

ФОРМУВАННЯ РЕПРЕЗЕНТАТИВНОЇ СТАТИСТИЧНОЇ ВИБІРКИ ЗА КОТИРУВАЛЬНИМИ ДАНИМИ УКРАЇНСЬКОГО ФОНДОВОГО РИНКУ

Постановка проблеми. При моделюванні оптимального портфеля цінних паперів важливим є питання «чистоти» котирувальних даних, які використовуються для розрахунків дохідності та ризику портфеля.

Класична теорія портфеля побудована для розвинених високоліквідних фондових ринків Заходу, тому її застосування до українських цінних паперів потребує певної адаптації з урахуванням специфіки функціонування вітчизняного фондового ринку.

На відміну від ліквідних фондових ринків з великими обсягами угод, високим рівнем інформаційної «прозорості», операції на яких здійснюються за реальними ринковими цінами, український фондовий ринок час від часу характеризується занадто екстремальними або незвичайними значеннями у рядах дохідностей акцій (так званими «викидами»), які можуть спотворювати справжню картину. Тому, реалії вітчизняного ринку цінних паперів вимагають застосування спеціальних математичних методів, які дозволяють перед використанням даних у подальших розрахунках «очистити» їх шляхом виключення аномальних цін угод.

З огляду на те, що фондові ринки країн, що розвиваються, привертають до себе все більшу увагу через великі можливості потенційного заробітку актуальність обраної теми не викликає заперечень.

Зазначимо, що дана стаття є продовженням публікації результатів авторських досліджень в цій сфері. Перша спроба аналізу котирувальних цін на організованому українському фондово-

му ринку була здійснена Л.Б. Долінським ще у 2000 році [1]. Нажаль, більш ніж через десятиріччя з того часу вітчизняний фондовий ринок, все ще характеризується недостатньою ліквідністю та значними коливаннями котирувальних цін, тому дослідження, спрямовані на формування репрезентативної статистичної вибірки, не втратили актуальності.

Крім того, у попередній статті Д.О. Пріми [2] виникла необхідність визначення обсягу вибірки доходностей акцій на українському фондовому ринку, усереднених за декілька торговельних днів, до якого можна застосовувати класичний інструментарій моделювання оптимального портфеля цінних паперів, заснований на властивостях нормального розподілу.

Деякі питання дослідження якості даних, що використовуються для експерименту, перевірки їх повноти, а також здійснення їх очистки вже піднімалися у роботах Тамрапарні Дасу та Теодора Джонсона [3]. Що стосується обсягу вибірки, достатнього для застосування методів заснованих на характеристиках нормального розподілу, то це питання обговорювалось у роботах як вітчизняних авторів (Гіхман, Скороход, Ядренко, Черняк і т.д.), так і зарубіжних (Брейс (Brase), Майерс (Myers), Мур (Moore) і т.д.).

Ціль статті - здійснення «очистки» рядів доходностей акцій на українському фондовому ринку та визначення допустимого обсягу вибірки доходностей акцій на українському фондовому ринку, необхідного для застосування класичного інструментарію моделю-

Леонід Долінський
к.е.н., доцент
кафедри економіко-математичного моделювання, ДВНЗ «Київський національний економічний університет імені Вадима Гетьмана»
Дарина Пріма
аспірант кафедри економіко-математичного моделювання, ДВНЗ «Київський національний економічний університет імені Вадима Гетьмана»

РИЗИК-МЕНЕДЖМЕНТ

вання оптимального портфеля цінних паперів, заснованого на характеристиках нормального розподілу.

Для досягнення поставленої цілі розв'язано наступні завдання:

- ідентифікація «викидів» у рядах доходностей для кількох спеціально відібраних емітентів на українському фондовому ринку та їх видалення;
- аналіз характеру «викидів» за датами для кожного емітента;
- аналіз рядів доходностей після очищення;
- перевірка автокореляції на оновлених даних;
- перевірка початкових та усереднених даних на відповідність нормальному розподілу.

Виклад основного матеріалу. Ідентифікація викидів здійснюється дослідником суб'єктивно для кожного конкретного випадку. Деякі дослідники використовують чисельні методи видалення викидів. Так, наприклад, за класичними підходами математичної статистики [4] виключають значення, що лежать за межами стандартних відхилень. Для будь-якого набору даних, незалежно від форми розподілу, відсоток спостережень, що лежать на відстані, що не перевищує стандартних відхилень від математичного сподівання має складати не менше: Звідси, як мінімум 75% всіх спостережень з будь-якого набору даних міститься в інтервалі, принаймні 88,89% спостережень міститься в інтервалі, і як мінімум 93,75% спостережень міститься в інтервалі. Цю важливу властивість стандартного відхилення відкрили Чебишев і Бьєнаме незалежно один від одного більше ніж 100 років тому.

