Wiadomości Statystyczne. The Polish Statistician, 2023, *68*(10), 1–23 https://doi.org/10.59139/ws.2023.10.1
Statystyka w praktyce / Statistics in practice
Zgłoszony/submitted: 20.09.2022, poprawiony/revised: 5.07.2023, zaakceptowany/accepted: 26.08.2023

Zastosowanie jednowskaźnikowego semiparametrycznego modelu ekonometrycznego w analizie ryzyka inwestycyjnego[1]

Dominik Krężołek[a]

Streszczenie. Jednym z zadań ekonometrii stosowanej i statystyki jest estymacja funkcji średniej warunkowej w założonym modelu. Dostępne metody estymacji takiej funkcji i wyniki oszacowania zależą głównie od przyjętych *a priori* założeń dotyczących populacji lub procesu, który generuje dane. Celem badania omówionego w artykule jest ocena możliwości zastosowania jednowskaźnikowego semiparametrycznego modelu ekonometrycznego do pomiaru ryzyka inwestycyjnego na Giełdzie Papierów Wartościowych w Warszawie (GPW). Taki model charakteryzuje się tym, że niektóre jego założenia są mniej restrykcyjne niż założenia modeli parametrycznych dla funkcji średniej warunkowej (takich jak model liniowy i binarny model probitowy), a jednocześnie zachowuje wiele pożądanych cech modelu liniowego i metody najmniejszych kwadratów.

W przeprowadzonym badaniu semiparametryczny model jednowskaźnikowy został wykorzystany do pomiaru ryzyka dla 10 wybranych spółek sektora informatycznego notowanych na GPW od 2018 r. do 2021 r. Dane pobrano z serwisu finansowego Bloomberg. Jako czynniki ryzyka przyjęto stopy zwrotu dwóch indeksów giełdowych: WIG20 (Polska) i S&P 500 (Stany Zjednoczone). Uzyskane wyniki wskazują, że polski rynek znacznie silniej niż amerykański determinuje zmienność stóp zwrotu badanych spółek. Z badania wynika również, że modele semiparametryczne są bardziej elastyczne niż modele parametryczne w odniesieniu do założeń teoretycznych, co w warunkach napływu ogromnej ilości informacji może ułatwiać podejmowanie właściwych decyzji inwestycyjnych.

Słowa kluczowe: semiparametryczny model jednowskaźnikowy, analiza ryzyka, Giełda Papierów Wartościowych w Warszawie, GPW

JEL: C14, C50, C58

Application of single-index semiparametric econometric model in investment risk analysis

Abstract. One of the tasks of applied econometrics and statistics is the estimation of the conditional mean function of the assumed model. The available methods for estimating such a function and the estimation results depend mostly on the *a priori* assumptions about the

[1] Artykuł został opracowany na podstawie referatu wygłoszonego na XV Międzynarodowej Konferencji Naukowej im. Profesora Aleksandra Zeliasia „Modelowanie i prognozowanie zjawisk społeczno-gospodarczych", która odbyła się w dniach 9–12 maja 2022 r. w Zakopanem. / The article is based on a paper delivered at the 15th Professor Alexander Zelias International Conference on 'Modeling and Forecasting of Socio-economic Phenomena', held on 9–12 May 2022 in Zakopane.

[a] Uniwersytet Ekonomiczny w Katowicach, Wydział Informatyki i Komunikacji, Polska / University of Economics in Katowice, Faculty of Informatics and Communication, Poland.
ORCID: https://orcid.org/0000-0002-4333-9405. E-mail: dominik.krezolek@ue.katowice.pl.

population or process that generates the data. The aim of the research presented in this paper is to apply single-index semiparametric econometric model to measure investment risk on the Warsaw Stock Exchange (WSE). Such a model is characterised by some assumptions less restrictive than in the case of other parametric models for the conditional mean function, such as a linear model or a binary probit model. At the same time, a single-index model retains many of the desirable features of a linear model and a least squares method.

The presented model was used to measure investment risk for 10 IT companies quoted on the WSE in the period from 2018 to 2021. The data came from the Bloomberg financial service. Rates of return of two stock market indices, WIG20 (Poland) and S&P 500 (USA), were adopted as risk factors. The results indicate that the Polish market determines the volatility of returns of the analysed companies to a much larger extent than is the case with the US market. Furthermore, semiparametric models proved more flexible than the parametric ones regarding theoretical assumptions, which in the event of a large inflow of information might facilitate making correct investment decisions.

Keywords: semiparametric single-index model, risk analysis, Warsaw Stock Exchange

1. Wprowadzenie

Metody nieparametryczne to techniki statystyczne i ekonometryczne, które nie wymagają od badacza określenia postaci funkcji dla szacowanych modeli. Dane prezentowane w formie nieparametrycznej informują o kształtowaniu się jakiegoś zjawiska. Najprostszą taką formą jest histogram empiryczny (w metodach parametrycznych dopasowuje się do niego odpowiednią funkcję gęstości). W przypadku modeli regresji podejście nieparametryczne jest znane jako *regresja nieparametryczna* lub *wygładzanie nieparametryczne*. Wśród metod nieparametrycznych szczególną rolę odgrywa estymacja jądrowa, która pozwala empirycznie oszacować funkcję gęstości lub dystrybuantę analizowanej zmiennej losowej.

Metody nieparametryczne są coraz szerzej wykorzystywane w analizie danych użytkowych. Najlepiej sprawdzają się w pracy z dużymi zbiorami danych, co pierwotnie spotykało się z krytyką, głównie ze względu na złożoność obliczeniową, niedokładność obliczeń czy brak statystycznej istotności wyników (Wasserman, 2006). Z tych powodów z metod nieparametrycznych często korzysta się, gdy powszechnie stosowane specyfikacje parametryczne okazują się nieodpowiednie w przypadku badanego problemu. Dzieje się tak zwłaszcza wtedy, gdy formalne odrzucenie modelu parametrycznego na podstawie testów specyfikacji nie daje żadnych wskazówek co do kierunku poszukiwań ulepszonego modelu parametrycznego. Popularność metod nieparametrycznych wynika z tego, że nie wymagają spełnienia niektórych sztywnych założeń nałożonych na proces generowania danych, co jest konieczne w przypadku modeli parametrycznych. Innymi słowy, metody nieparametryczne pozwalają, aby to dane same określiły odpowiedni model.

W ciągu ostatnich kilkudziesięciu lat metody nieparametryczne, jak również semiparametryczne stały się przedmiotem dużego zainteresowania statystyków. Model semiparametryczny stanowi połączenie modeli parametrycznego i niepara-

metrycznego – jedna jego część jest określona przez znane parametry (jak w modelach parametrycznych), a druga – elastyczna i nieokreślona przez stałą liczbę parametrów (jak w modelach nieparametrycznych).

W latach 50. XX w. opublikowano pierwszą pracę na temat estymacji jądrowej (Rosenblatt, 1956). Rozwiązanie to zaproponowano w raporcie technicznym Sił Powietrznych Stanów Zjednoczonych jako sposób na uwolnienie analizy dyskryminacyjnej od sztywnych specyfikacji parametrycznych (Fix i Hodges, 1951). Na przestrzeni lat nastąpił gwałtowny rozwój omawianej dziedziny, zarówno w wymiarze teoretycznym, jak i praktycznym.

Celem badania omawianego w artykule jest ocena możliwości zastosowania jednowskaźnikowego semiparametrycznego modelu ekonometrycznego do pomiaru ryzyka inwestycyjnego na Giełdzie Papierów Wartościowych w Warszawie (GPW). Analizę przeprowadzono dla wybranych spółek sektora informatycznego notowanych na GWP w latach 2018–2021. Spółki te są składowymi indeksu WIG-informatyka, a jako czynniki ryzyka przyjęto stopy zwrotu dwóch indeksów giełdowych: polskiego WIG20 i amerykańskiego S&P 500. Badanie przeprowadzono na podstawie dziennych logarytmicznych stóp zwrotu analizowanych walorów. Postawiono hipotezę, że semiparametryczny model jednowskaźnikowy umożliwia szacowanie ryzyka determinowanego zmiennością rynku przy relatywnie niskich wartościach średniego błędu absolutnego (ang. *mean absolute error* – MAE). Ponadto przyjęto, że spółki sektora informatycznego w Polsce w większym stopniu zależą od zmienności obserwowanej na giełdzie polskiej niż amerykańskiej.

