

Warsaw University of Life Sciences – SGGW
Faculty of Applied Informatics and Mathematics
Department of Econometrics and Statistics

**QUANTITATIVE METHODS
IN ECONOMICS**

**METODY ILOŚCIOWE W BADANIACH
EKONOMICZNYCH**

Volume XX, No. 2

Warsaw 2019

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 Kukuła (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 2019, Volume XX, No. 2

The original version is the paper version

Journal homepage: qme.sggw.pl

Published by Warsaw University of Life Sciences Press

SPIS TREŚCI

Małgorzata Kobylińska – Metoda zanurzania regresji w przypadku występowania obserwacji nietypowych	83
Jan Malchar, Hanna Dudek – Mikroekonometryczne modelowanie zawierania umów kredytowych na cele konsumpcyjne	93
Dominika Urbańczyk – Analiza czasu trwania na rynku na przykładzie przedsiębiorstw z województwa mazowieckiego	105
Dorota Witkowska – Zastosowanie trójczynnika modelu Famy-Frencha i CAPM dla wybranych spółek notowanych na GPW w Warszawie w latach 2007-2017	116
Marta Wójcik – Wielowymiarowa analiza porównawcza warunków oferowanych dla małych i średnich przedsiębiorstw w polskich województwach	128
Tomasz Wójtowicz – Wpływ informacji z USA na śróddzienną sezonowość na przykładzie spółki KGHM	138

METODA ZANURZANIA REGRESJI W PRZYPADKU WYSTĘPOWANIA OBSERWACJI NIETYPOWYCH

Małgorzata Kobylańska  <https://orcid.org/0000-0001-9674-5418>

Wydział Nauk Ekonomicznych
Uniwersytet Warmińsko-Mazurski w Olsztynie
e-mail: gosiak@uwm.edu.pl

Streszczenie: W pracy przedstawiono wykorzystanie metody maksymalnego zanurzenia regresji do szacowania parametrów strukturalnych funkcji regresji liniowej. Dla zbiorów dwuwymiarowych, zawierających obserwacje nietypowe, oszacowane zostały funkcje regresji wykorzystując klasyczną metodę najmniejszych kwadratów oraz metodę opartą na koncepcji zanurzenia obserwacji w próbie. Zauważyć można w jaki sposób obserwacje nietypowe wpływają na oszacowane modele.

Słowa kluczowe: funkcja regresji liniowej, klasyczna metoda najmniejszych kwadratów, zanurzenie obserwacji w próbie, zanurzenie regresji liniowej

JEL classification: C18, C40

WSTĘP

W literaturze coraz częściej poruszane są zagadnienia dotyczące statystyk odpornych oraz obserwacji nietypowych. Odporność jest rozumiana jako brak wrażliwości odnoszącej się do niespełnienia założeń dotyczących zgodności z rozkładem normalnym lub odporności na występowanie w danym zbiorze danych obserwacji nietypowych [Hampel 2000].

Założenie jednorodności zbioru danych jest jednym z podstawowych założeń dotyczących stosowania metod regresji. Zbiór punktów przestrzeni dwuwymiarowej nazywamy jednorodnym w sensie regresji liniowej, jeżeli tworzy w tej przestrzeni figurę o takim kształcie, że można ją aproksymować wykorzystując regresję liniową [Jajuga 1993].

Obserwacje nietypowe mogą zmieniać charakter zależności pomiędzy zmiennymi, co ma istotne znaczenie w badaniu zjawisk ekonomicznych. Z innej

strony mogą dostarczać ważnych informacji dotyczących zmian zachodzących w przypadku tych zjawisk. Jeżeli w danym w zbiorze występują obserwacje nietypowe (odstające) można zdecydować się na ich wyeliminowanie lub zastosowanie odpowiednich metod analizy statystycznej [Barnett, Lewis 1978].

Klasyczne metody statystycznej analizy wielowymiarowej nie są odporne na występowanie obserwacji nietypowych. Wyniki analizy, uzyskane z wykorzystaniem tych metod, mogą prowadzić do błędnych wniosków lub mogą być konsekwencją konstrukcji modelu funkcji regresji, który nieprawidłowo będzie opisywał badane zjawiska. Na szczególną uwagę zasługuje problem odporności w przypadku modeli regresji. Metodę regresji można uznać za odporną, jeżeli model regresji nie jest wrażliwy na występowanie obserwacji nietypowych. Wykazuje on tendencję, która jest reprezentowana przez większość obserwacji zbioru danych. Zagadnienia dotyczące regresji odpornej opisane zostały między innymi w pracach takich autorów jak Andersen [2008], Kosiorowski [2012] lub Rousseeuw i Leroy [1987].

Metody analizy regresji, oparte na koncepcji zanurzania obserwacji w próbie, należą do metod odpornych. W pracy przedstawiono wykorzystanie estymatora maksymalnego zanurzania funkcji regresji do szacowania parametrów modelu regresji dla danych dwuwymiarowych [Hubert, Rousseeuw 1999]. Modele regresji zostały oszacowane dla zbiorów liczbowych zawierających obserwacje nietypowe. W celu zilustrowania użyteczności metod opartych na zanurzeniu obserwacji w próbie, w przypadku występowania obserwacji odstających, oszacowane zostały również funkcje regresji wykorzystując klasyczną metodę najmniejszych kwadratów.

METODA BADAWCZA

W badaniu zjawisk ekonomicznych często należy określić powiązania pomiędzy zmienną zależną a zmiennymi niezależnymi. W tym celu wykorzystać można analizę regresji. Jako narzędzie badawcze pozwala ona zrozumieć i opisać analizowane zjawiska, określić siłę, kształt i kierunek zależności pomiędzy analizowanymi zmiennymi. Zmienne mogą być powiązane pomiędzy sobą za pomocą zależności funkcyjnej. W przypadku dwuwymiarowym każdej wielkości zmiennej niezależnej odpowiada określona jednoznacznie wartość zmiennej zależnej. Analiza regresji umożliwia między innymi określenie siły i rodzaju wpływu jednej zmiennej na drugą lub określenie wartości jednej zmiennej wykorzystując założone lub znane wartości drugiej zmiennej [Stanisz 2007].

Oszacowana funkcja regresji liniowej w przypadku dwuwymiarowym jest postaci:

$$\hat{y}_i = ax_i + b, \quad (1)$$

gdzie a i b są odpowiednio estymatorami parametru strukturalnego przy zmiennej objaśniającej i wyrazu wolnego tej funkcji [Luszniewicz, Słaby 2008].

Różnicę pomiędzy rzeczywistymi wartościami zmiennej objaśnianej (y_i) oraz odpowiadającymi im teoretycznymi wartościami tej zmiennej (\hat{y}_i) nazywamy składnikami resztowymi (e_i) modelu funkcji regresji. Wyznaczone są one według wzoru [Kot i in. 2007]:

$$e_i = y_i - \hat{y}_i. \quad (2)$$

Zanurzanie funkcji regresji liniowej zostało wprowadzone przez Hubert i Rousseeuw [1999]. Metoda oparta na tej koncepcji pozwala na oszacowanie parametrów strukturalnych równania regresji liniowej, która jest najbardziej zanurzona w stosunku do danych empirycznych oraz na określeniu stopnia jej zanurzenia w tym zbiorze. Estymator najbardziej zanurzonej regresji, może być wykorzystywane w przypadku występowania obserwacji nietypowych. Posługując się odpowiednim algorytmem, oprócz estymacji parametrów strukturalnych regresji liniowej, możliwe jest wyznaczenie wartości zanurzania modelu regresji w analizowanym zbiorze danych [Hubert, Rousseeuw 1999]. Wykorzystanie zanurzania w regresji liniowej przedstawione zostało między innymi w pracach Kobylińskiej [2011], Van Aelst, Rousseeuw [2000], Van Aelst i in. [2002]. W pracy definicje podane zostały dla przypadku dwuwymiarowego.

Niech P_n^2 określa zbiór dwuwymiarowy o liczebności n oraz niech (x_i, y_i) dla $i = 1, 2, \dots, n$ będzie obserwacją należącą do P_n^2 . Funkcję regresji liniowej $\hat{y}_i = ax_i + b$ nazywamy niedopasowaną do zbioru P_n^2 , jeżeli istnieje taka liczba rzeczywista $v_y = v$, która nie pokrywa się z żadną wartością x_i , czyli $x_i \neq v$, że spełnione są następujące warunki

$$e_i < 0 \text{ dla każdego } x_i < v \text{ i } e_i > 0 \text{ dla każdego } x_i > v$$

lub

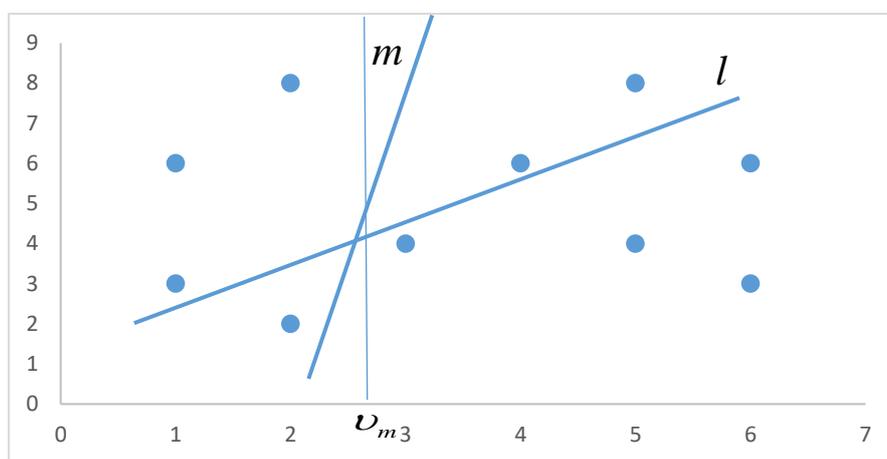
$$e_i > 0 \text{ dla każdego } x_i < v \text{ i } e_i < 0 \text{ dla każdego } x_i > v.$$

Zanurzeniem funkcji regresji liniowej $rzan(\hat{y}_i, P_n^2)$ w zbiorze dwuwymiarowym P_n^2 określamy najmniejszą liczbę obserwacji zbioru P_n^2 , które powinny zostać wyeliminowane ze zbioru P_n^2 , żeby funkcja \hat{y}_i stała się niedopasowana do danego zbioru. Należy zauważyć, że jeżeli funkcja \hat{y}_i , położona jest powyżej lub poniżej wszystkich obserwacji danego zbioru, zawsze będzie niedopasowane do danych empirycznych [Hubert, Rousseeuw 1998].

Na rysunku 1 przedstawiono wykres korelacyjny zbioru P_{10}^2 o liczebności 10. Umieszczone zostały dwie funkcje regresji, spośród których m jest

niedopasowana do danych tego zbioru. Zauważyć można, że w przypadku tej funkcji istnieje liczba ν_m taka, że możemy obrócić funkcję m do pozycji pionowej w ten sposób, że nie przejdzie ona przez żaden punkt P_{10}^2 . W przypadku prostej l , $rzan(m, P_{10}^2) = 4$, gdyż należy usunąć cztery punkty, żeby prosta l stała się niedopasowana do zbioru P_{10}^2 . Inaczej mówiąc, zanurzenie funkcji regresji jest równe najmniejszej liczbie reszt, które muszą zmienić znak, żeby prosta regresji stała się niedopasowana do danych empirycznych.

Rysunek 1. Wykres korelacyjny zbioru P_{10}^2 oraz dwie funkcje, gdzie m jest niedopasowana do zbioru P_{10}^2



Źródło: opracowanie własne na podstawie Hubert i Rousseeuw [1999]

Maksymalne zanurzenie funkcji regresji w zbiorze P_n^2 spełnia nierówność $\max rzan(\hat{y}_i, P_n^2) \geq \lceil n/3 \rceil$, gdzie $\lceil A \rceil$ jest częścią całkowitą liczby A . Dla każdego $P_n^2 \subset R^2$ zachodzi $\lceil n/3 \rceil \leq rzan(\hat{y}_i, P_n^2) \leq n$. Jeżeli wszystkie punkty P_n^2 leżą na prostej regresji \hat{y}_i , dla $i = 1, 2, \dots, n$, wtedy $rzan(\hat{y}_i, P_n^2) = \max rzan(\hat{y}_i, P_n^2) = n$. Własności dotyczące zanurzania regresji liniowej omówione zostały między innymi w pracy Hubert i Rousseeuw [1998].

PRZYKŁAD EMPIRYCZNY

W celu przedstawienia zastosowania metody zanurzania regresji liniowej wykorzystano dane liczbowe dwuwymiarowe. Do oszacowania funkcji regresji wykorzystano metodę najmniejszych kwadratów (obliczenia przeprowadzono

z wykorzystaniem pakietu STATISTICA PL) oraz metodę zanurzania regresji liniowej, dla której obliczenia przeprowadzono wykorzystując pakiet środowiska R „DepthProc”¹. W skład pakietu wchodzi zestaw procedur statystycznych opartych na koncepcji zanurzania obserwacji w próbie. Informacje na temat pakietu oraz możliwości jego zastosowania w przypadku wielowymiarowym zamieszczone zostały między innymi w pracy Kosiorowskiego i Zawadzkiego [2014].

Na rysunkach 2-5 przedstawiono dane liczbowe dwuwymiarowe przyjmując za każdym razem inną konfigurację zbiorów (zbiory $P_{30}^2(xy)$, $P_{30}^2(x_1, y_1)$, $P_{30}^2(x_4, y_4)$, $P_{30}^2(x_5, y_5)$). Dla każdego z tych zbiorów oszacowane zostały dwie funkcje regresji liniowej wykorzystując dane metody. Równania funkcji regresji zamieszczone zostały w tabeli 1.

Tabela 1. Oceny równań regresji z wykorzystaniem danych metod

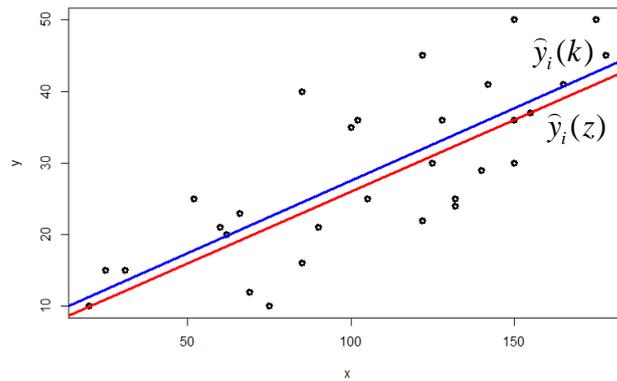
	Metoda zanurzania regresji	Metoda najmniejszych kwadratów
Ocena funkcji regresji	$\hat{y} = 6,00 + 0,20x$	$\hat{y} = 7,24 + 0,20x$
	$\hat{y}_1 = 6,00 + 0,20x_1$	$\hat{y}_1 = 11,59 + 0,15x_1$
	$\hat{y}_4 = 10,89 + 0,17x_4$	$\hat{y}_4 = 46,36 - 0,07x_4$
	$\hat{y}_5 = 8,67 + 0,20x_5$	$\hat{y}_5 = 7,98 + 0,32x_5$

Źródło: opracowanie własne

Nadmienić można, że zbiór danych $P_{30}^2(xy)$ (rysunek 2) tworzy zwartą figurę o kształcie pozwalającym aproksymować ją wykorzystując funkcję regresji liniowej. Zauważyć można, że funkcje regresji oszacowane dla tego zbioru położone są „blisko” siebie. Wynika to z faktu, że nie występują w tym przypadku obserwacje nietypowe, które mogą przyczynić się do zmiany położenia funkcji regresji. Współczynnik korelacji liniowej Pearsona wynosi dla zmiennych tego zbioru $r_{xy} = 0,78$ i jest to najwyższa wartość spośród współczynników korelacji wyznaczonych dla wszystkich rozpatrywanych zbiorów, które wynoszą odpowiednio $r_{x_1, y_1} = 0,74$, $r_{x_4, y_4} = 0, -0,18$ oraz $r_{x_5, y_5} = 0,41$.

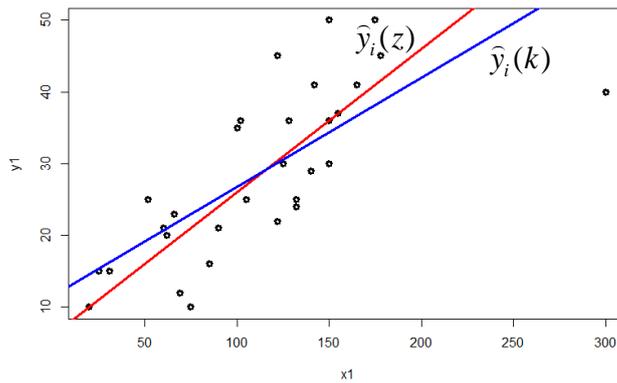
¹ <https://cran.rstudio.com/web/packages/DepthProc/index.html>

Rysunek 2. Wykres korelacyjny zbioru $P_{30}^2(xy)$ oraz funkcje regresji liniowej oszacowane metodą najmniejszych kwadratów ($\hat{y}_i(k)$) oraz metodą maksymalnego zanurzenia regresji ($\hat{y}_i(z)$)



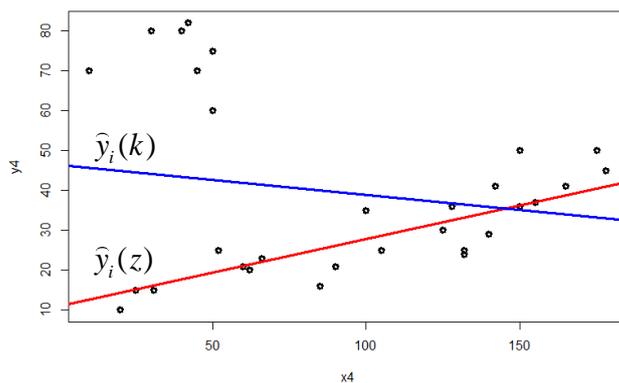
Źródło: opracowanie własne

Rysunek 3. Wykres korelacyjny zbioru $P_{30}^2(x_1 y_1)$ oraz funkcje regresji liniowej oszacowane metodą najmniejszych kwadratów ($\hat{y}_i(k)$) oraz metodą maksymalnego zanurzenia regresji ($\hat{y}_i(z)$)



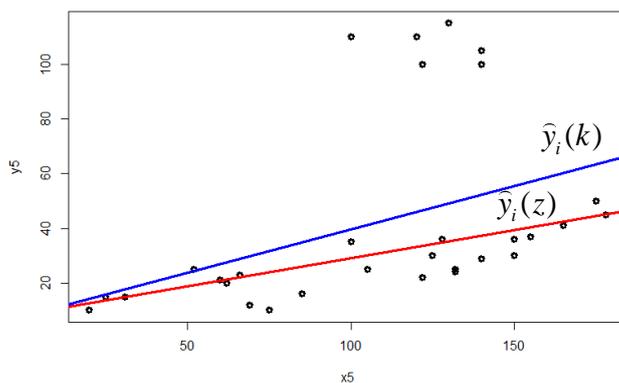
Źródło: opracowanie własne

Rysunek 4. Wykres korelacyjny zbioru $P_{30}^2(x_4, y_4)$ oraz funkcje regresji liniowej oszacowane metodą najmniejszych kwadratów ($\hat{y}_i(k)$) oraz metodą maksymalnego zanurzenia regresji ($\hat{y}_i(z)$)



Źródło: opracowanie własne

Rysunek 5. Wykres korelacyjny zbioru $P_{30}^2(x_5, y_5)$ oraz funkcje regresji liniowej oszacowane metodą najmniejszych kwadratów ($\hat{y}_i(k)$) oraz metodą maksymalnego zanurzenia regresji ($\hat{y}_i(z)$)



Źródło: opracowanie własne

Na rysunkach 3-5 zauważyć można występowanie obserwacji nietypowych, które nie pasują do konfiguracji rozpatrywanych zbiorów danych. Zmieniają one charakter zależności pomiędzy zmiennymi [Zeliaś 1996]. Występowanie w zbiorze

$P_{30}^2(x_1, y_1)$ obserwacji, która jest odstająca w kierunku osi OX (rysunek 3), spowodowała zmianę nachylenia funkcji regresji, oszacowanej klasyczną metodą najmniejszych kwadratów. Regresja liniowa oszacowana z wykorzystaniem metody opartej na koncepcji zanurzania obserwacji w próbie, nie zmieniła swojego położenia. Oceny parametrów strukturalnych funkcji oszacowanej tą metodą są takie same jak w przypadku zbioru $P_{30}^2(xy)$.

Na rysunku przedstawiającym zbiór $P_{30}^2(x_4, y_4)$ (rysunek 4) zauważyć można występowanie podzbioru punktów nietypowych, którym odpowiadają niskie wartości zmiennej X oraz wysokie wartości drugiej zmiennej. Spowodowało to zmianę siły i kierunku zależności pomiędzy analizowanymi zmiennymi. Ocena współczynnika kierunkowego funkcji regresji liniowej oszacowanej metodą najmniejszych kwadratów jest ujemna (tabela 1). W przypadku wykorzystania metody najbardziej zanurzonej regresji nadal przyjmuje on wartość dodatnią. Zauważyć można, że występowanie podzbioru obserwacji nietypowych, nie ma wpływu na zmianę kierunku regresji liniowej w tym przypadku. W zbiorze $P_{30}^2(x_5, y_5)$ (rysunek 5) występuje podzbiór obserwacji nietypowych ze względu na wysokie wartości zmiennej Y. Funkcja regresji oszacowana klasyczną metodą najmniejszych kwadratów umieszczona jest powyżej funkcji oszacowanej drugą metodą.

PODSUMOWANIE

Przedstawione rozważania prowadzą do wniosku, że dużym ograniczeniem, które spotykamy podczas szacowania parametrów funkcji regresji, jest problem występowania obserwacji nietypowych. W pracy przedstawiono użyteczność metod opartych na zanurzaniu obserwacji w próbie w szacowaniu parametrów strukturalnych funkcji regresji liniowej w przypadku występowania w zbiorze danych obserwacji znacznie odbiegających od pozostałych.

Wykorzystanie metody zanurzania regresji umożliwia oszacowanie regresji, dla większości punktów zbiorów danych. Na oceny parametrów regresji nie ma wpływu występowanie obserwacji nietypowych, modele regresji wykazują znacznie bliższe położenie w stosunku do większości punktów danych zbiorów. Estymator uzyskany z wykorzystaniem tej metody wykazuje się niską wrażliwością na zmiany wartości obserwacji w zbiorze danych [Hubert, Rousseeuw 1998].