Початкова статистична вибірка. При формуванні статистичної вибірки даних на українському фондовому ринку доцільно розглядати лише ті цінні папери, угоди з якими здійснюються систематично та в достатніх для оцінювання обсягах.

Дані для аналізу було взято із сайту Української біржі. Хоча вони мають більш коротку ретроспективу, порівняно з іншими вітчизняними фондовими біржами, зокрема біржею ПФТС, тим не менш, на думку багатьох експертів, вони є більш якісними.

Статистичну вибірку було сформовано з набору емітентів, що входять до

індексного кошику Української біржі, тобто з «блакитних фішок» вітчизняного фондового ринку. Було взято щоденні котирування акцій з 01.04.09 по 11.08.11 таких компаній: ALMK (Алчевський металургійний комбінат), AZST (комбінат «Азовсталь»), BAVL (Райффазен Банк Аваль), CEEN (Центренерго), ENMZ (Єнакієвський металургійний завод), UNAF (Укрнафта), USCB (Укрсоцбанк), ZAEN (Західенерго).

Оскільки на Українській біржі акціями вище наведених емітентів почали торгувати лише з 2009 року, тому досліджувані вибірки, по відношенню до котирувань Української біржі, можна розглядати як генеральні сукупності. Отже, генеральні сукупності за кожним емітентом склали 586 спостережень.

«Очищення» рядів доходностей. За котируваннями кожного емітента було розраховано ряди доходностей. Саме на основі рядів доходностей було проведено подальші дослідження. З метою виключення значень, які не відображають реальні ринкові курси цінних паперів за оцінюваний період, було розраховано основні статистики (математичне сподівання, стандартне відхилення, медіана, скіс, ексцес), побудовано інтервали, побудовано гістограми у пакеті Eviews.

Проілюструємо хід проведення аналізу на прикладі акцій ALMK (Алчевський металургійний комбінат). На рис.1 чітко видно, що декілька спостережень знаходяться досить далеко від основного масиву даних, тому можна припустити, що ці спостереження є викидами, які варто видалити. Для цього було розраховано інтервал. Після аналізу часового ряду доходностей виявилось, що необхідно відкинути 12 спостережень, як такі, що не входять в інтервал. Після очистки даних знову було побудовано гістограму (рис.2). Візуально можна побачити, що діаграма після очистки даних набула більш характерного для нормального розподілу вигляду. При аналізі отриманих результатів (табл.1) використовувались теоретичні властивості нормального розподілу, а саме те, що мода і медіана мають співпадати, скіс дорівнює нулю, а ексцес дорівнює трьом.

Отже, у таблиці 1 можна побачити, що доходності акцій семи емітентів до

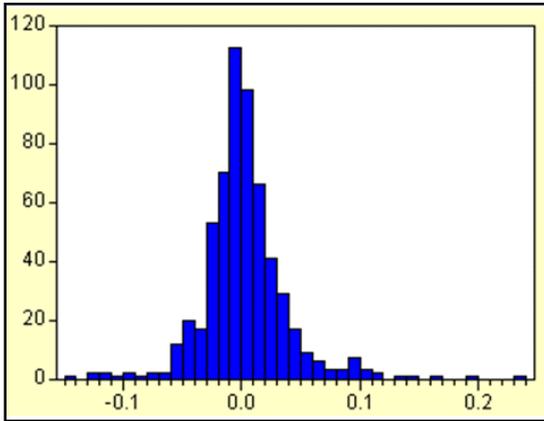


Рис.1 Гістограма доходностей акцій Алчевського металургійного комбінату до очистки даних

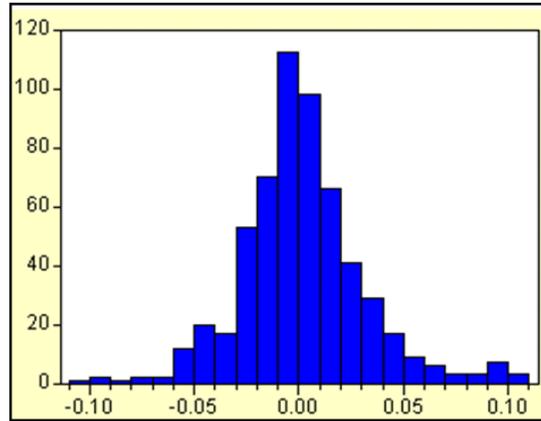


Рис.2 Гістограма доходностей акцій Алчевського металургійного комбінату після очистки даних

