

ПРЕЦИЗИРАНЕ НА РЕГРЕСИОННАТА ЗНАЧИМОСТ ЧРЕЗ ФИКТИВНИ РЕГРЕСОРИ, ПРОДУКТ НА УИНСОРИЗАЦИЯ НА УСЛОВНИ ПРОМЕНЛИВИ

Никола Илийчев Илиев

Стопанска академия „Д. А. Ценов“ – гр. Свищов
Катедра „Финанси и кредит“

Резюме: Регресионната връзка често е изправена пред редица проблеми, сред които липса на статистическа значимост, наличие на висока статистическа грешка и неувереност в резултатите. Преодоляването на въпросните става чрез усложняване на модела или направата на тежки допускания, всяко от които ограничава свободата на действие. Не такава е ситуацията при наличието на екстремни стойности, тъй като същите не позволяват усложняване на модела или направата на допускания. Те изискват използването на алтернативна методология, каквато е уинсоризацията. Разработена за нуждите на предварителния анализ на данните, уинсоризацията е приложима и при извеждане на регресионната връзка. Независимо от същото, уинсоризацията се характеризира с ключов недостатък – използването ѝ върху данни, за които се приема, че са безусловни, макар същите да се включват в регресионна връзка. Необходим се оказва модел, изразяващ регресионната връзка, отчитайки наличието на екстремуми и преодолявайки същите, в случая чрез фиктивни променливи.

Ключови думи: регион на доверителност, анализ на основните компоненти, айген-стойности, айген-вектори, уравнение на елипсоид.

JEL: C15, C24, C34, C58.

REGRESSION SIGNIFICANCE REFINING BY USING FICTIONAL REGRESSORS, THAT RESULT FROM WINSORIZATION OF CONDITIONAL VARIABLES

Nikola Iliychev Iliev

D. A. Tsenov Academy of Economics – Svishtov
The Department of Finance and Credit

Abstract: The regression relationship often faces a number of problems, including lack of statistical significance, presence of high statistical error and result uncertainty. To overcome this problems one can complicate the model, make new restrictive assumptions, each one restricting the degree of model application freedom. But the situation is different when there is a presence of extreme values, as they do not allow complication of the model or the making of assumptions. The said extreme values demand the use of an alternative methodology, like the winsORIZATION. Developed for the needs of the preliminary data analysis, the winsORIZATION is also applicable when using a regression relationship. Independently thereof the winsORIZATION is characterized with a major deficiency – its use on data, assumed to be unconditional, even if said data is included in a conditional regression relationship. There's a need for a model, describing

said regression relationship, that not only considers the presence of extreme values, but also overcomes them, which in this case is through the use of fictional variables.

Keywords: Confidence Region, Principal Component Analysis, Eigen-Values, Eigen-Vectors, Ellipsoid equation.

JEL: C15, C24, C34, C58.

Въведение

В основата на едно емпирично изследване стоят анализът и обработката на статистическа информация (Захариев, 2015). Същото изисква задълбоченото познание на свойствата и особеностите на въпросната информация, резултиращи от статистическата и иконометричната теория, въз основа на която същите се изграждат (Ангелова, 2016).

Традиционно, един от най-разпространените и най-използвани инструменти, които статистиката и иконометрията използват, е регресионният анализ. Въпросният представлява математически модел, позволяващ извеждането на уравнение на функция, описваща връзката между минимум две, максимум неограничен брой финансови променливи (Радков, 2015). В случая на две променливи връзката е между едната, която се определя като зависима, и втора, която се определя като влияеща. В случая на максимум възможни променливи, първата отново е в ролята на зависима, а всички останали – на влияещи променливи.

Въпросното уравнение на функция традиционно има вида:

$$(1) \quad Y = \alpha + \beta * X + \varepsilon,$$

където Y е зависимата финансова променлива; X – влияещата финансова променлива; α – регресионна интерцепта или алфа, описваща мащаба на връзката между зависимата и влияещата променлива; β – регресионен наклон или бета, описващ посоката и силата на връзката между зависимата и влияещата променлива; ε – регресионна грешка или епсилон, описващ частта от стойността на наблюденията на зависимата променлива, необяснени от стойността на наблюденията на влияещата променлива.

Ако връзката е на една зависима финансова променлива, спрямо множество влияещи финансови променливи (фактори), връзката се изменя в:

$$(2) \quad Y = \alpha + \sum_{k=1}^k \beta_k * X_k + \varepsilon,$$

където всички обозначения са като в предишното уравнение; β_k – регресионната бета на зависимата променлива спрямо влияещия фактор k ; X_k – влияещият фактор k .

Възможните влияещи фактори са най-разнообразни – от макроикономически данни, до финансови променливи, динамика на различни показатели и т.н. Една огромна група фактори са например т.нар. фиктивни променливи, популярни в една разновидност на регресионните модели, а именно моделите за панелни данни (Брусарски, 2015).

Същите модели изхождат от специфичните измерения на регресионната връзка, а именно регресионна връзка във времето – когато се изследва как влияещите фактори влияят върху зависимата променлива във времето или регресионна връзка в пространството – когато се изследва как влияещите фактори влияят върху зависимата променлива в пространството (в извадка). Обединяването на двете създава т.нар. панелен модел, при който връзката се извежда едновременно във времето и пространството.

Същото става именно чрез фиктивни променливи – показатели, чиито наблюдения приемат бинарни стойности – нула или единица, на база удовлетворение на даден критерий. Пример за такава фиктивна променлива във времето е например условието, дали стойността на дадено наблюдение е в точно определен период, например ако цялостната регресионна връзка е за периода 2010–2016, а фиктивната променлива е за годината 2013. В случая фиктивната променлива ще има стойност нула за всяко наблюдение, което не е за 2013 година, и единица, за всяко наблюдение, което е за 2013 година.

Друг пример (в пространството) за фиктивна променлива е например полъта, когато се изследва връзката между височината и теглото на извадка от хора. В случая, ако височината е на мъж, стойността на фиктивната променлива е нула, а в случай, че е на жена, стойността на фиктивната променлива е единица. Добавяйки въпросните фиктивни променливи в регресионния модел, същият се трансформира от времеви/пространствен в панелен и позволява квантифицирането на допълнителната връзка, която се създава на база дадена условност.

Научната разработка предлага използването на нов вид фиктивна променлива, а именно такава, която придобива стойност единица, когато стойността на дадено наблюдение е екстремна, и нула, когато стойността на дадено наблюдение не е екстремна.

Традиционно екстремните стойности се разполагат в двата края на вероятностно разпределение на дадена променлива и макар да описват действителността, от статистическа гледна точка създават т.нар. смущения под формата на бял шум. Въпросният бял шум, от гледна точка на регресионната връзка, видоизменя мащаба, силата и посоката на същата, следователно не позволява правилното ѝ установяване.

Решението на проблема традиционно е чрез два метода – трункация, целяща премахването на екстремните стойности, или уинсоризация, целяща коригирането на екстремните стойности. И двата преследват едно и също – чрез влияние върху белия шум да изчистят изкривяването на регресионната връзка към действителното.

Макар имащи своите преимущества, и двата метода имат и своите недостатъци – трукцията видоизменя извадката на вероятностното разпределение на наблюденията в дадена финансова променлива, докато уинсоризацията създава изкуствено т.нар. дебели опашки, чрез увеличаване на наблюденията с една и съща стойност, разполагаща се в граничната част на вероятностното разпределение, вследствие на самата уинсоризация.

Вниманието на научната разработка пада върху недостатъка на уинсоризацията, търсейки метод за неговото решаване. Същото изхожда от една проста грешка, която традиционно практикуващите уинсоризацията пропускат, а именно каузалната връзка между финансовите променливи. Въпросната грешка изхожда от извършването на уинсоризацията независимо, а именно – първоначално за едната променлива, впоследствие за следващата и т.н., докато уинсоризирани са всички променливи.

Въпросното е грешка, тъй като, от една страна, допуска независимост на променливите, при положение че за същите се допуска зависимост поради включването им в регресионен модел. От друга страна, проблемът може да се опише графично, чрез представянето на стойностите на двете променливи преди и след независимата им уинсоризация. Преди уинсоризацията същите заемат класическо разпределение (не в смисъла на статистическото значение на думата разпределение, а в смисъла на разполагане на самите стойности), което отчита в себе си регресионната връзка, белия шум от остатъчната ѝ грешка и корелационната връзка между променливите.

След уинсоризацията същите заемат странно разпределение (т.е. разполагане), което се описва като „квадрат“, вследствие на ограничаването на екстремни стойности в двата края на разпределението на едната и другата променлива. Въпросният квадрат показва много ясно проблема на независимата уинсоризация на финансови променливи, а именно – макар да се преследва изчистване на белия шум, същият се изчиства само отчасти, но се увеличава чрез несправедливото третиране на различни стойности в разполагането на двете променливи – един тип третиране на стойностите в близост до регресионната линия и действителната връзка, и друг тип третиране на стойностите в частта на нейната ортогонална проекция. Визуализацията на същия проблем ще се даде малко по-късно в разработката.

