

Warsaw University of Life Sciences – SGGW  
Institute of Economics and Finance  
Department of Econometrics and Statistics

**QUANTITATIVE METHODS  
IN ECONOMICS**

**METODY ILOŚCIOWE W BADANIACH  
EKONOMICZNYCH**

**Volume XXI, No. 3**

Warsaw 2020

## **EDITORIAL BOARD**

Editor-in-Chief: Bolesław Borkowski

Vice-Editor-in-Chief: Hanna Dudek

Managing Editor: Grzegorz Koszela

Theme Editors:

Econometrics: Bolesław Borkowski

Multidimensional Data Analysis: Wiesław Szczesny

Mathematical Economy: Zbigniew Binderman

Analysis of Labour Market: Joanna Landmesser

Financial Engineering: Monika Krawiec

Data Science: Michał Gostkowski

Theory of Risk: Marek Andrzej Kociński

Statistical Editor: Wojciech Zieliński

Technical Editors: Jolanta Kotlarska, Elżbieta Saganowska

Language Editor: Agata Cienkusz

Native Speaker: Yochanan Shachmurove

Editorial Assistant: Luiza Ochnio

## **SCIENTIFIC BOARD**

Adnene Ajimi (University of Sousse, Tunisia)

Heni Boubaker (University of Sousse, Tunisia)

Vasily Dikussar (Doradnicyn Computing Centre of the Russian Academy of Sciences, Russia)

Peter Friedrich (University of Tartu, Estonia)

Paolo Gajo (University of Florence, Italy)

Agnieszka Gehringer (University of Göttingen, Germany)

Anna Maria Gil-Lafuente (University of Barcelona, Spain)

Jaime Gil-Lafuente (University of Barcelona, Spain)

Vasile Glavan (Moldova State University, Moldova)

Francesca Greselin (The University of Milano-Bicocca, Italy)

Ana Kapaj (Agriculture University of Tirana, Albania)

Jirawan Kitchaicharoen (Chiang Mai University, Thailand)

Yuriy Kondratenko (Black Sea State University, Ukraine)

Vassilis Kostoglou (Alexander Technological Educational Institute of Thessaloniki, Greece)

Karol Kukula (University of Agriculture in Krakow, Poland)

Kesra Nermend (University of Szczecin, Poland)

Nikolas N. Olenev (Doradnicyn Computing Centre of the Russian Academy of Sciences, Russia)

Alexander N. Prokopenya (Brest State Technical University, Belarus)

Yochanan Shachmurove (The City College of The City University of New York, USA)

Mirbulat B. Sikhov (al-Farabi Kazakh National University, Kazakhstan)

Marina Z. Solesvik (Nord University, Norway)

Ewa Syczewska (Warsaw School of Economics, Poland)

Achille Vernizzi (University of Milan, Italy)

Andrzej Wiatrak (University of Warsaw, Poland)

Dorota Witkowska (University of Lodz, Poland)

ISSN 2082-792X

e-ISSN 2543-8565

© Copyright by Department of Econometrics and Statistics WULS – SGGW  
(Katedra Ekonometrii i Statystyki SGGW)

Warsaw 2020, Volume XXI, No. 3

The original version is the paper version


Journal homepage: [qme.sggw.pl](http://qme.sggw.pl)

Published by Warsaw University of Life Sciences Press

## CONTENTS

Domańska Sylwia Anna – Dopasowanie modeli GARCH a jakość uzyskanych prognoz .....	121
Pietrych Łukasz – Identyfikacja powiatów o największym ryzyku rozprzestrzeniania się choroby COVID-19 i obciążenia służby zdrowia w Polsce .....	134
Orzechowski Arkadiusz – Pricing European Options in Selected Stochastic Volatility Models .....	145
Binderman Zbigniew, Borkowski Bolesław, Szczesny Wiesław – Wykorzystanie wzorców pośrednich do budowy mierników syntetycznych .....	157
Pietrzykowski Robert – Wykorzystanie współczynnika Morana do klasyfikacji obiektów przestrzennych na przykładzie cen gruntów rolniczych .....	168

## DOPASOWANIE MODELI GARCH A JAKOŚĆ UZYSKANYCH PROGNOZ

Sylwia Anna Domańska  <https://orcid.org/0000-0002-9418-778X>  
Wydział Nauk Ekonomicznych, Uniwersytet Warszawski  
e-mail: sdomanska@wne.uw.edu.pl

**Streszczenie:** Analiza dopasowania modeli GARCH przedstawiona w artykule składa się z trzech części. Po pierwsze, dokonano rozważań teoretycznych w tym zakresie, odnosząc się do zaleceń wskazanych w literaturze przedmiotu. Po drugie, przedstawiono praktyczne aspekty właściwego doboru modelu GARCH dla danego szeregu czasowego na podstawie wartości kryteriów informacyjnych. Po trzecie, wskazano, za pomocą narzędzi analizy ilościowej, związek jakości dopasowania modelu z jakością uzyskanych za jego pomocą prognoz. Celem niniejszego artykułu jest sprawdzenie czy lepsze dopasowanie modelu skutkuje lepszą jakością uzyskanej prognozy.

**Słowa kluczowe:** prognozowanie, modele GARCH, szeregi czasowe, dopasowanie

**JEL classification:** C32, C53, C58, G10, G17

### WSTĘP

Istotną cechą rynków finansowych, odgrywającą ważną rolę przy ich wycenie, jest zmienność cen instrumentów finansowych. Wysoka zmienność cen to z jednej strony szansa na ponadprzeciętny zysk, z drugiej możliwość dużej straty. W literaturze przedmiotu wymienia się dwie główne cechy zmienności, które komplikują wycenę instrumentów finansowych. Po pierwsze, zmienność cen instrumentów finansowych ewoluuje w czasie oraz posiada tendencję do tworzenia skupień. Jeśli w jakimś okresie pojawi się większa wartość wariancji, to najprawdopodobniej pociągnie ona za sobą kolejną większą wartość i odwrotnie. Po drugie, cechą zmienności jest jej nieobserwowalność, która wymusza poszukiwanie różnego rodzaju miar zmienności. Ważnym sposobem prognozowania zmienności jest wykorzystanie modeli GARCH.

<https://doi.org/10.22630/MIBE.2020.21.3.12>

Analiza dopasowania modeli GARCH przedstawiona w artykule składa się z trzech części. Po pierwsze, dokonano rozważań teoretycznych w tym zakresie, odnosząc się do zaleceń wskazanych w literaturze przedmiotu. Po drugie, przedstawiono praktyczne aspekty właściwego doboru modelu GARCH dla danego szeregu czasowego na podstawie wartości kryteriów informacyjnych. Po trzecie, wskazano, za pomocą narzędzi analizy ilościowej, związek jakości dopasowania modelu z jakością uzyskanych za jego pomocą prognoz. Celem niniejszego artykułu jest sprawdzenie czy lepsze dopasowanie modelu skutkuje lepszą jakością uzyskanej prognozy.

## CHARAKTERYSTYKA SZEREGÓW FINANSOWYCH

Finansowe szeregi czasowe różnią się pod wieloma względami od typowych danych statystycznych. Kursy akcji czy indeksów zawsze odnoszą się do zysków lub strat określonej grupy inwestorów. Wpływ na nie mają często wydarzenia gospodarcze, ogólnospołeczne, polityczne czy też nieracjonalne zachowania inwestorów.

W badaniach ekonometrycznych dotyczących finansowych szeregów czasowych modeluje się najczęściej tempo wzrostu instrumentów finansowych, klasycznie definiowane jako<sup>1</sup>:

$$r_t = (y_t - y_{t-1}) / y_{t-1} \quad (1)$$

gdzie  $y_t$  jest kursem danego instrumentu finansowego. W literaturze finansowej  $r_t$  nazywane jest stopą zwrotu. Alternatywnie stosuje się przyrost logarytmu kursu  $y_t$ :

$$r_t = \ln y_t - \ln y_{t-1} = \ln(y_t / y_{t-1}) \quad (2)$$

Charakter szeregów różni się w zależności od segmentów, z którego pochodzą. Zwroty cen akcji mają inne własności niż zwroty kursów walutowych, bądź zwroty notowań indeksów<sup>2</sup>. Możemy jednak wyróżnić grupę cech charakterystyczną dla większości finansowych szeregów zwrotów logarymicznych.

Po pierwsze bardzo często pojawia się zjawisko grupowania wariancji, czyli seryjne występowanie wysokich i niskich wartości wariancji kursu danego instrumentu finansowego, w kolejnych następujących po sobie okresach. W literaturze przedmiotu można znaleźć wiele hipotez próbujących wyjaśnić przyczynę występowania tego efektu. Najbardziej popularna dotyczy charakteru samego procesu napływu informacji na rynek. W zależności od rodzaju informacji oraz jej subiektywnej oceny inwestorzy mogą sprzedawać lub kupować

---

<sup>1</sup> Brzeszczyński J., Kelm R. (2002) Ekonometryczne modele rynków finansowych, WIG-Press, Warszawa, s. 37.

<sup>2</sup> Doman M., Doman R. (2009) Modelowanie zmienności i ryzyka, Oficyna Wolters Kluwer, Kraków, s. 22.

instrumenty finansowe. Występują więc po stronie popytu, bądź podaży kształtując ceny giełdowe.

Zjawisko grupowania wariancji pojawia się jako skutek reakcji rynku na napływ informacji, które powodują rodzaj zaburzenia. Efektem jest wzrost, bądź spadek danego zjawiska, co w konsekwencji prowadzi do zwiększenia jego zmienności<sup>3</sup>. Pojawienie się informacji może spowodować krótkookresowe wzrosty, bądź spadki wartości kursu, które nie powodują zmiany dotychczasowego trendu. Zazwyczaj taka sytuacja pojawia się w relatywnie krótkich okresach, kiedy występuje niepewność związana z efektem oddziaływania danej informacji w najbliższej przyszłości.

Alternatywna hipoteza głosi, że zjawisko grupowania wariancji pochodzi z otoczenia makroekonomicznego, w którym funkcjonuje rynek. Czynniki determinującymi zmienność cen giełdowych są bowiem: regulacje prawne, zmiany podaży pieniądza, wielkość dywidendy, fazy cyklu gospodarczego, jak również zmiany stóp procentowych<sup>4</sup>. Czynniki te oddziałują na rynek papierów wartościowych, przyczyniając się do powstawania efektu grupowania wariancji.

Zgodnie z inną hipotezą zjawisko grupowania wariancji powstaje jako efekt grupowania kolejnych wartości obrotu giełdowego dla danego instrumentu finansowego<sup>5</sup>. Badania empiryczne wskazały bowiem na wysoką korelację pomiędzy zmiennością indeksu a wartością obrotu na giełdzie. Pokazano również, że duże zmiany kursów akcji poszczególnych spółek są powiązane z wysoką wartością obrotów tymi papierami wartościowymi.

Drugą charakterystyczną cechą szeregów finansowych jest ich leptokurtyczność. Przeprowadzone analizy wskazują na występowanie grubych ogonów, zaś typowe oszacowania kurtozy mieszczą się w granicach od 4 do 50, co świadczy o dużym odstępstwie od rozkładu normalnego. Na leptokurtyczność wskazuje również indeks ogona rozkładu, który przyjmuje wartości między 2 a 5<sup>6</sup>. Oznacza to, że prawdopodobieństwo wystąpienia nietypowych zmian kursów jest większe niż w przypadku rozkładu normalnego. Warto jednak zauważyć, że zwiększając jednostkę skali czasowej, względem której obliczane są zwroty, rozkład zwrotów upodabniają się do normalnego.

Rozkłady zwrotów instrumentów finansowych są również prawostronnie skośne, czyli wzrosty kursów występują częściej niż spadki. W szeregach notowań indeksów giełdowych i akcji, rzadziej w szeregach kursów walutowych, zauważyć

---

<sup>3</sup> Brzeszczyński J., Kelm R. (2002) Ekonometryczne modele rynków finansowych, WIG-Press, Warszawa, s. 41.

<sup>4</sup> Bollerslev T., Chou R. Y., Kroner K. F. (1992) ARCH Modeling in Finance: A Selective Review of the Theory and Empirical Evidence. *Journal of Econometrics*, 52, 5-59.

<sup>5</sup> Gallant A., Rossi P., Tauchen G. (1992) Stock Prices and Volume. *The Review of Financial Studies*, 5(2), 199-242.

<sup>6</sup> Doman M., Doman R. (2009) Modelowanie zmienności..., op.cit., s. 24.

można asymetrię spadków i wzrostów. Oznacza to, że znacznie częściej pojawiają się ruchy w dół co do wartości bezwzględnej większe niż znaczne wzrosty.

Kolejną cechą finansowych szeregów finansowych jest występowanie efektu dźwigni. Zmiany kursów są ujemnie skorelowane ze zmiennością ich wariancji. Wraz ze spadkiem wartości danych kursów ich zmiany względne są bowiem wyższe, o ile zmiany bezwzględne są w przybliżeniu stałe.

Finansowe szeregi czasowe cechują się również występowaniem autokorelacji stóp zwrotu. Stopy zwrotu w danym okresie zależą bowiem od ich realizacji w okresach poprzednich. Autokorelacja może dotyczyć większego rzędu opóźnień niż jeden. Efekt ten określa się jako AR(p) gdzie p – to rząd opóźnień. Badanie empiryczne wykazują, że autokorelacja może dotyczyć także wariancji. Także i w tym przypadku oznacza to, że zmienność w jednym podokresie zależy od zmienności zrealizowanej w okresach poprzednich. Własność tą oznacza się ją jako efekt ARCH(q) gdzie q także i w tym przypadku – określa rząd opóźnień. Na rynkach finansowych często występuje zależność między wariancją tempa wzrostu kursów i ich autokorelacją. Autokorelacja występuje zazwyczaj w przypadku małej zmienności kursowej, zaś duża wariancja powoduje brak autokorelacji.

## SPECYFIKACJE MODELI GARCH

Ogromne znaczenie w zarządzaniu ryzykiem odgrywa możliwość skutecznego prognozowania przyszłej zmienności poszczególnych instrumentów finansowych lub złożonych z nich portfeli. Przez zmienność ceny instrumentu finansowego rozumie się zazwyczaj warunkową wariancję zwrotu. Najczęściej stosowanym narzędziem prognozowania zmienności są uogólnione modele warunkowej autoregresyjnej heteroskedastyczności – GARCH.

Powszechność użycia modeli GARCH wynika z faktu, że uwzględniają one większość empirycznych własności finansowych szeregów czasowych. Na uwagę zasługuje również łatwość estymacji ich parametrów oraz łatwość rozszerzania modeli<sup>7</sup>. Ze względu na bardzo dużą liczbę uogólnień, najlepszym sposobem wyboru odpowiedniej specyfikacji jest analizowanie wybranych modeli GARCH.

Uogólniony model autoregresyjny warunkowej heteroskedastyczności minimalizuje liczbę parametrów przy dużej liczbie opóźnień oraz dobrze opisuje rozkłady o dużych ogonach. Został on zaproponowany przez Bollersleva [1986] jako uogólnienie procesu ARCH wprowadzonego przez Engle'a w 1985r.

Model GARCH(p, q) można zdefiniować jako proces stochastyczny  $\{\varepsilon_t, t \in \mathbb{Z}\}$  spełniający następujące warunki<sup>8</sup>:

<sup>7</sup> Fiszeder P. (2009) Modele klasy GARCH w empirycznych badaniach finansowych. Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń, s. 21.

<sup>8</sup> Pipień M. (2006) Wnioskowanie bayesowskie w ekonometrii finansowej. Zeszyty Naukowe, Akademia Ekonomiczna w Krakowie. Seria Specjalna. Monografie. 176, s. 52.

- 1)  $\varepsilon_t = z_t \sqrt{h_t}$ ,
- 2)  $h_t = a_0 + \sum_{i=1}^q a_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}$ ,  $a_0 > 0$ ,  $a_i \geq 0$ , dla  $i = 1, \dots, q$  oraz  $\beta_j \geq 0$  dla  $j = 1, \dots, p$ ,
- 3)  $\{z_t, t \in \mathbb{Z}\}$  jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie prawdopodobieństwa, spełniającym warunki:  $E(z_t) = 0$ ,  $E(z_t^2) = 1$ ,  $E(z_t^3) = 0$ ,  $E(z_t^4) = \lambda > 0$ .

Równanie wariancji warunkowej  $h_t$  uzależnia zmienność zarówno od jej wartości z przeszłości, jak również od wartości zaobserwowanych kwadratów stóp zwrotu. Parametry  $a_i$  decydują więc o wpływie, jaki na zmienność mają nowe napływające informacje zawarte w kwadratach  $\varepsilon_{t-i}^2$ . Parametry  $\beta_j$  charakteryzują zaś tę część dynamiki, która opisuje oczekiwania rynku odnośnie tego, że w przyszłości proces zmienności będzie przebiegał podobnie jak dotychczas.

Typowym i najczęściej spotykanym w badaniach empirycznych procesem jest model GARCH(p, q) dla  $p=1$  i  $q=1$ . Przyjęcie bowiem takich rzędów opóźnień umożliwia w sposób oszczędny uzależnienie  $h_t$  od całej historii  $\varepsilon_t$ . Oryginalny, zgodny z definicją Bollersleva<sup>9</sup>, proces GARCH(1, 1) ma warunkowy rozkład normalny o zerowej wartości oczekiwanej i wariancji  $h_t$ :

$$\varepsilon_t | \psi_{t-1}(\varepsilon) \sim N(0, h_t), \text{ jeśli } z_t \sim iN(0, 1),$$

$$h_t = a_0 + a_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, a_0 \geq 0, a_1 > 0 \text{ oraz } \beta_1 > 0, \beta_1 < 1. \quad (3)$$

Rozkład bezwarunkowy  $\varepsilon_t$  ma zerową wartość oczekiwaną oraz wariancję równą  $a_0(1 - a_1 - \beta_1)^{-1}$  dla  $a_1 + \beta_1 < 1$ . Rozkład ten ma ogony grubsze niż rozkład warunkowy. Dodatkowo jeśli przyjmiemy, że  $z_t$  ma rozkład normalny, to rozkład bezwarunkowy modeli GARCH (1, 1) ma ogony cięższe niż rozkład normalny. Bollerslev dowiódł również, że dla  $a_1 + \beta_1 < 1$  GARCH(1, 1) jest stacjonarny w sensie kowariancji. D. Nelson<sup>10</sup> wyprowadził natomiast warunek konieczny i wystarczający ścisłej stacjonarności modeli typu GARCH. Zgodnie z tym, jeśli parametry modelu GARCH (1, 1) spełniają warunek:  $E(\ln(a_1 z_t^2 + \beta_1)) < 0$ , to jest on ściśle stacjonarny.

Bardzo często w badaniach empirycznych suma ocen parametrów  $a_1$  i  $\beta_1$  w modelu GARCH(1, 1) jest bliska jedności. Tę własność można potraktować jako typową dla szeregów finansowych. Przypadek, gdy  $a_1 + \beta_1 = 1$ , jest określany jako zintegrowany model GARCH(1, 1) – *Integrated GARCH(1, 1)*; IGARCH. Model IGARCH nie jest procesem stacjonarnym kowariancyjnie, jednak jest on

<sup>9</sup> Bollerslev T. (1986) Generalised Autoregressive Conditional Heteroscedasticity. Journal of Econometrics, 31(3), s. 311.

<sup>10</sup> Nelson D. (1990) Stationarity and Persistence in GARCH(1, 1) Model. Econometric Theory, 6, s.324.



nadal procesem ściśle stacjonarnym. W przypadku modeli GARCH stacjonarność kowariancyjna jest więc własnością znacznie mocniejszą niż ścisła stacjonarność<sup>11</sup>.

Zgodnie z definicją model IGARCH(p, q) to taki proces stochastyczny, który dla każdego  $t \in \mathbb{Z}$  spełnia następujące warunki<sup>12</sup>:

$$\varepsilon_t = z_t \sqrt{h_t}, \quad (4)$$

$$h_t = a_0 + \varepsilon_{t-i}^2 + \sum_{i=2}^q a_i (\varepsilon_{t-i}^2 - \varepsilon_{t-1}^2) + \sum_{j=1}^p \beta_j (h_{t-j} - \varepsilon_{t-1}^2) \quad (5)$$

$$\sum_{i=1}^q a_i + \sum_{j=1}^p \beta_j = 1 \quad (6)$$

gdzie:  $\{\varepsilon_t\} \sim iid(0, 1)$ ,  $a_0 > 0$ ,  $a_i \geq 0$ , dla  $i = 1, \dots, q$ ,  $\beta_j \geq 0$ ,  $\beta_j \geq 0$  dla  $j = 1, \dots, p$ ,  $a_q \neq 0$ ,  $\beta_p \neq 0$ .

Najczęściej stosowanym modelem do opisu długiej pamięci w zmienności jest model FIGARCH. Jest to proces niestacjonarny kowariancyjnie (nie istnieje wariancja bezwarunkowa  $\varepsilon_t$ ) oraz procesem ściśle stacjonarnym. Model FIGARCH(p, d, q) został zdefiniowany przez Baillie'a, Bolersleva i Mikkelsena następująco<sup>13</sup>:

$$\varphi(L)(1-L)^d \varepsilon_t^2 = a_0 [1 - \beta(L)] v_t \quad (7)$$

gdzie:  $v_t = \varepsilon_t^2 - h_t$ ,  $\varphi(L) = 1 - \sum_{j=1}^q \varphi_j L^j$ ,  $\beta(L) = \sum_{j=1}^p \beta_j L^j$ ,  $L^s \varepsilon_t = \varepsilon_{t-s}$ , dla  $0 < d < 1$ , a wszystkie pierwiastki równania  $\varphi(L) = 0$  leżą poza kołem.