BIBLIOGRAFIA

- Andersen R. (2008) *Modern Methods for Robust Regression*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-152.
- Barnett V., Lewis T. (1978) *Outliers in Statistical Data*. Wiley and Sons, New York.
- Jajuga K. (1993) *Statystyczna analiza wielowymiarowa*. PWN, Warszawa.
- Hampel F. (2000) *Robust Inference*, Research Report 93. Seminar für Statistik, ETH Zürich. To appear: *Encyclopedia of Environmetrics*, Wiley. <https://www.researchcollection.ethz.ch/bitstream/handle/20.500.11850/145174/eth-24068-01.pdf?sequence=1>.
- Hubert M., Rousseeuw P. (1998) The Catline for Deep Regression. *Journal of Multivariate Analysis*, 66, 270-296.
- Kobylińska M. (2011) Zanurzenie w regresji liniowej. *Metody Ilościowe w Badaniach Ekonomicznych*, Warszawa, 12(2), 202-209.
- Kosiorowski D. (2012) *Statystyczne funkcje głębi w odpornej analizie ekonomicznej*. Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie.
- Kosiorowski D., Zawadzki Z. (2014) *DepthProc: An R Package for Robust Exploration of Multidimensional Economic Phenomene*. <http://arxiv.org/pdf/1408.4542.pdf>.
- Kot S., Jakubowski J., Sokołowski A. (2007) *Statystyka. Podręcznik dla studentów ekonomicznych*. Difin, Warszawa.
- Luszniewicz A., Słaby T. (2008) *Statystyka z pakietem komputerowym STATISTICA PL. Teoria i zastosowania*, Wydawnictwo C.H.Beck, Warszawa.
- Stanisz A., (2007) *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Modele liniowe i nieliniowe*, StatSoft, Kraków.
- Rousseeuw P., Hubert M. (1999) Regression Depth. *Journal of the American Statistical Association*, 94, 388-402.
- Rousseeuw P., Leroy A. (1987) *Robust Regression and Outlier Detection*. Wiley, New York.
- Van Aelst S., Rousseeuw P. J. (2000) Robustness of Deepest Regression. *Journal of Multivariate Analysis*, 73, 82-106.
- Van Aelst S., Rousseeuw P. J., Hubert M., Struyf A. (2002) The Deepest Regression method. *Journal of Multivariate Analysis*, 81, 138-166.
- Zeliaś A. (1996) *Metody wykrywania obserwacji nietypowych w badaniach ekonomicznych*. *Wiadomości Statystyczne*, 8, 16-27.
- <https://cran.rstudio.com/web/packages/DepthProc/index.html> [dostęp 10.08.2018]

REGRESSION DEPTH METHOD FOR UNUSUAL OBSERVATIONS

Abstract: This paper presents the application of the regression maximum depth for the estimation of linear regression function structural elements. For two-dimensional sets including unusual observations, regression functions were developed using the classical least squares method and a method based on the concept of observation depth measure in a sample. The effect of unusual observations on the estimated models has been noted.

Keywords: linear regression function, classical least squares method, observation depth measure in a sample, linear regression depth

JEL classification: C18, C40

MIKROEKONOMETRYCZNE MODELOWANIE ZAWIERANIA UMÓW KREDYTOWYCH NA CELE KONSUMPCYJNE

Jan Malchar  <https://orcid.org/0000-0001-5230-8670>

Hanna Dudek  <https://orcid.org/0000-0001-8261-2745>

Wydział Zastosowań Informatyki i Matematyki
Szkola Główna Gospodarstwa Wiejskiego w Warszawie
e-mail: janek.malchar@wp.pl; hanna_dudek@sggw.pl

Streszczenie: Celem artykułu jest identyfikacja czynników wpływających na skłonność do zaciągania przez gospodarstwa domowe kredytów i pożyczek na cele konsumpcyjne oraz weryfikacja zasadności zastosowania modelu probitowego z efektami losowymi. Na podstawie analizy danych z badania „Diagnoza Społeczna” wskazano determinanty analizowanego zjawiska. Wnioski z przeprowadzonej analizy wyraźnie potwierdziły zasadność wykorzystania w modelowaniu panelowej struktury danych, umożliwiające wyszczególnienie nieobserwowalnych efektów indywidualnych analizowanych gospodarstw.

Słowa kluczowe: kredyt konsumpcyjny, pożyczka, gospodarstwa domowe, dane panelowe, model probitowy z efektami losowymi

JEL classification: H31, C23, C25

WPROWADZENIE

Możliwość zaciągania pożyczek i kredytów bankowych przez gospodarstwa domowe, zarówno w Polsce jak i na świecie, ma dosyć istotne znaczenie. Między innymi umożliwia członkom rodzin zaspakajanie aktualnych potrzeb konsumpcyjnych, których zakup musiałby się wiązać z wcześniejszym gromadzeniem oszczędności. Kształtuje także ich ogólną sytuację majątkową [Wałęga 2013]. Ponadto, udzielanie kredytów stanowi ważny element rozwoju gospodarczego [Folwarski 2016]. Spowodowane jest to zwiększeniem konsumpcji oraz lepszą spłacalnością zobowiązań. Warto jednak pamiętać, że rozwój społeczno-gospodarczy ma charakter złożony i na jego poziom wpływają

<https://doi.org/10.22630/MIBE.2019.20.2.10>

różnorodne czynniki ekonomiczne, społeczne i geograficzno-przyrodnicze [Malchar, Zielińska-Sitkiewicz 2017].

Z uwagi na istotną wagę rozpatrywanego zjawiska, zagadnienie kredytowania gospodarstw domowych wielokrotnie poddawano analizie. W literaturze oprócz prac poświęconych teoretycznym aspektom zaciągania pożyczek i kredytów [Waliszewski 2007; Paczóska 2008; Główka 2009; Bieniasz, Czerwińska-Kayzer 2010; Bywalec 2012; Skikiewicz 2014] można znaleźć również opracowania, w których przedstawiono wyniki analizy danych empirycznych dotyczących modelowania prawdopodobieństwa zaistnienia zadłużenia [Crook 1996; Wałęga 2010; Rusnak 2011; Czapiński, Panek 2015]. W opracowaniach tych stosowano głównie modele logitowe oraz probitowe oszacowane na podstawie danych przekrojowych. Do nielicznych prac wykorzystujących dane panelowe należy zaliczyć opracowanie Crooka i Hochguertela, w którym podjęto próbę uwzględnienia nieobserwowalnej heterogeniczności gospodarstw [Crook, Hochguertel 2007].

Celem naszej pracy jest identyfikacja determinant wpływających na zaciąganie pożyczek i kredytów konsumpcyjnych przez gospodarstwa domowe oraz weryfikacja hipotezy, że uwzględnienie panelowej struktury danych poprawia dopasowanie modelu do danych empirycznych. Wykorzystano dane zebrane w ramach badania „Diagnoza społeczna” z lat 2009–2015. W analizie statystycznej, oprócz modelu probitowego z efektami losowymi rozpatrzono zwykły model probitowy typu „pooled regression”.

DETERMINANTY KREDYTOWANIA NA CELE KONSUMPCYJNE

Kredyty konsumpcyjne związane są z zakupem dóbr trwałego użytku, papierów wartościowych, funkcjonowaniem kart kredytowych oraz bieżącymi wydatkami [Waliszewski 2007; Wałęga 2013]. Do tego segmentu należą m.in. pożyczki na stałe opłaty, kształcenie, leczenie, wypoczynek oraz na pozostałe cele członków gospodarstw. Kredyty konsumpcyjne dotyczą także kredytów na zakupy w systemie ratalnym oraz spłatę poprzednich zadłużeń.

W ostatnich latach można dostrzec znaczny wzrost popytu gospodarstw domowych na korzystanie z usług rynku kredytowego [Utzig 2012]. Na to zjawisko w pozytywny sposób wpływa m.in. wysokie tempo wzrostu gospodarczego, spadek bezrobocia oraz większe wynagrodzenia. Optymistyczna ocena własnej sytuacji finansowej oraz zmniejszające się obawy przed utratą pracy i spadkiem dochodów służą podejmowaniu decyzji o zaciąganiu kredytów [Harasim 2005; Wałęga 2010; Bieniasz, Czerwińska-Kayzer 2010; Folwarski 2016]. Do sprzyjających czynników zaliczyć należy również łagodzenie przez banki polityki kredytowej, mniejsze koszty finansowania kredytów oraz większą liczbę ofert w sektorze bankowym, co wpływa na konkurencyjność oraz warunki udzielania kredytów [Nieto 2007; Paczóska 2008; Skikiewicz 2014].

Gospodarstwa domowe zadłużają się z różnych powodów. Niekiedy potrzeba zaciągnięcia kredytu odnosi się do sytuacji, gdy wydatki mieszkańców przerastają wartość bieżących zarobków. Dotyczy to przede wszystkim gospodarstw o niskich dochodach, które by przeżyć są zmuszone do zaciągania długów [Bywalec 2012]. Istnieją również okoliczności, w których korzystanie z pożyczek związane jest z zakupem kosztownych lub nawet ekskluzywnych dóbr. Znaczna część wzrostu spożycia gospodarstw domowych w Polsce jest finansowana właśnie ze środków zewnętrznych, czyli m.in. z kredytów [Grzybowska 2012]. Zjawisko to towarzyszy w szczególności gospodarstwom domowym w krajach średnio oraz wysoko rozwiniętych. Motywami do zadłużania przez gospodarstwa domowe mogą być także różnego typu zdarzenia losowe, o pozytywnym i negatywnym charakterze. Są to okoliczności zagrażające bytowi całego gospodarstwa lub któregoś z jego członków (np. pożar mieszkania, kradzież, ciężka choroba, śmierć bliskiej osoby), a także korzystne, nagłe i niepowtarzalne okazje, do których zalicza się możliwość zakupu taniego mieszkania, domu, gruntu bądź samochodu [Bywalec 2012]. Oznacza to, iż podjęcie decyzji o zapożyczaniu w bankach może nastąpić w wyniku nagłego impulsu.

Kredytowanie gospodarstw domowych jest konsekwencją oddziaływania wielu czynników. Na popyt poszczególnych jednostek mają wpływ między innymi cechy społeczne, demograficzne, ekonomiczne, finansowe oraz psychologiczne [Crook 1996; Stone i in. 2006; Bywalec 2012; Skikiewicz 2014; Wałęga, Wałęga 2015]. Do cech socjo-demograficznych zaliczyć należy między innymi rodzaj miejsca zamieszkania (miasto, wieś), wielkość gospodarstwa domowego, wiek głowy gospodarstwa domowego, a także jej wykształcenie i status zatrudnienia. Jak potwierdzają przeprowadzone badania [Wałęga, Wałęga 2015], kredytowanie gospodarstw domowych jest silnie uzależnione od wielkości zamieszkiwanej miejscowości, przy czym osoby z większych miejscowości wykazują większą skłonność do zapożyczania niż osoby z mniejszych miejscowości, w szczególności ze wsi. Znaczący okazuje się także typ gospodarstwa domowego oraz stan cywilny jego głowy. Gospodarstwa domowe osób będących w związkach małżeńskich oraz gospodarstwa posiadające dzieci wykazują większą skłonność do zapożyczania. Ma to związek m.in. z większym zapotrzebowaniem na wydatki wynikające z funkcjonowania rodziny. Cele na jakie gospodarstwa domowe przeznaczają uzyskiwane środki kredytowe zmieniają się w trakcie fazy ich istnienia. We wczesnych etapach cyklu życia rodziny, gospodarstwa wykazują większy popyt na towary oraz chęć życia na wyższym poziomie konsumpcji. Często kredytowanie jest wówczas nieuniknione, gdyż właśnie w tym czasie zapotrzebowanie na środki finansowe jest największe [Świecka, Koziński 2014]. Analizy wskazują, iż jedną z determinant wpływających na zadłużenie jest wiek głowy gospodarstwa. Wiele analiz potwierdza, iż popyt na kredyty rośnie wraz z wiekiem, ale mniej więcej do 30-45 roku życia [Duca, Rosenthal 1993; Crook 2003; Magri 2007; Wałęga 2010]. Stopień wykształcenia głowy gospodarstwa

domowego również ma wpływ na finansowanie konsumpcji ze środków bankowych. Osoby z wyższym jego poziomem cechują się lepszymi uwarunkowaniami na rynku pracy, ponadto są na ogół ostrożniejsze, dzięki czemu mają większe predyspozycje do spłaty zobowiązań, a co za tym idzie – lepszą zdolność kredytową. Poza tym, osoby z wyższym wykształceniem zwykle lepiej rozumieją procesy zachodzące we współczesnej gospodarce, co ma wpływ na większą inklinację do korzystania z rynku kredytowego [Crook 1996; Wałęga 2010; Wałęga, Wałęga 2015].

Oprócz ogólnej sytuacji ekonomicznej występującej w kraju, ważnym czynnikiem współwystępującym przy podejmowaniu decyzji o korzystaniu z usług rynku kredytowego przez gospodarstwa domowe jest obecna kondycja materialna ich członków. Poziom dochodu jest głównym kryterium zadłużenia gospodarstwa domowego i poziomu spłat z tego tytułu. Z jednej strony bieżący dochód stwarza zdolność do zaciągania pożyczek. Z drugiej jednak strony, założenie, że podobny poziom dochodu utrzyma się w kolejnych okresach, zachęca gospodarstwa domowe do przejścia na ścieżkę wyższego spożycia [Wałęga, Wałęga 2015]. Długi gospodarstw mniej zamożnych służą na ogół zaspokajaniu podstawowych potrzeb, do których należą między innymi zaopatrzenie w żywność lub odzież, zapewnienie dostępu do usług zdrowotnych oraz zakup mieszkania. Natomiast gospodarstwa domowe charakteryzujące się średnią lub wysoką zamożnością, zadłużają się w celu wyraźnej poprawy swojego poziomu życia [Bywalec 2012]. Uzyskane środki przeznaczone są głównie na zakup dóbr wyższego rzędu, takich jak luksusowe mieszkania, domy, samochody czy wakacje za granicą. Oznacza to, iż wzrost dochodów wywiera istotny wpływ na strukturę wydatków, przy czym powoduje zmniejszanie wydatków na dobra podstawowe.

Ostatnią wyszczególnianą grupą czynników, determinujących kredytowanie gospodarstw domowych, są psychologiczne uwarunkowania do konsumpcji. Wyróżnia się rozmaite cechy członków gospodarstw domowych skłaniające do użytkowania dóbr. Należą do nich m. in. motyw, postrzeganie świata, osobowość, uczenie się, nawyki, zwyczaje, styl życia oraz skłonności do podejmowania ryzyka [Piekut 2008]. Na decyzję o kredytowaniu również może wywierać wpływ optymistyczny lub pesymistyczny stosunek do życia, ugodowość, niestabilność emocjonalna, czy niska sumienność. W analizach trudno jest uwzględnić przytaczane cechy psychologiczne, dlatego nazywa się je czynnikami nieobserwowalnymi.

WYKORZYSTANE DANE ORAZ METODYKA BADAŃ

W pracy do analizy zjawiska kredytowania gospodarstw domowych wykorzystano dane pochodzące z badania „Diagnoza Społeczna”, opracowanego przez Radę Monitoringu Społecznego [Czapiński, Panek 2015]. W analizie ekonometrycznej wykorzystano dane z czterech ostatnich rund badań przeprowadzonych w latach 2009-2015, uwzględniając dane z 22 340 gospodarstw

domowych. Kwestionariusze badania wypełniały gospodarstwa uczestniczące w poprzednich rundach oraz nowe jednostki wybierane w sposób zapewniający reprezentatywność próby. Dzięki takiemu rozwiązaniu badanie miało charakter panelowy, przy czym pozyskane dane mają charakter panelu niezbilansowanego.

Binarna zmienna objaśniana wykorzystywana w modelowaniu została utworzona na podstawie odpowiedzi na pytanie: „*Czy Pana(i) gospodarstwo domowe ma obecnie do spłacenia pożyczki lub kredyty?*”. Dodatkowo uwzględniono informacje dotyczące zapożyczenia w instytucjach bankowych oraz celu przeznaczenia zapożyczonych środków. Za kredyty na cele konsumpcyjne uznano przypadki, w których otrzymane środki przekazywano na: bieżące wydatki, stałe opłaty (np. mieszkaniowe), zakup dóbr trwałego użytku, kształcenie, leczenie, wypoczynek, zakup papierów wartościowych, zakup lub dzierżawa narzędzi pracy, rozwój działalności gospodarczej, a także spłaty wcześniejszych długów. W ten sposób utworzono binarną zmienną objaśnianą określającą czy gospodarstwo domowe posiada kredyt lub pożyczkę na cele konsumpcyjne w instytucji bankowej, przy czym wartość 1 odpowiada sytuacji, gdy gospodarstwo jest zadłużone, natomiast wartość równa 0 odnosi się do sytuacji gdy gospodarstwo nie posiada żadnego produktu kredytowego.

W realizacji badania wykorzystano panelową strukturę danych umożliwiającą przeprowadzenie bardziej obszernych i wnikliwych analiz niż w przypadku użycia powszechnie stosowanych danych przekrojowych. Główną zaletą danych panelowych jest kontrolowanie jednostek pod kątem ich indywidualnych zachowań. Gospodarstwa domowe charakteryzują się dużą heterogenicznością. Z tego powodu badanie szeregów czasowych i przekrojowych bez rozważania zróżnicowania jednostek wiąże się z ryzykiem uzyskania nieobiektywnych wyników [Hsiao 2003]. Wykorzystywanie metod analizy danych panelowych dla gospodarstw domowych pozwala na uwzględnienie ich nieobserwowalnych efektów indywidualnych [Dudek 2016]. Tego typu dane są w stanie lepiej identyfikować i mierzyć efekty, które są niewykrywalne w przypadku zastosowania wyłącznie danych przekrojowych lub czasowych. Wynika to z faktu iż niektóre cechy charakterystyczne dla jednostek są stałe w czasie. Więcej zalet oraz przykładów potwierdzających atuty stosowania danych panelowych opisano w literaturze [Baltagi 2005].

Ponieważ, podczas analizy czynników skłaniających do zapożyczenia gospodarstw dostrzeżono potencjalną rolę cech ukrytych, postanowiono iż do estymacji zjawiska wykorzystany zostanie model probitowy z efektami losowymi. W pracy oszacowano również „zwykły” model probitowy, a następnie sprawdzono czy uwzględnienie efektów indywidualnych jest uzasadnione oraz czy poprawia jakość estymacji analizowanego zjawiska. Szczegóły dotyczące wykorzystywanych metod badawczych opisano w literaturze [Borkowski i in. 2003; Verbeek 2004; Greene 2012; Gruszczyński 2012; Józwiak, Podgórski 2012; Dudek 2016; Bland, Cook 2019].

WYNIKI ESTYMACJI MODELI

Na podstawie analizy literatury dotyczącej determinant kredytowania na cele konsumpcyjne przystąpiono do wyboru potencjalnych czynników mogących z merytorycznego punktu widzenia wpływać na analizowane zjawisko. W pierwszej kolejności postanowiono rozważyć w jaki sposób najlepiej zaprezentować kondycję materialną. Ponieważ przy porównywaniu dobrobytu gospodarstw domowych, oprócz dochodu należy uwzględnić również ich podstawowe cechy demograficzne [Dudek 2011] w pracy wykorzystano powszechnie stosowaną w Unii Europejskiej skalę ekwiwalentności OECD 50/30. W celu urealnienia uzyskiwanych dochodów w latach 2009–2015 uwzględniono także wskaźnik inflacji, przyjmując rok 2009 za okres bazowy. Taki sposób reprezentacji dochodów gospodarstw domowych zapewnił porównywalność danych w badanym okresie oraz uwzględnił strukturę demograficzną analizowanych jednostek. Pozostałe cechy mogące wpływać na fakt zadłużania się gospodarstw domowych dotyczą informacji na temat głowy gospodarstwa domowego, miejsca zamieszkania, typu biologicznego gospodarstwa oraz roku ankietowania. Skategoryzowano je poprzez zastosowanie zmiennych binarnych, przyjmujących wartość 1, jeśli dany wariant cechy występuje oraz wartość 0 w przeciwnym przypadku. W celu uniknięcia współliniowości pominięto jedną ze zmiennych binarnych odpowiadającą wariantowi referencyjnemu (ref.). Dodatkowo przed przystąpieniem do procesu estymacji zdecydowano się zweryfikować czy rozpatrywane potencjalne zmienne objaśniające nie są zbyt silnie ze sobą skorelowane. Analiza korelacji nie wykazała istnienia silnych zależności, co oznacza, że dla wybranego zbioru cech nie zachodzi zjawisko powielania informacji.

Wyniki estymacji parametrów modelu probitowego z efektami losowymi oraz modelu typu „pooled regression” przedstawiono w tabeli 1. Ponieważ na podstawie estymowanych parametrów można określić jedynie kierunek oddziaływania zmiennych objaśniających, zdecydowano się na wyznaczenie efektów krańcowych, obliczonych na podstawie średnich wartości zmiennych objaśniających.

Tabela 1. Wyniki oszacowań parametrów modeli oraz efektów krańcowych

Zmienna	Model probitowy (pooled regresion)			Model probitowy z efektami losowymi		
	Oszacowanie parametru	Błąd standardowy	Efekty krańcowe	Oszacowanie parametru	Błąd standardowy	Efekty krańcowe
Dochód na jedn. ekwi.	-0,027**	0,011	-0,008	-0,031**	0,015	-0,007
<i>Wiek głowy gospodarstwa domowego</i>						
do 29 lat	0,148**	0,061	0,045	0,166*	0,090	0,039
od 30 do 44 lat	0,07***	0,025	0,022	0,063	0,038	0,015
od 45 do 59	ref.	ref.	ref.	ref.	ref.	ref.
powyżej 60 lat	-0,203***	0,030	-0,062	-0,268***	0,044	-0,064

Zmienna	Model probitowy (pooled regresion)			Model probitowy z efektami losowymi		
	Oszacowanie parametru	Błąd standardowy	Efekty krańcowe	Oszacowanie parametru	Błąd standardowy	Efekty krańcowe
<i>Stan cywilny głowy gospodarstwa domowego</i>						
kawaler / panna	-0,072	0,055	-0,022	-0,154*	0,085	-0,037
mąż / żona	ref.	ref.	ref.	ref.	ref.	ref.
separacja / rozwód	0,123**	0,057	0,038	0,126	0,086	0,030
wdowiec / wdowa	-0,021	0,058	-0,006	-0,049	0,087	-0,012
<i>Wykształcenie głowy gospodarstwa domowego</i>						
podstawowe lub niższe	0,072*	0,038	0,022	0,068	0,062	0,016
zawodowe	0,096***	0,031	0,030	0,108**	0,050	0,026
średnie	0,1***	0,030	0,031	0,12**	0,050	0,029
wyższe lub policealne	ref.	ref.	ref.	ref.	ref.	ref.
<i>Źródło utrzymania głowy gospodarstwa domowego</i>						
pracownicze	0,057*	0,032	0,017	0,068	0,044	0,016
rolnictwo	0,322***	0,040	0,099	0,343***	0,057	0,082
rachunek własny	0,039	0,038	0,012	-0,008	0,054	-0,002
emerytura / renta	0,151***	0,056	0,046	0,164**	0,079	0,039
niezarobkowe	ref.	ref.	ref.	ref.	ref.	ref.
<i>Klasa miejscowości</i>						
bardzo małe miasto	0,107***	0,030	0,033	0,125**	0,052	0,030
małe miasto	0,077***	0,027	0,023	0,084*	0,046	0,020
średnie miasto	0,268***	0,039	0,082	0,355***	0,067	0,084
duże miasto	0,217***	0,036	0,067	0,258***	0,062	0,061
bardzo duże miasto	0,246***	0,041	0,075	0,303***	0,069	0,072
wieś	ref.	ref.	ref.	ref.	ref.	ref.
<i>Typ biologiczny gospodarstwa domowego</i>						
małżeństwo z 1 dziec.	0,140***	0,030	0,043	0,176***	0,045	0,042
małżeństwo z 2 dzieci	0,140***	0,033	0,043	0,169***	0,050	0,040
małżeństwo z 3+ dzieci	0,203***	0,039	0,062	0,282***	0,061	0,067
małżeństwo bez dzieci	ref.	ref.	ref.	ref.	ref.	ref.
jednoosobowe	-0,184*	0,104	-0,057	-0,261*	0,151	-0,062
niepełna rodzina	0,073	0,055	0,023	0,086	0,080	0,021
wielorodzinne	0,267***	0,039	0,082	0,279***	0,059	0,066
wieloosobowe	0,319**	0,162	0,098	0,39*	0,220	0,093
<i>Rok badania</i>						
2009	ref.	ref.	ref.	ref.	ref.	ref.
2011	-0,121***	0,030	-0,037	-0,167***	0,035	-0,040
2013	-0,239***	0,029	-0,073	-0,311***	0,035	-0,074
2015	-0,275***	0,028	-0,084	-0,358***	0,034	-0,085
Stała	-0,780***	0,057	-	-1,021***	0,085	-
σ_u	-	-	-	1,358***	0,035	-
ρ	-	-	-	0,480***	0,013	-
Log wiarygodności	-12109	-	-	-11352	-	-

* oznacza istotność na poziomie 0,1; ** na poziomie 0,05; *** na poziomie 0,01;

σ_u – odch. stand. efektów indywidualnych; ρ – udział wariancji efektów indywidualnych w łącznej wariancji składnika losowego

Źródło: obliczenia własne w SAS oraz R

Przyjmując w pracy poziom istotności 0,05, niezależnie od metody estymacji, stwierdzono statystyczną istotność zmiennej odnoszącej się do realnego dochodu na jednostkę ekwiwalentną. Analiza jakościowych cech wielomianowych

wymagała przeprowadzenia testu logarytmu wiarygodności LR. Na podstawie rezultatów uzyskanych dla obydwu typów modeli potwierdzono istotność każdej z rozpatrywanych jakościowych zmiennych objaśniających (por. tabela 2).

