

Witold Orzeszko
Uniwersytet Mikołaja Kopernika w Toruniu

Zastosowanie testu Kaplana do identyfikacji ekonomicznych szeregów czasowych

Streszczenie Identyfikacja zależności w szeregach czasowych jest jednym z głównych problemów badawczych ekonometrii. Ze względu na wyższy stopień rozwoju metodologii, kwestię tę często sprowadza się w praktyce do identyfikacji zależności o charakterze liniowym. Wydaje się jednak, że rozszerzenie analizy o poszukiwanie zależności również nieliniowych, może być bardzo obiecującym nurtem w ekonometrii. Zagadnienie to rozwija się w ostatnich latach bardzo dynamicznie, czego efektem jest powstanie wielu nowych metod nieliniowej analizy danych.

W pracy zaprezentowano test Kaplana – nieparametryczną metodę identyfikacji nieliniowości w szeregach czasowych. Jego użyteczność zweryfikowano w oparciu o dane symulowane. Ponadto, zastosowano test Kaplana do analizy indeksu WIG.

1. Test Kaplana

Test Kaplana (1994) jest nieparametryczną metodą identyfikacji szeregów czasowych. Z samej konstrukcji, służy on do zbadania istnienia w szeregu zależności deterministycznych, przy założeniu, że opisane są one funkcją ciągłą. Jednak w praktyce, możliwości stosowania testu są dużo szersze – można go wykorzystać do identyfikacji struktury różnego rodzaju, również stochastycznej.

Niech (x_t) będzie dowolnym szeregiem czasowym. Celem testu Kaplana jest zweryfikowanie hipotezy o istnieniu ciągłego odwzorowania f generującego dynamikę szeregu, tzn. spełniającego zależność:

$$x_{t+1} = f(x_t),$$

dla każdego t . Jeśli takie odwzorowanie istnieje, to, na mocy definicji ciągłości, dowolne bliskie sobie argumenty – x_i oraz x_j , mają również bliskie sobie obrazy, tzn. x_{i+1} oraz x_{j+1} .

W celu formalnego zaprezentowania procedury testowania, oznaczmy przez $\delta_{i,j}$ odległość między obserwacjami x_i oraz x_j , natomiast przez $\varepsilon_{i,j}$ – odległość między x_{i+1} oraz x_{j+1} , tzn.:

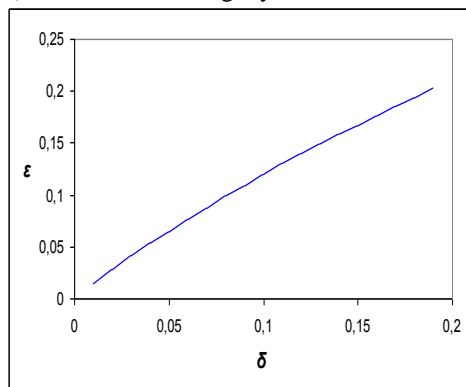
$$\delta_{i,j} = |x_i - x_j|, \quad \varepsilon_{i,j} = |x_{i+1} - x_{j+1}|.$$

Ponadto, dla dowolnej liczby $\delta > 0$, niech $\varepsilon(\delta)$ oznacza średnią z odległości $\varepsilon_{i,j}$, dla których indeksy i, j spełniają warunek $\delta_{i,j} < \delta$. Przytoczona wcześniej własność, dotycząca podobieństwa obrazów bliskich sobie argumentów, oznacza, iż dla funkcji ciągłych $\varepsilon(\delta) \rightarrow 0$ przy $\delta \rightarrow 0$. W przeciwnym wypadku, tzn., gdy szereg jest realizacją procesu stochastycznego, $\varepsilon(\delta) \rightarrow E$ (przy $\delta \rightarrow 0$), gdzie E jest pewną liczbą dodatnią.

