

Uniwersytet Ekonomiczny w Krakowie

Dziedzina nauki: Nauki społeczne

Dyscyplina naukowa: Ekonomia i finanse

Jakub Michańków

**Metody uczenia głębokiego w prognozowaniu
finansowych szeregów czasowych**

Rozprawa doktorska

Promotor: prof. UEK dr hab. inż. Janusz Morajda

Promotor pomocniczy: dr Łukasz Kwiatkowski

Kraków, 2022

Spis treści

Wstęp	1
Wprowadzenie do problematyki pracy	1
Cele i hipotezy badawcze	2
Metodyka badań i źródła informacji	4
Streszczenie zawartości poszczególnych rozdziałów	6
1 Podstawowe podejścia do predykcji finansowych szeregów czasowych w badaniach naukowych i wdrożeniach praktycznych	9
1.1 Rynki finansowe i ich prognozowanie	9
1.1.1 Rynki i aktywa finansowe	9
1.1.2 Prognozowanie instrumentów finansowych	12
1.2 Stan badań w nurcie ekonometrii finansowej	14
1.3 Stan badań w zakresie sieci neuronowych	15
2 Ekonometryczne modelowanie i prognozowanie jednowymiarowych finansowych szeregów czasowych (z wykorzystaniem modeli klasy ARMA-GARCH)	19
2.1 Procesy stochastyczne w ekonometrycznej analizie szeregów czasowych	19
2.2 Empiryczne własności finansowych szeregów czasowych	20
2.2.1 Szeregi czasowe cen	20
2.2.2 Proste i logarytmiczne Stopy zwrotu – ich definicje i własności .	21
2.2.3 Fakty empiryczne	22
2.3 Zmienność finansowych szeregów czasowych	23
2.3.1 Zmienność historyczna i zrealizowana	24
2.3.2 Estymatory zmienności bazujące na zakresie cen	25
2.3.3 Transformacje estymatorów zmienności	27
2.3.4 Zmienność implikowana	29
2.4 Modele klasy ARMA-GARCH	29
2.4.1 Definicja procesu ARMA-GARCH	29

2.4.2	Procesy ARMA	30
2.4.3	Opis struktury GARCH	33
2.4.4	Uogólnienia i modyfikacje podstawowej specyfikacji procesu GARCH	37
2.4.5	Estymacja modeli klasy ARMA-GARCH	39
2.4.6	Alternatywne specyfikacje modeli zmienności	41
2.4.7	Wielowymiarowe uogólnienia modeli ARMA-GARCH i ARMA-SV-GARCH	42
2.5	Prognozowanie stóp zwrotu i ich zmienności w ramach modeli klasy ARMA-GARCH	43
2.5.1	Predykcja punktowa	43
2.5.2	Mierniki trafności prognoz punktowych	44
2.5.3	Ocena trafności predykcji punktowej stóp zwrotu w kontekście strategii inwestycyjnych	47
2.5.4	Wyznaczanie i ocena trafności prognoz probabilistycznych	49
2.6	Ocena zdolności predykcyjnych modeli w kontekście szacowania ryzyka kapitałowego	53
2.6.1	Wartość zagrożona - VaR	54
2.6.2	Oczekiwany niedobór - ES	56
2.6.3	Testy weryfikacyjne modeli ryzyka	58
3	Sieci neuronowe uczenia głębokiego – koncepcja i wykorzystanie w budowie modeli predykcyjnych	61
3.1	Wybrane architektury sieci neuronowych	61
3.1.1	Wielowarstwowe sieci MLP jako podstawowa forma sieci neuronowych	62
3.1.2	Rekurencyjne sieci LSTM i GRU	64
3.1.3	Sieci konwolucyjne	70
3.1.4	Inne typy sieci głębokich	75
3.2	Podstawowe zagadnienia związane z uczeniem sieci	78
3.2.1	Sposoby uczenia sieci neuronowych	78
3.2.2	Algorytmy oparte na metodzie gradientu	79
3.2.3	Generalizacja i regularyzacja	82
3.2.4	Podział zbioru danych w procesie uczenia i testowania sieci	86
3.2.5	Przygotowanie danych	89
3.3	Dostrajanie hiperparametrów sieci	90
3.3.1	Liczba warstw i neuronów sieci	90
3.3.2	Optymalizatory i inicjalizatory	91

3.3.3	Funkcje aktywacji	92
3.3.4	Pozostałe hiperparametry	94
3.3.5	Automatyzacja procesu dostrajania	95
3.4	Sieci uczenia głębokiego w prognozowaniu szeregów czasowych	96
4	Predykcyjne modele hybrydowe łączące metody ekonometryczne i techniki uczenia głębokiego	99
4.1	Przegląd i analiza możliwości predykcyjnych modeli hybrydowych	99
4.2	Propozycja nowych modeli	102
4.2.1	Hybrydowy model punktowych prognoz zmienności	103
4.2.2	Model prognoz probabilistycznych	107
5	Empiryczna ewaluacja modeli predykcyjnych	113
5.1	Charakterystyka analizowanych zbiorów danych	113
5.2	Budowa oprogramowania i algorytmów testujących	122
5.3	Optymalizacja hiperparametrów modeli sieciowych	124
5.4	Prognozowanie punktowe stóp zwrotu	127
5.4.1	Wyniki dla strategii wykorzystującej sieci MLP	130
5.4.2	Wyniki dla strategii wykorzystującej sieci CNN	132
5.4.3	Wyniki dla strategii wykorzystującej sieci LSTM	134
5.5	Prognozowanie zmienności i ryzyka	140
5.5.1	Ocena trafności prognoz punktowych zmienności	142
5.5.2	Ocena trafności prognoz ryzyka uzyskanych przy pomocy modeli hybrydowych	145
5.6	Prognozowanie rozkładów prawdopodobieństw	150
5.6.1	Ocena trafności prognoz probabilistycznych	150
5.6.2	Ocena trafności prognoz ryzyka uzyskanych przy pomocy modeli prognozowania probabilistycznego	152
6	Szczegółowa analiza porównawcza wyników empirycznych uzyskanych za pomocą wybranych modeli	157
6.1	Porównanie uzyskanych wyników	157
6.1.1	Modele punktowych prognoz stóp zwrotu	157
6.1.2	Modele hybrydowe punktowych prognoz zmienności	160
6.1.3	Modele prognozowania probabilistycznego	167
6.2	Podsumowanie wyników	170
6.2.1	Realizacja celów badawczych	171
6.2.2	Odniesienie do hipotez badawczych	171

6.3	Zalety i ograniczenia wynikające ze stosowania metod uczenia głębokiego	174
6.4	Proponowane kierunki dalszych prac badawczych	176
	Zakończenie	179
	Aneks	181
	Bibliografia	203
	Spis rysunków	227
	Spis tabel	229

Wstęp

Wprowadzenie do problematyki pracy

W XXI. wieku, zwłaszcza w ostatnich latach, można zaobserwować znaczący wzrost popularności i zastosowań metod sztucznej inteligencji w szeroko pojętej analizie danych. Wiele modeli, powstałych jako mające ogólne zastosowanie w różnych dziedzinach sztucznej inteligencji, daje się wykorzystywać także jako narzędzia wspierające przy modelowaniu i prognozowaniu zmian cen aktywów finansowych. Nie są to jednak narzędzia, które pozwoliłyby niedoświadczonym użytkownikom przewidzieć notowania giełdowe czy kursy walut. Natomiast mają one na celu służyć jako pomoc dla analityków danych i inwestorów w podejmowaniu decyzji inwestycyjnych. Dotyczy to także ekonometrii finansowej, gdzie wiodącym obszarem zainteresowań badaczy jest modelowanie i prognozowanie szeregów czasowych cen i stóp zwrotu różnorodnych instrumentów finansowych.

Głównymi obszarami wykorzystania sztucznej inteligencji w badaniach rynków finansowych, i zarazem jednymi z wiodących zagadnień ekonometrii finansowej są: prognozowanie szeregów czasowych, zarządzanie ryzykiem oraz optymalizacja portfela inwestycyjnego. W tej pracy uwaga zostanie skoncentrowana przede wszystkim na prognozowaniu stóp zwrotu, zmienności oraz szacowaniu ryzyka. Pomimo licznych prac naukowych na ten temat, problem prognozowania wciąż pozostaje otwarty i aktualny. Analiza i optymalizacja istniejących modeli może przyczynić się zarówno do poprawy jakości istniejących już narzędzi handlu aktywami, jak i budowy nowych. Dodatkowe korzyści może przynieść także stworzenie modeli hybrydowych, opartych na sieciach neuronowych w połączeniu z modelami ekonometrii finansowej oraz metodami prognozowania probabilistycznego.