Отгук същата (разработката) предлага модифициране на уинсоризацията за решаване на проблема чрез преследване на гранична връзка, при която „квадратът“ се избягва, т.е. уинсоризацията на различните променливи не е независима. Това става чрез описването на екстремните стойности чрез общо уравнение, отчитащо стойността на наблюденията във всяка променлива, като въпросното уравнение ще описва елипса при две променливи (една зависима и една влияеща) или елипсоид при повече от две променливи (една зависима и множество влияещи).

След използване на същото уравнение е възможно извеждането на нови екстремни стойности, които независимата уинсоризация пропуска, както и изключването на други стойности, които независимата уинсоризация приема за екстремни. В резултат на това е възможно създаването на фиктивна променлива, по описаната по-горе методология, чието включване в регресионния модел да позволи от една страна премахване на белия шум, в резултат на екстремуми, а от друга – прецизиране на изискването за това, дадена стойност да бъде екстремум в условна връзка, а не независимо.

Обект на изследването са случайни времеви редове, описващи динамиката на финансови променливи във времето и пространството, като същите задължително съдържат ясно разграничени екстремуми. **Предметът** на разработката са така описаните методологични особености, които разработката прави за нуждите на изпълняване на задачите и изпълняване целта на научния текст.

Накратко обобщена въз основа на останалите елементи на научната разработка е **тезата** – използването на фиктивни регресори, резултати от условни екстремуми ще доведе до прецизиране на регресионната значимост. Вследствие на описаната концепция научната разработка може да изведе следното: като **цел** на разработката може да се посочи необходимостта от апробиране на методологията такава, каквато ще се специфицира впоследствие, а именно – прецизиране на регресионната връзка, респективно значимост, на база използването на фиктивни променливи, в ролята на фиктивни регресори, конструирани на база използване на модифицирана уинсоризация.

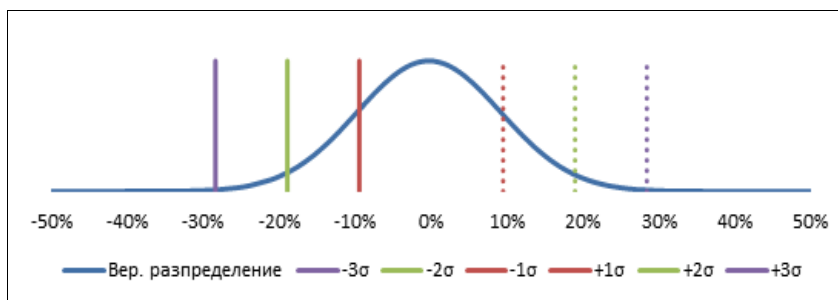
Задачите, които разработката поставя, се свързват с първоначалното специфициране методологията за извеждане на условни екстремуми; тяхната последваща модифицирана уинсоризация, на база допускането за условна връзка между отделните променливи; конструирането на фиктивни регресори, които да отчетат същите условни екстремуми; включването им в регресионен модел; и не на последно място сравнителен анализ между резултатите от класически регресионен модел и получения на база използването на въпросните фиктивни регресори.

Структурата на разработката е, както следва: в част 1 – Теоретични концепции – се преминава последователно през статистическо-иконометричната обосновка на необходимостта за прилагане на методологията; в част 2 – Методология – се описват моделите за класическа уинсоризация, модифицирана уинсоризация и конструиране на фиктивни регресори към регресионния модел; в част 3 – Емпирична част, научната разработка преминава през приложението на моделите и представяне на резултатите от въпросното приложение; в част 4 – Заключение – се правят изводи и обобщения относно получените резултати, прави се проверка, дали научната разработка съумява аргументирано да откаже или отхвърли тезата, както и се прави препратка към последващи разработки и изследвания, надграждащи върху направеното.

1. Теоретични концепции

Изхождайки от статистическата същност на финансовата променлива, същата е множество от наблюдения, всяко от които притежава индивидуална стойност. Същата не е ексклузивна, а може да се повтаря сред различните наблюдения. Традиционно в статистиката въпросното е не само възможно, а вероятно, тъй като стойностите, подлежащи на дадена зависимост, описваща даден процес, често следват определена динамика. Въпросната динамика се характеризира като вероятността, дадено наблюдение да притежава дадена стойност. Най-висока вероятност е налице, стойността на наблюдението да е например x , като вероятността намалява пропорционално с увеличаване или намаляване на стойността на наблюдението, респективно на $x + k$ и $x - k$. Още по-малка е вероятността, стойността на наблюдението да е $x + t$ и $x - t$, където $t < k$. Въпросната взаимовръзка описва всъщност т.нар. вероятностно разпределение на променливата.

Въпросното е фундаментална за статистиката концепция, въплъщаваща в себе си множество статистически особености, описващи същността, структурата и свойствата на данните в конкретната променлива. Такива са например средната, стандартното отклонение, асиметрията, ексцесът, наличието на дебели опашки и не на последно място – наличието на екстремуми. Вероятностното разпределение може да се представи чрез следващата фигура:



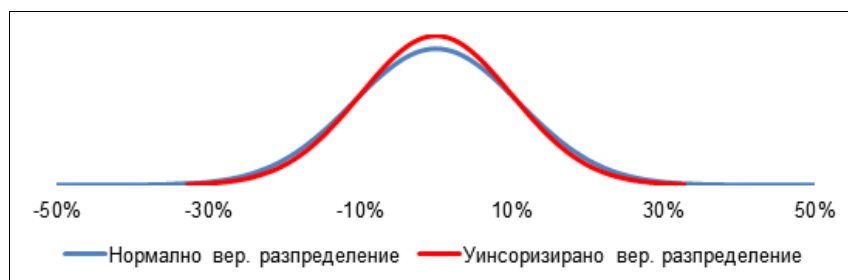
Фигура 1. Вероятностно разпределение на финансова променлива

Вниманието привлича формата на разпределението, традиционно определяна като камбанка. На следващо място стои средната, която макар необозначена на фигурата, се разполага точно в нейната среда, а именно – в най-високата част на кривата, описваща вероятностно разпределение. В случая същата кореспондира на стойността 0, а тълкуването ѝ е – средната на конкретната променлива е (разбира се) равна на нула. Впоследствие, около средната, последователно се разполагат т.нар. „брой стандартни отклонения“ от средната. Всяко от въпросните е гранична стойност на

интервал, в който, се разполагат определен брой наблюдения, без значение размера на извадката от стойности на разгледаната финансова променлива.

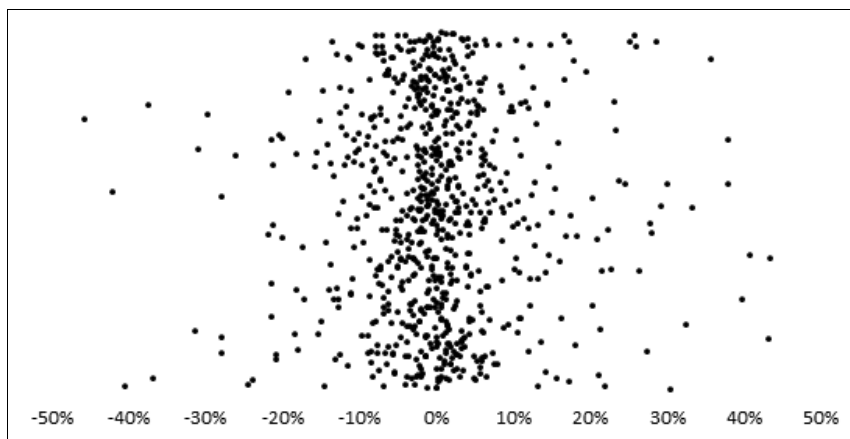
Диапазонът от едно стандартно отклонение от средната, в случая кореспондиращ на интервала -10% ; $+10\%$, описва 90% от стойностите в извадката, респективно две стандартни отклонения, в диапазона от -20% ; 20% описва 95% , а три стандартни отклонения, в диапазона от -30% ; 30% описва $99,9\%$ от стойностите в извадката. Въпросите са ключови за статистиката и по-конкретно разработката, тъй като позволяват точното определяне на това, кои наблюдения в извадката са екстремуми. А въпросите традиционно са тези, които са на разстояние, по-голямо от три стандартни отклонения от средната. Същите в действителност са едва $0,01\%$ от общия брой наблюдения в извадката, но макар изглеждащи малко, същите оказват достатъчно голямо влияние върху променливата, за да е възможно, същата да е замърсена от гореописания „бял шум“, дължащ се на статистическата неточност, ниска статистическа значимост и висока статистическа грешка.