Jeżeli  $d = 1$ , to model FIGARCH jest modelem IGARCH, zaś dla  $d = 0$  jest procesem GARCH. Należy jednak podkreślić, że proces FIGARCH(p, 0, q) nie zawsze redukuje się do procesu GARCH(p, q). Wpływ obecnej zmienności na jej prognozowane wartości w przypadku kowariancyjnie stacjonarnego modelu GARCH maleje do zera w tempie wykładniczym. Dla modelu IGARCH bieżąca zmienność ma nieskończony wpływ na prognozę wariancji warunkowej. W przypadku modeli FIGARCH wpływ ten maleje do zera zgodnie z funkcją hiperboliczną. Warto również zauważyć, że funkcja autokorelacyjna dla  $\varepsilon_t^2$  w przypadku procesów GARCH oraz IGARCH wygasa w tempie wykładniczym, zaś dla FIGARCH – w tempie hiperbolicznym<sup>14</sup>. Proces IGARCH charakteryzuje

<sup>11</sup> Osiewalski J., Pipień M. (1999) Bayesowskie wnioskowanie o stacjonarności procesów GARCH(1, 1). *Dynamiczne Modele Ekonometryczne*, Materiały na VI Ogólnopolskie Seminarium Naukowe, UMK, Toruń.

<sup>12</sup> Pajor A. (2010) Wielowymiarowe procesy wariancji stochastycznej w ekonometrii finansowej. *Ujęcie bayesowskie*. Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, s. 40.

<sup>13</sup> Baillie R. T., Bollerslev T., Mikkelsen H. O. (1996) Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 74, 3-30.

<sup>14</sup> Ding Z., Granger C. W. J. (1996) Modeling Volatility Persistence of Speculative Returns: A New Approach. *Journal of Econometrics*, 73, 185-215.

się krótką pamięcią, natomiast FIGARCH dla  $0 < d < 1$  ma długą pamięć w zmienności.

Ogólne modele GARCH zakładają, że zarówno pozytywne jak i negatywne zmiany cen mają ten sam wpływ na przyszłą zmienność. Tymczasem zauważyć można, że ich wpływają one asymetrycznie na poziom przyszłej zmienności. Asymetrię tę uwzględnia model APARCH, którymi szczególnymi przypadkami są m.in. modele: TS-GARCH, NARCH, Log-ARCH oraz GJR-GARCH.

Model APARCH(p, q) autorstwa Dinga, Grangera i Engle'a został określony w następujący sposób<sup>15</sup>:

$$z_t = \varepsilon_t \sqrt{h_t} \quad (8)$$

$$h_t^{\delta/2} = a_0 + \sum_{i=1}^q a_i (|z_{t-i}| - \gamma_i z_{t-i})^\delta + \sum_{j=1}^p \beta_j h_{t-j}^{\delta/2} \quad (9)$$

gdzie:  $\delta > 0$ ,  $\gamma \in (-1; 1)$ ,  $i = 1, 2, \dots, q$ .

Cechą charakterystyczną tego modelu, oprócz obecności współczynników asymetrii, jest możliwość dopasowania wykładnika  $\delta$ , gwarantującego istnienie bezwarunkowego momentu rzędu  $\delta$  dla procesu  $\sigma_t$ .

Szczególnym przypadkiem modelu APARCH jest zaproponowany przez Glostena, Jagannathana i Runkle'a model GJR-GARCH(p, q). Wariancja warunkowa w tym modelu określona jest równaniem:

$$h_t = a_0 + \sum_{i=1}^q a_i \varepsilon_{t-i}^2 + \sum_{i=1}^q \omega_i I_{t-i} \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j} \quad (10)$$

gdzie  $I(\cdot)$  jest funkcją wskaźnikową.

Warto zauważyć, że model GJR-GARCH(1, 1) jest kowariancyjnie stacjonarny, gdy spełnia warunek:

$$a_1 + \beta_1 + \frac{1}{2} \gamma_1 < 1 \quad (11)$$

Modelem, który pozwala opisać zależność między oczekiwaną stopą zwrotu a ryzykiem wyrażonym za pomocą warunkowej wariancji, bądź warunkowego odchylenia standardowego jest model GARCH-M. Proces ten został zaproponowany przez Engle'a, Liliena i Robinsa<sup>16</sup> jako odpowiedź na postulowaną w licznych teoriach finansowych bezpośrednią zależność pomiędzy wielkością zwrotu a ryzykiem. Model GARCH-M(p, q) ma następującą postać:

$$r_t = x_t' \xi + \delta g(h_t) + \varepsilon_t \quad (12)$$

$$\varepsilon_t = z_t \sqrt{h_t} \quad (13)$$

<sup>15</sup> Ding Z., Granger C. W. J., Engle R. F. (1993) A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, 1, s. 99.

<sup>16</sup> Engle R. F., Lilien D. M., Robins R. P. (1987) Estimating Time Varying Risk Premia in the Term Structure: the ARCH-M Model. *Econometrica*, 55, s. 394.

$$h_t = a_0 + \sum_{i=1}^q a_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j} \quad (14)$$

gdzie:  $x_{t-}$  wektor zmiennych objaśniających, zaś  $\delta$  zgodnie z jego interpretacją ekonomiczną nosi nazwę parametru premii za ryzyko.

Jak widać powyższy model uwzględnia w równaniu średniej warunkowej zmiennej reprezentującej zwrot, funkcję, której argumentem jest zmienność.

Tabela 1. Specyfikacje modeli GARCH

Model	Postać modelu	Objaśnienia zmiennych oraz warunki zapewniające dodatniość $h_t$
GARCH(p, q)	$h_t = a_0 + \sum_{i=1}^q a_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}$	$a_0 > 0$ , $a_i \geq 0$ , dla $i = 1, \dots, q$ oraz $\beta_j \geq 0$ dla $j = 1, \dots, p$ .
GJR-GARCH(p, q)	$h_t = a_0 + \sum_{i=1}^q a_i \varepsilon_{t-i}^2 + \sum_{i=1}^q \omega_i I_{t-i} \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}$	$I_{t-i} = 1$ gdy $\varepsilon_{t-i} \leq 0$ i $I_{t-i} = 0$ gdy $\varepsilon_{t-i} > 0$ , $a_0 > 0$ , $a_i \geq 0$ , $a_i + \omega_i \geq 0$ , dla $i = 1, \dots, q$ , $\beta_j \geq 0$ dla $j = 1, \dots, p$ .
GARCH-M(p, q)	$r_t = x_t' \xi + \delta g(h_t) + \varepsilon_t$ $h_t = a_0 + \sum_{i=1}^q a_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j h_{t-j}$	$g(h_t) = h_t$ lub $g(h_t) = \sqrt{h_t}$ lub $g(h_t) = \ln(h_t)$ , $x_{t-}$ wektor zmiennych objaśniających oraz $a_0 > 0$ , $a_i \geq 0$ , dla $i = 1, \dots, q$ , $\beta_j \geq 0$ dla $j = 1, \dots, p$ .
IGARCH(p, q)	$h_t = a_0 + \varepsilon_{t-i}^2 + \sum_{i=2}^q a_i (\varepsilon_{t-i}^2 - \varepsilon_{t-1}^2) + \sum_{j=1}^p \beta_j (h_{t-j} - \varepsilon_{t-1}^2)$	$a_0 > 0$ , $a_i \geq 0$ , dla $i = 2, \dots, q$ , $\beta_j \geq 0$ dla $j = 1, \dots, p$ oraz $\sum_{i=2}^q a_i + \sum_{j=1}^p \beta_j \leq 1$ .
FIGARCH(p, d, q)	$\varphi(L)(1-L)^d \varepsilon_t^2 = a_0 [1 - \beta(L)] v_t$	$v_t = \varepsilon_t^2 - h_t$ , $\varphi(L) = 1 - \sum_{j=1}^q \varphi_j L^j$ , $\beta(L) = \sum_{j=1}^p \beta_j L^j$ , $L^s \varepsilon_t = \varepsilon_{t-1}$ , $0 < d < 1$ , wszystkie pierwiastki równania $\varphi(L) = 0$ leżą poza kołem jednostkowym.

Model	Postać modelu	Objaśnienia zmiennych oraz warunki zapewniające dodatniość $h_t$
APARCH(p, q)	$h_t^{\delta/2} = a_0 + \sum_{i=1}^q a_i ( z_{t-i}  - \gamma_i z_{t-i})^\delta + \sum_{j=1}^p \beta_j h_{t-j}^{\delta/2}$	$\delta > 0, a_0 > 0, a_i \geq 0,$ dla $i = 1, \dots, q,$ $-1 < \gamma_i < 1, \beta_j \geq 0$ dla $j = 1, \dots, p.$

Źródło: opracowanie własne na podstawie Fiszeder P. [2009] Modele klasy GARCH..., op.cit, Toruń, s. 25

### TESTOWANIE JAKOŚCI DOPASOWANIA MODELU ORAZ OCENA JAKOŚCI PROGNOZ

Miary syntetyczne pozwalają zmierzyć precyzję i obciążenie prognoz. Precyzję mierzymy najczęściej za pomocą miar absolutnych: średniego błędu bezwzględnego (*mean absolute error* – *MAE*) i średniego bezwzględnego błędu procentowego (*mean absolute percentage error* – *MAPE*):

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_{T+i} - \hat{y}_{T+i}| \quad (15)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_{T+i} - \hat{y}_{T+i}|}{y_{T+i}} \quad (16)$$

oraz kwadratowych, m.in. pierwiastka błędu średniokwadratowego (*root mean square error* – *RMSE*):

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{T+i} - \hat{y}_{T+i})^2} \quad (17)$$

Obciążenie, czyli średnie systematyczne odchylenie się prognoz od wartości empirycznych jest mierzone natomiast za pomocą błędu średniego (*mean error* – *ME*), wyrażonego wzorem:

$$ME = \frac{1}{N} \sum_{i=1}^N (y_{T+i} - \hat{y}_{T+i}) \quad (18)$$

Porównanie odpowiednich miar absolutnych i kwadratowych informuje o nietypowo dużych błędach prognoz. Porównując miary precyzji i obciążenia

możemy wywnioskować, czy prognozy nie są w sposób systematyczny obciążone<sup>17</sup>.

Najpopularniejszą miarą precyzji modelu jest błąd średniokwadratowy (*mean square error – MSE*):

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_{T+i} - \hat{y}_{T+i})^2 \quad (19)$$

Z kwadratowym błędem prognozy związany jest również współczynnik Theila, wyrażony wzorem:

$$TIC = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (y_{T+i} - \hat{y}_{T+i})^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^N y_{T+i}^2 + \frac{1}{N} \sum_{i=1}^N \hat{y}_{T+i}^2}} \quad (20)$$

Jego pierwiastek informuje o całkowitym względnym błędzie prognozy w okresie testowania. Współczynnik Theila przyjmuje wartość zero dla prognoz idealnie trafnych. Im jego wartość jest mniejsza, tym mniejsze są różnice między prognozami a wartościami rzeczywistymi zmiennej prognozowanej. Współczynnik ten można rozdzielić na sumę trzech składników, dzięki czemu można ocenić źródła błędów predykcji:

$$TIC^2 = I_1^2 + I_2^2 + I_3^2 \quad (21)$$

Poszczególne elementy równania określone są następująco:

- 1)  $I_1^2 = \frac{N(\bar{y}_{T+i} - \hat{y}_{T+i})^2}{\sum_{i=1}^N y_{T+i}^2}$ , gdzie  $\bar{y}_{T+i} = I_1^2$  odzwierciedla obciążenie predykcji, czyli w jakim stopniu nie udało się odgadnąć średniej wartości zmiennej prognozowanej.
- 2)  $I_2^2$  określa, w jakim stopniu zmienność prognozy i zmiennej prognozowanej są do siebie zbliżone. Nieodgadnięcie wahań zmiennej prognozowanej na skutek wygładzania szeregów związane jest na ogół z mniejszą zmiennością prognoz niż wartości empirycznych. Składnik ten ma związek z precyzją prognozowania.
- 3)  $I_3^2$  określa, czy wystąpiła zgodność kierunków zmian prognoz z rzeczywistym kierunkiem zmian zmiennej prognozowanej.  $I_3^2 = 0$ , gdy współczynnik korelacji między nimi wynosi jeden. Nieodgadnięcie kierunku tendencji w szeregu świadczy o słabej przydatności modelu do odwzorowania tych tendencji.

<sup>17</sup> Doman M., Doman R. (2009) Modelowanie zmienności..., op.cit., s. 94.

## PROGNOZOWANIE ZMIENNOŚCI PRZY UŻYCIU PAKIETU SAS

Do analizy empirycznej wykorzystano logarytmiczne stopy zwrotu indeksu WIG w okresie od 8 listopada 2010 do 6 listopada 2020. Dane z okresu od 8 listopada 2010 do 6 listopada 2019 wykorzystano do estymacji modeli. Na okres od 6 listopada 2019 do 6 listopada 2020 dokonano prognozy. Do zbadania zmienności stóp zwrotu wykorzystano różne modele z rodziny GARCH. Na podstawie analizy statystyk opisowych ustalono, że najlepiej dopasowanymi modelami GARCH będą modele z rozkładem błędu t-Studenta. Wyniki testu Engle'a przesądziły o wyborze modeli z opóźnieniem 1. Niezależnie od jakości dopasowania oraz poziomu istotności parametrów wyznaczono prognozy.

W wyniku estymacji modeli GARCH otrzymano następujące wartości kryteriów informacyjnych:

Tabela 2. Kryteria informacyjne

Obs	Model	SBC	AIC
1	ar 1 garch 1 1	14922,4866	14896,8615
2	st garch 1 1	14947,8663	14928,6475
3	qgarch 1 1	14957,3193	14931,6943
4	tgarch 1 1	14966,2723	14938,6473
5	pgarch 1 1	14972,0388	14940,0075
6	garch 1 1	14976,0503	14956,8315
7	garchm 1 1	14983,2484	14957,6234
8	egarch 1 1	14988,1548	14962,5297
9	igarch 1 1	16512,2964	16505,8901

Źródło: opracowanie własne na podstawie estymacji w SAS

Ocena jakości dopasowania na podstawie kryteriów informacyjnych pozwala uszeregować modele w rankingu. Najlepszy okazał się model ar(1)-garch(1, 1), wartość kryteriów informacyjnych jest bowiem najniższa. Kolejny jest model st-garch(1, 1). Najgorzej dopasowany okazał się model igarch (1,1).

Kolejny etap polega na sprawdzeniu, czy z najlepiej dopasowanych modeli otrzymamy najlepszej jakości prognozy. Do oceny jakości prognoz zastosowano porównanie na podstawie wymienionych wcześniej błędów prognozy (tabela 3).

Tabela 3. Ocena jakości prognoz

Obs	Model	MSE	MAE	MAPE
1	ar 1 garch 1 1	0,0003054	0,01063896	57,69
2	st garch 1 1	0,0002648	0,02066475	60,26
3	qgarch 1 1	0,0003123	0,01631352	58,02
4	tgarch 1 1	0,0003326	0,02252550	52,49
5	pgarch 1 1	0,0002622	0,01075550	48,22
6	garch 1 1	0,0003054	0,02407896	62,41

Obs	Model	MSE	MAE	MAPE
7	garchm 1 1	0,0002826	0,01785352	61,47
8	egarch 1 1	0,0003748	0,02650297	65,59
9	igarch 1 1	0,0003194	0,01208901	58,62

Źródło: opracowanie własne na podstawie wyliczeń w SAS

Jak widać nie ma bezpośredniego związku pomiędzy jakością dopasowania modelu a jakością uzyskanych z niego prognoz. Różnice zarówno w dopasowaniu poszczególnych modeli, jak również jakością uzyskanych prognoz są niewielkie.

## BIBLIOGRAFIA

- Baillie R. T., Bollerslev T., Mikkelsen H. O. (1996) Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 74, 3-30.
- Bollerslev T. (1986) Generalised Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3), 307-327.
- Bollerslev T., Chou R. Y., Kroner K. F. (1992) ARCH Modeling in Finance: A Selective Review of the Theory and Empirical Evidence. *Journal of Econometrics*, 52, 5-59.
- Brzeszczyński J., Kelm R. (2002) *Ekonometryczne modele rynków finansowych*, WIG-Press, Warszawa.
- Ding Z., Granger C. W. J., Engle R. (1993) A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, 1, 83-106.
- Ding Z., Granger C. W. J. (1996) Modeling Volatility Persistence of Speculative Returns: A New Approach. *Journal of Econometrics*, 73, 185-215.
- Doman M., Doman R. (2009) *Modelowanie zmienności i ryzyka*, Oficyna Wolters Kluwer, Kraków.
- Engle R. F., Lilien D. M., Robins R. P. (1987) Estimating Time Varying Risk Premia in the Term Structure: the ARCH-M Model. *Econometrica*, 55(2), 391-407.
- Fiszeder P. (2009) *Modele klasy GARCH w empirycznych badaniach finansowych*. Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń.
- Gallant A., Rossi P., Tauchen G. (1992) Stock Prices and Volume. *The Review of Financial Studies*, 5(2), 199-242.
- Nelson D. (1990) Stationarity and Persistence in GARCH(1, 1) Model. *Econometric Theory*, 6(3), 318-334.
- Osiewalski J., Pipień M. (1999) Bayesowskie wnioskowanie o stacjonarności procesów GARCH(1, 1). *Dynamiczne Modele Ekonometryczne*, Materiały na VI Ogólnopolskie Seminarium Naukowe, UMK, Toruń.
- Pajor A. (2010) *Wielowymiarowe procesy wariancji stochastycznej w ekonometrii finansowej. Ujęcie bayesowskie*. Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie.
- Pipień M. (2006) *Wnioskowanie bayesowskie w ekonometrii finansowej*. Zeszyty Naukowe, Akademia Ekonomiczna w Krakowie. Seria Specjalna. Monografie, 176.

**THE ADAPTATION OF GARCH MODELS  
AND THE QUALITY OF OBTAINED FORECASTS**

**Abstract:** Analysis of the adaptation of GARCH models presented in the paper consists of three parts. Firstly, the theoretical considerations in this regard was made, referring to the recommendations identified in the literature. Secondly, practical aspects of the proper selection of GARCH model for the time series based on the values of information criteria was presented. Thirdly, the relation between the quality of adaptation and the quality of obtained forecasts was indicated, using the tools of quantitative analysis. The aim of this paper is to verify if the improvement of the goodness of GARCH models results in better volatility forecasts.

**Keywords:** forecasting, GARCH models, time series, adaptation

**JEL classification:** C32, C53, C58, G10, G17



## IDENTYFIKACJA POWIATÓW O NAJWIĘKSZYM RYZYKU ROZPRZESTRZENIANIA SIĘ CHOROBY COVID-19 I OBCIĄŻENIA SŁUŻBY ZDROWIA W POLSCE

Lukasz Pietrych  <https://orcid.org/0000-0001-5053-3890>

Instytut Ekonomii i Finansów

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

e-mail: lukasz\_pietrych@sggw.edu.pl

**Streszczenie:** Celem artykułu była identyfikacja powiatów charakteryzujących się najwyższym ryzykiem rozprzestrzeniania choroby COVID-19 oraz w konsekwencji obciążenia dla służby zdrowia. W badaniu oszacowano syntetyczny wskaźnik Hellwiga, który pozwolił na uporządkowanie regionów oraz wyszczególnienie tych z największym ryzykiem nadmiernego rozprzestrzeniania się epidemii. W artykule uwzględniono zestaw cech prostych charakteryzujących dany powiat, szczególnie istotna okazała się zmienna określająca natężenie ruchu turystycznego. Ponadto wyniki wskazują na nierówności pomiędzy obszarami wiejskimi a miejskimi.

**Słowa kluczowe:** wskaźnik syntetyczny Hellwiga, COVID-19, powiaty, zmienne społeczno-demograficzne

**JEL classification:** G11, G17

### WSTĘP

Koronawirusa SARS-CoV-2 został w Polsce po raz pierwszy zidentyfikowany 4 marca 2020 r. Choroba wywołana przez ten wirus określana jako COVID-2019, charakteryzuje się wysokim stopniem zaraźliwości. Szacowany do tej pory współczynnik reprodukcyjnej  $R_0$  waha się w różnych krajach na świecie w przedziale od 1,4 do 6,49 [Liu et al. 2020; Shim et al. 2020; Kuniya 2020; Prem et al. 2020; Huang et al. 2020; Wangping; Coronavirus disease 2019]. Jednocześnie szacowany ogólny współczynnik śmiertelności dla tej choroby oszacowano na poziomie 1,4% [Wu et al. 2020], przy czym wskaźnik ten jest znacząco zróżnicowany w poszczególnych grupach wiekowych. Od początku marca 2020 r.

<https://doi.org/10.22630/MIBE.2020.21.3.13>

patogen rozprzestrzenił się w całym kraju i zgodnie ze stanem na grudzień 2020 r. skumulowana zapadalność na COVID-19 w Polsce była jedną z najwyższych w Europie i obejmuje ponad 1,2 mln potwierdzonych przypadków.

