



AKADEMIA GÓRNICZO-HUTNICZA IM. STANISŁAWA STASZICA W KRAKOWIE

Wydział Zarządzania

Projekt EFID - Projekt 6B

Autorzy: *Patrycja Piła, Mateusz Strojek, Julia Szutka, Magdalena Wnuk*

Kierunek studiów: *Informatyka i Ekonometria*

Prowadzący: *dr hab. Tomasz Wójtowicz*

Kraków, 2025

1. Wstęp

Celem analizy jest zbadanie właściwości szeregu czasowego dziennych stóp zwrotu indeksu WIG20 z okresu od stycznia 2017 roku do chwili obecnej oraz ocena zdolności progностycznej modeli klasy GARCH w różnych wariantach.

W drugiej części analizy badanie rozszerzono o podejście z oknami przesuwanyimi, co pozwala ocenić stabilność parametrów modelu i jego zdolność progностyczną w czasie. Analizę przeprowadzono dla okien o długości 100, 200 oraz 500 obserwacji, przesuwanych tak, aby koniec pierwszego okna przypadał na ostatni dzień 2018 roku. Dla każdego roku ponownie dobierano optymalny model ARMA opisujący średnią warunkową, a następnie osadzano go w modelach warunkowej wariancji: GARCH, EGARCH oraz GJR-GARCH, z trzema różnymi rozkładami składnika losowego (normalnym, t-Studenta oraz skośnym t-Studenta), co daje 9 kombinacji na każde okno.

W każdym oknie obliczano 95% przedziały ufności dla prognozowanej kolejnej stopy zwrotu i sprawdzano, czy rzeczywista wartość mieści się w tym przedziale. Na tej podstawie określano odsetek przypadków nietrafienia oraz analizowano, czy błędy prognoz występują częściej w określonych okresach i czy ich rozkład był symetryczny. Dodatkowo zbadano, jak zgodność reszt z przyjętym rozkładem warunkowym zmieniała się w czasie i czy wybór modelu wariancji oraz rozkładu miał wpływ na jakość prognoz.

2. Eksploracja danych

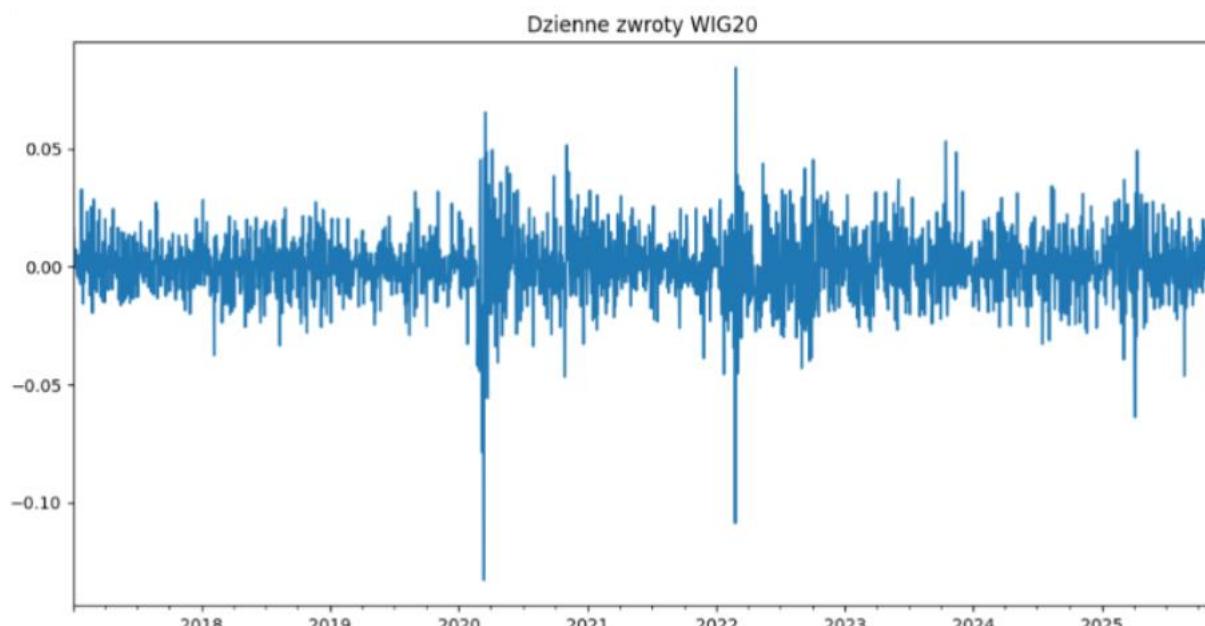
2.1 Statystyki opisowe

Tabela 1. Statystyki opisowe dziennych stóp zwrotu indeksu WIG20

Statystyka	Wartość
Liczba obserwacji	2305
Średnia	0,0003
Mediana	0,0000
Odchylenie Standardowe	0,0138
Minimum	-0,1328
Maksimum	0,0844
Rozstęp	0,2171
Skośność	-0,482
Eksces	7,328

W tabeli 1 przedstawiono podstawowe statystyki opisowe zwykłych dziennych stóp zwrotu indeksu WIG20 w okresie od 1 stycznia 2017 do 5 listopada 2025 roku. Próba obejmuje 2305 obserwacji.

Średnia dzienna stopa zwrotu wynosi 0,0003, a mediana 0,0000, co oznacza, że przeciętne dzienne zmiany wartości indeksu były bardzo niewielkie i oscylowały wokół zera. Wskazuje to, że w badanym okresie nie występował wyraźny trend wzrostowy ani spadkowy. Ujemna skośność (-0,48) oznacza, że więcej dni przynosi spadki niż wzrosty, a rozkład stóp zwrotu jest lekko przesunięty w stronę wartości ujemnych. Z kolei wysoka eksces (7,33) potwierdza obecność grubych ogonów i zwiększonej liczby obserwacji skrajnych w porównaniu z rozkładem normalnym.



Rysunek 1. Dzienne zwroty WIG20

Analiza statystyk opisowych wskazuje, że dzienne zwykłe stopy zwrotu indeksu WIG20 w latach 2017-2025 cechowały się niewielką przeciętną stopą zwrotu oraz asymetrycznym, leptokurycznym rozkładem, typowym dla danych finansowych, w których często występują epizody podwyższonej niepewności i gwałtownych zmian rynkowych.

2.2. Test ADF

W celu sprawdzenia czy szereg dziennych stóp zwrotu indeksu WIG20 jest stacjonarny, przeprowadzono test Augmented Dickey-Fullera (ADF). Test ten pozwala ocenić, czy w szeregu występuje pierwiastek jednostkowy, czyli składnik powodujący niestacjonarność.

Hipotezy testu:

- $H_0: X_t \sim I(1)$,

- $H_1: X_t \sim I(0)$.

Tabela 2. Wyniki testu ADF dla dziennych stóp zwrotu indeksu WIG20

Statystyka	Wartość p	Krytyczna wartość na poziomie 1%	Krytyczna wartość na poziomie 5%	Krytyczna wartość na poziomie 10%
-11,1823	0,0000	-3,4332	-2,8628	-2,5674

W tabeli 2 otrzymana statystyka testowa (-11,1670) jest mniejsza od wszystkich wartości krytycznych, a wartość $p = 0,00$ jest znacznie niższa od poziomu istotności 0,05. Oznacza to, że odrzucamy hipotezę zerową o niestacjonarności szeregu.

Szereg dziennych stóp zwrotu WIG20 jest stacjonarny, co oznacza, że jego średnia i wariancja nie zmieniają się w czasie.

3. Okno 100 obserwacji

3.1 Wybór rzędu modelu ARMA dla kolejnych lat

Dla analizy opartej na kroczącym oknie o rozmiarze 100 obserwacji, w każdym analizowanym roku i dla każdego okna, procedura selekcji rzędu modelu ARMA konsekwentnie wskazywała na $p = 0$ i $q = 0$, co skutkowało wyborem modelu ARMA(0,0).

Łącznie wygenerowano 1786 okien, co dało 16074 pojedyncze prognozy. Postęp obliczeń przesuwał się równomiernie przez lata i wskazuje, że każde kolejne okno było poprawnie przesuwane o jedną obserwację.

