Ключові слова: *авторегресія, різницеві рівняння, процес з трендом, гетероскедастичний процес.*

На сьогоднішній день в спеціальній літературі описано досить велика кількість методів прогнозування лінійних та нелінійних процесів на основі використання часових рядів. Однак, практика показує, що одного, навіть досить універсального методу, недостатньо для досягнення повноти аналізу процесу. Так, коректний аналіз гетероскедастичних процесів вимагає застосування моделей спеціальної структури для описання умовної дисперсії, що не забезпечує МГВА [1, с.132]. Кожний метод має також свої недоліки і переваги щодо обчислювальних витрат та характеристик точності прогнозу. Так, висока точність прогнозу за допомогою МГВА або нейромережі іноді досягається за рахунок високих обчислювальних витрат. Суттєвий виграш можна досягти в такому випадку за допомогою простіших моделей авторегресії: АР, АРКС та АРІКС, перевагами яких є простота структури та можливості оперативної адаптації до характеристик процесу в реальному часі [5, с.116].

Метою даної роботи є проведення огляду найбільш актуальних моделей для побудови моделей фінансово економічних процесів, що можуть бути використані для оцінювання прогнозів та подальшого прийняття рішень.

Досліджувані процеси

Нехай процес описується змінною вигляду $x(k)$, $k=\overline{1,n}$, де $k-$ дискретний час.

У випадку, коли $E\big(x(k)\big)\neq const$, тобто математичне сподівання змінюється в часі, то такий процес називають процесом з трендом або інтегрованим процесом (по аналогії із характером зміни сигналу на виході інтегратора) або процесом з одиничними коренями (відповідного характеристичного рівняння).

При цьому під трендом будемо розуміти поточне середнє значення процесу, яке може бути отримане за допомогою процедури цифрової фільтрації. В простому випадку це може бути формула для обчислення поточного середнього значення.

Наявність нелінійного детермінованого тренду в процесі можна визначити шляхом оцінювання рівняння:

$$y(k)=a_m x^m + a_{m-1}x^{m-1} + \ldots + a_1 x + a_0 \qquad (1)$$

яке представляє собою поліном порядку m відносно часу. Якщо хоча б один із коефіцієнтів $a_i, i=1,\ldots,m$ є статистично значимим, то гіпотеза щодо відсутності тренду відхиляється [3, с. 124]. Крім того, тренд може бути логарифмованим, експоненціальним, і загалі, будь-якою функцією від часу. У даній роботі розглядається поліноміальний та експоненціальний тренд. Якщо тренд відносно швидко змінює свій напрям розвитку і для нього важко знайти адекватне функціональне описання, то застосовують моделі випадкових трендів, які ґрунтуються на комбінаціях випадкових величин.

Інший тип нестаціонарних процесів — гетероскедастичні — це процеси із змінною в часі дисперсією, а гомоскедастичними — процеси із сталою дисперсією на відрізку часу, що розглядається при моделюванні та прогнозуванні.

Дисперсію та стандартне відхилення часто використовують як міру ризику при дослідженні фінансово-економічних процесів, а тому цій проблемі приділяється значна увага в спеціальній літературі. При дослідженні фінансових процесів дисперсію та стандартне відхилення використовують як міру волатильності (мінливості) процесу.

В технічних процесах дисперсія — міра розсіювання вимірів. Вона може характеризувати поточний стан механізмів, технологічних процесів, інтенсивність випадкових збурень, що впливають на технічні системи; інтенсивність похибок (шумів) вимірів і т. ін., тому це надзвичайно важливий статистичний параметр з точки зору дослідження поточного та прогнозування майбутнього стану системи [3, с.125].

В соціальних дослідженнях дисперсію і стандартне відхилення використовують для порівняння між собою різних соціальних груп або різних множин індивідуумів, що відносяться до однієї й тієї ж соціальної групи.