Na gruncie metod sztucznej inteligencji jednym z najbardziej dynamicznie rozwijających się w ciągu ostatnich lat obszarów jest uczenie głębokie (ang. *deep learning*). Stanowi ono pewne rozwinięcie klasycznych już metod uczenia maszynowego, a w szczególności sieci neuronowych. O uczeniu głębokim można mówić też jako o pewnej "nowej epoce" w rozwoju prac nad metodami sztucznej

inteligencji, w której, dzięki dostępności coraz szybszego sprzętu komputerowego i wzrostowi mocy obliczeniowej, nastąpiło znaczne przyspieszenie postępu badań nad nowymi rozwiązaniami i narzędziami w obszarze sieci neuronowych i uczenia maszynowego. Rozwój ten, z kolei, spowodował znaczny wzrost popularności tych metod zarówno wśród badaczy, jak i wśród końcowych użytkowników tych rozwiązań.

Badania dotyczące wykorzystania sieci uczenia głębokiego w prognozowaniu instrumentów finansowych zaczęły pojawiać się dopiero w ciągu ostatnich kilku lat i wciąż pozostawiają wiele miejsca na poszukiwanie nowych rozwiązań. Niewiele jest też badań łączących metody oparte na sieciach głębokich z metodami ekonometrycznymi. W nurcie tym mieści się także niniejsza rozprawa.

W pracy szczególny nacisk położony zostanie na synergię metod uczenia głębokiego oraz metod ekonometrycznych, która możliwa jest do osiągnięcia poprzez rozwiązania hybrydowe, zastosowane przede wszystkim do prognozowania zmienności i ryzyka, a także w kontekście prognozowania probabilistycznego. Proponowane w kolejnych rozdziałach rozwiązania wykorzystują aspekty obu podejść, łącząc modele uczenia głębokiego z modelami ekonometrycznymi w celu uzyskania efektywniejszych narzędzi. Zdaniem autora, takie podejście przyczyni się korzystniej do rozwoju badań, niż jedynie próba wskazania lepszych metod, których jednoznaczne określenie często jest niemożliwe.

Cele i hipotezy badawcze

Głównym celem pracy jest opracowanie modeli predykcyjnych, opartych na metodach uczenia głębokiego, oraz wskazanie możliwości i potencjalnych kierunków ich wykorzystania, a także przeprowadzenie analizy ich efektywności, rozumianej jako trafność predykcji w prognozowaniu szeregów czasowych (logarytmicznych) stóp zwrotu aktywów finansowych oraz ich zmienności.

Z kolei do celów cząstkowych dysertacji zaliczyć można:

1. Zbadanie efektywności modeli predykcyjnych opartych na uczeniu głębokim w prognozowaniu punktowym stóp zwrotu oraz zmienności.
2. Sprawdzenie jakości prognoz punktowych stóp zwrotu w kontekście strategii inwestycyjnych.
3. Porównanie wyników strategii inwestycyjnych w zależności od poziomu rozwinięcia rynków finansowych wybranych krajów.

4. Propozycja nowych rozwiązań w postaci modeli hybrydowych, łączących metody uczenia głębokiego oraz narzędzia ekonometryczne, w celu prognozowania zmienności instrumentów finansowych.
5. Wykorzystanie prognoz zmienności uzyskanych przez badane modele do szacowania ryzyka kapitałowego.
6. Propozycja nowych rozwiązań umożliwiających wykorzystanie głębokich sieci neuronowych w prognozowaniu probabilistycznym.
7. Porównanie modeli opartych na sztucznej inteligencji z metodami ekonometrycznymi opartymi na modelach klasy ARMA-GARCH pod względem trafności predykcji punktowej i probabilistycznej.
8. Stworzenie autorskich kodów umożliwiających przeprowadzenie badań z wykorzystaniem sieci uczenia głębokiego w zakresie powyższych celów badawczych.

Główna hipoteza badawcza zakłada, że modele predykcyjne oparte na metodach uczenia głębokiego można efektywnie wykorzystywać do prognozowania finansowych szeregów czasowych w postaci logarytmicznych stóp zwrotu, oraz ich zmienności. Natomiast hipotezy szczegółowe sformułowane zostały następująco:

1. Modele oparte na uczeniu głębokim generują lepsze efekty predykcji w porównaniu z modelami wykorzystującymi klasyczne metody sztucznej inteligencji w kontekście punktowych prognoz stóp zwrotu.
2. Prognozy punktowe stóp zwrotu uzyskiwane przez modele oparte na uczeniu głębokim można wykorzystać do budowy skutecznych strategii inwestycyjnych.
3. Prognozy stóp zwrotu uzyskane na podstawie danych z rynków wschodzących dają lepsze wyniki w kontekście strategii inwestycyjnych od prognoz z rynków rozwiniętych.
4. Połączenie metod ekonometrycznych z metodami uczenia głębokiego w ramach modeli hybrydowych przyczynia się do poprawy efektywności prognoz zmienności.
5. Prognozy zmienności uzyskane w ramach modeli hybrydowych przyczyniają się do poprawy prognoz ryzyka kapitałowego.
6. Sieci neuronowe uczenia głębokiego można wykorzystać jako narzędzie prognozowania rozkładów prawdopodobieństw.

7. Prognozy probabilistyczne uzyskane za pomocą sieci uczenia głębokiego przynoszą lepsze rezultaty w porównaniu z modelami klasy GARCH.

Metodyka badań

Punktem wyjścia w zakresie metodycznym niniejszej rozprawy jest przegląd literatury dotyczącej badań naukowych, przeprowadzonych w ostatnich latach, analizujących strukturę oraz skuteczność modeli opartych na klasycznych metodach sztucznej inteligencji oraz metodach ekonometrycznych. Przedstawia on opis obecnego stanu wiedzy, ze wskazaniem problemów, które do tej pory nie zostały zbadane lub pozostają w sferze zagadnień otwartych. W przeglądzie skupiono się na najnowszych publikacjach dotyczących badań naukowych w zakresie wykorzystania sieci uczenia głębokiego, modeli klasy GARCH oraz prognozowania probabilistycznego w analizie finansowych szeregów czasowych (konkretnie stóp zwrotu i ich zmienności). W dalszych podrozdziałach prezentowana uprzednio literatura została poszerzona także o badania bardziej specyficzne dla danego tematu, wymagające często wcześniejszego wprowadzenia teoretycznego.

W zakresie metod ekonometrycznych praca skupia się na modelowaniu jednowymiarowych szeregów czasowych (stóp zwrotu oraz ich zmienności) z wykorzystaniem modeli klasy GARCH (ARMA-GARCH), szacowanych za pomocą metody największej wiarygodności (MNW). Głównymi specyfikacjami klasy GARCH rozważanymi w tym opracowaniu będą podklasy EGARCH, GJR-GARCH i APARCH. Wskazane są także alternatywne modele zmienności: modele SV oraz modele hybrydowe (M)SV-(M)GARCH.

Głównym narzędziem wykorzystywanym w tej pracy są sieci neuronowe uczenia głębokiego. Do budowy modeli w tym zakresie stosowane są dwa typy sieci głębokich: sieci rekurencyjne LSTM (ang. *Long Short Term Memory*) oraz sieci konwolucyjne (ang. *Convolutional Neural Networks* - CNN). W celu zapewnienia prawidłowego działania sieci i optymalizacji ich hiperparametrów zastosowany jest także szeroki zakres narzędzi pomocniczych, szerzej opisanych w rozdziale 3. Jako dodatkowy poziom porównania metod uczenia głębokiego z klasycznymi sieciami neuronowymi wykorzystane są wielowarstwowe sieci perceptronowe (ang. *Multi-Layer Perceptron* - MLP).

Metody ekonometryczne połączone są z metodami uczenia głębokiego w modele hybrydowe, stanowiące autorskie rozwiązanie mające na celu poprawę wyników uzyskiwanych przez indywidualne modele składowe przynależące w całości bądź do grupy narzędzi ekonometrii finansowej, bądź sztucznej inteligencji.

Przedstawione powyżej metody wykorzystane są przede wszystkim do prognozowania finansowych szeregów czasowych. Przedmiotem prognoz punktowych są logarytmiczne stopy zwrotu wybranych instrumentów finansowych - i przede wszystkim - ich zmienność. Obok prognoz punktowych, badane narzędzia wykorzystane są także do uzyskania prognoz probabilistycznych, ze względu na wzrost znaczenia takiego podejścia w ostatnim czasie. W szczególności, do prognozowania probabilistycznego dostosowane będą modele bazujące na sieciach uczenia głębokiego.

