

Необходимост от използване и познавателен смисъл на моментите от по-висок ред (коефициенти на асиметрия и ексцес)

доц. д-р Тодор Калоянов

*УНСС, катедра „Статистика
и иконометрия”*

тел: 0889804713

e-mail: todornik@abv.bg

Резюме: Основната част от статистическите методи за изследване на факторни влияния се основава на сравняване на условни разпределения. Обичайна практика е анализите да се ограничават само до измерване и тълкуване на различията между стойностите на средната аритметична величина и средното квадратично отклонение, което се дължи на допускането, че разпределенията са нормални. Средната аритметична величина и стандартното отклонение са напълно достатъчни за дефиниране на нормалното разпределение. Много често в практиката се срещат статистически съвкупности, чиито единици нямат нормално разпределение. В тези случаи е необходима допълнителна информация за формата на разпределението, която средната аритметична величина и стандартното отклонение не могат да дадат.

В статията се обосновава необходимостта от използване на моменти от по-висок

ред – трети и четвърти. Дискутира се познавателният смисъл на параметрите на статистическото разпределение, конструирани на тяхна основа – коефициентите на асиметрия и ексцес, както и на тези, които се основават на позиционните средни величини.

Ключови думи: моменти, разпределения, асиметрия, ексцес.

JEL: C10, C16.

Увод

Основната част от статистическите методи за изследване на факторни влияния се основават на сравняване на условни разпределения. Това сравнение и оценката на различията може да се осъществи по различен начин – чрез тестове за съответствие с дадено теоретично разпределение, чрез сравняване на значимостта на съответните разлики между разпределения, когато данните са получени от представителни изважки. По-различна е ситуацията, когато е необходимо да се изследва динамиката на дадено явление. В този случай зада-

чата се свежда до сравняване на безусловни разпределения.

Внимание заслужават възможностите, които предлага анализът на динамиката на стойностите на параметрите на статистическото разпределение – средната аритметична величина, средното квадратично отклонение, коефициентите на асиметрия и ексцес. Значението на този подход се определя и от необходимостта да се извършват подобни сравнения, когато не са приложими методите на теорията на статистическите изводи и заключения, т.е. данните са получени от изчерпателни изследвания.

Обичайна практика е анализите да се ограничават само до установяване и тълкуване на средната аритметична величина и средното квадратично отклонение, което се дължи на допускането, че разпределенията са нормални.

Средната аритметична величина и стандартното отклонение са напълно достатъчни за дефиниране на нормалното разпределение. Но много често в практиката се срещат статистически съвкупности, чиито единици нямат нормално разпределение. В тези случаи е необходима допълнителна информация за формата на разпределението, която средната аритметична величина и стандартното отклонение не могат да дадат. Поради характера на своята конструкция тези параметри не могат да покажат посоката и силата на онези систематични и несистематични въздействия, които търпят само част от единиците. Действително измерителите на статистическото разсейване, в частност стандартното отклонение, се влияят и от въздействия върху част от единиците, но те не могат да отделият и измерят техния ефект. Тази задача се решава от коефициента на асиметрия. За този параметър обикновено само се казва,

че измерва степента на отклонение на емпиричното разпределение от съответното нормално по отношение на симетрията [Венецкий, Венецкая, 1979; Гамев, 1986, с. 210]. Това обяснение може да се смята за недостатъчно, в смисъл че остава открит въпросът за част от въздействията, които единиците са изпитвали в процеса на тяхното възникване и формиране на значенията им по даден признак. В тази насока е важно да се отбележи, че с помощта на коефициента на асиметрия се установява:

- наличието и силата на въздействията, на които са подложени само част от единиците на изследваната съвкупност.
- сумарната посока на тези въздействия, в резултат от които се получава лява (отрицателна) или дясна (положителна) асиметрия.

Коефициентът на ексцес е друг важен параметър на статистическото разпределение. Като правило в статистическата литература и за него само се посочва, че измерва отклонението на емпиричното разпределение от съответното нормално разпределение по отношение на неговата върхова изтегленост [Гамев, 1986; Mansfield, 1987]. В действителност различните стойности на този параметър също са признак за различни въздействия върху единиците на дадената съвкупност, но в друг смисъл [Стефанов, Томев, 1960]. Ако разпределението е симетрично, високата стойност на коефициента на ексцес може да означава отрицателна корелация между отделните реализации на индивидуалните въздействащи фактори. В този случай единиците се концентрират около центъра на разпределението в по-голяма степен, отколкото при нормалното разпределение, докато „полите“ на емпиричното разпределение излизат извън пределите на нормалното разпределение. В случай на положителни серийни корелации

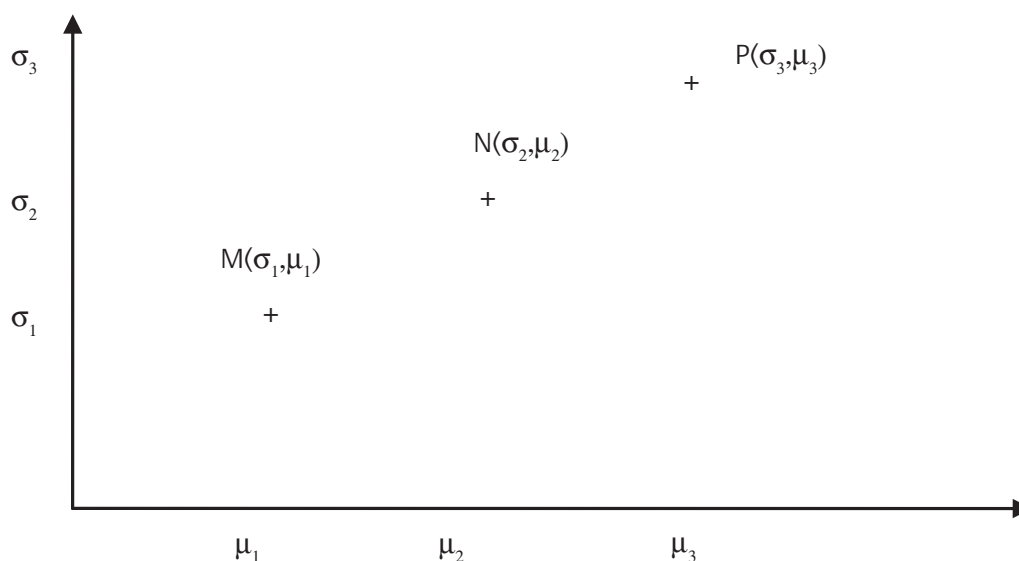
симетричният ексцес се оказва (при равни други условия) тъп, отрицателен. Чрез коефициента на ексцес, отново благодарение на неговата конструкция, се дава отговор на въпроса, дали има „свърхконцентрация“ на единиците в симетричното разпределение, или обратно – „свърхразсредоточаване“ под влияние на определени фактори.