Оскільки дисперсія може змінюватись у часі, то важливо мати таку математичну модель, яка дозволить коректно описувати поведінку дисперсії та прогнозувати її значення на один або більше кроків наперед. Це дасть можливість покращувати якість рішень по відношенню до управління процесами із змінною в часі дисперсією. Наприклад, коректно формулювати правила прийняття рішень щодо купівлі/продажу акцій на біржі, коректно оцінювати міру кредитного та інших ризиків в банківській діяльності. У технічних системах дисперсія та стандартне відхилення — це міри відхилення розмірів від заданих, ступінь зношеності механізму, рівень шуму і т.ін. Так, дисперсію та стандартне відхилення використовують в системах аналізу якості продукції при побудові контрольних карт, при визначенні точності розмірів і т. ін. [2, с.42].

Формально процес вважається гетероскедастичним (ГСП), якщо

$$Var\left[x(k)\right] \neq const \quad (2)$$

Припущення щодо гомоскедастичності означає, що варіація кожної випадкової величини $x(k)$ навколо її математичного сподівання залишається сталою величиною незалежно від значень факторів. Тобто дисперсія процесу не є функцією часу [3, с. 135].

Різницеві рівняння

Різницеві рівняння — зручний інструмент для аналізу та прогнозування економетричних часових рядів. Так склалося історично, що різницеві рівняння посідають, фактично, центральне місце в процесі розв'язку практичних економічних задач. Застосування різницевих рівнянь забезпечує наступні переваги:

- простота оцінювання структури і параметрів математичних моделей за допомогою стандартних пакетів прикладних програм на персональних комп'ютерах;
- наявність широкого спектру рекурсивних методів для обчислювання оцінок параметрів рівнянь;
- відносна простота знаходження та аналізу розв'язків отриманих рівнянь;
- можливість використання різницевих рівнянь для короткочасного та довгострокового прогнозування;
- побудова моделей в просторі станів з метою синтезу систем оптимального та адаптивного управління;
- відносна простота аналізу багатовимірних часових рядів [3, с. 138].

Деякі типи регресійних та різницевих рівнянь

Авторегресія (АР(p)): рівняння авторегресії описує пам'ять процесу, тобто вплив значень попередніх станів на його поточний стан:

$$y(k) = a_0 + \sum_{i=1}^{p} a_i y(k-i) + \varepsilon(k), \quad (3)$$

де $a_i, i = 1..p$ — коефіцієнти моделі, які оцінюються на основі значень часового ряду;

p — порядок авторегресії, який визначається числом затриманих в часі значень ряду, що використовуються в правій частині рівняння для описання динаміки змінної в момент;

$k, k = 1,2,\ldots$ – дискретний час;

$\varepsilon(k)$ – випадкова величина, поява якої зумовлена наступними причинами:

- вплив випадкових збурень на процес, що моделюється;
- похибки рівняння, зумовлені неточно вибраною структурою (можливо, що не враховано деякі регресори, введено непотрібні незалежні змінні або робиться спроба моделювати нелінійний процес за допомогою лінійного рівняння);
- методичні і обчислювальні похибки, які з'являються при обчисленні оцінок коефіцієнтів рівняння [4].

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

$$y(k) = a_0 + \sum_{i=1}^{p} a_i y(k-i) + \sum_{j=1}^{q} b_j \varepsilon(k-i) + \varepsilon(k), \quad (4)$$

де q – порядок ковзного середнього, b_j – коефіцієнтиковзного середнього.

Авторегресія з інтегровним ковзним середнім (АРІКС(p, d, q)):

$$\Delta^d y(k) = a_0 + \sum_{i=1}^{p} a_i \Delta^d y(k-i) + \sum_{j=1}^{q} b_j \varepsilon(k-i) + \varepsilon(k), \quad (5)$$

де Δ^d — оператор різниці часового ряду порядку d (послідовне взяття d разів різниць першого порядку — спочатку від часового ряду, потім від отриманих різниць і т.д.)

Авторегресійне рівняння з умовною гетероскедастичністю (АРУГ(p)):

$$\sigma_k^2 = a_0 + \sum_{k=1}^{p} a_k \varepsilon^2(k-i), \quad (6)$$

де σ_k^2 – дисперсія процесу у момент часу k, $a_k, k = 0..p$ – коефіцієнти моделі, $\varepsilon^2(k)$ – залишки моделі АР(p).

Для оцінки параметрів моделі використаємо метод найменших квадратів (МНК).