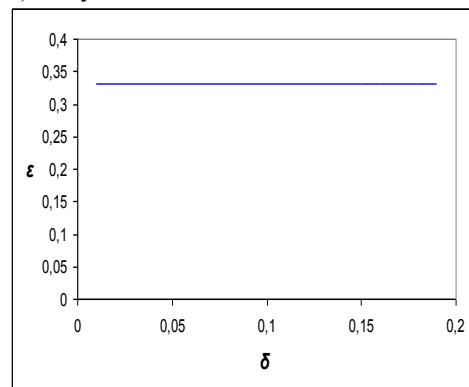
Dla zilustrowania powyższej własności, badaniu poddano szereg wygenerowany z odwzorowania logistycznego, postaci $x_{t+1} = 4x_t(1 - x_t)$, gdzie $x_0 = 0,7$ (2000 obserwacji). Analizowany szereg jest chaotyczny (por. np. Orzeszko, 2005) i, jak widać, jego kolejne obserwacje powiązane są deterministyczną, ciągłą zależnością. Dla porównania, testowaniu poddano również szereg białoszumowy, o rozkładzie jednostajnym na odcinku jednostkowym (2000 obserwacji).

Na rysunku 1 przedstawiono wykresy, ilustrujące zależność $\varepsilon(\delta)$ od różnych poziomów δ , dla badanych szeregów.

Rys. 1. Wyniki testu Kaplana
a) Odwzorowanie logistyczne



b) Biały szum



Źródło: obliczenia własne.

Zgodnie z oczekiwaniami, w przypadku szeregu deterministycznego, zauważalna jest zbieżność $\varepsilon(\delta) \rightarrow 0$ przy $\delta \rightarrow 0$. Spodziewany rezultat

otrzymano również dla szeregu białoszumowego, tzn. wraz ze spadkiem δ , obliczone wartości $\varepsilon(\delta)$ nie zbiegają do zera lecz do pewnej dodatniej wartości E . Można udowodnić analitycznie, że dla analizowanego szeregu losowego teoretyczną wartością liczby E jest $\frac{1}{3}$ (Kaplan 1994). Widoczna na rysunku 1b) krzywa rzeczywiście zbiega z dużą dokładnością do oczekiwanej wartości.

2) Wykorzystanie testu Kaplana do identyfikacji szeregów ekonomicznych

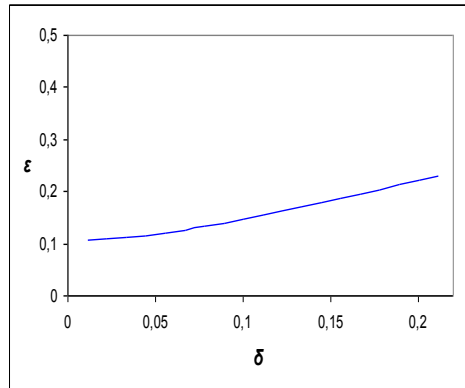
Wydaje się, że prosta dychotomia determinizm – losowość nie jest właściwa w przypadku analizy rzeczywistych procesów ekonomicznych. Abstrahując od filozoficznej kwestii, czym jest w istocie losowość, w praktyce można przyjąć, że dane ekonomiczne zawsze zawierają szum losowy. Reprezentuje on przede wszystkim błędy pomiaru, a także czynniki egzogeniczne, wpływające na dynamikę systemu, których odrębna identyfikacja jest niemożliwa lub niecelowa.

W takiej sytuacji podstawowy cel testu Kaplana – weryfikacja determinizmu, staje się co najmniej wątpliwy. Okazuje się jednak, że w praktyce, zakres możliwości stosowania tej metody jest dużo szerszy. Oprócz ciągłych zależności deterministycznych, identyfikuje ona również wiele struktur innego typu. Z tego powodu test Kaplana może być traktowany jako narzędzie weryfikacji hipotezy o niezależności danych. Oczywiście, ze względu na swoją nieparametryczność, test ten nie prowadzi bezpośrednio do określenia charakteru zidentyfikowanych zależności. Jednak filtrowanie danych modelem określonego typu może prowadzić do uzyskania tej ważnej informacji. Przykładowo, dane, w których w oparciu o test Kaplana zidentyfikowano strukturę, można przefiltrować modelem liniowym. Odrzucenie hipotezy o losowości reszt modelu, będzie świadczyło wówczas o nieliniowym charakterze zidentyfikowanych zależności.