W związku z powyższym, aby ograniczyć ryzyko pojawiania się lokalnych ognisk epidemii ważna jest identyfikacja powiatów o dużym potencjale rozprzestrzeniania się tej choroby. Wpływ na to może mieć wiele czynników geograficznych, demograficznych a nawet społeczno-gospodarczych, które mogą znacznie sprzyjać pojawianiu się nowym ognisk a w konsekwencji spowodować niekontrolowany wzrost liczby zarażonych. Oprócz czynników związanych z gęstością zaludnienia, wydarzenia z 2020 r. pokazują, iż czynnikiem niosącym ryzyko niekontrolowanego wybuchu pandemii może być przyjęcie rodzinne (wesela). Wzmożenie ruchu turystycznego w okresie przerw od nauki, może również generować dodatkowe ryzyko. Celem artykułu było opracowanie rankingu powiatów pod względem ryzyka rozprzestrzeniania się choroby COVID-19 oraz nadmiernego obciążenia dla służby zdrowia.

Oprócz identyfikacji powiatów o dużym potencjale wybuchu epidemii ze względu na każdą z wymienionych zmiennych osobno zastosowano wskaźnik syntetyczny Hellwiga w celu wyselekcjonowania powiatów, w których łącznie wszystkie czynniki mogą sprzyjać rozwojowi epidemii. Oznacza to, że powiaty z najwyższą wartością wskaźnika mogą wymagać ciągłego monitorowania w celu niedopuszczenia do niekontrolowanego wybuchu epidemii.

## MATERIAŁY ŹRÓDŁOWE I METODY BADAWCZE

W badaniu wykorzystano dane zgromadzone w Banku Danych Lokalnych, który jest największą w Polsce bazą danych o gospodarce, społeczeństwie i środowisku [Bank Danych Lokalnych – GUS]. Wykorzystano następujące cechy proste: gęstość zaludnienia, udział ludności w wieku 60 lat i więcej w populacji ogółem, podmioty skupiające w swojej działalności duże grupy osób (przedsiębiorstwa zatrudniających co najmniej 250 osób, liczba szkół podstawowych i ponadpodstawowych oraz liczba domów pomocy społecznej), liczba turystów korzystających z noclegów na 1 tys. ludności oraz liczba zawartych małżeństw w danym powiecie w 2019 roku w przeliczeniu na 1 tys. mieszkańców.

W literaturze przedmiotu zaproponowano szereg metod porządkowania liniowego, celem tego artykułu nie było jednak porównywanie poszczególnych metod. W związku z powyższym, do badania wybrano jedną z pierwszych propozycji porządkowania liniowego przedstawioną przez Hellwiga [1968]. Należy jednak pamiętać, że oprócz licznych zastosowań metody, pojawiły się też uwagi krytyczne związane z ograniczeniami zastosowania tej metody. Przykładowo, występowanie obserwacji oddalonych cech prostych, ma istotny wpływ na zmniejszenie zakresu zmienności wartości konstruowanej cechy syntetycznej

i w rezultacie może powodować problemy dotyczące identyfikacji poziomów rozwoju zjawiska [Głowicka-Wołoszyn, Wysocki 2018].

Metody wielowymiarowej analizy statystycznej, a w tym metody porządkowania liniowego, są często stosowanymi narzędziami umożliwiającymi porównywanie obiektów niejednorodnych, których rozwój warunkowany jest różnymi zmiennymi, a literatura przedmiotu jest dość bogata (zob. [Łuniewska, Tarczyński 2006; Młodak 2006; Kompa 2014]). Miernik syntetyczny wzorcowej metody Hellwiga wyznaczany jest jako [Nowak 1990]:

$$\mu_i = 1 - \frac{d_{i0}}{d_0} \quad (i = 1, 2, \dots, n), \quad (1)$$

gdzie

$$d_{i0} = \left[ \sum_{j=1}^m (z_{ij} - z_{0j})^2 \right]^{0.5} \quad (2)$$

jest odległością każdego obiektu  $z_{ij}^1$  od obiektu wzorca  $z_{0j}$  wyznaczonego jako wartość maksymalna dla cech będących stymulantami:

$$d_0 = \bar{d}_0 + 2s(d_0), \quad (3)$$

przy czym

$$\bar{d}_0 = \frac{1}{n} \sum_{i=1}^n d_{i0}, \quad (4)$$

zaś

$$s(d_0) = \left[ \frac{1}{n} \sum_{i=1}^n (d_{i0} - \bar{d}_0)^2 \right]^{0.5}. \quad (5)$$

## WYNIKI BADAŃ

### Charakterystyka społeczno-demograficzna powiatów

Do pomiaru gęstości zaludnienia w poszczególnych powiatach wykorzystano cechę liczba ludności/km<sup>2</sup>. Na obszarach o dość dużym zagęszczeniu ludności pojawienie się osoby zarażonej może prowadzić do dość szybkiej transmisji wirusa. Do powiatów z najwyższą gęstością zaludnienia (z wyłączeniem miast na prawach powiatu) należy zaliczyć powiaty zlokalizowane w centralnej i południowej części kraju, a w szczególności powiaty:

- wodzisławski, mikołowski, będziński, oświęcimski, bieruńsko-lędziński, bielski, rybnicki chrzanowski, wielicki oraz legionowski.

<sup>1</sup>  $z_{ij}$  – to zmienna poddana normalizacji zgodnie ze wzorem  $z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}$ .

Biorąc pod uwagę ten czynnik powiaty (poza kilkoma wyjątkami) w województwach podlaskim, warmińsko-mazurskim oraz lubuskim cechują najmniejszym ryzykiem (rys. 1).

W celu identyfikacji powiatów z niekorzystną strukturą wiekową wykorzystano cechę opisującą udział ludności w wieku 60 i więcej lat. Powiaty z najbardziej niekorzystną strukturą wiekową zlokalizowane są w województwach świętokrzyskim, łódzkim i podlaskim. W szczególności są to powiaty (bez miast na prawach powiatów):

- hajnowski, skarżyski, ostrowiecki, krasnostawski, będziniński, kłodzki, zawierciański, dzierzoniowski, bielski oraz pińczowski.

Powiaty o niesprzyjającej rozwoju epidemii strukturze wiekowej znajdują się prawie w całym województwie pomorskim oraz małopolskim (rys. 2).

Do identyfikacji powiatów, w których zlokalizowanych jest najwięcej podmiotów skupiających duże grupy osób (najczęściej w jednym budynku) wykorzystano cechy opisujące liczbę podmiotów zatrudniających powyżej 250 osób, liczbę szkół podstawowych oraz ponadpodstawowych<sup>2</sup> oraz liczbę domów pomocy społecznej (DPS-ów)<sup>3</sup>. Powiaty, w których sumarycznie zlokalizowanych jest najwięcej tego typu podmiotów znajdują się głównie w województwach małopolskim i śląskim. W szczególności są to:

- powiat krakowski, poznański, nowotarski, tarnowski, nowosądecki, kielecki, rzeszowski, cieszyński, kłodzki oraz żywiecki.

W tym przypadku nie da się wyróżnić żadnego województwa, w których powiaty mogą nie być zbyt narażone działaniem tego czynnika (rys. 3).

Do zmierzenia natężenia ruchu turystycznego w poszczególnych powiatach wykorzystano cechę prezentującą liczbę turystów korzystających z noclegów na 1 tys. ludności. Biorąc pod uwagę tę cechę należy stwierdzić, że do grupy powiatów obarczonych największym ryzykiem rozprzestrzeniania się choroby należy zaliczyć powiaty turystyczne zlokalizowane na wybrzeżu oraz południowych krańcach kraju, a w szczególności powiaty:

- tatrzański, jeleniogórski, kołobrzeski, leski, kamieński, bieszczadzki, mrągowski, gryficki, nowodworski oraz pucki.

---

<sup>2</sup> Nie uwzględniono uczelni wyższych, ponieważ są one zlokalizowane głównie w dużych ośrodkach miejskich. Do największych „miasteczek akademickich” należy zaliczyć miasteczka w Krakowie oraz Lublinie.

<sup>3</sup> Samych DPS-ów najwięcej zlokalizowanych jest w powiatach (z wyłączeniem miast na prawach powiatu): cieszyńskim, krakowskim, tarnowskim, kłodzkim, tczewskim oraz myśliborskim.

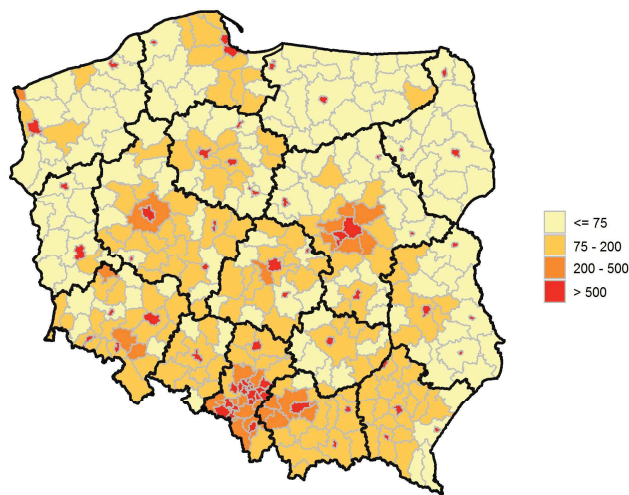
Należy zwrócić uwagę, że powiaty te mogą stanowić miejsca zakażeń oraz rozprzestrzeniania się choroby w głąb kraju tym bardziej, że najbliższe tygodnie to sezon urlopowy (rys. 4).

Informacje o ostatnich ogniskach zarażenia, które wystąpiły po zezwoleniu na organizację przyjęć rodzinnych, wskazują na kolejny czynnik zarażeń, jakim są przyjęcia weselne. W celu identyfikacji powiatów o możliwie dużym natężeniu takich przyjęć (przy założeniu, że liczba zawieranych małżeństw jest stała w czasie), wykorzystano cechę odzwierciedlającą liczbę zawartych małżeństw w danym powiecie w przeliczeniu na 1 tys. mieszkańców w 2019 roku. Biorąc pod uwagę wyżej wymieniony czynnik należy stwierdzić, że do grupy powiatów obarczonych największym ryzykiem rozprzestrzeniania się choroby należy zaliczyć następujące powiaty:

- limanowski, kartuski, rycki, bocheński, lubaczowski, kołobrzegi, nowosądecki, suski, łukowski oraz siedlecki.

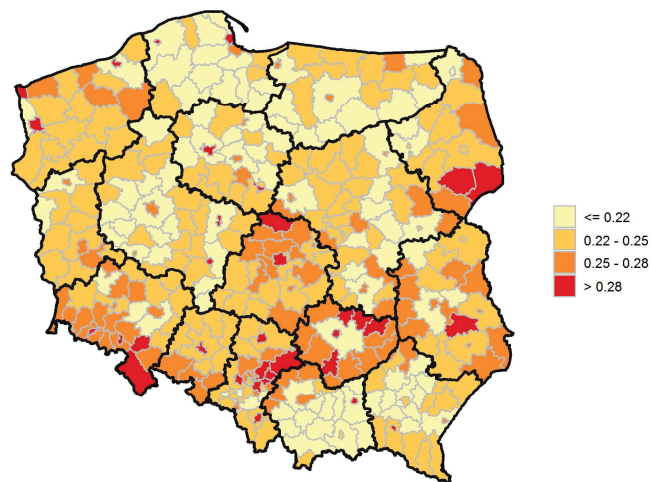
Przy założeniu, iż liczba zawieranych małżeństw nie zmienia się w kolejnych latach powiaty, w których notowana może być zwiększona liczba przyjęć weselnych znajdują się m.in. w województwie małopolskim oraz wielkopolskim (rys. 5).

Rysunek 1. Gęstość zaludnienia w układzie powiatów w Polsce w 2019 roku



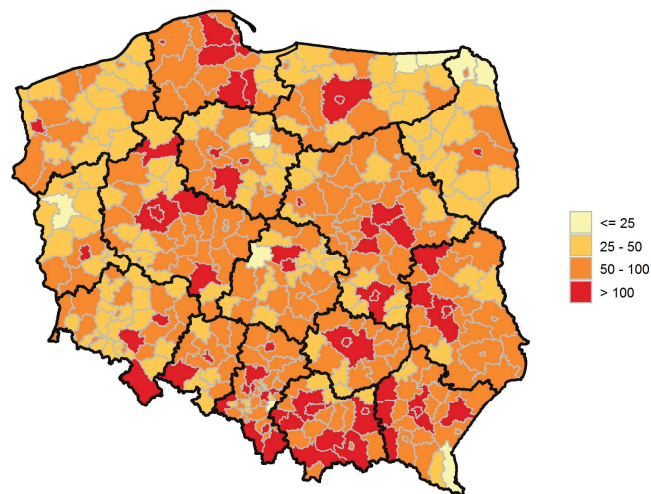
Źródło: opracowanie własne na podstawie danych BDL GUS [dostęp: 20.11.2020]

Rysunek 2. Udział ludności w wieku 60 lat i więcej w ludności ogółem w układzie powiatów w 2018 r.



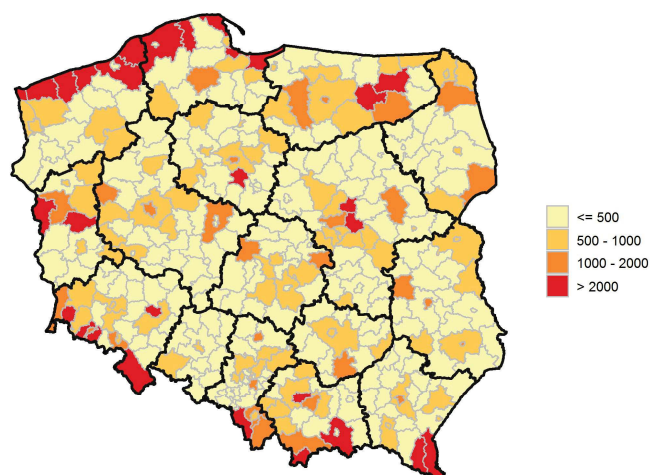
Źródło: opracowanie własne na podstawie danych BDL GUS [dostęp: 20.11.2020]

Rysunek 3. Powiaty z największą liczbą potencjalnych skupisk ludności w układzie powiatów w 2019 r.



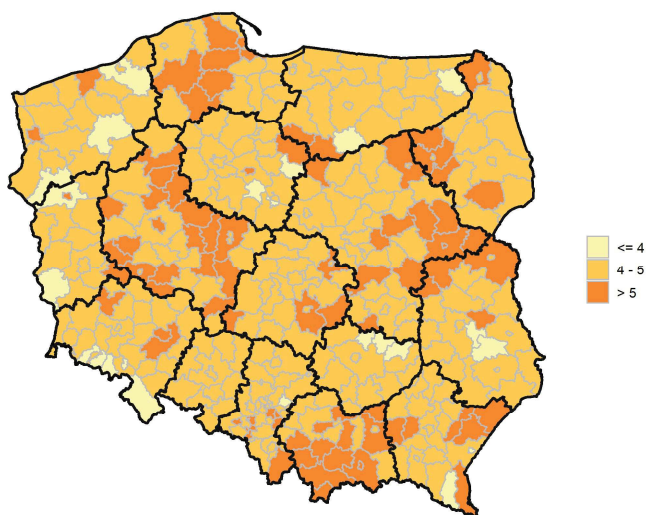
Źródło: opracowanie własne na podstawie danych BDL GUS [dostęp: 20.11.2020]

Rysunek 4. Liczba turystów korzystających z noclegów na 1 tys. ludności w układzie powiatów 2018 r.



Źródło: opracowanie własne na podstawie danych BDL GUS [dostęp: 20.11.2020]

Rysunek 5. Rozkład liczby zawartych małżeństw w przeliczeniu na 1 tys. mieszkańców w układzie powiatów w 2019 r.



Źródło: opracowanie własne na podstawie danych BDL GUS [dostęp: 20.11.2020]

### Syntetyczna ocena poziomu ryzyka rozprzestrzeniania się choroby COVID-19 w układzie powiatów

Wykorzystując wyżej wspomniane cechy do próby identyfikacji powiatów o szczególnych cechach, które mogą sprzyjać powstawaniu ognisk zarażenia, dokonano ich agregacji i wyznaczono syntetyczny wskaźnik Hellwiga. Wskaźnik ten ma na celu wyznaczyć miarę, która pozwoli porównać wszystkie powiaty między sobą biorąc pod uwagę wszystkie czynniki łącznie. Powiaty z najwyższym wskaźnikiem należy zaliczyć do grupy powiatów o dużym ryzyku rozprzestrzeniania się choroby, są to w szczególności:

- duże miasta wojewódzkie oraz miasta na prawach powiatu,
- powiaty z największym ryzykiem dominują w województwach: małopolskim, śląskim oraz łódzkim,
- po wyłączeniu miast na prawach powiatów, powiaty z największym ryzykiem to m.in.: powiat kołobrzeski, tatrzański, jeleniogórski, cieszyński, kamieński, kłodzki, bieszczadzki, mrągowski, leski oraz gryficki.

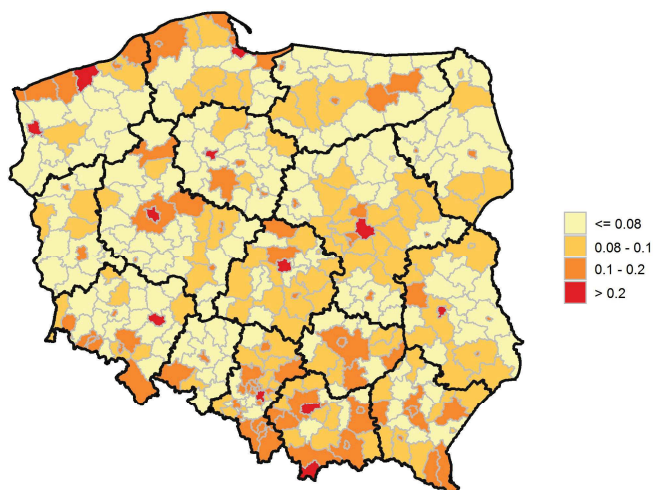
Tabela 1. Cechy charakterystyczne wybranych powiatów z największym ryzykiem rozprzestrzeniania się choroby

Powiat	Województwo	Cecha charakterystyczna
powiat kołobrzeski	zachodniopomorskie	wysokie natężenie ruchu turystycznego, wysoka względna liczba małżeństw oraz wysoki udział ludności w wieku 60+
powiat tatrzański	małopolskie	wysokie natężenie ruchu turystycznego
powiat jeleniogórski	dolnośląskie	wysokie natężenie ruchu turystycznego oraz wysoki udział ludności w wieku 60+
powiat cieszyński	śląskie	duża liczba potencjalnych skupisk ludności, wysokie natężenie ruchu turystycznego oraz wysoka gęstość zaludnienia
powiat kamieński	zachodniopomorskie	wysokie natężenie ruchu turystycznego oraz wysoki udział ludności w wieku 60+
powiat kłodzki	dolnośląskie	duża liczba potencjalnych skupisk ludności, wysokie natężenie ruchu turystycznego oraz wysoki udział ludności w wieku 60+
powiat bieszczadzki	podkarpackie	wysokie natężenie ruchu turystycznego
powiat mrągowski	warmińsko-mazurskie	wysokie natężenie ruchu turystycznego
powiat leski	podkarpackie	wysokie natężenie ruchu turystycznego
powiat gryficki	zachodniopomorskie	wysokie natężenie ruchu turystycznego

Źródło: opracowanie własne



Rysunek 6. Powiaty z największym ryzykiem rozprzestrzeniania się choroby (najwyższy wskaźnik)



Granice klas topologicznych stanowią kwartyle.

Źródło: opracowanie własne

## PODSUMOWANIE

Ponieważ epidemia może obciążyć systemy opieki zdrowotnej, identyfikacja regionów, w których spodziewane obciążenie chorobami może być wysokie w porównaniu z resztą kraju, ma kluczowe znaczenie dla umożliwienia optymalnej dystrybucji zasobów opieki medycznej. W badaniu zidentyfikowano kilka regionów szczególnie narażonych na nadmierne ryzyko rozprzestrzeniania się choroby (biorąc od uwagę gęstość zaludnienia, strukturę wiekową, potencjalne skupiska ludności, natężenie ruchu turystycznego oraz liczbę zawieranych małżeństw w roku poprzednim). Do tej grupy należy zaliczyć powiaty: kołobrzegi, tatrzański, jeleniogórski, cieszyński, kamieński, kłodzki, bieszczadzki, mrągowski, leski oraz gryficki. Taki stan zdeterminowany jest głównie wysokim natężeniem ruchu turystycznego. Ponadto wyniki wskazują na nierówności pomiędzy obszarami wiejskimi a miejskimi. Przetawione badania w kolejnym kroku powinny być uzupełnione o przestrzenną analizę dostępności do opieki medycznej, co umożliwi identyfikację regionów z największym ryzykiem i obciążeniem dla opieki medycznej choroba COVID-19.