2. Semiparametryczny model jednowskaźnikowy

Metody semiparametryczne to jedne z najpopularniejszych metod bazujących na estymacji elastycznej, odnoszącej się do wyznaczenia funkcji, która – inaczej niż w tradycyjnych modelach parametrycznych – nie jest ściśle określona przez stałą liczbę parametrów. Funkcja jest elastyczna i można ją dostosowywać do skomplikowanych wzorców w zbiorze danych bez konieczności określania *a priori* jej formy. Dzięki temu w modelu można uwzględnić nieliniowe zależności w zbiorze danych, co mogłoby się okazać trudne w przypadku zastosowania typowych modeli parametrycznych (Silverman, 1986). Modele semiparametryczne, powstałe dzięki połączeniu modeli parametrycznych i nieparametrycznych, są przydatne w sytuacjach, w których narzędzia w pełni nieparametryczne nie dają właściwych oszacowań. Do takiej sytuacji dochodzi, gdy nadwymiarowość analizowanego zbioru danych prowadzi do bardzo zróżnicowanych oszacowań lub gdy nie jest znana postać funkcyjna określonej zależności w odniesieniu do podzbioru regresorów lub rozkładu składnika losowego. Wspomniana nadmiarowość dotyczy zbiorów danych charakteryzujących się dużą liczbą cech i zmiennych. Może się też zdarzyć, że niektóre regresory

występują jako funkcja liniowa względem zmiennych, ale postać funkcyjna parametrów w odniesieniu do innych zmiennych nie jest znana, lub że funkcja regresji jest nieparametryczna, ale struktura składnika losowego ma postać parametryczną (Li i Racine, 2007).

Modele semiparametryczne są postrzegane jako rodzaj kompromisowego rozwiązania między specyfikacjami w pełni nieparametrycznymi i w pełni parametrycznymi. Opierają się na założeniach parametrycznych i dlatego mogą być błędnie określane, tak jak ich parametryczne odpowiedniki. Znanych jest wiele metod semiparametrycznych. W dalszej części artykułu ograniczono się do opisu modelu typu regresyjnego, a dokładnej – modelu jednowskaźnikowego.

Dokonując analizy dużych zbiorów danych, badacze niejednokrotnie spotykają się ze zjawiskiem nadwymiarowości w zbiorze danych. Jednym ze sposobów jej redukcji jest zastosowanie modelu jednowskaźnikowego. W modelu tego typu występuje jednowymiarowy wskaźnik (skalar), który stanowi element nieznanej funkcji nieparametrycznej. Modele jednowskaźnikowe zawdzięczają swoją popularność temu, że działają podobnie jak wiele znanych modeli parametrycznych, np. regresja Poissona, modele logitowe, probitowe i tobitowe (Härdle i Stoker, 1989).

Model jednowskaźnikowy to jeden z najpopularniejszych modeli semiparametrycznych stosowanych w naukach ilościowych. Estymacja wektora $\boldsymbol{\beta}$ współczynników modelu była analizowana przez wielu badaczy, którzy rozważali kwestię efektywności. Wśród wykorzystywanych do tego celu technik należy wymienić metody: średniej pochodnej (Powell i in., 1989; Stoker, 1986), odwróconej regresji krokowej (Duan i Li, 1991; Li, 1991), najmniejszych kwadratów (Härdle i in., 1993; Li i in., 2014; Yang i Song, 2014), minimalnej uśrednionej warunkowej wariancji (Xia, 2006; Xia i in., 2002) i empirycznego prawdopodobieństwa (Xue, 2013; Xue i Zhu, 2006; Zhu i Xue, 2006). Model jednowskaźnikowy należy uznać za atrakcyjne i efektywne podejście do nieparametrycznego wielowymiarowego modelowania, ponieważ redukuje on problem nadwymiarowości zbioru danych przy jednoczesnym zachowaniu interpretowalności modeli liniowych.

Semiparametryczne modele jednowskaźnikowe mają przewagę nad klasycznymi modelami liniowymi w kilku kluczowych aspektach (Pagan i Ullah, 1999). Pierwszy z nich to elastyczność dotycząca kształtu funkcji. W modelu liniowym zakłada się zachodzenie liniowej relacji między zmiennymi objaśniającymi a zmienną zależną, co może prowadzić do niedoszacowania lub przeszacowania efektów zmiennych objaśniających, jeśli rzeczywista relacja jest nieliniowa. W jednowskaźnikowym semiparametrycznym modelu ekonometrycznym można uwzględnić nieliniowe zależności między zmiennymi, co pozwala na dokładniejsze oszacowanie efektów zmiennych objaśniających. Istotną cechą tego modelu jest także brak założeń dotyczących rozkładu składnika losowego. W modelu liniowym wymaga się, aby błędy

miały rozkład normalny, co w niektórych przypadkach może być niewłaściwe. W jednowskaźnikowym semiparametrycznym modelu ekonometrycznym taki warunek nie jest wymagany, co czyni ten model bardziej odpornym na nieodpowiednie założenia. Inną właściwością podejścia semiparametrycznego jest brak założeń dotyczących homoskedastyczności składnika losowego. W modelu liniowym zakłada się, że błędy mają stałą wariancję. Ten warunek może okazać się nieodpowiedni w przypadku, gdy wariancja błędów zależy od wartości zmiennych objaśniających (Racine, 2008). Jednowskaźnikowy semiparametryczny model ekonometryczny nie wymaga takiego założenia i może dostarczyć bardziej wiarygodnych oszacowań, gdy homoskedastyczność wariancji składnika losowego nie występuje. Ponadto istnieje możliwość uwzględnienia efektów nieliniowych zmiennych w wariancji błędów, a także nieliniowych zależności między zmiennymi objaśniającymi a błędami, co może się przyczyniać do zwiększania dokładności oszacowań. Za kolejną zaletę modelu jednowskaźnikowego należy uznać to, że oszacowania składnika parametrycznego są asymptotycznie zbieżne z rzeczywistymi w tempie podobnym do modelu w pełni parametrycznego. Te korzyści wiążą się jednak z pewnymi kosztami, np. trudnościami z włączeniem do wskaźnika (skalara) zmiennych dyskretnych lub w ogóle brakiem takiej możliwości czy problemami z określeniem właściwej struktury modelu, głównie w kontekście zastosowanej funkcji łączącej i funkcji składowych takiego modelu (Henderson i Parmeter, 2015).

Ogólna postać modelu jednowskaźnikowego jest następująca (Horowitz, 2009):

$$y_i = m\big(\varphi(\boldsymbol{x}_i, \boldsymbol{\beta})\big) + u_i, \tag{1}$$

gdzie:

$m(\cdot)$ – nieznana funkcja wygładzająca,

$\varphi(\cdot,\cdot)$ – znana funkcja parametryczna q zmiennych objaśniających $\boldsymbol{x}_i$ i p-wymiarowego wektora $\boldsymbol{\beta}$ współczynników,

u_i – składnik losowy nieskorelowany z $\boldsymbol{x}_i$,

$i = 1, 2, \dots, n$.

Nawet jeśli $\varphi(\boldsymbol{x}_i, \boldsymbol{\beta})$ jest funkcją nieliniową, to uważa się ją za pojedynczy wskaźnik, ponieważ $\varphi(\boldsymbol{x}_i, \boldsymbol{\beta})$ jest skalarem.

W większości badań ekonomicznych przyjmuje się, że mamy do czynienia z pojedynczym liniowym wskaźnikiem (skalarem), a zatem funkcja $\varphi(\cdot,\cdot)$ przyjmuje postać:

$$\varphi(\boldsymbol{x}_i, \boldsymbol{\beta}) = \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_q x_{qi}, \tag{2}$$

gdzie liczba zmiennych niezależnych jest równa liczbie szacowanych parametrów, tj. $p = q$.

Postać macierzowa modelu jednowskaźnikowego ma więc formę (Ramanathan, 1989):

$$y_i = m(\boldsymbol{x}_i\boldsymbol{\beta}) + u_i, \tag{3}$$

gdzie:
$\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{qi})$,
$\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_q)'$,
$i = 1, 2, \ldots, n$.

Warto zauważyć, że w modelu (2) nie występuje wyraz wolny, ponieważ nie można zidentyfikować punktu przecięcia bez ograniczeń nałożonych na nieznaną funkcję, a wektor parametrów $\boldsymbol{\beta}$ zazwyczaj nie może być zidentyfikowany bez normalizacji (podobnie jak w przypadku modeli logitowych i probitowych). W literaturze najczęściej występują dwie normalizacje. Pierwsza z nich polega na ustawieniu współczynnika pierwszego regresora na poziomie 1 (przy założeniu, że prawdziwy parametr jest różny od 0), a w drugiej zakłada się, że wektor $\boldsymbol{\beta}$ ma długość jednostkową (tj. jego norma euklidesowa wynosi 1). Więcej szczegółowych informacji na temat normalizacji parametrów w modelach jednowskaźnikowych można znaleźć w pracy Camerona i Trivediego (2005).

Podstawowa różnica pomiędzy modelem parametrycznym a semiparametrycznym polega na wykorzystaniu funkcji wygładzającej $m(\cdot)$. W modelu parametrycznym ta funkcja jest znana i możliwe jest zastosowanie estymacji nieliniowej metodą najmniejszych kwadratów lub estymacji metodą największej wiarygodności. W przypadku modelu semiparametrycznego funkcja $m(\cdot)$ jest nieznana i należy ją oszacować. Warto zauważyć, że w sensie ogólnym $m(\cdot)$ jest średnią warunkową y względem $\boldsymbol{x\beta}$. Wynika z tego wybór przeprowadzenia nieparametrycznej regresji y względem $\boldsymbol{x\beta}$. Gdyby wektor $\boldsymbol{\beta}$ był znany, taka regresja byłaby bardzo prosta, jednak w praktyce jest on nieznany, dlatego należy oszacować zarówno $m(\cdot)$, jak i $\boldsymbol{\beta}$.