3.2 Analiza skuteczności prognoz

Tabela 3. Skuteczność trafień w 95% przedziału ufności dla modeli GARCH w oknie 100 obserwacji

Model	Rozkład	Trafienia	Odsetek trafień (%)	Poniżej przedziału	Powyżej przedziału	Średnie AIC	Zgodność z rozkładem (%)
EGARCH	normalny	1615	90,43	83	88	391006,6191	96,75
	skośny t-Studenta	1670	93,51	56	60	-452,1111	93,28
	t-Studenta	1666	93,28	59	61	-531,5438	98,49
GARCH	normalny	1669	93,45	56	61	-579,4067	98,15
	skośny t-Studenta	1713	95,91	34	39	-578,9020	99,55
	t-Studenta	1706	95,52	38	42	-580,1872	100,00
GJR-GARCH	normalny	1664	93,17	62	60	-580,3333	100,00
	skośny t-Studenta	1705	95,46	39	42	-579,0238	100,00

t-Studenta	1693	94,79	44	49	-580,2490	100,00
------------	------	-------	----	----	-----------	--------

Tabela 3 przedstawia skuteczność prognoz w oknie 100 obserwacji dla trzech kombinacji modeli warunkowej wariancji (GARCH, EGARCH, GJR-GARCH) i trzech rozkładów warunkowych (normalny, skośny t-Studenta, t-Studenta).

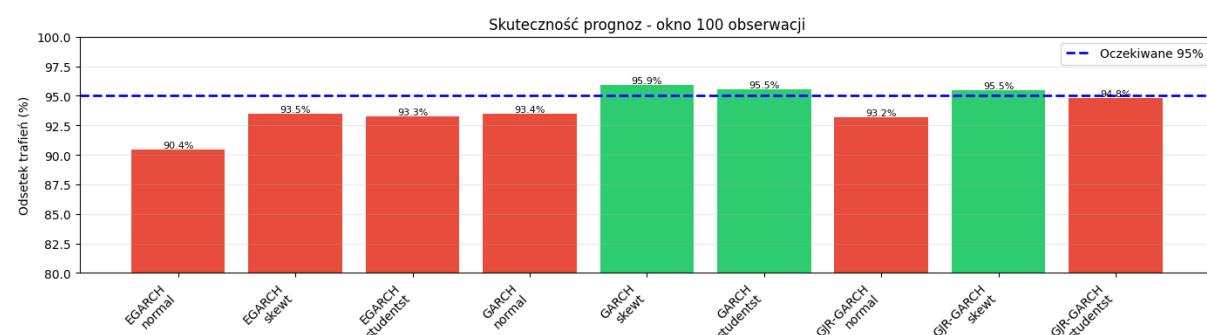
Najniższą skuteczność uzyskał model EGARCH z rozkładem normalnym (90,43% trafień), co pokazuje, że rozkład normalny nie radzi sobie tak dobrze z grubymi ogonami i asymetrią stóp zwrotu jak reszta rozkładów. Zastosowanie bardziej „elastycznych” rozkładów jak t-Studenta czy skośnego t-Studenta poprawia wyniki, podnosząc skuteczność do poziomu kolejno 93,28% i 93,51% w modeli EGARCH.

Najlepiej wypadają modele GARCH i GJR-GARCH z rozkładami t-Studenta i skośnym t-Studenta (ok. 94,79-95,91%).

Pod względem kryterium AIC zdecydowanie najgorszy okazał się EGARCH z rozkładem normalnym, osiągając wartość aż 391006,6191, co wskazuje na skrajnie słabe dopasowanie tego modelu. Z kolei najlepsze wynik uzyskały modele GARCH oraz GJR-GARCH, które - niezależnie od zastosowanego rozkładu - osiągały bardzo zbliżone wartości AIC w przedziale około -578 do -580, co świadczy o stabilnym i porównywalnie dobrym dopasowaniu.

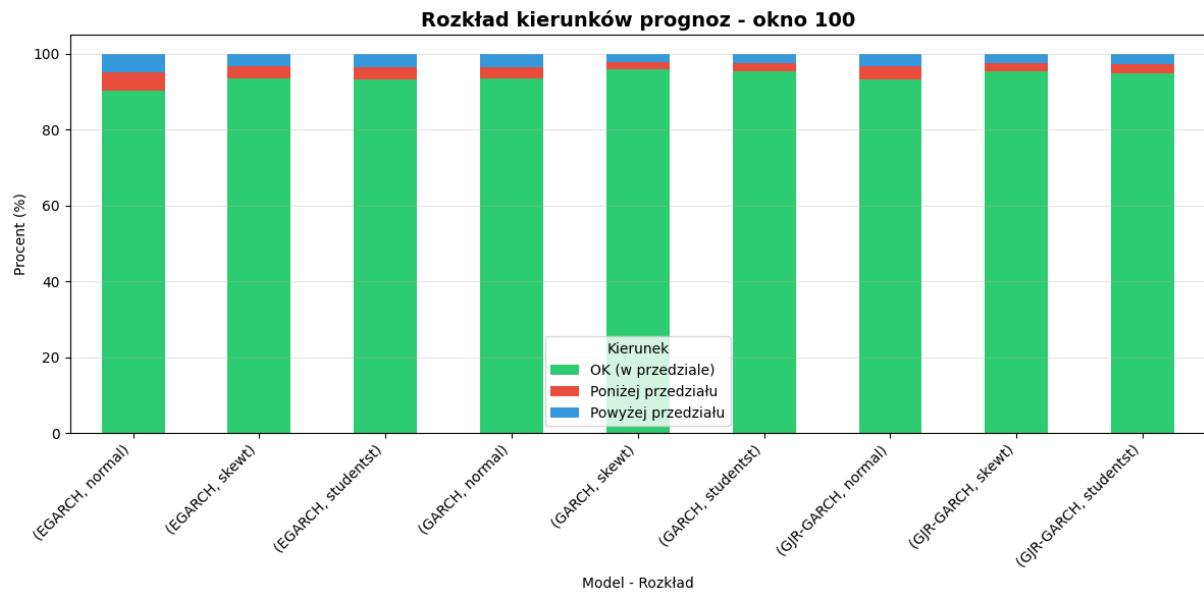
Wśród modeli EGARCH najlepszy wynik uzyskała wersja z rozkładem t-Studenta (-531,5438) a po nim EGARCH z skośnym rozkładem t-Studenta (-452,1111). Mimo to oba te modele pozostają słabsze od wariantów GARCH i GJR-GARCH.

Kombinacje modeli GARCH i GJR-GARCH z rozkładami t-Studenta oraz skośnym t-Studenta wykazują bardzo wysoką zgodność składników losowych z założonym rozkładem (99,55% dla GARCH ze skośnym t-Studentem oraz 100% dla GARCH i GJR-GARCH z rozkładem t-Studenta i jego wersją skośną). Świadczy to o tym, że reszty tych modeli rzeczywiście mają właściwości zgodne z przyjętymi rozkładami, co potwierdza dobre dopasowanie modeli.



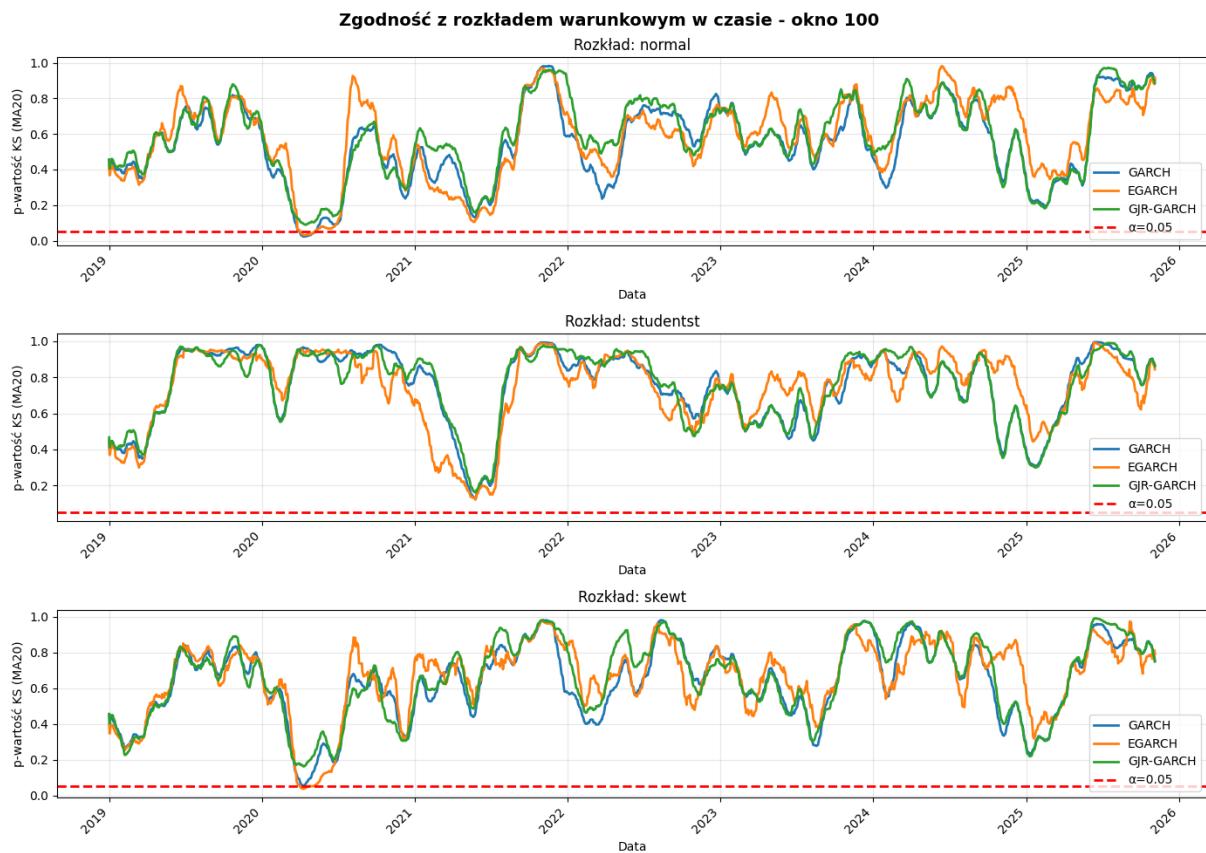
Rysunek 2. Skuteczność trafień w 95% przedziału ufności dla modeli GARCH w oknie 100 obserwacji