Уинсоризацията има за цел промяната на стойността на всяко наблюдение извън диапазона на най-близката му кореспондираща гранична стойност – в случая плюс или минус три стандартни отклонения от средната. Описана по този начин, уинсоризацията звучи не само логично, но и правдоподобно. Ситуацията става по-сложна обаче при разполагане на вероятностното разпределение на уинсоризираната променлива спрямо това на неуинсоризираната. Същото е представено на следващата фигура:



Фигура 2. Вероятностни разпределения преди и след уинсоризация

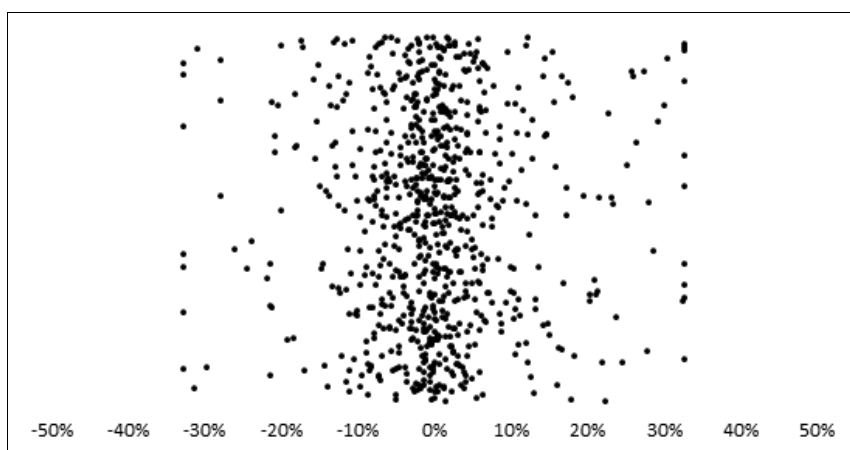
Тенденцията е, след уинсоризация вероятността да се увеличава за наблюденията, разполагащи се по-близо до средната, както и да намалява за наблюденията, разполагащи се около 2 стандартни отклонения от средната. Основната причина за това се дължи на промяната на цялостната структура на разпределението, именно в резултат на уинсоризацията. Проблемът е обаче по-скоро „козметичен“ и не оказва толкова голямо влияние колкото един друг проблем, който може да се представи на следващата фигура.



Фигура 3. Вероятностно разпределение, разположено спрямо случайна променлива

Идеята за представянето чрез въпросната фигура е алтернативното представяне на вероятностното разпределение. В случая подобно на „камбанката“ тук стойностите с най-голяма вероятност са около средната (равна на нула), а с отдалечаване от средната вероятността им намалява. Упоменавайки първата фигура, стойностите на разстояние над три стандартни отклонения от средната (в случая извън диапазона -30% , 30%) са екстремуми.

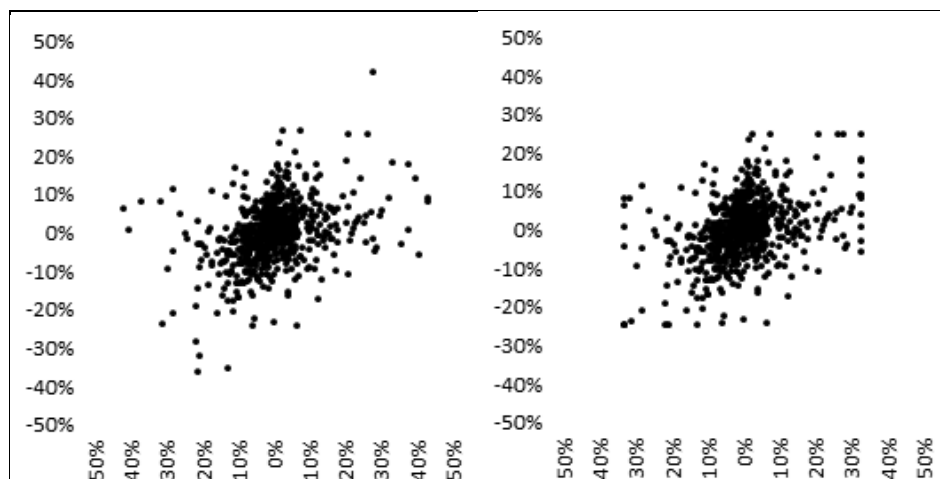
От така представената информация е трудно да се заключи какъв точно е недостатъкът на уинсоризацията, до момента, в който вероятностното разпределение след уинсоризация не се представи по същия начин, а именно:



Фигура 4. Вероятностно разпределение, разположено спрямо случайна променлива, след уинсоризация

Проблемът се вижда относително ясно – след уинсоризация екстремумите се разполагат по права линия, в резултат на факта, че граничните стойности са точки при класически представено вероятностно разпределение (камбанка), но се трансформират в прави, при разполагането на същото в $n + 1$ измерното пространство (в конкретната ситуация двуизмерното пространство).

Продължавайки примера в ситуацията, когато уинсоризацията се извършва едновременно за две, приемани за независими, финансови променливи, първоначалното разположение на вероятностните им разпределения в двуизмерното пространство преди и след уинсоризация търпи следната трансформация:



Фигура 5. Вероятностни разпределения на независими финансови променливи преди и след независима уинсоризация

Вляво на графиката ясно може да се види случайният характер на част от стойностите, описващи двете променливи – случаен характер, който липсва вдясно, заменен от хипотетична граница, описвана като квадрат (макар визуално да не изглежда като такъв, същият се описва от диапазон от три стандартни отклонения от средната и за двете променливи, при размер на същите, аналогичен за двете променливи, а именно $-30\%, 30\%$).

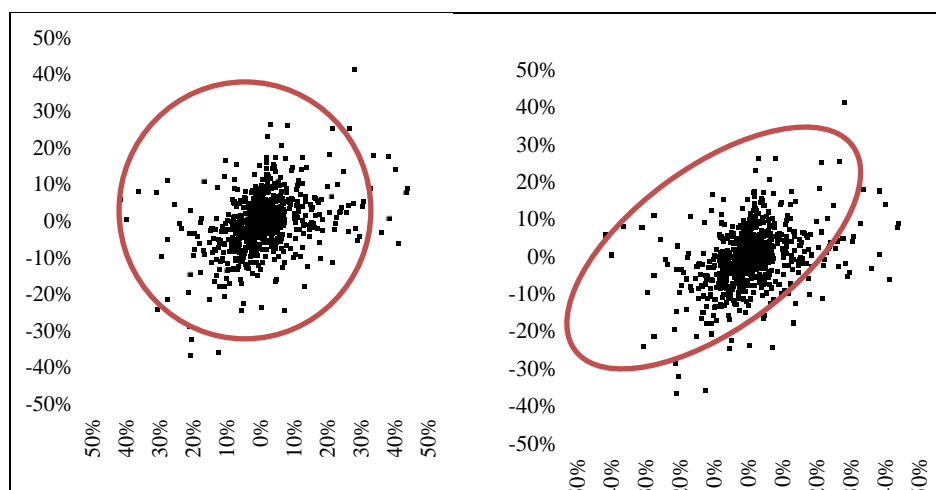
Проблемът е ясно забележим – налице е субективна оценка за финансовите променливи, резултат от независимото им изследване. Причината – стойностите на някои наблюдения, разположени в дадена част от графиката, се третираат различно от други, разположени в друга част на графиката. В конкретната ситуация това са наблюденията в горния ляв и долния десен ъгъл на „квadrата“ за сметка на наблюденията в долния ляв и горния десен ъгъл. Причината, това да са наблюденията в точно тези

части на квадрата, е проста – връзката между двете променливи най-ясно може да се изрази чрез регресионна линия, преминаваща през диагонала от долния ляв до горния десен ъгъл.

Оттук всяко наблюдение по диагонала се приема за присъщо на връзката, а всяко наблюдение, ортогонално на диагонала, се приема за неприсъщо на връзката. Оттук упоменатата по-рано логика – ако дадено наблюдение е присъщо на връзката, същото трябва да се оценява като екстремум в по-широки граници, докато ако дадено наблюдение не е присъщо на връзката, същото трябва да се оценява като екстремум по-строго. Същото ще „разтегли“ описаната като „квадрат“ гранична функция на уинсоризацията на две независими променливи в „ромб“, чийто голям диагонал ще лежи върху линията на регресионната връзка.

Макар същото да „модифицира“ отчасти уинсоризацията на двете променливи, единственото, което се случва в действителност, е промяната на теглото, с което връзката третира едната променлива за сметка на другата. Същинската модификация ще стане когато уинсоризацията третира двете променливи като действително условно свързани, но равнопоставено. Същото може да стане едва когато уинсоризацията на стойността на дадено наблюдение се прави на база не само въпросната променлива, но и на база стойността на същото наблюдение при другата променлива.

Казано по друг начин, от значение вече не е разстоянието на стойността на дадено наблюдение в променливата от средната (като брой стандартни отклонения), но и разстоянието на стойността на същото наблюдение в другата променлива от съответстващата му средна (отново като брой стандартни отклонения). Представянето на идеята при липса на каузална връзка има вида на фигурата вляво, докато представянето при наличие на каузална връзка има вида на фигурата вдясно.