Результати

Проведемо дослідження на прикладі дослідження курсу валют долар до гривні [6]. Були узяті денні спостереження за останні 60 днів починаючи з 01.05.2016). Нижче приведений графік процесу:

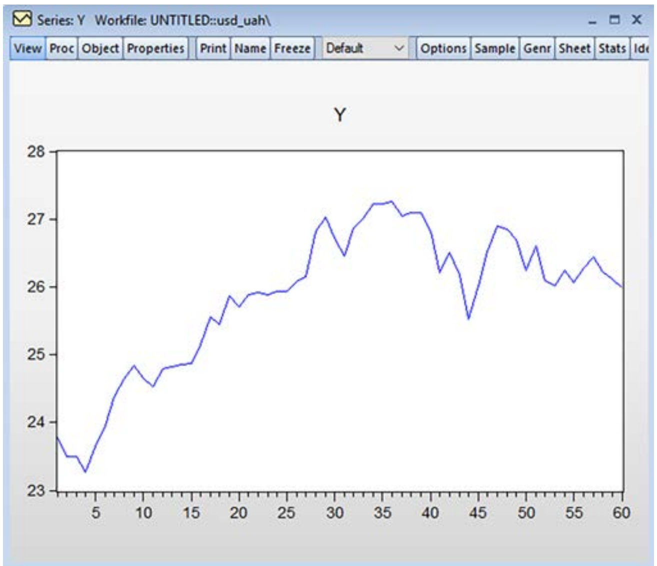
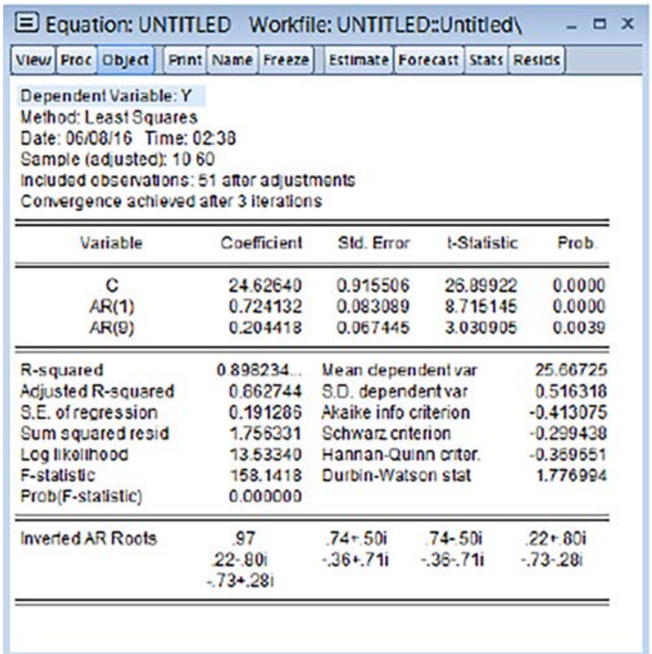


Рисунок 1. Нестаціонарний процес курсу валют долара до гривні

Побудуємо модель авторегресії: аналіз ЧАКФ (часткової автокореляційної функції) вказує про доцільність включення лагу 2 та 9 порядку:



Equation: UNTITLED Workfile: UNTITLED::Untitled\

Dependent Variable: Y
Method: Least Squares
Date: 06/08/16 Time: 02:38
Sample (adjusted): 10 60
Included observations: 51 after adjustments
Convergence achieved after 3 iterations