Na rysunku 2 można zaobserwować że najwięcej nietrafień - zarówno poniżej dolnej, jak i powyżej górnej granicy przedziału - pojawiało się konsekwentnie w modelach EGARCH, niezależnie od zastosowanego rozkładu. Modele GARCH i GJR-GARCH wypadały pod tym względem stabilniej, a różnice między nimi były niewielkie. Dodatkowo najlepiej zachowywały się wersje oparte na rozkładach t-Studenta i skośnym t-Studenta, podczas gdy rozkład normalny generował zauważalnie więcej nietrafień.



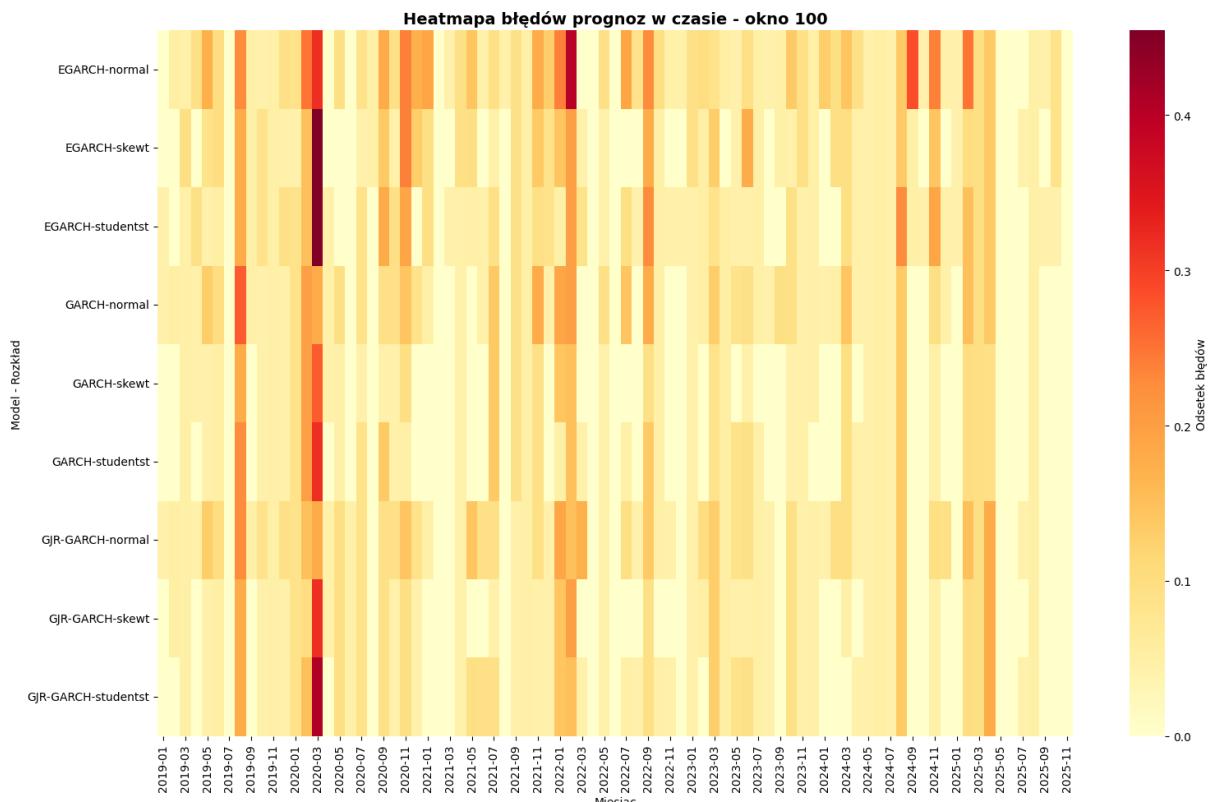
Rysunek 3. Rozkład trafień i nietrafień prognoz względem przedziału ufności - okno 100

Na rysunku 3, który przedstawia wykres rozkładu kierunków prognoz widać, że modele EGARCH generują najwięcej nietrafień, zarówno poniżej dolnej, jak i powyżej górnej granicy przedziału ufności - i dotyczy to każdego z użytych rozkładów. Modele GARCH i GJR-GARCH wypadają pod tym względem znacznie stabilniej, a rozkład błędów jest w nich wyraźnie korzystniejszy. Najmniej nietrafień pojawia się w modelu GARCH z rozkładem skośnym t-Studenta zarówno poniżej, jak i powyżej przedziału.



Rysunek 4. Zgodność z rozkładem warunkowym w czasie – okno 100

Jak można zauważyć na rysunku 4 rozkład normalny wypada najsłabiej - modele oparte na normalności częściej spadają w okolice progu istotności, co potwierdza, że stopy zwrotu WIG20 mają wyraźnie grube ogony i normalność nie jest adekwatnym założeniem. Znacznie lepsze dopasowanie zapewniają rozkład t-Studenta oraz jego skośna wersja, przy czym rozkład t-Studenta daje najwyższą i najbardziej stabilną zgodność w całym badanym okresie. Najlepszą kombinacją (najwyższa zgodność i najmniejsza zmienność w czasie) są modele GARCH lub GJR-GARCH z rozkładem t-Studenta lub skośnym t-Studenta, natomiast najgorszą - EGARCH z rozkładem normalnym.

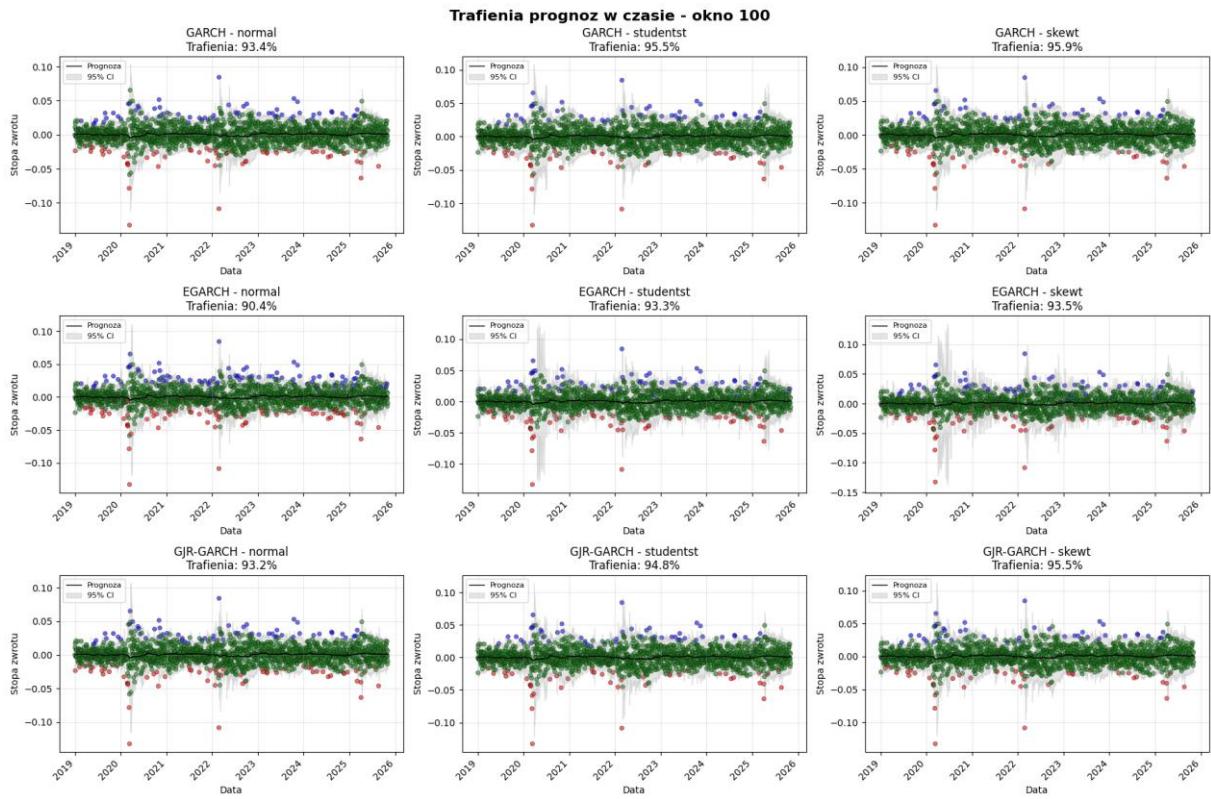


Rysunek 5. Mapa cieplna błędów prognoz w czasie – okno 100

Rysunek 5 przedstawia odsetek błędów prognozy w czasie dla wszystkich kombinacji modeli GARCH, EGARCH i GJR-GARCH oraz trzech rozkładów warunkowych. Ciemniejsze kolory oznaczają wyższy udział przypadków, w których rzeczywista stopa zwrotu wypadła poza 95% przedziału ufności.