Tabela 2. Wyniki testu logarytmu wiarygodności dla zmiennych jakościowych

Jakościowa zmienna objaśniająca	Liczba stopni swobody	Wartość statystyki LR	
		Model probitowy (pooled regression)	Model probitowy z efektami losowymi
Wiek głowy gosp. domowego	3	76,41***	50,21***
Stan cywilny głowy gosp. domowego	3	51,94***	44,89***
Wykształcenie głowy gosp. domowego	3	56,79***	38,24***
Źródło utrzymania głowy gosp. domowego	4	194,42***	139,33***
Klasa miejscowości	5	85,67***	46,54***
Typ biologiczny gosp. domowego	7	192,13***	153,1***
Rok badania	3	116,8***	126,84***

Źródło: obliczenia własne w SAS oraz R, oznaczenia jak w tabeli 1

Kolejnym etapem analizy była ocena zasadności zastosowania modelu probitowego z efektami losowymi. Stwierdzono, iż wartości oszacowań odchylenia standardowego efektów indywidualnych σ_u oraz udział wariancji efektów indywidualnych w łącznej wariancji składnika losowego ρ są statystycznie istotne (por. tabela 1). Odrzucenie w teście logarytmu wiarygodności hipotezy zerowej oznajmiającej, iż wartość parametru ρ jest równa zeru świadczy o istotności występowania indywidualnych efektów wyrażających nieobserwowalne cechy gospodarstw domowych. Oznacza to, że dla wykorzystywanej próbki, w analizie kredytowania gospodarstw domowych na cele konsumpcyjne, zastosowanie modeli probitowych z efektami losowymi jest bardziej zasadne niż budowa zwykłych modeli probitowych nieuwzględniających panelowej struktury danych. Powyższe wnioski potwierdza również porównanie trafności klasyfikacji zbudowanych modeli (por. tabela 3). Przyjmując wartość progową jako udział wszystkich kredytowanych gospodarstw w całej próbie modelowej, zliczeniowy R^2 jest większy w przypadku wykorzystania efektów indywidualnych o ponad 27,2 p.p. Zastosowanie modelu probitowego typu „pooled regression” skutkuje zatem słabszą jakością dopasowania do danych empirycznych.

Tabela 3. Wyniki trafności klasyfikacji binarnej zmiennej objaśnianej przy zastosowaniu wartości odcinającej adekwatnej do udziału kredytowanych gosp. domowych

Zliczeniowy R^2 (dla wartości progowej $p^* = 0,242$)	Model probitowy (pooled regression)	Model probitowy z efektami losowymi
		43,58%

Źródło: obliczenia własne w SAS oraz R

Mając pewność, że w utworzonych modelach wszystkie zmienne charakteryzujące zjawisko kredytowania są istotne statystycznie przystąpiono

do interpretacji oszacowanych efektów krańcowych, gwarantujących bardziej szczegółową interpretację rezultatów. Ponieważ udowodniono, iż zastosowanie modelu uwzględniającego panelową strukturę danych zapewnia lepsze wyniki estymacji, skupiono się na interpretacji efektów krańcowych tylko dla modelu probitowego z efektami losowymi (por. tabela 1).

Na podstawie uzyskanych rezultatów stwierdzono, iż przy założeniu *ceteris paribus*, wzrost dochodu o 1 tys. zł powodował spadek prawdopodobieństwa zadłużenia średnio o 0,007. Zaistniałe zjawisko może być spowodowane faktem, iż kredyty i pożyczki umotywowane celami konsumpcyjnymi dotyczą głównie gospodarstw mniej zamożnych. W wyniku modelowania otrzymano także rezultat świadczący, iż gospodarstwa, w których głowa rodziny miała wykształcenie podstawowe, zawodowe lub średnie częściej zadłużyły się na cele konsumpcyjne, niż w przypadku gdy osoba prowadząca gospodarstwo ukończyła różnego rodzaju studia. Warto nadmienić iż poziom wykształcenia w dużej mierze określa uwarunkowania na rynku pracy, a wyniki uzyskane dla cech określających poziom wykształcenia są spójne z rezultatami otrzymanymi dla realnego dochodu na jednostkę ekwiwalentną. Godne uwagi jest również oszacowanie parametru odnoszącego się do utrzymywania głowy gospodarstwa domowego z rolnictwa. Przy założeniu *ceteris paribus*, w porównaniu do źródła niezarobkowego, prawdopodobieństwo kredytowania rolników było przeciętnie większe o 0,082. Zaobserwowane zjawisko może być spowodowane faktem, iż wiele rodzin dziedziczących z pokolenia na pokolenie gospodarstwo rolne wraz z domem nie jest zmuszone do zaciągania kredytów hipotecznych, a odnowienie bądź przebudowa domów odbywa się wówczas ze środków kredytów konsumpcyjnych. Zaobserwowano także, że w porównaniu do gospodarstw domowych prowadzonych przez osoby w stanie wolnym, większe prawdopodobieństwo zadłużenia cechowało gospodarstwa, w których głowa rodziny była żonata bądź zamężna. Kategorie wiekowe głowy gospodarstwa domowego odniesiono do wieku w przedziale od 45 do 59 lat. W porównaniu do grupy referencyjnej, wiek poniżej 29 roku życia wpływał na większe prawdopodobieństwo zadłużenia w instytucji bankowej. Wynik ten może mieć związek z większymi potrzebami konsumpcyjnymi wynikającymi z założenia rodziny. W pracy udowodniono również, iż wielkość miejscowości zamieszkania istotnie różnicuje prawdopodobieństwo zaciągnięcia kredytów na cele konsumpcyjne. Dowiedziono, że w odniesieniu do wiejskich gospodarstw domowych, największy wzrost prawdopodobieństwa kredytowania występuje w miastach, w których liczba ludności mieści się w przedziale od 100 – 200 tys. mieszkańców. Ostatnią rozpatrywaną cechą charakteryzującą gospodarstwa był ich typ biologiczny. Zaobserwowano, że przy założeniu *ceteris paribus* w stosunku do gospodarstw małżeństw bez dzieci, prawdopodobieństwo zadłużenia dla małżeństw z co najmniej dwójką dzieci było większe średnio o 0,067. W modelu uwzględniono także informacje o roku badania. Wyniki oszacowań parametrów potwierdziły, iż z roku na rok liczba gospodarstw domowych

kredytowanych na cele konsumpcyjne spadała. Przy założeniu *ceteris paribus*, w stosunku do roku 2009, w roku 2015 prawdopodobieństwo zadłużenia na cele konsumpcyjne było mniejsze średnio o 0,085.

PODSUMOWANIE

Zawieranie umów kredytowych na cele konsumpcyjne umożliwia gospodarstwom domowym zaspokajanie aktualnych potrzeb oraz poprawia ich ogólną sytuację majątkową. W pracy podjęto się identyfikacji determinant wpływających na zaciąganie pożyczek i kredytów konsumpcyjnych przez gospodarstwa domowe oraz weryfikacji hipotezy, że uwzględnienie panelowej struktury danych jest uzasadnione w modelowaniu rozpatrywanego zjawiska. W tym celu wykorzystano dane z badania „Diagnoza Społeczna” z lat 2009–2015. Wyniki z przeprowadzonej analizy wykazały, iż zastosowanie modelu probitowego z efektami losowymi znacząco poprawia jakość klasyfikacji w porównaniu do modelu typu „pooled regression”. Dodatkowo potwierdzono istotność występowania indywidualnego efektu, reprezentującego nieobserwowalną heterogeniczność. Warto przy tym nadmienić, że w w większości prac odnoszących się do modelowania zjawiska kredytowania polskich gospodarstw domowych stosowano regresję logistyczną, której parametry estymowano na podstawie danych przekrojowych. Ponadto należy podkreślić, że w dotychczasowych opracowaniach z tego zakresu nie uwzględniano zawierania umów wyłącznie na cele konsumpcyjne. Dlatego też niniejsza praca stanowi ważne uzupełnienie dotychczasowych analiz.

BIBLIOGRAFIA

- Baltagi B. (2005) *Econometric Analysis of Panel Data*. John Wiley & Sons Ltd, Wyd. 3, Chippenham, Wiltshire.
- Bieniasz A., Czerwinska-Kayzer D. (2010) Zadłużenie gospodarstw domowych w sektorze bankowym w Polsce. *Journal of Agribusiness and Rural Development*, 1(15), 17-28.
- Bland J. R., Cook A. (2019) Random Effects Probit and Logit: Understanding Predictions and Marginal Effects. *Applied Economics Letters*, 26(2), 116-123.
- Borkowski B., Dudek H., Szczesny W. (2003) *Ekonometria. Wybrane zagadnienia*. Wydawnictwo Naukowe PWN, Warszawa.
- Bywalec C. (2012) *Ekonomika i finanse gospodarstw domowych*. Wydawnictwo Naukowe PWN, Wyd. 2, Warszawa.
- Crook J. (1996) Credit Constraints and US Households. *Applied Financial Economics*, 6(6), 477-485.
- Crook J. (2003) *The Demand and Supply for Household Debt: A Cross Country Comparison*. Credit Research Centre, University of Edinburgh.
- Crook J., Hochguertel S. (2007) *US and European Household Debt and Credit Constraints: Comparative Micro Evidence from the Last 15 Years*. Tinbergen Institute Discussion Paper, 2007-087/3.

- Czapiński J., Panek T. (red.) (2015) *Diagnoza społeczna 2015. Warunki i jakość życia Polaków*. www.diagnoza.com [dostęp w dniu 11.01.2018].
- Diagnoza społeczna: zintegrowana baza danych. www.diagnoza.com [dostęp w dniu 11.01.2018].
- Duca J., Rosenthal S. (1993) Borrowing Constraints, Household Debt, and Racial Discrimination in Loan Markets. *Journal of Financial Intermediation*, 3(1), 77-103.
- Dudek H. (2016) Ocena trudności finansowych w zakresie nabywania żywności - zastosowanie modelu probitowego ze zmiennymi efektami. *Metody Ilościowe w Badaniach Ekonomicznych*, 17(1), 31-41.
- Dudek H. (2011) Skale ekwiwalentności - estymacja na podstawie kompletnych modeli popytu. Wydawnictwo SGGW, Warszawa.
- Folwarski M. (2016) Wpływ kredytów bankowych na wzrost gospodarczy w Polsce. *Zarządzanie Finansami i Rachunkowość*, 4(3), 5-14.
- Główka G. (2009) Instrumenty finansowe a rynek mieszkaniowy. *Problemy Rozwoju Miast*, 3, 11-26.
- Greene W. H. (2012) *Econometric Analysis. Seventh Edition*. Pearson Education Limited, Harlow.
- Gruszczynski M. (2012) *Mikroekonometria. Modele i metody analizy danych indywidualnych*. Wolters Kluwer Polska Sp. z o.o, Wyd. 2, Warszawa.
- Grzybowska A. (2012) Gospodarstwo domowe jako podmiot współczesnej bankowości. *Zeszyty Naukowe Wyższej Szkoły Bankowej we Wrocławiu*, 31, 79-89.
- Harasim J. (2005) Bankowość detaliczna w Polsce. CeDeWu, Warszawa, 49-55.
- Hsiao C. (2003) *Analysis of Panel Data*. Cambridge University Press, Wyd. 2.
- Jóźwiak J., Podgórski J. (2012) *Statystyka od podstaw*. PWE, Warszawa, Wyd. 7.
- Magri S. (2007) Italian Households' Debt: the Participation to the Debt Market and the Size of the Loan. *Empirical Economics*, 33(3), 401-426.
- Malchar J., Zielińska-Sitkiewicz M. (2017) Metody klasyfikacji w analizie porównawczej rozwoju społeczno-gospodarczego polskich województw w latach 2010 i 2014 - wpływ procedury normalizacji na wynik rankingu. *Metody Ilościowe w Badaniach Ekonomicznych*, 18(4), 643-652.
- Nieto F. (2007) The Determinants of Household Credit in Spain. *Documentos de Trabajo*, 0716, Banco de España.
- Paczóska A. (2008) Rynek kredytów hipotecznych w Polsce. *Zeszyty Naukowe / SGH, Kolegium Gospodarki Światowej*, 23, 261-274.
- Piekut M. (2008) Polskie gospodarstwa domowe: dochody, wydatki i wyposażenie w dobra trwałego użytkowania. Wydawnictwo SGGW, Warszawa.
- Rusnak Z. (2011) Regresja logistyczna a analiza oszczędności i zobowiązań finansowych gospodarstw domowych. *Zeszyty Naukowe Wyższej Szkoły Bankowej we Wrocławiu*, 25, 169-188.
- Skikiewicz R. (2014) Determinanty popytu na usługi bankowe na przykładzie kredytów konsumpcyjnych i mieszkaniowych w latach 2004-2013. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 353, 54-65.
- Stone B., Maury R. V. (2006) Indicators of Personal Financial Debt Using a Multi-Disciplinary Behavioral Model. *Journal of Economic Psychology*, 27, 543-556.

- Świecka B., Koziński M. (2014) Rola perspektywy czasowej w zadłużeniu gospodarstw domowych. Ujęcie behawioralne. *Zeszyty Naukowe Uniwersytetu Szczecińskiego, Problemy zarządzania, finansów i marketingu*, 34, 217-230.
- Utzig M. (2012) Zadłużenie gospodarstw domowych w monetarnych instytucjach finansowych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu. Ekonomia*, 245, 579-589.
- Verbeek M. (2004) *A Guide to Modern Econometrics*. Wiley & Sons, Ltd, Chichester.
- Wałęga G. (2010) Determinanty zadłużenia gospodarstw domowych w Polsce w świetle wybranych teorii konsumpcji. [w:] Dach Z. (red.) *Otoczenie ekonomiczne a zachowania podmiotów rynkowych*. Polskie Towarzystwo Ekonomiczne, Kraków, 114-133.
- Wałęga G. (2013) *Kredytowanie gospodarstw domowych*. Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Wałęga G., Wałęga A. (2015) Determinants of Loan Repayments: Evidence from Household Budget Survey in Poland. *Proceedings of 9th International Days of Statistics and Economics, MELANDRIUM*, Prague, 1774-1783.
- Waliszewski K. (2007) Działalność instytucji pośrednictwa kredytowego na rynku kredytów konsumpcyjnych w Polsce. *Zeszyty Studiów Doktoranckich, Akademia Ekonomiczna w Poznaniu, Wydział Ekonomii*, 31, 49-71.

MICROECONOMETRIC MODELING OF CONCLUDING CONSUMER CREDITS BY HOUSEHOLDS

Abstract: The purpose of the article is to identify factors affecting the willingness to take loans and credits for consumption by households and to verify the validity of using the probit model with random effects. Based on the analysis of data from the "Social Diagnosis" survey, it is found which determinants influence the analyzed phenomenon. The conclusions from the analysis clearly indicate the appropriateness of using the panel data structure in modeling, which allows to specify the unobservable effects of individual analyzed households.

Keywords: consumer credit, loan, households, panel data, random effects probit model

JEL classification: H31, C23, C25

ANALIZA CZASU TRWANIA NA RYNKU NA PRZYKŁADZIE PRZEDSIĘBIORSTW Z WOJEWÓDZTWA MAZOWIECKIEGO

Dominika Urbańczyk  <https://orcid.org/0000-0002-5134-4693>

Wydział Zastosowań Informatyki i Matematyki
Szkoła Główna Gospodarstwa Wiejskiego w Warszawie
e-mail: dominika_urbanczyk@sggw.pl

Streszczenie: Praca poświęcona jest analizie czasu trwania przedsiębiorstw na rynku. Próbę badawczą stanowiła obserwowana przez 10 lat kohorta firm z województwa mazowieckiego, które rozpoczęły działalność w 2007 r. Wykorzystano nieparametryczne oraz parametryczne metody analizy przeżycia, w tym modele hazardu. Zwrócono również uwagę na problem cenzurowania danych. Wyniki uzyskane w obu podejściach są zbieżne. Zauważono, że największe ryzyko zakończenia działalności występuje pomiędzy 2 a 5 rokiem istnienia firmy. Istotny wpływ na czas trwania przedsiębiorstw mają takie czynniki jak siedziba, forma własności, a także sekcja działalności według PKD.

Słowa kluczowe: analiza przeżycia, nieparametryczne i parametryczne modele hazardu, czas trwania przedsiębiorstw na rynku, dane cenzurowane, estymator Kaplana-Meiera

JEL classification: C4

WSTĘP

Kondycja firm jest kluczowym aspektem w ocenie sytuacji ekonomicznej i planowaniu polityki gospodarczej. Rozwój przedsiębiorczości jest korzystny nie tylko ze względu na poprawę wyników finansowych, ale także – przyczyniając się do większej aktywności zawodowej – wspomaga rozwiązywanie problemu bezrobocia. Nietrywialnym problemem jest również prawidłowa ocena sytuacji przedsiębiorstw w kontekście ich zdolności kredytowej lub też konstrukcji oferty ubezpieczeń. Tłumaczy to zainteresowanie zagadnieniem przetrwania firm na rynku [Boratyńska 2013; Dominiak, Mazurkiewicz 2011; Hadasik 1998; Kisielińska 2016;

<https://doi.org/10.22630/MIBE.2019.20.2.11>

Korol, Prusak 2015; Markowicz, Bieszk-Stolorz 2006; Markowicz 2016; Mikulec 2017; Ptak-Chmielewska 2012; Ptak-Chmielewska 2016].

Biorąc pod uwagę znaczenie sytuacji firm dla całej gospodarki, niepokojące wydają się publikowane przez Główny Urząd Statystyczny wyniki wskazujące na to, że duża część spośród małych i średnich przedsiębiorstw (MŚP) nie jest w stanie przetrwać na rynku pierwszych pięciu lat. W celu wprowadzenia skutecznych rozwiązań konieczne jest zdobycie rzetelnej wiedzy.

Zdolność prawidłowego prognozowania czasu trwania przedsiębiorstwa na rynku jest niewątpliwie istotnym problemem badawczym. Jednak konieczne jest również poznanie czynników warunkujących zdolność firm do kontynuowania działalności gospodarczej. Warty uwagi jest również problem identyfikacji okresu cyklu życia przedsiębiorstwa, w którym ryzyko zakończenia działalności jest największe. Kolejnym bardzo ważnym aspektem jest zapotrzebowanie na wiedzę o cechach populacji firm.

Podjmując temat przetrwania przedsiębiorstw na rynku należy mieć jednak świadomość występujących w tego rodzaju badaniach ograniczeń i trudności. Należy zwrócić uwagę na charakter danych dotyczących wszystkich czasów trwania. Przede wszystkim występuje tutaj problem niepełnej informacji oraz konieczność pracy z danymi uciętymi i cenzurowanymi. Brak wstępnej wiedzy na temat badanej populacji przedsiębiorstw uniemożliwia sformułowanie założeń *a priori* dotyczących rozkładu czasu trwania. Dodatkowo należy brać pod uwagę, że nietypowe indywidualne cechy firm (np. szczególne warunki otoczenia podmiotu) mogą mieć zaburzający wpływ na wynik i wnioski dotyczące całej populacji przedsiębiorstw.

Celem pracy jest zbadanie czasu trwania dla populacji firm z województwa mazowieckiego z wykorzystaniem podejścia opartego na analizie przeżycia – w tym modelach hazardu. Tego rodzaju metody statystyczne są zaprojektowane do badania nieujemnych zmiennych losowych, a w szczególności czasów trwania. Ich główną zaletą w stosunku do innych stosowanych w tego rodzaju badaniach metod jest możliwość uwzględnienia występowania danych cenzurowanych i uciętych, co pozwala na uzyskanie wyników obarczonych mniejszym błędem.

PROBLEM CENZUROWANIA DANYCH

Niezależnie od aspektów będących przedmiotem zainteresowania, w pracach poświęconych badaniu czasów trwania (czyli czasu pomiędzy wystąpieniem określonych zdarzeń) pojawia się problem uwzględniania obserwacji, o których nie ma pełnej informacji. Mówimy wtedy o obserwacjach cenzurowanych i uciętych. Należy tutaj głównie zwrócić uwagę na przypadki, kiedy jedno ze zdarzeń ma miejsce przed rozpoczęciem obserwacji (lewostronne cenzurowanie), badanie zakończy się przed wystąpieniem zdarzenia końcowego, bądź też w trakcie prowadzenia badań stracimy możliwość obserwacji danego obiektu (cenzurowanie

prawostronne). W przypadku, kiedy występują oba zjawiska mówimy o cenzurowaniu obustronnym [Sokołowski 2010].

W takich sytuacjach nie możemy dokładnie określić czasu trwania dla danego epizodu, ale mamy częściową informację o tym, że w populacji wystąpiły takie zdarzenia i dla tych obserwacji możemy podjąć próbę określenia dolnej granicy czasu trwania. Zupełne pominięcie tego rodzaju obserwacji wydaje się traceniem cennych informacji o populacji. Z drugiej strony nie można uciętego czasu trwania traktować w sposób równoznaczny z dokładnie zmierzonymi czasami trwania dla innych obserwacji. Pozostaje więc pytanie jak radzić sobie z danymi cenzurowanymi i uciętymi, czy uwzględniać je w badaniach i jeśli tak to w jaki sposób.