## BIBLIOGRAFIA

- European Centre for Disease Prevention and Control (2020) Coronavirus Disease 2019 (COVID-19) Pandemic: Increased Transmission in the EU/EEA and the UK – Seventh Update, 25, Stockholm: ECDC.
- Głowicka-Wołoszyn R., Wysocki F. (2018) Problem identyfikacji poziomów rozwoju w zagadnieniu konstrukcji cechy syntetycznej. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 508, 56-65.
- Hellwig, Z. (1968) Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę wykwalifikowanych kadr. *Przegląd Statystyczny*, 4, 307-327.
- Huang Y., Yang L., Dai H., Tian F., Chen K. (2020) Epidemic Situation and Forecasting of COVID-19 in and outside China. [Preprint]. *Bull World Health Organ.* doi: 10.2471/BLT.20.255158.
- Kompa K. (2014) Zastosowanie mierników taksonomicznych do oceny efektywności finansowej giełd europejskich w latach 2002-2011. *Metody Ilościowe w Badaniach Ekonomicznych* 15(4), 52-61.
- Kuniya T. (2020) Prediction of the Epidemic Peak of Coronavirus Disease in Japan, 2020. *J Clin Med.*, 9(3):789. doi: 10.3390/jcm9030789.
- Liu Y., Gayle A., Wilder-Smith A., Rocklöv J. (2020) The Reproductive Number of COVID-19 is Higher Compared to SARS Coronavirus. *Journal of Travel Medicine*, 27(2), 1-7. doi: 10.1093/jtm/taaa021.
- Luniewska M., Tarczyński W. (2006) *Metody wielowymiarowej analizy porównawczej na rynku kapitałowym*, Warszawa: Wydawnictwo Naukowe PWN.
- Młodak A. (2006) *Analiza taksonomiczna w statystyce regionalnej*, Warszawa: Difin.
- Nowak E. (1990) *Metody taksonomiczne w klasyfikacji obiektów społeczno-gospodarczych*. Państwowe Wydawnictwo Ekonomiczne, Warszawa.
- Prem K., Liu Y., Russell T. W., Kucharski A. J., Eggo R. M., Davies N., Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Jit M., Klepac P. (2020) The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health*, 5(5) 261-270. doi: 10.1016/S2468-2667(20)30073-6.
- Shim E., Tariq A., Choi W., Lee Y., Chowell G. (2020) Transmission Potential and Severity of COVID-19 in South Korea. *International Journal of Infectious Diseases*, 93, 339-344. doi: 10.1016/j.ijid.2020.03.031.
- Wangping J., Ke H., Yang S., Wenzhe C., Shengshu W., Shanshan Y., Jianwei W., Fuyin K., Penggang T., Jing L., Miao L., Yao H. (2020) Extended SIR Prediction of the Epidemics Trend of COVID-19 in Italy and Compared With Hunan, China. *Front Med (Lausanne)*, 6, 7:169. doi: 10.3389/fmed.2020.00169.
- Wu J.T., Leung K., Bushman M. et al. (2020) Estimating Clinical Severity of COVID-19 from the Transmission Dynamics in Wuhan, China. *Nat Med* 26, 506-510. doi: 10.1038/s41591-020-0822-7.

---

**IDENTIFICATION OF COUNTIES WITH THE HIGHEST RISK  
OF SPREADING COVID-19 DISEASE IN POLAND**

**Abstract:** The aim of the article was to identify counties with the highest risk of spreading the COVID-19 disease and, consequently, burdening the health service. The study estimated the synthetic Hellwig index, which allowed for the ordering of the regions and for identifying those with the highest risk of over-spreading epidemics. The article takes into account a set of simple features characterizing a given county, the variable determining the intensity of tourist traffic turned out to be particularly important. Moreover, the results show inequalities between rural and urban areas.

**Keywords:** Hellwig's synthetic index, COVID-19, counties, socio-demographic variables

**JEL classification:** G11, G17

## PRICING EUROPEAN OPTIONS IN SELECTED STOCHASTIC VOLATILITY MODELS

Arkadiusz Orzechowski<sup>1</sup>  <https://orcid.org/0000-0003-2872-189X>

Institute of Risk and Financial Markets  
Warsaw School of Economics – SGH, Poland  
e-mail: aorzec@sgh.waw.pl

**Abstract:** In this paper four methods of calculating characteristic functions and their application to selected stochastic volatility models are considered. The methods applied are based on the assumption that the prices of European calls are evaluated numerically by means of the Gauss-Kronrod quadrature. Such approach is used to investigate computational efficiency of pricing European calls. Particular attention in this matter is paid to the speed of generating theoretical prices of the analyzed contracts.

**Keywords:** option pricing, the Heston model, the Bates model, characteristic functions

**JEL classification:** C02, G13

### INTRODUCTION

The first model of pricing European options in continuous time was introduced by Black & Scholes [1973]. Although the model is still widely used by many practitioners its structure has long ceased to meet the requirements of modern financial markets. Such view should be considered correct for at least two reasons.

Firstly, there is enough evidence suggesting the existence of so-called stylized facts. According to Cont [2001] this term should be associated with seemingly random variations of asset prices that have non-trivial statistical properties. A different definition of stylized facts is proposed by Challet et al. [2001], who

---

<sup>1</sup> The views expressed in the article by the author are the personal views of the author and do not express the official position of the institution in which he is employed.

identify the above term with empirical statistical regularities in financial data (which are not explained in terms of cause and effect - ed. by the author). Taylor [2005] suggests a similar approach stating that stylized facts are general properties expected to be present in any set of returns. Rogers & Zhang [2011] present a different opinion about it. According to them, stylized facts are a set of independently recognized characteristics relating to various instruments, markets and periods of time. According to R. Cont [2001], stylized facts are properties that are common across a wide range of instruments, markets and periods of time.

The following stylized facts are most often analyzed in the financial literature:

- absence of autocorrelation, except for very short time intervals where microstructural effects start to play a role,
- (conditional) heavy tails in the (conditional) distribution of returns,
- gain/loss asymmetry (prices of financial assets and stock index values are slower to go up and faster to go down; the opposite relationship is observed for exchange rates),
- aggregational gaussianity (as one increases the time interval over which returns are calculated, the distribution of returns converges to a normal one),
- intermittency of returns and volatility clustering (time series of returns are irregular and volatility shows tendency to cluster over time),
- slow decay of autocorrelation in absolute returns,
- leverage effect (there is a negative relationship between stock returns and both historical and implied volatility),
- volume/volatility correlation (trading volume and volatility are negatively correlated),
- asymmetry in time scales (coarse-grained measures of volatility predict a fine-scale volatility better than vice versa).

It is worth noting that stylized facts are not tantamount to market anomalies. Stylized facts refer to the immanent properties of the financial market forming the foundation for building some scientific theories. The market anomalies are regarded as inexplicable phenomena contradicting the existing concepts and views. Such statements seem to be in line with the opinion offered by T. Lux [2009]. It is important to notice that although microstructural effects, heavy tails in the distributions of returns, volatility clustering etc. are treated as immanent properties of the variables which do not fit within the restrictive framework of some existing models, e.g. the Black-Scholes model, they are sometimes captured by other models, e.g. stochastic volatility models.

Secondly, as a result of the constant technological progress, the computing power of computers is systematically increasing. It means that the range of computational techniques that can be used in practice for the valuation of capital assets or derivatives is expanding. The mathematical tools that have gained particular importance in this context are the characteristic functions. They can be applied to many option pricing models, including the stochastic volatility models.

The aim of the article is to show that the process of pricing European options in selected stochastic volatility models can be improved in terms of computational speed. The article is organized as follows. In the first section two stochastic volatility models of pricing European options are investigated, i.e. the Heston model [Heston 1991] and the Bates model [Bates 1996]. In the second section, some approaches to calculating characteristic functions are presented. In the third section, the computational speed of pricing European options is investigated. Finally, in the fourth section, this article has been summarized and some major conclusions have been drawn.

## SELECTED STOCHASTIC VOLATILITY MODELS

In this section, two stochastic volatility models of pricing European options are analyzed: the Heston model and the Bates model.

### The Heston model

In the Heston model, the dynamics of the underlying asset price  $S$  and the volatility process  $\sigma^2$  are governed by two stochastic differential equations:

$$dS_t = \mu S_t dt + \sqrt{\sigma_t^2} S_t dW_{1,t}. \quad (1)$$

$$d\sigma_t^2 = \kappa(\theta - \sigma_t^2)dt + v\sqrt{\sigma_t^2}dW_{2,t}. \quad (2)$$

where:  $S_t$  denotes the spot price of the underlying asset at time  $t$ ,  $\mu$  is the constant (drift),  $\sigma_t^2$  is the instantaneous variance,  $\theta$  is the long-term variance,  $\kappa$  is the mean-reversion rate, and  $v$  is the volatility of the variance process. It is worth noting that under Feller's condition, i.e.  $2\kappa\theta \geq v^2$ , the variance process remains always positive. The Brownian motions  $W_1$  and  $W_2$  are correlated with a constant  $\rho$ .

The formula for the price of a European call in the Heston model takes the following form:

$$C^H(s_t, \sigma_t^2, t) = S_t P_1^H(s_t, \sigma_t^2, \tau) - e^{-r\tau} K P_2^H(s_t, \sigma_t^2, \tau). \quad (3)$$

where:  $\tau = T - t$ ,  $r$  is the risk-free rate,  $K$  is the exercise price,  $P_1^H(s_t, \sigma_t^2, \tau)$  and  $P_2^H(s_t, \sigma_t^2, \tau)$  are the probabilities of expiring European call in-the-money.

Although  $P_1^H(s_t, \sigma_t^2, \tau)$  and  $P_2^H(s_t, \sigma_t^2, \tau)$  are not known they can be easily extracted from the characteristic functions, for  $j = 1, 2$ , i.e.:

$$P_j^H(s_t, \sigma_t^2, \tau) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re \left( \frac{e^{-\mathbb{I}\xi \ln K} \phi^{j,H}(\xi, s_t, \sigma_t^2)}{\mathbb{I}\xi} \right) d\xi. \quad (4)$$

where:  $\Re(\cdot)$  is the real part of the subintegral function,  $\mathbb{I}$  is the imaginary unit of the complex number,  $\phi^{j,H}(\xi, s_t, \sigma_t^2)$  is the characteristic function of  $s_t = \ln S_t$  (corresponding to  $P_j^H(s_t, \sigma_t^2, \tau)$ ). The remaining notation is the same as previously introduced.

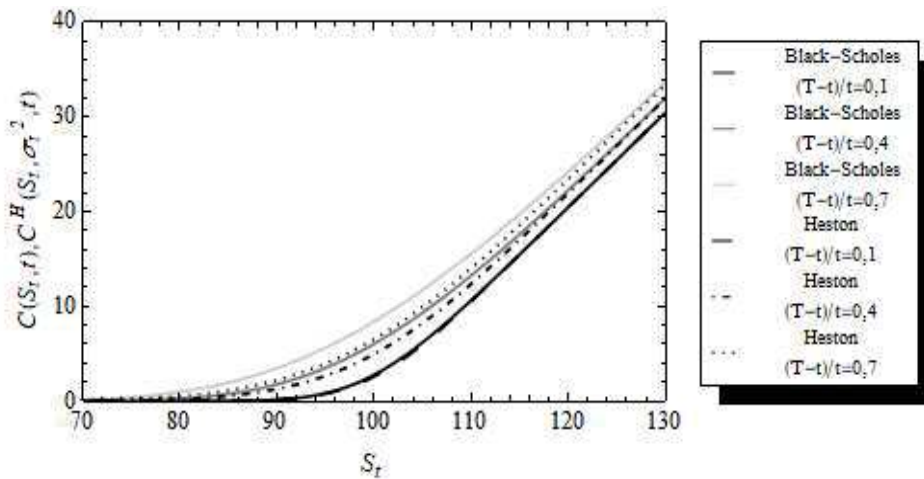
Theoretical price of a European call can be obtained under the assumption that the general form of the characteristic functions of  $s_t$  (corresponding to probabilities  $P_j^H$ , for  $j = 1, 2$ ), is the following:

$$\phi^{j,H}(\xi, s_t, \sigma_t^2) = e^{C_j(\xi, \tau) + D_j(\xi, \tau) \sigma_t^2 + \mathbb{I} \xi s_t}. \quad (5)$$

where:  $C_j(\xi, \tau) = r \mathbb{I} \xi \tau + \frac{a}{v^2} \left[ (b_j - v \rho \mathbb{I} \xi + d_j) \tau - 2 \ln \left( \frac{1 - g_j e^{d_j \tau}}{1 - g_j} \right) \right]$ ,  $D_j(\xi, \tau) = \frac{b_j - v \rho \mathbb{I} \xi + d_j}{v^2} \left( \frac{1 - e^{d_j \tau}}{1 - g_j e^{d_j \tau}} \right)$ ,  $u_1 = \frac{1}{2}$ ,  $u_2 = -\frac{1}{2}$ ,  $a = \kappa \theta$ ,  $b_1 = \kappa + \lambda - v \rho$ ,  $b_2 = \kappa + \lambda$ ,  $g_j = \frac{b_j - v \rho \mathbb{I} \xi + d_j}{b_j - v \rho \mathbb{I} \xi - d_j}$ , zaś  $d_j = \sqrt{(v \rho \mathbb{I} \xi - b_j)^2 - v^2 (2u_j \mathbb{I} \xi - \xi^2)}$ .

The figure below presents the payoff functions of a European call in the Heston model ( $C^H(S_t, \sigma_t^2, t)$ ) compared to the payoff functions of a European call in the Black-Scholes model ( $C(S_t, t)$ ) assuming that:  $S_t \in [70, 130]$ ,  $K = 100$ ,  $\sigma_t = 0.2$ ,  $r = 5\%$ ,  $v = 0.04$ ,  $\kappa = 1.5$ ,  $\lambda = 3$ ,  $\theta = 0.04$ ,  $\rho = 0.8$  for different periods remaining to expiration, i.e.  $\frac{T-t}{T} \in \{0.1; 0.4; 0.7\}$ .

Figure 1. Payoff functions of a European call in the Heston model and the Black-Scholes model assuming that:  $S_t \in [70, 130]$ ,  $K = 100$ ,  $\sigma_t = 0.2$ ,  $r = 5\%$ ,  $\frac{T-t}{T} \in \{0.1; 0.4; 0.7\}$ ,  $v = 0.04$ ,  $\kappa = 1.5$ ,  $\lambda = 3$ ,  $\theta = 0.04$ ,  $\rho = 0.8$



Source: own preparation

It is worth noting that pricing of options using the method described above is inefficient. In eq. 3-4 there are two integrals and each of them has to be evaluated numerically. It means that the computational effort necessary for the valuation of options is huge. A straightforward solution to the problem lies in the application of

alternative methods in which only one characteristic function is used. This issue will be discussed in next section of this article.

### The Bates model

In the Bates model [Bates 1996], the underlying asset price process is determined by two equations, i.e.:

$$dS_t = (\mu - \lambda\mu_J)S_t dt + \sqrt{\sigma_t^2}S_t dW_{1,t} + JS_t dN_t. \quad (6)$$

$$d\sigma_t^2 = \kappa(\theta - \sigma_t^2)dt + v\sqrt{\sigma_t^2}dW_{2,t}. \quad (7)$$

where:  $S_t$  denotes the spot price of the underlying asset at time  $t$ ,  $\mu$  is the instantaneous expected rate of appreciation of the underlying asset,  $\lambda$  is the annual frequency of jumps,  $J$  is the random percentage jump conditional on a jump occurring,  $N_t$  is the Poisson counter with intensity  $\lambda$ ,  $\sigma_t^2$  is the instantaneous variance of the price process. It is worth noting that:  $1 + J \sim \mathcal{LN}(\mu_S, \sigma_S^2)$ . Moreover, the Brownian motions  $W_1$  and  $W_2$  are correlated with a constant  $\rho$  and the relationship between  $\mu_J$  and  $\mu_S$  is the following:  $\mu_J = e^{\left(\mu_S + \frac{\sigma_S^2}{2}\right)} - 1$ .

The price of a European call in the Bates model can be evaluated in the same manner as in the Heston model. Thus, the theoretical framework discussed herein is simple on the one hand, while, on the other hand, the drawbacks of the applied approach concerning computational inefficiency remain the same.

Similarly to the Heston model, the methodology of pricing options in the Bates model is based on characteristic functions. For the purpose of further analysis it is assumed that the characteristic function needed for the valuation of European options is as follows:

$$\phi^{2,B}(\xi, S_t, \sigma_t^2) = e^{C_2(\xi, \tau) + D_2(\xi, \tau)\sigma_t^2 + \mathcal{P}(\xi)\lambda\tau + \mathbb{I}\xi S_t} \quad (8)$$

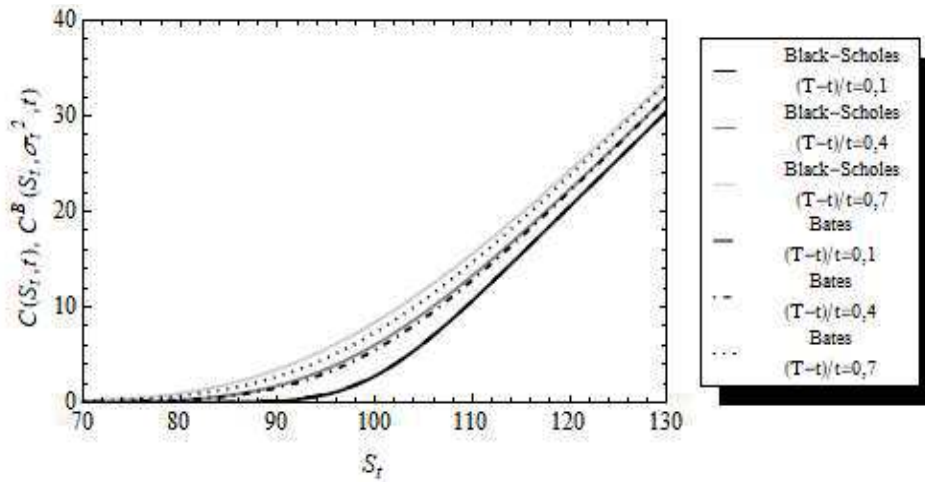
where:  $\mathcal{P}(\xi) = -\mu_J \mathbb{I}\xi + \left( (1 + \mu_J)^{\mathbb{I}\xi} e^{\sigma_S^2 \left(\frac{\mathbb{I}\xi}{2}\right) (\mathbb{I}\xi - 1)} - 1 \right)$ , and both  $C_2(\xi, \tau)$  and  $D_2(\xi, \tau)$  can be concluded from the Heston model. The remaining notation is the same as in the previous case.

The formulas that can be used for pricing European options are presented in the next section of this article.

The figure below presents the payoff functions of a European call in the Bates model ( $C^B(S_t, \sigma_t^2, t)$ ) compared to the payoff functions of a European call in the Black-Scholes model ( $C(S_t, t)$ ) assuming that:  $S_t \in [70, 130]$ ,  $K = 100$ ,  $\sigma_t = 0.2$ ,  $r = 5\%$ . Additionally, it is assumed that:  $v = 0.04$ ,  $\kappa = 1.5$ ,  $\lambda = 3$ ,  $\theta = 0.04$ ,  $\rho = 0.8$ ,  $\mu_S = -0.05$ , and  $\sigma_S = 0.00004$  for different periods remaining to expiration, i.e.  $\frac{T-t}{T} \in \{0.1; 0.4; 0.7\}$ .



Figure 2. Payoff functions of a European call in the Bates and the Black-Scholes model assuming that:  $S_t \in [70, 130]$ ,  $K = 100$ ,  $\sigma_t = 0.2$ ,  $r = 5\%$ ,  $\frac{T-t}{T} \in \{0.1; 0.4; 0.7\}$ ,  $v = 0.04$ ,  $\kappa = 1.5$ ,  $\lambda = 3$ ,  $\theta = 0.04$ ,  $\rho = 0.8$ ,  $\mu_S = -0.05$  and  $\sigma_S = 0.00004$



Source: own preparation

Some alternative methods of pricing European options by characteristic functions are presented below.

## CHARACTERISTIC FUNCTIONS

There are numerous ways of deriving characteristic functions for the purpose of pricing European options. In this article particular attention is paid to the formulas developed by Carr & Madan [1999], Attari [2004], Bates [2006] and Orzechowski [2018], only. On the basis of the formulas, the theoretical values of European calls in the Heston model and the Bates model are determined. It should be noted that in all equations presented below it is assumed that  $t = 0$ . The remaining notation remains consistent with the one introduced previously.

### The Heston model

1. The Carr & Madan approach [Carr, Madan 1999] for  $\alpha = 1$ :

$$C^H(S_0, \sigma_0^2, 0) = \frac{e^{-\alpha k}}{\pi} \int_0^\infty \Re \left( e^{-\mathbb{I}\xi k} \frac{e^{-rT} \phi^{2,H}(\xi - (\alpha+1)\mathbb{I}, S_0, \sigma_0^2)}{\alpha^2 + \alpha - \xi^2 + \mathbb{I}(2\alpha+1)\xi} \right) d\xi. \quad (9)$$

2. The Attari approach [Attari 2004]:

$$C^H(S_0, \sigma_0^2, 0) = S_0 \left( 1 + \frac{e^l}{\pi} \int_0^\infty \Re \left( \frac{e^{-\mathbb{I}\xi l}}{\mathbb{I}(\xi + \mathbb{I})} \psi^{2,H}(\xi, s_0, \sigma_0^2) \right) d\xi \right) + e^{-rT} K \left( \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re \left( \frac{e^{-\mathbb{I}\xi l}}{\mathbb{I}\xi} \psi^{2,H}(\xi, s_0, \sigma_0^2) \right) d\xi \right). \quad (10)$$

where:  $l = \ln \left( \frac{K}{S_0 e^{rT}} \right)$ .