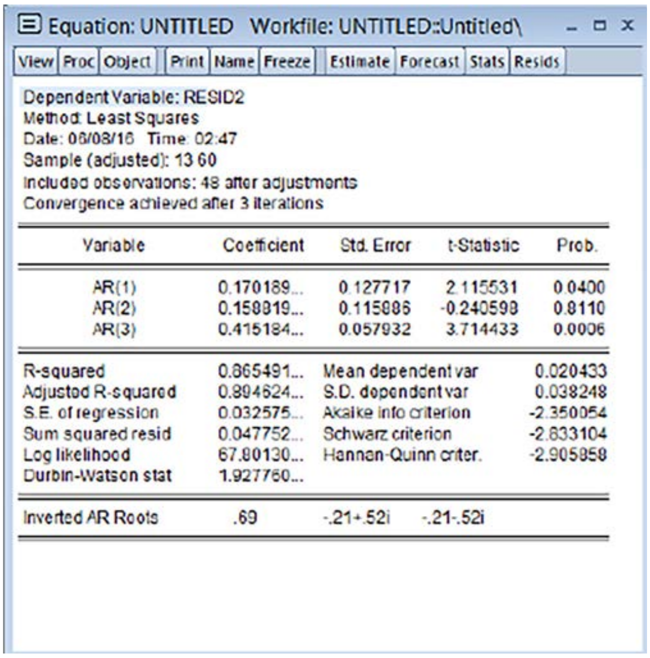
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	24.62640	0.915506	26.89922	0.0000
AR(1)	0.724132	0.083089	8.715145	0.0000
AR(9)	0.204418	0.067445	3.030905	0.0039

R-squared	0.898234...	Mean dependent var	25.66725
Adjusted R-squared	0.862744	S.D. dependent var	0.516318
S.E. of regression	0.191286	Akaike info criterion	-0.413075
Sum squared resid	1.756331	Schwarz criterion	-0.299438
Log likelihood	13.53340	Hannan-Quinn criter.	-0.369651
F-statistic	158.1418	Durbin-Watson stat	1.776994
Prob(F-statistic)	0.000000		

Inverted AR Roots .97 .74+.50i .74-.50i .22+.80i
.22-.80i -.36+.71i -.36-.71i -.73-.28i
-.73+.28i

Рисунок 2. АР(9) модель

Тест ряду з квадратів залишків моделі АР(9) вказує на наявність гетероскедастичності, тому побудуємо модель АРУГ (3), опишемо умовну дисперсію вихідного процесу, оцінивши ряд з квадратів залишок моделі АР(9).



Equation: UNTITLED Workfile: UNTITLED::Untitled\

Dependent Variable: RESID2
Method: Least Squares
Date: 06/08/16 Time: 02:47
Sample (adjusted): 13 60
Included observations: 48 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.170189...	0.127717	2.115531	0.0400
AR(2)	0.158819...	0.115886	-0.240598	0.8110
AR(3)	0.415184...	0.057932	3.714433	0.0006

R-squared	0.865491...	Mean dependent var	0.020433
Adjusted R-squared	0.894624...	S.D. dependent var	0.038248
S.E. of regression	0.032575...	Akaike info criterion	-2.350054
Sum squared resid	0.047752...	Schwarz criterion	-2.833104
Log likelihood	67.80130...	Hannan-Quinn criter.	-2.905858
Durbin-Watson stat	1.927760...		

Inverted AR Roots .69 -.21+.52i -.21-.52i

Рисунок 3. АРУГ(3) модель

Таким чином, АРУГ(3) модель можна описати наступним рівнянням:

$$\widehat{\sigma^2}(k) = 0.1702\varepsilon^2(k-1) + 0.1588\varepsilon^2(k-2) + 0.4152\varepsilon^2(k-3) \quad (7)$$

Для оцінки адекватності прогнозу візьмемо 54 значення і зробимо прогноз на наступні 6 днів (з 27.05.2005 по 01.06.2016). Результати приведені у таблиці:

Таблиця 1

АРУГ(3) модель

Спостереження	Реальне значення умовної дисперсії	EViews	
		Прогноз, грн.	Похибка, грн.
27.05.2016	0.027	0.0265	-0.0005
28.05.2016	0.019	0.0256	0.0064
29.05.2016	0.031	0.0305	-0.0001
30.05.2016	0.036	0.0304	-0.0060
31.05.2016	0.033	0.0267	-0.0061
1.06.2016	0.022	0.0215	-0.0002
САП			1.3e-5
САПП, %			35.4

Проведемо дослідження процесу ВВП США узявши поквартальні значення з 01.01.1947 по 01.01.2016 [7]. Нижче представлений графік процесу:

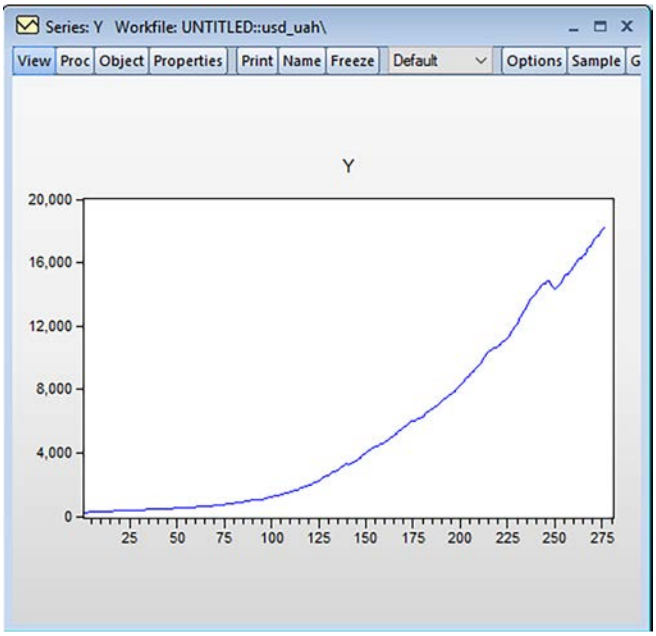


Рисунок 4. Процес ВВП США

З малюнку видно чітко виражений тренд, побудуємо поліномуальну модель його опису та прогноз (перевірочна вибірка — 274 значення, наступні 6 — тестова).

Статистичний пакет EViews з використанням методу оцінки МНК дав наступні результати:

$$y(k) = 150.59 + 17.09k - 0.3765k^2 + 0.0038k^3 - 0.000067k^4, \quad k = 1..284 \qquad (8)$$

Оцінка адекватності прогнозу виявилась наступною:

Таблиця 2

Поліноміальна модель

Спостереження	Реальне значення, млрд $	МНК	
		Прогноз, млрд. $	Похибка, млрд. $
2015–1	17913.7	17770.8	-142.9
2015–2	18060.2	17893.8	-166.4
2015–3	18164.8	18016.3	-148.5
2015–4	18229.5	18138.3	-91.2
2016–1		18259.7	
2016–2		18380.6	
2016–3		18500.8	
САП		137,2	
САШ, %		0.76%	

Таким чином, можна зробити висновок, що використання тієї чи іншої моделі певним чином залежить від конкретної задачі та попереднього аналізу нестаціонарного процесу. Авторегресійні моделі є найбільш простими з точки зору побудови та об'єму обчислювальних операцій і при цьому у більшості випадках модель процесу та прогнозу виявляється адекватною для особи, що приймає рішення.

Список використаних джерел

1. Бідюк П. І. Система підтримки прийняття рішень для аналізу і прогнозування стану підприємства / Бідюк П. І., Кожухівський А. Д., Кожухівська О. А. // Радіоелектроніка, інформатика, управління. — 2013. С. 128–136.

2. Киселев М. Средства добычи знаний в бизнесе и финансах / Киселев М., Соломатин Е. // Открытые системы. — 1997. — № 4. — С. 41–44.

3. Бідюк П. І. Аналіз часових рядів (навчальний посібник) [Текст] / Бідюк П. І., Романенко В. Д., Тимощук О. Л. — К.: Політехніка, 2010. — 317 с.

4. Бідюк П. І. Аналіз та моделювання економічних процесів перехідного періоду. — Київ: Інститут прикладного системного аналізу при НТУУ «КПІ», 2000. — 230 с.

5. Бидюк П. И. Системный подход к построению регрессионной модели по временным рядам / Бидюк П. И., Баклан И. В. // Системні дослідження та інформаційні технології, 2002 — с. 114–131.

6. Курс валют долар/гривня. [Електроний ресурс]. — Режим доступу: http: //investfunds. ua/markets/indicators/usduahnbu/?f s%5Bidx%5D=0&f s%5Bsdate%5D=01.06.2005& Bedate%5D=01.06.2016

7. Federal Reserve Bank of St. Louis [Електронний ресурс]. — Режим доступу: https://research.stlouisfed.org/fred2/series/GDP#