Najbardziej „problematyczne” są modele EGARCH, zwłaszcza z rozkładem normalnym i skośnym t-Studenta - pojawiają się u nich wyraźne, ciemniejsze „piki”, świadczące o większej liczbie nietrafień. Modele GARCH i GJR-GARCH mają jaśniejszy profil, co wskazuje na lepszą skuteczność prognoz. Różnice między nimi są niewielkie, a błędy najczęściej pojawiają się w tych samych okresach rynkowych.

Na wykresie wyraźnie widać okresy, w których zgodność z rozkładem warunkowym szczególnie spada. Najciemniejszy i najbardziej problematyczny fragment przypada na marzec 2020 roku. Podwyższoną liczbę odchyleń można również zauważać w sierpniu 2019 oraz lutym 2022 roku.



Rysunek 6. Trafienia prognoz w czasie – okno 100

Wizualna analiza prognoz na rysunku 6 pokazuje wyraźną przewagę modeli GARCH i GJR-GARCH, które tworzą stabilne i spójne 95% przedziały ufności. Szare pasmo wokół prognozy jest u nich stosunkowo wąskie i dobrze dopasowane, a większość rzeczywistych obserwacji (zielone punkty) utrzymuje się w jego zakresie. Nietrafienia (czerwone i niebieskie punkty) pojawiają się głównie w momentach gwałtownej zmienności rynku, takich jak marzec 2020 czy luty 2022. Modele EGARCH wypadają wyraźnie gorzej: ich przedziały są bardziej niestabilne, często nadmiernie reagują na skoki zmienności.

Najbardziej stabilne wizualnie i przejrzyste wykresy uzyskują modele GARCH i GJR-GARCH z rozkładami t-Studenta i skośnym t-Studentem - widać tam równomierny układ zielonych punktów i dobrze skalibrowane przedziały. Rozkład normalny, niezależnie od modelu, częściej prowadzi do ciaśniejszych przedziałów.

4. Okno 200 obserwacji

4.1 Wybór rzędu modelu ARMA dla kolejnych lat

Dla analizy opartej na kroczącym oknie o rozmiarze 200 obserwacji, w każdym analizowanym roku, procedura selekcji rzędu modelu ARMA znowu konsekwentnie wskazywała na $p = 0$ i $q = 0$, co skutkowało ciągłym wyborem modelu ARMA(0,0).

Przeprowadzono łącznie 1786 prognoz dla każdego rozważanego wariantu modelu różniącego się założeniami co do rozkładu składnika losowego. Postępowanie było analogiczne do przypadku okna 100-obserwacyjnego.

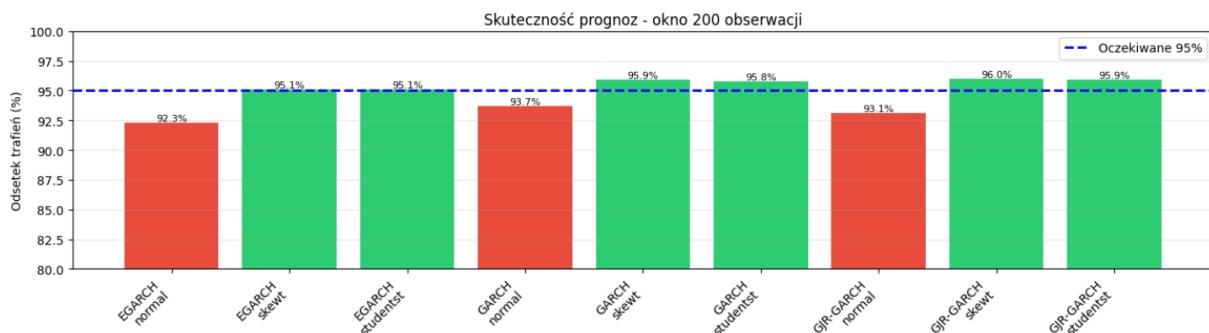
4.2 Analiza skuteczności prognoz

Tabela 4. Skuteczność trafień w 95% przedziału ufności dla modeli GARCH w oknie 200 obserwacji

Model	Rozkład	Trafienia	Odsetek trafień (%)	Poniżej przedziału	Powyżej przedziału	Średnie AIC	Zgodność z rozkładem (%)
EGARCH	normalny	1649	92,33	69	68	363443,5312	91,15
	skośny t-Studenta	1698	95,07	47	41	-733,0457	90,26
	t-Studenta	1698	95,07	47	41	-1076,7507	98,66
GARCH	normalny	1673	93,67	55	58	-1160,7063	91,99
	skośny t-Studenta	1713	95,91	34	39	-1163,8401	96,14
	t-Studenta	1711	95,80	37	38	-1165,2931	100,00
GJR-GARCH	normalny	1663	93,11	59	64	-1165,3371	96,98
	skośny t-Studenta	1715	96,02	33	38	-1166,7821	99,27
	t-Studenta	1713	95,91	35	38	-1168,0599	100,00

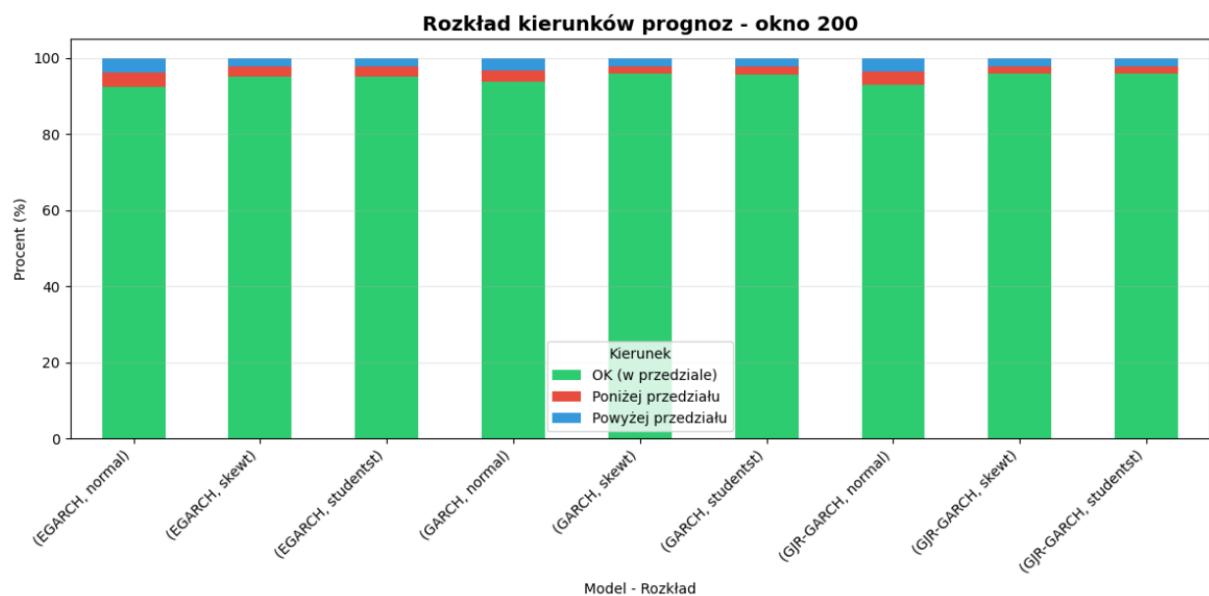
Biorąc uwagę to, że skuteczność prognoz jest mierzona odsetkiem trafień w 95% przedziale ufności, odsetek trafień jest w większości na odpowiednim poziomie, oscylując w zakresie od 92,33% do 96,02%. Przedział ufności dla proporcji 95% na poziomie istotności 0,05 wynosi (0,934; 0,960). Zatem model EGARCH z resztami z warunkowego rozkładu normalnego, jak i GJR-GARCH z resztami z tego samego rozkładu, jest słabej jakości. Najwyższą skuteczność zanotowano dla modelu GJR-GARCH ze skośnym rozkładem t-Studenta, co sugeruje, że uwzględnienie asymetrii i grubych ogonów w rozkładzie reszt poprawia jakość prognoz przedziałowych. Warto zauważyć, że dla wszystkich trzech rodzajów modeli zastosowanie rozkładów alternatywnych względem normalnego prowadziło do wyższej skuteczności predykcji. Modele z rozkładem t-Studenta charakteryzują się najwyższą zgodnością z rozkładem teoretycznym. Dla GARCH i GJR-GARCH zgodność wyniosła 100%.