W przypadku prowadzenia badań kohortowych, kiedy badanie danej populacji rozpoczyna się w momencie początku czasu trwania, eliminujemy problem cenzurowania lewostronnego. Natomiast nadal występują obserwacje cenzurowane prawostronnie (patrz rysunek 1). Wybrana metoda badawcza powinna więc umożliwiać uwzględnienie tego rodzaju zjawiska.

Rysunek 1. Czas trwania cenzurowany prawostronnie



Źródło: opracowanie własne

ANALIZA PRZEŻYCIA

Analiza przeżycia jest zbiorem metod statystycznych opracowanych do badania nieujemnych zmiennych losowych, w szczególności czasów trwania. Niech badany czas trwania będzie zmienną losową T , przyjmującą wartości z przedziału $(0, \infty)$. Wówczas funkcję S nazywamy funkcją przeżywalności i definiujemy następująco:

$$S(t) = P(T > t) = 1 - P(T \leq t) = 1 - F(t) = \int_t^{\infty} f(z) dz, \quad (1)$$

gdzie F jest dystrybuantą, a f funkcją gęstości prawdopodobieństwa zmiennej losowej T . Bezpośredni związek funkcji przeżywalności z dystrybuantą i funkcją gęstości prawdopodobieństwa wskazuje na to, że wyznaczenie funkcji S pozwala w pełni określić rozkład zmiennej losowej oraz stosować metody wnioskowania statystycznego w odniesieniu do tej funkcji [Klein i Moeschberger 2003].

Dla znanej funkcji przeżywalności wartość oczekiwaną zmiennej losowej T , czyli oczekiwany czas trwania definiuje się w następujący sposób:

$$E(T) = \int_0^{\infty} S(t) dt. \quad (2)$$

Ryzyko zajścia oczekiwanego zdarzenia końcowego wzdłuż rozkładu zmiennej T , może być opisywane poprzez analizę zmienności funkcji hazardu $h(t)$ (inaczej funkcji intensywności procesu) [Sokołowski 2010]. Jest ona wyrażona poprzez granicę prawdopodobieństwa warunkowego w następujący sposób [Klein i Moeschberger 2003]:

$$h(t) = \lim_{dt \rightarrow 0} \frac{P(t \leq T < t + dt | T \geq t)}{dt}. \quad (3)$$

Interpretacja wartości funkcji wykorzystywanych w analizie przeżycia w kontekście zagadnienia przetrwania przedsiębiorstw na rynku jest następująca: dla danego czasu t ryzyko wyjścia przedsiębiorstwa z rynku jest tym większe, im większą wartość przyjmuje funkcja hazardu $h(t)$. Natomiast prawdopodobieństwo, że firma będzie prowadziła działalność dłużej niż czas t jest tym większe, im większa jest wartość funkcji przeżywalności $S(t)$.

Tablice trwania życia

Jedną z najstarszych i najbardziej podstawowych metod analizy przeżycia jest metoda tablic trwania życia. W tej metodzie wartości czasu trwania dzieli się na przedziały jednakowej długości. Dla każdego z takich przedziałów można wyznaczyć liczbę obserwacji na początku przedziału (n_i) oraz liczbę obserwacji zakończonych (z_i) i uciętych (c_i) w danym przedziale. Na tej podstawie dla kolejnych przedziałów wyznacza się warunkowe prawdopodobieństwa przeżycia $\frac{n_{i+1}}{n_i}$, gdzie dla każdego t : $n_{i+1} = n_i - z_i - c_i$.

Estymator Kaplana-Meiera funkcji przeżywalności

Estymator Kaplana-Meiera jest nieparametrycznym estymatorem największej wiarygodności dla funkcji przeżywalności. Metoda ta nie wymaga konstruowania podziału czasu obserwacji na klasy. Niech T będzie zmienną losową opisującą czas trwania. Wówczas dla analogicznych oznaczeń jak w przypadku tablic trwania życia estymator jest postaci:

$$\hat{S}(t) = \prod_{j=1}^i P(T > t_j | T \geq t_{j-1}) = \prod_{j=1}^i \hat{p}_j = \prod_{j=1}^i \left(1 - \frac{z_j}{n_j} \right). \quad (4)$$

Estymator pozwala na uzyskanie funkcji przeżywalności bez wcześniejszej znajomości rozkładu. Wyodrębnienie podgrup wewnątrz próby może być

zrealizowane poprzez estymację funkcji przeżywalności dla każdej z grup oddzielnie i porównanie ich postaci.

Modele regresyjne

Modele regresyjne wymagają zastosowania założeń *a priori* dotyczących rozkładu zmiennej losowej. Ich zaletą jest jednak możliwość włączenia do modelu zmiennych objaśniających, co pozwala na ilościową ocenę wpływu poszczególnych czynników na czas trwania. W ogólnej postaci model regresyjny dla funkcji przeżywalności jest postaci:

$$\hat{S}(t) = P(T > t | Z) = \hat{S}_0(t) \exp(\hat{\beta}Z), \quad (5)$$

gdzie $\hat{S}_0(t)$ jest bazową funkcją przeżycia, Z jest wektorem zmiennych objaśniających, a β wektorem parametrów.

WYNIKI BADAŃ

Wykorzystana baza danych

W pracy wykorzystano dane z rejestru REGON (prowadzonego przez Główny Urząd Statystyczny) dla przedsiębiorstw z województwa mazowieckiego, które rozpoczęły działalność w 2007 r. Obserwację zakończono na koniec 2017 roku. Badania zostały zrealizowane dla całej populacji, która zawierała 32 788 przedsiębiorstw. Spośród informacji zawartych w rejestrze REGON na potrzeby badania wybrano następujące charakterystyki dla każdej z firm:

Tabela 1. Zmienne wykorzystane w modelach

Nazwa zmiennej	Opis i wartości
data_roz poczecia	data dzienna rozpoczęcia działalności
data_zakonczenia	data dzienna zakończenia działalności
siedziba	adres siedziby firmy (na potrzeby badania przyjęto Warszawa dla siedziby w Warszawie, inne – dla pozostałych obszarów woj. mazowieckiego)
wlasnosc	forma własności (os. fizyczna, os. prawna, jedn. org. bez os. prawnej)
PKD2007	sektor działalności (wartości według głównych sekcji PKD 2007)

Źródło: opracowanie własne

Metody nieparametryczne

Analizę czasu trwania przedsiębiorstw na rynku rozpoczęto od zastosowania metod nieparametrycznych ze względu na brak konieczności stosowania założeń *a priori*, co było bardzo istotne, gdyż na początku badania Autor nie dysponował wiedzą na temat rozkładu czasu trwania.

W pierwszym etapie skonstruowano tablicę trwania dla firm w województwie mazowieckim (tabela 2).

Tabela 2. Tablica trwania firm w woj. mazowieckim (czas podano w miesiącach)

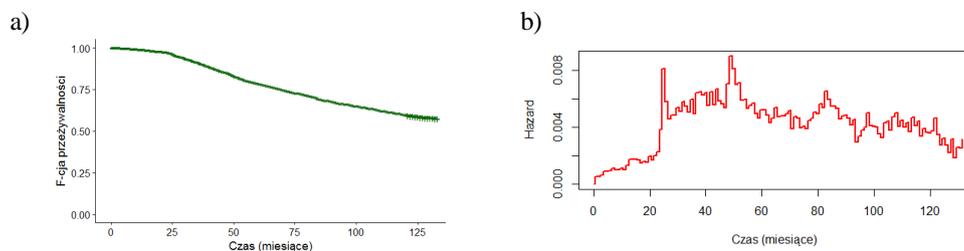
t	n _t	z _t	c _t	\hat{S}_t
0	32788	113	0	0,997
12	32367	42	0	0,986
24	31712	121	0	0,963
36	29729	146	0	0,902
48	27677	195	0	0,838
60	25648	133	0	0,778
72	24209	114	0	0,735
84	22850	135	0	0,693
96	21605	81	0	0,656
108	20574	102	0	0,624
120	19579	70	0	0,595
121	19509	7	32	0,593
122	19405	86	1345	0,590
126	12515	32	1460	0,583

Źródło: opracowanie własne na podstawie danych z REGON

Warto zauważyć, że ze względu na specyfikę próby badawczej (kohorta firm rozpoczynających działalność w 2007 roku) pierwsze obserwacje cenzurowane pojawiają się dla firm, które przetrwały minimum 10 lat, czyli dotrwały od 2007 roku do zakończenia obserwacji.

Następnie metodą Kaplana-Meiera oszacowano postać funkcji przeżywalności oraz hazardu dla całej próby przedsiębiorstw (rysunek 2).

Rysunek 2. Funkcja przeżywalności (a) i hazard (b) oszacowane metodą Kaplana-Meiera



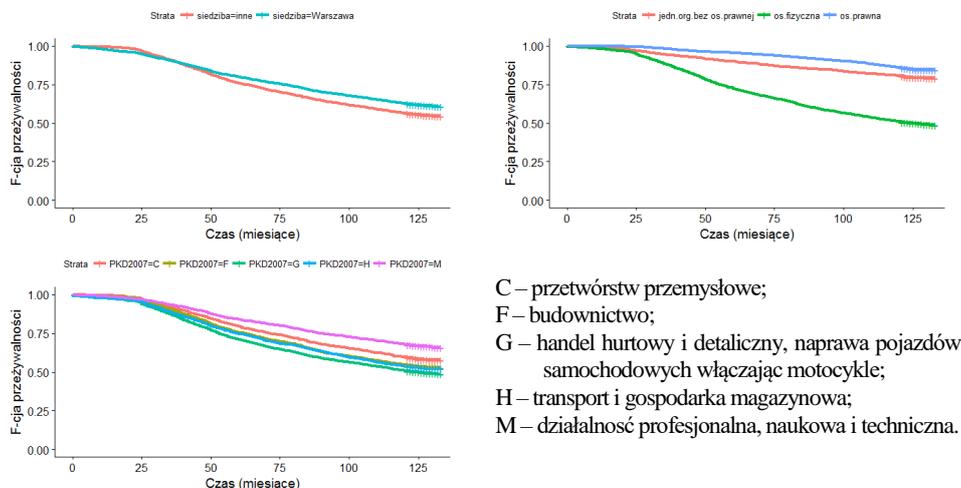
Źródło: opracowanie własne na podstawie danych z REGON

Widać, że w początkowym okresie trwania (do około 24 miesiąca) niewielka część przedsiębiorstw kończy działalność. Natomiast zarówno na podstawie wykresu funkcji przeżywalności jak i funkcji hazardu można zauważyć, że największe ryzyko zakończenia działalności występuje w okresie od 2 do 5 roku trwania firmy, po czym zaczyna maleć.

W kolejnym etapie badania w celu dokładniejszej analizy wpływu cech przedsiębiorstwa na ich zdolność przetrwania na rynku, przeprowadzono estymację funkcji przeżywalności dla podgrup. Jako czynniki podziału zastosowano siedzibę przedsiębiorstwa, formę własności oraz sektor działalności według PKD (rysunek 3). Prezentowane są wyniki dla wybranych pięciu sekcji działalności według PKD

2007 o dużej liczności i znacząco różnych krzywych przeżycia: C – przetwórstwo przemysłowe, F – budownictwo, G – handel hurtowy i detaliczny oraz naprawa pojazdów samochodowych (włączając motocykle), H – transport i gospodarka magazynowa, M – działalność profesjonalna, naukowa i techniczna. Wyniki nieparametrycznych testów na równość rozkładów [Sokołowski 2010] (zastosowano testy typu log-rank, test Wilcozona oraz test Fleminga-Harringtona) potwierdziły, że rozkłady dla podgrup wyodrębnionych ze względu na te zmienne są istotnie różne.

Rysunek 3. Funkcja przeżywalności dla podgrup



Źródło: opracowanie własne na podstawie danych z REGON

Metody parametryczne

Informacje na temat postaci funkcji hazardu uzyskane na podstawie wyników modeli nieparametrycznych, wykorzystano podczas estymacji funkcji przeżywalności i hazardu metodami parametrycznymi. W pierwszym etapie oszacowano modele w podstawowej postaci bez uwzględnienia zmiennych objaśniających. Ze względu na to, że oszacowana w modelach nieparametrycznych funkcja hazardu ma charakter niemonotoniczny zastosowano rozkład logarytmiczno-logistyczny, logarytmiczno-normalny oraz gamma. Uzyskano następujące wyniki:

Rozkład logarytmiczno-logistyczny	$S(t) = \frac{1}{1 + \lambda t^\alpha}$	$\alpha = 1,419$ $\lambda = 158,661$	$\ln L = -88358,24$ $AIC = 176720,5$
Rozkład logarytmiczno-normalny	$S(t) = 1 - \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right)$	$\mu = 5,241$ $\sigma = 1,494$	$\ln L = -89820,75$ $AIC = 179645,5$
Rozkład gamma	$S(t) = 1 - \int_0^t u^{\beta-1} \exp(-u) du / \Gamma(\beta), \quad u = \lambda t$	$\beta = 1,326$ $\lambda = 0,007$	$\ln L = -88464,15$ $AIC = 176932,3$

Następnie, biorąc pod uwagę kryteria dopasowania modelu (najniższa wartość AIC oraz najwyższa wartość $\ln L$), dla rozkładu logarytmiczno-logistycznego oszacowano model hazardu z uwzględnieniem zmiennych objaśniających. Dla tego rozkładu funkcja hazardu jest opisana wzorem: $h(t) = \frac{\lambda \alpha t^{\alpha-1}}{1 + \lambda t^\alpha}$. W modelu regresyjnym jest postaci $\hat{h}(t) = \hat{h}_0(t) \exp(\hat{\beta}X)$, gdzie h_0 jest hazardem bazowym, β – wektorem parametrów, a X – wektorem zmiennych objaśniających. W celu estymacji modelu dokonano analogicznego jak w podejściu nieparametrycznym wyboru sekcji działalności według PKD. Uzyskano następujące parametry oszacowanego modelu:

Tabela 3. Estymatory parametrów dla oszacowanego modelu hazardu

Zmienna	Parametr	$\exp(\beta)$
α	1,5511*	-
λ	324,9659*	-
siedziba = Warszawa	0,0848***	1,08850
własność = os.fizyczna	0,9007***	2,46133
własność = os.prawna	-0,3606***	0,69726
PKD2007= F	0,1728***	1,18863
PKD2007= G	0,2493***	1,28313
PKD2007= H	0,0600*	1,06184
PKD2007= M	-0,2479***	0,78044

Grupy referencyjne dla cech to siedziba = inne, własność = jedn. org. bez os. prawnej, PKD2007 = C). AIC = 116220, $\ln L = -58101$, ozn. poziomu istotności: * to 0,1, ** to 0,01, *** to 0,001

Źródło: opracowanie własne na podstawie danych z REGON

Ze względu na spodziewany różny wpływ poszczególnych sekcji działalności w zależności od siedziby przedsiębiorstwa, oszacowano również osobne modele dla firm z Warszawy i pozostałej części województwa mazowieckiego.

Tabela 4. Estymatory parametrów dla modelu hazardu oszacowanego dla firm z Warszawy

Zmienna	Parametr	$\exp(\beta)$
α	1,44522*	-
λ	182,19801*	-
PKD2007= F	-0,04074*	0,96008
PKD2007= G	0,32635***	1,38590
PKD2007= H	0,40123***	1,49366
PKD2007= M	-0,10716**	0,89838

Grupa referencyjna to PKD2007 = C. AIC = 46169, $\ln L = -23078$, oznaczenie poziomu istotności: * to 0,1, ** to 0,01, *** to 0,001

Źródło: opracowanie własne na podstawie danych z REGON

Tabela 5. Estymatory parametrów dla modelu hazardu oszacowanego dla firm poza Warszawą

Zmienna	Parametr	$\exp(\beta)$
α	1,6563*	-
λ	141,8764*	-
PKD2007= F	0,2046***	1,22703
PKD2007= G	0,2145***	1,23924
PKD2007= H	0,0110	1,01106
PKD2007= M	-0,2870***	0,75051

Grupa referencyjna to PKD2007 = C.

AIC = 70889, lnL = -35438, ozn. poziomu istotności: * to 0,1, ** to 0,01, *** to 0,001.

Źródło: opracowanie własne na podstawie danych z REGON

Zarówno w przypadku podejścia nieparametrycznego jak i oszacowanych modeli regresyjnych zaobserwowano różnice pomiędzy czasem trwania dla firm z siedzibą w Warszawie i poza granicami administracyjnymi miasta stołecznego. W pierwszych miesiącach prowadzenia działalności większą przeżywalnością charakteryzują się firmy spoza stolicy, natomiast w późniejszym okresie zauważalnie wyższą przeżywalność obserwowano dla firm z Warszawy (rysunek 3). Znacznie „bezpieczniejsze na rynku” są przedsiębiorstwa z osobowością prawną, co może być związane z tym, że są to większe firmy, często z sektora publicznego. Natomiast szczególnie narażone na ryzyko zakończenia działalności są raczej małe i średnie przedsiębiorstwa. W całym województwie mazowieckim wśród analizowanych sektorów największą zdolnością przetrwania charakteryzują się firmy zajmujące się działalnością profesjonalną, naukową i techniczną (tabela 3: dla tej zmiennej najniższa i mniejsza od 1 wartość $\exp(\beta)$, czyli ilorazu hazardu w przypadku rozważania wpływu prowadzenia działalności w tym sektorze przy założeniu *ceteris paribus*). Natomiast najbardziej narażoną na ryzyko sekcją działalności w Warszawie jest transport i gospodarka magazynowa (tabela 4), podczas gdy poza Warszawą nie zaobserwowano istotnego wpływu prowadzenia działalności tego typu na zdolność przetrwania na rynku (tabela 5). W pozostałej części województwa największe ryzyko zaobserwowano dla firm z sektora handlu, naprawy pojazdów samochodowych oraz budownictwa.

PODSUMOWANIE

Sytuacja firm staje się obecnie coraz istotniejszym zagadnieniem. Celem pracy była analiza czasu trwania przedsiębiorstw na rynku na przykładzie podmiotów gospodarczych z województwa mazowieckiego. Zrealizowano go przy wykorzystaniu metod analizy przeżycia, w szczególności modeli hazardu. Porównano wyniki podejścia nieparametrycznego oraz parametrycznego. Należy zauważyć, że istotną zaletą wybranych metod była możliwość uwzględnienia cenzurowania czasów trwania, co stanowi istotny problem w tego rodzaju badaniach.

Zastosowane metody nieparametryczne pozwoliły na oszacowanie postaci funkcji przeżywalności i hazardu. Na tej podstawie zauważno, że największe ryzyko zakończenia działalności występuje pomiędzy 2 a 5 rokiem istnienia przedsiębiorstwa, przy czym obserwuje się wyraźny wzrost tego ryzyka po zakończeniu 24 miesiąca. Biorąc pod uwagę duży udział w populacji podmiotów z sektora MŚP, można wnioskować, że wzrost ten jest powiązany z końcem preferencyjnego okresu dla nowych firm (np. obniżonej składki ZUS).

Zastosowane metody badawcze pozwoliły na zbadanie czasu trwania na rynku przedsiębiorstw z województwa mazowieckiego. Wyniki metod nieparametrycznych wykorzystano w dalszej części pracy, przy wyborze rozkładu podczas konstrukcji modeli parametrycznych. Wnioski sformułowane na podstawie metod nieparametrycznych dotyczące wpływu zmiennych na postać funkcji przeżywalności i hazardu zostały potwierdzone i uzupełnione przez ilościową analizę wykonaną za pomocą metod parametrycznych. Istotnymi czynnikami wpływającymi na czas trwania na rynku są siedziba przedsiębiorstwa, forma własności oraz sektor działalności według klasyfikacji PKD.

Dalsze prace badawcze poświęcone będą przeprowadzeniu analiz z uwzględnieniem różnych form zakończenia działalności firmy (modele ryzyk konkurencyjnych), co pozwoli na pełniejsze zrozumienie mechanizmów warunkujących zdolność przedsiębiorstwa do przetrwania na rynku.

BIBLIOGRAFIA

- Boratyńska K. (2013) Selected Methods of Corporate Bankruptcy Risk Assessment – the Cases of the Polish Chosen Meat Sector Companies. *Roczniki Naukowe Stowarzyszenia Ekonomistów Rolnictwa i Agrobiznesu*, XV, 27-32.
- Dominiak P., Mazurkiewicz M. (2011) Analysis of the Risk of Company's Bankruptcy in Polish Food and Beverage Production Sector using the Cox Regression. *Operations Research and Decisions*, 21(1), 19-31.
- Hadasik D. (1998) Upadłość przedsiębiorstw w Polsce i metody jej prognozowania. *Zeszyty Naukowe. Akademia Ekonomiczna w Poznaniu, Seria 2, Prace Habilitacyjne 1230-6673 z. 153*, 187-196.
- Kisielińska J. (2016) Skuteczność modeli upadłości przedsiębiorstw. *Studia Ekonomiczne i Regionalne*, 9(1), 5-17.
- Klein J. P., Moeschberger M. L. (2003) *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York, 21-90.
- Korol T., Prusak B. (2015) Upadłość przedsiębiorstw a wykorzystanie sztucznej inteligencji. *CeDeWu*, Warszawa, 5-126.
- Markowicz I., Bieszk-Stolorz B. (2006) Wykorzystanie analizy historii zdarzeń do konstrukcji tablic żywotności firm. *Wiadomości Statystyczne*, 4, 1-7.
- Markowicz I. (2016) Analiza trwania firm w woj. zachodniopomorskim. *Wiadomości Statystyczne*, 1, 44-61.
- Mikulec A. (2017) Kohortowe tablice trwania przedsiębiorstw w województwie łódzkim – ujęcie kwartalne. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 468, 147-160.

- Ptak-Chmielewska A. (2012) Wykorzystanie modeli przeżycia i analizy dyskryminacyjnej do oceny ryzyka upadłości przedsiębiorstw. *Ekonometria*, 4, 157-172.
- Ptak-Chmielewska A. (2016) Determinanty przeżywalności mikro i małych przedsiębiorstw w Polsce. Oficyna Wydawnicza SGH. Warszawa, 10-200.
- Sokołowski A. (2010) Jak rozumieć i wykonywać analizę przeżycia. *StatSoftPolska*, www.statsoft.pl/czytelnia.html, 33-48.

SURVIVAL ANALYSIS OF ENTERPRISES ON THE EXAMPLE OF THE MAZOWIECKIE VOIVODESHIP

Abstract: This paper is devoted to investigate the enterprises duration on the market. The research sample consisted of a cohort of companies from the Mazowieckie voivodeship, which started their economic activity in 2007 (observed for 10 years). Non-parametric and parametric methods of survival analysis (including hazard models) were applied. Attention was also paid to the problem of censored data. The results obtained in both approaches are convergent. It was noted that the greatest risk of termination of business occurs between 2 and 5 years of the company existence. The factors such as the headquarters, ownership form and the business section have a significant impact on the duration of enterprises.