Коефициентите на асиметрия и ексцес могат да дадат твърде ценна информация за въздействията, на които са подложени единиците на статистическата съвкупност и да подпомогнат вземането на решение [Калоянов, 1998, с. 59-66]. Неслучайно още преди десетки години тези два коефициента се използват за контрол на производствения процес [Бородачев, 1946, 1950]. Коефициентът на асиметрия (известен и като нормиран трети централен момент) се прилага и като един от начините за сравнение на различни семейства криви за разкриване на разликата между тях и за използването им като основа на емпиричния анализ [Цонев, 1971; Кокс, Оукс, 1988]. Двата измерителя са в основата на теста Jarque-Berra за проверка на съвпадението с нормалното разпределение.

При анализа на емпиричните разпределения трябва да се има предвид, че допускането, че голяма част от разпределенията са нормални или се доближават до нормалното, е довело до използване при изследванията главно на средната аритметична величина и на измерителите на разсейване. Това допускане се е отразило при създаването на различните статистически методи за изследване на връзки и зависимости. Моментите от по-висок порядък – трети и четвърти, не се използват в тези методи.

Необходимост от използване и познавателен смисъл на моментите от по-висок ред (коефициенти на асиметрия и ексцес)

Финансите, по-конкретно изборът на портфолио е една от областите, в която статистиката и допускането за нормално разпределение намират голямо приложение. През 1952 г. излизат две публикации, които откриват нова епоха на съвременното портфолио. Първата работа е на Roy, A. D. „Safety



Фигура 1. Варианти на портфолио с различна възвръщаемост и риск

first and the holding of assets", а втората е на Harry M. Markowitz и е със заглавие „Избор на портфолио“. Според финансистите работата на Markowitz прави революция в областта на съвременната инвестиционна теория и практика. Същността на идеята на автора се свежда до това, при избор на инвестициите да се имат предвид едновременно възвръщаемостта и вариацията на възвръщаемостта на портфолиото като цяло – фигура 1.

Това означава при дадена възвръщаемост да се минимизира вариацията или да се осигури максимална възвръщаемост при определено равнище на риска (вариацията). В случая се приема, че дисперсията не е постоянна, както е в редица статистически методи, а че тя е променлива и се търси оптимално съотношение между стойностите на средната и дисперсията. Във финансите възвръщаемостта се измерва с математическото очакване (средна аритметична величина), а вариацията – със средноквадратичното отклонение или дисперсията. Математическото очакване (очакваната норма на възвръщаемост) се изчислява по формулата:

$$E(r) = \sum_{s=1}^n \text{Pr}(s)r_s, \quad (1)$$

където:

$s=1, 2, \dots, n$ са възможните изходи в портфейла;

r_s – нормата на възвръщаемост за изхода s ;

$\text{Pr}(s)$ – вероятността за събъждане на изхода s .

Дисперсията, измерваща риска, се изчислява по следната формула, в която символите са познати:

$$\sigma^2 = \sum_{s=1}^n \text{Pr}(s)[r_s - E(r)]^2, \quad (2)$$

Финансистите са възприели да работят с понятията първи, втори, трети и четвърти момент, които са им по-удобни поради по-честото използване на вероятностите

като мезла, вместо честотите. С помощта на вариацията, независимо дали се има предвид средноквадратичното отклонение (стандартното) или вариацията (дисперсията), те измерват риска на съответната инвестиция. Вариацията се изчислява като разлика между очакваната и реализираната възвръщаемост. Очевидно колкото по-малка е разликата между очакваната и реализираната, действителна възвръщаемост, толкова по-малък е рискът, и обратно.

На практика оценката и сравненията между различни варианти на инвестиционен портфейл се осъществяват на основата на неговите две характеристики – очакваната стойност и стандартното отклонение на възвръщаемостите на портфейлите. На основата на тези две характеристики е изграден и критерият „средна стойност – стандартно отклонение“ (очакване – дисперсия, mean – variance или M-V). На с. 131 Боди З., Кейн А. и Маркърс А. Дж. (2000) формулират критерия по следния начин: „А е по-добър от В, ако

$$E(r_A) \geq E(r_B) \text{ и } \sigma_A \leq \sigma_B$$

и поне едно от равенствата е строго (т.е. няма равенство)“.

Критерият е конструиран на основата само на две характеристики на вероятностното разпределение – математическото очакване и стандартното отклонение. И ако се приеме, че математическото очакване (средната аритметична величина) е достатъчно ясно само по себе си, стандартното отклонение има доста богато съдържание. В случая то се приема като измерител на риска и според авторите на „Инвестиции“ идеята е да се опишат вероятностите и големините на „изненадите“ (отклоненията от средната стойност) с минимално необходимия за постигане на нужната точност набор от ста-

мистику“ [2000; с. 140]. Боди З., Кейн А. и Маркър А. Дж. обръщат внимание, че дисперсията не осигурява пълно описание на риска.