3. The Bates approach [Bates 2006]:

$$C^H(S_0, \sigma_0^2, 0) = S_0 - e^{-rT} K \left( \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re \left( \frac{e^{-\mathbb{I}\xi \ln \left( \frac{K}{S_0} \right)}}{\mathbb{I}\xi(1 - \mathbb{I}\xi)} \right) \varphi^{2,H}(\xi, s_0, \sigma_0^2) d\xi \right). \quad (11)$$

4. The Orzechowski approach [Orzechowski 2018]:

$$C^H(S_0, \sigma_0^2, 0) = \frac{1}{2} S_0 + e^{-rT} \frac{1}{\pi} \int_0^\infty \Re \left( e^{-\mathbb{I}\xi k} \frac{\phi^{2,H}(\xi - \mathbb{I}, s_0, \sigma_0^2)}{\mathbb{I}\xi(\mathbb{I}\xi + 1)} \right) d\xi. \quad (12)$$

It is worth noting that:  $\phi^{1,H}(\xi, s_0, \sigma_0^2)$ ,  $\phi^{2,H}(\xi, s_0, \sigma_0^2)$ ,  $\psi^{2,H}(\xi, s_0, \sigma_0^2)$  and  $\varphi^{2,H}(\xi, s_0, \sigma_0^2)$  are characteristic functions corresponding to  $P_1^H(s_t, \sigma_t^2, \tau)$  and  $P_2^H(s_t, \sigma_t^2, \tau)$ , respectively. The functions corresponding to  $P_2^H(s_t, \sigma_t^2, \tau)$  are determined by the following equations:

$$\phi^{2,H}(\xi, s_0, \sigma_0^2) = e^{C_2(\xi, T) + D_2(\xi, T)\sigma_0^2 + \mathbb{I}\xi s_0}. \quad (13)$$

$$\psi^{2,H}(\xi, s_0, \sigma_0^2) = \phi^{2,H}(\xi, s_0, \sigma_0^2) e^{-\mathbb{I}\xi s_0 - \mathbb{I}\xi rT}. \quad (14)$$

$$\varphi^{2,H}(\xi, s_0, \sigma_0^2) = \phi^{2,H}(\xi, s_0, \sigma_0^2) e^{-\mathbb{I}\xi s_0}. \quad (15)$$

### The Bates model

1. The Carr & Madan approach [Carr, Madan 1999] for  $\alpha = 1$ :

$$C^B(S_0, \sigma_0^2, 0) = \frac{e^{-\alpha k}}{\pi} \int_0^\infty \Re \left( e^{-\mathbb{I}\xi k} \frac{e^{-rT} \phi^{2,B}(\xi - (\alpha + 1)\mathbb{I}, s_0, \sigma_0^2)}{\alpha^2 + \alpha - \xi^2 + \mathbb{I}(2\alpha + 1)\xi} \right) d\xi. \quad (16)$$

2. The Attari approach [Attari 2004]:

$$C^B(S_0, \sigma_0^2, 0) = S_0 \left( 1 + \frac{e^l}{\pi} \int_0^\infty \Re \left( \frac{e^{-\mathbb{I}\xi l}}{\mathbb{I}(\xi + \mathbb{I})} \psi^{2,B}(\xi, s_0, \sigma_0^2) \right) d\xi \right) + e^{-rT} K \left( \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re \left( \frac{e^{-\mathbb{I}\xi l}}{\mathbb{I}\xi} \psi^{2,B}(\xi, s_0, \sigma_0^2) \right) d\xi \right). \quad (17)$$

where:  $l = \ln\left(\frac{K}{S_0 e^{rT}}\right)$ .

3. The Bates approach [Bates 2006]:

$$C^B(S_0, \sigma_0^2, 0) = S_0 - e^{-rT} K \left( \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re \left( \frac{e^{-\mathbb{I}\xi \ln\left(\frac{K}{S_0}\right)}}{\mathbb{I}\xi(1-\mathbb{I}\xi)} \right) \varphi^{2,B}(\xi, s_0, \sigma_0^2) d\xi \right). \quad (18)$$

4. The Orzechowski approach [Orzechowski 2018]:

$$C^B(S_0, \sigma_0^2, 0) = \frac{1}{2} S_0 + e^{-rT} \frac{1}{\pi} \int_0^\infty \Re \left( e^{-\mathbb{I}\xi k} \frac{\phi^{2,B}(\xi - \mathbb{I}, s_0, \sigma_0^2)}{\mathbb{I}\xi(\mathbb{I}\xi + 1)} \right) d\xi. \quad (19)$$

It is worth noting that:  $\phi^{2,B}(\xi, s_0, \sigma_0^2)$ ,  $\psi^{2,B}(\xi, s_0, \sigma_0^2)$  as well as  $\varphi^{2,B}(\xi, s_0, \sigma_0^2)$  are characteristic functions determined by the following equations:

$$\phi^{2,B}(\xi, s_0, \sigma_0^2) = e^{C_2(\xi, T) + D_2(\xi, T)\sigma_0^2 + \mathcal{P}(\xi)\lambda T + \mathbb{I}\xi s_0} \quad (20)$$

$$\psi^{2,B}(\xi, s_0, \sigma_0^2) = \phi^{2,B}(\xi, s_0, \sigma_0^2) e^{-\mathbb{I}\xi s_0 - \mathbb{I}\xi rT}. \quad (21)$$

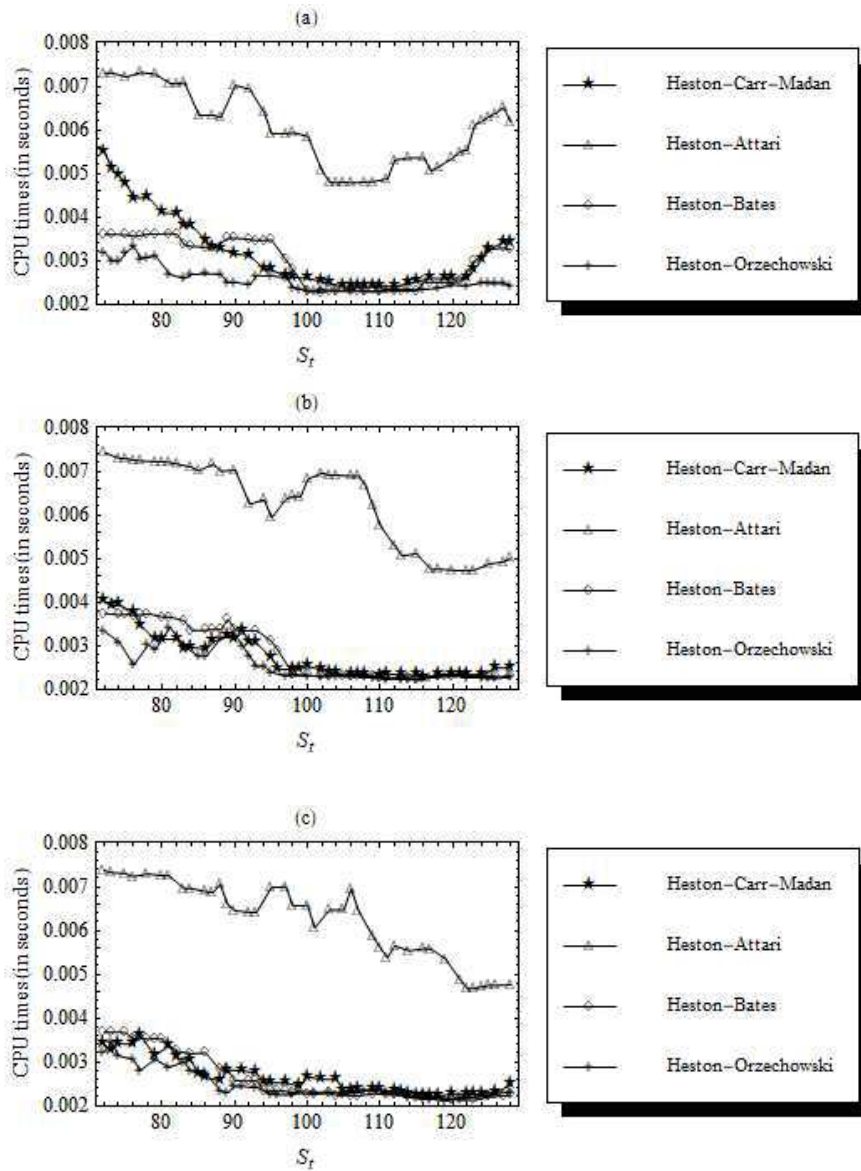
$$\varphi^{2,B}(\xi, s_0, \sigma_0^2) = \phi^{2,B}(\xi, s_0, \sigma_0^2) e^{-\mathbb{I}\xi s_0}. \quad (22)$$

## RESULTS

In order to investigate which of the approaches proposed in the previous section is the most efficient one in terms of computational speed, appropriate codes are developed in the Mathematica 10.2. It is assumed that integrals in the formulas for the prices of European calls are evaluated numerically by means of the Gauss-Kronrod quadrature. Graphs are smoothed by averaging runs of five elements. The package is launched on a computer with Intel i7- 1070 CPU @ 2.90 GHz processor with RAM memory of 32 GB. Cache memory is cleared before starting codes allowing for the valuation of options.

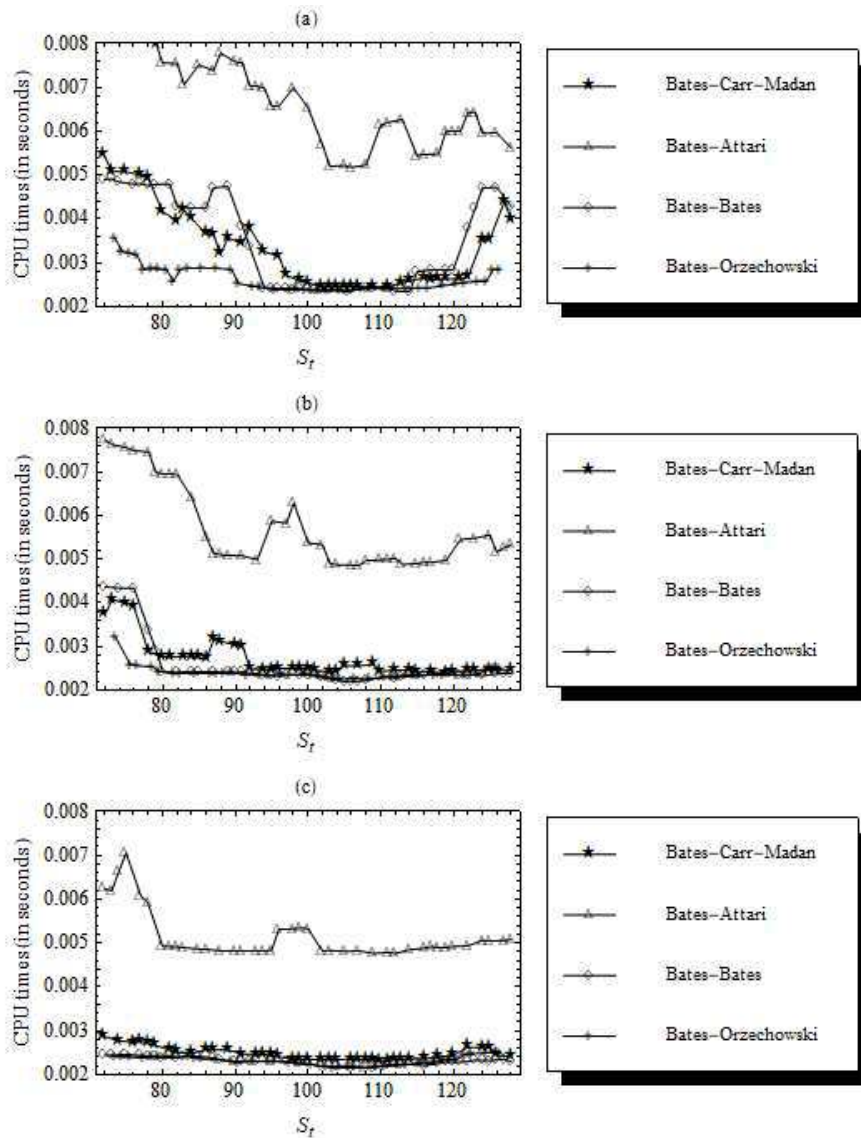
The results of the research carried out are shown in the graphs below - see Figures 3 and 4.

Figure 3. Computational speed in the Heston model assuming that  $S_t \in [70, 130]$ ,  $K = 100$ ,  $\sigma_t = 0.2$ ,  $r = 5\%$ ,  $v = 0.04$ ,  $\kappa = 1.5$ ,  $\lambda = 3$ ,  $\theta = 0.04$ ,  $\rho = 0.8$  for (a)  $\frac{T-t}{T} = 0.1$ , (b)  $\frac{T-t}{T} = 0.4$  and (c)  $\frac{T-t}{T} = 0.7$



Source: own preparation

Figure 4. Computational speed in the Bates model assuming that  $S_t \in [70, 130]$ ,  $K = 100$ ,  $\sigma_t = 0.2$ ,  $r = 5\%$ ,  $v = 0.04$ ,  $\kappa = 1.5$ ,  $\lambda = 3$ ,  $\theta = 0.04$ ,  $\rho = 0.8$ ,  $\mu_S = -0.05$  and  $\sigma_S = 0.00004$  for (a)  $\frac{T-t}{T} = 0.1$ , (b)  $\frac{T-t}{T} = 0.4$  and (c)  $\frac{T-t}{T} = 0.7$



Source: own preparation

The results obtained show that the speed of pricing European options in the Heston model and the Bates model depends on the way the characteristic functions are calculated. The closer the time to expiration the better (in terms of computational speed) is the method developed by Orzechowski [Orzechowski 2018]. The closer the moment of writing European options, the more ambiguous become the differences between alternative methods of pricing derivatives. An exception to the rule applies to the Attari method. In both option pricing models this approach allows for the slowest pricing of the contracts being considered (regardless the time to expiration).

## SUMMARY

In this article some selected methods of determining characteristic functions were applied to the valuation of options. Based on the results obtained it can be concluded that the characteristic function developed by Orzechowski [Orzechowski 2018] allows for the fastest pricing of European options in the Heston model and the Bates model. This is true under the assumption that the prices of the contracts are evaluated numerically by means of the Gauss-Kronrod quadrature. It should be noted that although the results are vulnerable to the time to expiration, nevertheless abovementioned conclusion should be maintained in its current form.

Further research should focus on analyzing the speed of pricing European options in other stochastic volatility models. Particular attention in this matter should be paid to the selection of both numerical methods used to approximate theoretical values of the analyzed instruments and characteristic functions of  $s_t$ .

## REFERENCES

- Attari M. (2004) Option Pricing Using Fourier Transform: A Numerically Efficient Simplification. [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=520042](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=520042), [dostęp: 20.12.2020].
- Bates D. S. (1996) Jumps and Stochastic Volatility: Exchange Rate Process Implicit in Deutsche Mark Option, *The Review of Financial Studies*, 9(1), 69-107.
- Bates D. S. (2006) Maximum Likelihood Estimation of Latent Affine Processes, *Review of Financial Studies*, 19(3), 909-965.
- Black F., Scholes M. (1973) The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3), 637-654.
- Challet D., Marsili M., Zhang Y. -C. (2001) Stylized Facts of Financial Markets and Market Crashes in Minority Games, *Physica A: Statistical Mechanics and Its Applications*, 294(3-4), 514-524.
- Carr P., Madan D. B. (1999) Option Valuation Using the Fast Fourier Transform. *Journal of Computational Finance*, 2(4), 61-73.
- Cont R. (2001) Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues, *Quantitative Finance*, 1(2), 223-236.

- 
- Heston S. (1993) A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6(2), 327-343.
- Lux T. (2009) Stochastic Behavioral Asset - Pricing Models and the Stylized Facts, [in:] T. Hens, K. R. Schenk-Hopp (Ed.), *Handbook of Financial Markets: Dynamics and Evolution*, North-Holland, Amsterdam, The Netherlands.
- Orzechowski A. (2018) Pricing Correlation Options: from the P. Carr and D. Madan Approach to the New Method Based on the Fourier Transform. *Economics and Business Review*, 4(1), 16-28.
- Rogers L. C. G., Zhang L. (2011) Understanding Asset Returns. *Mathematics and Financial Economics*, 5(2), 101-119.
- Taylor S. J. (2005) *Asset Price Dynamics, Volatility and Prediction*. Princeton University Press, Princeton, USA.

## WYKORZYSTANIE WZORCÓW POŚREDNICH DO BUDOWY MIERNIKÓW SYNTETYCZNYCH

**Zbigniew Binderman**  <https://orcid.org/0000-0003-2917-4381>

**Bolesław Borkowski**  <https://orcid.org/0000-0001-6073-6173>

Instytut Ekonomii i Finansów

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

**Wiesław Szczesny**  <https://orcid.org/0000-0002-8083-4624>

Instytut Informatyki Technicznej

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

e-mail: [zbigniew\\_binderman@sggw.pl](mailto:zbigniew_binderman@sggw.pl); [boleslaw\\_borkowski@sggw.edu.pl](mailto:boleslaw_borkowski@sggw.edu.pl);

[wieslaw\\_szczesny@sggw.edu.pl](mailto:wieslaw_szczesny@sggw.edu.pl)

**Streszczenie:** W pracy zaproponowano użycie oprócz wzorców negatywnego i pozytywnego dodatkowego wzorca pośredniego do budowy miernika syntetycznego, przeznaczonego do porządkowania i podziału obiektów na grupy. Na przykładzie rzeczywistego podziału funduszu premiowego pomiędzy jednostki biznesowe w dużej korporacji, zaprezentowano użyteczność takiego podejścia. Przeprowadzone badania wykazały dużą przydatność omówionej metody do porządkowania i grupowania obiektów.

**Słowa kluczowe:** trójwzorcowe mierniki syntetyczne, mediana Webera, metoda TOPSIS, klasyfikacja, funkcja użyteczności

**JEL classification:** C1, G2

### WSTĘP

Podstawowym celem Wielowymiarowej Analizy Porównawczej (WAP) jest dokonanie porządkowania, a następnie grupowania obiektów (jednostek), które są opisane wektorami z wielowymiarowej przestrzeni cech. Do porządkowania i grupowania stosowanych jest wiele metod [zob. Binderman 2006; Binderman i inni 2008, 2010; Borkowski i inni 2020; Cieślak 1976; Borys 1978; Hellwig 1968; Kukuła 2000; Malina 2004; Młodak 2006; Panek 2010; Gatnar, Walesiak 2009; Zeliaś 2000], które zazwyczaj dają różne wyniki. Metody te są często stosowane

<https://doi.org/10.22630/MIBE.2020.21.3.15>



w praktyce do oceny pracy (także do jej premiowania) zespołów ludzkich w złożonych strukturalnie korporacjach. Jednakże wybór wskaźnika syntetycznego będącego podstawą oceny, a szczególnie podziału zespołów na tzw. grupy premiowe budzi niejednokrotnie wiele dyskusji. Jest to spowodowane faktem, że niejednokrotnie niewielkie zmiany w sposobie budowy wskaźnika wpływają na znaczące zmiany w klasyfikacji poszczególnych zespołów do grup premiowych. Dlatego ważnym zagadnieniem jest ograniczenie tego typu wrażliwości. Z tego powodu celem tej pracy było wskazanie możliwości ograniczenia wrażliwości wyników grupowania opartych o wartości różnych mierników/wskaźników wykorzystywanych do oceny i klasyfikacji badanych obiektów. Jednym ze sposobów osiągnięcia tego celu jest - według naszych badań - oparcie definicji wskaźników o więcej niż 2 wzorce, czyli zastosowanie tzw. wzorców pośrednich.

## METODY I TECHNIKI BUDOWY WSKAŹNIKA SYNTETYCZNEGO

Ogólnie w naukach przyrodniczych (fizyce matematycznej, matematyce) uważa się, że zagadnienie (brzegowe, początkowe) jest *poprawnie postawione*, jeżeli:

- przy określonych warunkach istnieje rozwiązanie tego zagadnienia,
- rozwiązanie to jest jednoznaczne,
- rozwiązanie to zależy w sposób ciągły od zadanych warunków (jest *stabilne*).

W naukach ekonomicznych zagadnienie poprawnie postawione jest różnie opisywane. Na przykład, według określeń podanych w serii prac Jacksona [1969] zagadnienie jest *poprawnie postawione*, jeżeli:

- w wyniku zastosowanego algorytmu otrzymujemy jeden wynik,
- otrzymana klasyfikacja (uporządkowanie) jest stabilna. (Czyli otrzymana klasyfikacja nie może być „rażąco odmienna” od klasyfikacji otrzymanych przy małych zmianach w danych wejściowych).
- użyty algorytm musi być niezmienniczy względem permutacji zarówno zmiennych i nazw obiektów, które mają być klasyfikowane. Oznacza to, że algorytm musi być niezależny od etykietowania zmiennych i obiektów.
- użyty algorytm musi być niezależny od skali. Oznacza to, że algorytm musi być niezmienniczy przez mnożenie macierzy podobieństwa przez stałą dodatnią, różną od zera.