W celu zilustrowania wpływu szumu losowego na wyniki testu Kaplana, do szeregu logistycznego dodano obserwacyjny szum losowy, którego współczynnik $SNR = 10$.¹ Na rysunku 2 przedstawiono wyniki testu dla analizowanego szeregu.

¹ SNR (*signal-to-noise ratio*) liczony jest jako stosunek odchylenia standardowego szeregu wygenerowanego przez część deterministyczną do odchylenia szumu losowego.

Rys. 2. Wyniki testu Kaplana dla szeregu logistycznego z szumem



Źródło: obliczenia własne.

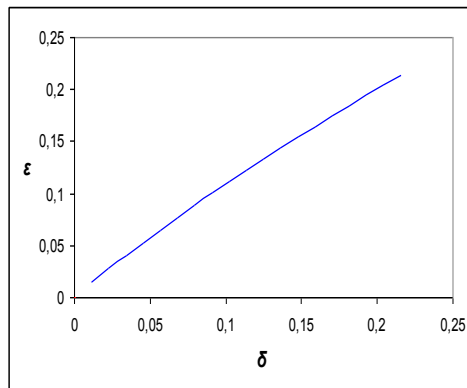
Porównując otrzymany wykres z rysunkiem 1a), można zauważyć podstawową różnicę – wartości $\varepsilon(\delta)$ nie zbiegają do 0 przy $\delta \rightarrow 0$. Mimo to, w oparciu o test Kaplana, możliwe jest zidentyfikowanie struktury, obecnej w analizowanym szeregu. W tym celu, w pierwszej kolejności, należy dla badanego szeregu skonstruować m -wymiarowe wektory opóźnień, postaci $\hat{x}_t^m = (x_t, x_{t+1}, \dots, x_{t+(m-1)})$, gdzie m jest ustaloną z góry liczbą naturalną, nazywaną wymiarem zanurzenia. Następnie, analogicznie do procedury przedstawionej w rozdziale 1, należy obliczyć $\delta_{i,j} = \|\hat{x}_i^m - \hat{x}_j^m\|$, $\varepsilon_{i,j} = \|\hat{x}_{i+1}^m - \hat{x}_{j+1}^m\|$ (gdzie $\|\cdot\|$ jest m -wymiarową normą euklidesową) i wyznaczyć graficznie zależność między $\varepsilon(\delta)$ a δ .²

Na rysunkach 3a)-c) zilustrowano wyniki testu Kaplana z przyjętą wartością wymiaru zanurzenia $m = 2$ dla szeregu logistycznego, białego szumu i szeregu logistycznego z dodanym szumem losowym.

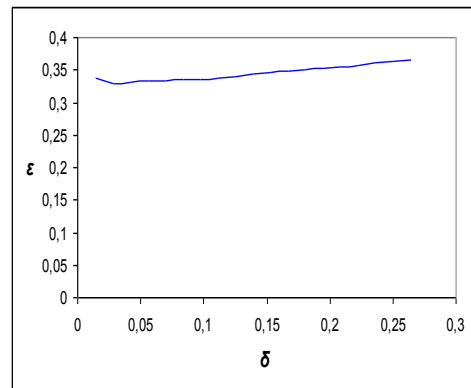
² Wartość parametru $m=1$ prowadzi do procedury opisanej w rozdziale pierwszym.

Rys. 3. Wyniki testu Kaplana dla $m=2$

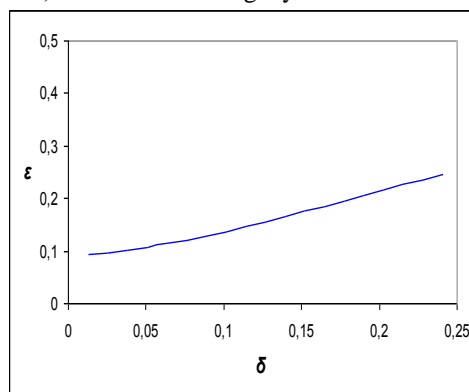
a) odwzorowanie logistyczne



b) biały szum



c) odwzorowanie logistyczne z szumem



Źródło: obliczenia własne.

Jak widać, uwzględnienie wyższego wymiaru zanurzenia nie zmienia zasadniczych własności otrzymanych krzywych, tzn. nadal zbiegają one do wartości 0 dla struktury deterministycznej i do $E > 0$ dla danych stochastycznych.