Основният въпрос, който вълнува не само теоретичните, но и практичните финансиста, е кога и при какви условия анализът „средна – вариация“ е приложим. Както пише Семюелсън в самото начало на статията си „Основна теорема на портфейлния анализ за апроксимацията със средни стойности, дисперсии и моменти от по-висок ред“ (1970), „James Tobin (1958, 1965), Harry Markowitz (1952, 1959) и много други автори направиха много ценни приноси за решаване на оптималния риск с ударение върху анализа средни и вариации. Тези автори разбраха, че резултатите могат да бъдат само приближение, но също разбраха, че приближението и изчислимите резултати са нещо по-добро, отколкото нищо... Но аз мисля, че е важно да се подчертае отново един аспект на модела „средна – вариация“, който изглежда не е получил достатъчно внимание в съвременната дискусия, и това е полезността на „средната“ и „вариацията“ в ситуации, съдържащи все по-малко и по-малко риск – това, което аз наричам „плътни“ вероятности. Настоящата работа формулира и доказва две основни теореми. Тя осигурява защита на анализа „средна – вариация“ и според мен е най-тежката защита, която е давана. (В икономикса съответните вероятностни разпределения не са близки до Гаусовото и квадратичната полезност в повечето случаи. Води до добре известни абсурди). Но тъй като аз усъвършенствам метода „средна – вариация“ и показвам неговите точни ограничения – заедно с тези за всеки модел с r -моменти – работата също може да се отнесе и като критика на този подход“.

Резултатите в работата на Семюелсън са теоретичната основа за използването на анализа „средна – вариация“.

В своите действия инвеститорите допускат, че условията за приложение на анализа „средна – дисперсия“ са изпълнени и пренебрегват по-високите моменти. Подобно на редица други области се допуска, че разпределението, с което се работи, в случая на възвръщаемостите на активите, е нормално. Удобрствата, които създава това допускане, са известни. Но в същото време са известни и проблемите, които произтичат от приемането на това допускане, когато то не е вярно. Както отбелязват Боди З., Кейн А. и Маркър А. Дж. [2000; с. 143] „Към допускането, че възвръщаемостите от отделните видове акции са нормално разпределени, остават и теоретични възражения. При положение че цените на акциите не могат да бъдат отрицателни, нормалното разпределение не може да бъде напълно представително за поведението на нормата на възвръщаемост за периода на задържане, защото допуска всякакви резултати, включително отрицателни цени“.

Основните критики към метода „средна – дисперсия“ според Byrne, P. и Lee, S. (1997) се свеждат до:

- Полезността като основа на подхода притежава сериозни ограничения, които го правят неспособен да описва поведението на големи групи от инвеститори. Използването на подхода за избор на ефективно портфолио предполага в почти всички случаи квадратична функция на полезност, като се твърди, че тя апроксимира достатъчно добре много други функции. Но тази функция поставя съществени ограничения за практическо приложение на подхода „средна – дисперсия“.
- Допускането за нормално разпределение на възвръщаемостите не е валидно за много ценни книжа, включително и недвижимо имущество. Когато разпределението на възвръщаемостите не е нормално, в общия

случай не е възможно да се намери точно оптималното решение на проблемите с разпределението (формирането на портфолио).

Все повече изследователи търсят начини за преодоляване на ограниченията, които налага допускането за нормалност на разпределенията. Като пряка последица от това допускане е и използването само на средната и дисперсията, в общия случай първи и втори начален, централен и смесен момент. Удобрствата, които създава нормалното разпределение, а именно, че то се дефинира еднозначно само с два параметъра, се реализират при решаването само на част от множеството практически задачи.

Един от възможните начини за по-вярното представяне на действителността и намиране на по-точни решения е използването на моменти от по-висок от втори ред. Традиционно тези моменти са били прилагани във физическите и биологичните науки. Според Donohо (2000) „чрез възглеждането в трети, четвърти и даже по-високи моменти може да бъде спечелено ново философско проникновение в широк кръг от дисциплини, от анализа на гените до случайното движение на финансовите данни. В днешно време включването на моменти от трети и по-висок ред става задължително“. Harvey и Siddique (2000) стигат до извода, че „пазарът взема под внимание наличието на асиметрия при оценка на активите и инвеститорите искат компенсация за поддържането на активи с отрицателна асиметрия“. До аналогичен извод достигат при своите изследвания и Ang и Bekaert (2001), според които „пазарът оценява степеня на асиметрична зависимост, създавайки асиметрично портфолио“. Tsaiing (1972) твърди, че „включването на по-високи моменти е желателно, в ред и условия подобни на метода „средна – дисперсия“; степеня на положителна асиметрия трябва да бъде използвана при определяне на предпочитанията. Използ-

ването на моменти от по-висок ред води до обогатяване на множеството от функции на полезност, които се прилагат при формиране на портфолио“.

Сред многото изследвания, посветени на този въпрос, внимание заслужава „*Skewness and the Bubble*“ на Conrad, Dittmar и Ghysels (2007). Авторите изследват възможността по-високите моменти на разпределението на възвръщаемостта да се окажат важни при обясняване на възвръщаемостта от ценни книжа. На основата на данни за периода от 1965 до 2005 г. те оценяват индивидуалните вариации на ценните книжа, тяхната асиметрия и ексцес. Установява се, че има значима отрицателна зависимост между асиметрията и възвръщаемостта – книжата с положителна или с по-малка отрицателна асиметрия имат по-малка възвръщаемост през следващите месеци. Достига се до извода, че има положителна зависимост между ексцеса и възвръщаемостта. Оказва се, че в зависимост от сектора на производство, за който са закупени акциите, чувствителността към асиметрията е различна. Така например за секторите на производство на хардуер, софтуер и полупроводникова индустрия чувствителността към положителната асиметрия е по-голяма. Интересен е изводът, подкрепен от практически резултати, че вариацията и асиметрията намаляват при нарастване на ексцеса. Според тях това отново акцентира върху необходимостта оценката на връзката възвръщаемост и моменти от по-висок ред да става едновременно. Техните резултати показват също, че акциите с високо разсейване и голяма асиметрия имат по-малка последваща възвръщаемост, докато по-високият ексцес е свързан с по-голяма последваща възвръщаемост.