очистки даних мали невеликий додатний скіс (окрім Західенерго, у якого спостерігався від'ємний скіс), ексцес відхилявся від 3, медіана і мода (там де її вдалося розрахувати) приблизно співпадали. Тобто доходності даних акцій схильні змінюватися більш різко, ніж це було б при нормальному розподілі. Після очистки даних від надто екстремальних значень, скоси та ексцеси усіх акцій значно наблизилась до відповідних характеристик нормаль-

ного розподілу. Тобто якби вже після очистки ми застосовували до даних акцій метод моделювання оптимального портфеля, що заснований на параметрах нормального розподілу, то імовірно, що ми б отримали менш суттєві відхилення розрахованих результатів від фактичних, ніж при використанні даних до очистки.

В результаті очистки даних з генеральних сукупностей було видалено від 1,71% до 2,39% спостережень (табл. 2).

Таблиця 1

Основні статистики доходностей акцій до та після очистки генеральних сукупностей

Емітент	Тікер		Математичне сподівання	Медіана	Мода	Стандартне відхилення	Скіс	Ексцес
Алчевський металургійний комбінат	ALMK	до очистки	0,0021	-0,0004	-	0,0360	0,9564	9,4377
		після очистки	0,0013	-0,0005	-	0,0294	0,4080	4,7497
Комбінат «Азовсталь»	AZST	до очистки	0,0023	0,0011	0,0364	0,0321	0,5582	7,4126
		після очистки	0,0020	0,0011	0,0364	0,0279	0,2223	4,1092
Райффайзен Банк Аваль	BAVL	до очистки	0,0001	-0,0001	0,0043	0,0053	0,5351	7,9126
		після очистки	-0,0001	-0,0001	0,0043	0,0043	-0,1977	4,0077
Центренерго	CEEN	до очистки	0,0021	0,0002	-	0,0283	0,7964	9,9792
		після очистки	0,0015	0,0001	-	0,0228	0,1542	4,4401
Єнакієвський металургійний завод	ENMZ	до очистки	0,0015	0,0005	-	0,0329	1,3304	17,5753
		після очистки	0,0015	0,0005	-	0,0257	0,4315	4,2839
Укрнафта	UNAF	до очистки	0,0041	0,0013	0,0000	0,0302	0,5593	8,0813
		після очистки	0,0035	0,0013	0,0000	0,0246	0,0283	4,3744
Укрсоцбанк	USCB	до очистки	0,0028	0,0002	-	0,0367	1,6506	12,1878
		після очистки	0,0003	-0,0001	-	0,0267	0,0899	5,0833
Західенерго	ZAEN	до очистки	0,0004	-0,0004	0,0000	0,0238	-0,2199	9,0518
		після очистки	0,0001	-0,0006	0,0000	0,0190	0,1204	4,8545

Результати очистки генеральних сукупностей

Емітент	Тікер	Кількість видалених спостережень	Частка видалених спостережень	Обсяг генеральної сукупності після очистки
Алчевський металургійний комбінат	ALMK	12	2,05%	97,95%
Комбінат «Азовсталь»	AZST	10	1,71%	98,29%
Райффайзен Банк Аваль	BAVL	12	2,05%	97,95%
Центренерго	CEEN	12	2,05%	97,95%
Єнакієвський металургійний завод	ENMZ	10	1,71%	98,29%
Укрнафта	UNAF	14	2,39%	97,61%
Укрсоцбанк	USCB	14	2,39%	97,61%
Західенерго	ZAEN	14	2,39%	97,61%

При аналізі кожного спостереження, які було відкинуто з генеральних сукупностей виявилось, що у деяких емітентів дати екстремальних значень співпадають. Наприклад, 07.05.2009 чотири з восьми досліджуваних емітентів мали занадто екстремальні для них значення (всі вони в цей день зросли). Так само наприклад, 25.05.2010 - сім з восьми емітентів мали занадто екстремальні для них значення (всі мали від'ємні доходності). Очевидно, що екстремальні значення, дати яких співпали одночасно у декількох емітентів - виникли внаслідок певних об'єктивних економічних чи політичних причин (занадто погані чи навпаки дуже хороші новини у нашій країні чи у світі). І саме вони спричинили екстремальне падіння чи зростання відповідних акцій. Наприклад, Азовсталь, Райффазен Банк Аваль, Укрсоцбанк мали лише по одній даті, яка не перетнулася з жодним іншим емітентом. Отже, поведінка цих акцій повністю відображала поведінку усього ринку.