Do oceny trafności predykcji stosowane jest podejście wieloaspektowe. Prognozy punktowe ocenione są przy pomocy typowych mierników błędu, takich jak błąd średniokwadratowy. Jako zmienność "zaobserwowana" rozumiana jest zmienność oszacowana przy pomocy estymatorów opartych na zakresie cen. Dodatkowym poziomem oceny prognoz punktowych stóp zwrotu są strategie inwestycyjne, opisane przy pomocy krzywych kapitałowych oraz dodatkowych mierników. W końcu, do oceny trafności probabilistycznych stosowane są kryteria specyficzne dla tej problematyki, w szczególności LPS (ang. *Log Predictive Score*), CRPS (ang. *Continous Ranked Probability Score*) oraz PIT (ang. *Probability Integral Transform*).

Uzyskane prognozy ocenione są także w kontekście ryzyka kapitałowego. Do tego celu wykorzystane są prognozy wartości zagrożonej (ang. *Value at Risk* - VaR) oraz oczekiwanego niedoboru (ang. *Expected Shortfall* - ES). Ocena prognoz VaR i ES przeprowadzona zostanie zarówno w kontekście punktowych prognoz zmienności jak i modeli prognoz probabilistycznych.

Do oceny trafności prognoz modeli wykorzystywane są symulacje komputerowe, z których pomocą badana jest efektywność wybranych modeli predykcyjnych opartych na uczeniu głębokim i metodach ekonometrycznych, z wykorzystaniem algorytmów testowania wstecznego.

Do implementacji modeli i tworzenia algorytmów wykorzystywane są narzędzia programistyczne i biblioteki dostępne w środowiskach Python oraz R, wspomagające uczenie głębokie i metody ekonometryczne. Jako podstawowe biblioteki do przetwarzania i wizualizacji danych używane są pakiety Pandas (McKinney [2010]), Numpy (Harris i in. [2020]) oraz Matplotlib (Hunter [2007]). Spośród narzędzi wspomagających klasyczne uczenie maszynowe oraz w ramach uczenia głębokiego wykorzystane zostaną biblioteki TensorFlow (Abadi i in. [2016]) oraz Keras (Chollet i in. [2015]). Do budowy modeli klasy ARMA-GARCH, oraz do testowania prognozowanych poziomów wartości zagrożonej i oczekiwanego niedoboru wykorzystany będzie pakiet rugarch (Ghalanos [2020]).

Badania prowadzone są z wykorzystaniem danych w postaci logarytmicznych stóp zwrotu. Dane obejmują sześć indeksów giełdowych: S&P 500, DAX, NIKKEI 225,

WIG, KOSPI oraz BOVESPA, reprezentujących kraje na różnych poziomach rozwoju gospodarczego (rynkı rozwinęte oraz wschodzące) i z różnych rejonów geograficznych.

Wyniki empiryczne poddane są analizie porównawczej, polegającej na zestawieniu najlepszych specyfikacji modelowych, a także wskazaniu zalet i ograniczeń wynikających ze stosowania metod opartych na uczeniu głębokim. W odniesieniu do prognoz punktowych stóp zwrotu i zbudowanych na ich podstawie strategii inwestycyjnych porównane zostaną klasyczne sieci neuronowe MLP, oraz sieci uczenia głębokiego LSTM i CNN. W kontekście punktowych prognoz zmienności i szacowania ryzyka kapitałowego porównane zostaną poszczególne specyfikacje modeli klasy GARCH z wybranymi rozkładami warunkowymi, oraz modele hybrydowe, łączące modele GARCH z sieciami uczenia głębokiego. Ocenie w kontekście szacowania ryzyka kapitałowego poddane zostaną również modele prognoz probabilistycznych.

Streszczenie zawartości poszczególnych rozdziałów

W pierwszym rozdziale, w ramach wprowadzenia, omówiono podstawowe kwestie związane z rynkami finansowymi oraz ich prognozowaniem. Przedstawiono definicje oraz strukturę podziału rynków finansowych oraz wskazano podstawowe narzędzia służące do ich prognozowania. W tej części pracy przedstawiono także najważniejsze publikacje naukowe, z zakresu ekonometrii finansowej oraz sieci neuronowych, dotyczące prognozowania instrumentów finansowych.

W rozdziale drugim szczegółowo przedstawiono narzędzia ekonometryczne stosowane do modelowania i prognozowania instrumentów finansowych. Omówiono empiryczne własności finansowych szeregów czasowych, oraz najważniejsze modele służące do analizy i prognozowania stóp zwrotu oraz zmienności, ze szczególnym naciskiem położonym na modele klasy ARMA-GARCH. Przedstawiono podstawy konstrukcji tych modeli, metody estymacji, oraz ich możliwości predykcyjne, zarówno w kontekście prognoz punktowych jak i probabilistycznych. Wskazano także sposoby oceny trafności predykcji oraz możliwości ich wykorzystania przy budowie strategii inwestycyjnych oraz w szacowaniu ryzyka kapitałowego.

Trzeci rozdział dotyczy koncepcji związanych z sieciami neuronowymi uczenia głębokiego oraz ich wykorzystania jako modeli predykcyjnych. Przedstawiono w nim wybrane architektury sieci neuronowych, w szczególności sieci uczenia głębokiego, kwestie związane z uczeniem sieci, dotyczące stosowanych w tym procesie algorytmów oraz odpowiedniego przygotowania danych, a także problematykę dotyczącą dostrajania hiperparametrów sieci. Omówiono także wybrane publikacje naukowe dotyczące wykorzystania sieci uczenia głębokiego w prognozowaniu finansowych

szeregów czasowych.

W rozdziale czwartym zaprezentowano autorskie propozycje modeli wykorzystujące elementy sieci uczenia głębokiego oraz metod ekonometrycznych. W pierwszej kolejności wskazano dotychczasowe propozycje modeli hybrydowych, wykorzystujące wspomniane metody, proponowane w literaturze przedmiotu. Biorąc pod uwagę zalety i niedociągnięcia tych rozwiązań, zaproponowano autorską architekturę hybrydowego modelu punktowych prognoz zmienności, ARMA-GARCH-LSTM. Przedstawiono także propozycje wykorzystania sieci neuronowych uczenia głębokiego w prognozowaniu probabilistycznym, poprzez prognozowanie parametrów rozkładów prawdopodobieństwa finansowych szeregów czasowych.

Rozdział piąty dotyczy empirycznej ewaluacji omawianych specyfikacji modelowych. W ramach tego rozdziału omówiono statystyczne charakterystyki badanych zbiorów danych, narzędzia programistyczne stosowane do wdrożenia analizowanych narzędzi, a także kwestie związane z doбором i optymalizacją hiperparametrów modeli sieciowych. Zaprezentowano tu wyniki empiryczne uzyskane przez modele wykorzystujące sieci neuronowe w kontekście prognoz punktowych stóp zwrotu oraz zbudowanych z ich wykorzystaniem strategii inwestycyjnych. W dalszej kolejności omówiono wyniki uzyskane z wykorzystaniem modeli punktowych prognoz zmienności: poszczególnych specyfikacji ekonometrycznych modeli ARMA-GARCH oraz modeli hybrydowych ARMA-GARCH-LSTM. Wyniki oceniono wykorzystując klasyczne mierniki prognoz punktowych oraz w kontekście szacowania ryzyka kapitałowego. W kolejnej części zaprezentowano wyniki uzyskane przez modele prognoz probabilistycznych, które oceniono pod kątem specyficznych miar i kryteriów oceny prognoz probabilistycznych, oraz, podobnie jak w przypadku modeli punktowych prognoz zmienności, w odniesieniu do ryzyka kapitałowego.

W rozdziale szóstym przeprowadzono analizę porównawczą wyników uzyskanych przez wybrane najlepsze specyfikacje modelowe. Porównano tutaj modele punktowych prognoz stóp zwrotu, modele hybrydowe punktowych prognoz zmienności oraz sieciowe modele prognozowania probabilistycznego. Omówiono także sposoby realizacji założonych w pracy celów badawczych oraz poszczególne hipotezy badawcze. Przedstawiono tu także zalety i wady wynikające z zastosowania metod uczenia głębokiego oraz propozycje dalszych prac badawczych w tym zakresie.