2.1. Estymacja parametrów semiparametrycznego modelu jednowskaźnikowego

Przy założeniu, że funkcja wygładzająca $m(\cdot)$ jest znana, oszacowanie wektora $\boldsymbol{\beta}$ można uzyskać za pomocą nieliniowej metody najmniejszych kwadratów. W takim przypadku należy minimalizować funkcję celu postaci:

$$\frac{1}{n}\sum_{i=1}^{n}[y_i - m(\boldsymbol{x}_i\boldsymbol{\beta})]^2 \tag{4}$$

względem wektora $\boldsymbol{\beta}$ (Ichimura, 1993). Jednak w sytuacji, kiedy funkcja $m(\cdot)$ nie jest znana, należy oszacować zarówno $m(\cdot)$, jak i $\boldsymbol{\beta}$. W pracy Ichimury (1993) zaproponowano wykorzystanie w tym celu estymatora typu *leave-one-out* średniej warunkowej zmiennej $\boldsymbol{y}$ względem $\boldsymbol{x\beta}$ przy jednoczesnej minimalizacji sumy kwadratów reszt względem q-wymiarowego wektora $\boldsymbol{\beta}$. W statystyce estymacja typu *leave-one-out* odnosi się do specyficznego schematu estymacji, w którym pojedyncza obserwacja jest pomijana przy estymacji parametru. Następnie tej pominiętej obserwacji używa się do obliczenia błędu estymacji.

Estymator typu *leave-one-out* średniej warunkowej $\boldsymbol{y}$ względem $\boldsymbol{x\beta}$ ma następującą postać:

$$E_{-i}(y_i|\boldsymbol{x}_i\boldsymbol{\beta}) = \frac{\sum_{j\neq i}^{n} y_j K_h(\boldsymbol{x}_j\boldsymbol{\beta}, \boldsymbol{x}_i\boldsymbol{\beta})}{\sum_{j\neq i}^{n} K_h(\boldsymbol{x}_j\boldsymbol{\beta}, \boldsymbol{x}_i\boldsymbol{\beta})}, \tag{5}$$

gdzie:

$K_h(\boldsymbol{x}_j\boldsymbol{\beta}, \boldsymbol{x}_i\boldsymbol{\beta}) = \prod_{d=1}^{q} k\left(\frac{x_{dj}\boldsymbol{\beta}_d - x_{di}\boldsymbol{\beta}_d}{h_d}\right)$ – iloczyn funkcji jądrowej $k(\cdot)$[2],

h_d – szerokość pasma (parametr wygładzania).

Problem optymalizacyjny dany wzorem (4) przyjmuje ostatecznie postać:

$$\frac{1}{n}\sum_{i=1}^{n}\left[y_i - \frac{\sum_{j\neq i}^{n} y_j K_h(\boldsymbol{x}_j\boldsymbol{\beta}, \boldsymbol{x}_i\boldsymbol{\beta})}{\sum_{j\neq i}^{n} K_h(\boldsymbol{x}_j\boldsymbol{\beta}, \boldsymbol{x}_i\boldsymbol{\beta})}\right]^2. \tag{6}$$

Estymator wektora $\boldsymbol{\beta}$ jest definiowany jako wartości wektora oszacowań $\widehat{\boldsymbol{\beta}}$, które minimalizują funkcję celu określoną wzorem (6).

W badaniach teoretycznych nad modelem jednowskaźnikowym funkcja celu określona wzorem (6) różni się na ogół w dwóch aspektach od funkcji danej wzorem (4). Po pierwsze, biorąc pod uwagę, że lokalnie stały estymator najmniejszych kwadratów jest zbieżny jednostajnie po zbiorach zwartych, niezbędna jest pewna funkcja ucinająca, aby wyeliminować małe wartości w mianowniku. Po drugie, funkcja celu szacowana jest często za pomocą funkcji wagowej. Ichimura (1993) sugeruje wykorzystanie jako funkcji wagowej odwrotności wariancji niejednorodnej. Funkcja wagowa nie jest niezbędna do osiągnięcia tempa zbieżności rzędu $\sqrt{n}$ dla współczynników kierunkowych, gdy składnik losowy jest heteroskedastyczny, ale jest wymagana do osiągnięcia granicy efektywności dla oszacowań semiparametrycznych.

[2] Dla ciągu $X_1, X_2, \ldots, X_n$ realizacji zmiennej losowej X funkcja $k\left(\frac{x-X_i}{h}\right)$ jest funkcją jądrową o szerokości pasma h.

2.2. Wybór optymalnej szerokości pasma semiparametrycznego modelu jednowskaźnikowego

Funkcja jądrowa $k(\cdot)$ występująca we wzorze (5) jest funkcją nieujemną i powinna spełniać następujące warunki:

- $\int_{-\infty}^{+\infty} k(\psi)d\psi = 1$;
- $\int_{-\infty}^{+\infty} \psi\, k(\psi)d\psi = 0$;
- $\int_{-\infty}^{+\infty} \psi^2\, k(\psi)d\psi = \kappa_2 < \infty$, gdzie κ_2 to moment centralny rzędu drugiego.

Dodatkowo funkcja $k(\psi)$ powinna być symetryczna i przyjmować duże (małe) wartości dla małych (dużych) ψ. Wśród najpopularniejszych funkcji jądra wyróżnia się funkcje (Henderson i Parmeter, 2015):

- gaussowską: $k(\psi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\psi^2}{2}\right)$;
- Epanecznikowa: $k(\psi) = \frac{3}{4\sqrt{5}}\left(1 - \frac{\psi^2}{5}\right)$ dla $|\psi| < \sqrt{5}$;
- trójkątną: $k(\psi) = 1 - |\psi|$ dla $|\psi| < 1$;
- jednostajną: $k(\psi) = \frac{1}{2}$ dla $|\psi| < 1$;
- dwuwagową: $k(\psi) = \frac{15}{16}(1 - \psi^2)^2$ dla $|\psi| < 1$;
- trójwagową: $k(\psi) = \frac{35}{32}(1 - \psi^2)^3$ dla $|\psi| < 1$.

Szerokość pasma h_d występująca we wzorze (5) określa stopień wygładzenia funkcji $m(\cdot)$. Optymalna szerokość pasma h to taka, która minimalizuje określoną miarę dokładności estymatora, np. błąd średniokwadratowy (ang. *mean squared error* – MSE) lub scałkowany błąd średniokwadratowy (ang. *integrated mean squared error* – IMSE). W semiparametrycznym modelu jednowskaźnikowym wybór szerokości pasma jest dość skomplikowany. Nie jest oczywiste, że optymalna szerokość pasma najlepsza do oszacowania średniej warunkowej jest jednocześnie najlepsza do estymacji parametrów modeli parametrycznych. Nie jest również ściśle określone, czy szerokość pasma należy szacować w jednym kroku, czy w większej liczbie kroków. W pracy Härdle'a i in. (1993) zaproponowano procedurę wyboru optymalnej wartości h. Wykazano przy tym, że równoczesna estymacja szerokości pasma h i wektora $\boldsymbol{\beta}$ jest bardzo zbliżona do minimalizacji sumy kwadratów reszt oraz zwykłej funkcji walidacji krzyżowej dla osobnej estymacji h i $\boldsymbol{\beta}$. W rzeczywistości prowadzi to do uzyskania $\sqrt{n}$-zgodnych oszacowań wektora $\boldsymbol{\beta}$ współczynników, jak również asymptotycznie optymalnych oszacowań szerokości pasma h.

Funkcja celu wykorzystana do wyznaczenia optymalnej szerokości pasma otrzymywana jest metodą walidacji krzyżowej i przyjmuje postać

$$CV(\widehat{\boldsymbol{\beta}},\, h) = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - m_{-i}(\boldsymbol{x}_i\widehat{\boldsymbol{\beta}})\right]^2, \tag{7}$$

gdzie $m_{-i}(\cdot)$ jest estymatorem typu *leave-one-out* funkcji wygładzającej $m(\cdot)$.