Zidentyfikowanie testem Kaplana obecności struktury stochastycznej w szeregu jest możliwe dzięki porównaniu otrzymanego wykresu z oczekiwanym dla szeregu pozbawionego zależności. Należy jednak podkreślić, że szereg losowy, który ma być bazą do porównania musi

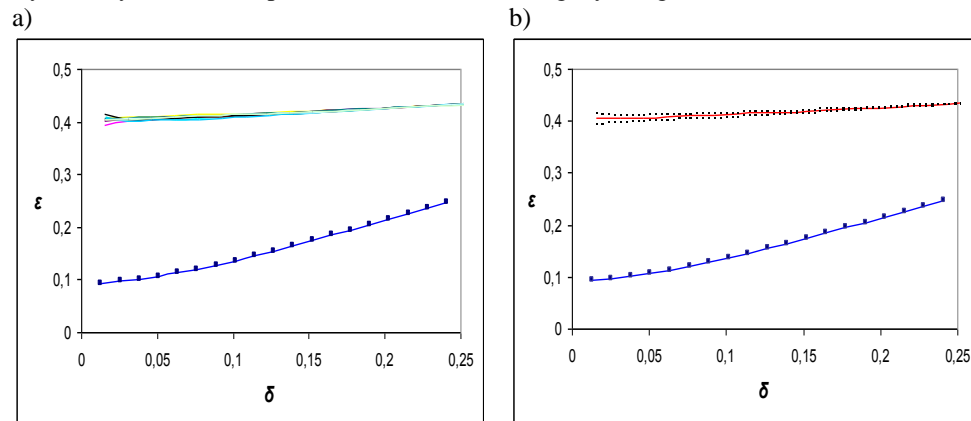
mieć ten sam histogram, co badane dane. Można go uzyskać poprzez losowe wymieszanie obserwacji w pierwotnym szeregu³.

Na rysunku 4a) przedstawiono wykres testu Kaplana dla odwzorowania logistycznego z szumem (linia ze znacznikami) oraz wiązkę 10 wykresów, otrzymanych dla szeregów wymieszanych. Jak widać, krzywa dla szeregu oryginalnego jest położona znacząco niżej, co oznacza, że test zidentyfikował w nim pewną strukturę.

Analizę tę można uczynić bardziej precyzyjną, wyznaczając, w oparciu o uśrednione wartości $\varepsilon(\delta)$ szeregów wymieszanych, pojedynczej krzywej oczekiwanej, a także otoczenia tej krzywej o szerokości będącej ustaloną wielokrotnością odchylenia standardowego tych wartości. Jeśli krzywa otrzymana dla szeregu oryginalnego wykracza poza zakres wyznaczony przez krzywą oczekiwaną (szczególnie dla małych poziomów $\delta > 0$), wówczas należy odrzucić hipotezę o niezależności.

Na rysunku 4b) linią ciągłą zaznaczono krzywą oczekiwaną, a przerywanymi – granice jej otoczenia o szerokości $\pm 2\sigma$, wyznaczone w oparciu o 10 szeregów wymieszanych. Jak widać, test Kaplana skutecznie wykrył istniejące w badanym szeregu zależności.

Rys. 4. Wyniki testu Kaplana dla odwzorowania logistycznego z szumem ($m=2$)⁴



Źródło: obliczenia własne.

³ W teście Kaplana procedura mieszania danych ma sens tylko dla $m > 1$.

⁴ W przypadku każdego z wygenerowanych szeregów dokonano analizy dla $m=2, 3, 4, 5$. Ze względu na ograniczoną objętość artykułu zdecydowano się na prezentację wyników jedynie dla $m=2$. Jednak w każdym z pozostałych przypadków, otrzymano takie same wnioski.