Rozkład obserwacji wykraczających poza przedział ufności był asymetryczny. Modele częściej wykazywały więcej obserwacji powyżej przedziału niż poniżej, co oznacza, że nasze modele nieco pesymistyczne (oprócz modelu EGARCH). Najniższą wartość kryterium AIC uzyskał model GJR-GARCH z rozkładem t-Studenta (-1168,0599), co wskazuje na jego najlepsze dopasowanie do danych spośród wszystkich rozważanych specyfikacji. Różnice w wartościach AIC między modelami są relatywnie niewielkie, oprócz modelu EGARCH, gdzie dla normalnych reszt średnie AIC wyniosło aż 363443,5312, co może mocno niepokoić.



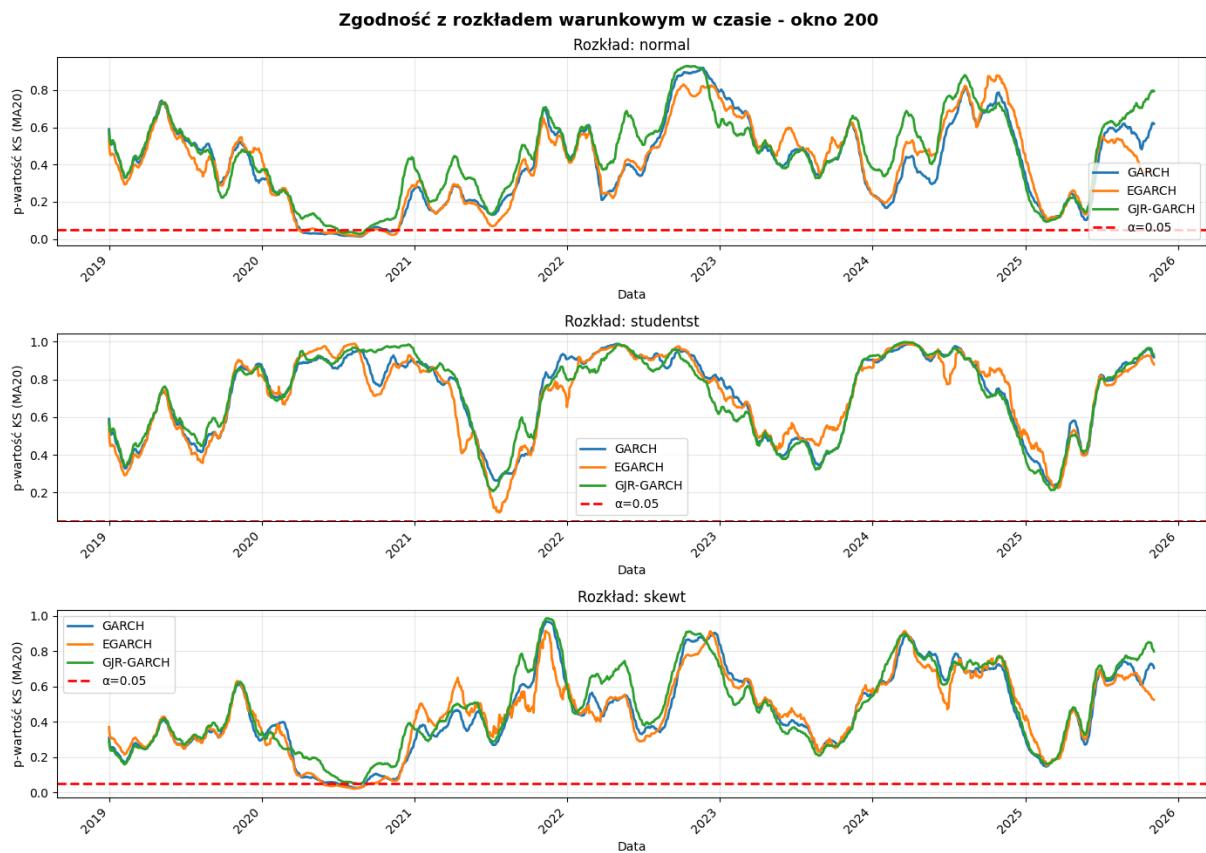
Rysunek 7. Skuteczność trafień w 95% przedziału ufności dla modeli GARCH w oknie 200 obserwacji

Na rysunku 7 można zaobserwować że najwięcej nietrafień - zarówno poniżej dolnej, jak i powyżej górnej granicy przedziału - pojawiało się konsekwentnie w modelach z resztami z rozkładu normalnego. Tak jak wspomniano wcześniej, najgorzej poradził sobie model EGARCH ze wspomnianym rozkładem. Modele z wyjątkiem EGARCH, których reszty nie były z rozkładu normalnego, poradziły sobie na podobnym poziomie z ok. 96% skutecznością prognoz.



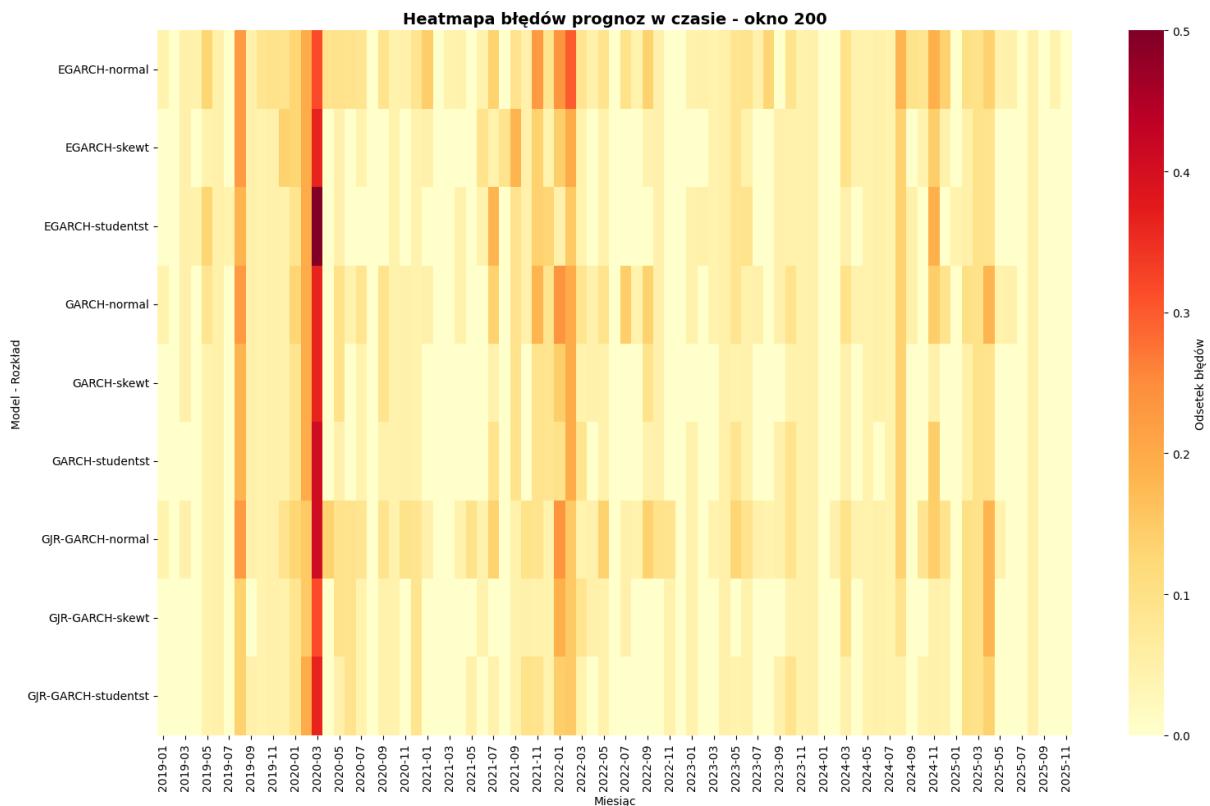
Rysunek 8. Rozkład trafień i nietrafień prognoz względem przedziału ufności - okno 200

Na rysunku 8, który przedstawia wykres rozkładu kierunków prognoz widać, że modele, których reszty są z rozkładu normalnego, dają najgorsze rezultaty. Poza tym widać lekką przewagę pomyłek spowodowanych pesymizmem modeli.