Funkcja celu $CV(\widehat{\boldsymbol{\beta}}, h)$ jest minimalizowana zarówno względem $\widehat{\boldsymbol{\beta}}$, jak i względem h. Udowodniono (Härdle i in., 1993), że funkcję daną wzorem (7) można przybliżyć jako

$$CV(\widehat{\boldsymbol{\beta}}, h) = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - m_{-i}(\boldsymbol{x}_i\widehat{\boldsymbol{\beta}})\right]^2 \approx \frac{1}{n}\sum_{i=1}^{n}\left[y_i - m(\boldsymbol{x}_i\widehat{\boldsymbol{\beta}})\right]^2 + \\ + \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{m}_{-i}(\boldsymbol{x}_i\widehat{\boldsymbol{\beta}}) - m(\boldsymbol{x}_i\widehat{\boldsymbol{\beta}})\right]^2. \tag{8}$$

Analizując wzór (8), można zauważyć, że pierwszy składnik jest problemem optymalizacyjnym najmniejszych kwadratów dla oszacowania wektora parametrów, gdy funkcja $m(\cdot)$ jest znana, a drugi składnik jest problemem optymalizacyjnym najmniejszych kwadratów dla oszacowania szerokości pasma nieznanej funkcji $m(\cdot)$, przy założeniu że wektor parametrów jest znany. Funkcja walidacji krzyżowej zachowuje się więc tak, jakby każdy składnik modelu estymowano osobno.

2.3. Ocena semiparametrycznego modelu jednowskaźnikowego

Zastosowanie semiparametrycznego modelu jednowskaźnikowego wymaga oceny stopnia, w jakim dopasowuje się on do danych. Wśród najważniejszych kryteriów takiej oceny można wymienić miary jakości, zgodności i odporności modelu.

Modele parametryczne mają ustalony zestaw parametrów opisujących ich strukturę, np. w modelu regresji liniowej parametrami są współczynniki nachylenia i przesunięcia. Modele semiparametryczne odznaczają się większą elastycznością i w większym stopniu mogą dostosowywać się do danych – niektóre części ich struktury są parametryczne, a inne nieparametryczne. Oznacza to, że pewne elementy modelu nie są ograniczone do określonych parametrów, co pozwala na lepszą adaptację do różnych form danych (Horowitz, 2009).

Jeśli chodzi o miary jakości, to różnica między rozpatrywanymi typami narzędzi może wynikać z elastyczności modeli semiparametrycznych. Modele parametryczne charakteryzują się bowiem zazwyczaj mniejszą liczbą parametrów i ściśle określoną strukturą, co może prowadzić do uproszczeń w zakresie interpretacji i oceny modelu. Jeśli jednak dane są bardziej skomplikowane lub nie spełniają założeń modelu parametrycznego, to modele semiparametryczne mogą się do nich lepiej dostosować (Yatchew, 2003).

Znanych jest kilka miar dobroci dopasowania, które mogą być wykorzystywane do oceny efektywności semiparametrycznego modelu jednowskaźnikowego. Wybrane z nich przedstawiono w zestawieniu (Henderson i Parmeter, 2015).

Zestawienie wybranych miar dobroci dopasowania

Miary dobroci dopasowania	Charakterystyka
Średni kwadrat błędu (MSE)	Miara średniej kwadratowej różnicy między wartościami przewidywanymi a rzeczywistymi; niskie wartości MSE świadczą o lepszym dopasowaniu
Kryterium informacyjne Akaikego (ang. *Akaike information criterion* – AIC)	Miara relacji między dopasowaniem modelu a jego złożonością; niższa wartość AIC wskazuje na lepszą równowagę między tymi cechami
Kryterium informacyjne Bayesa (ang. *Bayesian information criterion* – BIC)	Miara podobna do AIC, ale nakłada większą karę za złożoność modelu; niższa wartość BIC wskazuje na lepsze dopasowanie
R^2	Miara jakości dopasowania określająca proporcję zmienności zmiennej zależnej, która jest wyjaśniona przez zmienne niezależne; wyższy R^2 wskazuje na lepsze dopasowanie
Miara rozbieżności Kullbacka-Leiblera	Miara różnicy pomiędzy prawdziwym rozkładem danych a założonym rozkładem teoretycznym; niższe wartości miary wskazują na lepsze dopasowanie
Średni błąd bezwzględny (MAE)	Miara średniej różnicy między wartościami przewidywanymi a rzeczywistymi; niższe wartości MAE wskazują na lepsze dopasowanie modelu

Źródło: opracowanie własne na podstawie: Pagan i Ullah (1999).

Należy podkreślić, że zastosowanie jakiejkolwiek pojedynczej miary nie pozwala w pełni uchwycić dobroci dopasowania modelu, a zastosowanie różnych miar może prowadzić do odmiennych wniosków na temat wydajności modelu. Do oceny wydajności modelu zaleca się zatem stosowanie kilku miar.

Znane są również bardziej zaawansowane miary jakości, wykorzystywane do oceny dopasowania semiparametrycznych modeli ekonometrycznych do danych o bardziej skomplikowanej strukturze, czyli takich, które nie mogą być łatwo modelowane przy użyciu standardowych metod parametrycznych z powodu ich różnorodności, nieliniowości lub interakcji zmiennych. Należą do nich miary oparte na estymatorach jądrowych, metody bootstrapowe i różne testy nieparametryczne (Horowitz, 2009). Miary oparte na estymatorach jądrowych, takie jak jądrowa funkcja gęstości, wykorzystuje się do oceny dopasowania modelu do rozkładu danych, np. estymator jądrowy może być użyty do porównania empirycznego rozkładu danych z rozkładem przewidywanym przez model semiparametryczny. Metody bootstrapowe, czyli techniki próbkowania z powtórzeniami, pozwalają na estymację rozkładu próbkowego w przypadku analizowanych statystyk i są stosowane w semiparametrycznych modelach ekonometrycznych do oszacowania przedziałów ufności dla parametrów lub innych statystyk modelu. Testy nieparametryczne opierają się z kolei na rozkładzie empirycznym danych, który jest porównywany z rozkładem przewidywanym przez model. Za przykład mogą w tym wypadku posłużyć testy dopasowania rozkładu (np.

test Kołmogorowa-Smirnowa – K-S) lub testy porównujące grupy (np. test rangowy Wilcoxona-Manna-Whitneya). Wymienione miary jakości stosuje się w różnych kontekstach, w zależności od konkretnego semiparametrycznego modelu ekonometrycznego i charakterystyki danych. Wybór odpowiedniej miary jakości zależy od celu analizy i struktury danych.

Inną grupą miar, za pomocą których można dokonywać oceny modeli semiparametrycznych, są miary zgodności. Pomagają one określić, w jakim stopniu model odpowiada danym, czyli jak dalece pasuje do kształtowania się obserwacji. Miary zgodności stosowane w przypadku modeli semiparametrycznych mogą się różnić od tych wykorzystywanych w przypadku modeli parametrycznych. Wśród miar zgodności dla modeli semiparametrycznych można wyróżnić m.in. krzywą ROC (ang. *receiver operating characteristic*) i krzywą Kaplana-Meiera. Krzywa ROC znajduje zastosowanie m.in. w przypadku modeli klasyfikacyjnych, np. logistycznych. Im krzywa ROC jest położona wyżej i im większe jest pole powierzchni pod krzywą (AUC – *area under the curve*), tym większa zgodność modelu z danymi. Krzywa Kaplana-Meiera może być z kolei wykorzystana w analizie historii zdarzeń (analizie przeżycia). Przedstawia ona szanse przeżycia w zależności od czasu. Miara zgodności w postaci testu log-rank i różne metody porównawcze mogą być natomiast stosowane do porównywania różnych semiparametrycznych modeli przeżycia. Aby ocenić, który model semiparametryczny lepiej pasuje do danych, warto zastosować testy porównawcze, np. porównanie funkcji straty (ang. *loss function comparison*; Henderson i Parmeter, 2015).

Semiparametryczne modele ekonometryczne różnią się od modeli parametrycznych również pod względem miar odporności. Miary te służą do oceny stopnia stabilności modelu i jego niezależności od odstępstw od założeń lub skrajnych obserwacji w zbiorze danych. W przypadku modeli parametrycznych – o ściśle określonych założeniach dotyczących struktury modelu i parametrów – miary odporności mogą być wykorzystywane do identyfikacji obserwacji odstających, które mają duży wpływ na estymowane parametry lub dopasowanie modelu. Warto tutaj wymienić współczynnik wpływu Cooka, statystykę DFBETA (mierzącą zmianę w estymowanych parametrach modelu po usunięciu poszczególnych obserwacji) i błąd standardowy Hubera-White'a (odporna miara wariancji estymatora parametrów, która uwzględnia obserwacje odstające).

W przypadku semiparametrycznych modeli ekonometrycznych – o większej elastyczności w strukturze modelu i możliwości dopasowywania się do różnych form danych – miary odporności mogą być wykorzystywane w podobny sposób. W tym przypadku warto wykorzystać ocenę reszt lub odchyleń standardowych. Analiza reszt modelu może ujawnić obserwacje odstające lub niezgodności z założeniami modelu, a duże wartości odchylenia standardowego mogą wskazywać na obserwacje,

które mają znaczący wpływ na model. Inne miary bazują na estymacji kwantylowej, która pozwala na modelowanie rozkładu zmiennych zależnych w bardziej elastyczny sposób. Do analizy wpływu obserwacji ekstremalnych na model i estymowane parametry można stosować różne kwantyle, zależnie od problemu badawczego. W modelach semiparametrycznych przydatne są także techniki wykorzystujące modele odporne (takie jak odporna regresja lub odporna estymacja jądrowa), które minimalizują wpływ obserwacji odstających na estymację parametrów modelu. Ma to szczególne znaczenie w przypadku danych nietypowych lub obserwacji odstających, które mogą istotnie wpływać na wyniki estymacji modelu (Pagan i Ullah, 1999).