Rozdział 1

Podstawowe podejścia do predykcji finansowych szeregów czasowych w badaniach naukowych i wdrożeniach praktycznych

W pierwszej kolejności przedstawione zostaną podstawowe pojęcia i definicje związane z rynkami finansowymi, a także różne podejścia stosowane w ich prognozowaniu. Zaprezentowany zostanie także przegląd podstawowej literatury z zakresu ekonometrii finansowej i sieci neuronowych. Informacje zawarte w tym rozdziale mają na celu zaznajomienie czytelnika z przedmiotem niniejszej rozprawy, oraz zwrócenie uwagi na popularność problematyki związanej z prognozowaniem instrumentów finansowych,

1.1 Rynki finansowe i ich prognozowanie

1.1.1 Rynki i aktywa finansowe

Głównym przedmiotem badań w niniejszej pracy, a zarazem źródłem danych empirycznych są wybrane instrumenty finansowe. Jako instrument finansowych rozumieć można kontrakt regulujący zobowiązania finansowe pomiędzy stronami, powodujące powstanie aktywów finansowych u jednej ze stron i zobowiązania finansowego lub instrumentu kapitałowego u drugiej (za Kudła [2011], oraz art. 3 ust. 1 pkt 23 ustawy o rachunkowości (Dz.U. z 2021 r. poz. 217)). Transakcje kupna i sprzedaży różnych form kapitału pieniężnego, w oparciu o różne instrumenty finansowe, zawierane są na rynkach finansowych. Te z kolei podzielić można na

następujące segmenty (zob. Kudła [2011]):

- rynek pieniężny, na którym zawierane są transakcje krótkoterminowe (do jednego roku) o niskim poziomie ryzyka i dużej płynności,
- rynek kapitałowy, na którym zawierane są transakcje średnio- i długoterminowe, dotyczące różnych instrumentów finansowych emitowanych i nabywanych przez podmioty, takich jak akcje i obligacje,
- rynek walutowy, dotyczący obrotu walutami obcymi, na którym stosunek ceny między walutami odzwierciedla kurs walutowy,
- rynek depozytowo-kredytowy, dotyczący transakcji między bankami a ich klientami, charakteryzujący się dużym ryzykiem i małą płynnością,
- rynek terminowy (instrumentów pochodnych), na którym zawierane są kontrakty terminowe, czyli umowy na operacje finansowe z terminem wykonania w przyszłości.

W obrębie przedstawionych powyżej segmentów rynku finansowego, wyróżnia się dalsze podziały i struktury. Przykładowo, rynek kapitałowy, który jest w tej pracy głównym przedmiotem zainteresowania, dodatkowo podzielić można (za Kachniewski i in. [2008]) na rynek pierwotny i wtórny (z uwzględnieniem przepływów finansowych), rynek giełdowy i pozagiełdowy (w zależności od stopnia zorganizowania rynku), oraz na rynek regulowany i nieregulowany (w zależności od stopnia regulacji). Na rynku pierwotnym dokonywane są transakcje emitowanych papierów wartościowych (np. akcji spółek giełdowych) bezpośrednio między spółkami a inwestorami, zanim trafią na rynek wtórny (Giełdę Papierów Wartościowych w Warszawie - GPW), na którym transakcje zawierane są między inwestorami. GPW należy wg podanego podziału do rynków giełdowych i regulowanych. Przykładami rynków nieregulowanych są alternatywne systemy obrotu (ASO), z pośród których wymienić można działające obecnie w Polsce rynek NewConnect oraz Catalyst¹ (zob. Kachniewski i in. [2008] oraz Kozdra [2015]).

Jak wspomiano wcześniej, przedmiotem transakcji na rynkach finansowych są różnego rodzaju instrumenty finansowe. Ich zakres jest bardzo szeroki, dlatego poniżej ograniczymy się do przedstawienia kilku typów instrumentów, które często wybierane są jako przedmiot badań finansowych.

¹Rynek Catalyst działa w kilku wariantach, jako regulowany rynek ASO dostępny dla klientów detalicznych, prowadzony przez GPW, oraz rynek pozagiełdowy, dostępny tylko dla klientów hurtowych, prowadzony przez BondSpot S.A

Jednym z podstawowych instrumentów na rynku kapitałowym, a jednocześnie jednym z najpopularniejszych aktywów finansowych są akcje spółek giełdowych. Jako akcje rozumieć można (za Kudła [2011]) „*tytuł własności do ułamkowej części spółki akcyjnej, lub komandytowo-akcyjnej, wyrażający wniesiony wkład i prawa członkowskie, a jednocześnie papier wartościowy (imienny lub na okaziciela), który może podlegać obrotowi*”. Ceny akcji na rynku wtórnym kształtowane są poprzez popyt i podaż wśród inwestorów. Oprócz zleceń kupna i sprzedaży akcji, dla dużych spółek (m.in. o wartości rynkowej powyżej 100 mln euro) możliwa jest także tzw. sprzedaż krótka — sprzedaż akcji pożyczonych z przeznaczeniem odkupu w przyszłości (zob. Kudła [2011]).

Zmiany kursów koszyka akcji odzwierciedla indeks giełdowy. Jest on statystycznym wskaźnikiem, podsumowującym zmiany kursów akcji wybranych spółek podczas sesji giełdowej (zob. Kachniewski i in. [2008]). Indeksy giełdowe mogą dostarczać informacji o konkretnych segmentach rynku, zależnych przykładowo od wielkości spółek lub wybranej branży, przez co mogą służyć jako punkt odniesienia przy ocenie efektywności portfeli (lub funduszy) inwestycyjnych. Bezpośredni handel indeksami nie jest możliwy, odbywa się on z wykorzystaniem instrumentów pochodnych, takich jak kontrakty terminowe, opcje lub certyfikaty funduszy inwestycyjnych typu ETF (ang. *Exchange-Traded Fund*), dla których indeksy stanowią instrument bazowy.

Do najpopularniejszych indeksów giełdowych (zob. Bodie i in. [2018]) należą indeksy DJIA (ang. *Dow Jones Industrial Average*), S&P 500 (ang. *Standard and Poor's 500*), w skład których wchodzi największe spółki notowane na Nowojorskiej Giełdzie Papierów Wartościowych (ang. *New York Stock Exchange - NYSE*) oraz NASDAQ (ang. *National Association of Securities Dealers Automated Quotations*) czy indeks FTSE (ang. *Financial Times Stock Exchange*) złożony ze spółek notowanych na Giełdzie Papierów Wartościowych w Londynie (ang. *London Stock Exchange*). W Polsce, od 1991, roku na Giełdzie Papierów Wartościowych notowany jest indeks WIG (Warszawski Indeks Giełdowy), będący najstarszym indeksem w kraju.

Sposoby obliczania wartości indeksów giełdowych (oraz kryteria udziału spółek) są różne, w zależności od danego kraju i giełdy. Przykładowo, dla indeksu WIG notowanego na GPW stosowana jest następująca formuła (Kachniewski i in. [2008]):

$$WIG = \frac{\sum P_i \cdot S_i}{\sum (P_0 \cdot S_0) \cdot K_t} \cdot 1000, \quad (1.1)$$

gdzie S_i to pakiet uczestnika indeksu i na danej sesji, P_i to kurs uczestnika indeksu i

na danej sesji, P_0 oraz S_0 to odpowiednio pakiet i kurs uczestnika w dniu bazowym, zaś K_t jest współczynnikiem korygującym indeksu na danej sesji.

Zakres instrumentów, które podlegają handlowi na rynkach finansowych jest bardzo szeroki i nie sposób je tu wszystkie opisać. Do najważniejszych grup aktywów finansowych należą także (zob Kudła [2011]):

- kursy walutowe - rozumiane mogą być jako instrumenty finansowe, odzwierciedlające stosunek ceny między walutami, wynikający z popytu i podaży na rynku walutowym,
- derywatywy - instrumenty pochodne, których wycena zależy od ceny instrumentów bazowych; do najważniejszych przykładów należą tutaj kontrakty terminowe oraz opcje,
- towary handlowe (ang. *commodities*) - produkty występujące w naturalny sposób w ziemi lub uprawiane w ramach rolnictwa, takie jak ropa naftowa, gaz ziemny, pszenica; handel towarami odbywa się najczęściej poprzez instrumenty pochodne,
- kryptowaluty - rozproszone systemy wirtualnych walut, bazujące na kryptografii, cieszące się w ostatnich latach rosnącą popularnością wśród inwestorów detalicznych; jako najważniejsze wymienić można kryptowaluty Bitcoin oraz Ethereum,
- inne instrumenty finansowe, w tym obligacje skarbowe, bony skarbowe.