Необходимостта от използване на моменти от по-висок ред е осъзната отдавна. Причините за това имат и теоретичен, и чисто

практически характер. Общото и за двата вида причини е необходимостта да се работи с разпределения, различни от нормалното, независимо от въздействията, обусловили съответното разпределение.

Карл Пирсън (1895) обръща внимание на гамаразпределението като модел за асиметрия¹. Интерес към асиметричните разпределения проявява и Pareto (1897), тъй като част от разпределенията в икономиката са асиметрични. Както посочват Groeneveld, R. A. и Meeden, G. [1984, с. 391], независимо от многото области, в които се срещат и работят с асиметрични разпределения, „идеята и измерването на асиметрията остават неточни“.

Основният въпрос, който се поставя почти във всяка работа, е какво измерват измерителите с различна конструкция и как да се тълкуват получените резултати. Очевидно е, че между отделните измерители на асиметрията ще има различия, но важното при тълкуването на получените резултати е да се държи сметка за особеностите на използваните в конкретния случай.

За разлика от показателя за асиметрия, разбирането за ексцеса претърпява сериозна еволюция. Развитието се осъществява в три основни направления. Първото е свързано със същността на измерителя и неговия познавателен смисъл – какво означава наличието на ексцес, какво всъщност се измерва с помощта на коефициента. Второто направление е свързано с разработването на различни измерители на ексцес и техните качества. Третото направление е свързано с първите две и се свежда до използването на ексцеса в теорията и практиката.

През периода от 1906 до 1910 г. в статии в списание „*Biometrika*“ са осъществявани сравнения между честотни разпределения за много големи съвкупности от различни области и нормалното разпределение с помощта на ексцеса и асиметрията.

Стюгент в своята работа „*Errors of Routine Analysis*“ (1927) на с. 160 дава следното описание на ексцеса, показано на фигура 2.

* In case any of my readers may be unfamiliar with the term “kurtosis” we may define mesokurtic as “having β_2 equal to 3,” while platykurtic curves have $\beta_2 < 3$ and leptokurtic > 3 . The important property which follows from this is that is that platykurtic curves have shorter “tails” than the



normal curve of error and leptokurtic longer “tails.” I myself bear in mind the meaning of the words by the above *memoria technica*, where the first figure represents platypus, and the second kangaroos, noted for “lepping,” though, perhaps, with equal reason they should be hares!

Фигура 2.

¹ Хронологията е по Seier, E., Celebrating 100 years of Kurtosis 1905-2005.

Среща се и допускането, че ако за една и съща стойност на аргумента $x=0$ функция на плътност е по-висока, то и съответното разпределение ще има по-висок ексцес. Въвежда се и понятието „density crossing” като достатъчно условие за едно разпределение да има по-висок ексцес β_2 от друго разпределение [Dyson, 1943; Funican, 1963]. Допуска се, че ако две функции на плътността на вероятностните разпределения с еднаква вариация се пресичат два пъти от всяка страна спрямо нулата (центъра на разпределението), едното ще има по-голям ексцес от другото.

Най-често се приема, че ексцесът е мярка на това, дали разпределението е по-остро или по-тъпо (сплескано) спрямо съответното нормално разпределение. Разпределението с висок ексцес проявява тенденция да има връх близо до средната аритметична величина, намалява бързо и има тежки опашки. По-високият ексцес означава, че по-голяма част от вариацията се дължи на рядко срещаните екстремални отклонения в противовес на умерените отклонения с голямо честота. Разпределението с ниска стойност на ексцес проявява тенденция да има по-скоро плосък връх близо до средната аритметична величина, отколкото остър връх.

Друго определение за ексцеса е, че той представлява степен на заостреност на разпределението, определена като нормирана форма на четвърти централен момент на разпределението. Срещат се няколко начина на неговото представяне. Така например той се отбелязва като β_2 (Abramovitz and Stegun 1972, с. 928) или α_4 (Kenney and Keeping, 1951, с. 27; Kenney and Keeping, 1961, с. 99-102),

$$\text{където } \beta_2 = \frac{\mu_4}{\mu_2^2}, \quad (3)$$

като с μ_i се отбелязва i -тият централен момент.

Същите автори представят и следната разновидност на формулата:

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3, \quad (4)$$

която се използва по-често, тъй като чрез нея се измерва ексцесът спрямо нормалното разпределение. Чрез изваждането на числото 3 се получава стойност нула за разпределение с нормален ексцес.

Van Zwet W. R. (1964) Въвежда за клас симетрични разпределения подреждане $\leq_s$ дефинирано чрез $F \leq_s G$, ако $R_{F,G}(x) = G^{-1}(F(x))$ е изпъкнала за $x > m_F$, където m_F е точката на симетрия на F . От допускането за симетрия на разпределенията следва, че $R_{F,G}(x)$ е изпъкнала за $x > m_F$, ако тя е вдлъбната за $x < m_F$. $F \leq_x G$ е в сила, ако случайната променлива X с разпределение F може да бъде отнесена към случайната променлива Y с разпределение G чрез нарастваща вдлъбнато-изпъкнала функция относно медианата. Van Zwet определя ексцеса като подредба на симетрични разпределения и според него ексцесът не може да се представя с един измерител. Предлага се метод за подреждане на две разпределения съгласно асиметрията. Идеята на автора получава развитие и в литературата са известни подрежданията на Loh, Oja, Lawrence. Вдлъбнато-изпъкналите функции намират приложение при разработване на методите за портфолио анализи.