3. Pomiar ryzyka rynkowego z wykorzystaniem miar wrażliwości

Funkcjonowanie jednostki gospodarczej na rynku wiąże się z procesem podejmowania decyzji. Ze względu na posiadane informacje problemy decyzyjne można podzielić na trzy grupy (Jajuga, 2003):

- decyzje podejmowane w warunkach pewności, gdy każda decyzja pociąga za sobą określone, znane konsekwencje;
- decyzje podejmowane w warunkach ryzyka, gdy każda decyzja pociąga za sobą więcej niż jedną konsekwencję oraz znany jest zbiór możliwych konsekwencji i prawdopodobieństwo ich wystąpienia;
- decyzje podejmowane w warunkach niepewności, gdy prawdopodobieństwo wystąpienia konsekwencji decyzji nie jest znane.

W praktyce gospodarczej decyzje najczęściej podejmowane są w warunkach ryzyka i niepewności. Przez niepewność należy rozumieć obawę związaną z niewiedzą dotyczącą otaczającego świata i konsekwencjami tej niewiedzy. Może być ona wynikiem wielu czynników, np. niedostatecznej i nie w pełni transparentnej informacji, braku umiejętności interpretacji zjawisk gospodarczych – hipotetycznie związanych z rozważanym zakresem działalności rynkowej – lub indywidualnego podejścia każdego uczestnika rynku. Najczęstsza prawidłowość jest taka, że im bardziej złożone jest zjawisko rynkowe, tym większa niepewność i nieprzewidywalność dotycząca jego realizacji w przyszłości. Pojawia się bowiem ryzyko, że oczekiwania okażą się odmienne od nieznanej rzeczywistości (Krężołek, 2020).

Termin *ryzyko* najczęściej kojarzy się z sytuacją, w której realizacja podjętych działań przebiega na niższym poziomie, niż oczekiwano. Znane są dwa sposoby rozumienia tego terminu: neutralne i negatywne. W podejściu neutralnym przyjmuje się, że rzeczywista wielkość badanego zjawiska różni się od jej zakładanego poziomu (nie jest istotny znak tej różnicy), a w podejściu negatywnym rzeczywista wielkość badanego zjawiska jest mniejsza od jej oczekiwanego poziomu. Warto wskazać paradoks wynikający z neutralnego podejścia do rozumienia ryzyka – brak istotności

znaku różnicy poziomu badanego zjawiska zakłada, że w przypadku różnicy dodatniej (co z inwestycyjnego punktu widzenia oznacza zysk) uczestnik rynku nie postrzega takiej sytuacji jako ryzykownej. Stąd wniosek, że ryzyko jest zazwyczaj postrzegane negatywnie. W literaturze podano wiele definicji i kryteriów klasyfikacji ryzyka z uwzględnieniem jego możliwych konsekwencji (Alexander, 2008; Daníelsson, 2011; Dowd, 2005; Trzpiot, 2010).

Z punktu widzenia funkcjonowania gospodarki szczególnie ważną kategorią ryzyka jest ryzyko rynkowe (ang. *market risk*). Pojawia się ono wtedy, gdy na realizację transakcji kupna-sprzedaży pomiędzy stronami popytową i podażową wpływają różne czynniki rynkowe, najczęściej niezależne od realizowanego przedsięwzięcia. Najogólniej mówiąc, ryzyko rynkowe związane jest z prawdopodobieństwem wystąpienia straty, która stanowi efekt nieprzewidywalnego kierunku zmiany ceny przedmiotu transakcji. Stąd wniosek, że ryzyko rynkowe związane jest z każdym rynkiem, na którym prowadzi się wymianę handlową (materialną lub niematerialną), w tym z rynkiem kapitałowym (akcje, obligacje), instrumentów pochodnych (opcji, kontraktów terminowych na różnego rodzaju produkty bazowe), towarów i surowców, a także ze zmianami kursów walut (Holliwell, 2001; Jajuga, 2003).

Aby efektywne kontrolowanie ryzyka było możliwe, należy postrzegać to zagadnienie jako proces. Wyróżnia się następujące elementy procesu zarządzania ryzykiem (Christoffersen, 2012):

- identyfikacja ryzyka;
- pomiar i analiza ryzyka;
- opracowanie strategii ograniczającej lub minimalizującej ryzyko;
- implementacja strategii;
- kontrola, monitorowanie i korygowanie strategii ograniczającej lub minimalizującej ryzyko.

Każdy z powyższych elementów należy traktować równorzędnie.

W literaturze przedmiotu opisano wiele metod pomiaru ryzyka. Jedną z nich jest analiza wrażliwości, polegająca na dokonaniu oceny stopnia reakcji zmiennej ryzyka na zmienność określonych czynników, które to ryzyko determinują, za pomocą miary wrażliwości. Miary wrażliwości umożliwiają przeprowadzenie analizy przyczyn ryzyka, ponieważ obrazują wpływ czynników na zmienną ryzyka dzięki pewnej funkcji zależności od zmiennej ryzyka (Van Groenendaal, 1995). Matematyczna postać funkcji zależności zmiennej ryzyka od czynników ryzyka przyjmuje zatem postać

$$R = g(F_1, F_2, \dots, F_i, \varepsilon), \tag{9}$$

gdzie:

R – zmienna ryzyka (zmienna losowa),

F_i – i-ty czynnik ryzyka ($i = 1, 2, \dots, q$),

$g(\cdot)$ – funkcja określająca zależność pomiędzy zmienną ryzyka a czynnikami ryzyka,
ε – składnik losowy.

Funkcja $g(\cdot)$, określająca zależność między zmienną ryzyka a czynnikami ryzyka, może być liniowa lub nieliniowa. W przypadku liniowej postaci funkcji $g(\cdot)$ zmienna ryzyka przyjmuje postać

$$R = \beta_1 F_1 + \beta_2 F_2 + \cdots + \beta_i F_i + \varepsilon. \tag{10}$$

Miara wrażliwości powiązana z funkcją $g(\cdot)$ definiowana jest jako pierwsza pochodna tej funkcji względem i-tego czynnika ryzyka:

$$\beta_i = \frac{\partial R}{\partial F_i}. \tag{11}$$

W szczególności:

- jeśli $g(\cdot)$ jest funkcją liniową, to miara wrażliwości dla i-tego czynnika jest parametrem występującym przy i-tym czynniku ryzyka w modelu liniowym;
- jeśli $g(\cdot)$ jest funkcją nieliniową, to wykorzystując rozwinięcie ΔR w szereg Taylora, miarę wrażliwości dla i-tego czynnika wyznacza się jako $\Delta R \approx \frac{\partial R}{\partial F_i}\Delta F_i$.

Należy zwrócić uwagę na podobieństwo funkcji definiujących semiparametryczny model jednowskaźnikowy (1) i (2) z funkcją $g(\cdot)$, określającą zależność między zmienną ryzyka a czynnikami ryzyka (9), co w nomenklaturze modelu semiparametrycznego można zapisać jako

$$y_i = m\big(\varphi(\boldsymbol{x}_i, \boldsymbol{\beta})\big) + u_i \Rightarrow R = m\big(\varphi(\boldsymbol{F}, \boldsymbol{\beta})\big) + \varepsilon_i. \tag{12}$$

Stąd wniosek, że zmienna ryzyka może zostać zapisana przy wykorzystaniu nieznanej funkcji wygładzającej $m(\cdot)$, znanej funkcji parametrycznej $\varphi(\cdot,\cdot)$, q-wymiarowego wektora czynników $\boldsymbol{F}$ i p-wymiarowego wektora $\boldsymbol{\beta}$. Tę własność wykorzystano w przeprowadzonym badaniu.

4. Analiza ryzyka wybranych spółek notowanych na GPW

4.1. Metoda badania

W badaniu empirycznym przeprowadzono analizę ryzyka z wykorzystaniem semiparametrycznego modelu jednowskaźnikowego, uwzględniając koncepcję miar wrażliwości. Do analizy wybrano 10 spółek sektora informatycznego notowanych na GPW

od 4 stycznia 2018 r. do 30 grudnia 2021 r.: Ailleron, Atende, Betacom, Comarch, Ifirmę, LiveChat, LSI, NTT, Talex i Wasko. Dane zaczerpnięto z serwisu finansowego Bloomberg. Wyboru spółek dokonano na podstawie kryterium zbieżności liczby dni notowań danej spółki z liczbą notowań WIG20 w badanym okresie. Jako czynniki ryzyka wskazano dwa indeksy giełdowe: WIG20 i S&P 500. Na podstawie dziennych cen zamknięcia wyznaczono dzienne logarytmiczne stopy zwrotu zgodnie ze wzorem $R_t = 100 \ln \frac{p_t}{p_{t-1}}$ (gdzie p_t oznacza cenę zamknięcia w dniu t, a p_{t-1} – cenę zamknięcia w dniu $t-1$). Estymację semiparametrycznego modelu jednowskaźnikowego przeprowadzono, wykorzystując procedurę Ichimury (1993), w której zastosowano estymator typu *leave-one-out* średniej warunkowej R względem $\boldsymbol{F\beta}$ przy jednoczesnej minimalizacji sumy kwadratów reszt względem q-wymiarowego wektora $\boldsymbol{\beta}$.