Фигура 6. Теоретични граници на модифицирана уинсоризация при вероятностни разпределения на финансови променливи

Макар ситуацията на лявата част от фигурата да не подпомага цялостната идея на модифицирането на уинсоризацията, същата дава визуална представа за описания по-горе проблем, а именно – при наличие на независима уинсоризация, същата се извършва за всяка променлива поотделно, което води до създаване на гранична функция, чието разстояние от средната се променя с отдалечаването от абсцисата и ординатата, достигайки максимум при диагонала на координатната система. Същото е парадокс, тъй като уинсоризацията изисква разстояние от средната точно равно на три стандартни отклонения, докато в точката на диагонала същото достига до стойността от $3\sqrt{2} = 4,25$ стандартни отклонения от средната.

Решаването на парадокса става чрез допускането, че между двете променливи съществува каузалност, т.е. от значение е вече не независимото разстояние на всяко наблюдение на дадена стойност от респективната му средна, а „общото“ разстояние, където същото може да се определи като радиус от началото на координатната система (представляващо стойността, определяна от стойността на двете средни) до дадена стойност, определяна от стойността на наблюденията ѝ в двете променливи. Въпросното създава представената окръжност, за която се приема, че е новата „каузална“ уинсоризация, позволяваща определянето на „каузални“ екстремуми.

Въпросните не са просто наблюдения, чиято стойност е на по-голямо от три стандартни отклонения от средната разстояние. Същите са наблюдения, чието абсолютно разстояние от началото на координатната система е по-голямо от корен от сумата на квадратите на три стандартни отклонения от средните и на двете променливи. Връзката се извежда чрез уравнението за окръжност, резултат от Питагоровата теорема и тригонометричното изражение на връзката между катетите и хипотенузата на правоъгълен триъгълник.

Последната стъпка от цялостното модифициране на уинсоризацията изисква отчитането както на каузалната, така и на регресионната връзка чрез същата трансформация на идеята (от квадрат към ромб), но за описаната окръжност (към елипса). Въпросната трансформация ще позволи постигането на представената в дясната част на фигурата гранична функция, при която екстремум е вече стойност извън елипса, вместо окръжност, където елипсата е продукт на описаната по-горе математическа връзка – корен от сумата на квадратите на три стандартни отклонения от средните и на двете променливи, но този път коригирани с различните радиуси на елипсата, където същите са успореден и ортогонален на регресионната линия.

Изразена по последващия начин, отново връзката може да се изведе чрез уравнение, този път на елипса, резултат от чиста тригонометрия

на произволен, неправоеъгълен триъгълник и връзката между две от страните му и третата, най-дълга, страна.

Както може да се види на двете фигури, стойности, които едната функция приема за екстремуми, не са екстремуми според другата функция, и обратно. Въпросното е описаният по-горе момент на оценяване на екстремумите в по-широки и по-тесни граници на база разполагането им спрямо линията на регресионната връзка и ортогоналната ѝ проекция.

Разполагането с така специфицираната теоретична концепция зад „модифицирането“ на процеса по уинсоризиране на финансови променливи, пред научната разработка стои единствено специфицирането на методология за извършване на същото, както и продължаването ѝ за нуждите на изпълняването на целта и задачите на разработката чрез емпирично доказване на поставената теза.

2. Методология

Традиционно извършването на класическа уинсоризация става чрез изпълняването на следния логически оператор:

$$(3) \quad x_i = \begin{cases} [(-3 * \sigma_x) + \bar{x}], & x_i < [(-3 * \sigma_x) + \bar{x}] \\ x_i, & [(-3 * \sigma_x) + \bar{x}] < x_i < [(+3 * \sigma_x) + \bar{x}] \\ [(+3 * \sigma_x) + \bar{x}], & [(+3 * \sigma_x) + \bar{x}] < x_i \end{cases}$$

Обяснението на същия е следното – ако стойността на дадено наблюдение е по-ниска от три стандартни отклонения вляво от средната, то същото променя стойността си на три стандартни отклонения вляво от средната; обратно, ако стойността му е по-висока от три стандартни отклонения вдясно от средната, същото се променя на три стандартни отклонения вдясно от средната. Във всеки друг случай стойността е в диапазона три стандартни отклонения вляво от средната и три стандартни отклонения вдясно от средната.

Така специфицирана, уинсоризацията е класическа, т.е. не разглежда нито каузалната връзка на променливата x с всяка друга променлива, например y , нито регресионната връзка между двете.

За да е налице каузалната уинсоризация, специфицирана от описаната окръжна гранична функция, логическият оператор се видоизменя в:

$$(4) \quad A(x_i, y_i) = \begin{cases} A(x_i, y_i), & \sqrt{x_i^2 + y_i^2} < \sqrt{C_x + C_y} \\ A(x_i * C_{xy}, y_i * C_{xy}), & \sqrt{x_i^2 + y_i^2} > \sqrt{C_x + C_y} \end{cases}$$

където $C_x = [(3 * \sigma_x) + \bar{x}]^2$, $C_y = [(3 * \sigma_y) + \bar{y}]^2$ са координатите на функцията на окръжността; $C_{xy} = \frac{C_x + C_y}{x_i^2 + y_i^2}$ – коефициент на корекция на екстремумите към проекциите им върху функцията на окръжността.

За разлика от предишния логически оператор, третиращ уинсоризацията на финансовата променлива независимо от втората променлива, тук ситуацията е вече обратна – двете уинсоризации се извършват едновременно, като всъщност стойностите на наблюденията не се променят като такива, а като координати на точки, разположени в координатна система. Граничната функция се дефинира, както стана дума, като окръжност, резултираща от функцията на описаната чрез съкратените уравнения връзки. Обобщението на каузалната уинсоризация е следното – ако точката лежи в окръжността, координатите ѝ остават непроменени, но ако точката лежи извън окръжността, координатите ѝ се променят пропорционално, така че същата да се проектира върху самата окръжност.

Продължавайки логиката от каузална към модифицирана уинсоризация, логическият оператор отново се трансформира, придобивайки финалната си формулировка, която има вида:

$$(5) \quad A(x_i, y_i) = \begin{cases} A(x_i, y_i), & \sqrt{(x_i^2/a^2) + (y_i^2/b^2)} < \sqrt{E_x + E_y} \\ A(x_i * E_{xy}, y_i * E_{xy}), & \sqrt{(x_i^2/a^2) + (y_i^2/b^2)} > \sqrt{E_x + E_y} \end{cases}$$

където $E_x = \frac{[(3 * \sigma_x) + \bar{x}]^2}{a^2}$, $E_y = \frac{[(3 * \sigma_y) + \bar{y}]^2}{b^2}$ са координатите на функцията на елипсата; $E_{xy} = \frac{E_x + E_y}{(x_i^2/a^2) + (y_i^2/b^2)}$ – коефициент на корекция на екстремумите към проекциите им върху функцията на елипсата.

Отново, подобно на последната ситуация, екстремумите не се изследват независимо, като наблюдения, а зависимо – като координати на точки от координатната система. За разлика от каузалната уинсоризация при която входящите данни са достатъчни за цялостното специфициране на граничната функция и определянето на екстремумите, в случая на елипсата, необходими се оказват още две променливи. Същите са стойностите a и b , които участват в уравненията. Въпросните изразяват съотношението на двата радиуса, които характеризират елипсата.

Следващият въпрос, който стои пред успешната реализация на модифицираната уинсоризация, е успешното определяне каква е дължината на същите. Традиционно това може да се извърши на база изчисляването

на стандартните отклонения, които променливите имат, но при база, различна от базата X, Y . Необходима е база, при която абсцисата съвпада с регресионната линия, а ордината е нейна ортогонална проекция.

Същото е възможно чрез извършването на ротация на координатната система. За ротацията на координатната система необходим се оказва ъгъл, спрямо който същата да се осъществи. Въпросният може да се изведе на база наклона на регресионната линия (бета). В случай че стойността на бета коефициента е например единица, ъгълът е 45° . Бета коефициент, равен на две, съответства на ъгъл от 60° , и т.н. до теоретично достигане на ъгъл от 90° , когато бета коефициентът клони към $+\infty$. В обратна посока бета коефициент при ъгъл от 0° е налице, когато бета коефициентът е равен на 0.

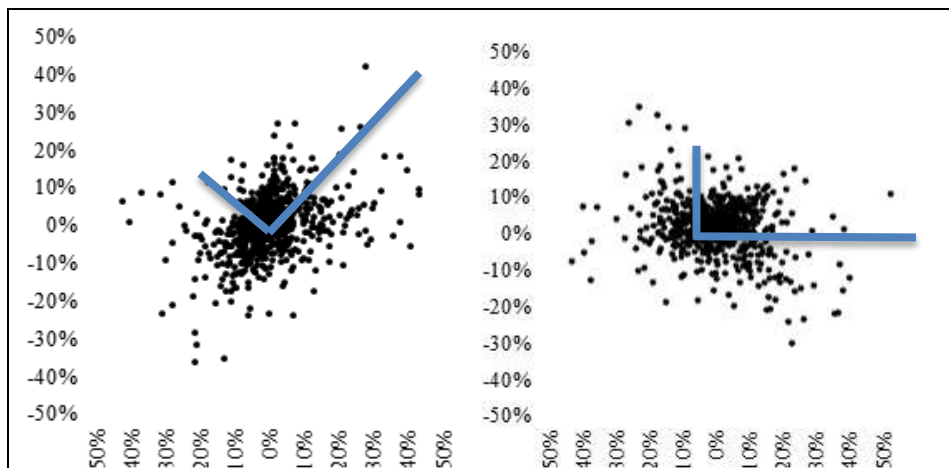
Проблемът при използването на бета коефициента за нуждите на ротацията е, че същият възплъщава в себе си всички недостатъци на регресионна връзка, замърсена от бял шум, резултиращ от екстремуми. За да се избегне това, е необходимо прибегването до използване на алтернативен ъгъл на ротация, независим от наличните екстремуми. Същият резултира от т.нар. анализ на основните компоненти, използващ теоретичната концепция за айген-стойности и айген-вектори.