Keywords: survival analysis, non-parametric and parametric hazard models, enterprises duration on the market, censored data, Kaplan-Meier estimator

JEL classification: C4

ZASTOSOWANIE TRÓJCZYNNIKOWEGO MODELU FAMY-FRENCHA I CAPM DLA WYBRANYCH SPÓŁEK NOTOWANYCH NA GPW W WARSZAWIE W LATACH 2007-2017

Dorota Witkowska  <https://orcid.org/0000-0001-9538-9589>

Wydział Zarządzania

Uniwersytet Łódzki

e-mail: dorota.witkowska@uni.lodz.pl

Streszczenie: Celem prezentowanego badania jest próba oceny czy na polskim rynku kapitałowym trójczynnikiowy model Famy-Frencha lepiej opisuje stopy zwrotu spółek niż klasyczny model wyceny aktywów kapitałowych. Analizy przeprowadzono dla 30 spółek notowanych na GPW w Warszawie w latach 2007-2017, dzieląc okres badania na siedem podokresów, reprezentujących odmienną sytuację na giełdzie. W badaniu wykorzystano dane dzienne, a analizy porównawcze przeprowadzono dla trzech portfeli składających się ze spółek należących do indeksów giełdowych WIG20, mWIG40 i sWIG80.

Słowa kluczowe: model wyceny aktywów kapitałowych (CAPM), trójczynnikiowy model Famy-Frencha, polski rynek kapitałowy

JEL classification: C01, G10, G12

WPROWADZENIE

W latach sześćdziesiątych ubiegłego stulecia opracowano jednoczynnikowy model Sharpe'a oraz model wyceny aktywów kapitałowych (CAPM), których zastosowanie wzbudza szereg wątpliwości, ponieważ każdorazowo należy: (a) określić długość interwału do wyznaczania stóp zwrotu, (b) ustalić okres, będący próbą estymacyjną, (c) wybrać reprezentanta rynku i instrumentu wolnego od ryzyka. Dochodzi do tego krytyka związana z: (1) estymacją MNK modeli, które zazwyczaj nie spełniają założeń Gaussa-Markowa, (2) niestabilnością parametru

beta oraz (3) ze zbyt uproszczoną postacią modeli, zawierających de facto jedną zmienną objaśniającą tj. albo indeks rynku albo premię za ryzyko¹.

W odpowiedzi na krytykę CAPM, Fama i French (1992, 1993) zaproponowali uzupełnienie go o dodatkowe, poza premią za ryzyko, czynniki ryzyka tj.: wielkością spółki wyrażoną jej kapitalizacją oraz stosunkiem wartości księgowej (*Book Value* – BV) do wartości rynkowej (*Market Value* – MV). Pierwszy, określany jest jako efekt skali *SMB* (*Small Minus Big*) i wyjaśnia różnicę pomiędzy stopą zwrotu spółek o małej kapitalizacji a stopą zwrotu spółek o dużej kapitalizacji. Bowiem, jak zauważyli Autorzy, małe firmy często wykazują sezonowość w osiąganiu przychodów i utrzymywaniu płynności, natomiast duże spółki są mniej podatne na takie zmiany. Drugi – *HML* (*High Minus Low*), tłumaczy anomalie między stopą zwrotu z akcji spółek o wysokiej wartości współczynnika BV/MV, a stopą zwrotu z akcji spółek o niskiej wartości tego współczynnika. Spółki, które w horyzoncie kilku lat utrzymują stosunek wartości księgowej do rynkowej na wysokim poziomie, osiągają zazwyczaj niższe stopy zwrotu, niż spółki o niskiej wartości BV/MV. Trójczynnikowy model Famy-Frencha jest zatem postaci [Fama, French 1996]:

$$R_i - R_f = \alpha + \beta_M(R_M - R_f) + \beta_{SMB}SMB + \beta_{HML}HML + \varepsilon, \quad (1)$$

gdzie: R_i – oczekiwana stopa zwrotu z i -tego portfela, R_f – stopa wolna od ryzyka; R_M – stopa zwrotu z czynnika rynkowego; *SMB* – efekt skali, *HML* – różnica między średnimi stopami zwrotów spółek o wysokim i niskim współczynnika BV/MV; α , β_M , β_{SMB} , β_{HML} – parametry modelu; ε – składnik losowy.

Pojawia się zatem pytanie, dlaczego ten ulepszony model nie jest powszechnie stosowany. Okazuje się bowiem, że 73,5% dyrektorów finansowych (CFOs) spośród 392 przedsiębiorstw działających w USA zawsze lub prawie zawsze używa CAPM do wyceny kapitału [Graham, Harvey 2001]. Wyniki podobnego badania przeprowadzonego w 313 europejskich firmach wykazały, że 45% CFOs całkowicie polega na wskazaniach modelu CAPM [Brounen i in. 2004]. Wyjaśnienie tego zjawiska może być różne. Po pierwsze, praktycy nie znają trójczynnikowego modelu Famy-Frencha (F-F). Po drugie, jego zastosowanie wymaga znacznie więcej danych i pracy do ich przetworzenia. Po trzecie, praktycy nie są przekonani o wyższości modelu F-F nad CAPM. Trudno wyrokować która ze wzmiankowanych przyczyn jest najbardziej istotna. Wydaje się jednak, że najbardziej prawdopodobną przyczyną relatywnie małego zainteresowania modelem trójczynnikowym również wśród badaczy jest prosty rachunek zysków i kosztów. Bowiem oszacowanie modelu F-F wymaga znacznie większego (w stosunku do CAPM) nakładu pracy związanego z pozyskaniem i przetwarzaniem

¹ Szeroką dyskusję na temat krytyki towarzyszącej tym modelom przedstawiono m.in. w pracy [Tarczyński i in. 2013, s. 45-56].

dotychczasowych danych, a efekt w postaci lepszego dopasowania modelu trójczynnikowego do danych empirycznych nie zawsze jest satysfakcjonujący.

Model F-F był weryfikowany dla różnych rynków (por. Witkowska 2019), w tym również szacowany dla spółek notowanych na GPW w Warszawie (por. [Kowerski 2010, Czapkiewicz, Skalna 2010 i 2011, Schmidt 2013, Zaremba 2014, Mościbrodzka 2014, Czapiewski 2015, Redlicki i Borowski 2017, Urbański 2017]). W pracach tych za czynnik rynkowy zazwyczaj przyjmuje się indeks giełdowy WIG, a za instrument wolny od ryzyka stopę procentową WIBOR lub 52-tygodniowe bony skarbowe.

Kowerski [2010], Urbański [2017] i Zaremba [2014] potwierdzają lepsze dopasowanie modelu trójczynnikowego F-F do danych empirycznych niż CAPM w przypadku polskiego rynku kapitałowego. Redlicki i Borowski [2017] stwierdzili dobre dopasowanie modelu trójczynnikowego do uzyskanych stóp zwrotu z portfeli skonstruowanych na podstawie ich kapitalizacji i stosunku BV do MV w latach 2005-2015. Czapkiewicz i Skalna [2010], potwierdzają wyższość tego modelu nad CAPM w oparciu o badania spółek publicznych od grudnia 2002 do stycznia 2010. Jednakże Autorki, na podstawie próby przedłużonej do grudnia 2010, wykazały, że model F-F jest adekwatny do opisu rynku kapitałowego w okresie hossy ale nieprzydatny w okresie bessy [Czapkiewicz, Skalna 2011]. Badania Czapiewskiego [2015], dotyczące spółek z giełdy warszawskiej w latach 2000-2013 potwierdzają błędne wyceny oparte na obu modelach. Mościbrodzka [2014] analizowała spółki w latach 2009-2013 i na ich podstawie konkluduje, że szacownie zwrotów z portfeli o wysokim lub średnim wskaźniku BV/BM można oprzeć na czynnikach ryzyka związanych ze zmiennymi w modelu F-F, natomiast w przypadku portfeli tworzonych ze spółek o niskim wskaźniku odnotowano niestabilność modelu.

Celem prezentowanych badań jest próba oceny czy trójczynnikowy model Fama-Frencha lepiej opisuje stopy zwrotu spółek niż klasyczny model wyceny aktywów kapitałowych. Analizy przeprowadzono dla 30 spółek notowanych na GPW w Warszawie w latach 2007-2017, który to okres podzielono na siedem podokresów reprezentujących odmienną sytuację na rynku kapitałowym w Polsce. Wykorzystano codzienne logarytmiczne stopy zwrotu, a analizy porównawcze przeprowadzono dla trzech portfeli składających się ze spółek należących do indeksów giełdowych: WIG20, mWIG40 i sWIG80. We wszystkich testach przyjęto odrzucenie hipotez zerowych na poziomie istotności 0,05.

PRZYJĘTE ZAŁOŻENIA BADAWCZE

Prezentowane w niniejszym opracowaniu badania zasadniczo różnią się od tych, które są prezentowane we wzmiankowanych pozycjach literatury, w których rozpatruje się portfele akcji zbudowane na podstawie przeprowadzonej klasyfikacji spółek według czynników ryzyka związanych z wielkością kapitalizacji i wartością współczynnika BV/MV , a same modele szacowane są na podstawie miesięcznych

stóp zwrotu. W naszych rozważaniach przyjęto założenie, że jeśli model trójczynnika jest ulepszoną wersją modelu wyceny aktywów kapitałowych, to można go wykorzystać zarówno do opisu premii za ryzyko uzyskiwanej dla akcji poszczególnych spółek, jak i dla portfeli akcji. W związku z tym oszacowano modele CAPM i Famy-Frencha dla akcji pojedynczych spółek, z których komponowano portfele według ich przynależności do wyróżnionych indeksów giełdowych. Warto dodać, że podobne podejście można znaleźć w pracy [Bartholdy, Peare 2005].

W badaniach uwzględniono dzienne logarytmiczne stopy zwrotu cen akcji spółek z rynku głównego GPW w Warszawie na zamknięcie dnia sesyjnego (2363 obserwacje), które zostały skorygowane o efekt splitów i wypłaconych dywidend. Wszystkie dane pochodzą z serwisu stooq.pl. Za indeks rynku przyjęto WIG, a za stopę wolną od ryzyka - stopę procentową WIBOR 1Y.

Analizy obejmują okres od 15.10.2007 do 31.03.17, który charakteryzował się zróżnicowaną sytuacją na rynku finansowym. W związku z tym cały okres badawczy został podzielony na siedem podokresów (tabela 1). Wynika to z faktu, że współczynniki beta są zazwyczaj niestabilne w czasie i do ich estymacji należy brać okresy krótsze, na co wskazują m.in. [Bartholdy, Peare 2005]. Dotyczy to zwłaszcza rynków mniejszych i słabiej rozwiniętych, na których obserwuje się znaczną dynamikę zmian [Tarczyński i in. 2013, s. 50-51]. Również zmieniająca się sytuacja rynkowa wpływa na niestabilność współczynników beta, bowiem ceny akcji spółek różniące się np. wielkością lub branżą w odmienny sposób reagują na zmieniającą się koniunkturę [Tarczyński i in. 2013, s. 81, 137-176].

Tabela 1. Wyróżnione okresy badania

Symbol okresu i rodzaj tendencji rynkowej	Data		Liczba obserwacji	
	rozpoczęcia	zakończenia	w dniach	(%)
B1 - bessza	15.10.2007	31.10.2008	262	11
H1 - hossza	3.11.2008	5.08.2011	693	29
N1 - stagnacja	8.08.2011	5.06.2012	207	9
H2 - hossza	6.06.2012	31.10.2013	351	15
N2 - stagnacja	4.11.2013	30.04.2015	367	16
B2 - bessza	4.05.2015	15.01.2016	177	7
H3 - hossza	18.01.2016	31.03.2017	306	13

Źródło: opracowanie własne na podstawie [Rogowicz 2017, s. 48]

Do badania wybrano spółki, które spełniały dwa kryteria: (1) były nieprzerwanie notowane na rynku głównym w całym horyzoncie badania oraz (2) należały do portfeli indeksów WIG20, mWIG40 lub sWIG80 w latach 2006-2017 przynajmniej 10-krotnie. Lista spółek uwzględnionych w badaniu została przedstawiona w tabeli 2. Warto zauważyć, że w okresie 2006-2017 zaledwie 7 spółek było nieprzerwanie uwzględnianych w kompozycji indeksu WIG20 i 6 spółek w przypadku mWIG40, a dla indeksu sWIG80 żadna ze spółek nie spełniła

tęgo warunku. Z drugiej strony, 35 spółek znalazło się w portfolio któregośkolwiek z tych indeksów, co wynika z okresowych zmian składu indeksów (tabela 2).

Badania własności statystycznych stóp zwrotu przeprowadzono w oparciu o testy statystyczne. Na podstawie uzyskanych wyników stwierdzono, że stopy zwrotu z utworzonych portfeli porównywane w tych samych podokresach nie różnią się istotnie. Natomiast ryzyko jest wyraźnie wyższe dla portfeli utworzonych ze spółek średnich tj. należących do mWIG40 [Witkowska 2019].

Porównania oczekiwanych stóp zwrotu w dwóch sąsiadujących ze sobą okresach nie wskazują na występowanie istotnych różnic z wyjątkiem istotnie mniejszych zwrotów w czasie trwania pierwszej bessy w porównaniu z pierwszą hossą. Również stopy zwrotu są istotnie mniejsze od zera jedynie w pierwszym analizowanym podokresie i to z wyjątkiem portfela utworzonego ze spółek sWIG80. Widoczne są natomiast istotne różnice ryzyka w sąsiadujących podokresach [Witkowska 2019].

Tabela 2. Lista spółek uwzględnionych w badaniu i ich obecność w portfelach indeksów

Liczba wystąpień spółek w portfolio indeksu:					
Spółki	WIG20	Spółki	mWIG40	Spółki	sWIG80
KGHM	12 (12)	AMREST	12 (12)	DEBICA	11 (11)
MBANK	12 (12)	BUDIMEX	12 (12)	JWCONSTR	11 (11)
ORANGEPL	12 (12)	ECHO	12 (12)	LENTEX	11 (12)
PEKAO	12 (12)	INGBSK	12 (12)	RAFAKO	11 (12)
PGNIG	12 (12)	MILLENNIUM	12 (12)	SNIEZKA	11 (11)
PKNORLEN	12 (12)	ORBIS	12 (12)	COMP	10 (10)
PKOBP	12 (12)	CIECH	11 (12)	DOMDEV	10 (12)
LOTOS	11 (12)	EMPERIA	11 (11)	FAMUR	10 (11)
ASSECOPOL	10 (12)	KETY	11 (12)	PELION	10 (12)
BZWBK	10 (10)	NETIA	11 (12)	POLICE	10 (11)
Liczba wystąpień	Liczba spółek w portfelach indeksów w latach 2006-2016				
	WIG20	mWIG40	sWIG80	przynajmniej w jednym	
12	7	6	0	35	
11	1	4	5	20	
10	2	0	5	10	

Uwaga: W nawiasach podano liczbę wystąpień danej spółki we wszystkich indeksach.

Źródło: opracowanie własne na podstawie [Rogowicz 2017, s. 51-52]

BUDOWA MODELU FAMY-FRENCHA

W celu konstrukcji i estymacji modelu F-F należy określić czynniki: *SMB* oraz *HML*. Zatem przeprowadzono klasyfikację spółek pod względem wielkości kapitalizacji oraz relacji wartości księgowej do rynkowej, dzieląc je na te o dużej lub małej kapitalizacji oraz o wysokiej, umiarkowanej i niskiej wartości współczynnika *BV/MV*. Do grupy małych spółek zaklasyfikowano te o wartości kapitalizacji poniżej mediany, pozostałe zaś określono jako duże. Podział pod

względem kryterium stosunku wartości księgowej do rynkowej pozwolił zaklasyfikować spółki do trzech grup, które utworzyło 30% spółek o największej wartości współczynnika, 40% o jego średniej wartości oraz pozostałe 30% o małej wartości. Mając na celu odwzorowanie faktu opierania decyzji inwestorów na jak najbardziej aktualnych danych, dostępnych w danym momencie, reklasyfikację spółek do danej kategorii przeprowadzano na koniec każdego roku. Utworzono zatem sześć portfeli, które ze względu na specyfikę przyporządkowania, w różnych latach mogły cechować się odmienną liczbą spółek, co przedstawiono w tabeli 3, w której symbole portfeli określają klasyfikację wg dwóch kryteriów: SL (small/low), SM (small/medium), SH (small/high), BL (big/low), BM (big/medium) i BH (big/high).

Wykorzystując utworzone portfele, dla każdego dnia notowań wyznaczono średnie stopy zwrotu z portfeli, rozumiane jako średnia arytmetyczna stóp zwrotu wszystkich spółek tworzących dany portfel. Następnie, na ich podstawie wyznaczono wartości czynników *SMB* i *HML* modelu Famy-Frencha postaci:

$$SMB = \frac{SL+SM+SH}{3} - \frac{BL+BM+BH}{3}, \quad (2)$$

$$HML = \frac{SH+BH}{2} - \frac{SL+BL}{2}, \quad (3)$$

gdzie: *SL, SM, SH, BL, BM, BH* oznaczają stopy zwrotu z portfeli o określonym symbolu.

Tabela 3. Liczba spółek w poszczególnych portfelach w kolejnych latach

Rok	SL	SM	SH	BL	BM	BH	Rok	SL	SM	SH	BL	BM	BH
2007	5	4	5	5	8	3	2013	4	6	5	6	6	3
2008	4	7	4	6	5	4	2014	4	5	6	5	7	3
2009	4	6	5	5	6	4	2015	4	6	5	6	6	3
2010	5	6	4	5	6	4	2016	3	7	5	6	5	4
2011	3	6	5	6	6	4	2017	3	6	6	6	6	3
2012	3	7	5	7	5	3							

Źródło: opracowanie własne

Badając średnie wartości wszystkich zmiennych modelu Famy-Frencha dla każdego rozpatrywanego podokresu zauważa się, że czynnik *HML* jest zawsze ujemny, czyli stopy zwrotu spółek o wysokim wskaźniku *BV/MV* są średnio niższe niż spółek o niskich wartościach tego miernika. Widoczne jest też, że w okresach hossy uśredniona (po wszystkich obserwacjach i spółkach) premia za ryzyko jest dodatnia, a w okresach bessy – ujemna (por. [Witkowska 2019]).

WYNIKI ESTYMACJI MODELI JEDNO- I TRÓJCZYNNIKOWYCH

Modele F-F zostały oszacowane dla wszystkich wyodrębnionych podokresów koniunktury, osobno dla każdej z 30 rozpatrywanych spółek. We wszystkich modelach parametry stojące przy premii za ryzyko były statystycznie

istotne, natomiast w 104 modelach (na 210 oszacowanych, czyli w 49,5%) wszystkie zmienne były statystycznie istotne. Przy czym w 34 modelach (czyli 16,2%) obie dodatkowe (w stosunku do modelu CAPM) zmienne były statystycznie nieistotne. Oznacza to, że w niemal 84% oszacowanych modeli jedna ze zmiennych SMB lub HML przyczyniała się do lepszego opisu stóp zwrotu badanej spółki w wyróżnionym okresie. Przy czym najwięcej modeli, w których wszystkie zmienne były statystycznie istotne zaobserwowano w podokresie H2 i N2 (odpowiednio 20 i 18 modeli na 30 szacowanych w każdym podokresie), a najmniej w B2 i H3 (odpowiednio 11 i 12 modeli). Z kolei najwięcej modeli, w których obie zmienne SMB i HML były statystycznie nieistotne, odnotowano w podokresach N1 i B2 (takich modeli było odpowiednio 9 i 8 na 30), a najmniej w H1 i B1 (tj. 2 lub 3). Biorąc pod uwagę przynależność spółek do wyróżnionych indeksów giełdowych, to najwięcej modeli, w których istotne były wszystkie zmienne objaśniające było wśród modeli spółek należących do sWIG80 – 43 modele na 70 oszacowanych, natomiast w przypadku spółek należących do WIG20 i mWIG40 takich modeli było odpowiednio 30 i 31 (porównaj tabela 4).

Tabela 4. Informacja o istotności zmiennych SMB i HML oraz współczynnikach determinacji w oszacowanych modelach Famy-Frencha

Okresy	R ² (%)	Liczba modeli, w których zmienne				Modele spółek należących do portfeli indeksów giełdowych		Liczba modeli, w których wszystkie zmienne są statystycznie istotne
		wszystkie są statystycznie istotne	SMB	HML	SMB i HML			
			są statystycznie nieistotne				Indeks	
B1	41,8	14	7	12	3			
H1	33,7	16	7	9	2			
N1	37,5	13	15	11	9			
H2	23,5	18	6	10	4			
N2	24,6	20	6	8	4	WIG20	44,2	
B2	26,5	11	11	16	8	mWIG40	23,4	
H3	24,2	12	9	13	4	sWIG80	23,2	
Suma		104	61	79	34	Suma		

Uwaga: podane wartości R² są wartościami uśrednionymi dla modeli oszacowanych dla poszczególnych spółek w rozpatrywanych podokresach i portfelach indeksów giełdowych.

Źródło: opracowanie własne

Jakość dopasowania modelu do danych empirycznych opisywana jest przez współczynnik determinacji. Porównując poszczególne okresy badania stwierdzono, że relatywnie najlepiej dopasowane były modele w pierwszej hossie H1 i pierwszym okresie stagnacji N1. Biorąc pod uwagę podział spółek wg ich przynależności do indeksów giełdowych widoczne jest najwyższe dopasowanie (zarówno minimalne, jak i maksymalne oraz średnie) wśród spółek należących do portfela WIG20 (tabela 4) – maksymalne dla banków PEKAO (75,9%) i PKOBP (75,2% maksymalne i 62,9% średnie), a minimalne dla ORANGEPL (8,7% minimalne i 25% średnie), dla której zróżnicowanie współczynników determinacji

w rozpatrywanych siedmiu okresach było największe ($V=47,7\%$). Wśród spółek należących do portfela mWIG40 maksymalną wartość współczynnika R^2 zaobserwowano dla CIECHu (51,6%), a największą średnią dla MILLENNIUM (37,2%). Z kolei najniższe wartości minimalne i średnie osiągnął ORBIS (odpowiednio 6,2% i 13,7%), który charakteryzował się największą zmiennością stopnia objaśnienia ($V=53,2\%$). W ostatnim z analizowanych portfeli najlepiej objaśnione były stopy zwrotu spółek JWCONSTR i LENTEX, w przypadku których maksymalne wartości R^2 wyniosły odpowiednio 54,8% i 53,8%. Jednocześnie pierwsza z wymienionych spółek charakteryzowała się największą średnią wartością tego współczynnika (36,7%). Najniższe wartości współczynnika determinacji odnotowano dla spółki: DEBICA (2,9% minimum i 13,9% średnie). Drugą spółką o najniższym R^2 jest LENTEX (4,7%), charakteryzujący się przy tym największą zmiennością stopnia objaśnienia stóp zwrotu ($V=69,3\%$), w czym wyprzedza spółkę DEBICA ($V=68,7\%$).