Powyższa klasyfikacja nie stanowi pełnej klasyfikacji wszystkich typów i podtypów instrumentów finansowych. Szersze informacje na ten temat przedstawione zostały w opracowaniach Kudła [2011], Bodie i in. [2018] oraz Babu [2006].

1.1.2 Prognozowanie instrumentów finansowych

Modelowanie instrumentów finansowych leży u podstaw finansów empirycznych. Obok oczywistych korzyści finansowych, odpowiednie prognozowanie wartości stóp zwrotu oraz zmienności jest kluczowe w konstruowaniu narzędzi wyceny opcji, szacowaniu ryzyka, tworzeniu strategii hedgingowych czy też w obszarze inżynierii finansowej.

Pomimo powszechnego przekonania, że szeregi czasowe cen instrumentów finansowych mają charakter losowy (zob. Cootner [1964], Fama [1965b]), istnieją pewne charakterystyczne własności wspomnianych szeregów stóp zwrotu oraz ich zmienności, wskazujące na to, że modelowanie szeregów w tej postaci jest uzasadnione (zob. punkt 2.2.3).

Wśród rozmaitych podejść stosowanych w modelowaniu i prognozowaniu finansowych szeregów czasowych, jako najważniejsze należy wskazać metody ekonometrii finansowej. Narzędzia rozwijane w ramach tej dyscypliny skupiają się na analizie danych finansowych z wykorzystaniem metod ilościowych, w szczególności z wykorzystaniem różnego typu modeli zmienności. Rozwój ekonometrii finansowej zyskał na popularności na początku lat siedemdziesiątych ubiegłego stulecia, kiedy pojawiły się pierwsze modele wyceny opcji (Black i Scholes [1973]). Do dalszego rozwoju dyscypliny przyczyniły się prace Mandelbrot [1963], Fama [1965a], Engle [1982] oraz Bollerslev [1986] dotyczące własności oraz modelowania szeregów stóp zwrotu i ich zmienności. Tematyce tej poświęcona jest znaczna część tej pracy (w szczególności rozdziały 2, 5 oraz 6), zaś szczegółowe omówienie narzędzi ekonometrycznych znaleźć można w licznych opracowaniach literaturowych, których przegląd przedstawiony jest w podrozdziale 1.2.

W ciągu ostatniej dekady, na popularności zyskały metody bazujące na wykorzystaniu sieci neuronowych i uczenia maszynowego. Co prawda możliwości prognostyczne sieci neuronowych badane były już wcześniej, jednak efektywność zbudowanych z ich wykorzystaniem modeli była ograniczona, głównie ze względu na znacznie niższą niż obecnie moc obliczeniową². Wraz z dynamicznym rozwojem dostępnego na rynku sprzętu komputerowego i dostępności danych, rezultaty badań wykorzystujących te narzędzia uległy znacznej poprawie, głównie za zasługą sieci neuronowych uczenia głębokiego, stosowanych w obszarze rozpoznawania obrazów i przetwarzania języka naturalnego. Narzędzia te z powodzeniem zaczęto stosować także w analizie i prognozowaniu danych finansowych, co jest też głównym tematem tej pracy. Szeroki zakres literatury dotyczącej prognozowania z wykorzystaniem metod bazujących na sztucznej inteligencji przedstawiony został w podrozdziałach 1.3, 3.4 oraz 4.1.

Obok przedstawionych powyżej podejść bazujących na ekonometrii finansowej i sztucznej inteligencji, problematyka związana z prognozowaniem instrumentów finansowych jest przedmiotem badań także w innych dziedzinach, takich jak:

- matematyka finansowa,
- inżynieria finansowa,
- analiza techniczna,
- teoria chaosu,

²W zasadzie argument ten można zastosować także w odniesieniu do innych prezentowanych tutaj podejść.

- analiza falkowa,
- modele logiki rozmytej,

Pomimo, że podejścia te nie będą animalizowane w tej pracy, wskazać w tym miejscu można wybrane publikacje z tych dziedzin, do których należą: Kodogiannis i Lolis [2002], Ravi i in. [2017], Farias Nazário i in. [2017] czy też Zolfaghari i Gholami [2021]. Opublikowane zostało także wiele artykułów przeglądowych, prezentujących różne podejścia do analizy szeregów czasowych, z pośród których wymienić można prace: Mallikarjuna i Rao [2019], Majid [2018] oraz Timmermann [2018].

Wielość i różnorodność przedstawionych powyżej podejść świadczy o wadze i popularności problematyki prognozowania instrumentów finansowych. Wciąż prowadzone są nowe badania nad rozwojem metod prognostycznych, pozwalających na dalszą poprawę uzyskiwanych wyników, o czym świadczyć może szeroki zakres literatury przedstawiony w kolejnych podrozdziałach.

Zaznaczyć należy także, że obszar zastosowania wymienionych metod, w tym także tych opracowywanych w ramach tej pracy, nie musi ograniczać się do rynków finansowych. Proponowane modele można stosować także w innych dziedzinach, przykładowo do prognozowania cen i zużycia cen energii elektrycznej.

1.2 Stan badań w nurcie ekonometrii finansowej

W poniższym podrozdziale przedstawiony został przegląd podstawowej (zarówno zagranicznej, jak i krajowej) literatury dotyczącej wykorzystania metod ekonometrii finansowej w modelowaniu stóp zwrotu i zmienności instrumentów finansowych.

Największą popularnością wśród narzędzi wykorzystywanych do szacowania i predykcji zmienności są modele autoregresyjnej warunkowej heteroskedastyczności (ang. *AutoRegressive Conditional Heteroskedasticity* - ARCH), zaproponowane przez Engla [1982]. Podstawowa postać modelu rozszerzona została następnie przez Bollersleva [1986] oraz Taylora [1986], do modeli GARCH (ang. *Generalized Autoregressive Conditional Heteroskedasticity*), które znaczenie częściej stosowane są w badaniach rynków finansowych.³ Modele tej klasy zostały szerzej opisane w punkcie 2.4.3.

Szeroki i dokładny opis metod wykorzystywanych w badaniu finansowych szeregów czasowych, zarówno od strony teoretycznej, jak i praktycznej, przedstawił w swojej pracy Tsay [2010]. Natomiast w literaturze polskojęzycznej tymi

³Engle za swoją pracę nad metodami analizowania szeregów czasowych w 2003 roku otrzymał nagrodę im. Alfreda Nobla.

zagadnieniami zajmowali się Doman i Doman [2004, 2009] oraz Fiszeder [2009], który w swojej rozprawie habilitacyjnej szczegółowo przedstawia różne parametryzacje, uogólnienia i modyfikacje modeli klasy GARCH (jedno- i wielowymiarowe) oraz ich zastosowanie w empirycznych badaniach finansowych, w tym także w prognozowaniu procesów finansowych. Porównanie możliwości predykcyjnych różnych specyfikacji modeli GARCH w prognozowaniu zmienności instrumentów finansowych znaleźć można także między innymi w pracach: Trück i Liang [2012], Nilsson [2017], Costa [2017], Tong i in. [2020], oraz Tay i in. [2020].

Na gruncie ekonometrii finansowej silnie rozwijane są także podejścia bazujące na wykorzystaniu modeli SV (ang. *Stochastic Volatility*) (zob. Clark [1973]), również w połączeniu z modelami GARCH, także w postaci wielowymiarowej. W literaturze polskojęzycznej problematyka ta szczegółowo opisana została w pracach Doman i Doman [2004], [2009], Osiewalski i in. [2004], Pajor ([2003, 2010]) oraz Osiewalski i Pajor [2018].

Prognozy zmienności mają bardzo istotne znaczenie w większości problemów z obszaru finansów empirycznych, m. in. w kontekście szacowania ryzyka kapitałowego, mierzonego przy pomocy wartości zagrożonej oraz oczekiwanego niedoboru. Modele klasy GARCH w szacowaniu wartości zagrożonej (VaR) wykorzystywali w swoich badaniach m.in. Tong i in. [2020], Piontek [2001], Piontek i Papla [2005], Fiszeder [2007], [2009], Aloui i Mabrouk [2010], Ardia i Hoogerheide [2014], Będowska-Sójka [2015], Slim i in. [2017], Bams i in. [2017], Laporta i in. [2018], Zhang i in. [2018a]. Małecka [2016] z kolei szczegółowo przedstawia sposoby oceny prognoz VaR oraz ES.

Obszar zastosowań modeli ekonometrycznych dodatkowo można rozszerzyć o prognozowanie całego rozkładu prawdopodobieństwa przyszłych cen oraz ich zmian. Szczegółowy opis tych metod, a także miar i sposobów ewaluacji wykorzystywanych w prognozowaniu probabilistycznym, przedstawili w swoich pracach Gneiting i in. [2007], Gneiting i Raftery [2007], Jordan i in. [2018] a także Nowotarski i Weron [2018] w pracy dotyczącej prognozowania cen energii elektrycznej.