Айген-стойностите и айген-векторите представляват специфични числа, респективно матрици, позволяващи еквивалентна трансформация на всеки масив от данни. Един масив от данни притежава точно определени айген-стойности и айген-вектори, броя на които съответства на броя променливи, които същия включва. В случая на две променливи броят на айген-стойности е две, респективно броят на айген-вектори – два. Айген-векторите на свой ред включват определен брой стойности, който в случая също кореспондира на броя променливи, а именно две, следователно обединени айген-векторите създават матрица с размер два реда и две колони. Същата е именно ротационната матрица, която може да се използва вместо бета коефициента за нуждите на ротацията.

Извършвайки въпросната ротация, данните се трансформират от такива, разположени в координатна система X, Y , в такива в координатна система V, W . В случая V е успоредна на теоретичната регресионна линия, която би съществувала, ако променливите не включват екстремуми. Същата е продукт, както стана дума, на анализа на основните компоненти, айген-стойностите и айген-векторите. Обратно, W е ортогонална на въпросната теоретична регресионна линия (перпендикулярна на V). Върху въпросните две се разполагат два компонентът, които анализът извежда – компонентът V и компонента W , всеки от които описва част от връзката между двете.

В случая компонентите са необходими именно за определяне на едно конкретно нещо – стойностите на a и b , където същите са 3 стандартни отклонения на компонентите V и W (аналогично на методологията за

уинсоризация) от средната, която макар наличието на ротацията е точката на самата ротация, следователно както преди, така и след нея, притежава координати 0,0.



Фигура 7. Ротация на финансовите променливи чрез използване матрицата на айген-векторите

В изпълнение на модифицираната уинсоризация, след калкулиране стойностите на a и b , е налице всичко необходимо за нейното извършване. След приключването му финансовите променливи са вече уинсоризирани, както от гледна точка каузалната, така и от гледна точка теоретичната регресионна връзка.

Описана дотук, методологията за модифицирана уинсоризация изхожда от две променливи – зависима и влияеща. Възможно е обаче цялостното ѝ трансформиране в многоизмерното пространство, в което връзката е между n на брой променливи – една зависима и множество $(n - 1)$ влияещи. В този случай единственото, което се трансформира, е математическото еволюиране на модела – от точка с координати стойностите на наблюденията на две финансови променливи, всяка точка вече е с координати стойностите на наблюдения на n на брой финансови променливи. Оттук граничната функция на елипса се видоизменя в гранична функция на елипсоид.

В обобщение на методологията, изхождайки от класическата уинсоризация, същата изхожда от т.нар. интервал на доверителност, който вероятностното разпределение на дадена финансова променлива има. В случая на модифицираната уинсоризация същата изхожда от т.нар. регион на доверителност, който съвместното вероятностно разпределение на множество финансови променливи имат.

Разполагайки с това определение, методологията може да продължи към създаването на т.нар. фиктивни променливи, позволяващи прецизиране на регресионния модел. Една такава променлива ще се създаде на база изпълняването на двата противоположни логически оператора:

$$(6) \quad D_{i_1} = \begin{cases} 1, & \sqrt{(x_i^2/a^2) + (y_i^2/b^2)} < \sqrt{E_x + E_y} \\ 0, & \sqrt{(x_i^2/a^2) + (y_i^2/b^2)} > \sqrt{E_x + E_y} \end{cases}$$

$$(7) \quad D_{i_2} = \begin{cases} 0, & \sqrt{(x_i^2/a^2) + (y_i^2/b^2)} < \sqrt{E_x + E_y} \\ 1, & \sqrt{(x_i^2/a^2) + (y_i^2/b^2)} > \sqrt{E_x + E_y} \end{cases}$$

Всеки от въпросните логически оператори води до конструирането на противоположни фиктивни променливи, първата от които изпълнява ролята на втора регресионна интерцепта, докато втората – ролята на измерител на влиянието на екстремните стойности върху финансовите променливи. Независимо от привидно липсата на необходимост от използването и на двете, същото е задължително, тъй като без отчитането на втората интерцепта е невъзможно отчитането на влиянието на екстремните стойности.

А въпросното отчитане става просто чрез включването им като фактори в многофакторен модел, с общ брой фактори $k + 2$, където k е броят фактори преди модифицираната уинсоризация, т.е., ако факторът е само един, с добавянето на двете фиктивни променливи моделът става трифакторен.

Тълкуването на резултатите от въпросната регресия е специфично. От една страна, анализаторът не се интересува до такава степен от стойностите на регресионните бета коефициенти на фиктивните променливи – стойността на първата ще е винаги нула, тъй като същата изпълнява ролята на втора интерцепта, докато стойността на втората реално измерва самия, по-горе споменат, бял шум, квантифициращ влиянието на екстремумите. От значение е по-скоро не стойността на въпросния бета коефициент, а статистическата му значимост (Пътев, 2014). Ако същата е налице, моделът съумява успешно да обхване влиянието на екстремумите върху

регресионната връзка, респективно изкривяването на същата, вследствие наличието им във финансовите променливи.

Пред анализатора стои и друг въпрос – как да използва информацията, която регресионното уравнение, отчитащо фиктивните променливи, му дава. Най-простото тълкуване е, информацията да се използва отчасти, т.е. например при изграждане на прогноза за бъдещото развитие динамиката на финансовите променливи, същото да не включва частта от модела, описваща влиянието на екстремумите. Причината за същото – връщането на белия шум в прогнозата, при възможност същият да бъде избегнат.

3. Емпирична част

За нуждата на емпиричното изследване се избират финансови променливи, притежаващи всички описани характеристики както от гледна точка предимства, така и от гледна точка недостатъци, които да възплават. В случая това са редове от доходности на публични акции за периода 2009–2015, калкулирани на седмична база. Избрани са осем акции, между всяка двойка от които е направена регресия, в единия случай, когато едната акция е зависима, а другата влияеща, и обратно, когато другата акция е зависима, а първата – влияеща. Регресиите последователно са направени без и с включването на фиктивната променлива, изведена на база модифицирана уинсоризация при гранична функция – регионът на доверителност, имащ формата на елипса.

Представянето на резултатите започва чрез таблици, съдържащи статистическата значимост на регресиите, представена чрез коефициента R^2 . За да е налице подобряване на регресионната връзка, резултиращо от прецизирането ѝ чрез включване на фиктивна променлива, е необходимо, стойностите на коефициента R^2 да са по-високи, ако в регресиите са включени фиктивни променливи, спрямо регресиите, в които същите не са включени. Въпросната калкулация се смята като натурален логаритъм от съотношението на R^2 при наличие спрямо R^2 при липса на фиктивна променлива. Резултатите са представени в Таблица 1.

Вниманието пада върху третата част от таблицата, а именно относителната промяна на коефициента на детерминация. Същата е винаги положителна, т.е. налице е повишаване на статистическата значимост във всеки един разгледан случай, в който се сравнява същата на регресия без и с отчитане на фиктивна променлива, отчитаща екстремуми.

Таблица 1.