Następnym etapem przeprowadzonych symulacji było zbadanie szeregów (2000 obserwacji), będących realizacjami następujących procesów stochastycznych:

a) ARMA(2,1):

$$x_t = 0,8x_{t-1} + 0,15x_{t-2} + u_t + 0,3u_{t-1},$$

gdzie $x_0 = 1$, $x_1 = 0,7$,

b) GARCH(1,1):

$$x_t = \sqrt{h_t} u_t,$$

$$h_t = 1 + 0,1x_{t-1}^2 + 0,8h_{t-1},$$

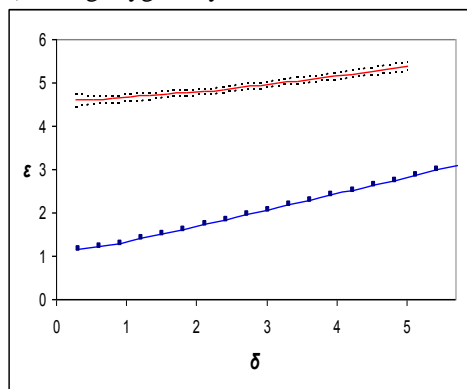
gdzie $h_0 = 1$ i $x_0 = 0$.⁵

Na rysunku 5a) przedstawiono wynik testu Kaplana dla szeregu wygenerowanego z procesu ARMA(2,1). Jak widać, wynik testu jednoznacznie wskazuje, że w danych istnieje pewna struktura. W celu określenia jej charakteru, szereg przefiltrowano modelem liniowym typu ARMA. W oparciu o kryterium Schwarza odpowiednim modelem okazał się proces ARMA(1,1).

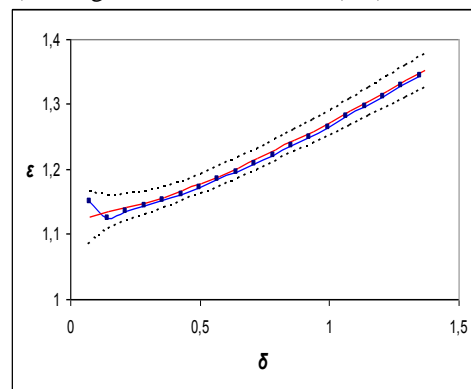
Na rysunku 5b) zaprezentowano wynik testu Kaplana dla reszt oszacowanego modelu linowego. Tym razem test wykazał brak istnienia jakiegokolwiek struktury w analizowanych danych. Oznacza to, że test poprawnie wykazał, że zidentyfikowane w oryginalnym szeregu zależności są liniowe.

Rys. 5. Wyniki testu Kaplana dla szeregu ARMA(2,1) ($m=2$)

a) szereg oryginalny



b) szereg reszt modelu ARMA(1,1)



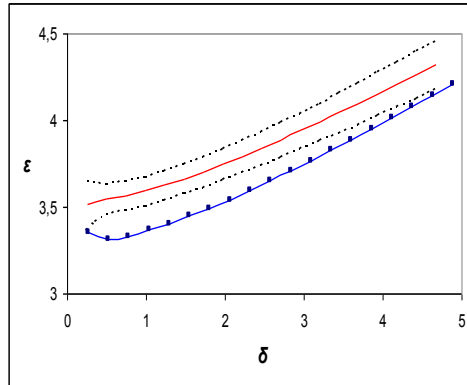
Źródło: obliczenia własne.

⁵ W obu procesach (u_t) jest gaussowskim białym szumem.

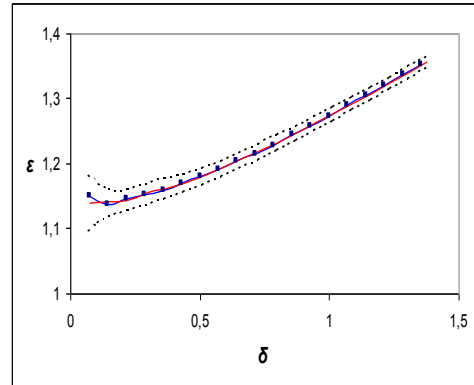
Następnie testowaniu poddano szereg wygenerowany z procesu GARCH(1,1). Widoczny na rysunku 6a) wykres świadczy, że w badanym szeregu obecne są pewne zależności.

Rys. 6. Wyniki testu Kaplana dla szeregu GARCH(1,1) ($m=2$)

a) szereg oryginalny



b) szereg reszt modelu GARCH(1,1)



Źródło: obliczenia własne.