В той самий час, у деяких емітентів таких як Укрнафта і Західенерго сім з чотирнадцяти і відповідно шість з чотирнадцяти викидів за своїми датами виникнення не перетнулись з жодною з семи інших компаній. Тобто відповідні викиди відбулися нібито без причини. Даний факт наштовхує на думку, що ці акції стали об'єктами маніпуляцій великих гравців фондового ринку, що домовились різко підвищити або

навпаки зменшити ціни на відповідні акції.

Звичайно, можна було б видаляти занадто екстремальні значення у доходностях акцій, але для цього їх необхідно групувати за більш менш схожими датами викидів. Адже, щоб використати їх в одному портфелі - усі спостереження необхідно звести до відповідності, тобто спостереження мають відповідати одним і тим самим датам. В такому випадку в одному портфелі можна буде використовувати лише невелику кількість емітентів. Так само активи які ми підозрюємо на спекулятивність також не варто комбінувати з усіма іншими.

Визначення необхідного обсягу вибірки. У попередній статті Пріми Дарини [2] постало питання яким має бути обсяг вибірки усереднених даних, щоб отримати адекватні результати застосування класичного інструментарію моделювання оптимального портфеля цінних паперів.

Всі розрахунки перевірки наявності автокореляції було оновлено для останніх даних. Діапазон вибірки склав 728 спостережень для кожного емітента (від 01.04.2009 до 01.03.2012). Результати перевірки початкових та усереднених даних на наявність автокореляції на 20 лагах наведено в табл. 3.

З таблиці 3 видно, що у початкових не усереднених доходностях акцій восьми емітентів спостерігається автокореляція високих порядків. Далі, було усереднено ціни акцій за 2, 3,

Результати перевірки наявності автокореляції

Емітент	Початкові дані	Порядки автокореляції для усереднених даних										
		За 2 дні	За 3 дні	За 4 дні	За 5 днів	За 7 днів	За 8 днів	За 9 днів	За 10 днів	За 15 днів	За 20 днів	
ALMK	1, 4, 6, 8, 10	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2	1
AZST	1, 2, 4, 5, 9, 10, 19	1, 2, 6	1, 2	1	1	1, 2	1, 2	1, 2	1	1, 8	1	1
BAVL	1, 2, 4, 5, 6, 8, 10, 17, 19	1, 2, 13	1, 2, 9	1, 2, 7	1, 2, 6	1, 2, 4	1, 4	1	1, 3	1	1	1
CEEN	1, 4, 5, 7, 8, 9, 10, 16, 19, 20	1, 6	1, 2	1	1	1	1	1	1, 2	1, 19	1	1
ENMZ	1, 4, 6, 8, 9, 10, 12, 14, 15, 16	1, 2	1, 2	1, 2	1	1, 2	1	1, 2	1	1	1	1
UNAF	1, 2, 4, 5, 6, 9, 10, 15	1, 2, 6	1, 2, 4	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2	1, 2
USCB	1, 2, 4, 6, 9, 10, 12, 13	1, 2, 13	1, 2, 9	1, 2, 7	1	1, 2	1	1	1	1	1	1
ZAEN	1, 2, 4, 5, 7, 9, 10, 11, 12, 13, 14, 17, 18	1, 2, 7	1, 2, 10	1, 2	1, 2	1	1, 18	1	1	1	1	1

4, 5, 6, 7, 8, 9, 10, 15 і 20 днів і на їх основі знову розраховано доходності. Вже для доходностей усереднених за 2 дні порядки автокореляції значно скоротились і продовжували поступово скорочуватись для кожного наступного рівня усереднення. Але навіть для доходностей за 20 торгових днів (тобто за місяць) автокореляція повністю не зникла.

Початкові ряди доходностей, а також ряди усереднених доходностей було перевірено на відповідність нормальному розподілу за допомогою тесту Жарка-Берра. Результати наведено в таблиці 4.

В таблиці 4 для кожного емітента наведено коефіцієнт асиметрії, коефіцієнт ексцесу, статистику Jarque-Bera і P-value. Статистика Jarque-Bera використовується для перевірки того, чи відповідає часовий ряд нормально-

му розподілу. Формула розрахунку статистики Jarque-Bera:

$$JB = \frac{n}{6} \left(s^2 + \frac{(k-3)^2}{4} \right),$$

де n - кількість спостережень, s - вибірковий коефіцієнт асиметрії, k - вибірковий коефіцієнт ексцесу.