Warunki te są wyraźnie nie wystarczające, ponieważ pomijają zupełnie problem nierozwiązanych kwestii adekwatności konfiguracji poszczególnych klas w stosunku do szczegółowych związków między obiektami, czy do szczególnych potrzeb. Do powyższego problemu odnosi się praca Binderman [2015]. Z kolei Breiman [1994] w swojej pracy stwierdza, że pojedynczy klasyfikator (miernik syntetyczny) może być daleki od optymalnego, natomiast kombinacje wielu dają

klasyfikator bliski optymalnemu i stabilny. Niestety w przypadku "słabych" klasyfikatorów w wyniku kombinacji można otrzymać klasyfikator jeszcze gorszy.

Ogólnie z literatury przedmiotu wynika, że nie ma jednego uniwersalnego klasyfikatora i zazwyczaj zalecane jest stosowanie mierników będących funkcją różnych klasyfikatorów. Oczywiście każdy badacz dąży do ideału, czyli wyników, które nie zależałyby od doboru cech, sposobu normalizacji zmiennych, wyboru miary odległości (niepodobieństwa) i techniki budowy miernika syntetycznego.

W polskojęzycznej literaturze wielu autorów nie zawsze uwzględnia powyższe zalecenia. Skłoniło nas to podjęcia próby poprawy technik konstrukcji nowych mierników, jak również do poprawy już używanych mierników. W pracy ograniczyliśmy się do mierników, które mają charakter funkcji użyteczności w warunkach niedosytu [Binderman, Borkowski, Szczesny 2008]. W tym celu wprowadziliśmy pewne oznaczenia formalne.

Niech  $\mathbf{X}=\mathcal{R}^n$ ,  $\mathcal{R}=(-\infty,+\infty)$ ,  $n\in\mathcal{N}$  oznacza  $n$ -wymiarową przestrzeń wektorową. Rozważmy problem polegający na klasyfikacji  $m\in\mathcal{N}$  obiektów  $\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_m$  badanego zjawiska za pomocą  $n\in\mathcal{N}$  zmiennych (cech). Niech wektor  $\mathbf{x}_i=(x_{i1}, x_{i2}, \dots, x_{in})\in X$ ,  $i=1, 2, \dots, m$ , opisuje  $i$ -ty obiekt. Jeżeli  $x_{ik} > x_{jk}$  ( $x_{ik} \geq x_{jk}$ ) dla  $k=1, 2, \dots, n$ , to pisać będziemy

$$\mathbf{x}_i > \mathbf{x}_j, (\mathbf{x}_i \geq \mathbf{x}_j),$$

gdzie  $i, j \in [1, m]$ .

Nietrudno zauważyć, że jeżeli  $\mathbf{x}_i > \mathbf{x}_j$  i  $\mathbf{x}_i \neq \mathbf{x}_j$  to w niektórych przypadkach naturalnym jest nazywać obiekt  $\mathbf{x}_i$  lepszym (wyżej ocenianym) od obiektu  $\mathbf{x}_j$ . Oznacza to, że żadna ze składowych wektora  $\mathbf{x}_i$  nie jest mniejsza od odpowiednich składowych wektora  $\mathbf{x}_j$ , a przynajmniej jedna z nich ma wartość większą, tj. istnieje takie  $k \in [1, n]$ , że  $x_{ik} > x_{jk}$ .

Przyjmijmy następujące oznaczenia:

$$x_{0,k} = \min_{1 \leq i \leq m} x_{i,k}, \quad x_{m+1,k} = \max_{1 \leq i \leq m} x_{i,k}, \quad k=1, 2, \dots, n$$

$$x_{med,k} = \text{mediana}(x_{1,k}, x_{2,k}, \dots, x_{m,k}), \quad k = 1, 2, \dots, n,$$

$$\mathbf{x}_0 := (x_{0,1}, x_{0,2}, \dots, x_{0,n}), \quad \mathbf{x}_{m+1} := (x_{m+1,1}, x_{m+1,2}, \dots, x_{m+1,n}),$$

$$\mathbf{x}_{med} := (x_{med,1}, x_{med,2}, \dots, x_{med,n}).$$

Niech obiekt  $\mathbf{O}_0$  będzie opisany przez wektor  $\mathbf{x}_0$ , obiekt  $\mathbf{O}_{m+1}$  będzie opisany przez wektor  $\mathbf{x}_{m+1}$ , zaś  $\mathbf{O}_{med}$  jest obiektem opisany przez wektor „medianowy”  $\mathbf{x}_{med}$ . Oczywiście jest, że obiekty:  $\mathbf{O}_0$ ,  $\mathbf{O}_{m+1}$  (być może fikcyjne) są odpowiednio nie lepsze, nie gorsze od pozostałych obiektów  $\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_m$ , tj.

$$\mathbf{x}_{m+1} \geq \mathbf{x}_i \text{ oraz } \mathbf{x}_i \geq \mathbf{x}_0 \text{ dla każdego } i: m \geq i \geq 1.$$

W związku z powyższym oznaczmy przez

$$\langle \mathbf{x}_0, \mathbf{x}_{m+1} \rangle := \{ \mathbf{x} \in \mathfrak{R}_+^n : \mathbf{x}_0 \leq \mathbf{x} \leq \mathbf{x}_{m+1} \}.$$

Oczywiście,  $\mathbf{x}_{med} \in \langle \mathbf{x}_0, \mathbf{x}_{m+1} \rangle$

W przypadku gdy obiekty  $\mathbf{O}_0$  i  $\mathbf{O}_{m+1}$  są różne od rozważanych obiektów  $\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_m$ , to obiekty te spełniają rolę **obektu najlepszego**, **obektu najgorszego**, odpowiednio i będą traktowane, jako **wzorcy**. Podobnie jako wzorec będziemy traktowali **obekt medianowy**  $\mathbf{O}_{med}$ . Oczywiście w zależności od praktycznych zastosowań zamiast obiektu  $\mathbf{O}_{med}$  można wziąć inny obiekt modelowy (np. rzeczywisty) na podstawie ustaleń eksperckich. Można także wybrać więcej takich obiektów wzorcowych. W pracy prezentujemy efekt przyjęcia jednego obiektu, ale czytelnik w łatwy sposób może wykorzystać tę ideę na większą liczbę obiektów, określając jakie wartości powinien przyjąć wskaźnik syntetyczny w przypadku tych dodatkowo wybranych obiektów.

Funkcję użyteczności  $u$ , która spełnia warunek:

$$u(\mathbf{x}_0) = 0 \text{ i } u(\mathbf{x}_{m+1}) = 1,$$

nazywać będziemy *znormalizowaną* funkcją użyteczności [Panek 2000].

Zauważmy, że jeżeli dana funkcja użyteczności  $u$  indukuje relacje preferencji w zbiorze  $m+3$  obiektów

$$\mathcal{W} := \{ \mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m, \mathbf{x}_{m+1}, \mathbf{x}_{med} \},$$

to funkcja złożona  $g(u(\mathbf{x}))$ , gdzie  $g: \mathfrak{R} \rightarrow \mathfrak{R}$  jest funkcją rosnącą, jak również funkcją użyteczności, generującą tą samą relację preferencji w zbiorze obiektów  $\mathcal{W}$  co funkcja  $u(\mathbf{x})$ .

Wykorzystując powyższą własność celowe jest unormowanie funkcji użyteczności polegające na wybraniu takiej funkcji,  $g$  aby jej wartość dla obiektu najgorszego  $\mathbf{x}_0$  wynosiła 0, wartość zaś dla obiektu najlepszego  $\mathbf{x}_{m+1}$  wynosiła 1, tj., aby:

$$1. g(u(\mathbf{x}_0)) = 0,$$

$$2. g(u(\mathbf{x}_{m+1})) = 1.$$

Korzystając ze znanych postaci funkcji  $u(\mathbf{x})$  w naszej pracy (zdając sobie sprawę, że takich funkcji jest nieskończenie wiele) podajemy przykład takiej funkcji jednej zmiennej  $g(u)$ , która spełniałaby jeszcze jeden warunek:

$$3. g(U(\mathbf{x}_{\text{med}})) = 1/2.$$

Najprostszą funkcją spełniającą te trzy warunki jest funkcja kawałkami liniowa

$$g(u) = \begin{cases} \frac{1}{2\alpha} u & \text{dla } 0 \leq u < \alpha \\ \frac{(u-1)}{2(1-\alpha)} + 1 & \text{dla } \alpha \leq u \leq 1 \end{cases} \quad (1),$$

$$\alpha := U(\mathbf{x}_{\text{med}}).$$

Dużą klasę mierników syntetycznych wykorzystywanych w badaniach naukowych i praktyce gospodarczej można wyrazić jako odległość od jednego lub dwu wzorców. Jeśli symbolem  $d$  oznaczymy miarę odległości między obiektami, symbolami  $\mathbf{x}_{\text{min}}$ ,  $\mathbf{x}_{\text{max}}$  wektory opisujące, odpowiednio obiekt *najgorszy* i *najlepszy*, charakteryzowane za pomocą cech będących stymulantami, to dla  $\mathbf{x} \in \langle \mathbf{x}_{\text{min}}, \mathbf{x}_{\text{max}} \rangle$  znane z literatury są następujące mierniki syntetyczne:

$$\mathcal{M}_1(\mathbf{x}) = \frac{d(\mathbf{x}_{\text{min}}, \mathbf{x})}{d(\mathbf{x}_{\text{min}}, \mathbf{x}_{\text{max}})},$$

$$\mathcal{M}_2(\mathbf{x}) = 1 - \frac{d(\mathbf{x}_{\text{max}}, \mathbf{x})}{d(\mathbf{x}_{\text{min}}, \mathbf{x}_{\text{max}})},$$

$$\mathcal{M}_3(\mathbf{x}) = \frac{\mathcal{M}_1(\mathbf{x}) + \mathcal{M}_2(\mathbf{x})}{2} = \frac{1}{2} + \frac{d(\mathbf{x}_{\text{min}}, \mathbf{x}) - d(\mathbf{x}_{\text{max}}, \mathbf{x})}{2d(\mathbf{x}_{\text{min}}, \mathbf{x}_{\text{max}})},$$

$$\mathcal{M}_4(\mathbf{x}) = \sqrt{\mathcal{M}_1(\mathbf{x}) \mathcal{M}_2(\mathbf{x})}$$

lub w postaci tradycyjnej czyli „w języku odległości”

$$\mathcal{M}_4(\mathbf{x}) = \frac{1}{d(\mathbf{x}_{\text{min}}, \mathbf{x}_{\text{max}})} \sqrt{d(\mathbf{x}_{\text{min}}, \mathbf{x}) d(\mathbf{x}_{\text{min}}, \mathbf{x}_{\text{max}}) - d(\mathbf{x}_{\text{max}}, \mathbf{x})},$$

$$\mathcal{M}_5(\mathbf{x}) = \frac{\mathcal{M}_1(\mathbf{x})}{1 + \mathcal{M}_1(\mathbf{x}) - \mathcal{M}_2(\mathbf{x})} = \frac{1}{2} + \frac{d(\mathbf{x}_{\text{min}}, \mathbf{x})}{d(\mathbf{x}_{\text{min}}, \mathbf{x}) + d(\mathbf{x}_{\text{max}}, \mathbf{x})},$$

Mierniki  $\mathcal{M}_1$  i  $\mathcal{M}_2$  wykorzystują w zasadzie tylko jeden wzorec, natomiast  $\mathcal{M}_3$ ,  $\mathcal{M}_3$  i  $\mathcal{M}_5$  wykorzystują dwa wzorce. Mierniki te można potraktować jako narzędzia rozwiązywania wielokryterialnych problemów decyzyjnych. Mierniki  $\mathcal{M}_1$  i  $\mathcal{M}_2$  wykorzystują jedno kryterium natomiast  $\mathcal{M}_3$ ,  $\mathcal{M}_3$  i  $\mathcal{M}_5$  wykorzystują dwa kryteria.

Podane wyżej mierniki są znormalizowane tj.

$$0 \leq \mathcal{M}_k(\mathbf{x}) \leq 1 \text{ dla } \mathbf{x} \in \langle \mathbf{x}_0, \mathbf{x}_{m+1} \rangle \text{ dla } k = 1, 2, \dots, 5.$$

Zauważmy ponadto, że:

$$\mathcal{M}_k(\mathbf{x}_0) = 0, \mathcal{M}_k(\mathbf{x}_{m+1}) = 1 \text{ dla } k = 1, 2, \dots, 5.$$

W pracy Hellwiga 1968 podany miernik obiektów przy zastosowaniu standaryzacji cech, wykorzystuje obiekt najlepszy  $\mathbf{x}_{max}$  i miernik  $\mathcal{M}_2$ .

Miernik  $\mathcal{M}_5$  związany jest z metodą TOPSIS (Technique for order Preference by Similarity to Ideal Solution) [Hwang, Yoon 1981]. Metodę TOPSIS można zakwalifikować do metod rankingowych służących do podejmowania decyzji wielokryterialnych. Metodami tego typu są również znane metody SAW (*Simple Additive Weighting*) [Chen, Hwang 1992; Neuman 1998; Zhang 2004] oraz AHP (*Analytical Hierarchy Process*) [Saaty 1980, 1995].

W szeregu prac m.in. Binderman [2006] podano teorie i zastosowania miernika  $\mathcal{M}_1$  ( $\mathcal{M}_2$ ) w ocenie regionalnego zróżnicowania rolnictwa w Polsce, pokazano, że stosowanie metod opartych tylko na jednym wzorcu, w wielu przypadkach prowadzi do otrzymania błędnych wyników, które nie spełniają warunków poprawności, według Jacksona [1969], zobacz również Kukuła, Luty [2017].

W pracy proponujemy zmodyfikowaną technikę budowy miernika syntetycznego, poprzez dodanie co najmniej jednego dodatkowego wzorca, czyli dołożenie w technologii budowy wskaźnika przekształcenia jego wartości według wzoru (1) lub bardziej rozbudowanej funkcji  $g$  jeśli chcemy włączyć więcej niż jeden punkt wzorcowy. Rekomendujemy metody oparte o co najmniej trzy wzorce: wzorec minimalny, maksymalny i wzorec pośredni oparty np. na medianach danych, czyli miernik o postaci:

$$\mathcal{M}_6(*) = \mathbf{g}(u(*)), \text{ gdzie } u(\mathbf{x}) = \mathcal{M}_1(\mathbf{x}).$$

Nowy miernik  $\mathcal{M}_6$  ma następujące własności:

$$\mathcal{M}_6(\mathbf{x}_0) = 0, \mathcal{M}_6(\mathbf{x}_{med}) = 1/2, \mathcal{M}_k(\mathbf{x}_{m+1}) = 1,$$

gdzie  $\mathbf{x}_0, \mathbf{x}_{med}, \mathbf{x}_{m+1}$  są określonymi wcześniej wektorami wzorcowymi.

## BADANIE EMPIRYCZNE

W tabeli 1 zamieszczono 5 stymulant służących do oceny efektów biznesowych pracy 16 oddziałów pewnego dużego Banku (archiwalne dane rzeczywiste). Wartości stymulant zostały znormalizowane przy wykorzystaniu

techniki nazywanej unitaryzacją zerowaną [Kukuła 2000]. Oddziałom nadano sztuczne nazwy  $\mathbf{O}01, \dots, \mathbf{O}16$ .

W standardowej procedurze premiowania oddziały są dzielone na 4-ry uporządkowane grupy, i każda grupa otrzymuje ustalony procent funduszu płac jako fundusz premiowy. Oczywiście współczynniki stanowiące procent funduszu płac są zazwyczaj dość znacząco zróżnicowane w zależności od grupy, do której przydzielono poszczególne oddziały (największy otrzymują oddziały z grupy 1 a najmniejszy z grupy 4). W tabeli tej zostały podane także wartości 6-ciu wskaźników syntetycznych otrzymanych jako odległość od jednego lub dwu wzorców (kolumny  $W1, \dots, W6$ ) oraz informacje o podziale oddziałów na 4-ry grupy w zależności od wartości poszczególnych mierników/wskaźników w stosunku do 3-ech następujących progów:  $m_i - S_i$ ,  $m_i$ ,  $m_i + S_i$  dla  $i=1, \dots, 6$  gdzie  $m_i$  oznacza średnią wartość  $W_i$ , a  $S_i$  odchylenie standardowe  $W_i$ . Jako wzorce w tym przypadku posłużyły wektory

$$\mathbf{x}_0 = (0,0,0,0,0) \text{ oraz } \mathbf{x}_{17} = (1,1,1,1,1).$$

Natomiast wskaźniki zostały określone jako:  $W1$  - odległość od pozytywnego wzorca według metryki Minkowskiego z  $p=1$ ,  $W2$  oraz  $W3$  – odpowiednio odległość euklidesowa od negatywnego i pozytywnego wzorca,  $W4$ - to odległość euklidesowa od obu wzorców według formuły Topsis, a  $W5=(W2+W3)/2$ ,  $W6=(W2*W3)^{0,5}$ .

Z przeprowadzonej analizy wynika, iż podziały na grupy zależą od wybranego miernika/wskaźnika syntetycznego (por. tabela 1). Można oczywiście przeanalizować np. macierz podobieństw dotycząca tych 6-ciu podziałów i wybrać reprezentanta. Jednak wtedy rozpocznie się dyskusja jaką miarę podobieństwa wektorów wybrać. Naszym zdaniem lepszym rozwiązaniem jest wybranie dodatkowego pośredniego wzorca (lub więcej wzorców pośrednich) i ustalenie wartości naszego miernika. W tabeli 2 przedstawiliśmy wyniki dla tych 6-ciu wskaźników w przypadku, gdy zastosujemy funkcję opisaną wzorem (1).

W oparciu o przeliczone wartości mierników dokonaliśmy podziału na cztery grupy. Wektor odpowiadający dodatkowemu wzorcowi to  $\mathbf{x}_{med} = (0,378; 0,406; 0,343; 0,362; 0,195)$ , a wartości wskaźników  $W1, \dots, W6$  dla tego wzorca wynoszą odpowiednio 0,337; 0,345; 0,333; 0,509; 0,506 oraz 0,480. Po takim przekształceniu wartości mierników syntetycznych w tym przypadku otrzymaliśmy pełną zgodność podziału na grupy. Oczywiście nie zawsze tak się musi wydarzyć. Jednak zazwyczaj dołożenie wzorca pośredniego zmniejszy różnorodność podziałów na grupy. W bardziej skomplikowanych przypadkach należy zastanowić się nad użyciem większej niż jeden liczby wzorców pośrednich. Innym rodzajem wzorca pośredniego może być zamiast  $\mathbf{x}_{med}$  wektor  $\mathbf{x}_{web}$  indukowany przez mediany Webera danych. Wektor ten minimalizuje sumę odległości od obiektów  $\mathbf{O}_{01}, \mathbf{O}_{02}, \dots, \mathbf{O}_{16}$  [Młodak 2006].

W przypadku danych podanych w tabeli 2 i odległości euklidesowej jest to wektor  $\mathbf{x}_{\text{web}} = (0,374; 0,434; 0,384; 0,408; 0,307)$ . Sprawdziliśmy, że w przypadku wykorzystania tego wzorca w procesie podziału na grupy do danych z tabeli 1 otrzymujemy podział identyczny z tym zapisanym w tabeli 2. Zmieniają się oczywiście wartości w kolumnach W1-W6.