Chissom B. S. (1970) на основата на експериментирание с добавяне и намаляване на случаи показва как според него трябва да се тълкува коректно коефициентът на ексцес (има се предвид измерителят, основаващ се на четвърти централен момент). За целта той използва три разпределения – приблизително нормално, правоъгълно и бимодално.

В заключението се извежда на преден план фактът, че: „важно е да се помни, че ексце-

съм зависи от върха на разпределението и неговите опашки. Ударението обаче пада на опашките на разпределението при определяне на четвъртия момент" [1970, с. 22].

През 1970 г. се поставя началото на дискусия дали измерителите на ексцеса откриват наличие на бимодалност. В основата на тази дискусия е работата на Richard Darlington „Is Kurtosis Really „Peakedness?““ публикувана в *The American Statistician*. Според автора терминът „заостреност“ (peakedness) е грешен и за описване на ексцеса по-добър е терминът „бимодалност на разпределението“ [1970, с. 19]. За да докаже това твърдение, Darlington използва нормираните отклонения z вместо оригиналните стойности на X . По този начин първоначалната формула на коефициента на ексцес

$$k = \frac{N^{-1} \sum (x - m)^4}{s^4} \quad (5)$$

се трансформира до

$$k = N^{-1} \sum z^4. \quad (6)$$

Според Darlington това означава, че k може да се тълкува като „мярка за степента, в която стойностите на z^2 се групират около тяхната средна със стойност 1; по-голямото групиране (сгрупиране) означава по-ниско k . Тъй като z е равно на $+1$ или -1 , когато $z^2 = 1$, k може да се тълкува като мярка на степента, в която z стойностите на разпределението се групират около $+1$ и -1 . Най-добре описващата характеристика на подобно групиране е „бимодалност““ [1970, с. 20]. Вниманието на тази идея е отделено в публикацията на David K. Hildebrand (1971). В нея той представя семейство двойно гамма-разпределения, които са бимодални и с коефициент на ексцес между -2 и $+3$. Оказва се обаче, че бимодалните разпределения могат да имат висок ексцес, когато модите не са близо до стойностите $z = \pm 1$. Според

J. J. A. Moors това е причината, поради която идеята на Darlington не е получила вниманието, което е заслужавала. Moors [1986, с. 283] изразява становището, че „ексцесът измерва дисперсията около две стойности $\mu \pm \sigma$, което е една обратна мярка на концентрацията в тези две точки. Висок ексцес може да възникне в две ситуации: а) концентрация на вероятностна маса близо до μ (съответстваща на изострено едномодално разпределение), и б) концентрация на вероятностната маса в опашките на разпределенията. Съществуването на тези две възможности обяснява объркването при тълкуването на ексцеса“.

Според David Rupert [1987, с. 1] „ексцесът често се разглежда като мярка на тежестта на опашките на разпределението по отношение на нормалното разпределение. Според други автори той измерва връхната изостреност близо до центъра на разпределението“. Според Rupert трактовката на ексцеса е твърде кратка и обикновено опитите да се направи по-разбираема са неуспешни. На същата страница той посочва, че „основният проблем, както Bickel и Lehmann (1975) отбелязват, е, че няма съгласие относно това, какво измерва ексцесът. Някои автори (напр. Kendall and Buckland 1971 и Levin 1984) твърдят, че ексцесът разграничава „връхната източеност“ от „сплеснатостта“. Darlington акцентира на това, че противоположността на „източен“ е бимодалност. Сплеснатото разпределение е между тези две крайности. Но термините „източеност“, „сплеснатост“ и „бимодалност“ подценяват зависимостта на ексцеса от поведението на опашките. Chissom обръща внимание, че „главното ударение трябва да се постави на опашките на разпределението при определянето на четвъртия момент“.

Johnson и Kotz приемат, че „ексцесът е мярка на отклонението от нормалността, за-

Висещо от относителната честота както на стойностите близо до средната, така и на тези, далече от нея спрямо тези, намиращи се на средно разстояние от средната” [1985, с. 22]. Предлагат се алтернативни измерители на ексцеса, като много от тях са разработени да измерват само източността или само тежестта на опашките, докато според Rupert ексцесът характеризира и двете.

С помощта на функцията на Влияние на Hampel (1968, 1974) Rupert доразвива дискусията на Darlington. Той прави сравнителен анализ между следните три измерителя на ексцес:

- Класически $k(F) = \mu_4(F) = \frac{\mu_4(F)}{\sigma^4(F)}$.
- Измерител на Hodg (1974), за който се предполага, че измерва тежестта на опашките на основата на „крайните значения” посредством средните на опашките:

$$\bar{U}_p(F) = \int_{q_{1-p}(F)}^{\infty} x dF(x), \quad (7)$$

средна, измерваща тежестта на горната опашка;

$$\bar{L}_p(F) = \int_{-\infty}^{q_p(F)} x dF(x), \quad (8)$$

средна, измерваща тежестта на долната опашка.

Както се посочва,

$$Q_p(F) = \frac{\bar{U}_p(F) - \bar{L}_p(F)}{\bar{U}_{0,5}(F) - \bar{L}_{0,5}(F)} \quad (9)$$

е инвариантен по отношение на местоположението и на скалата и измерва тежестта на опашките. В цитираната работа Hodg (1974) е използвал Вторите квантили на функцията F , при условие че $0 < p \leq 0,5$. За този измерител се установява, че не е устойчив, тъй като горните средни са чувстви-

телни на екстремалните стойности (*outliers*) макар и не толкова, колкото класическият измерител, основаващ се на четвъртия момент.