Статистика Jarque-Bera відображає різницю коефіцієнтів асиметрії і ексцесу даного ряду від значень аналогічних показників, характерних для нормального розподілу. Коефіцієнти асиметрії та ексцесу для нормального розподілу дорівнюють відповідно 0 та 3.

P-value - це ймовірність, пов'язана зі статистикою Jarque-Bera. Вона показує чи приймається нульова гіпотеза про нормальний розподіл часового ряду. Якщо P-value менше за 0,05, то нульо-

**Результати перевірки рядів доходностей
на відповідність нормальному розподілу**

		ALMK	AZST	BAVL	CEEN	ENMZ	UNAF	USCB	ZAEN
Для даних без усереднення	Асиметрія	1,116	0,520	1,421	0,692	1,128	0,388	1,496	0,347
	Експес	15,186	14,022	11,466	9,032	12,782	7,657	11,965	13,335
	Jarque-Bera	4655,70	3717,89	2416,51	1162,15	3057,14	675,33	2709,77	3183,55
	P-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Для даних усереднених за 5 днів	Асиметрія	0,800	0,242	-0,423	0,222	0,244	0,825	-0,088	0,229
	Експес	3,372	2,620	1,968	2,137	2,874	2,308	1,973	2,139
	Jarque-Bera	16,32	2,288	10,767	5,699	1,541	19,355	6,562	5,627
	P-value	0,000	0,318	0,005	0,058	0,462	0,000	0,038	0,059
Для даних усереднених за 10 днів	Асиметрія	0,771	0,229	-0,445	0,203	0,236	0,816	-0,102	0,236
	Експес	3,293	2,625	2,006	2,111	2,889	2,279	1,981	2,121
	Jarque-Bera	7,391	1,053	5,348	2,863	0,707	9,554	3,242	2,947
	P-value	0,025	0,591	0,069	0,238	0,702	0,008	0,197	0,229
Для даних усереднених за 15 днів	Асиметрія	0,753	0,202	-0,447	0,190	0,215	0,811	-0,118	0,221
	Експес	3,281	2,571	2,005	2,087	2,859	2,266	1,949	2,160
	Jarque-Bera	4,697	0,695	3,583	1,953	0,411	6,350	2,317	1,763
	P-value	0,095	0,706	0,166	0,376	0,814	0,042	0,314	0,414
Для даних усереднених за 16 днів	Асиметрія	0,760	0,205	-0,450	0,192	0,221	0,812	-0,104	0,236
	Експес	3,277	2,594	2,025	2,111	2,881	2,263	1,971	2,107
	Jarque-Bera	4,477	0,621	3,301	1,756	0,393	5,966	2,066	1,871
	P-value	0,106	0,732	0,191	0,415	0,822	0,052	0,355	0,392

ва гіпотеза про нормальний розподіл даного часового ряду відхиляється при рівні значущості 5%.

За результатами таблиці 4 видно, що для початкових рядів доходностей гіпотеза про нормальність за тестом Jarque-Bera відхиляється при рівні значущості 5% для всіх восьми емітентів, оскільки всі P-value є меншими за 0,05. Вже для даних, усереднених за 5 днів, для таких емітентів як AZST (Азовсталь), CEEN (Центрэнерго), ENMZ (Снакіївський металургійний завод) і ZAEN (Західенерго) можна прийняти нульову гіпотезу про відповідність їх рядів доходностей нормальному розподілу. Далі, посту-

пово на кожному наступному кроці (за 10, 15 днів) кількість таких рядів збільшується. А при розрахунках для даних усереднених за 16 днів, для рядів доходностей всіх восьми емітентів можна прийняти гіпотезу про відповідність нормальному розподілу при рівні значущості 5%. При аналізі коефіцієнтів асиметрії та експесу для цих рядів ми бачимо, що вони справді є дуже близькими до нуля і трьох відповідно. Отже, розподіл доходностей цих рядів є дуже близьким до нормального.

За центральною граничною теоремою, якщо розподіл даних є невідомим, але не надто сильно відрізняється

ся від нормального розподілу, то обсяг вибірки необхідний для застосування інструментарію заснованого на характеристиках нормального розподілу має бути не менше 30 спостережень [4].