Tabela 1. Ocena 16 obiektów opisanych 5-ciu wskaźnikami cząstkowymi za pomocą 6-ciu wybranych wskaźników syntetycznych

	wartości wskaźników cząstkowych					wartości wskaźników						grupy					
	X1	X2	X3	X4	X5	W1	W2	W3	W4	W5	W6	W1	W2	W3	W4	W5	W6
O01	0	0,586	0,4	0,324	0,119	0,286	0,353	0,256	0,322	0,305	0,301	3	3	3	4	4	4
O02	0,344	0,401	0	0,203	0,515	0,293	0,342	0,271	0,319	0,306	0,304	3	3	3	4	4	4
O03	0,341	0	0,154	0,304	0,039	0,168	0,216	0,156	0,204	0,186	0,184	4	3	4	4	4	4
O04	0,204	0,305	0,255	0,352	0,355	0,294	0,300	0,292	0,297	0,296	0,296	3	3	3	4	4	4
O05	0,385	0,304	0,146	0,536	0,104	0,295	0,335	0,277	0,317	0,306	0,305	3	3	3	4	4	4
O06	0,591	0,487	0,194	0,399	0,497	0,434	0,454	0,418	0,438	0,436	0,436	2	2	2	3	3	3
O07	0,398	0,702	0,582	0,496	0	0,436	0,497	0,387	0,448	0,442	0,439	2	2	2	3	3	3
O08	0,674	0,971	1	0,79	0,729	0,833	0,843	0,788	0,799	0,815	0,815	1	1	1	1	1	1
O09	0,37	0,266	0,163	0,371	0,125	0,259	0,278	0,252	0,271	0,265	0,265	3	3	3	4	4	4
O10	0,449	0,56	0,848	0,479	1	0,667	0,702	0,602	0,638	0,652	0,650	1	1	1	2	2	2
O11	0,441	0,544	0,455	0	0,053	0,299	0,374	0,263	0,337	0,319	0,314	3	3	3	3	4	3
O12	1	0,411	0,507	0,169	0,265	0,470	0,552	0,397	0,478	0,474	0,468	2	2	2	3	3	3
O13	0,144	0,187	0,208	0,247	0,079	0,173	0,182	0,171	0,180	0,177	0,177	4	4	4	4	4	4
O14	0,087	0,023	0,356	0,63	0,43	0,305	0,379	0,270	0,341	0,324	0,320	3	3	3	3	3	3
O15	0,438	1	0,543	1	0,552	0,707	0,747	0,619	0,662	0,683	0,680	1	1	1	2	2	1
O16	0,119	0,202	0,33	0,229	0,045	0,185	0,209	0,179	0,203	0,194	0,194	4	3	4	4	4	4
m	0,374	0,434	0,384	0,408	0,307	0,381	0,384	0,380	0,503	0,502	0,488						
S	0,239	0,282	0,259	0,240	0,283	0,192	0,192	0,174	0,172	0,182	0,181						
V	0,639	0,649	0,675	0,588	0,921	0,504	0,500	0,457	0,342	0,362	0,372						
Q1	0,159	0,218	0,171	0,234	0,060	0,266	0,284	0,253	0,278	0,273	0,273						
Q2	0,378	0,406	0,343	0,362	0,195	0,297	0,364	0,274	0,329	0,313	0,309						
Q3	0,447	0,580	0,534	0,526	0,511	0,462	0,538	0,413	0,470	0,466	0,461						

Źródło: opracowanie własne

Na zakończenie analizy warto zwrócić uwagę, iż rozkłady wartości 6-ciu wskaźników syntetycznych po zastosowaniu wzorca pośredniego stały się mniej skośne oraz ich średnia i mediana są w pobliżu wartości 0,5. Wydaje się, iż takie

przeskalowanie wartości wskaźników syntetycznych prowadzi do bardziej czytelnego przedstawienia efektów oceny poszczególnych obiektów. Takie postępowanie jest także bardziej odporne na zaburzenia, spowodowane dodaniem jednego lub więcej obiektów, które mają znacząco lepsze wyniki niż pozostałe obiekty (czyli na zjawisko znacznego spłaszczenia wyników pozostałych obiektów).

Tabela 2. Wartości wskaźników z tabeli 1 po przekształceniu przez funkcję  $g(u)$

	wartości wskaźników						rangi						grupy					
	W1	W2	W3	W4	W5	W6	W1	W2	W3	W4	W5	W6	W1	W2	W3	W4	W5	W6
O01	0,425	0,506	0,386	0,316	0,301	0,314	12	9	12	9	11	11	3	3	3	3	3	3
O02	0,435	0,497	0,407	0,314	0,303	0,317	11	10	9	10	9	10	3	3	3	3	3	3
O03	0,249	0,314	0,235	0,201	0,184	0,192	16	14	16	14	15	15	4	4	4	4	4	4
O04	0,437	0,435	0,439	0,292	0,292	0,308	10	12	7	12	12	12	3	3	3	3	3	3
O05	0,438	0,486	0,417	0,311	0,302	0,318	9	11	8	11	10	9	3	3	3	3	3	3
O06	0,573	0,583	0,564	0,430	0,431	0,454	6	6	4	6	6	6	2	2	2	2	2	2
O07	0,575	0,616	0,541	0,440	0,437	0,457	5	5	6	5	5	5	2	2	2	2	2	2
O08	0,874	0,880	0,841	0,795	0,813	0,822	1	1	1	1	1	1	1	1	1	1	1	1
O09	0,385	0,404	0,379	0,266	0,262	0,276	13	13	13	13	13	13	3	3	3	3	3	3
O10	0,749	0,773	0,702	0,632	0,648	0,664	3	3	3	3	3	3	1	1	1	1	1	1
O11	0,444	0,523	0,396	0,331	0,315	0,327	8	8	11	8	8	8	3	3	3	3	3	3
O12	0,601	0,658	0,548	0,469	0,469	0,488	4	4	5	4	4	4	2	2	2	2	2	2
O13	0,257	0,265	0,257	0,177	0,175	0,184	15	16	15	16	16	16	4	4	4	4	4	4
O14	0,453	0,526	0,406	0,336	0,320	0,333	7	7	10	7	7	7	3	3	3	3	3	3
O15	0,779	0,807	0,715	0,656	0,679	0,693	2	2	2	2	2	2	1	1	1	1	1	1
O16	0,275	0,303	0,270	0,199	0,192	0,202	14	15	14	15	14	14	4	4	4	4	4	4
m	0,497	0,536	0,469	0,385	0,383	0,454												
S	0,179	0,174	0,167	0,171	0,181	0,182												
V	0,359	0,324	0,356	0,445	0,474	0,401												
Q1	0,395	0,412	0,381	0,273	0,270	0,284												
Q2	0,441	0,514	0,412	0,324	0,309	0,322												
Q3	0,594	0,648	0,560	0,462	0,461	0,480												

Źródło: opracowanie własne

## PODSUMOWANIE I WNIOSKI

Zazwyczaj w praktycznych zastosowaniach podział ocenianych obiektów na uporządkowane grupy realizuje się w oparciu o arbitralne progi lub arbitralny sposób



ich ustalania. Jednakże taki podział zależy od wybranego do oceny miernika/wskaźnika. Wynika to z faktu, iż rozkład wartości poszczególnych tradycyjnie wykorzystywanych mierników może znacząco się różnić. Jeśli prowadzimy ocenę z uwagi na pewien aspekt (kryterium) przez kilka okresów sprawozdawczych, to zazwyczaj mamy wyobrażenie jaką wartość wskaźnika należałoby nadać w przypadku dla pewnych szczególnych, wybranych wektorów. Na przykład część analityków uważa, że miernik powinien przyjmować wartość 0,5 w przypadku, gdy mamy do czynienia z obiektem wzorcowym, który jest opisany wektorem generowanym przez medianę Webera, w badanym zbiorze wektorów opisujących oceniane obiekty. W naszym przypadku gdybyśmy użyli metryki euklidesowej i wektora  $\mathbf{X}_{web}$  (mediana Webera) zamiast użytego w  $\mathbf{X}_{med}$  do budowy tabeli 2, to podział na grupy pozostał by identyczny jak w tabeli 2 - zmieniłyby się tylko wartości wskaźników  $W_1, \dots, W_6$ .

W podsumowaniu należy dodać, że im więcej użyje się pośrednich wzorców uzgodnionych z ekspertami z danej dziedziny, to dokładniej wyskaluje się tworzony miernik według wymagań dostosowanych do badanego (ocenianego) problemu. Dalsze prace powinny być skoncentrowane na symulacjach zjawisk gospodarczych w poszukiwaniu optymalnych rozwiązań (najlepszego miernika syntetycznego).

## BIBLIOGRAFIA

- Binderman A. (2006) Klasyfikacja obiektów oparta na dwóch wzorcach. *EiOGŻ, Zeszyty Naukowe SGGW*, 60, Warszawa, 25-37.
- Binderman Z. (2015) Zagadnienia poprawnie postawione w ekonomii i zarządzaniu, (klasyfikacja i porządkowanie obiektów, pomiar koncentracji). *Розвиток національної економіки: теорія і практика, Матеріали Міжнародної науково-практичної конференції 3-4 квітня 2015 року Частина 2 Івано-Франківськ*, s. 318-321.
- Binderman Z., Borkowski B., Szczesny W. (2008) O pewnej metodzie porządkowania obiektów na przykładzie regionalnego zróżnicowania rolnictwa. *Metody Ilościowe w Badaniach Ekonomicznych*. Warszawa. Wydawnictwo SGGW, s. 39-48.
- Borkowski B., Wiliński A, Binderman Z., Szczesny W. (2020) Mathematical Analysis of Synthetic Measures Based on Radar Charts. *Mathematical Modelling and Analysis*, 25(3), 473-489.
- Breiman L. (1994) Bagging Predictors. Technical Report 420, Department of Statistics, University of California, USA.
- Cieślak M. (1993) Ekonomiczne zastosowanie mierników syntetycznych ze zmiennym wzorcem. [w:] *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, AE, Kraków.
- Hellwig Z. (1968) Zastosowanie metody taksonomicznej do typologicznego podziału krajów ze względu na poziom ich rozwoju oraz zasoby i strukturę kwalifikowanych kadr. *Przegląd Statystyczny*, 4, 307-327.
- Hwang. C. L., Yoon K. (1981) *Multiple Attribute Decision Making: Methods and Applications*. Springer-Verlag, New York.

- Gatnar E., Walesiak M. (2009) Statystyczna analiza danych z wykorzystaniem programu R. PWN, Warszawa.
- Jackson D. H. (1969) Comparison of Classification. Academic Press. Cambridge University.
- Kukuła K. (2000) Metoda unitaryzacji zerowanej. PWN, Warszawa.
- Kukuła K., Luty L. (2017) Jeszcze o procedurze wyboru metody porządkowania liniowego. *Przegląd Statystyczny*, 64(2), 163-176.
- Młodak A. (2006) Analiza taksonomiczna w statystyce regionalnej. DIFIN, Warszawa.
- Panek E. (2000) *Ekonomia matematyczna*. AE Poznań.
- Panek T. (2009) Statystyczne metody wielowymiarowej analizy porównawczej. Wyd. SGH, Warszawa.
- Zeliaś A. (2000) Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym. Wydawnictwo AE, Kraków.


**AN APPLICATION OF INDIRECT MODELS  
FOR THE CONSTRUCTION OF SYNTHETIC MEASURES  
OF CLASSIFICATION**

**Abstract:** The work is a direct continuation of the series of authors articles concerning the construction of new indicators of classification. In the present paper, a manner of classification of objects which is based on three model objects is proposed. An example of the actual distribution of a bonus fund among business units in a large corporation demonstrates the usefulness of such an approach. Studies have shown that this method is very useful for organizing and grouping objects.

**Keywords:** classification, synthetic measure, Weber median, TOPSIS method, utility function

**JEL classification:** C1, G2

## WYKORZYSTANIE WSPÓLCZYNNIKA MORANA DO KLASYFIKACJI OBIEKTÓW PRZESTRZENNYCH NA PRZYKŁADZIE CEN GRUNTÓW ROLNICZYCH

**Robert Pietrzykowski**  <https://orcid.org/0000-0002-6122-3043>  
Instytut Ekonomii i Finansów  
Szkola Główna Gospodarstwa Wiejskiego w Warszawie  
e-mail: robert\_pietrzykowski@sggw.edu.pl

**Streszczenie:** W artykule przedstawiono możliwości wykorzystania globalnej statystyki Morana do klasyfikacji obiektów przestrzennych. Przestrzenna autokorelacja jest wykorzystywana przede wszystkim do określenia podobieństwa obiektów, które znajdują się w różnych odległościach od siebie. Sąsiedztwo badanych obiektów, określamy przez jego rząd czyli obiekty, które sąsiadują bezpośrednio ze sobą traktujemy jako sąsiadów pierwszego rzędu. Natomiast dalsi sąsiedzi określani są jako sąsiedztwo drugiego, trzeciego i dalszych rzędów. Współczynniki autokorelacji globalne i lokalne pozwalają określić wielkość zależności przestrzennej, ale służą przede wszystkim do wnioskowania o strukturze przestrzennej. Wykorzystując globalne miary statystyki przestrzennej możemy określić istnienie reżimów przestrzennych, a w konsekwencji klasyfikować badane obiekty w klastry.

**Słowa kluczowe:** autokorelacja przestrzenna, ceny gruntów rolniczych

**JEL classification:** C14, R3

### WSTĘP

Współczynniki autokorelacji przestrzennej służą do oceny korelacji zmiennych w odniesieniu do lokalizacji przestrzennych. Przestrzenna autokorelacja w rozumieniu geograficznym określa podobieństwo obiektów, które znajdują się w różnych odległościach od siebie. Współczynniki autokorelacji pozwalają określić wielkość zależności przestrzennej, ale służą przede wszystkim do wnioskowania o strukturze przestrzennej. Wykorzystując miary statystyki przestrzennej możemy określić istnienie reżimów przestrzennych a w konsekwencji klasyfikować badane

<https://doi.org/10.22630/MIBE.2020.21.3.16>

obiekty w klastry. Do globalnych statystyk autokorelacji zaliczamy: współczynnik Morana, C-Geara i Joint-count. Natomiast do lokalnych: współczynnik lokalny Morana, a także wchodzący w skład tzw. LISA (ang. Local Indicator of Spatial Association, Anselin 1995) CL-Geara oraz statystyka G-Getisa [Ord i Getis 1995]. Poszerzeniem miary lokalnej G jest zaproponowana przez Getisa i Orda [2012] miara LOSH (ang. Local spatial heteroscedansity), która pozwala na bardziej szczegółowe analizy przestrzenne.

Celem pracy było pokazanie możliwości wykorzystania globalnej statystyki Morana do klasyfikacji obiektów przestrzennych na przykładzie cen gruntów rolniczych.

## BADANIE ZALEŻNOŚCI

Ze względu na rodzaj obserwowanych zmiennych oraz charakteryzujących je rozkładów do badania zależności wykorzystuje się zwykle współczynniki korelacji Pearsona, Spearmana, Kendalla i V-Cramera. Klasyczny współczynnik korelacji Pearsona jest miarą liniowej zależności i dotyczy analizy cech ciągłych, które charakteryzują się dwuwymiarowym rozkładem normalnym. Do badania wielowymiarowej normalności można wykorzystać test zaproponowany przez Mardia w 1970 r., lub jego modyfikację zaproponowaną przez Hanusz i Tarasińską w 2014. Przy niespełnieniu tych założeń należy wykorzystać rangowe współczynniki korelacji lub testy nieparametryczne.

W analizie przestrzennej do badania korelacji wykorzystuje się dwa typy miar, a mianowicie miary globalne i lokalne. Miary globalne są wskaźnikiem autokorelacji przestrzennej, który dotyczy ogólnego podobieństwa regionów. Natomiast statystyki lokalne pozwalają na określenie podobieństwa badanego regionu w stosunku do jego sąsiadów. Punktem wyjścia konstrukcji globalnych współczynników korelacji oprócz ogólnej koncepcji liniowego współczynnika korelacji Pearsona jest statystyka gamma, która łączy ze sobą informacje dotyczące macierzy odległości między badanymi obiektami i macierzy korelacji między wartościami badanej zmiennej w poszczególnych lokalizacjach. Statystyka gamma nazywana jest również indeksem Mantela [1967], który jako pierwszy opracował i zastosował test korelacji między dwiema macierzami w zagadnieniach epidemiologicznych. W literaturze zagadnienie to znane jest również jako „paradygmat analiz kwadratowych” QAP (ang. Quadratic Analysis Paradigm) [Hubert i in. 1985, Anselin 1992]. W ogólnej postaci statystykę gamma zapiszemy jako:

$$\Gamma = \mathbf{WR} = \sum_{i=1}^N \sum_{j=1}^N w_{ij} r_{ij}, \quad (1)$$

gdzie:  $\mathbf{W}$  jest macierzą odległości, a  $\mathbf{R}$  jest macierzą korelacji obserwowanych zmiennych. Statystyka gamma jest funkcją iloczynów krzyżowych miar odległości w badanej przestrzeni ( $w_{ij}$ ) i miar współzależności ( $r_{ij}$ ) wartości zmiennej Y. Chcąc przedstawić globalną statystykę Morana w postaci miary typu QAP najlepiej dokonać standaryzacji wartości zmiennych zgodnie z formułą:

$$z_i = \frac{y_i - \bar{y}}{s_y}, \quad (2)$$

co w konsekwencji prowadzi do uzyskania statystyki QAP Morana, którą zapisujemy jako:

$$\Gamma_M = \sum_{i=1}^N \sum_{j=1}^N w_{ij} (z_i \cdot z_j), \quad (3)$$

Ostateczna postać statystyki Morana różni się od statystyki QAP Morana (formuła 3) ponieważ jej postać uzyskuje się przez podzielenie konkretnej statystyki przez wariancję z próby w celu uzyskania odpowiedniego przedziału liczbowego dla miernika korelacji (-1; 1):

$$M_{QAP} = \frac{\Gamma_M}{n \cdot \text{var}(Y)} = \frac{\sum_{i=1}^N \sum_{j=1}^N w_{ij} (z_i \cdot z_j)}{n \cdot \text{var}(Y)} = \frac{\mathbf{z}' \mathbf{W} \mathbf{z}}{\mathbf{z}' \mathbf{z}} \quad (4)$$

gdzie:  $\mathbf{z}$  jest wektorem kolumnowym o elementach  $z_i = y_i - \bar{y}$ , kiedy rozpatrujemy statystykę Morana dla zmiennych  $Y$ , które są zbiorem dotyczącym obserwacji w wielu lokalizacjach  $y_i$  dla ( $i = 1, \dots, N$ ).

Mając niestandardyzowaną macierz odległości  $\mathbf{W}$  oraz nietransformowane wartości obserwacji, statystyka Morana obliczana z zastosowaniem miar podobieństwa w postaci iloczynów krzyżowych odchyłeń od średniej ma postać:

$$M_{gl} = \frac{1}{\sum_i^n \sum_j^n w_{ij}} \cdot \frac{\sum_i^n \sum_j^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\frac{1}{N} \sum_i^n (y_i - \bar{y})^2} = \frac{N}{S_0} \cdot \frac{\mathbf{z}' \mathbf{W} \mathbf{z}}{\mathbf{z}' \mathbf{z}}, \quad (5)$$

gdzie:  $S_0 = \sum_i^n \sum_j^n w_{ij}$  jest sumą wszystkich elementów macierzy  $\mathbf{W}$ .

## WYKORZYSTANE DANE

Dane wykorzystane w pracy pochodziły z Systemu Analiz i Monitorowania Rynku Obrotu Nieruchomościami (AMRON). Zgromadzone dane dotyczą ich lokalizacji i cen transakcyjnych z aktów notarialnych, wartości wycen, ofert lub innych w pełni weryfikowalnych źródeł. Baza spełnia najwyższe standardy jakości zdefiniowane przez Komisję Nadzoru Finansowego. W analizach wykorzystano ceny gruntów rolniczych na poziomie powiatów i gmin z lat 2004 – 2012. Dokładny opis danych przedstawił Pietrzykowski w monografii „Przestrzenne różnicowanie cen ziemi rolniczych w Polsce” [2019].

## METODA BADAWCZA

W pracy wykorzystano globalny współczynnik autokorelacji Morana [1948] do klasyfikacji badanych obiektów. Globalny współczynnik autokorelacji Morana określa zależności uwzględniające wzajemne położenie obiektów. Dla danych przestrzennych mówimy zatem o opóźnieniu przestrzennym, które jest spowodowane przez kryterium sąsiedztwa, a tym samym powoduje autokorelację przestrzenną. Konsekwencją istnienia zależności przestrzennej jest grupowanie się

podobnych wartości w klastry, co wykorzystano w badaniach. Związki przestrzenne mogą wynikać z istnienia korelacji w obrębie całego badanego obszaru i wtedy mówimy o globalnej korelacji. Chcąc stwierdzić, czy badane obiekty są rozmieszczone w całkowicie losowy sposób pomiędzy poszczególnymi lokalizacjami należy zweryfikować hipotezę o braku autokorelacji przestrzennej o postaci:

$$H_0: \rho = 0, H_1: \rho \neq 0, \quad (6)$$

dla dostatecznie dużych  $N$  unormowana statystyka Morana  $Z(M_{gl})$  o postaci:

$$Z(M_{gl}) = \frac{M_{gl} - E(M_{gl})}{\sqrt{\text{var}(M_{gl})}} \sim N(0; 1), \quad (7)$$

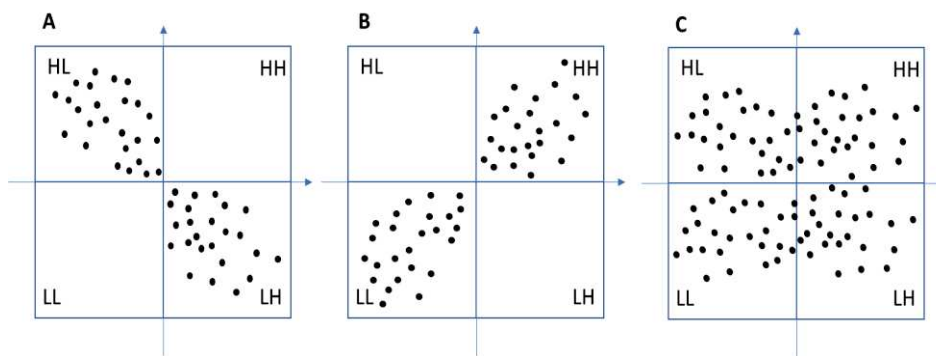
ma rozkład normalny o wartości oczekiwanej 0 i wariancji 1 [Cliff i Ord 1973]. Odrzucenie hipotezy zerowej na rzecz hipotezy alternatywnej ( $H_1$ ) oznacza istnienie autokorelacji przestrzennej i implikuje wniosek, że wartości obserwowanej zmiennej nie są rozmieszczone w sposób losowy pomiędzy poszczególnymi lokalizacjami. Do weryfikacji hipotezy zerowej wykorzystuje się globalny współczynnik korelacji Morana zapisany formułą 5. W pracy wykorzystano macierz wag pierwszego rzędu określającą jedynie najbliższych sąsiadów:

$$W = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1j} \\ w_{21} & w_{22} & \dots & w_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ w_{i1} & w_{i2} & \dots & w_{ij} \end{bmatrix}, \quad (8)$$

gdzie: poszczególne elementy macierzy uzyskano wstawiając  $w_{ij} = 1$  w przypadku kiedy obiekt  $i$ -ty był sąsiadem obiektu  $j$ -tego. Natomiast  $w_{ij} = 0$ , kiedy obiekt  $i$ -ty nie był sąsiadem obiektu  $j$ -tego oraz na diagonalu macierzy  $\mathbf{W}$ , ponieważ dany obiekt nie może być swoim własnym sąsiadem. Tak zapisana macierz pierwszego rzędu nazywana jest również macierzą binarną. Współczynnik korelacji Morana i diagram Morana posłużył do określenia struktury związków przestrzennych. Budując diagram Morana dane oryginalne poddaje się standaryzacji. Na osi  $x$ -ów występuje zatem standaryzowana wartość cechy, a na osi  $y$ -ów opóźnienie przestrzenne. Diagram Morana dzielimy na cztery części względem wartości zerowych. Rozłożenie punktów w przestrzeni wykresu świadczy o obserwowanej autokorelacji. W prowadzonych badaniach zostało to wykorzystane do określenia i podziału obiektów (gmin) ze względu na badaną cechę na klastry. W celu poprawienia czytelności wyników, tak uzyskane grupy obiektów w diagramach Morana przeniesiono na mapę Polski (patrz rysunek 2). Rozłożenie punktów na wykresie Morana (rysunek 1), świadczy o obserwowanej autokorelacji i tak: punkty w kwadracie HL i LH wskazują na ujemną autokorelację, a w kwadratach LL i HH na dodatnią autokorelację. Równomierne rozłożenie punktów w czterech kwadratach świadczy o braku autokorelacji przestrzennej. Jak wspomniano wcześniej wykres Morana wykorzystano do określenia klastrow, które uzyskano odczytując rozmieszczenie punktów w poszczególnych kwadratach wykresu.