Wśród zwartych monografii wskazujących metodykę narzędzi ekonometrycznych w zakresie ekonometrii finansowej wyróżnić można: Osiewalski [2001], Brzeszczynski i Kelm [2002], Tsay [2010] oraz Danielsson [2011].

1.3 Stan badań w zakresie sieci neuronowych

Wraz ze wzrostem popularności sieci neuronowych uczenia głębokiego w ciągu ostatniej dekady, pojawiło się w tym okresie wiele publikacji naukowych poświęconych

metodom sztucznej inteligencji w modelowaniu szeregów czasowych. Poniższe zestawienie literatury skupia się przede wszystkim na przedstawieniu najnowszych publikacji, w dodatku z naciskiem na te badania, w których zastosowanie sieci służyło przede wszystkim modelowaniu stóp zwrotu, zmienności oraz szacowaniu ryzyka.

Jedną z publikacji, która szeroko opisuje problem zastosowania uczenia maszynowego do prognozowania rynków finansowych jest praca doktorska Fletchera [2012]. Można w niej znaleźć rzetelny i dobrze odzwierciedlający ówczesny stan badań przegląd metod i modeli uczenia maszynowego. Wykorzystanie klasycznych sieci neuronowych w tym obszarze opisane zostało także w monografii Dunis i in. [2016]. Zastosowanie poszczególnych narzędzi uczenia maszynowego do prognozowania szeregów czasowych instrumentów finansowych było także przedmiotem prac: Morajda [2007], Marcek i in. [2009], Devadoss i in. [2013], Bernal i in. [2012], Vui i in. [2013] Ding i in. [2015], Chen i Lee [2015], Fischer i Krauss [2018], Kurani i in. [2021].

Z kolei problematyka dotycząca wykorzystania sieci uczenia głębokiego opisana została m.in. w pracach: Fischer i Krauss [2018], Chong i in. [2017], Jiahong Li i in. [2017], Dixon i in. [2017], Di Persio i Honchar [2016], Hansson [2017], Chen i in. [2015], Chen i in. [2016], oraz Yang i in. [2019]. Dodatkowy przegląd literatury dotyczących stosowania narzędzi uczenia głębokiego w prognozowaniu finansowych szeregów czasowych można znaleźć w pracy Sezer i in. [2020]. Wyniki badań z tego zakresu dokładniej przeanalizowane zostały w podrozdziale 3.4. Z kolei porównanie narzędzi ekonometrycznych i metodami uczenia maszynowego omówione zostało m.in. w pracach: Hossain i Nasser [2011], Yu i Li [2018], Zhou i in. [2020] oraz Kijewski i Ślepaczuk [2020].

W literaturze znaleźć można także wiele publikacji na temat badań nad modelami hybrydowymi, łączącymi metody ekonometryczne z modelami bazującymi na sieciach neuronowych. Wśród prac dotyczących tej tematyki można wyróżnić publikacje: Matías i in. [2010], Monfared i Enke [2014], Lu i in. [2016], Kristjanpoller i Hernández [2017], Kristjanpoller i Minutolo [2018], Kim i Won [2018], Jeong i Lee [2019] oraz García i Kristjanpoller [2019]. Wyniki badań dotyczących wykorzystania modeli hybrydowych dokładniej przeanalizowane zostały w rozdziale 4.1. Na tle tych rezultatów wskazany następnie został wkład autora w tym nurcie.

Modele bazujące na sieciach neuronowych wykorzystywane są także w prognozowaniu całych rozkładów prawdopodobieństwa przyszłych stóp zwrotu. Problematyką tą zajmowali się między innymi Toubreau i in. [2019], Brusaferrri i in. [2019], Salinas i in. [2020], Chen i in. [2020]. Tematyka ta poruszana została także w monografiach Duerra 2020, oraz Murphy’ego [2012, 2022]. Opracowania te opisują problematykę prognozowania probabilistycznego z wykorzystaniem narzędzi uczenia

głębokiego na poziomie ogólnym, lub, jak ma to miejsce w przypadku wskazanych artykułów, dotyczą danych spoza obszaru związanego z rynkami finansowymi. Jednym z celów niniejszej pracy jest propozycja rozwiązań umożliwiających wykorzystanie tych metod w probabilistycznym prognozowaniu finansowych szeregów czasowych.

W tym miejscu zaznaczyć należy, że ze względu na dużą liczbę publikacji naukowych dotyczących prognozowania finansowych szeregów czasowych, zarówno w zakresie ekonometrii finansowej jak i sieci neuronowych, przedstawiona w powyższych podrozdziałach literatura dotyczy jedynie wybranych publikacji. W dalszej części rozprawy starano się możliwie często wskazywać dodatkowe odniesienia literaturowe, w miejscach gdzie wydawało się to zasadne, co ma na celu uzupełnienie obrazu obecnego stanu wiedzy.

Rozdział 2

Ekonometryczne modelowanie i prognozowanie jednowymiarowych finansowych szeregów czasowych (z wykorzystaniem modeli klasy ARMA-GARCH)

Poniższy rozdział opisuje podstawowe narzędzia wywodzące się przede wszystkim z dziedziny ekonometrii finansowej, stosowane w analizie szeregów czasowych. Ze względu na bardzo szeroki zakres poruszanych kwestii, szczególną uwagę poświęcono tutaj metodom, które zostaną wykorzystane w badaniach empirycznych wykonanych w ramach w tej pracy, przy czym wskazane zostały także inne narzędzia stosowane przez badaczy w zakresie poruszanej tematyki. Przedstawione w tym rozdziale definicje i wzory bazują przede wszystkim na pracach: Tsay [2010], Doman i Doman [2009], Fiszeder [2009] oraz Małecka [2016], w których znaleźć można znacznie szersze omówienie rozważanych metod.

2.1 Procesy stochastyczne w ekonometrycznej analizie szeregów czasowych

Powszechnie przyjmuje się, że procesy cen oraz stóp zwrotu mają charakter losowy (zob. Kendall i Hill [1953], Fama [1965b], Campbell i in. [1997]). Procesy te zatem traktować jako realizację procesów stochastycznych, pod pojęciem których rozumiemy (za Doman i Doman [2009]) rodzinę zmiennych losowych $(X_t)_{t \in \mathbb{T}}$,

określonych na wspólnej przestrzeni probabilistycznej (Ω, \mathcal{F}, P) , gdzie Ω jest zbiorem zdarzeń elementarnych, \mathcal{F} jest σ -ciałem podzbioru w zbiorze Ω , natomiast P jest σ -addytywną miarą probabilistyczną. W dalszym ciągu rozważane będą tylko procesy z czasem dyskretnym, czyli takie, gdzie zbiór indeksów \mathbb{T} jest przeliczalny i zawiera się w zbiorze liczb całkowitych ($\mathbb{T} = \mathbb{Z}$) lub całkowitych nieujemnych ($\mathbb{T} = \mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$). Szeregiem czasowym, oznaczanym symbolem (x_t) , nazywa się realizację procesu stochastycznego $(X_t)_{t \in \mathbb{T}}$, stanowiącą ciąg obserwacji x_1, x_2, \dots . W dalszej części pracy dla uproszczenia notacji jako (x_t) będzie oznaczany zarówno szereg czasowy jak i proces stochastyczny, z pominięciem także subskryptu $t \in \mathbb{T}$.

Na gruncie finansów empirycznych dwoma typami szeregów czasowych, które najczęściej podlegają analizie są szeregi czasowe cen (notowań) instrumentów finansowych oraz ich stopy zwrotu, opisane w poniższych podrozdziałach.

2.2 Empiryczne własności finansowych szeregów czasowych

Aby prawidłowo zrozumieć poruszaną w pracy problematykę, konieczne jest pewne teoretyczne wprowadzenie do tematyki związanej z modelowaniem finansowych szeregów czasowych, oraz przedstawienie podstawowych definicji i własności takich szeregów. Poniżej przedstawione zostaną podstawowe własności szeregów czasowych cen oraz, przede wszystkim, stóp zwrotu, które stanowią główny przedmiot badań ekonometrii finansowej. Znajomość tych własności jest kluczowa, ponieważ umożliwia wybór i odpowiednie zastosowanie narzędzi służących do ich modelowania.