- Според Rupert, за да бъде устойчив, измерителят на ексцес трябва да е конструиран на основата на отношението на две устойчиви функции от типа

$$R_{\eta,p}(F) = \frac{\{q_{1-p}(F) - q_p(F)\}}{\{q_{1-\eta}(F) - q_{\eta}(F)\}} \text{ за } 0 < p < \eta < 0,5. \quad (10)$$

Чрез функцията на Влияние (*influence function* IF) на Hampel се дава количествено разбиране на ексцеса. С помощта на тази функция е проведен сравнителен анализ на горните три измерителя на ексцес. Прави се опит да се установи каква е относителната важност на тежестта на опашките и връхната източност при всеки един от тях.

От особена важност както за теорията, така и за практиката са следните изводи, до които Rupert достига с помощта на функцията на Влияние:

- Ексцесът измерва както връхната източност, така и тежестта на опашките. Няма чисти измерители само на връхната източност или само на тежестта на опашките.
- Местоположението на екстремалните стойности (*outliers*) е толкова важно, колкото и техните честоти.
- Смушенията в центъра имат далеч по-малко Влияние в сравнение с тези в краищата на опашките. Измерителят на ексцес, основаващ се на четвъртия момент k , е преди всичко мярка на поведението на опашките и в по-малка степен на връхната източност.
- И трите сравнявани измерителя имат общи свойства.

J. J. A. Moors (1988) предлага измерител на ексцеса, конструиран на основата на „octiles”, т.е. на квантили, разделящи реда

на осем равни части. Формулата на измерителя е следната:

$$\frac{[(E_7 - E_5) + (E_3 - E_1)]}{(E_6 - E_2)}, \quad (11)$$

където:

E_i ($i = 1, 2, 3, 5, 6, 7$) са съответно първи, втори, трети, пети, шести и седми квантили.

Moors посочва следните положителни качества на този измерител на ексцес:

- възможно е да се използва дори и когато не съществуват моменти;
- не се влияе от (екстремалните стойности) опашките на разпределението;
- лесно се изчислява (може да се определи дори и графично).

Посочените качества са резултат от използването на позиционни средни, които, както е известно, не се изчисляват от всички значения на признака. Това е причината, поради която те се предпочитат при наличие на силно отклоняващи се значения.

През същата година Balanda и MacGillivray [1988, с. 111] дефинират ексцеса като **„сво-бодно от местоположението и скалата движение на вероятностната маса от раменете на разпределението към неговия център и опашки и можем да признаем, че е възможно той да се формализира по много начини“**. Това е дефиницията, която се възприема като една от най-сполучливите и често цитирани, тъй като се вземат под внимание стойностите и техните честоти не само в центъра на разпределението, но в неговите опашки и рамене. Според Balanda и MacGillivray моментите изпълняват важна роля в статистическите заключения, но те са слаби показатели за формата на разпределението. Това е причината, поради която изследователите

търсят различни конструкции на измерителя на ексцес. Според Balanda и MacGillivray основната част от работите са посветени на измерването на ексцеса при симетрични разпределения и по-малко внимание се отделя на асиметричните разпределения, в т.ч. и на връзката между асиметрията и ексцеса. Очевиден е фактът, че този проблем заслужава внимание в бъдеще. Мнението на авторите е, че формализацията на ексцеса следва да се търси чрез частичното подреждане на разпределенията на основата на концепцията на Van Zwet (1964).

Две години по-късно, през 1990 г., същите автори разширяват критерия на Van Zwet за несиметрични разпределения. Те въвеждат кохерентна структура на подреждане и измерители, които не изискват наличие на симетрия. Това става на базата на спрег функцията.

През същата година Hosking (1990) дефинира L ексцес, изчисляван по формулата:

$$\tau = L_4/L_2, \quad (12)$$

където L са моменти, които според Hosking представляват обобщаващи числови характеристики на вероятностни разпределения и данни от извадки.

Теорията включва такива процедури, като порядковите статистики на средната разлика на Джини, и дава обещаващи иновации като измерители на асиметрията и ексцеса, както и нови методи за оценка на параметрите на няколко разпределения.

Groeneveld R. A. (1998) проверява доколко са чувствителни към формата на симетричното разпределение измерителите, предложени от Groeneveld и Meeden (1984). Те са конструирани на основата на квантилите, дефинирани по следния начин [1998, с. 325]:

„За непрекъсната случайна променлива X с функция на разпределението F и непрекъсната плътност f , симетрична по отношение на нулата, ние се концентрираме на разпределението на $|X|$, което има функция на плътност $F_*(x) = 2F(x) - 1$ “. Измерителите са дефинирани чрез израза, който представява квантилна мярка на асиметрия:

$$\gamma_2(p, F) = \frac{F_*^{-1}(1-p) + F_*^{-1}(p) - 2F_*^{-1}(0,5)}{F_*^{-1}(1-p) - F_*^{-1}(p)} =$$

$$= \frac{F^{-1}(1-(p/2)) + F^{-1}((1+p)/2) - 2F^{-1}(0,75)}{F^{-1}(1-(p/2)) - F^{-1}((1+p)/2)},$$

$$0 < p < 0,5. \quad (13)$$

Groeneveld R. A. показва, че по своята същност измерителят е отношение на разлики от две положителни разстояния към сумата от тези разстояния:

$$\gamma_2(p, F) = \frac{[F_*^{-1}(1-p) - F_*^{-1}(0,5)] - [F_*^{-1}(0,5) - F_*^{-1}(p)]}{[F_*^{-1}(1-p) - F_*^{-1}(0,5)] + [F_*^{-1}(0,5) - F_*^{-1}(p)]} =$$

$$= \frac{d_2(p) - d_1(p)}{d_2(p) + d_1(p)} \quad (14)$$

В изследването Groeneveld поставя за решаване няколко задачи:

- да установи доколко $\gamma_2(p, F)$ е подходящ измерител на ексцеса;
- чрез функцията на влияние да сравни как $\gamma_2(p, F)$ и β_2 отразяват формата на разпределение;
- да установи доколко теста Shapiro-Wilk е подходящ за измерване на различието от нормалното разпределение.