Tabela 5. Oceny parametrów modeli Famy-Frencha i CAPM stojących przy wyróżnionych zmiennych dla portfeli, utworzonych ze spółek należących do różnych indeksów giełdowych

Modele	Famy-Frencha			CAPM	Famy-Frencha			CAPM
Zmienne	(R_M-R_f)	SMB	HML	(R_M-R_f)	(R_M-R_f)	SMB	HML	(R_M-R_f)
Okres	Portfel WIG20				Portfel mWIG40			
B1	1,0560	-0,2923	0,0649	1,1203	0,8557	0,2758	-0,0384	0,7962
H1	1,1220	-0,2667	0,0269	1,2261	0,7749	0,2741	0,0232	0,6637
N1	0,9947	-0,1723	-0,0007	1,0331	0,8452	0,2497	0,3046	0,8003
H2	1,0369	-0,1576	0,0265	1,0946	0,7695	0,2262	0,0543	0,6899
N2	1,0005	-0,2520	0,1027	1,0779	0,7995	0,1918	-0,0367	0,7322
B2	0,9066	-0,3783	0,0553	1,1152	0,7688	0,1680	-0,0018	0,6814
H3	1,0219	-0,2178	0,0590	1,1817	0,8079	0,1483	0,1124	0,6937
	Portfel sWIG80				Cały portfel			
B1	0,9917	0,8121	0,0613	0,8258	0,9678	0,2652	0,0293	0,9141
H1	0,9584	0,8139	-0,0138	0,6351	0,9518	0,2738	0,0121	0,8416
N1	0,9347	0,8570	0,1612	0,7492	0,9249	0,3115	0,1550	0,8609
H2	0,8742	0,7422	-0,0312	0,6055	0,8935	0,2703	0,0165	0,7967
N2	0,9666	0,8386	0,0849	0,6232	0,9222	0,2595	0,0503	0,8111
B2	0,8884	0,7559	0,0703	0,5131	0,8258	0,6744	0,0525	0,7699
H3	0,8546	0,1819	0,0412	0,3226	0,8852	0,2016	0,0746	0,7327

Źródło: opracowanie własne

W celu zbadania który z modeli lepiej opisuje stopy zwrotu, na podstawie tych samych danych (co F-F) oszacowano CAPM. Tabela 5 zawiera wartości parametrów modeli F-F i CAPM, oszacowanych dla 10-składnikowych portfeli utworzonych ze spółek należących do odpowiednich indeksów w równych udziałach oraz 30-składnikowego portfela zawierającego akcje wszystkich analizowanych spółek.

Jak można zauważyć, największe wartości ocen estymatorów parametrów modeli F-F, stojących przy premii za ryzyko, obserwuje się dla portfela WIG20,

a najmniejsze dla mWIG40, aczkolwiek różnice w wartościach nie wydają się być znaczące. W przypadku uwzględnienia zróżnicowania stóp zwrotu w zależności od wielkości firmy, to w przypadku portfela WIG20 parametry są ujemne dla wszystkich analizowanych podokresów, podczas gdy dla pozostałych portfeli są dodatnie, a największe wartości ocen estymatorów parametru stojącego przy tej zmiennej obserwuje się dla portfela WIG80. Natomiast zróżnicowanie stóp zwrotu wynikające z różnej relacji wartości księgowej do ceny rynkowej wydaje się mieć niewielki wpływ dla wszystkich portfeli. W przypadku CAMP, niemal we wszystkich 210 modelach bety były istotnie większe od zera z wyjątkiem modeli estymowanych dla COMP w podokresach B2 i H3, DEBICA i PELION w H3.

Portfel utworzony ze spółek należących do indeksu WIG20 miał charakter agresywny we wszystkich analizowanych podokresach, przy czym spółki: KGHM, MBANK, PEKAO i PKOBP charakteryzowały się betami większymi od jedności we wszystkich modelach, PKNORLEN – we wszystkich okresach z wyjątkiem B2, BZWBK – z wyjątkiem N1 i H2. PGNIG i LOTOS miały bety mniejsze od jedności w trzech (odpowiednio B1, H1 i N1 oraz B1, N2 i B2), a ASSECOPOLAND i ORANGEPL – we wszystkich analizowanych okresach. Portfele utworzone ze spółek z indeksów mWIG40 i sWIG80 miały defensywny charakter i tylko bank MILLENNIUM (z mWIG40) miał charakter agresywny we wszystkich okresach, a parametry beta w modelach oszacowanych dla spółki JWCONSTR (z sWIG80) były większe od zera w pięciu (na siedem badanych) podokresach.

Tabela 6. Porównanie wartości skorygowanych współczynników determinacji dla portfeli

Okres	FF	CAPM	różnica	FF	CAPM	różnica
	WIG20			mWIG40		
B1	0,5683	0,5219	-8,18%	0,3054	0,2627	-13,99%
H1	0,5454	0,5101	-6,48%	0,2270	0,1739	-23,41%
N1	0,4702	0,4538	-3,47%	0,3201	0,2699	-15,70%
H2	0,3545	0,3293	-7,10%	0,1571	0,1123	-28,53%
N2	0,3621	0,3138	-13,32%	0,1834	0,1329	-27,52%
B2	0,3854	0,3312	-14,07%	0,2042	0,1601	-21,60%
H3	0,3648	0,3405	-6,64%	0,1857	0,1314	-29,26%
Średnia dla portfela		WIG20	-8,47%		mWIG40	-22,86%
	sWIG80			całość		
B1	0,3613	0,2278	-36,93%	0,4112	0,3375	-17,94%
H1	0,2293	0,1323	-42,27%	0,3341	0,2719	-18,61%
N1	0,3082	0,1698	-44,92%	0,3658	0,2976	-18,64%
H2	0,1725	0,0554	-67,88%	0,2284	0,1656	-27,48%
N2	0,1741	0,0691	-60,34%	0,2398	0,1717	-28,37%
B2	0,1669	0,0669	-59,92%	0,1474	0,1864	26,47%
H3	0,2577	0,0371	-85,59%	0,2345	0,1693	-27,80%
Średnia dla portfela		sWIG80	-55,84%		całość	-16,05%

Uwaga: różnice to względne udziały procentowe, które zostały wyznaczone w odniesieniu do modelu Famy-Frencha, ich ujemne wartości oznaczają lepsze dopasowanie modelu trójczynnikowego do danych empirycznych niż dopasowanie modelu jednoczynnikowego.

Źródło: opracowanie własne

Warto dodać, że zróżnicowanie wartości bet dla poszczególnych spółek w kolejnych próbach estymacyjnych było najmniejsze w przypadku portfela WIG20 tj. między 24,3% w okresie N2 i 42,5% w N1 (liczone jako współczynnik zmienności bet). Dla spółek z portfela mWIG40 najmniejsze zróżnicowanie zaobserwowano w okresie B1 31,9%, a największe w H3 47%, a sWIG80 odpowiednio 36,5% w B1 i 56% w B2.

W dalszych badaniach porównano skorygowane stopniami swobody współczynniki determinacji wszystkich modeli wyznaczonych dla poszczególnych spółek. Generalnie w większości przypadków modele trójczynnika dają lepsze dopasowanie do danych empirycznych i tylko w 12 przypadkach (na 210, czyli 5,7%) modele CAPM okazały się lepsze niż modele Famy-Frencha. Najczęściej sytuacja taka miała miejsce w czasie trwania drugiej hossy (3 przypadki na 30) i dla spółki KGHM i INGBSK (po 3 przypadki na 7).

Z kolei największą poprawę dopasowania modelu w przypadku zastosowania modelu trójczynnika zaobserwowano dla spółek COMP (4 przypadki na 7) oraz SNIEZKA w 6 przypadkach na 7. Natomiast w 47 przypadkach (na 210, czyli 22%) przyrost stopnia dopasowania modelu Famy-Frencha w porównaniu z CAPM wyniósł ponad 50%. Najwięcej takich przypadków zaobserwowano w podokresach N2 i H3 (po 12 na 30), a najmniej w okresach B1, H1 i N1 (po 2 na 30). Biorąc pod uwagę najmniejsze przyrosty skorygowanego współczynnika determinacji stwierdzamy, że w 66 modelach (na 210 par szacowanych modeli, czyli w 31,4%) były one mniejsze od 10%, a spółkami o największym podobieństwie R^2 są banki PKOBP i PEKAO. Najwięcej podobnych wartości skorygowanych współczynników determinacji odnotowano w podokresie N1 (16 na 30).

Tabela 6 zawiera średnie wartości skorygowanych współczynników determinacji wyznaczone dla spółek tworzących portfele indeksów giełdowych w badanych podokresach. Potwierdzają one lepsze dopasowanie modelu trójczynnika do danych empirycznych niż klasyczne modele CAPM. Wynika z nich również, że największe zróżnicowanie stopnia wyjaśnienia danych rzeczywistych przez oba modele ma miejsce w przypadku spółek należących do indeksu sWIG80, a najmniejsze dla spółek z indeksu WIG20.

ZAKOŃCZENIE

Celem badań była ocena przydatności trójczynnika modelu Famy-Frencha do opisu stóp zwrotu akcji spółek notowanych na Giełdzie Papierów Wartościowych w Warszawie. Za model referencyjny przyjęto klasyczny model wyceny aktywów kapitałowych. Wybór spółek do badania został przeprowadzony w taki sposób, aby zapewnić zróżnicowanie analizowanych spółek, przy jednoczesnym spełnieniu warunku ciągłości notowań w całym horyzoncie badania oraz założonej minimalnej liczbie wystąpień w kompozycji indeksów WIG20,

mWIG40 oraz sWIG80. Przyjęty horyzont badania obejmował lata 2007-2017 i został podzielony na siedem podokresów o zróżnicowanej sytuacji rynkowej.

Uzyskane wyniki pozwalają sformułować następujące wnioski:

- rynkowy czynnik ryzyka ma większy wpływ na stopy zwrotu z portfela niż czynniki związane z wielkością spółki SMB i opierające się na wskaźniku wartości księgowej do rynkowej HML;
- wprowadzenie do równania dodatkowych czynników, wynikających z analizy fundamentalnej tj. wielkości kapitalizacji oraz stosunku wartości księgowej do wartości rynkowej, pozytywnie wpływa na poprawę jakości dopasowania modelu;
- poprawę stopnia dopasowania modeli odnotowuje się przede wszystkim dla spółek małych i średnich, natomiast w przypadku największych i najbardziej płynnych spółek zmiany te są niewielkie;
- szczególnie widoczna jest różnica w stopniu dopasowania obu modeli do danych empirycznych w segmencie małych spółek, w którym względne różnice współczynników determinacji przewyższają 10% w każdym okresie estymacji (dla spółek z mWIG40 te różnice są mniejsze i wynoszą ponad 5%).

BIBLIOGRAFIA

- Bartholdy J., Peare P. (2005) Estimation of Expected Return: CAPM vs. Fama and French. *International Review of Financial Analysis*, 14(4), 407-427.
- Brounen D., De Jong A., Koedijk K. (2004) Corporate Finance in Europe Confronting Theory with Practice. *Financial Management*, 33, 71-101.
- Czapiewski L. (2015) Model CAPM i trójczynnikowy model Famy-Frencha w analizie zdarzeń na polskim rynku kapitałowym. *ZN Uniwersytetu Szczecińskiego*, 854, 609-620.
- Czapkiewicz A., Skalna I. (2010) The CAPM and the Fama-French Models in Warsaw Stock Exchange. *Przegląd Statystyczny*, 57(4), 128-141.
- Czapkiewicz A., Skalna I. (2011) Użyteczność stosowania modelu Famy i Frencha w okresach hossy i bessy na rynku akcji GPW w Warszawie. *Bank i Kredyt*, 42(3), 61-80.
- Fama E., French K. (1992) The Cross-Section of Expected Stock Returns. *Journal of Finance*, 47, 427-465.
- Fama E., French K. (1993) Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 3, 3-56.
- Fama E., French K. (1996) Multifactor Explanations of Asset Pricing Anomalies. *Journal of Finance*, 53, 55-84.
- Graham J. R., Harvey C. R. (2001) The Theory and Practice of Corporate Finance: Evidence from the Field. *Journal of Financial Economics*, 60, 187-243.
- Kowerski M. (2010) The Analysis of an Investment Risk Within Emerging Capital Markets. The Case of the Warsaw Stock Exchange. *e-Finanse*, 6 (4), 1-23.

- Mościbrodzka M. (2014) Stabilność czynników ryzyka w modelu Fama-Frencha wyceny kapitału na GPW w Warszawie. *Finanse, Rynki Finansowe, Ubezpieczenia*, 66, 145-159.
- Redlicki M., Borowski K. (2017) Wykorzystanie trójczynnika modelu Fama-Frencha na GPW. *Studia i Prace Kolegium Zarządzania i Finansów*, ZN 153, 81-102.
- Rogowicz T. (2017) Model Fama-Frencha na Giełdzie Papierów Wartościowych w Warszawie. Praca magisterska, opiekun K. Kompa, SGGW, Warszawa.
- Schmidt B. (2013) Zastosowanie modelu trójczynnika w inwestowaniu na Giełdzie Papierów Wartościowych w Warszawie. *ZN Uniwersytetu Ekonomicznego w Krakowie*, 907, 157-169.
- Tarczyński W., Witkowska D., Kompa K. (2013) Współczynnik beta. Teoria i praktyka. Warszawa, Pielaszek Research.
- Urbański S. (2017) Comparison of a Modified and Classic Fama-French Model for the Polish Market. *Folia Oeconomica Stetinensia*, 17, 80-96.
- Witkowska D. (2019) Is the Three-Factor better than Single-Factor Capital Asset Pricing Model? Case of Polish Capital Market [w:] Tarczyński W., Nermend K. (red.), *Effective Investments on Capital Markets. 10-th Capital Market Effective Investment Conference (CMEI 2018), Springer Proceedings in Business and Economics*, Springer Nature Switzerland AG 2019, 225-237.
- Zaremba A. (2014) Cross-Sectional Asset Pricing Models for the Polish Market. SSRN working paper: <http://dx.doi.org/10.2139/ssrn.2396884>

**APPLICATION OF FAMA-FRENCH THREE-FACTOR MODEL
AND CAPM TO SELECTED COMPANIES LISTED ON WSE
IN YEARS 2007-2017**

Abstract: Presented research aims in evaluation if three-factor model better describes rates of return than single-factor capital asset pricing model. Investigation concerns 30 selected companies listed on WSE in years 2007-2017. The whole period of analysis is divided into seven samples according to observed market tendency in Poland. Research is conducted for daily rates of return whereas comparative analysis is provided for portfolios constructed from companies belonging to stock indexes WIG20, mWIG40 and sWIG80.

Keywords: capital asset pricing model (CAPM), Fama-French three-factor model, Polish capital market

JEL classification: C01, G10, G12

WIELOWYMIAROWA ANALIZA PORÓWNAWCZA WARUNKÓW OFEROWANYCH DLA MAŁYCH I ŚREDNICH PRZEDSIĘBIORSTW W POLSKICH WOJEWÓDZTWACH

Marta Wójcik  <https://orcid.org/0000-0003-3654-386X>

Wydział Zarządzania, Informatyki i Finansów
Uniwersytet Ekonomiczny we Wrocławiu
e-mail: 165382@student.ue.wroc.pl

Streszczenie: W opracowaniu przedstawiono wykorzystanie analizy skupień metodą Warda do zobrazowania warunków oferowanych względem małych i średnich przedsiębiorstw w poszczególnych województwach Polski. W badaniu wykorzystano dane z Głównego Urzędu Statystycznego. Wykorzystując hierarchiczne procedury aglomeracyjne dokonano grupowania województw na jednorodne klasy ze względu na zmienne charakteryzujące otoczenie jakie stwarza województwo dla małych i średnich przedsiębiorstw. Stworzono również ranking województw pod względem atrakcyjności dla biznesu.

Słowa kluczowe: analiza skupień, małe i średnie przedsiębiorstwa, metoda aglomeracyjna, metoda wzorca rozwoju

JEL classification: C38

WSTĘP

Małe i średnie przedsiębiorstwa to siła napędowa gospodarki. Ich liczebność, czyli 99,8% wszystkich przedsiębiorstw w Polsce w 2018 roku [Skowrońska i in. 2018] i odsetek zatrudnianych przez nie pracowników sprawia, że są ważnym elementem gospodarki. Sektor wytworzył w 2016 roku 49,8% PKB [Skowrońska i in. 2019] i zatrudniał 69% pracujących [Skowrońska i in. 2018].

Z tego względu istnieje mnogość inicjatyw odnoszących się bezpośrednio do niewielkich biznesów. Aby rozdysponować środki pomiędzy konkretne przedsiębiorstwa utworzono definicję małego i średniego przedsiębiorstwa:

<https://doi.org/10.22630/MIBE.2019.20.2.13>

Tabela 1. Definicja małego i średniego przedsiębiorstwa

Typ przedsiębiorstwa	Liczba zatrudnionych	Obrót roczny	Bilans roczny
Średnie	Poniżej 250	Poniżej 50 milionów euro	Poniżej 43 milionów euro
Małe	Poniżej 50	Poniżej 10 milionów euro	Poniżej 10 milionów euro
Mikro	Poniżej 10	Poniżej 2 milionów euro	Poniżej 2 milionów euro

Źródło: opracowanie własne

Zatrudniając poniżej 250 pracowników przedsiębiorstwa stają się niezmiernie elastyczne na zmiany i innowacje. Jednocześnie ich rozmiar sprawia, że silnie odczuwają nagłe wahania koniunkturalne [Skowrońska i in. 2017]. Z tego względu są swoistym pulsem gospodarki.

Znaczny wpływ na sektor mają decyzje podejmowane przez Unię Europejską i polski rząd. Przedsiębiorcy jako największe utrudnienia w prowadzeniu działalności wskazują między innymi wysokie składki ZUS, skomplikowane przepisy podatkowe, nadmierną biurokrację urzędniczą, czy wysokie podatki [Danielak 2017]. Należy mieć jednak na uwadze, że na niewielkie jednostki gospodarcze wpływa w dużym stopniu ich bezpośrednie otoczenie. Opłaty za lokal, niedobór specjalistów na rynku, czy konkurencja to czynniki często różniące się zależnie od regionu Polski [Ibidem]. Od wielu lat województwa należące do tak zwanej ściany wschodniej, czyli województwa: podlaskie, podkarpackie, świętokrzyskie, lubelskie i warmińsko-mazurskie, wypadają najslabiej pod względem rozwoju gospodarczego. Peryferyjny obszar Polski jest znacznie gorzej rozwinięty ekonomicznie ze względów zarówno historycznych jak i sąsiedztwa z krajami nienależącymi do Unii Europejskiej. Położenie obok granicy unijnej niesprzyja rozwojowi wymiany handlowej. Ponadto tamtejsze kraje graniczne, czyli Białoruś, Litwa, czy Ukraina nie należą do wysoko rozwiniętych państw. Mikroregion ten jest uznawany za jeden z najgorzej rozwiniętych w całej Unii [Polska Wschodnia 2014]. Państwo, poprzez dotowanie tego regionu funduszami pochodzącymi z funduszu państwa i Unii Europejskiej stara się zatrzeć tą granicę. Polska Agencja Rozwoju Przedsiębiorczości poza publikacjami i informacjami dotyczącymi małych i średnich przedsiębiorstw posiada osobną stronę internetową poświęconą Programowi Operacyjnemu Polska Wschodnia [PARP 2019]. Na stronie dostępne są informacje dla nowych przedsiębiorców, otwierania firm na platformach startowych z pomocą ekspertów, a także na temat udzielania ulg w spłacie należności, czy odnośnie internacjonalizacji przedsiębiorstwa. Program Operacyjny Polska Wschodnia posiada budżet równy 9,6 mld zł na lata 2014-2020. Koncentruje się on na poprawie sytuacji w sektorach:

rynku start-upów, wzornictwa, internacjonalizacji przedsiębiorstw, inwestycji w innowacje i infrastruktury dla rozwoju [Ibidem].

Celem artykułu było zbadanie warunków oferowanych małym i średnim przedsiębiorstwom przez poszczególne województwa Polski. Wyodrębniono województwa cechujące się podobnym otoczeniem dla biznesu, a także stworzono ranking. Poprzez analizę zweryfikowane zostanie także otoczenie przedsiębiorstw w województwach tak zwanej ściany wschodniej na tle pozostałych województw.

WYKORZYSTANE DANE

Podstawowe informacje

Do analizy warunków względem małych i średnich przedsiębiorstw w województwach wytypowano początkowo 12 zmiennych opisujących zjawisko. Wybór zmiennych do zbioru cech diagnostycznych oparto na kryteriach potrzebnych do przeprowadzenia analizy skupień:

- zmienności – wybrane zmienne powinny być nośnikami informacji różnicującym badane obiekty, z tego względu obliczono współczynnik zmienności, eliminując cechy, dla których osiągnięte wartości były niższe niż 0,1;
- stopnia skorelowania - silne powiązanie pomiędzy analizowanymi cechami powoduje, że nie dostarczają one nowych informacji na temat zjawiska. Obliczono współczynnik korelacji i odrzucono jedną ze zmiennych w przypadku, gdy para posiadała silną korelację. Wybór reprezentanta opierał się na względach merytorycznych [Ostasiewicz 1998].

W celu ujednoczenia miar zmiennych zdecydowano się na standaryzację według wzoru (zob. Balicki [2013]).

Ze względu na podane powyżej kryteria wybrano 7 zmiennych:

- X_1 – dochody na mieszkańca (w złotych; 2018),
- X_2 – środki z Unii Europejskiej na finansowanie programów i projektów unijnych na mieszkańca (w złotych; 2018),
- X_3 – środki z budżetu państwa na współfinansowanie programów realizowanych z udziałem środków z funduszy strukturalnych na mieszkańca (w złotych; 2018),
- X_4 – kredyty i pożyczki krajowe na mieszkańca (w tysiącach złotych; 2017),
- X_5 – przeciętny miesięczny dochód rozporządzalny na osobę (w złotych; 2017),
- X_6 – wymeldowania (w celu porównywalności województw przeliczone na mieszkańca; 2018),
- X_7 – nakłady na sektor B+R na mieszkańca (w złotych; 2017).

Dane zostały pobrane z portalu Głównego Urzędu Statystycznego. Są to najnowsze dostępne dane.

METODA BADAWCZA

Do przeprowadzenia analizy wybrano analizę skupień. Skorzystano z metody aglomeracji. Cechuje się ona łączeniem w podzbiory podobnych i rozdzielaniem różniących się między sobą obiektów. Początkowo każdy obiekt stanowi osobne skupienie. Kolejno łączy się ze sobą obiekty do siebie podobne. Aby przyporządkować element do danej grupy oblicza się odległość między obiektami. Każdy zbiór jest rozłączny i niepusty [Balicki 2013]. Ostatecznie otrzymuje się jedno skupienie zawierające wszystkie elementy skupień [Stanisz 2007]. W pracy jako sposób badania odległości pomiędzy obiektami wybrano metodę Warda. Sposób ten na podstawie analizy wariancji szacuje odległość pomiędzy skupieniami. Jego celem jest minimalizacja sumy kwadratów odchyień w środku skupień. W metodzie tej oblicza się błąd sumy kwadratów (ESS) według wzoru (zob. Stanisz [2017]).

Po iteracyjnym łączeniu ze sobą obiektów o najniższym błędzie sumy kwadratów otrzymuje się dendrogram.