2.2.1 Szeregi czasowe cen

Podstawowym źródłem informacji w badaniach z zakresu finansów empirycznych są ceny instrumentów finansowych. Same ceny rzadko są jednak przedmiotem badań, w praktyce najczęściej analizie poddaje się stopy zwrotu. Campbell i in. [1997] (za Tsay [2010]) zwracają uwagę na dwa powody ku temu: stopy zwrotu zawierają kompletne informacje o inwestycji, wyrażone w jednakowej skali, a także posiadają pewne (bardziej atrakcyjne dla badaczy) własności statystyczne, które umożliwiają ich modelowanie. Istnieją jednak obszary badań, w których podstawą analizy są jednak szeregi cen. Przykładowo, w modelowaniu wielowymiarowym (wielu instrumentów finansowych jednocześnie), w przypadku stóp zwrotu możliwe jest jedynie badanie krótkookresowych zależności pomiędzy aktywami. Chcąc natomiast modelować zależności długookresowe (przykładowo w ramach analizy kointegracji),

podstawę stanowią szeregi czasowe w postaci logarytmów cen badanych instrumentów (zob. Osiewalski i Osiewalski [2013], Pajor i Wróblewska [2017]). Szeregi czasowe cen instrumentów finansowych są także często przedmiotem badań w publikacjach dotyczących prognozowania z wykorzystaniem metod sztucznej inteligencji, cechujących się podejściem eklektycznym (zob. Lu i in. [2020], Shahi i in. [2020]). Jednakże, ze względu na przedmiot tej pracy (procesy jednowymiarowe) i teoretyczne podstawy silnie bazujące na ekonometrii finansowej, analizie poddane zostaną szeregi czasowe w postaci stóp zwrotu.

2.2.2 Proste i logarytmiczne Stopy zwrotu – ich definicje i własności

Można wyróżnić dwa rodzaje stóp zwrotu (za Fiszeder [2009]):

- zwykła (prosta) stopa zwrotu, wyrażona w punktach procentowych, określona wzorem:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \cdot 100, \quad (2.1)$$

- logarytmiczna stopa zwrotu, wyrażona w punktach procentowych, określona wzorem:

$$r_t = \ln P_t - \ln P_{t-1} \cdot 100, \quad (2.2)$$

gdzie P_t jest ceną instrumentu finansowego w chwili t . Doman i Doman [2004] zwracają uwagę, że w ekonometrii i matematyce finansowej logarytmiczne stopy zwrotu analizowane są częściej, głównie z uwagi na prostotę obliczania zwrotów długookresowych, możliwość przyjmowania przez nie dowolnych wartości ze zbioru liczb rzeczywistych (w przeciwieństwie do prostych stóp zwrotu, które są lewostronnie ograniczone przez -1), oraz częste występowanie procesu logarytmu ceny w stochastycznych równaniach różniczkowych wykorzystywanych modelach matematyki finansowej, stosowanych w przypadku analizy procesów z czasem ciągłym. Proste stopy zwrotu z kolei stosowane są przez inwestorów, w praktyce banków i instytucji finansowych oraz w analizie portfelowej, głównie ze względu na ich łatwiejszą interpretację. W dalszych badaniach najczęściej wykorzystywana będzie postać logarytmiczna.

Powszechnie zakłada się, że szeregi (logarytmicznych) stóp zwrotu instrumentów finansowych są kowariancyjnie (słabo) stacjonarne (zob. Doman i Doman [2009]). Oznacza to, że zmienne r_t mają skończony drugi moment $E(r_t^2) < \infty$, a dodatkowo

$$E(r_t) = \mu, \quad (2.3)$$

$$Cov(r_t, r_{t+l}) = \gamma_l, \quad (2.4)$$

dla dowolnych indeksów t oraz l , co oznacza, że auto-kowariancja jest funkcją tylko odległości na osi czasu, a nie samego czasu, oraz że wszystkie r_t mają tą samą, skończoną wariancję γ_0 .

Z kolei o ściślejszej stacjonarności procesu (r_t) mówimy w przypadku, kiedy łączny rozkład wektora losowego $(r_{t_1}, r_{t_2}, \dots, r_{t_n})$ jest taki sam jak rozkład wektora $(r_{t_1+k}, r_{t_2+k}, \dots, r_{t_n+k})$, co oznacza, że rozkłady te nie zmieniają się względem przesunięcia na osi czasu.

Gdy mamy do czynienia z procesem kowariancyjnie stacjonarnym, można posłużyć się empiryczną funkcją autokorelacji (ang. *AutoCorrelation Function* - ACF), w celu oszacowania współczynników autokorelacji z rzędów $l = 1, 2, \dots$, zadaną wzorem:

$$\hat{C}orr(r_t, r_{t-l}) = \hat{\rho}_l = \frac{\sum_{t=l+1}^T (r_t - \bar{r})(r_{t-l} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2}, \quad (2.5)$$

gdzie T oznacza długość szeregu czasowego, natomiast \bar{r} oznacza średnią z próby, $\bar{r} = \frac{1}{T} \sum_{t=1}^T r_t$. Należy się spodziewać, że dla procesów kowariancyjnie stacjonarnych wpływ przeszłych wartości procesu (r_t) na bieżące wartości r_t wygasa wraz ze wzrostem l .

Istotnym pojęciem w analizie szeregów czasowych, o którym należy tutaj wspomnieć, jest biały szum. Proces stochastyczny nazywamy białym szumem (ang. *White Noise*) gdy zmienne r_t tworzą ciąg nieskorelowanych zmiennych losowych o zerowej średniej i stałej, skończonej wariancji. Jeśli zmienne losowe są dodatkowo niezależne to proces taki nazywamy ścisłym białym szumem.

2.2.3 Fakty empiryczne

Wśród empirycznych własności danych finansowych wyróżnić można pewne prawidłowości, które są obserwowalne podczas analizy szeregów stóp zwrotu różnorodnych aktywów. Właściwości te określane są mianem faktów empirycznych (ang. *stylized facts*). Należą do nich następujące własności (za Doman i Doman

[2004]):

- brak autokorelacji - autokorelacja w szeregach stóp zwrotu jest często nieistotna statystycznie (z wyjątkiem zwrotów śróddziennych, gdzie może pojawić się efekt mikrostruktury rynku),
- grupowanie zmienności (ang. *volatility clustering*) - występujące po sobie okresy podwyższonej i obniżonej zmienności (rozumiane np. jako warunkowa wariancja),
- grube ogony rozkładów - częste występowanie obserwacji nietypowych (odstających) w porównaniu do rozkładu normalnego,
- spiczastość rozkładów - silniejsza koncentracja wokół modalnej,
- asymetria spadków i wzrostów - znaczne ujemne zmiany są często większe co do wartości bezwzględnej od znacznych wzrostów, przez co rozkłady są często lewostronnie asymetryczne,
- efekt dźwigni (ang. *leverage effect*) - ujemna korelacja pomiędzy stopami zwrotu a ich zmiennością,
- wahania zmienności - zmienne w czasie natężenie poziomu zmienności, niezależnie od przyjętego sposobu jej estymowania,
- powracanie zmienności do średniej - po okresach zwiększonej zmienności następuje powrót do pewnego jej „normalnego” poziomu,
- powolne zanikanie autokorelacji w szeregach wartości bezwzględnych stóp zwrotu, co może wskazywać na występowanie zależności długookresowej.

Własnością tym poświęcona jest szeroka literatura, z pośród której wskazać można tutaj takie publikacje jak Cont [2001], Engle i Patton [2001], Brzeszczyński i Kelm [2002] oraz Doman i Doman [2004, 2009].

2.3 Zmienność finansowych szeregów czasowych

Kluczową wielkością podlegającą modelowaniu i prognozowaniu zarówno w ekonometrii finansowej, jak i wśród praktyków rynku jest zmienności cen instrumentów finansowych. Najczęściej rozumiana jest jako niepewność dotycząca ceny lub stóp zwrotu tego instrumentu, a wyrażana jest jako jego warunkowa wariancja lub warunkowe odchylenie standardowe (zob. Fiszeder [2009]). Zmienność nie jest bezpośrednio obserwowana, przez co powstało wiele sposobów estymacji jej

wartości. Możliwie najlepszy opis i prognozowanie zmienności jest istotny ze względu na szeroki zakres jej zastosowania w analizie instrumentów finansowych, stanowiąc podstawy w takich obszarach jak konstruowanie portfela inwestycyjnego, szacowanie ryzyka kapitałowego czy wycena opcji.