4.2. Wyniki analizy empirycznej

Badanie rozpoczęto od przeprowadzenia analizy podstawowych charakterystyk stóp zwrotu badanych aktywów. Jej wyniki zawiera tabl. 1.

Tabl. 1. Statystyki opisowe dla szeregów stóp zwrotu analizowanych spółek notowanych na GWP w latach 2018–2021

Wyszczególnienie	*M*	S_d	W_{zm}	*S*	*K*	Min	Max	Test normalności	
								K-S	wartość *p**
WIG20	–0,008	1,410	–16924,324	–1,150	12,640	–14,246	6,335	0,069	<0,001
S&P 500	0,057	1,326	2334,836	–1,076	19,000	–12,765	8,968	0,127	<0,001
Ailleron	–0,038	3,322	–8685,027	0,308	5,604	–21,492	19,236	0,089	<0,001
Atende	0,036	2,392	6647,479	0,237	5,237	–14,540	13,353	0,151	<0,001
Betacom	–0,051	2,752	–5384,450	–0,426	1,707	–14,023	7,688	0,152	<0,001
Comarch	–0,003	2,073	–73124,217	–0,077	4,162	–12,851	9,345	0,074	<0,001
Ifirma	0,202	2,863	1418,183	0,101	2,920	–13,352	14,774	0,071	<0,001
LiveChat	0,120	2,581	2142,977	0,426	3,777	–11,118	14,228	0,089	<0,001
LSI	–0,009	2,573	–30197,148	0,203	9,833	–22,182	18,740	0,132	<0,001
NTT	0,096	2,906	3024,925	0,514	5,142	–19,416	15,207	0,130	<0,001
Talex	–0,009	2,654	–28883,459	–0,318	6,369	–18,469	14,898	0,157	<0,001
Wasko	–0,030	2,810	–9284,848	–0,381	7,090	–21,911	11,772	0,104	<0,001

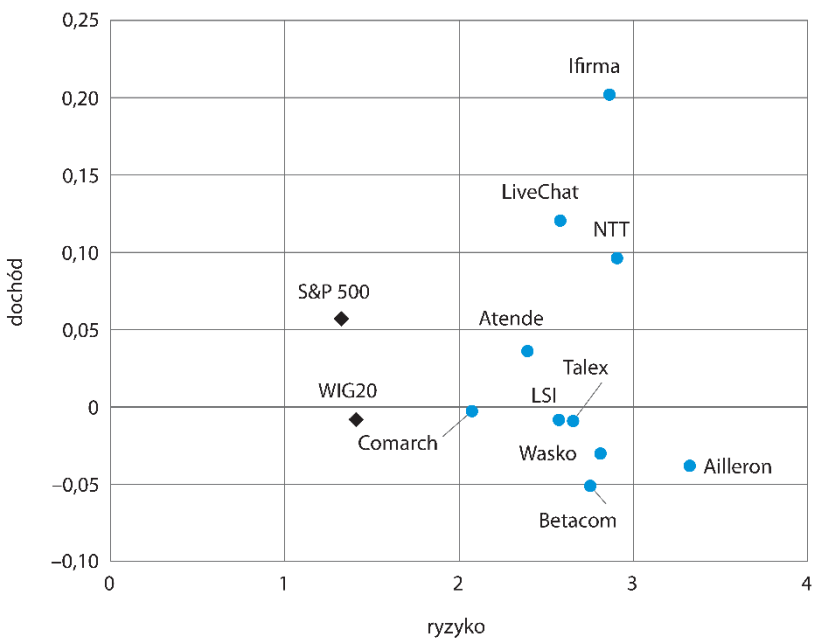
Uwaga. *M* – średnia, S_d – odchylenie standardowe, W_{zm} – współczynnik zmienności, *S* – skośność, *K* – kurtoza istotność, Min – minimum, Max – maksimum. * – istotność statystyczna na poziomie 0,01.

Źródło: obliczenia własne na podstawie danych z serwisu Bloomberg.

Zaobserwowano, że większość analizowanych walorów charakteryzowała się średnio ujemną stopą zwrotu. Ponadto inwestycje w WIG20 generowały w badanym okresie średniookresowe straty, a indeks S&P 500 uzyskał średnio dodatnią stopę zwrotu. Odnotowano także wysoką zmienność stóp zwrotu. Większość rozkładów wykazywała asymetrię prawostronną. Zauważono również silną leptokurtyczność i rozbieżność z rozkładem normalnym (odrzucono hipotezę zerową w teście zgodności K-S).

Z punktu widzenia potencjalnych inwestycji decydent jest zazwyczaj zainteresowany oczekiwanym poziomem zysku lub straty i związanym z tym ryzykiem. Na wykr. 1 zestawiono te dwie charakterystyki dla analizowanych walorów.

Wykr. 1. Mapa ryzyko – dochód dla analizowanych aktywów GPW w latach 2018–2021



Źródło: opracowanie własne na podstawie danych z serwisu Bloomberg.

Z informacji zamieszczonych na wykr. 1 wynika, że WIG20 i S&P 500 wykazywały niższy poziom ryzyka w porównaniu z pozostałymi walorami. Najbardziej ryzykowne spośród badanych aktywów były Ailleron, NTT i Ifirma, a najmniej – Comarch. Najbardziej dochodowe okazały się natomiast inwestycje w walory Ifirmy, a najmniej – Betacomu. Wszystkie spółki (poza Talexem) wykazywały także statystycznie istotną, dodatnią korelację z oboma indeksami (tabl. 2).

Tabl. 2. Współczynniki korelacji Pearsona pomiędzy analizowanymi aktywami GPW w latach 2018–2021

Wyszczególnienie	WIG20	S&P 500	Ailleron	Atende	Betacom	Comarch	Ifirma	LiveChat	LSI	NTT	Talex	Wasko
WIG20	1,000	0,497*	0,290*	0,177*	0,137*	0,280*	0,206*	0,251*	0,324*	0,102*	0,044	0,155*
S&P 500	0,497*	1,000	0,169*	0,109*	0,072*	0,207*	0,176*	0,139*	0,219*	0,071*	0,046	0,116*
Ailleron	0,290*	0,169*	1,000	0,111*	0,048	0,103*	0,087*	0,158*	0,172*	0,121*	–0,002	0,066*
Atende	0,177*	0,109*	0,111*	1,000	0,013	0,024	0,040	0,058	0,103*	0,121*	0,053	0,075*
Betacom	0,137*	0,072*	0,048	0,013	1,000	0,070*	0,022	0,033	0,110*	0,082*	0,019	0,053
Comarch	0,280*	0,207*	0,103*	0,024	0,070*	1,000	0,153*	0,123*	0,176*	–0,055	0,036	0,006
Ifirma	0,206*	0,176*	0,087*	0,040	0,022	0,153*	1,000	0,085*	0,154*	0,036	0,008	0,015
LiveChat	0,251*	0,139*	0,158*	0,058	0,033	0,123*	0,085*	1,000	0,097*	0,022	0,053	0,052
LSI	0,324*	0,219*	0,172*	0,103*	0,110*	0,176*	0,154*	0,097*	1,000	0,027	0,046	0,072*
NTT	0,102*	0,071*	0,121*	0,121*	0,082*	–0,055	0,036	0,022	0,027	1,000	0,010	0,005
Talex	0,044	0,046	–0,002	0,053	0,019	0,036	0,008	0,053	0,046	0,010	1,000	0,017
Wasko	0,155*	0,116*	0,066*	0,075*	0,053	0,006	0,015	0,052	0,072*	0,005	0,017	1,000

Uwaga. * – istotność statystyczna na poziomie 0,05.

Źródło: obliczenia własne na podstawie danych z serwisu Bloomberg.

Na kolejnym etapie badania oszacowano semiparametryczny model jednowskaźnikowy określany dla każdej analizowanej spółki sektora informatycznego GPW, gdzie czynnikiem ryzyka były stopy zwrotu indeksów WIG20 i S&P 500. Zastosowano procedurę estymacji Ichimury, wykorzystując gaussowską funkcję jądra. Wyniki oszacowań przedstawiono w tabl. 3.