Obiekty, które znalazły się w kwadracie LL i HH ze względu na badaną cechę w tym regionie przyjmują odpowiednio niskie lub wysokie wartości. Lewa kolumna obejmująca kwadraty HL i LL, zawiera wartości niskie w regionach sąsiadujących, a kolumna prawa zawierająca kwadraty HH i LH, dotyczy wysokich wartości cechy w regionach sąsiadujących. Analogicznie górny wiersz (kwadraty HL, HH) zawiera wartości wysokie w regionie i-tym, a wiersz dolny (kwadraty LL, LH) obejmuje wartości niskie w regionie i-tym [Anselin 2001, Kopczewska 2020].

Rysunek 1. Przykładowe diagramy Morana obrazujące występowanie ujemnej (A), dodatniej (B) i brak (C) autokorelacji przestrzennej



Źródło: opracowanie własne

W celu potwierdzenia grupowania się obiektów w układy klastrów wykorzystano test Kruskala—Wallisa [Kruskal 1952, Kruskal i Wallis 1952], który jest nieparametryczną wersją jednoczynnikowej analizy wariancji. Podobnie jak w jednoczynnikowej analizie wariancji (ANOVA), weryfikuje się hipotezę o równości średnich z tą różnicą, że zamiast średnich bierzemy pod uwagę mediany:

$$H_0: \theta_1 = \theta_2 = \dots = \theta_k,$$

$$H_1: \text{nie wszystkie } \theta_j \text{ są sobie równe } (j = 1, 2 \dots k) \quad (9)$$

gdzie:  $\theta_1, \theta_2, \dots, \theta_k$  to mediany badanej zmiennej w populacji. Statystykę testową zapisujemy następującą formułą:

$$KW_{emp} = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{(\sum_{i=1}^{n_j} G_{ij})^2}{n_i} - 3(N+1) \quad (10)$$

gdzie:  $G_{ij}$  określa rangi przypisane do wartości zmiennej dla  $i = 1, 2, \dots, n_j, j = 1, 2, \dots, k$ . Jedynym ograniczeniem do stosowania tego testu jest wielkość próby (dla  $n_j > 5$ ). Zaletą testu Kruskala – Wallisa jest to, że nie wymaga on spełnienia założeń dotyczących normalności rozkładów, a także stałości wariancji. W tym przypadku zakładamy, że badane populacje różnią się jedynie parametrem położenia, a zatem mają taki sam kształt i parametr skali. W przypadku odrzucenia hipotezy zerowej (formuła 9) należy zastosować procedurę porównań szczegółowych dla median

stosując testy porównań szczegółowych np. test Dunna [1964], Conovera-Imana [1999] lub Wilcoxona [1945]:

$$H_0: \theta_j = \theta_{j+1}, H_1: \theta_j \neq \theta_{j+1}, \quad (11)$$

Do weryfikacji serii hipotez (formuła 11) wykorzystano test Wilcoxona [Conover 1999] o postaci:

$$T = \min\{\sum G(-), \sum G(+)\}, \quad (12)$$

gdzie:  $\sum G(-)$ ,  $\sum G(+)$  oznaczają sumy rang ujemnych i dodatnich dla badanych prób. Przy prawdziwości hipotezy zerowej (formuła 11) statystyka T ma rozkład Wilcoxona, który przy dużych próbach można aproksymować testem t Studenta [Wilcoxon 1945].

## WYNIKI BADAŃ

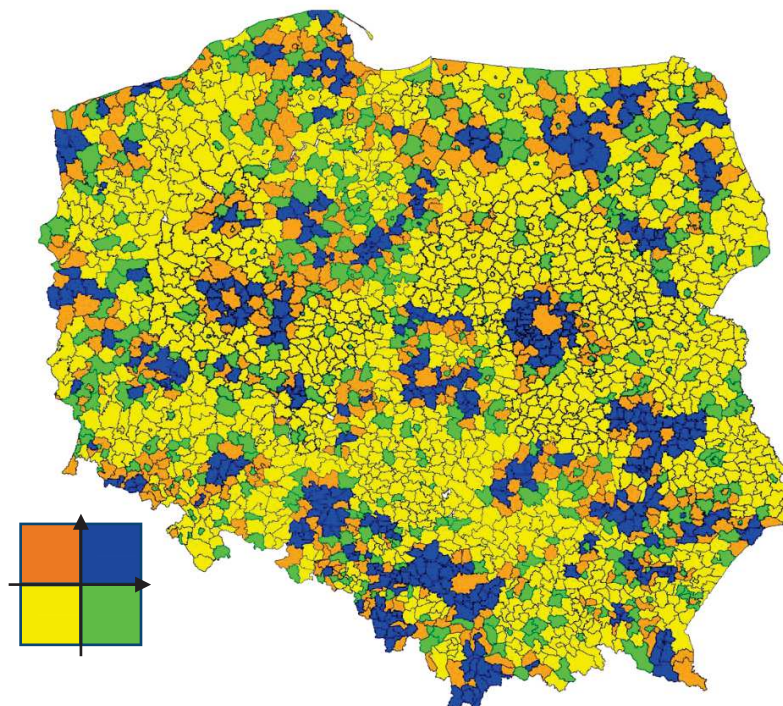
W pracy prowadzono badania obiektów zgodnie z klasyfikacją NUTS, która została wprowadzona w Polsce od 26 listopada 2005 roku. Poziom NUTS1 dzieli Polskę na makroregiony grupujące województwa. NUTS2 dzieli Polskę na województwa, a NUTS3 dzieli Polskę na podregiony grupujące powiaty. Rozporządzenie NUTS pozwala na wydzielenie lokalnych jednostek administracyjnych (LAU<sup>1</sup>), które są różne dla poszczególnych państw Unii Europejskiej. W Polsce gminy są zatem lokalnym poziomem jednostek administracyjnych LAU, na bazie których przeprowadzono analizy z wykorzystaniem statystyki Morana. Na rysunku 2 przedstawiono klasyfikację cen ziemi rolniczej w gminach w Polsce w roku 2004 na podstawie statystyki i diagramu Morana. Poszczególne kolory oznaczają odpowiednio ćwiartki diagramu Morana do których zostały przyporządkowane poszczególne obiekty (gminy). Kolor pomarańczowy oznacza gminy o niskich cenach ziemi otoczonych przez gminy w których zaobserwowano wysokie ceny gruntów (LH). Kolor niebieski oznacza te gminy, w których wartości cen gruntów są zarówno wysokie w regionie, jak i w jego otoczeniu (HH). Kolor żółty to gminy o niskich średnich wartościach w regionie jak i w jego otoczeniu. Kolor zielony to gminy o średnich niskich wartościach w regionie i wysokich w otoczeniu. Zauważmy, że duże miasta wojewódzkie są otoczone przez pierścienie (gminy) w których zaobserwowano wysokie ceny gruntów. W miarę oddalania się od centrum miasta średnie ceny ziemi w gminach są coraz niższe. Podobny podział na klastry uzyskano dla pozostałych lat tzn. od roku 2005 do 2012. W roku 2012 nie obserwowano tak silnego zarysowania pierścieni wokół miast, ale mogły mieć na to wpływ braki w danych.

---

<sup>1</sup> <https://ec.europa.eu/eurostat/web/nuts/local-administrative-units>



Rysunek 2. Klasyfikacja cen ziemi rolniczej w gminach w Polsce w roku 2004 na podstawie statystyki i diagramu Morana

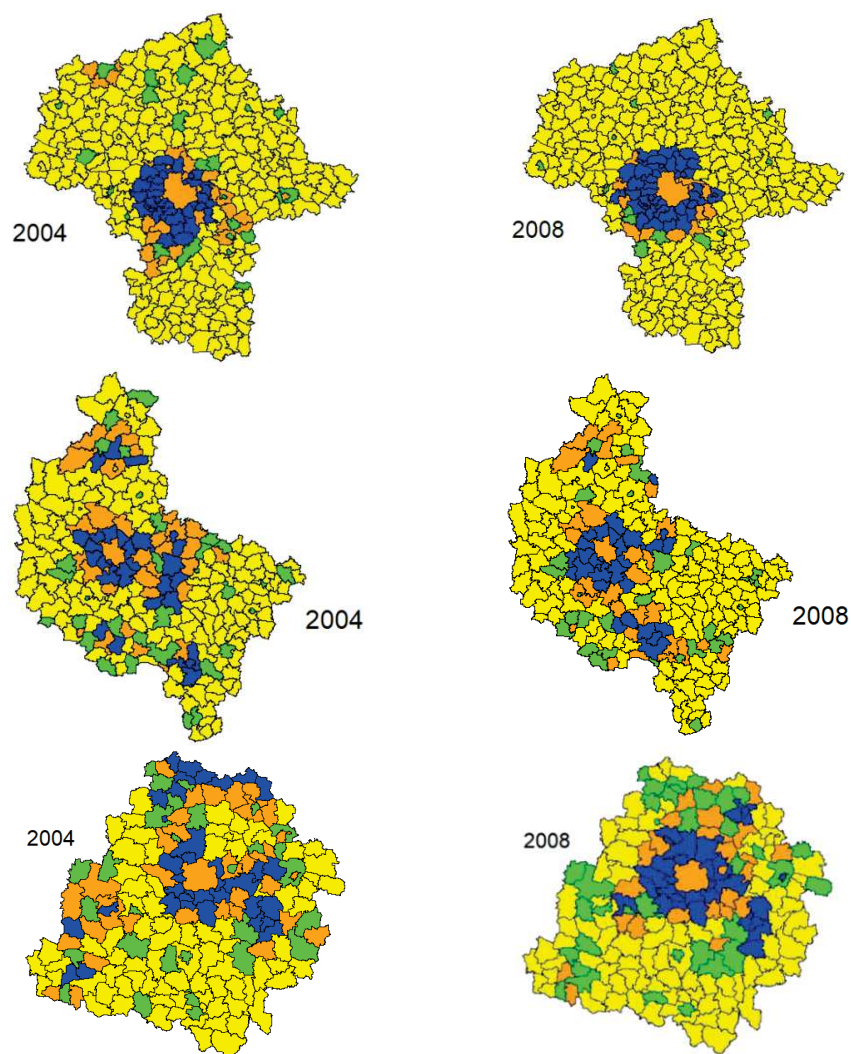


Źródło: opracowanie własne

Szczegółową analizę przeprowadzono dla wszystkich województw z podziałem na gminy. We wszystkich województwach uzyskano podobne wyniki grupowania gmin tzn. duże miasto wojewódzkie silnie oddziaływało na grupowanie badanych jednostek. Znajdowało to swoje potwierdzenie w poszczególnych latach badanego okresu tzn. 2004 – 2012. W niektórych województwach obserwowano podobne podziały dla mniejszych miast. Jednak najsilniejszy podział i oddziaływania miało duże miasto wokół, którego grupowały się gminy. W dalszej części rozdziału przedstawiono, jak wyglądało szczegółowe badanie dla poszczególnych gmin na przykładzie województwa mazowieckiego, łódzkiego i wielkopolskiego. Na rysunku 3 zamieszczono klasyfikację cen ziemi rolniczej w gminach w Polsce w roku 2004 i 2008 na podstawie statystyki i diagramu Morana w trzech województwach. Na rysunku 3 patrząc od góry mamy województwo mazowieckie, następnie w środku wielkopolskie i na dole łódzkie. Analizy dla innych polskich województw wyglądały podobnie, jednak ze względu na ograniczoną objętość pracy przedstawiono jedynie te trzy województwa. Przedstawione lata 2004 i 2008 nie są przypadkowe. Rok 2004 można traktować jako graniczny ze względu na przystąpienie Polski do Unii Europejskiej. Natomiast

rok 2008 jest czasem w którym wystąpił krach gospodarczy na świecie. Zauważmy, że nie miał on wpływu na zmiany cen gruntów rolniczych. Można stwierdzić, że podziały w roku 2008 są silniejsze niż w roku 2004.

Rysunek 3. Klasyfikacja cen ziemi rolniczej w gminach w Polsce w roku 2004 i 2008 na podstawie statystyki i diagramu Morana w województwach: mazowieckim, wielkopolskim i łódzkim



Źródło: opracowanie własne

W dalszej części pracy szukano potwierdzenia uzyskanych klasyfikacji cen ziemi rolniczej w gminach w Polsce na podstawie statystyki i diagramu Morana.

W tym celu wykorzystano test Kruskala–Wallisa i test Wilcozona. Jak wyjaśniono wcześniej, wykorzystanie klasycznej analizy wariancji nie było możliwe ze względu na niespełnienie założeń, zarówno dotyczących równości wariancji, jak i rozkładu normalnego. Dla poszczególnych podziałów weryfikowano następujące hipotezę:

$$H_0: \theta_1 = \theta_2 = \theta_3 = \theta_4, \quad (13)$$

gdzie:  $\theta_1, \theta_2, \theta_3, \theta_4$  oznaczają mediany dotyczące kolejnych klastrów (gmin, dla poszczególnych województw). Hipotezę zapisaną formułą 13 odrzucono dla każdego z badanych województw, co pozwoliło potwierdzić, że uzyskany podział nie jest losowy. W celu wydzielenia grup różniących się między sobą zastosowano test Wilcozona z uwzględnieniem poprawki na wielokrotne testowanie. Przykładowe wyniki porównań szczegółowych dla poszczególnych grup od 1 do 4, dla województwa mazowieckiego zapisano w tabeli 1. Wartości p-value zamieszczone w tabeli, dotyczą porównań szczegółowych dla badanych median, reprezentujących odpowiednie klastry w województwie mazowieckim.

**Tabela 1.** Wyniki podziału na grupy homogeniczne dla cen gruntów rolniczych w województwie mazowieckim w roku 2004, potwierdzające brak losowości w podziałach uzyskanych na podstawie współczynnika korelacji i diagramu Morana

Hipotezy	p-value	Mediany	Oznaczenie na wykresie	Grupy homogeniczne
$H_0: \theta_1 = \theta_2$	< 2.2e-16	$\theta_1$	Blue HH	A
$H_0: \theta_1 = \theta_3$	< 2.2e-16	$\theta_2$	Green HL	B
$H_0: \theta_1 = \theta_4$	< 2.2e-16	$\theta_3$	Yellow LL	C
$H_0: \theta_2 = \theta_3$	< 2.2e-16	$\theta_4$	Orange LH	C
$H_0: \theta_2 = \theta_4$	< 2.2e-16			
$H_0: \theta_3 = \theta_4$	0,2657			

Źródło: opracowanie własne.

Na podstawie przeprowadzonych badań (tabela 1) można stwierdzić występowanie różnic pomiędzy prawie wszystkimi grupami (wartości p-value są mniejsze od zadanego poziomu istotności 0,05). Nie udało się rozróżnić jedynie grup oznaczonych jako 3 i 4, czyli skupień (reżimów przestrzennych) znajdujących się w trzeciej i czwartej ćwiartce ( $0,05 < 0,2657$ ). Można zatem wyciągnąć wniosek, że najwyższe ceny ziemi lokalizują się wokół Warszawy, po czym następuje ich zmniejszanie. Struktura ta utrzymywała się w badanym okresie tzn od 2004 do 2012 roku. Analogiczne badania przeprowadzono dla pozostałych województw i uzyskano podobne wyniki. W województwie mazowieckim miasto stołeczne

można traktować jako metropolię, której oddziaływanie jest tak silne, że nie ma efektu grupowania się cen ziemi wokół małych miast.

## PODSUMOWANIE

Na podstawie przeprowadzonych badań stwierdzono, że globalna statystyka i diagram Morana mogą być wykorzystane do klasyfikacji obiektów przestrzennych. Uzyskane podziały potwierdzono wykorzystując nieparametryczną analizę wariacji i badania szczegółowe. Podział na klastry potwierdził niezmienny w czasie i przestrzeni popyt na ziemię, przede wszystkim wokół dużych miast. Można przypuszczać, że wysoka cena ziemi rolniczej wokół dużych miast związana jest z rosnącymi potrzebami i przekształcaniem ziemi rolniczej w działki o charakterze budowlanym. Zaobserwowano również wpływ dużych miast na ceny gruntów rolniczych. Najsilniej było to obserwowane w województwie mazowieckim, jednak ta sytuacja występuje również w pozostałych województwach. Ze względu na ograniczenia objętościowe w pracy nie można było przedstawić szczegółowych analiz dla wszystkich województw oraz lat. Można je jednak znaleźć w pracy Pietrzykowskiego z 2019 r.

## BIBLIOGRAFIA

- Anselin L. (1992) Space and Applied Econometrics. Special Issue: Regional Science and Urban Economics 22.
- Anselin L. (1995) Local Indicators of Spatial Association – LISA. *Geographical Analysis* 27(2), 93-115.
- Anselin L. (2001) *Spatial Econometrics*. Oxford: Basil Blackwell.
- Conover W. J. (1999) *Practical Nonparametric Statistics* (3rd ed.). New York: John Wiley and Sons.
- Dunn O. J. (1964) Multiple Comparisons using Rank Sums. *Technometrics*, (6), 241-252.
- Hanusz Z., Tarasińska J. (2014) On Multivariate Normality Tests using Skewness and Kurtosis. *Colloquium Biometricum*, 44, 139-148.
- Hubert L. J., Golledge R. G., Costanzo C. M., Gale N. (1985) Measuring Association between Spatially Defined Variables: an Alternative Procedure. *Geographical Analysis*, 17.
- Kopczewska K. (2020) *Przestrzenne metody ilościowe w R: statystyka, ekonometria, uczenie maszynowe, analiza danych*, Warszawa, CeDeWu.
- Kruskal W. H. (1952) A Nonparametric Test for the Several Sample Problem. *Annals of Mathematical Statistics*, 23, 525-540.
- Kruskal W. H., Wallis W. A. (1952) Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47, 583-621.
- Mardia K. V. (1970) Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika*, 57, 519-530.
- Moran P. (1948) The Interpretation of Statistical Maps. *Journal of the Statistical Royal Society, Series B*, 10, 243-251.

- Mantel N. (1967) The Detection of Disease Clustering and Generalized Regression Approach. *Cancer Research*, 27.
- Ord J. K., Getis A. (1995) Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*, 27(4), 286-306.
- Ord J. K., Getis A. (2012) Local Spatial Heteroscedasticity (LOSH). *The Annals of Regional Science*, 48, 529-539.
- Pietrzykowski R. (2019) *Zróżnicowanie przestrzenne cen ziemi rolniczej w Polsce*. Warszawa, Wydawnictwo SGGW.
- Wilcoxon F. (1945) Individual Comparisons by Ranking Methods. *Biometrics*, 1(6), 80-83.

**CLASSIFICATION OF SPATIAL OBJECTS ON BASE THE GLOBAL  
MORAN COEFFICIENT. ON THE EXAMPLE OF AGRICULTURAL  
LAND PRICES**

**Abstract:** The aim of the work was to show the possibility of using the global Moran statistics for the classification of spatial objects on the example of agricultural land prices. The analyses were conducted at the level of communes (in NUTS 3 subregions of the voivodeship) in the years 2004-2012. Apart from the global Moran coefficient and the Moran diagram, nonparametric statistics (one-way ANOVA on ranks) were used in the work.

**Keywords:** spatial autocorrelation, agricultural land prices

**JEL classification:** C14, R3