Metoda wzorca rozwoju stosowana jest w celu uporządkowania elementów opisywanych wielowymiarowo. Zmienne analizowane poprzez porządkowanie muszą być mierzone na skali silniejszej od nominalnej. Gdy znajdują się na skali mocniejszej od porządkowej należy je znormalizować, aby były porównywalne [Panek i in. 2013]. Przed przystąpieniem do analizy należy określić charakter zmiennych, to znaczy przydzielić je do stymulant, destymulant, lub nominant [Ostasiewicz 1998]. W przypadku nominanty należy taką zmienną przekształcić do postaci stymulanty. Następnie ujednicamy zmienne poprzez standaryzację lub unitaryzację, tworzymy wzorzec i antywzorzec. Wzorzec składa się z maksymalnych wartości dla danej zmiennej, a antywzorzec z najniższych jej wartości [Stanimir 2006]. Kolejno oblicza się odległość między każdym obiektem, a wzorcem za pomocą wzoru (zob. Stanimir [2006]).

Wartości odległości dla obiektów są bazą dla powstania miar rozwoju (M_i)¹.

$$M_i = 1 - \frac{d_{i0}}{d_0} \quad (1)$$

d_{i0} – odległość euklidesowa pomiędzy i-tym obiektem, a wzorcem,

d_0 – odległość euklidesowa pomiędzy wzorcem, a antywzorcem.

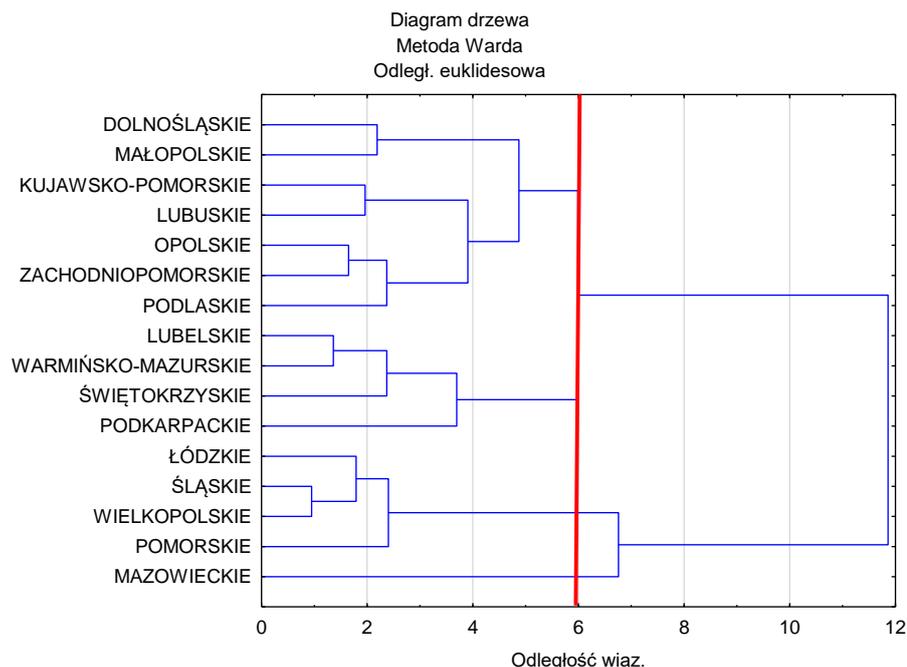
Miara ta zawiera się w przedziale [0;1]. Im wyższa jej wartość tym wyższy poziom zjawiska złożonego.

¹ Źródło: Stanimir A. (2006) Analiza danych marketingowych. Problemy, metody, przykłady. [w:] Wydawnictwo AE we Wrocławiu, 122.

WYNIKI BADAŃ

Dla zdefiniowanych zestandaryzowanych zmiennych dokonano klasyfikacji województw z wykorzystaniem metody Warda.

Rysunek 1. Wyniki analizy skupień metodą Warda



Źródło: opracowanie własne z zastosowaniem programu STATISTICA

W badaniu skorzystano z tak zwanej metody łokcia. Zdecydowano się na przecięcie dendrogramu na poziomie odległości wiązania równej ok.6. Wynikało to ze skoku jej wartości, oraz liczby powstałych w wyniku cięcia skupień. Analiza wyodrębniła następujące skupienia:

- I – województwa: dolnośląskie, małopolskie, kujawsko-pomorskie, lubuskie, opolskie, zachodniopomorskie, podlaskie,
- II – województwa: lubelskie, warmińsko-mazurskie, świętokrzyskie, podkarpackie,
- III – województwa: łódzkie, śląskie, wielkopolskie, pomorskie,
- IV – województwo mazowieckie.

Średnie skupień przedstawiono w tabeli 1. Wartości zostały przedstawione na skali szarości, ciemne barwy oznaczają wysokie wartości, jasne - niskie. Zmienna nr 6 - wymeldowania jest destymulantą, w jej przypadku wysokie

wartości oznaczono jasnymi barwami, jako niesprzyjające dla sytuacji małych i średnich przedsiębiorstw.

Tabela 2. Średnie skupień

	skupienie 1	skupienie 2	skupienie 3	skupienie 4
X1	549,85	371,37	474,13	414,28
X2	9,22	18,15	26,16	24,76
X3	13,79	17,62	35,00	19,77
X4	0,73	0,55	0,19	0,3
X5	1912,09	1616,86	1404,93	1364,34
X6	0,19%	0,25%	0,43%	0,28%
X7	1482,70	395,35	246,25	296,25

Źródło: opracowanie własne

Można zauważyć, iż województwo mazowieckie, które pozostało jednoosobową podgrupą, znacznie odbiega rozwojem od pozostałych. Dla zmiennych dochody na mieszkańca ogółem, kredyty i pożyczki krajowe, przeciętny miesięczny dochód rozporządzalny na osobę, wymeldowania i nakłady na sektor B+R osiągnęło najwyższe wartości. Zmienne druga i trzecia, czyli środki z UE na finansowanie projektów unijnych i środki z budżetu państwa na finansowanie programów z udziałem funduszy strukturalnych, posiadają najniższe wartości. Spowodowane jest to wyrównywaniem poziomu rozwoju pomiędzy polskimi województwami. Fakt ten potwierdzają najwyższe wartości w obu tych zmiennych w skupieniu trzecim. Skupienie to zgrupowało województwo lubelskie, podkarpackie, świętokrzyskie i warmińsko-mazurskie, a więc województwa ściany wschodniej, oprócz województwa podlaskiego. Województwa te otrzymują znacznie większą pomoc ze strony państwa i Unii Europejskiej w celu ujednolicenia rozwoju w polskich województwach. Analizując wyniki można dojść do wniosku, że najlepsze warunki do rozwoju małego i średniego biznesu można otrzymać zakładając biznes w województwie mazowieckim, kolejno w województwach skupionych w grupie nr 2, następnie nr 4. Najtrudniejsze warunki spotka się w województwach skupienia nr 3. Tendencja jest natomiast optymistyczna, ponieważ dofinansowanie tych regionów może wspomóc rozwój i poprawę sytuacji w tych regionach. Ze względu na niewielkie rozmiary małych i średnich przedsiębiorstw są one bardzo podatne na sytuację otoczenia, dlatego jako pierwsze powinny odczuć zmianę. Będąc przyszłym przedsiębiorcą nie posiadającym kapitału startowego pomysł startu biznesu w województwach ściany wschodniej może skutkować wyższym dofinansowaniem na start, co może przełożyć się w długim okresie na wzrost przedsiębiorczości w tym rejonie. Problemem może się natomiast okazać utrzymanie biznesu, ze względu na nadal odstający od pozostałych rozwój w tym regionie. Może to skutkować bezrobociem strukturalnym i niewielką ilością wykwalifikowanych specjalistów. Widoczne jest

to w zmiennej wymeldowania. Znacznie więcej wymeldowań występuję w uznawanych za gorzej rozwinięte skupieniach numer 3 i 4. Zakładając, iż spory odsetek emigrantów to osoby wykwalifikowane, szukające lepszej pracy nowi przedsiębiorcy mogą napotkać problemy ze skompletowaniem profesjonalnej kadry. Do grupy uznawanej za ścianę wschodnią należą województwa: lubelskie, podkarpackie, podlaskie, warmińsko-mazurskie i świętokrzyskie. Wszystkie te województwa, oprócz podlaskiego, zostały zgrupowane w jedno skupienie. Województwo to według danych pierwotnych cechuje się wyższym dochodem na mieszkańca, czy stopą bezrobocia. Głębsza analiza regionu pod kątem sektora pozwoliłaby ocenić i sprawdzić przyczynę tego zjawiska.

Aby poszerzyć analizę zdecydowano się na przeprowadzenie rankingu województw metodą wzorca rozwoju z zastosowaniem szerszego zakresu zmiennych, powiększonego o odrzucone przy analizie skupień zmienne. Porządkowanie nie posiada tych samych założeń, potrzebnych do jej przeprowadzenia, dlatego ze względu na ważność tych cech w rozważaniu zdecydowano się je uwzględnić. W efekcie dodano:

- X_8 – nakłady inwestycyjne na mieszkańca (w złotych; 2017),
- X_9 – stopa bezrobocia (%; pierwszy kwartał 2019),
- X_{10} – czas poszukiwania pracy (w miesiącach; 2018),
- X_{11} – liczba pracujących w przedsiębiorstwach do 49 osób (2018),
- X_{12} – absolwenci uczelni wyższych na 10 tys. ludności (2018).

Zmienne: wymeldowania, stopa bezrobocia i czas poszukiwania pracy w miesiącach uznano za destymulanty i stworzono ranking widoczny w tabeli 2. Każda ze zmiennych została tak samo ważona. Wartości w tabeli zaokrąglono do dwóch miejsc po przecinku.

Tabela 3. Ranking województw pod względem sytuacji małych i średnich przedsiębiorstw

Lp.	Województwo	Miara rozwoju
1.	Dolnośląskie	0,50
2.	Mazowieckie	0,48
3.	Pomorskie	0,47
4.	Małopolskie	0,47
5.	Opolskie	0,42
6.	Zachodniopomorskie	0,41
7.	Podlaskie	0,40
8.	Łódzkie	0,40
9.	Wielkopolskie	0,38
10.	Podkarpackie	0,36
11.	Warmińsko-Mazurskie	0,35
12.	Lubelskie	0,35

Lp.	Województwo	Miara rozwoju
13.	Śląskie	0,32
14.	Świętokrzyskie	0,31
15.	Kujawsko-pomorskie	0,29
16.	Lubuskie	0,25

Źródło: opracowanie własne

Najwyższe wartości osiągnęły województwa dolnośląskie i mazowieckie, najniższe - kujawsko-pomorskie i lubuskie. Województwa ściany wschodniej zajmują miejsca w drugiej części rankingu. Wyjątkiem, tak jak i w przypadku analizy skupień, jest województwo podlaskie będące w połowie stawki. Wyróżnia się ono znacznie wyższym dochodem na mieszkańca od pozostałych województw ściany wschodniej. Województwo mazowieckie, ze względu na wysoki rozwój gospodarczy, nie otrzymuje znacznych funduszy ze skarbu państwa, czy Unii, są one kierowane w rejony o niższym poziomie rozwoju. Z tego względu województwo dolnośląskie znacznie zyskało w rankingu przedstawionym w tabeli 3. Jest ono połączeniem dość wysokiego rozwoju gospodarczego i wyższych dotacji, jakie mogą otrzymać przedsiębiorcy. Województwa zajmujące dwa ostatnie miejsca - kujawsko-pomorskie i lubuskie otrzymują jedne z najniższych dofinansowań ze Skarbu Państwa na współfinansowanie programów z udziałem funduszy strukturalnych. Lubuskie ponadto posiada zdecydowanie najniższą liczbę absolwentów uczelni wyższych na 10 tys. mieszkańców- 39, gdzie średnia krajowa to 91,44. Wyniki te wpłynęły na pozycję w rankingu tych województw.

PODSUMOWANIE

W oparciu o przeprowadzoną analizę skupień można stwierdzić, iż nadal występuje niewielka dysproporcja pomiędzy czterema województwami tak zwanej ściany wschodniej (podkarpackie, lubelskie, warmińsko-mazurskie, świętokrzyskie), a pozostałymi województwami Polski. Być może intensywne inwestowanie w ten mikroregion poprawi sytuację ekonomiczną co niemal natychmiastowo odbije się na sytuacji małych i średnich przedsiębiorstw. W tej analizie znacznie wybija się województwo mazowieckie. Mając w swoich granicach stolicę przyciąga zagranicznych inwestorów, a także studentów studiujących na jednych z najlepszych uczelni w kraju.

Rangowanie województw potwierdziło powyższe wnioski, lokując województwo mazowieckie na drugim miejscu i cztery województwa ściany wschodniej w drugiej połowie stawki. Rozszerzenie zakresu zmiennych i zmiana metody analizy nieznacznie zmieniła jej wynik. Pierwszą lokatę osiągnęło województwo dolnośląskie, a ostatnie dwa zajęły województwa kujawsko-pomorskie i lubuskie. Może się to wiązać ze znacznie wyższymi środkami

przeznaczonymi na rozwój tak zwanej ściany wschodniej i niewielkimi dla województwa mazowieckiego.

Przeprowadzona analiza potwierdziła, iż występuje dysproporcja pomiędzy województwami Polski. Widoczny jest znaczny interwencjonizm ze strony państwa mający na celu wyrównanie rozwoju w poszczególnych województwach. Scentralizowane decyzje państwa starają się nakłonić przedsiębiorców do zakładania przedsiębiorstw w regionach wschodnich. Regiony te jednak nadal nie zachęcają występującą tam stopą bezrobocia, czy dochodem na osobę. W konsekwencji wysoka jest liczba wymeldowań. Może to być zagrożeniem dla znalezienia kadry profesjonalistów przez zachęconych inicjatywami na start przedsiębiorców.

BIBLIOGRAFIA

- Balicki A. (2013) Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne. Wydawnictwo Uniwersytetu Gdańskiego, 27, 211, 317-318.
- Danielak W., Mierzwa D., Bartczak K. (2017) Małe i średnie przedsiębiorstwa w Polsce. Szanse i zagrożenia rynkowe. Wydawnictwo Exante, 46-50.
- Ostasiewicz W. (1998) Statystyczne metody analizy danych. Wydawnictwo Akademii Ekonomicznej we Wrocławiu, 16, 102.
- Panek T., Zwierzchowski J. (2013) Statystyczne metody wielowymiarowej analizy porównawczej. Teoria i zastosowania. Oficyna Wydawnicza Szkoła Główna Handlowa w Warszawie, 57, 63.
- Skowrońska A., Tarnawa A. (2017) Sektor małych i średnich przedsiębiorstw w Polsce i Unii Europejskiej. Polska Agencja Rozwoju Przedsiębiorczości, 5-15.
- Skowrońska A., Tarnawa A. (2018) Raport o stanie sektora małych i średnich przedsiębiorstw w Polsce. Polska Agencja Rozwoju Przedsiębiorczości, 6-12.
- Skowrońska A., Zakrzewski R. (2019) Raport o stanie sektora małych i średnich przedsiębiorstw w Polsce. Polska Agencja Rozwoju Przedsiębiorczości, 5-17.
- Stanimir A. (2006) Analiza danych marketingowych. Problemy, metody, przykłady. Wydawnictwo AE we Wrocławiu, 121-122.
- Stanisz A. (2007) Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom 3. Analizy wielowymiarowe. StatSoft, 118-127.
- www.parp.gov.pl/storage/publications/pdf/PARP-POPW.pdf [dostęp: 16.08.2019].
- www.polskawschodnia.gov.pl/media/1294/Zal_nr_4_POPW_Diagonoza_wyzwan_potrzeb_potencjalow_obszarow_objetych_POPW.pdf [dostęp: 16.08.2019].
- www.popw.parp.gov.pl/ [dostęp: 16.08.2019].

**A COMPARATIVE ANALYSIS OF CONDITIONS OFFERED
FOR SMALL AND MEDIUM ENTERPRISES
IN THE POLISH VOIVODESHIPS**

Abstract: The study presents the use of cluster analysis using the Ward method to illustrate the conditions offered for small and medium enterprises in individual Polish voivodeships. The study used data from the Central Statistical Office. Using the hierarchical agglomeration procedures, the voivodeships were grouped into homogeneous classes due to the variables characterizing the environment that the voivodeship creates for small and medium enterprises. A ranking in terms of attractiveness for business was also created.

Keywords: cluster analysis, comparative analysis, small and medium enterprises, development pattern method, agglomeration method

JEL classification: C38

WPŁYW INFORMACJI Z USA NA ŚRÓDDZIENNĄ SEZONOWOŚĆ NA PRZYKŁADZIE SPÓŁKI KGHM¹

Tomasz Wójtowicz  <https://orcid.org/0000-0002-9742-3493>

Wydział Zarządzania

AGH Akademia Górniczo-Hutnicza im. S. Staszica w Krakowie

e-mail: twojtow@agh.edu.pl

Streszczenie: Intensywność handlu na rynkach akcji charakteryzuje się wyraźną śróddzienną sezonowością. W przypadku rynków europejskich na kształt tej sezonowości wpływ mają publikacje informacji dotyczących gospodarki USA. W artykule zostaną zaprezentowane wyniki badań oddziaływania tych publikacji na śróddzienną sezonowość wolumenu obrotów oraz zmienności cen akcji spółki KGHM w latach 2001-2016. Analiza dotyczy zarówno siły, jak i długości oddziaływania pojawiających się nowych, ważnych informacji.

Słowa kluczowe: ogłoszenia danych makroekonomicznych, dane śróddzienne, zmienność, śróddzienna sezonowość

JEL classification: G14, E44

WSTĘP

Silny i istotny wpływ gospodarki na rynki kapitałowe nie podlega dyskusji. Ponieważ stan gospodarki może być opisany za pomocą różnych wskaźników, stąd liczne prace empiryczne, w których próbowano zbadać siłę i kierunek oddziaływania wartości tych wskaźników na ceny akcji i zachowanie indeksów giełdowych. W związku z dużym znaczeniem dla inwestorów informacji o stanie gospodarki można również rozważać wpływ samego faktu publikacji wartości wskaźników makroekonomicznych na ceny akcji. Ze względu na swą wagę, tego

¹ Publikacja finansowana przez Akademię Górniczo-Hutniczą im. Stanisława Staszica w Krakowie (dotacja podmiotowa na utrzymanie potencjału badawczego).

typu publikacje powinny wzbudzać wśród inwestorów duże zainteresowanie i ich pojawienie się powinno znajdować odzwierciedlenie na rynkach kapitałowych.

Z uwagi na położenie geograficzne, na ceny akcji na giełdach w Europie najsilniej oddziałują publikacje danych opisujących gospodarkę Stanów Zjednoczonych. Wynika to z faktu, że informacje te pojawiają się w trakcie trwania fazy ciągłej notowań. Oczywiście siła ich oddziaływania wynika również z istotności znaczenia sytuacji w USA dla rynków finansowych na całym świecie.

Dotychczasowe badania [m.in. Harju i Hussain 2011; Gurgul i Wójtowicz 2014, 2015; Będowska-Sójka 2010] pozwoliły ustalić, że publikacje amerykańskich danych makroekonomicznych istotnie i silnie oddziałują na śróddzienne stopy zwrotu i ich zmienność na giełdach w Europie, w szczególności na GPW w Warszawie. Ich pojawienie się wywołuje istotne zmiany ceny oraz wzrost zmienności. Na skutek zmian intensywności handlu na początku i na końcu sesji obserwowana jest podwyższona zmienność cen akcji. Najniższa zmienność występuje w środkowej fazie sesji. Natomiast m.in. [Harju i Hussain 2011] pokazali, że tuż po publikacji danych makroekonomicznych z USA widoczny jest gwałtowny wzrost zmienności na rynkach europejskich. Podwyższona zmienność obserwowana jest również w czasie, gdy równocześnie z europejskimi, otwarte są giełdy w USA.

Dotychczasowe badania prowadzone były głównie z wykorzystaniem śróddziennych danych dotyczących indeksów lub dotyczyły stosunkowo krótkiego okresu czasu. Ponadto, zwykle prowadzone badania dotyczyły reakcji tuż po publikacji danych i nie dotyczyły siły istotnego oddziaływania.

W tym artykule zostanie zaprezentowane badanie oddziaływania publikacji wybranych wskaźników opisujących stan gospodarki USA na zmienność śróddziennych stóp zwrotu akcji KGHM oraz wolumen obrotu w okresie od początku 2001 do końca 2016. Spółka KGHM została wybrana nieprzypadkowo. Jest ona jedną z największych i najważniejszych spółek notowanych na GPW w Warszawie. Ponadto, jest ona notowana na GPW od 10 lipca 1997, a transakcje jej akcjami należą do jednych z najczęściej zawieranych w całym badanym okresie. To gwarantuje, że pojawiająca się nowa informacja powinna szybko znaleźć odzwierciedlenie w ich cenie, a uzyskane wyniki nie będą zniekształcone przez negatywny wpływ efektów mikrostruktury rynku.

W artykule zbadana zostanie nie tylko siła oddziaływania na zmienność i na wielkość obrotów publikacji danych makroekonomicznych z USA, ale również czas, w jakim taki istotny wpływ jest obserwowany.

Określenie siły i długości oddziaływania publikacji amerykańskich danych makroekonomicznych ma duże znaczenie m.in. przy badaniu własności i konstrukcji modeli opisujących śróddzienne stopy zwrotu. Jest to istotne zwłaszcza w przypadku stosowania modeli i testów zakładających np. stałość wariancji stóp zwrotu [por. m.in. Gurgul i Wójtowicz 2014].

WYKORZYSTANE DANE I METODY BADAWCZE

Badanie zostanie przeprowadzone z wykorzystaniem 5-minutowych stóp zwrotu KGHM z okresu od początku 2001 do końca 2016. Rozważane będą tylko stopy zwrotu pochodzące z fazy ciągłej notowań, tzn. z wyłączeniem tzw. stóp zwrotu overnight opisujących zmiany ceny akcji pomiędzy zamknięciem, a otwarciem następnego sesji. Jako miara wielkości obrotów zostanie zastosowany logarytm naturalny wolumenu obrotów w kolejnych 5-minutowych okresach w czasie fazy ciągłej notowań.

Ponieważ celem pracy jest ocena siły i długości oddziaływania informacji makroekonomicznych z USA, to na podstawie wyników dotychczasowych badań, m.in. [Andersen i Bollerslev 1997; Harju i Hussain 2011; Gurgul i Wójtowicz 2014, 2015] wybranych zostało trzynaście wskaźników opisujących różne aspekty amerykańskiej gospodarki:

- wskaźnik cen towarów i usług konsumpcyjnych (**CPI**),
- wskaźnik cen dóbr produkcyjnych (**PPI**),
- liczba miejsc pracy w sektorach pozarolniczych (**NFP**),
- liczba nowych wniosków o zasiłek dla bezrobotnych (**IJC**),
- liczba rozpoczętych budów domów (**HS**),
- sprzedaż nowych domów (**NHS**),
- sprzedaż istniejących domów (**EHS**),
- produkt krajowy brutto (**GDP**),
- sprzedaż detaliczna (**RS**),
- produkcja przemysłowa (**IP**),
- zamówienia na dobra trwałe (**DGO**),
- indeks ISM dla amerykańskiego sektora produkcyjnego (**ISM**),
- wskaźnik zaufania konsumentów wg Conference Board (**CCI**).