Tabl. 3. Oszacowania parametrów semiparametrycznych modeli jednowskaźnikowych dla analizowanych aktywów GWP w latach 2018–2021

Spółki	Współczynniki	Wartości współczynników	Wartość p^* dla testu t	Szerokość pasma	Funkcja celu	R^2	MAE
Ailleron	$\hat{\beta}_{WIG20}$	0,64347	<0,001	0,23865	3,43276	0,176	2,181
	$\hat{\beta}_{S\&P\ 500}$	0,08356	0,003				
Atende	$\hat{\beta}_{WIG20}$	0,27754	<0,001	0,28711	4,66361	0,204	2,356
	$\hat{\beta}_{S\&P\ 500}$	0,05016	0,001				
Betacom	$\hat{\beta}_{WIG20}$	0,26387	<0,001	0,21953	3,21873	0,193	2,729
	$\hat{\beta}_{S\&P\ 500}$	0,00900	0,001				
Comarch	$\hat{\beta}_{WIG20}$	0,34570	<0,001	0,20779	3,98047	0,147	1,986
	$\hat{\beta}_{S\&P\ 500}$	0,14046	0,005				
Ifirma	$\hat{\beta}_{WIG20}$	0,31893	<0,001	0,31653	5,29474	0,164	2,794
	$\hat{\beta}_{S\&P\ 500}$	0,21110	0,002				
LiveChat	$\hat{\beta}_{WIG20}$	0,44187	<0,001	0,27931	4,31885	0,221	1,874
	$\hat{\beta}_{S\&P\ 500}$	0,03700	0,001				
LSI	$\hat{\beta}_{WIG20}$	0,51985	<0,001	0,36271	5,96080	0,178	2,501
	$\hat{\beta}_{S\&P\ 500}$	0,15067	0,002				
NTT	$\hat{\beta}_{WIG20}$	0,18278	<0,001	0,53862	6,32957	0,148	2,893
	$\hat{\beta}_{S\&P\ 500}$	0,05961	0,001				
Talex	$\hat{\beta}_{WIG20}$	0,05308	<0,001	0,61496	6,99463	0,093	2,653
	$\hat{\beta}_{S\&P\ 500}$	0,06349	0,001				
Wasko	$\hat{\beta}_{WIG20}$	0,25670	<0,001	0,56328	6,23192	0,113	2,776
	$\hat{\beta}_{S\&P\ 500}$	0,10992	0,001				

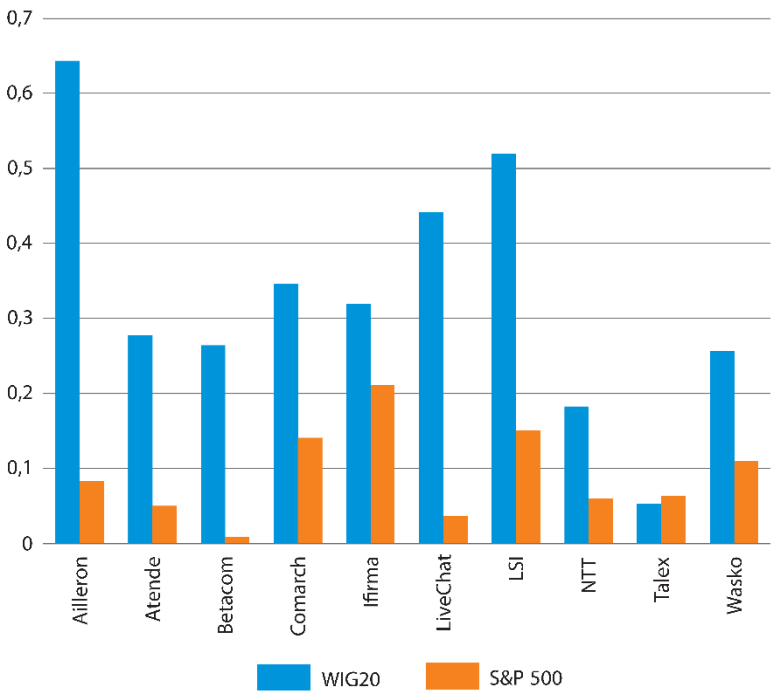
Uwaga. * – istotność statystyczna na poziomie 0,01.

Źródło: obliczenia własne na podstawie danych z serwisu Bloomberg.

Wyniki zamieszczone w tabl. 3 pokazują, że wybrane indeksy są statystycznie istotnymi czynnikami zmienności stóp zwrotu analizowanych spółek. Wartości oszacowań $\widehat{\boldsymbol{\beta}}$ wektora $\boldsymbol{\beta}$ dla WIG20 są wyższe niż dla S&P 500. To prawdopodobnie wynik powiązania analizowanych spółek z polskim rynkiem kapitałowym. W badanym okresie na zmienność WIG20 najwrażliwsze były stopy zwrotu Ailleronu, LSI i LiveChatu, a na zmienność S&P 500 – Ifirmy, LSI i Comarchu. Wszystkie oszacowane modele są istotne statystycznie, a wartości współczynnika R^2 wskazują, że jakość dopasowania jest niska (najmniejszą wartość współczynnika uzyskał Talex). Analizując wartości błędu MAE, zaobserwowano z kolei, że najmniejsze błędy wystąpiły w przypadku LiveChatu i Comarchu. Zauważono ponadto, że wartości funk-

cji celu wzrastają wraz ze zwyżką wartości parametru wygładzania (szerokości pasma). Oszacowania wartości wektora $\boldsymbol{\beta}$ zaprezentowano na wykr. 2.

Wykr. 2. Oszacowania wartości współczynnika wektora $\boldsymbol{\beta}$ semiparametrycznych modeli jednowskaźnikowych dla analizowanych aktywów GPW w latach 2018–2021



Źródło: opracowanie własne na podstawie danych z serwisu Bloomberg.

Jak widać na wykresie, największe różnice w oszacowaniach współczynników $\widehat{\boldsymbol{\beta}}$ w semiparametrycznym modelu jednowskaźnikowym dotyczyły Aillerona, LiveChatu i LSI, a najmniejsze – Talexu.

5. Podsumowanie

W artykule przedstawiono wyniki pomiaru ryzyka rynkowego z wykorzystaniem jednowskaźnikowego semiparametrycznego modelu ekonometrycznego. Badaniem objęto wybrane spółki sektora informatycznego notowane na GPW w latach 2018–2021. Analizowano dzienne logarytmiczne stopy zwrotu badanych aktywów, a jako czynniki ryzyka wskazano dwa indeksy giełdowe: polski WIG20 i amerykański S&P 500. Estymację semiparametrycznego modelu jednowskaźnikowego przepro-

wadzono z zastosowaniem procedury Ichimury bazującej na estymatorze typu *leave--one-out*. Ponadto jako funkcję jądra wykorzystano funkcję gaussowską.

Model semiparametryczny opiera się na estymacji elastycznej i powstaje dzięki połączeniu modeli parametrycznego i nieparametrycznego. Modele semiparametryczne są przydatne w sytuacjach, w których modele w pełni parametryczne i w pełni nieparametryczne nie dają właściwych oszacowań. Skutecznym sposobem redukcji zjawiska nadmiarowości, z którym można się spotkać podczas przeprowadzania analizy dużych zbiorów danych, jest wykorzystanie modelu jednowskaźnikowego. W takim modelu występuje jednowymiarowy wskaźnik, który stanowi element nieznanej funkcji nieparametrycznej. Modele jednowskaźnikowe działają podobnie jak wiele znanych modeli parametrycznych, co potwierdza ich użyteczność. Jednowskaźnikowy semiparametryczny model ekonometryczny ma tę przewagę nad modelem liniowym, że pozwala na uwzględnienie nieliniowych zależności między zmiennymi objaśniającymi a zmienną zależną, nie wymaga założeń dotyczących rozkładu i homoskedastyczności składnika losowego oraz umożliwia uwzględnienie efektów nieliniowych zmiennych w wariancji błędów. Ponadto szacunki składnika parametrycznego są asymptotycznie zbieżne z rzeczywistymi w tempie zbliżonym do modelu w pełni parametrycznego.

Na podstawie przeprowadzonego badania wykazano, że zarówno WIG20, jak i S&P 500 były istotnymi czynnikami zmienności stóp zwrotu badanych spółek sektora informatycznego w analizowanym okresie. Oszacowania współczynników $\widehat{\boldsymbol{\beta}}$ semiparametrycznych modeli jednowskaźnikowych dla indeksu polskiej giełdy były znacznie wyższe niż oszacowania dla S&P 500 (wyjątek stanowiła spółka Talex). Oba indeksy przyjmowały wartości dodatnie, poniżej progu 1. Świadczy to o tym, że stopy zwrotu badanych spółek zmieniały się w tym samym kierunku co stopy zwrotu indeksów, jednak ich reakcje na zmiany stopy zwrotu rynków polskiego i amerykańskiego były niewielkie. Nawet gdy spadek stóp zwrotu rynku był znaczny, stopy zwrotu badanych akcji zmniejszyły się tylko nieznacznie. Prowadzi to do wniosku, że akcje badanych spółek cechowało ryzyko niższe od rynkowego (są to akcje defensywne).