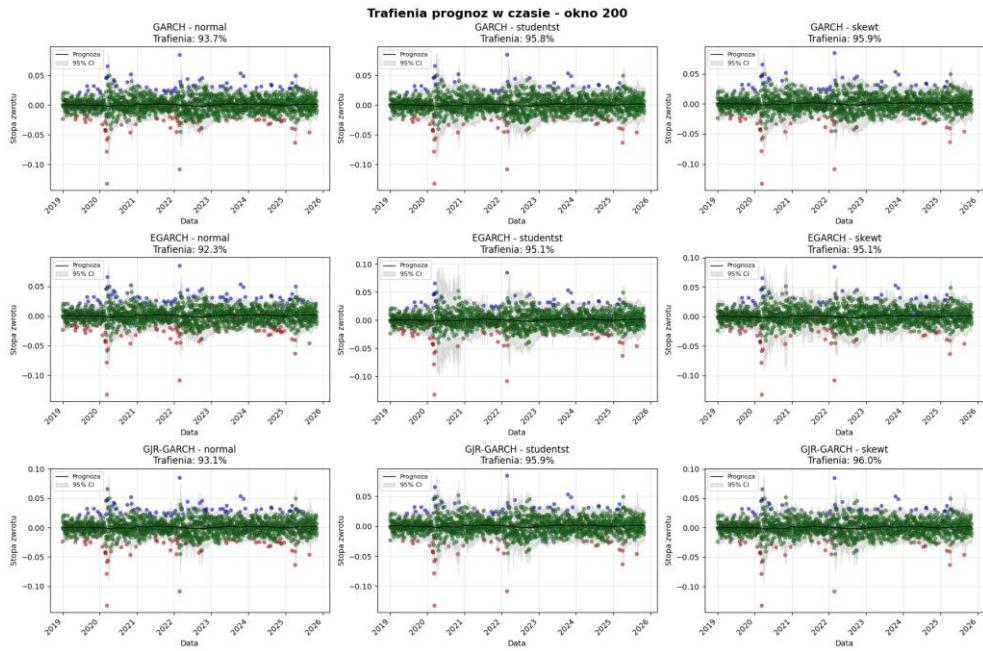
Rysunek 9. Zgodność z rozkładem warunkowym w czasie – okno 200

Jak ilustruje rysunek 9, zgodność z rozkładem warunkowym wykazywała znaczącą zmienność w czasie. Najwyraźniejsze pogorszenie jakości prognoz wystąpiło w okresie pandemii COVID-19, kiedy gwałtowne reakcje rynków znaczco utrudniały precyzyjne przewidywanie. Zjawisko to jest szczególnie widoczne dla modeli z rozkładem normalnym oraz skośnym t-Studenta. Niezależnie od specyfikacji modelu, wzorce zmienności zgodności z rozkładem warunkowym pozostały zbliżone – kluczowa różnica dotyczyła przede wszystkim kształtu trajektorii w zależności od przyjętego założenia o rozkładzie reszt.



Rysunek 10. Mapa cieplna błędów prognoz w czasie – okno 200

Heatmapa błędów prognoz dla okna 200 obserwacji ujawnia wyraźną koncentrację najwyższych błędów w okresie pandemii COVID-19 (marzec-kwiecień 2020), co zaznaczone jest ciemnoczerwonymi obszarami dla wszystkich specyfikacji modeli. Po okresie startu pandemii w Polsce obserwuje się poprawę jakości prognoz we wszystkich modelach, choć sporadyczne wzrosty błędów pojawiają się w 2022 roku. Było to spowodowane rozpoczęciem wojny na Ukrainie. Niezależnie od typu modelu (EGARCH, GARCH, GJR-GARCH), wzorce błędów w czasie są zbliżone, co sugeruje, że źródło problemów tkwi w ekstremalnej zmienności rynku w danych okresach.



Rysunek 11. Trafienia prognoz w czasie – okno 200

Tak jak wspomniano powyżej, niedoszacowanie zmienności pojawiają się głównie w momentach gwałtownej zmienności rynku, takich jak marzec 2020 czy luty 2022. Widać, że modele z normalnymi resztami miały najwęższe przedziały ufności. Z kolei, modele z resztami o warunkowym rozkładzie t-studenta charakteryzowały się tym, że w wyniku dużej zmienności, szeroki przedział ufności utrzymywał się dłużej, niż w pozostałych modelach. Można to zauważyć w szczególności w 2020 roku. Pozostałe trzy modele cechowały się tym, że szerokość przedziałów w momencie wysokiej zmienności nie utrzymywała się tak długo jak w przypadku modeli o resztach z warunkowym rozkładem t-studenta. Jednakże, jakość predykcji była dość zbliżona.

5. Okno 500 obserwacji

5.1 Wybór rzędu modelu ARMA dla kolejnych lat

Tak jak w przypadku dwóch poprzednich okien obserwacji, tak i teraz dla analizy opartej na kroczącym oknie o rozmiarze 500 obserwacji, w każdym analizowanym roku, procedura selekcji rzędu modelu ARMA wskazywała na $p = 0$ i $q = 0$, co skutkowało wyborem modelu ARMA(0,0).

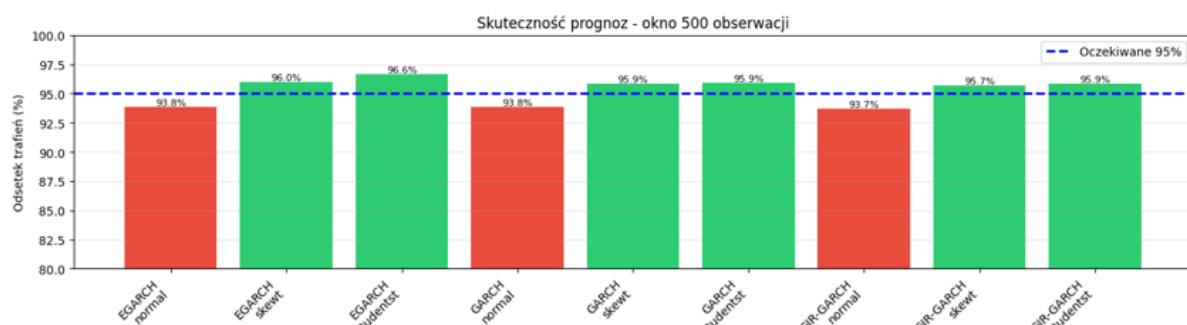
5.2 Analiza skuteczności prognoz

Tabela 5. Skuteczność trafień w 95% przedziału ufności dla modeli GARCH w oknie 500 obserwacji

Model	Rozkład	Trafienia	Odsetek trafień (%)	Poniżej przedziału	Powyżej przedziału	Średnie AIC	Zgodność z rozkładem (%)
EGARCH	normalny	1676	93,84	52	58	-2920,5511	72,12
	skośny t-Studenta	1715	96,02	31	40	-2917,4088	70,66
	t-Studenta	1726	96,64	31	29	-2916,2892	99,94