Oceniając wpływ jaki mogą mieć publikacje tych wskaźników należy wziąć pod uwagę nie tylko ich znaczenie w opisie sytuacji gospodarczej w USA, ale również to, czy publikowane wartości zawierają nową informację dla inwestorów. Wszystkie rozważane wskaźniki z wyjątkiem IJC, publikowane są co miesiąc i zwykle ich wartość dotyczy miesiąca poprzedniego. Istotnym wyjątkiem są publikacje informacji o produkcie krajowym brutto (GDP), gdzie w kolejnych miesiącach pojawiają się kolejne przybliżenia PKB z ostatniego kwartału. Z drugiej strony, IJC jest publikowany zwykle w każdy czwartek i opisuje sytuację na rynku pracy w poprzednim tygodniu.

Większość z wybranych wskaźników (tzn. CPI, PPI, NFP, IJC, HS, GDP, RS i DGO) jest publikowana o 8:30 EST². Z kolei informacje o NHS, EHS, ISM,

² EST (*Eastern Standard Time*) - Standardowy Czas Wschodni. Jest to strefa czasowa, w której znajduje się wschodnie wybrzeże USA (w tym Nowy Jork). CET (*Central*

CCI są publikowane o 10:00 EST. Tylko ogłoszenia IP pojawiają się o 9:15 EST. Terminom tym odpowiadają zwykle godziny 14:30, 16:00 i 15:15CET³.

Należy jednak zaznaczyć, że nie wszystkie publikacje wspomnianych powyżej wskaźników będą brane pod uwagę w zaprezentowanym badaniu. Ponieważ cenotwórczy charakter mają tylko nieoczekiwane informacje, to uwzględnione w badaniu zostaną tylko te spośród publikacji wskaźników, które zawierały wiadomości odmienne od oczekiwań inwestorów. Oczekiwania te zostały oszacowane na podstawie wartości konsensusu opublikowanego przez agencję Bloomburga kilka dni przed planowanym terminem ogłoszenia poszczególnych wskaźników makroekonomicznych. Jako nieoczekiwane zostały zaklasyfikowane wszystkie publikacje, w których ogłaszana wartość wskaźnika różniła się od konsensusu.

Jedną z bardziej popularnych metod estymacji i opisu śróddziennej sezonowości zmienności jest zastosowanie elastycznej formy Fouriera (FFF) [por. Andersen i Bollerslev 1997]. Przyjmijmy, że sesja jest podzielona na N 5-minutowych podokresów, a 5-min logarytmiczną stopę zwrotu $R_{t,n}$ obliczoną w n -tym okresie dnia t możemy zapisać w postaci:

$$R_{t,n} = E(R_{t,n}) + s_{t,n} \frac{\sigma_t}{\sqrt{N}} Z_{t,n} \quad (1)$$

gdzie innowacje $Z_{t,n}$ są i.i.d. $(0,1)$, σ_t/\sqrt{N} jest czynnikiem opisującym wpływ zmienności danych dziennych na śróddzienną zmienność, a $s_{t,n}$ opisuje wartość śróddziennej składowej cyklicznej zmienności w n -tym okresie. Wtedy podstawą elastycznej formy Fouriera jest założenie, że składowa $\ln s_{t,n}$ jest sumą trendu kwadratowego i pewnej liczby czynników rozwinięcia Fouriera uzupełnioną o dodatkowe sztuczne zmienne i można ją estymować z modelu:

$$2 \ln \frac{|R_{t,n} - \bar{R}|}{(\hat{\sigma}_t/\sqrt{N})} = c + \delta_1 \frac{n}{N_1} + \delta_2 \frac{n^2}{N_2} + \sum_{k=1}^D \lambda_k I_k(t, n) + \sum_{p=1}^P \left(\delta_{c,p} \cos\left(\frac{2\pi p}{N} n\right) + \delta_{s,p} \sin\left(\frac{2\pi p}{N} n\right) \right) + \varepsilon_{t,n}, \quad (2)$$

gdzie N_1 i N_2 są stałymi normującymi, p jest rzędem rozwinięcia w szereg Fouriera, \bar{R} jest średnią 5-min stóp zwrotu, $\hat{\sigma}_t$ jest oszacowaniem dziennego warunkowego odchylenia standardowego (np. z modelu GARCH), a $I_k(t, n)$ są sztucznymi zmiennymi zero-jedynkowymi opisującymi m.in. wpływ publikacji amerykańskich danych makroekonomicznych. Parametry modelu (1) można

European Time) – Czas Środkowoeuropejski. Większość godzin w tym artykule będzie podawanych jako czas środkowoeuropejski.

³ Ze względu na różne terminy przechodzenia w USA i Europie z czasu letniego na zimowy i odwrotnie, terminy te mogą ulec przesunięciu o jedną godzinę.

estymować za pomocą MNK, jednak ze względu na heteroskedastyczność i autokorelację reszt, w badaniu ich istotności wykorzystane zostaną odporne estymatory macierzy wariancji-kowariancji Neweya-Westa [Newey i West 1987].

W analogiczny sposób będzie estymowana składowa deterministyczna śróddziennych obrotów, z tym, że w tym przypadku lewa strona (2) będzie równa logarytmowi 5-minutowego wolumenu skorygowanego o prognozę logarytmu wolumenu dla danych dziennych.

Kluczowym elementem badań zaprezentowanych w tym artykule będzie odpowiedni dobór sztucznych zmiennych I_k i analiza istotności parametrów λ_k . W celu odpowiedniego ujęcia wpływu informacji z USA na zmienność śróddziennych stóp zwrotu rozważane będą następujące sztuczne zmienne zero-jedynkowe:

- I_{NYSE} - opisująca wpływ otwarcia giełd w USA; przyjmuje ona wartość 1 począwszy od 15:30, aż do końca sesji w dniach otwarcia giełdy w Nowym Jorku,
- $I_{wsk}(1), \dots, I_{wsk}(12)$ - opisujące wpływ publikacji poszczególnych wskaźników (indeks *wsk*) w ciągu godziny po ich ogłoszeniu (tzn. w 12 okresach o długości 5 minut). Każda ze zmiennych przyjmuje wartość 1 w danym 5-min okresie po publikacji wskaźnika, np. zmienna $I_{GDP}(4)$ jest równa 1 pomiędzy 15 a 20 minutą po terminie publikacji informacji o PKB Stanów Zjednoczonych.

Oprócz zmiennych opisanych powyżej, w modelu (2) uwzględnione zostaną również zmienne opisujące potencjalny efekt dnia tygodnia. Dodatkowo, w celu uniknięcia niekorzystnego wpływu wzrostu zmienności na początku i na końcu sesji estymacja modelu (2) zostanie przeprowadzona z wyłączeniem danych z pierwszych 15 minut i ostatnich 15 minut fazy ciągłej⁴.

WPŁYW OGŁOSZEŃ NA ZMIENNOŚĆ

Na podstawie 5-minutowych stóp zwrotu KGHM z okresu od 2 stycznia 2001 do 31 grudnia 2016 wyestymowano parametry regresji (2) uwzględniającej sztuczne zmienne opisane powyżej. Na podstawie kryterium informacyjnego Akaike wybrany został model z $P = 3$. Jako przykład, na rysunku 1 zaprezentowano wykres oszacowanych w ten sposób wartości składowej sezonowej $s_{t,n}$ w dniach 17-21 listopada 2011. Wtedy miały miejsce publikacje odbiegających od oczekiwań wartości następujących wskaźników:

- 18 listopada: PPI (godz. 14:30),

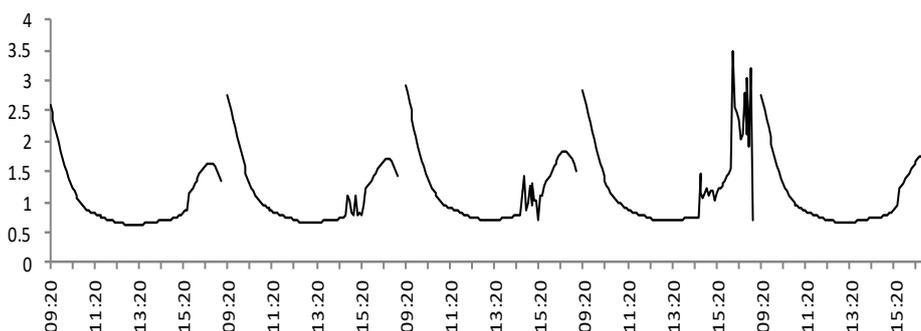
⁴ W procesie estymacji należy oczywiście uwzględnić zmiany harmonogramu sesji, które miały miejsce na GPW kilka razy w całym badanym okresie.

- 19 listopada: CPI i HS (godz. 14:30),
- 20 listopada: IJC (godz. 14:30), EHS (godz. 16:00).

W każdym dniu można o 15:30 zaobserwować wzrost zmienności stóp zwrotu KGHM wywołany otwarciem giełdy w Nowym Jorku. Ponadto, w dniach publikacji wskaźników makroekonomicznych widoczny jest ich wyraźny wpływ na składową deterministyczną zmienności. W zależności od publikowanego wskaźnika, gwałtowny wzrost zmienności tuż po ogłoszeniu występuje razem z mniejszymi lub większymi wahaniami widocznymi w kolejnych 5-minutowych okresach.

Uzyskane w wyniku przeprowadzonej estymacji oszacowania trendu kwadratowego i rozwinięcia Fouriera w (2) okazały się istotne. Istotny na poziomie 1% jest również parametr przy zmiennej I_{NYSE} . Jego wartość (0,46) wskazuje, że otwarcie giełdy w Nowym Jorku powoduje średni wzrost odchylenia standardowego 5-min stóp zwrotu KGHM o ok. 30%.

Rysunek 1. Składowa deterministyczna $s_{t,n}$ śróddziennej sezonowości zmienności stóp zwrotu KGHM w dniach 17-21 listopada 2011



Źródło: obliczenia własne

Z punktu widzenia celu tego artykułu kluczowa jest jednak analiza wartości i istotności parametrów stojących przy zmiennych opisujących oddziaływanie publikacji danych makroekonomicznych. Ponieważ nie wszystkie spośród tych 156 zmiennych okazały się istotne, to w celu uproszczenia analizy uzyskanych wyników, w tabeli 1 zebrane zostały najważniejsze informacje. Zaprezentowane zostały w niej wartości istotnych (na poziomie 1%) parametrów stojących przy zmiennych $I_{wsk}(1)$, tzn. opisujących zmiany wariancji w pierwszych pięciu minutach po publikacji. W trzeciej kolumnie podany został czas nieprzerwanego istotnego oddziaływania publikacji poszczególnych wskaźników. Ze względu na to, że dla każdego ze wskaźników rozważanych było tylko 12 sztucznych zmiennych I_{wsk} , to oszacowanie to jest z góry ograniczone do 60 minut.

Tabela 1. Siła reakcji zmienności stóp zwrotu KGHM i czas oddziaływania publikacji wskaźników makroekonomicznych

Wskaźnik	Wartość $I_{wsk}(1)$	Czas oddziaływania (w min)
CCI	1,23	15
DGO	1,11	10
ISM	1,20	5
IJC	1,30	60
NFP	2,53	25
GDP	1,71	10
RS	1,16	5

Źródło: obliczenia własne

Na podstawie wyników zamieszczonych w tabeli 1 można stwierdzić, że w całym badanym okresie istotny wzrost zmienności w ciągu pierwszych pięciu minut po publikacji wywołują nieoczekiwane informacje dotyczące siedmiu spośród rozważanych wskaźników. Są to informacje o: rynku pracy (NFP i IJC), produkcie krajowym brutto (GDP), sprzedaży detalicznej (RS), zamówieniach na dobra trwałe (DGO) oraz wskaźnik aktywności w przemyśle (ISM) i wskaźnik nastrojów konsumentów (CCI). Przeprowadzona analiza śróddziennej sezonowości pokazuje również, że w całym badanym okresie nieistotny wpływ na zmienność stóp zwrotu KGHM miały informacje o typowych wskaźnikach inflacji (CPI i PPI), a także informacje z rynku mieszkań (HS, NHS, EHS).

Najsilniejszą reakcję inwestorów przejawiającą się prawie trzyipółkrotnym wzrostem odchylenia standardowego, wywołują publikacje liczby miejsc pracy w sektorach pozarolniczych. Na podstawie dotychczasowych wyników [m.in. Andersen i in. 2007; Gurgul i Wójtowicz 2014, 2015; Suliga i Wójtowicz 2013] ten wynik nie jest zaskoczeniem, gdyż publikacje NFP są jednymi z najsilniej oddziałujących na inwestorów informacji o stanie gospodarki USA. Wynika to z tego, że wartość NFP jest częścią „Raportu o zatrudnieniu”, który jest publikowany zwykle w pierwszy piątek miesiąca i opisuje rynek pracy w USA w miesiącu wcześniejszym. Jest to, więc pierwsze poważne opracowanie dotyczące aktualnej sytuacji makroekonomicznej USA. Wartości kolejnych wskaźników, m.in. GDP, inflacji itp., są publikowane dopiero w okolicy połowy miesiąca i mogą być już częściowo prognozowane na podstawie znajomości informacji o rynku pracy. Niezależnie jednak od tego, te późniejsze publikacje również wywołują istotny wzrost zmienności cen akcji KGHM. Przykładowo, informacje o kolejnych przybliżeniach wartości produktu krajowego brutto wywołują w ciągu pierwszych 5 minut ponad dwukrotny wzrost odchylenia standardowego stóp zwrotu KGHM.

Inwestorzy zwracają również uwagę na publikowane co tydzień informacje o liczbie nowych bezrobotnych (IJC). Dzieje się tak niezależnie od dużego zaszumienia tych wartości wynikającego właśnie z cotygodniowego charakteru publikacji.

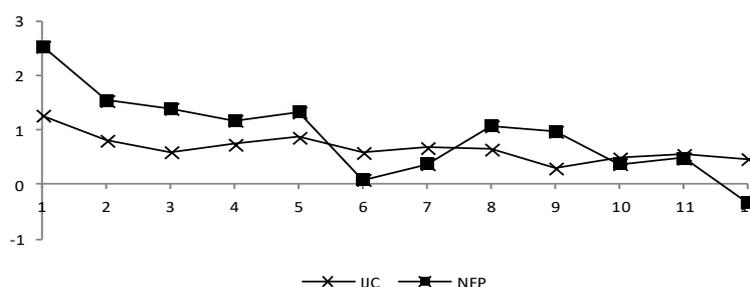
Oprócz siły oddziaływania publikowanych wskaźników na zmienność stóp zwrotu ważne jest, by określić, jak długo pojawiające się informacje powodują istotne zmiany. Wyniki z ostatniej kolumny tabeli 1 wskazują, że w zależności od wskaźnika istotny wpływ na zmienność trwa od 5 minut (RS i ISM) aż do godziny (IJC). W przypadku NFP ta istotność (na poziomie 1%) jest widoczna nieprzerwanie przez pierwsze 25 minut. Jeżeli jednak rozważymy istotność na poziomie 5%, to można ją zaobserwować aż do 45 minut po publikacji.

W przypadku większości wskaźników, tych publikowanych o 14:30, badanie długości oddziaływania ich publikacji z konieczności musi być ograniczone do 60 minut. Po tym czasie, tzn. o 15:30, gdy rozpoczyna się sesja w Nowym Jorku, wpływ ten jest zakłócony przez wtórne sygnały powstające w wyniku obserwacji reakcji inwestorów na amerykańskim rynku akcji.

Aby dokładniej zaprezentować zmiany siły oddziaływania nieoczekiwanych informacji zawartych w amerykańskich danych makroekonomicznych w kolejnych minutach po ich publikacji, na rysunku 2 przedstawiono wartości parametrów przy zmiennych $I_{NFP}(1), \dots, I_{NFP}(12)$ oraz $I_{IJC}(1), \dots, I_{IJC}(12)$. Jak już wcześniej wspomniano wartości te opisują zmiany wariancji stóp zwrotu KGHM w kolejnych 5-minutowych okresach po publikacji.

Na tej podstawie można zauważyć powolne osłabianie powstałego impulsu. Niezależnie jednak od tego (tam, gdzie zmiany są istotne, co najmniej na poziomie 5%) publikowane informacje powodują wzrost zmienności. W przypadku NFP, nawet po 25 minutach od terminu ogłoszenia obserwujemy prawie dwukrotny wzrost odchylenia standardowego stóp zwrotu. Z kolei publikacje wartości IJC powodują, że nawet po godzinie od publikacji odchylenie standardowe stóp zwrotu KGHM jest podwyższone o ok. 30%.

Rysunek 2. Siła oddziaływania publikacji IJC i NFP w kolejnych 5-minutowych okresach



Źródło: opracowanie własne

WPLYW OGŁOSZEŃ NA ZMIENNOŚĆ

Estymacja elastycznej formy Fouriera dla wolumenu obrotów pozwoliła potwierdzić istotny wpływ otwarcia rynków akcji w USA na intensywność handlu na GPW w Warszawie. W dniach otwarcia NYSE obserwowany jest od 15:30 wzrost wielkości obrotów akcjami KGHM średnio o ok. 30%.

W tabeli 2, podobnie jak wcześniej, zebrane zostały wartości współczynników przy zmiennych $I_{wsk}(1)$ istotnych, co najmniej na poziomie 5%, chociaż zdecydowana większość z nich jest istotna na poziomie 1%. Trzecia kolumna zawiera informacje o czasie istotnego oddziaływania tych publikacji na wielkość obrotów akcjami KGHM.

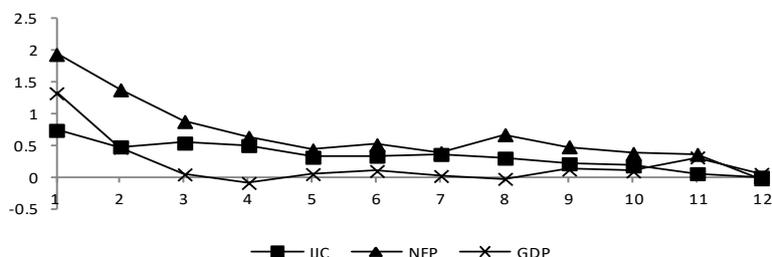
Uzyskane wyniki wskazują, że istotny wzrost wolumenu obrotów wywoływany jest przez większą liczbę wskaźników niż miało to miejsce w przypadku zmienności. Oprócz wskaźników wymienionych w tabeli 1 istotny wzrost zmienności wywołują publikacje informacji o produkcji przemysłowej (IP), rozpoczętych budowach domów jednorodzinnych (HS) i wskaźniku cen dóbr produkcyjnych (PPI). Niemniej jednak, najsilniejsze zainteresowanie inwestorów powodują publikacje liczby miejsc pracy w sektorach pozarolniczych (NFP). W pierwszych 5 minutach po ogłoszeniu powodują one blisko siedmiokrotny wzrost wolumenu. Co więcej, istotnie podwyższone obroty wywołane tymi publikacjami utrzymują się przez 55 minut. Porównując wykresy przedstawione na rysunku 3 można zauważyć, że oddziaływanie publikacji wskaźników makroekonomicznych prawie całkowicie zanika po 60 minutach niezależnie od swojej pierwotnej siły.

Tabela 2. Siła reakcji śróddziennej wielkości obrotów akcji KGHM i czas oddziaływania publikacji wskaźników makroekonomicznych

Wskaźnik	Wartość $I_{wsk}(1)$	Czas oddziaływania (w min)
CCI	0,53	5
DGO	0,92	5
HS	0,58	25
IP	0,41	5
ISM	0,58	5
IJC	0,74	50
NFP	1,93	55
PPI	0,67	5
GDP	1,32	10
RS	0,69	5

Źródło: obliczenia własne

Rysunek 3. Siła oddziaływania publikacji IJC, NFP i GDP na wolumen w kolejnych 5-minutowych okresach



Źródło: obliczenia własne

PODSUMOWANIE

Zaprezentowane w tym artykule badanie oddziaływania publikacji wybranych wskaźników makroekonomicznych opisujących gospodarkę USA pokazało, że te informacje mają istotny wpływ na śróddzienną zmienność stóp zwrotu KGHM. Co więcej, w przypadku niektórych z rozważanych ogłoszeń (np. NFP lub IJC) ten istotny wpływ, tzn. podwyższona zmienność, utrzymuje się przez dłuższy czas po terminie publikacji. Analogiczny, istotny wpływ, publikacje tych wskaźników mają na wolumen obrotów akcjami KGHM. W tym przypadku również najsilniej i najdłużej oddziałują nieoczekiwane informacje o liczbie nowych miejsc pracy w sektorach pozarolniczych (NFP). Uzyskane wyniki mają duże praktyczne znaczenie dla modelowania śróddziennych stóp zwrotu. Gwałtowne zmiany wariancji i wolumenu obrotów powinny zostać uwzględnione przy konstrukcji odpowiednich modeli lub w testach wykorzystujących dane śróddzienne.

BIBLIOGRAFIA

- Andersen T. G., Bollerslev T. (1997) Intraday Periodicity and Volatility Persistence in Financial Markets. *Journal of Empirical Finance*, 4, 115-158.
- Andersen T. G., Bollerslev T., Diebold F., Vega C. (2007) Real-Time Price Discovery in Global Stock, Bond and Foreign Exchange Markets. *Journal of International Economics*, 73, 251-277.
- Będowska-Sójka B. (2010) Intraday CAC40, DAX and WIG20 Returns when the American Macro News Is Announced. *Bank i Kredyt*, 41(2), 7-20.
- Gurgul H., Wójtowicz T. (2014) The Impact of US Macroeconomic News on the Polish Stock Market. The Importance of Company Size to Information Flow. *Central European Journal of Operations Research*, 22, 795-817.
- Gurgul H., Wójtowicz T. (2015) The Response of Intraday ATX Returns to U.S. Macroeconomic News. *Finance a úvěr – Czech Journal of Economics and Finance*, 65(3), 230-253.

- Harju K., Hussain S. (2011) Intraday Seasonalities and Macroeconomic News Announcements. *European Financial Management*, 17, 367-390.
- Newey W., West K. (1987) A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3), 703-708.
- Suliga M., Wójtowicz T. (2013) The Reaction of the WSE to U.S. Employment News Announcements. *Managerial Economics*, 14, 39-60.

**THE IMPORTANCE OF INFORMATION
FROM THE US ECONOMY FOR INTRADAY SEASONALITY.
EVIDENCE FOR KGHM**

Abstract: Intensity of trading on stock markets is characterized by a visible intraday seasonality pattern. In the case of European markets, this seasonality is strongly influenced by announcements of information about the US economy. In this paper we study the impact of these publications on intraday seasonality of trading volume and volatility of KGHM returns in the period from 2001 to 2016. The analysis concerns both the strength and the length of the impact of new, important information.

Keywords: macroeconomic news announcements, intraday data, volatility, intraday seasonality

JEL classification: G14, E44