В нашому випадку у рамках тесту Jarque-Bera було прийнято гіпотезу про відповідність нормальному розподілу на рівні значущості 5% для рядів доходностей всіх восьми емітентів, побудованих на основі цін, усереднених за 16 торгових днів. Крім того для цих рядів було перевірено таку важливу характеристику нормального розподілу, яка полягає в тому, що 68,27% вибірки має знаходитись в інтервалі «математичне сподівання стандартне відхилення». Для всіх восьми емітентів цей показник виявився досить близьким до значення, характерного для нормального розподілу: ALMK - 71,11%, AZST - 65,22%, BAVL - 66,66%, CEEN - 66,66%, ENMZ - 68,88%, UNAF - 71,11%, USCB - 68,88%, ZAEN - 66,66%.

З усіх отриманих результатів можна зробити висновок, що навіть при наявності незначної автокореляції у рядах доходностей усереднених за 16 торгових днів для конкретних восьми емітентів, імовірно їх можна об'єднувати в один портфель і застосувати інструментарій моделювання оптимального портфеля цінних паперів, заснований на характеристиках нормального розподілу. Обсяг вибірки при цьому має складати не менше 30 спостережень (в нашому випадку ця умова задовольняється, обсяг вибірки складає 45 спостережень).

ВИСНОВКИ

Отримані результати свідчать про те, що після очистки рядів доходностей акцій на українському фондовому ринку від занадто екстремальних значень, їх розподіл стає досить близьким до нормального. Отже, при застосуванні до таких даних класичного інструментарію моделювання оптимального портфеля, що заснований на властивостях нормального розподілу, імовірно, що отримані результати не будуть занадто відрізнятися від фак-

тичних показників. В той самий час існують певні обмеження щодо формування портфеля з акцій після очистки їх рядів доходностей. Адже викиди в даних можуть спостерігатися у різні моменти часу, а у портфелі доходності всіх акцій мають співпадати за датами. Таким чином, найбільш доцільним буде об'єднувати у портфель акції, дати екстремальних значень яких майже співпадають.

Також за отриманими результатами можна зробити висновок, що для того, щоб інструментарій моделювання оптимального портфеля акцій, заснований на властивостях нормального розподілу, дав адекватні результати, при його застосуванні на українському фондовому ринку, необхідно здійснювати розрахунки на попередньо усереднених рядах доходностей, так, щоб їх розподіл був близьким до нормального. Якщо їх розподіл буде близьким до нормального, то за центральною граничною теоремою нам буде достатньо не менше 30 спостережень. Як показало проведене дослідження, у випадку восьми емітентів, обраних нами, усереднення цін акцій має складати не менше, ніж 16 торгових днів, щоб акції всіх восьми емітентів можна було об'єднати в один портфель і застосувати до нього класичний інструментарій моделювання оптимального портфеля акцій. Обсяг вибірки у нашому випадку склав 45 спостережень для кожного емітента, отже ймовірно, що помилка результатів моделювання у порівнянні з реальними фактами не буде занадто великою.

Необхідно, звичайно, зазначити, що наявність автокореляції, хоча і в основному 1-го, 2-го порядків у випадку суттєвого усереднення даних, буде ймовірно мати свій вплив на результати моделювання при застосуванні класичного інструментарію. Адже, окрім, того, що дані мають бути нормально розподіленими, вони мають бути ще і незалежними. Отже, у майбутніх дослідженнях все ж таки необхідно шукати способи повного позбавлення від автокореляції, а не лише наближення даних до нормального розподілу.

ЛІТЕРАТУРА

1. Долінський Л.Б. Особливості розрахунку котирувань цінних паперів в умовах українського фондового ринку // Економіка та підприємництво: Зб. наук. праць молодих учених та аспірантів. - К.: КНЕУ, 2000. - Вип. 4. - С. 273-279.
2. Пріма Д.О. Аналіз серійної кореляції у даних на українському фондовому ринку // Моделювання та інформаційні системи в економіці. Зб. наук. праць. Вип. 83. - К.: КНЕУ, 2011. - С. 201 - 210.
3. Dasu T., Johnson T. Exploratory data mining and data cleaning. - New Jersey: Hoboken.: John Wiley & Sons, 2003. - с. 203
4. Гухман И.И., Скороход А.В., Ядренко М.И. Теория вероятностей и математическая статистика. - К.: Вища шк. Гол. Вид-во, 1988. - с. 654
5. Brase C.H., Brase C.P. Understandable statistics: concepts and methods, tenth ed. - Massachusetts: Boston.: Brooks/Cole, Cengage Learning, 2012. - с. 844