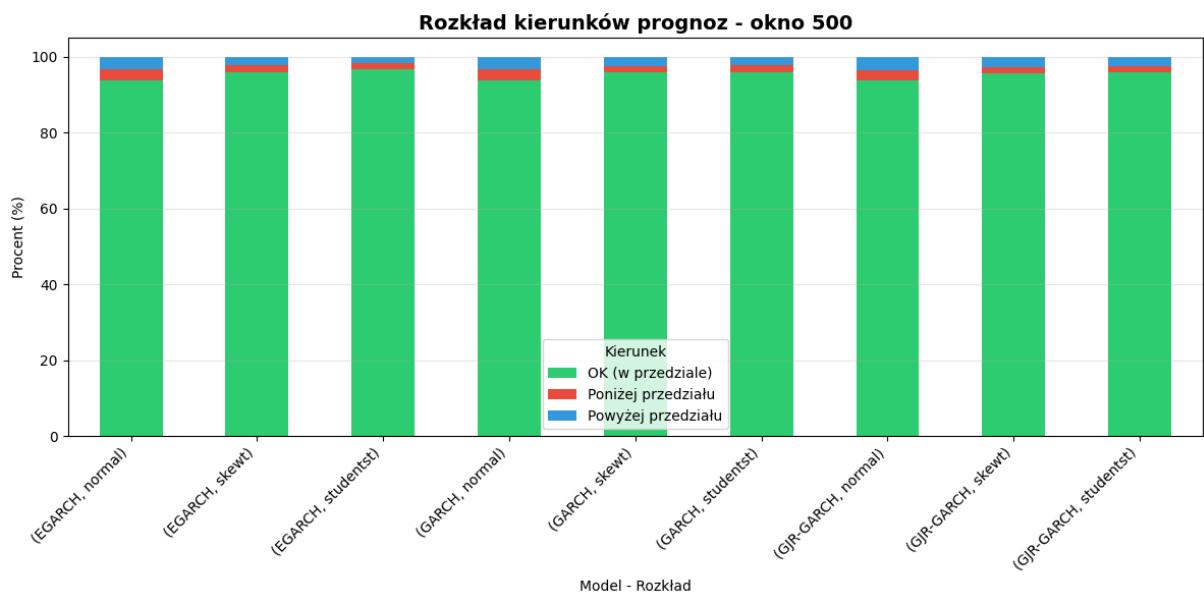
	normalny	1676	93,84	51	59	-2920,1489	73,80
GARCH	skośny t-Studenta	1712	95,86	32	42	-2935,1302	73,07
	t-Studenta	1713	95,91	36	37	-2936,7016	100,00
GJR-GARCH	normalny	1673	93,67	48	65	-2934,6945	88,02
	skośny t-Studenta	1709	95,69	30	47	-2943,8640	91,71
	t-Studenta	1712	95,86	29	45	-2945,5121	100,00

Prognozy wariancji oparte na modelach GARCH wykazują wyraźnie lepszą jakość, gdy wykorzystują rozkład t-Studenta. Modele zakładające normalność błędów systematycznie niedoszacowują zmienność, co skutkuje niższym odsetkiem trafień względem oczekiwanych 95%. Najwyższą skuteczność przedziałową osiąga EGARCH z rozkładem t-Studenta, z wynikiem najbliższym idealnemu poziomowi trafień. Z kolei najniższe wartości AIC wskazują, że najlepiej dopasowanym modelem jest GJR-GARCH z rozkładem t-Studenta. W wielu modelach obserwuje się lekką przewagę obserwacji powyżej przedziału, co sugeruje tendencję do niedoszacowania zmienności podczas silnych wzrostów.



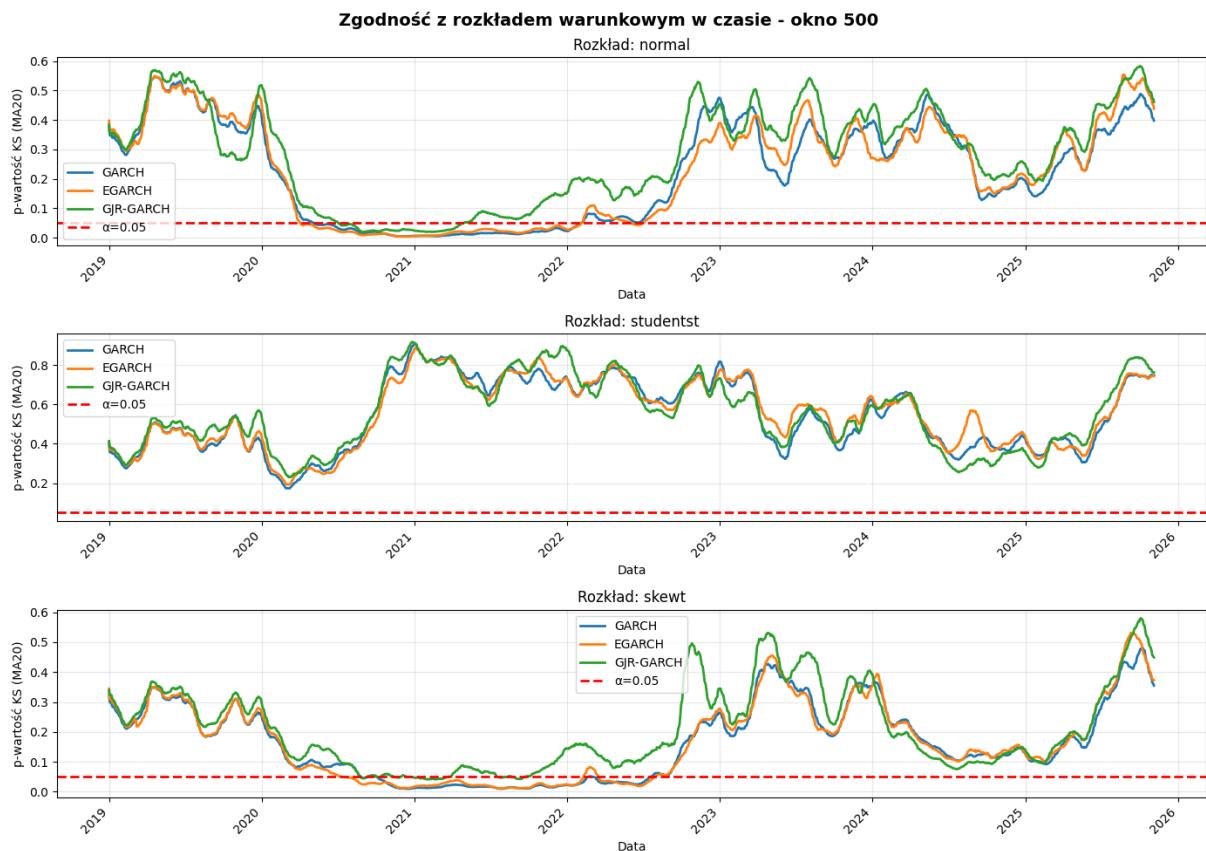
Rysunek 12. Skuteczność trafień w 95% przedziału ufności dla modeli GARCH w oknie 500 obserwacji

Wykres słupkowy pokazuje, że skuteczność prognoz 95% przedziałów ufności zależy przede wszystkim od przyjętego rozkładu błędów, a nie od konkretnej specyfikacji modelu GARCH. Wszystkie modele z rozkładem normalnym niedoszacowują zmienność. Z kolei modele z rozkładem t-Studenta i skośnym t-Studenta osiągają wyniki powyżej 95%, a najlepszy rezultat uzyskuje EGARCH z rozkładem t-Studenta, tak jak zauważono przy wynikach w tabeli.



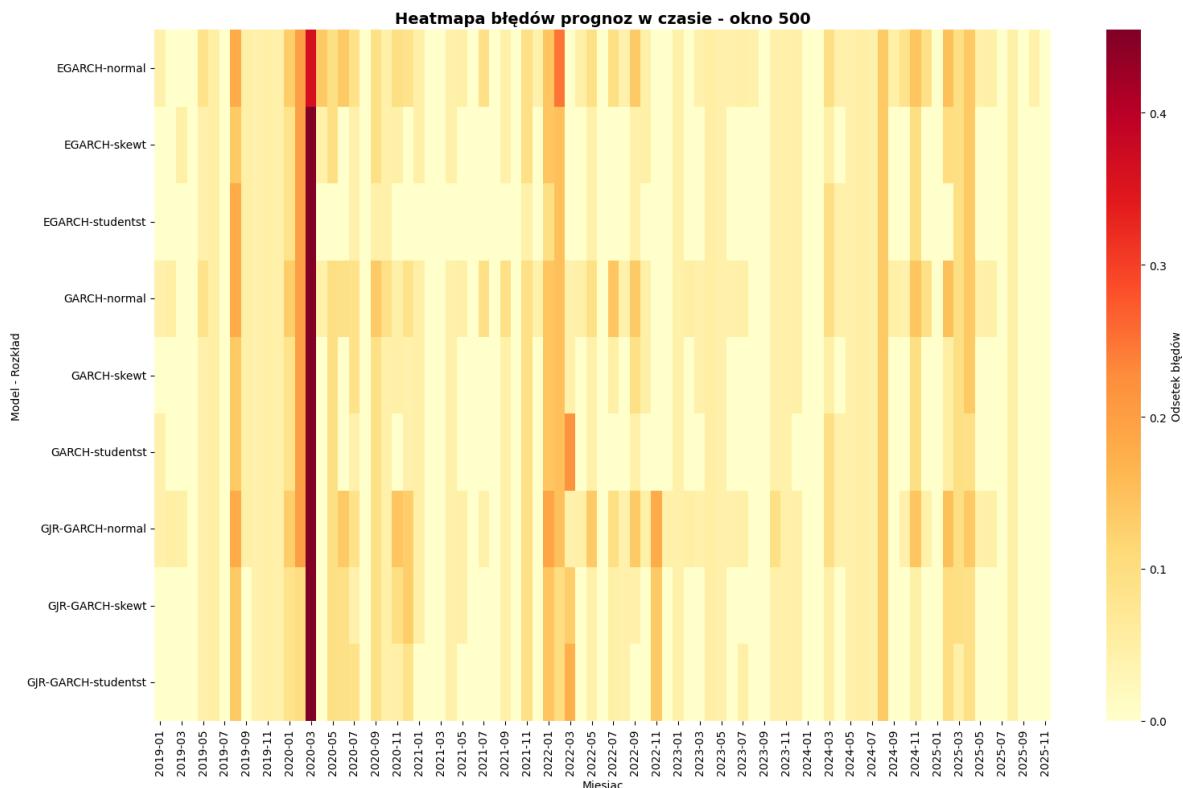
Rysunek 13. Rozkład trafień i nietrafień prognoz względem przedziału ufności - okno 500