Następnie analizowany szereg przefiltrowano modelem GARCH. Jak wynika z rysunku 6b, oszacowany w oparciu o kryterium Schwarzera model GARCH(1,1), dobrze opisuje zidentyfikowaną strukturę. Oznacza to, że test Kaplana poprawnie zidentyfikował badany proces jako GARCH.

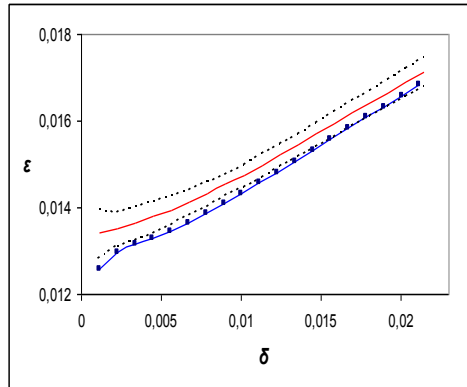
3) Analiza indeksu WIG

W niniejszym rozdziale zastosowano test Kaplana do analizy danych z Giełdy Papierów Wartościowych w Warszawie. Badaniu poddano logarytmiczne, dzienne stopy zmian indeksu WIG z okresu 2.01.2001-12.06.2008 (1867 obserwacji).

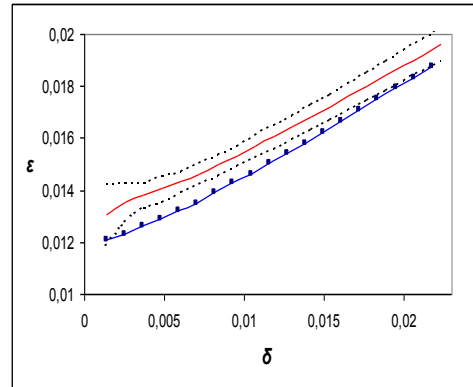
Na wykresach 7a)-d) zilustrowano wyniki testu Kaplana dla analizowanego szeregu czasowego.

Rys. 7. Wyniki testu Kaplana dla logarytmicznych stóp zmian indeksu WIG

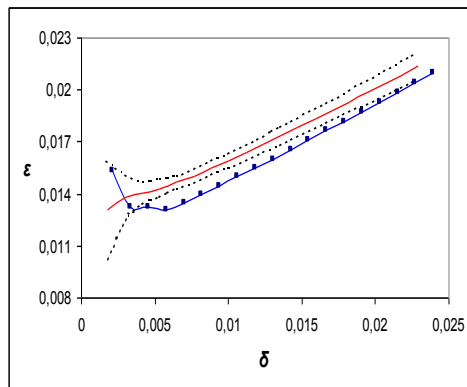
a) $m=2$



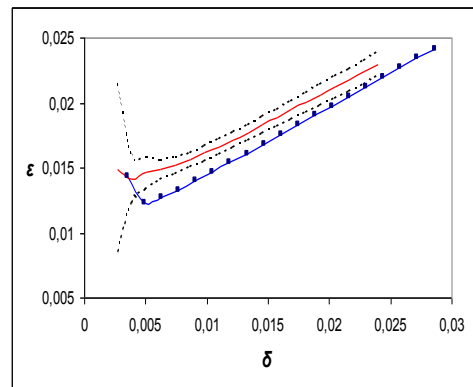
b) $m=3$



c) $m=4$



d) $m=5$



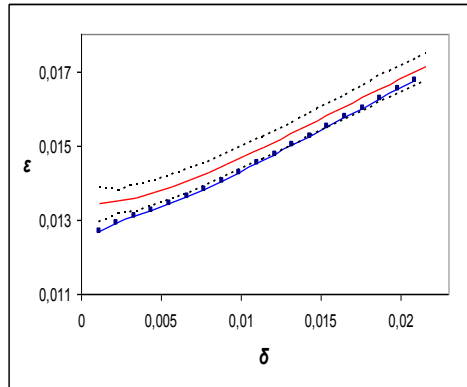
Źródło: obliczenia własne.

Otrzymane wyniki świadczą o istnieniu pewnej struktury w szeregu stóp zmian indeksu WIG. W celu zidentyfikowania charakteru znalezionych zależności, badany szereg przefiltrowano modelem ARMA. W oparciu o kryterium Schwarzera wybrano do tego celu model AR(1).