Статистическа значимост на регресиите без и с включването на фиктивна променлива

| Коефициент на детерминация | Eurocash S.A. | Grupa LOTOS | KGHM Polska Miedz | LPP | Orange Polska | Polski Koncern Naftowy ORLEN C.A. | Polskie Górnictwo Naftowe i Gazownictwo S.A. | Synthos |
|-------------------------------------|---------------|-------------|-------------------|-------|---------------|-----------------------------------|--|---------|
| а. Регресия без фиктивна променлива | | | | | | | | |
| Eurocash S.A | - | 12,7% | 14,3% | 7,9% | 4,9% | 15,3% | 11,9% | 15,1% |
| Grupa LOTOS | 12,7% | - | 29,8% | 12,6% | 6,7% | 48,8% | 26,9% | 18,9% |
| KGHM Polska Miedz | 14,3% | 29,8% | - | 20,4% | 7,0% | 35,6% | 26,9% | 22,5% |
| LPP | 7,9% | 12,6% | 20,4% | - | 3,4% | 16,0% | 18,7% | 10,8% |
| Orange Polska | 4,9% | 6,7% | 7,0% | 3,4% | - | 11,3% | 8,0% | 3,8% |
| P.K. Naftowy ORLEN | 15,3% | 48,8% | 35,6% | 16,0% | 11,3% | - | 27,9% | 19,7% |
| P. G. N. i G. | 11,9% | 26,9% | 26,9% | 18,7% | 8,0% | 27,9% | - | 14,2% |
| Synthos | 15,1% | 18,9% | 22,5% | 10,8% | 3,8% | 19,7% | 14,2% | - |
| б. Регресия с фиктивна променлива | | | | | | | | |
| Eurocash S.A | - | 12,9% | 34,5% | 8,3% | 30,1% | 22,0% | 12,1% | 15,2% |
| Grupa LOTOS | 14,8% | - | 40,7% | 12,8% | 22,5% | 50,0% | 28,0% | 18,9% |
| KGHM Polska Miedz | 14,4% | 31,4% | - | 24,3% | 17,6% | 35,7% | 28,0% | 22,9% |
| LPP | 9,0% | 12,7% | 39,5% | - | 25,7% | 22,5% | 19,0% | 10,8% |
| Orange Polska | 5,0% | 7,6% | 20,0% | 8,2% | - | 13,3% | 8,6% | 5,9% |
| P.K. Naftowy ORLEN | 16,3% | 49,0% | 46,4% | 16,4% | 27,9% | - | 27,9% | 19,7% |
| P. G. N. i G. | 13,1% | 27,1% | 41,0% | 18,8% | 25,2% | 31,8% | - | 14,9% |
| Synthos | 16,1% | 18,9% | 32,6% | 10,8% | 20,2% | 23,1% | 14,4% | - |
| в. Относителна промяна | | | | | | | | |
| Eurocash S.A | - | 1,8% | 87,9% | 4,7% | 182,0% | 36,4% | 1,9% | 0,1% |
| Grupa LOTOS | 15,4% | - | 31,1% | 1,5% | 121,6% | 2,5% | 4,0% | 0,1% |
| KGHM Polska Miedz | 0,2% | 5,1% | - | 17,4% | 92,7% | 0,4% | 3,9% | 1,9% |
| LPP | 13,5% | 0,3% | 66,1% | - | 202,5% | 34,2% | 1,6% | 0,2% |
| Orange Polska | 1,5% | 12,3% | 105,5% | 87,8% | - | 15,8% | 7,5% | 43,9% |
| P.K. Naftowy ORLEN | 6,7% | 0,5% | 26,5% | 2,5% | 89,9% | - | 0,0% | 0,0% |
| P. G. N. i G. | 10,2% | 0,9% | 42,1% | 0,5% | 114,9% | 13,3% | - | 4,9% |
| Synthos | 5,9% | 0,1% | 37,1% | 0,5% | 167,3% | 16,1% | 1,4% | - |

На следващо място е необходимо разглеждането на общата статистическа грешка на регресията, представена отново в таблица:

Таблица 2.

Статистическа грешка на регресиите без и с включването на фиктивна променлива

| Статистическа грешка | Eurocash S.A. | Grupa LOTOS Spółka Akcyjna | KGHM Polska Miedz Spółka Akcyjna | LPP Spółka Akcyjna | Orange Polska S.A. | Polski Koncern Naftowy ORLEN C.A. | Górnictwo Naftowe i Gazownictwo | Synthos S.A. |
|--|---------------|----------------------------|----------------------------------|--------------------|--------------------|-----------------------------------|---------------------------------|--------------|
| а. Регресия без фиктивна променлива | | | | | | | | |
| Eurocash S.A | - | 4,9% | 5,1% | 4,3% | 5,1% | 4,6% | 3,7% | 5,3% |
| Grupa LOTOS | 5,3% | - | 4,6% | 4,2% | 5,0% | 3,6% | 3,3% | 5,1% |
| KGHM Polska Miedz | 5,2% | 4,4% | - | 4,0% | 5,0% | 4,0% | 3,3% | 5,0% |
| LPP | 5,4% | 4,9% | 4,9% | - | 5,1% | 4,6% | 3,5% | 5,4% |
| Orange Polska | 5,5% | 5,1% | 5,3% | 4,4% | - | 4,7% | 3,8% | 5,6% |
| P.K. Naftowy ORLEN | 5,2% | 3,7% | 4,4% | 4,1% | 4,9% | - | 3,3% | 5,1% |
| P. G. N. i G. | 5,3% | 4,5% | 4,7% | 4,0% | 5,0% | 4,2% | - | 5,3% |
| Synthos | 5,2% | 4,7% | 4,8% | 4,2% | 5,1% | 4,5% | 3,6% | - |
| б. Регресия с фиктивна променлива | | | | | | | | |
| Eurocash S.A | - | 4,9% | 4,4% | 4,3% | 4,3% | 4,4% | 3,7% | 5,3% |
| Grupa LOTOS | 5,2% | - | 4,2% | 4,2% | 4,6% | 3,5% | 3,3% | 5,2% |
| KGHM Polska Miedz | 5,2% | 4,3% | - | 3,9% | 4,7% | 4,0% | 3,3% | 5,0% |
| LPP | 5,4% | 4,9% | 4,3% | - | 4,5% | 4,4% | 3,5% | 5,4% |
| Orange Polska | 5,5% | 5,0% | 4,9% | 4,3% | - | 4,6% | 3,8% | 5,6% |
| P.K. Naftowy ORLEN | 5,2% | 3,7% | 4,0% | 4,1% | 4,4% | - | 3,3% | 5,1% |
| P. G. N. i G. | 5,3% | 4,5% | 4,2% | 4,1% | 4,5% | 4,1% | - | 5,3% |
| Synthos | 5,2% | 4,7% | 4,5% | 4,2% | 4,6% | 4,4% | 3,6% | - |
| в. Относителна промяна | | | | | | | | |
| Eurocash S.A | - | 0,0% | -13,3% | 0,0% | -15,2% | -4,0% | 0,0% | 0,2% |
| Grupa LOTOS | -1,1% | - | -8,2% | 0,1% | -9,1% | -1,0% | -0,6% | 0,2% |
| KGHM Polska Miedz | 0,1% | -1,0% | - | -2,3% | -5,9% | 0,1% | -0,6% | -0,1% |
| LPP | -0,5% | 0,1% | -13,6% | - | -13,0% | -3,9% | 0,0% | 0,2% |
| Orange Polska | 0,1% | -0,3% | -7,4% | -2,4% | - | -0,9% | -0,2% | -0,9% |
| P.K. Naftowy ORLEN | -0,5% | -0,1% | -9,0% | -0,1% | -10,2% | - | 0,2% | 0,2% |
| P. G. N. i G. | -0,6% | 0,0% | -10,6% | 0,1% | -10,2% | -2,7% | - | -0,2% |
| Synthos | -0,4% | 0,2% | -6,8% | 0,1% | -9,2% | -2,0% | 0,1% | - |

Статистическата грешка показва смесени резултати – налице са множество регресии, при които грешката намалява с включване на фиктивна променлива, но това не е налице навсякъде. Факт е обаче, че в ситуациите, когато грешката се увеличава, увеличението е с много по-малка стойност спрямо стойностите, с които същата намалява, във всеки друг случай. Казано по друг начин, докато намалението е средно $-4,30\%$, увеличението е средно $0,11\%$ (калкулирани съответно за всички отрицателни спрямо всички положителни стойности на относителната промяна на статистическата грешка).

На следващо място стои разглеждането на регресионната алфа:

Таблица 3.