Wykres słupkowy pokazuje, że wszystkie modele GARCH osiągają wysoką skuteczność trafień, jednak rozkłady t-Studenta i skośny t-Studenta wyraźnie redukują odsetek błędów względem modeli z rozkładem normalnym. W wariantach normalnych nietrafienia są częstsze. Choć EGARCH i GJR-GARCH uwzględniają asymetrię wariancji, błędy prognozy nadal wykazują przesunięcie w stronę wartości powyżej przedziału. W efekcie wykres potwierdza, że rozkłady t-Studenta najsukuczniej odwzorowują ryzyko i minimalizują błędy, ale nawet one nie eliminują całkowicie asymetrii wynikającej z obecności gwałtownych szoków rynkowych.



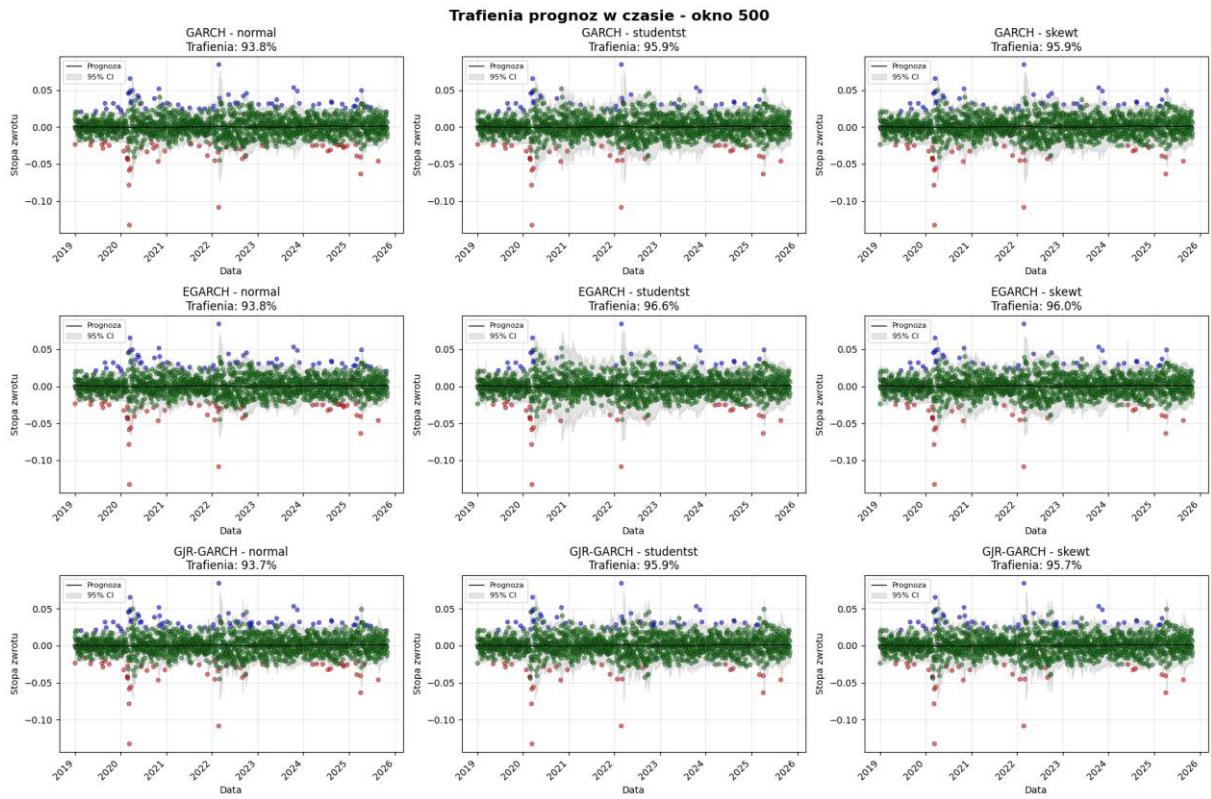
Rysunek 14. Zgodność z rozkładem warunkowym w czasie – okno 500

Wykres p wartości testu Kołmogorowa-Smirnowa jednoznacznie pokazuje, że kluczowym elementem poprawnego modelowania ryzyka w modelach GARCH jest właściwy wybór rozkładu błędów, a nie sama specyfikacja modelu. Rozkład normalny konsekwentnie nie przechodzi testu KS przez długie okresy, co potwierdza jego nieadekwatność wobec danych. Z kolei rozkład t-Studenta zapewnia najwyższą i najstabilniejszą zgodność: p-wartości dla wszystkich modeli pozostają przez cały okres nad poziomem istotności, co oznacza, że test KS nigdy nie odrzuca hipotezy zgodności. Rozkład skośny t-Studenta poprawia dopasowanie względem normalnego, lecz nadal wykazuje okresowe odrzucenia.



Rysunek 15. Mapa cieplna błędów prognoz w czasie – okno 500

Modele z rozkładem normalnym generują systematycznie wyższy poziom nietrafień, co widać w ciemniejszych odcieniach, natomiast modele oparte na rozkładach t-Studenta i skośnym t-Studenta charakteryzują się znacznie jaśniejszymi wierszami, wskazującymi na niższe, bardziej stabilne błędy. Szczególnie wyraźnym odstępstwem od poprawnych prognoz jest marzec 2020, kiedy ekstremalna zmienność związana z pandemią COVID-19 spowodowała drastyczny wzrost błędów we wszystkich modelach, sięgający nawet 40%. Podwyższony poziom błędów widoczny jest również w listopadzie 2022, szczególnie w modelach z rozkładem normalnym i skośnym t-Studenta. Poza okresami szoków modele wykorzystujące rozkład t-Studenta zapewniają dużą stabilność i utrzymują odsetek błędów blisko oczekiwanych 5%, podczas gdy modele normalne pozostają wyraźnie mniej precyzyjne.



Rysunek 16. Trafienia prognoz w czasie – okno 500

Modele GARCH z normalnymi resztami wyraźnie niedoszacowują zmienność w okresach gwałtownych wahań, takich jak marzec 2020 czy luty 2022, co skutkuje zbyt wąskimi przedziałami ufności i licznymi nietrafieniami. Z kolei modele z resztami o rozkładzie t-Studenta tworzą szersze i bardziej realistyczne przedziały, które utrzymują się dłużej po szokach rynkowych. Widać to szczególnie w 2020 roku, gdy szerokość przedziałów w tych modelach lepiej odzwierciedlała utrzymującą się wysoką zmienność. Pozostałe modele również poprawnie reagowały na wzrost zmienności, choć nie tak trwale jak modele z rozkładem t-Studenta.

6. Podsumowanie

Wszystkie testowane modele ARMA-GARCH wykazały wysoką skuteczność w konstruowaniu przedziałów ufności, zbliżoną do nominalnego poziomu 95%, co potwierdza ich przydatność w prognozowaniu zmienności stóp zwrotu WIG20.

Najlepsze wyniki uzyskano dla rozkładu t-Studenta i skośnego t-Studenta (skuteczność 94-96%), które lepiej radzą sobie z grubymi ogonami rozkładu niż rozkład normalny (93-94%). Model EGARCH jest najlepszy przy dłuższym oknie (500 obserwacji), osiągając najwyższą skuteczność 96,64%, podczas gdy przy krótszych oknach lepiej sprawdzają się klasyczne modele GARCH i GJR-GARCH.

Marzec 2020 (początek pandemii COVID-19) był najbardziej problematycznym okresem dla wszystkich modeli (55-68% trafień), co wskazuje na ograniczenia modeli w przewidywaniu ekstremalnych szoków rynkowych. Asymetria błędów jest niemal

zrównoważona, z niewielką przewagą błędów "powyżej przedziału", co sugeruje, że modele nie wykazują systematycznego przeszacowania lub niedoszacowania ryzyka.

Rozmiar okna ma umiarkowany wpływ na wyniki – większe okna (500 dni) zapewniają nieco lepszą jakość modeli, ale różnice są stosunkowo niewielkie.