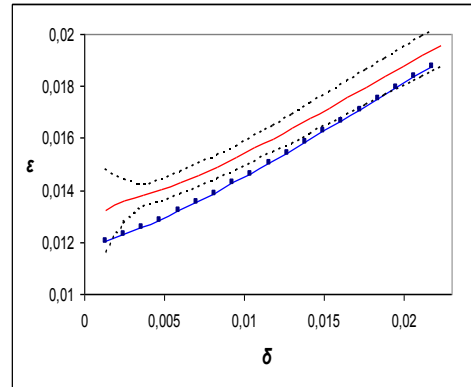
Na wykresach 8a)-d) zilustrowano wyniki testu Kaplana dla reszt z oszacowanego modelu AR(1).

Rys. 8. Wyniki testu Kaplana dla stóp zmian WIG przefiltrowanych modelem AR(1)

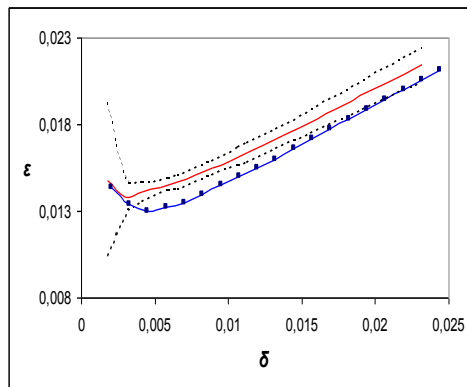
a) $m=2$



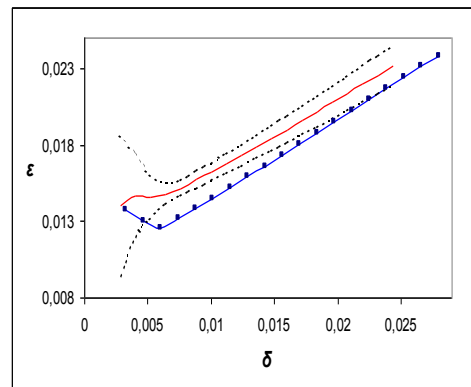
b) $m=3$



c) $m=4$



d) $m=5$



Źródło: obliczenia własne.

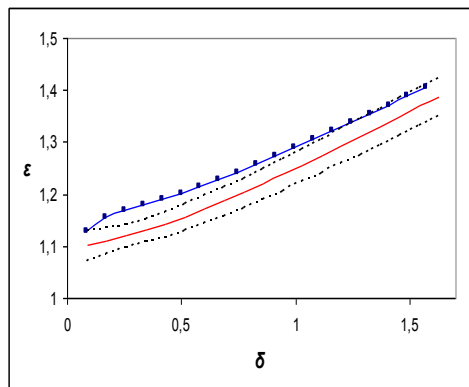
Jak widać, przefiltrowanie danych modelem AR(1) nie wyeliminowało wykrytych wcześniej zależności. Jest to szczególnie wyraźnie widoczne dla wymiaru zanurzenia $m = 3$.

W przypadku danych finansowych bardzo prawdopodobne jest, że powodem ten sytuacji jest obecność w resztach efektu ARCH. Z tego powodu, w następnej kolejności, stopy zmian indeksu WIG przefiltrowano modelem AR(1)-GARCH(1,1) i otrzymane reszty poddano testowaniu.

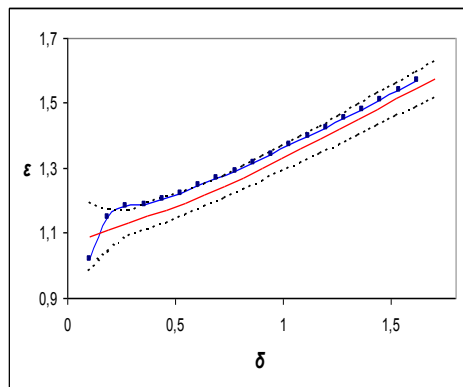
Na wykresach 9a)-d) zilustrowano wyniki testu Kaplana dla standaryzowanych reszt z modelu AR(1)-GARCH(1,1).