Отговорът на първата задача е, че за подходящи значения на p $\gamma_2(p, F)$ е полезен измерител на ексцеса.

Отговорът на втората задача е, че измерителят, основаващ се на моментите, дава по-голямо влияние на тежестта

на опашките и относително по-малко на масата в центъра на разпределението. „Функцията на влияние за β_2 в нормалния случай показва, че той се влияе силно от малките измествания на масата към опашките на разпределението. Този ефект чувствително намалява при квантилната мярка“ [1998, с. 329].

Симулацията, осъществена от автора, му е позволила да установи, че Shapiro-Wilk тестът е чувствителна мярка за отклонение от нормалното разпределение при малки стойности на p , $\gamma_2(p, F)$.

Резултатите са очаквани и логични, като се има предвид осреднителната процедура при измерителя, основаващ се на моментите. В числителя повдигането на четвърта степен дава по-голяма тежест на големите отклонения от центъра на разпределението и очевидно те компенсират по-големите честоти на стойности в центъра и близо до него.

През 2002 г. е дефиниран извадков g ексцес на основата на теста за нормалност на Geary (1936) за симетрични разпределения – $\hat{\tau}/\hat{\sigma}$.

$$g = \hat{w} = 13,29(\ln \hat{\sigma} - \ln \hat{\tau}), \quad (15)$$

$$\text{където } \hat{\tau} = (1/n) \sum_{i=1}^n |x_i - \bar{x}| \quad (16)$$

Seier и Bonett (2003) дефинират две семейства измерители като $E(g(z))$, където g е функция на стандартизираната променлива z ; а е стандартизирана променлива, избрана по такъв начин, че ексцесът за нормалното разпределение да бъде равен на 3.

$$g(z) = ab^{-|z|}, \quad 2 \leq b \leq 20 \quad (17)$$

$$\text{и } g(z) = a[1 - |z|^b], \quad 0,2 \leq b \leq 1 \quad (18)$$

За тези измерители се посочва, че придават по-голяма тежест на централния връх

В сравнение с тежестта на опашките, за разлика от измерителя, предложен от Карл Пирсън. Вследствие от това тестовете за нормалност, основаващи се на тези измерители, са по-чувствителни към върха на разпределението. Установено е също, че двата измерителя удовлетворяват подреждането на Van Zwet и са тясно свързани с L ексцеса и измерителя, основаващ се на квантилите.

Samuel Kotz и Edith Seier (2007) се връщат отново към квантилния измерител на ексцес, въведен от Groeneveld. Чрез заместване на разликата между трети и първи квантил в числителя $[Q_3(p) - Q_1(p)]$ с разликата между втори и първи квантил $[Q_2 - Q_1(p)]$ и с помощта на функцията на влияние, авторите изследват връзката между измерителите на ексцес, основаващи се на моментите, и тези на квантилите. Отново се дискутира въпросът за съотношението между центъра и опашките на разпределението. За целта са разгледани пет разпределения, характеризиращи се като едномодални и с тежки опашки.

Заклучение

За стогодишния период от по-общата дефиниция на Пирсън за разпределения с ексцес като относително повече или по-малко плосковърхи от нормалната крива се стига до схващането за ексцеса като „свободно от местоположението и скалата движение на вероятностната маса от раменете на разпределението към неговия център и опашки“. Между тези две разбирания се намират две други. Първото, което по-често се среща и днес, в това число и в литературата на български език, е, че ексцесът представлява върхната източеност на разпределението. В случая не се отдава дължимото на опашките на разпределението. Второто крайно разбиране е свързано с преувеличената роля на опашките и бимодалността на разпределението.

Главният въпрос, който се опитват да решат различните автори, занимаващи се с ексцеса, е кой измерител е най-подходящ. Част от авторите приемат измерителите, конструирани на базата на моментите (втори и четвърти централен) като достатъчно добри, независимо че според редица изследвания тези измерители придават по-голяма тежест на опашките, в т.ч. на екстремалните стойности, отколкото на тези в центъра и близо до центъра. Поради тази причина се предлагат измерители, основаващи се на квантилите в най-различни варианти. Според автора на настоящата работа те имат по-малка познавателна стойност поради това, че позиционните средни не отчитат стойностите на променливата. Те са функция само на честотите и използват по-малко количество информация. Безспорен е фактът, че те не се влияят от екстремалните стойности, но наличието на такива стойности невинаги може и следва да се пренебрегва. Измерителят на ексцеса е обобщаваща числова характеристика на върхната изтегленост на разпределението, която трябва да отчита стойностите на променливата в центъра и опашките на разпределението в единство с честотите. Такъв е измерителят, основаващ се на моментите. От особена важност е свързването на съответния измерител и неговия познавателен смисъл с поведението на единиците на съответната съвкупност и причините, които са обусловили съответното разпределение. Очевидно е, че поради различната логика, заложена в конструкциите на отделните измерители, те ще имат различни стойности за едно и също разпределение. Но по-важното е каква информация носи всеки от тях, как ще се тълкува и как може да бъде използван при аналитичната дейност.

Независимо от продължаващите теоретични търсения относно познавателния смисъл и най-подходящия измерител на ексцеса практиката налага неговото използване

при решаване на задачи с различен характер. Немалка част от изследователите използват наличието или липсата на ексцес, съвместно с асиметрията, като признак за отклонение или съвпадение с нормалното разпределение. Друга част изследователите използват ексцеса (обикновено заедно с асиметрията) за решаване на задачи от областта на медицината, инженерните науки, финансите и др.

Всяка статистическа характеристика има определен познавателен смисъл и не бива да се надценяват нейните възможности. Казаното важи с пълна сила и за коефициентите на асиметрия и ексцес. Те са свързани с формата на разпределение и имат смисъл главно тогава, когато разпределенията са едновърхови.