Регресионна алфа на регресиите без и с включването на фиктивна променлива

| Регресионна интер- цепта | Eurocash S.A. | Grupa LOTOS Spółka Akcyjna | KGHM Polska Miedz Spółka Akcyjna | LPP Spolka Akcyjna | Orange Polska S.A. | Polski Koncern Naftowy ORLEN S.A. | Polskie Górnictwo Naftowe i Gazownictwo S.A. | Synthos S.A. |
|--|---------------|-------------------------------|-------------------------------------|--------------------|--------------------|--------------------------------------|--|--------------|
| а. Регресия без фиктивна променлива | | | | | | | | |
| Eurocash S.A | - | 0,05% | 0,08% | 0,59% | -0,25% | 0,13% | 0,02% | 0,51% |
| Grupa LOTOS | 0,32% | - | 0,12% | 0,62% | -0,22% | 0,15% | 0,05% | 0,58% |
| KGHM Polska Miedz | 0,31% | 0,06% | - | 0,60% | -0,23% | 0,14% | 0,03% | 0,56% |
| LPP | 0,15% | -0,10% | -0,15% | - | -0,31% | -0,04% | -0,14% | 0,38% |
| Orange Polska | 0,43% | 0,22% | 0,27% | 0,70% | - | 0,32% | 0,15% | 0,70% |
| P.K. Naftowy ORLEN | 0,27% | -0,02% | 0,05% | 0,58% | -0,26% | - | 0,01% | 0,53% |
| P. G. N. i G. | 0,33% | 0,10% | 0,14% | 0,62% | -0,21% | 0,19% | - | 0,60% |
| Synthos | 0,14% | -0,09% | -0,08% | 0,51% | -0,29% | 0,01% | -0,06% | - |
| б. Регресия с фиктивна променлива | | | | | | | | |
| Eurocash S.A | - | 0,02% | 0,43% | 0,57% | 0,04% | 0,28% | 0,04% | 0,52% |
| Grupa LOTOS | 0,42% | - | 0,40% | 0,60% | 0,08% | 0,22% | 0,11% | 0,57% |
| KGHM Polska Miedz | 0,29% | -0,06% | - | 0,44% | 0,10% | 0,18% | -0,04% | 0,48% |
| LPP | 0,20% | -0,09% | 0,20% | - | -0,11% | 0,12% | -0,11% | 0,37% |
| Orange Polska | 0,45% | 0,15% | 0,57% | 0,61% | - | 0,41% | 0,11% | 0,59% |
| P.K. Naftowy ORLEN | 0,35% | -0,05% | 0,33% | 0,54% | 0,06% | - | 0,01% | 0,54% |
| P. G. N. i G. | 0,40% | 0,06% | 0,47% | 0,60% | 0,07% | 0,32% | - | 0,68% |
| Synthos | 0,21% | -0,08% | 0,23% | 0,49% | 0,00% | 0,15% | -0,03% | - |
| в. Промяна в знака | | | | | | | | |
| Eurocash S.A | - | He | He | He | Да | He | He | He |
| Grupa LOTOS | He | - | He | He | Да | He | He | He |
| KGHM Polska Miedz | He | Да | - | He | Да | He | Да | He |
| LPP | He | He | Да | - | He | Да | He | He |
| Orange Polska | He | He | He | He | - | He | He | He |
| P.K. Naftowy ORLEN | He | He | He | He | Да | - | He | He |
| P. G. N. i G. | He | He | He | He | Да | He | - | He |
| Synthos | He | He | Да | He | He | He | He | - |
| г. Относителна промяна | | | | | | | | |
| Eurocash S.A | - | 0,0% | 0,4% | 0,0% | 0,3% | 0,2% | 0,0% | 0,0% |
| Grupa LOTOS | 0,1% | - | 0,3% | 0,0% | 0,3% | 0,1% | 0,1% | 0,0% |
| KGHM Polska Miedz | 0,0% | -0,1% | - | -0,2% | 0,3% | 0,0% | -0,1% | -0,1% |
| LPP | 0,0% | 0,0% | 0,4% | - | 0,2% | 0,2% | 0,0% | 0,0% |
| Orange Polska | 0,0% | -0,1% | 0,3% | -0,1% | - | 0,1% | 0,0% | -0,1% |
| P.K. Naftowy ORLEN | 0,1% | 0,0% | 0,3% | 0,0% | 0,3% | - | 0,0% | 0,0% |
| P. G. N. i G. | 0,1% | 0,0% | 0,3% | 0,0% | 0,3% | 0,1% | - | 0,1% |
| Synthos | 0,1% | 0,0% | 0,3% | 0,0% | 0,3% | 0,1% | 0,0% | - |

Вниманието привличат две специфични особености – не само относителната промяна на интерцептата, но и промяната на знака, тъй като налице се наблюдават две специфични тенденции. Когато например промяна на знака е налице, относителният размер на промяна на интерцептата е относително голям и в посока на увеличение. Средно увеличението е близо 0,21%, докато средно увеличението, при непромяна на знака, е едва 0,05%.

На следващо място стои регресионната бета, разгледана както от гледна точка изменение с добавянето на фиктивна променлива, така и от гледна точка въпросното спрямо бетата на самата фиктивна променлива. Резултатите са представени в следващите таблици:

Таблица 4.

Регресионна бета на регресиите без и с включването на фиктивна променлива

| Регресионен наклон | Eurocash S.A. | Grupa LOTOS Spółka Akcyjna | KGHM Polska Miedz Spółka Akcyjna | LPP Spółka Akcyjna | Orange Polska S.A. | Polski Koncern Naftowy ORLEN C.A. | Polskie Górnictwo Naftowe i Gazownictwo S.A. | Synthos S.A. |
|-------------------------------------|---------------|----------------------------|----------------------------------|--------------------|--------------------|-----------------------------------|--|--------------|
| а. Регресия без фиктивна променлива | | | | | | | | |
| Eurocash S.A | - | 0,33 | 0,37 | 0,22 | 0,20 | 0,34 | 0,24 | 0,40 |
| Grupa LOTOS | 0,38 | - | 0,57 | 0,30 | 0,26 | 0,66 | 0,39 | 0,47 |
| KGHM Polska Miedz | 0,39 | 0,52 | - | 0,37 | 0,25 | 0,54 | 0,37 | 0,49 |
| LPP | 0,35 | 0,41 | 0,55 | - | 0,21 | 0,44 | 0,38 | 0,42 |
| Orange Polska | 0,24 | 0,26 | 0,28 | 0,16 | - | 0,32 | 0,21 | 0,21 |
| P.K. Naftowy ORLEN | 0,44 | 0,74 | 0,66 | 0,36 | 0,35 | - | 0,42 | 0,51 |
| P. G. N. i G. | 0,49 | 0,69 | 0,73 | 0,50 | 0,37 | 0,67 | - | 0,55 |
| Synthos | 0,38 | 0,40 | 0,46 | 0,26 | 0,18 | 0,39 | 0,26 | - |
| б. Регресия с фиктивна променлива | | | | | | | | |
| Eurocash S.A | - | 0,34 | 0,29 | 0,23 | 0,14 | 0,29 | 0,23 | 0,39 |
| Grupa LOTOS | 0,38 | - | 0,54 | 0,31 | 0,25 | 0,66 | 0,39 | 0,47 |
| KGHM Polska Miedz | 0,40 | 0,57 | - | 0,45 | 0,10 | 0,52 | 0,41 | 0,52 |
| LPP | 0,36 | 0,42 | 0,54 | - | 0,27 | 0,43 | 0,38 | 0,42 |
| Orange Polska | 0,22 | 0,30 | 0,11 | 0,25 | - | 0,25 | 0,24 | 0,28 |
| P.K. Naftowy ORLEN | 0,40 | 0,74 | 0,53 | 0,38 | 0,22 | - | 0,41 | 0,51 |
| P. G. N. i G. | 0,48 | 0,70 | 0,65 | 0,50 | 0,35 | 0,63 | - | 0,54 |
| Synthos | 0,38 | 0,40 | 0,42 | 0,26 | 0,20 | 0,37 | 0,26 | - |