Z analizy oszacowanych semiparametrycznych modeli jednowskaźnikowych wynika, że w przypadku badanych spółek parametr wygładzania zawierał się w przedziale 0,20–0,62. Dodatkowo wraz ze wzrostem wartości tego parametru zwiększała się wartość funkcji celu wykorzystywanej do oszacowywania współczynników $\widehat{\boldsymbol{\beta}}$ modeli. Warto jednak zwrócić uwagę, że jakość dopasowania określona za pomocą wartości współczynnika R^2 jest niezadowalająca. Największą wartość uzyskano dla modelu spółek LiveChat i Atende, a najmniejszą – dla Talexu. Zaobserwowano także małe wartości błędu MAE (najmniejsze dla LiveChatu i Comarchu). Te wyniki są zbieżne z wynikami oceny korelacji pomiędzy badanymi spółkami i indeksami.

Semiparametryczne modele jednowskaźnikowe mają wiele zalet przydatnych w analizie ryzyka na giełdzie. Przede wszystkim są bardziej elastyczne – w przeciwieństwie do modeli parametrycznych (np. CAPM) nie narzucają szczegółowych założeń dotyczących formy funkcyjnej relacji między zmiennymi. Dzięki temu mogą lepiej dopasowywać się do rzeczywistych danych i trafniej uchwycić złożone zależności zachodzące między zmiennymi. Ponadto pozwalają na modelowanie wpływu wielu indeksów giełdowych na stopę zwrotu akcji, co może prowadzić do precyzyjniejszych oszacowań ryzyka. Za ich inną zaletę należy uznać uwzględnianie zmienności czasowej, ponieważ współczynnik $\widehat{\boldsymbol{\beta}}$ jest funkcją zmiennych niezależnych i może zmieniać się w czasie. To pozwala na uwzględnienie zmienności ryzyka systematycznego w czasie, co jest istotne w dynamicznie zmieniającym się środowisku giełdowym.

Modele jednoindeksowe mają również wady. Estymacja semiparametryczna może okazać się bardziej złożona i czasochłonna niż estymacja parametryczna, a uzyskane wyniki – wrażliwsze na wybór metody estymacji i parametrów modelu. Ważne zatem, aby przed zastosowaniem tych modeli dobrze zrozumieć ich założenia i związane z nimi ograniczenia.

Z punktu widzenia inwestora korzystanie z semiparametrycznego modelu jednowskaźnikowego w analizie ryzyka giełdowego może przynieść wiele korzyści. Modele tej klasy pozwalają na dokładniejszą ocenę ryzyka, ponieważ w ich przypadku nie zakłada się konkretnej formy funkcjonalnej zależności między zmiennymi, dzięki czemu mogą one lepiej pasować do rzeczywistych danych i zapewniać dokładniejsze oceny ryzyka, a tym samym prowadzić do właściwszych decyzji inwestycyjnych. Semiparametryczne modele jednowskaźnikowe pozwalają także trafniej uchwycić złożone zależności między stopami zwrotu a indeksami giełdowymi, co powinno pomóc inwestorom zrozumieć, jakie czynniki wpływają na ryzyko ich inwestycji. Łatwo je dostosować do zmieniających się warunków giełdowych, co w konsekwencji pozwala inwestorom na dokonywanie bardziej świadomych decyzji inwestycyjnych. Dokładniejsza ocena ryzyka przekłada się ponadto na możliwość lepszego alokowania kapitału. Inwestorzy mają możliwość przeznaczenia większego kapitału na aktywa o niższym ryzyku i wyższym potencjalnym zwrocie, czym zwiększają swoje szanse na osiągnięcie większych korzyści przy określonym poziomie ryzyka.

Należy przy tym pamiętać, że korzystanie z semiparametrycznych modeli ekonometrycznych nie gwarantuje sukcesu inwestycyjnego. Ważne, aby inwestorzy mieli świadomość ograniczeń tych modeli i traktowali je jako jedno z wielu narzędzi analizy inwestycyjnej.

Bibliografia

Alexander, C. (2008). *Market Risk Analysis* (vol. 1–4). John Wiley & Sons.

Cameron, A. C., Trivedi, P. K. (2005). *Microeconometrics. Methods and Applications*. Cambridge University Press.

Christoffersen, P. F. (2012). *Elements of Financial Risk Management* (2nd edition). Elsevier.

Daníelsson, J. (2011). *Financial Risk Forecasting. The Theory and Practice of Forecasting Market Risk with Implementation in R and MATLAB*. John Wiley & Sons.

Dowd, K. (2005). *Measuring Market Risk* (2nd edition). John Wiley & Sons.

Duan, N., Li, K. C. (1991). Slicing regression: A link-free regression method. *The Annals of Statistics*, *19*(2), 505–530. https://doi.org/10.1214/aos/1176348109.

Fix, E., Hodges, J. L. (1951). *Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties*. https://apps.dtic.mil/sti/pdfs/ADA800276.pdf.

Härdle, W., Hall, P., Ichimura, H. (1993). Optimal smoothing in single-index models. *The Annals of Statistics*, *21*(1), 157–178. https://doi.org/10.1214/aos/1176349020.

Härdle, W., Stoker, T. M. (1989). Investigating Smooth Multiple Regression by the Method of Average Derivatives. *Journal of the American Statistical Association*, *84*(408), 986–995. https://doi.org/10.2307/2290074.

Henderson, D. J., Parmeter, C. F. (2015). *Applied Nonparametric Econometrics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511845765.

Holliwell, J. (2001). *Ryzyko finansowe. Metody identyfikacji i zarządzania ryzykiem finansowym*. Wydawnictwo Liber.

Horowitz, J. L. (2009). *Semiparametric and Nonparametric Methods in Econometrics*. Springer.

Ichimura, H. (1993). Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models. *Journal of Econometrics*, *58*(1–2), 71–120. https://doi.org/10.1016/0304-4076(93)90114-K.

Jajuga, K. (2003). *Metody statystyczne w finansach*. StatSoft Polska. https://media.statsoft.pl/_old_dnn/downloads/jajuga.pdf.

Krężołek, D. (2020). *Modelowanie ryzyka na rynku metali*. Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach.

Li, K. C. (1991). Sliced Inverse Regression for Dimension Reduction. *Journal of American Statistical Association*, *86*(414), 316–327. https://doi.org/10.2307/2290563.

Li, G., Peng, H., Dong, K., Tong, T. (2014). Simultaneous confidence bands and hypothesis testing for single-index models. *Statistica Sinica*, *24*(2), 937–955. http://dx.doi.org/10.5705/ss.2012.127.

Li, Q., Racine, J. S. (2007). *Nonparametric Econometrics. Theory and Practice*. Princeton University Press.

Pagan, A., Ullah, A. (1999). *Nonparametric Econometrics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511612503.

Powell, J. L., Stock, J. H., Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica, 57*, 1403–1430.

Racine J. S. (2008). Nonparametric Econometrics. *A Primer. Foundations and Trends in Econometrics, 3*(1), 1–88.

Ramanathan R. (1989). *Introductory Econometrics with Applications*. Harcourt Brace Jovanovich.

Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, *27*(3), 832–837. https://doi.org/10.1214/aoms/1177728190.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman & Hall.

Stoker, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica, 54*(6), 1461–1481.

Trzpiot, G. (red.). (2010). *Wielowymiarowe metody statystyczne w analizie ryzyka inwestycyjnego*. Polskie Wydawnictwo Ekonomiczne.

Van Groenendaal, W. J. H. (1995). Measuring Risk: Risk Analysis or Sensitivity Analysis?. *IFAC Proceedings Volumes*, *28*(7), 443–450. https://doi.org/10.1016/S1474-6670(17)47145-X.

Wasserman, L. (2006). *All of Nonparametric Statistics*. Springer. https://doi.org/10.1007/0-387-30623-4.

Xia, Y. (2006). Asymptotic distributions for two estimators of the single-index model. *Econometric Theory*, *22*(6), 1112–1137. https://doi.org/10.1017/S0266466606060531.

Xia, Y., Tong, H., Li, W. K. (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *64*(3), 363–410. https://doi.org/10.1111/1467-9868.03412.

Xue, L. (2013). Estimation and empirical likelihood for single-index models with missing data in the covariates. *Computational Statistics and Data Analysis*, *68*, 82–97. https://doi.org/10.1016/j.csda.2013.06.017.

Xue, L., Zhu, L. (2006). Empirical likelihood for single-index model. *Journal of Multivariate Analysis*, *97*(6), 1295–1312. https://doi.org/10.1016/j.jmva.2005.09.004.

Yang, Y., Song, Q. (2014). Jump detection in time series nonparametric regression models: a polynomial spline approach. *Annals of the Institute of Statistical Mathematics*, *66*(2), 325–344. https://doi.org/10.1007/s10463-013-0411-3.

Yatchew, A. (2003). *Semiparametric Regression for the Applied Econometrician*. Cambridge University Press. https://doi.org/10.1017/CBO9780511615887.

Zhu, L., Xue, L. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *68*(3), 549–570. https://doi.org/10.1111/j.1467-9868.2006.00556.x.