В практиката се срещат случаи, когато разпределенията са не само с положителен или отрицателен ексцес, но са и асиметрични. В литературата този въпрос почти не е разработен. Всяка ситуация изисква специално внимание за разкриване на причините, които са обусловили подобно разпределение.

Изследователите отделят внимание главно на конструкцията на измерителите на асиметрия и ексцес и на тяхното приложение. Недостатъчно се работи по въпроса за интерпретацията и познавателния смисъл на съответните измерители. В най-голяма степен това е направено в областта на финансите, биологията и някои технически науки, което обяснява наличието на част от предходното изложение. В икономиката, като правило, моментите от по-висок ред неоправдано се пренебрегват.

Литература

1. Боди З., А. Кейн и А. Дж. Маркър, Инвестиции, Натурела, С., 2000.
2. Борогачев, Н. А., Основные Вопросы теории точности производства, Москва, 1950.
3. Борогачев, Н. А., Обоснования методики расчета допусков и ошибок размеров и кинематической цепей, Москва, 1946.
4. Венецкий, И. Г., В. И. Венецкая, Основные математико-статистические понятия и формулы в экономическом анализе, Москва, 1979.
5. Гатев, К., Обща теория на статистиката, С., 1986.
6. Калоянов, Т., Относно познавателния смисъл на коефициентите на асиметрия и ексцес, сп. „Алтернативи“, 10/11, 1998.
7. Стефанов, И., А. Ю. Томев, Теория на статистиката, С., 1960.
8. Цонев, В., Основи на репрезентативното изучаване, С., 1971.
9. Byrne, P. и S. Lee, Real estate portfolio analysis under conditions of non-normality: The case of NCREIF, Journal of Real Estate Portfolio Management, 1997.
10. Conrad, J., R. Dittmar and E. Ghysels, Skewness and the Bubble, Preliminary Report, 2007.
11. Chissom, Brad S., Interpretation of the Kurtosis Statistics, The American Statistician, Vol. 34, No. 4, 1970, pp. 19-22.
12. Darlington, R. B., Is Kurtosis Really Peak-edness? The American Statistician, 24, 1970, pp. 19-22.
13. Dyson, F. J., A Note on Kurtosis, Journal of the Royal Statistical Society, 106, 4, 1943, pp. 360-361.

14. Edith Seier and D.G. Bonett, Two Families of Kurtosis Measures, *Metrika*, 58, 1, 2003, pp. 59-70.
15. Finucan, H. M., A Note on Kurtosis, *Journal of the Royal Statistical Society. Series B*, 26, 1, 1964, pp. 111-112.
16. Geary, R. C., Moments of the ratio of the mean deviation to the standard deviation for normal samples, *Biometrika*, 28, 1936, pp. 295-307.
17. Hampel, F. R., Contributions to the theory of robust estimation. Ph.D. dissertation, 1968.
18. Hildebrand, David K., Kurtosis Measures Bimodality?, *The American Statistician*, 25, No. 1, 1971, pp. 42-43.
19. Hogg, R. V., Adaptive robust procedures: A partial review and some suggestions for future applications and theory, *J. Am. Stat. Assoc.*, 69, No. 348, 1974, pp. 909-927.
20. Hosking, J. R. M., The theory of probability weighted moments, *Research Rep. RC 12210*, p. 160, IBM Research Division, Yorktown Heights, NY, 1986.
21. Groeneveld, R. A. and G. Meeden, Measuring Skewness and Kurtosis, *The Statistician*, 33, 1984, pp. 391-399.
22. Johnson, N. I. and S. Kotz, *Encyclopedia of Statistical Sciences*. Vol. 5, New York: John Wiley & Sons, Inc., 1985.
23. Kenney, J. F. and E. S. Keeping, Kurtosis, §7.12 in *Mathematics of Statistics*, Pt. 1, 3rd ed. Princeton, NJ: Van Nostrand, 1962, pp. 102-103.
24. MacGillivray, H. L., K. P. Balanda, The relationships between skewness and kurtosis, *Austral. J. Statist.* 30 (3), 1988, pp. 319-337.
25. Mansfield, E., *Statistics for Business and Economics. Methods and Applications*, Third ed. W. W. Norton & Company, Inc. New York, 1987.
26. Markovitz, H. M., Portfolio Selection, *The Journal of Finance*, 7, 1952.
27. Markovitz, H. M., *Portfolio Selection: Efficient Diversification of Investments*, New York, John Wiley & Sons, 1959.
28. Moors, J. J. A., A Quantile Alternative for Kurtosis, *The Statistician*, 37, 1, 1988, pp. 25-32.
29. Кокс, Д. И., А. Оукс, Анализ данных мина Времени жизни, Москва, 1988.
30. Pareto, V., *Cours d'Economie Politique*. Rouge and Cie, Lausanne and Paris, 1987.
31. Roy, A. D., Safety first and the holding of assets, *Econometrica*, 20, 1952, pp. 431-49.
32. Ruppert, D., What is Kurtosis? An Influence Function Approach, *Amer. Statist.* 41, 1987, pp. 1-5.
33. Student, Errors of Routine Analysis, *Biometrika*, 19, July 1927, pp. 151-164.
34. Tobin, J., Liquidity preference as behavior towards risk, *Review of Economic Studies*, February 1958, pp. 65-86.
35. Tobin, J., The Theory of Portfolio selection in F. H. Hahn and F. P. R. Brechling (eds.), *The Theory of Interest Rates*, London, Macmillan, 1965.
36. Tsiang, S. I., The Rationale of the Mean-Standard Deviation Analysis, Skewness Preferences, and the Demand for Money, *The American Economic Review*, 62, 1972, pp. 354-371.
37. Van Zwet, W. R., *Convex Transformations of Random Variables*, Mathematics Centre Tract 7, Mathematisch Centrum, Amsterdam, 1964. ~~VIA~~