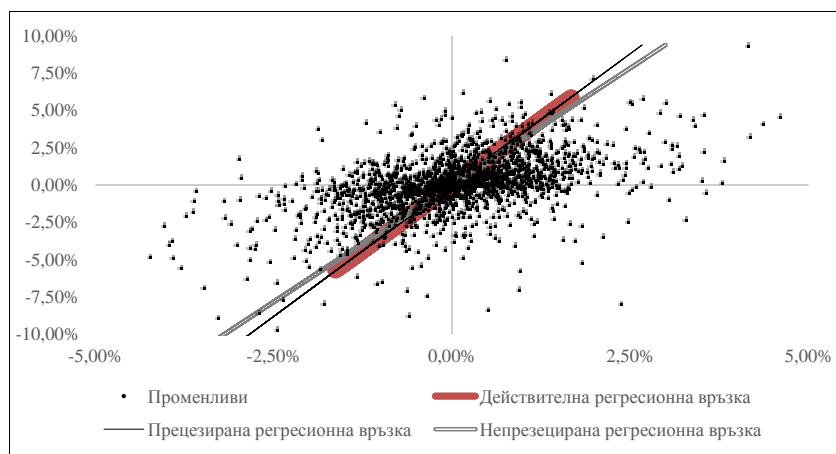
| | | | | | | | | |
|--|-------|--------|-------|-------|-------|-------|--------|-------|
| в. Относителна промяна | | | | | | | | |
| Eurocash S.A | - | 0,02 | -0,24 | 0,02 | -0,39 | -0,17 | -0,02 | 0,00 |
| Grupa LOTOS | 0,00 | - | -0,06 | 0,00 | -0,03 | -0,01 | 0,00 | 0,00 |
| KGHM Polska Miedz | 0,03 | 0,09 | - | 0,20 | -0,92 | -0,04 | 0,09 | 0,05 |
| LPP | 0,01 | 0,00 | -0,03 | - | 0,26 | -0,03 | 0,00 | 0,00 |
| Orange Polska | -0,08 | 0,15 | -0,95 | 0,47 | - | -0,26 | 0,12 | 0,28 |
| P.K. Naftowy ORLEN | -0,10 | 0,01 | -0,22 | 0,04 | -0,45 | - | -0,01 | 0,00 |
| P. G. N. i G. | -0,03 | 0,01 | -0,11 | 0,00 | -0,08 | -0,06 | - | -0,02 |
| Synthos | -0,01 | 0,00 | -0,08 | 0,00 | 0,11 | -0,05 | -0,01 | - |
| г. Регресионен наклон на фиктивната променлива | | | | | | | | |
| Eurocash S.A | - | 0,02 | -0,20 | 0,03 | -0,26 | -0,13 | -0,02 | -0,01 |
| Grupa LOTOS | -0,07 | - | -0,12 | 0,02 | -0,15 | -0,04 | -0,03 | 0,00 |
| KGHM Polska Miedz | 0,01 | 0,05 | - | 0,07 | -0,12 | -0,02 | 0,03 | 0,02 |
| LPP | -0,07 | -0,01 | -0,17 | - | -0,25 | -0,11 | -0,02 | 0,01 |
| Orange Polska | -0,02 | 0,04 | -0,14 | 0,11 | - | -0,06 | 0,03 | 0,06 |
| P.K. Naftowy ORLEN | -0,06 | 0,02 | -0,15 | 0,03 | -0,17 | - | 0,00 | 0,00 |
| P. G. N. i G. | -0,06 | 0,02 | -0,14 | 0,01 | -0,17 | -0,08 | - | -0,03 |
| Synthos | -0,04 | 0,00 | -0,11 | 0,01 | -0,15 | -0,07 | -0,01 | - |
| д. Промяна на знака на фиктивната бета спрямо нормалната бета | | | | | | | | |
| Eurocash S.A | - | He | He | He | He | He | He | He |
| Grupa LOTOS | He | - | He | He | He | He | He | He |
| KGHM Polska Miedz | He | He | - | He | He | He | He | He |
| LPP | Да | Да | He | - | Да | He | Да | Да |
| Orange Polska | He | He | He | He | - | He | He | He |
| P.K. Naftowy ORLEN | He | He | He | He | He | - | He | He |
| P. G. N. i G. | He | He | He | He | He | He | - | He |
| Synthos | He | Да | He | Да | Да | He | He | - |
| е. Относителна промяна на фиктивната бета спрямо нормалната бета | | | | | | | | |
| Eurocash S.A | - | 1,12 | 0,81 | 1,61 | 0,68 | 0,78 | 0,92 | 1,47 |
| Grupa LOTOS | 36,15 | - | 2,13 | 25,13 | 5,07 | 4,51 | 17,05 | 20,02 |
| KGHM Polska Miedz | 0,33 | 0,52 | - | 0,36 | 0,14 | 0,45 | 0,32 | 0,47 |
| LPP | -5,73 | -3,49 | 6,40 | - | -0,97 | 3,53 | -19,71 | -4,71 |
| Orange Polska | 0,23 | 0,26 | 0,15 | 0,24 | - | 0,24 | 0,21 | 0,23 |
| P.K. Naftowy ORLEN | 0,61 | 1,88 | 0,68 | 0,58 | 0,39 | - | 0,57 | 0,78 |
| P. G. N. i G. | 1,98 | 1,79 | 1,30 | 3,30 | 2,22 | 1,26 | - | 1,60 |
| Synthos | 2,96 | -21,33 | 1,34 | -4,34 | -1,36 | 1,40 | 0,83 | - |

Първоначално вниманието се насочва към относителната промяна, която е разнообразна, тъй като е възможно, регресионният наклон да се промени по най-различен начин, след като се премахне влиянието на екстремумите посредством фиктивната променлива.

В този момент обаче се добавят стойностите на регресионния наклон на самата фиктивна променлива, което показва различна тенденция. Относителната промяна почти винаги запазва знака си спрямо регресионната бета на фиктивната променлива. В редките случаи, в които знакът е различен, различна се оказва и самата промяна на фиктивната бета спрямо нормалната бета.

В обобщение на така направените констатации е възможно извеждането на следната тенденция – с добавянето на фиктивната променлива, промяната на регресионната бета традиционно се измества във фиктивната бета, а когато това не е налице, причината е, защото самата промяна е прекалено малка, за да е възможно включването на фиктивната бета да окаже значима разлика върху цялостния регресионен модел.

Дотук резултатите от множеството регресии показват преди всичко успеха в прецизирането на регресионната връзка с добавянето на фиктивна променлива, контролираща за наличните екстремуми, макар и въпросните да са условни. Интересно е обаче как изглежда резултатът, представен графично, показвайки в действителност, какво се случва преди, по време и след контролиране за екстремуми със самата регресионна връзка.



Фигура 8. Резултати от регресиите на финансови променливи с и без отчитане влиянието на екстремуми

Ясно може да се види, че е налице различен наклон (регресионна бета), в случай че екстремумите се отчетат, за сметка на ситуацията, в която същите се игнорират. Т.нар. прецизирана регресионна връзка има наклон, който е не само различен от този на непрецизираната регресионна връзка, но и съвпада с теоретичния наклон, който променливите трябва да имат. Въпросният теоретичен наклон може най-ясно да се изведе чрез използването на алтернативна методология, която директно прилага трун-

кация върху екстремумите, вместо да тества уинсоризацията по един или друг начин.

В резултат на това може да се види, че чрез прецизиране на регресионната връзка посредством използване на фиктивни променливи, третиращи условните екстремуми, е възможно осъществяването на по-добра регресия, получаването на по-стойностен резултат и извеждане на същинската връзка между финансови променливи, която регресията често пропуска заради влиянието на екстремуми.

Заклучение

Научната разработка си поставя амбициозната цел да трансформира икономическата логика, която стои зад извеждането на връзка между финансови променливи посредством регресионен модел. Причината за това е фактът, че често в научната практика много научни изследвания пропускат възможността, наличието на дори една прекалено екстремна стойност изцяло да създаде заблуда относно това, каква в действителност е връзката между различни променливи.

В случая на конкретната разработка това не се прави, а именно – преследва се обратната логика – как чрез трансформация на данните да се постигне именно по-добър резултат, да се изведат по-ясни изводи и да се направят по-убедителни заключения. Същото е възможно например чрез използване на описаната и тествана методология.

Налице е обаче амбициозната идея за продължаване на идеята на разработката в последващи бъдещи изследвания. Причината – възможността методологията допълнително да се трансформира, интегрирайки множество други идеи и концепции, всяка от които да решава различни проблеми на регресионните модели, но предварително – по време на фазата на анализа на данни, а не по време на прилагането на модела, чрез голям брой допускания, предположения и приемане на доказани от математиката, статистиката и иконометрията грешки, всяка от които да прави резултатите от емпиричното изследване все по-лоши и замърсени.

Типичен пример за такава ситуация е проблемът с хетероскедастичността на остатъчните грешки при регресията. Същата е добре позната от много години на науката, но традиционно се третира чрез набралите огромна популярност ARCH модели, стъпващи на идеята за автокорелационна зависимост. Възможна е обаче друга, алтернативна концепция, стъпваща върху апробираната в научната разработка методология.

Тъй като в разработката се допуска условна връзка между променливите, същата се трансформира в регион на доверителност, наподобяващ елипса. В действителност това е само частен случай, изхождащ от статичната стойност на корелацията между променливите. Възможна е например ситуацията, при която корелацията между променливите е нестационарна,

но се изменя константно, което създава ситуация за пропорционално изкривяване на региона на доверителност от трети към първи квадрант (Проданов, 2013). Казано по прост начин формата на региона на доверителност ще се измени от елипсоидна в капковидна структура (издължена в трети квадрант, но клоняща към окръжност в първи), популярна в математиката като лопатка на Джоуковски.

Същото изхожда именно от приемането на корелацията за нестационарна, което на свой ред видоизменя уравнението на корелационната връзка в такова, съдържащо компонента копула. Копулата представлява специфично математическо изражение на част от връзката между две променливи, която много ясно показва само частта на съвместното вероятно разпределение, от която е премахната същинската динамика на връзката (аналогично на трансформацията на ковариационна връзка в корелационна).

Успешното съчетаване на идеята за дефиниране на региона на доверителност на база отчитане формата на копулата ще позволи на анализа да определи ясно до каква степен във финансовите данни е налице стационарност, както и до каква степен – хетероскедастичност. Липсата на първата и наличието на втората в случая вече ще се окажат не проблем, а спецификация, задължаваща регресионният модел да отчете същите чрез например фиктивни променливи, изследващи условни екстремуми, този път дефинирани от уравнение, отчитащо стационарността и хетероскедастичността (както стана дума например във формата на капка/лопатка на Джоуковски).

Същото ще обезсмисли отчитането на т.нар. ARCH ефект, превърнал се в олицетворение на емпиричното моделиране на статистическа връзка през последните десетки години – възможност, която ще зарадва всеки научен работник, несъумяващ да разбере идеята на авторегресионните условни хетероскедастични модели, какъвто е например ... авторът на разработката.

Използвани източници:

Ангелова, Поля. *Статистически методи в емпиричните социологически изследвания*. Свищов: АИ Ценов, (2016).

Брусарски, Румен и др. (2015). *Финансова теория*. В. Търново: Фабер.

Захариев, Андрей и др. (2015). *Финансов анализ*. Свищов: АИ „Ценов“.

Проданов, Стоян и др. (2013). *Инвестиции*. Свищов: АИ „Ценов“.

Пътев, Пламен. (2014). *Международен финансов мениджмънт*. Свищов: АИ „Ценов“.

Радков, Радко и др. (2015). *Международни финанси*. АБАГАР.