Rys. 9. Wyniki testu Kaplana dla stóp zmian indeksu WIG przefiltrowanych modelem AR(1)-GARCH(1,1)

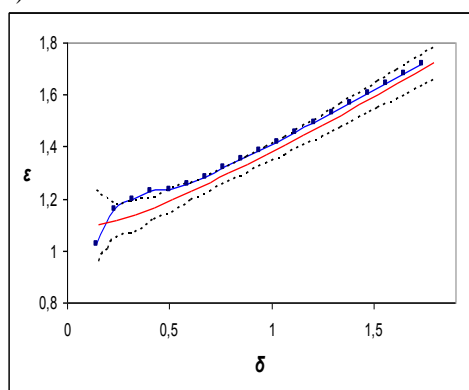
a) $m=2$



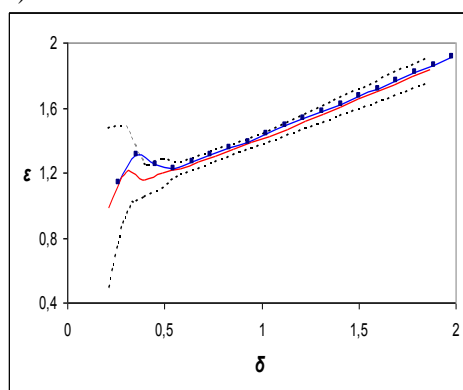
b) $m=3$



c) $m=4$



d) $m=5$



Źródło: obliczenia własne.

Dla $m = 2$ badana krzywa znajduje się poza obszarem wyznaczonym przez krzywą oczekiwaną. Oznacza to, że oszacowany model AR-GARCH nie był w stanie w pełni opisać istniejących w danych zależności. Świadczy to, że w analizowanym szeregu stóp zmian istnieją nieliniowe zależności innej natury niż GARCH.

4) Wnioski

Przeprowadzone symulacje potwierdziły przydatność testu Kaplana do identyfikacji zależności w szeregach czasowych. Skuteczność metody została potwierdzona nie tylko dla danych typowo deterministycznych, ale również posiadających strukturę stochastyczną.

W zastosowaniu do szeregu logarytmicznych stóp zmian indeksu WIG, test wykazał istnienie zależności, które tylko w ograniczonym stopniu udało się opisać modelem AR-GARCH.

Podobnie, jak inne, nieparametryczne metody testujące hipotezę o niezależności, test Kaplana nie daje wprost informacji o naturze wykrytych zależności. Z tego powodu otrzymane wyniki należy traktować jako potwierdzenie celowości dalszej analizy, prowadzącej do znalezienia odpowiedniego modelu dla badanych danych. Właściwe opisanie wykrytych zależności daje m.in. potencjalną możliwość znalezienia lepszego predyktora dla indeksu WIG niż model AR-GARCH.

Należy jednocześnie podkreślić, że potrzebne są dalsze prace nad udoskonaleniem testu Kaplana. Szczególnie potrzebne jest jego sformalizowanie, aby dokonywane wnioski były bardziej ściśle od strony statystycznej.

Literatura

- Barnett, W. A., Gallant A. R., Hinich M. J., Jungeilges J. A., Kaplan D., Jensen M. J. (1998), A single-blind controlled competition among tests for nonlinearity and chaos, *Journal of Econometrics*, vol. 82.1.
- Kaplan, D. (1994), Exceptional events as evidence for determinism, *Physica D*, vol. 73.
- Orzeszko, W. (2005), *Identyfikacja i prognozowanie chaosu deterministycznego w ekonomicznych szeregach czasowych*, seria: Nowe Trendy w Naukach Ekonomicznych, Fundacja Promocji i Akredytacji Kierunków Ekonomicznych, Polskie Towarzystwo Ekonomiczne, Warszawa.

Identification of economic time series using Kaplan's test

Identification of autodependencies in economic time series plays a crucial role in econometric modeling. Due to a higher degree of development, the linear framework has been dominating so far. However, an application of nonlinear analysis seems to be a very promising area in econometrics.

In this paper, the Kaplan's test for nonlinear dependencies in time series is presented. Its usefulness has been verified using simulated data. Moreover, the test has been applied to financial time series from the Warsaw Stock Exchange.