2.3.1 Zmienność historyczna i zrealizowana

Jednym z najprostszych sposobów mierzenia zmienności są kwadraty stóp zwrotu wyznaczanych na podstawie cen zamknięcia:

$$\hat{\sigma}_t^2 = r_t^2, \quad (2.6)$$

gdzie r_t zadane jest wzorem 2.2. Estymator w tej postaci jest jednak nieefektywny ponieważ bazuje tylko na jednej obserwacji.

Piontek [2002] oraz Fiszeder [2009] jako popularne estymatory zmienności wskazują wariancję i odchylenie standardowe stóp zwrotu, opisane wzorami:

$$\hat{\sigma}_t^2 = \frac{1}{n} \sum_{t=1}^n r_t^2 - \bar{r}^2, \quad (2.7)$$

gdzie \bar{r} jest średnią arytmetyczną stóp zwrotu z próby, zaś n ustaloną długością ruchomego okna obserwacji.

W praktyce, wśród badaczy związanych bliżej ze środowiskiem inwestycyjnym, często stosowaną miarą zmienności jest tzw. zmienność historyczna, (ang. *Historical Volatility* - HV), najczęściej przedstawiana w skali rocznej jako odchylenie standardowe. Zmienność historyczna obliczana jest za pomocą wzoru (Piontek [1999]):

$$HV_t = \sqrt{\frac{N}{n-1} \sum_{t=1}^n (r_t - \bar{r})^2}, \quad (2.8)$$

gdzie N liczbą okresów sesyjnych w ciągu roku (zazwyczaj 252 dni).

Fiszeder [2009] i Doman i Doman [2009] jako najlepszą miarę zmienności wskazują zmienność zrealizowaną (ang. *Realized Volatility* - RV), zaproponowaną przez Andersena i Bollersleva [1998], szacowaną na podstawie kwadratów stóp zwrotu o częstotliwości wyższej niż dzienna:

$$RV_t = \sum_{j=1}^d r_{t,j}^2, \quad (2.9)$$

gdzie $r_{t,j}$ oznacza śróddzienną stopę zwrotu o interwale j , zaś d to liczba obserwacji śróddziennych.

Na rynkach finansowych, które nie są notowane przez całą dobę, należy wziąć pod uwagę także nocną stopę zwrotu. Powstało kilka propozycji rozwiązania tego problemu, w tym popularne rozwiązanie zaproponowane przez Martensa 2002, polegające na wprowadzeniu dodatkowego współczynnika skalującego c :

$$RV_t = (1 + c) \sum_{d=1}^D r_{d,t}^2, \quad (2.10)$$

$$c = \frac{\sigma_{co}^2}{\sigma_{oc}^2}, \quad (2.11)$$

gdzie $\sigma_{co}^2 = Var(r_{n,t})$ oznacza wariancję nocnych stóp zwrotu, natomiast $\sigma_{oc}^2 = Var\left(\sum_{d=1}^D r_{d,t}\right)$ jest wariancją dziennych stóp zwrotu, przy czym D oznacza tutaj liczbę obserwacji o ustalonej częstotliwości w ciągu dnia.

Doman i Doman [2009] zwracają uwagę, że podczas wyboru częstotliwości obserwacji śróddziennych należałoby wybierać obserwacje o jak najwyższej częstotliwości, ze względu na to, że najbardziej zbliżone są one do procesu ciągłego. W praktyce jednak, dane o niskiej częstotliwości dla długich okresów często są słabej jakości lub niedostępne. Dodatkowo pojawiają się także problemy związane z tzw. efektem mikrostruktury rynku, wynikające z dynamiki i ułatwień możliwości zawierania transakcji (zob. także Bień-Barkowska [2016]). Zalicza się do nich między innymi niską płynność, koszty transakcyjne, różnice w ofertach kupna i sprzedaży oraz problemy związane z napływem informacji i nieregularnością handlu. Dla danych o wysokiej częstotliwości często występuje także silna autokorelacja. Autorzy, powołując się na literaturę, zalecają wybór częstotliwości pomiędzy 5-minutową i godzinną, natomiast Tsay [2010] dla aktywów o dużej płynności – częstotliwości od 4- do 15-minutowej (zob. także Andersen i in. [2000], Oomen [2001], Będowska-Sójka i Kliber [2010]).

2.3.2 Estymatory zmienności bazujące na zakresie cen

W związku z przedstawionymi powyżej problemami dotyczącymi danych o wysokiej częstotliwości, Fiszeder [2020] proponuje stosowanie estymatorów konstruowanych na podstawie zakresów cen (ang. *Range Estimators*), szczególnie w sytuacjach, gdy występują efekty mikrostruktury rynku lub gdy brakuje dobrej jakości danych o wysokiej częstotliwości.

Przyjmijmy następujące oznaczenia:

- O_t — cena otwarcia w okresie t ,
- C_t — cena zamknięcia w okresie t ,
- L_t — cena minimalna w okresie t ,
- H_t — cena maksymalna w okresie t .

Poniżej zestawiono wybrane estymatory bazujące na zakresach cen.

Estymator Parkinsona

Jednym z podstawowych estymatorów budowanych na podstawie zakresów cen jest estymator Parkinsona [1980], bazujący na cenach maksymalnej i minimalnej, zakładający zerową wartość dryfu:

$$\hat{\sigma}_t^2 = \frac{1}{4 \ln 2} \left(\ln \frac{H_t}{L_t} \right)^2. \quad (2.12)$$

Estymator Garmana-Klasa

Garman i Klass [1980] przedstawili estymator wprowadzający dodatkowo ceny otwarcia i zamknięcia, zakładający jednak brak skoku między cenami zamknięcia z okresu poprzedniego i otwarcia w bieżącym okresie, a także zerową wartość dryfu:

$$\hat{\sigma}_t^2 = \frac{1}{2} \left(\ln \frac{H_t}{L_t} \right)^2 - (2 \cdot \ln 2 - 1) \cdot \ln \left(\frac{C_t}{O_t} \right)^2. \quad (2.13)$$

Estymator Rogersa-Sachela

Chcąc umożliwić wykorzystanie niezerowej wartości dryfu, Rogers i Satchell [1991] jako rozszerzenie estymatora Parkinsona zaproponowali estymator przedstawiony wzorem:

$$\hat{\sigma}_t^2 = \ln \frac{H_t}{C_t} \cdot \ln \frac{H_t}{O_t} + \ln \frac{L_t}{C_t} \cdot \ln \frac{L_t}{O_t}. \quad (2.14)$$

Estymator ten może jednak przyjmować wartość zerową pomimo dużych zmian ceny w ciągu dnia. Dzieje się tak w przypadku gdy cena otwarcia jest równa cenie maksymalnej a cena zamknięcia cenie minimalnej.

Estymator Garmana Klasa zmodyfikowany o skoki pomiędzy cenami otwarcia i zamknięcia (GKYZ)

Yang i Zhang [2000] rozszerzyli podstawową postać estymatora Garmana-Klasa o różnicę pomiędzy cenami zamknięcia z dnia poprzedniego i otwarcia w dniu dzisiejszym (wciąż jednak zakładając zerową wartość dryfu):

$$\hat{\sigma}_t^2 = \left(\ln \frac{O_t}{C_{t-1}} \right)^2 + \frac{1}{2} \left(\ln \frac{H_t}{L_t} \right)^2 - (2 \cdot \ln 2 - 1) \cdot \ln \left(\frac{C_t}{O_t} \right)^2. \quad (2.15)$$

We wspomnianej pracy, Autorzy zaproponowali także własną wersję estymatora, który bierze pod uwagę skoki cen między zamknięciem i otwarciem, a także umożliwia wykorzystanie założenia o niezerowym dryfie. Estymator ten zakłada jednak stałość wariancji dla okresu dłuższego niż jeden dzień. Ponadto, niemożliwe jest także oszacowanie za jego pomocą wariancji dla pojedynczego dnia.

W literaturze znaleźć można także wiele innych propozycji estymatorów zmienności, których dokładne porównanie znaleźć można w książce Fiszeder [2020], poświęconej w całości tej tematyce, a także w pracach Fałdziński i Osińska [2016] oraz Shu i Zhang [2006]. W części empirycznej tej pracy wykorzystany zostanie estymator Garmana-Klasa zmodyfikowany przez Yanga i Zhanga (oznaczany w dalszej części pracy skrótem od nazwisk autorów - GKYZ), głównie ze względu na prostotę jego konstrukcji, brak założenia o stałości wariancji i wzięcie pod uwagę skoku między cenami otwarcia i zamknięcia.

2.3.3 Transformacje estymatorów zmienności

Jak zauważa Fiszeder [2005], [2009], [2020], estymatory zmienności konstruowane na podstawie zakresu cen często niedoszacowują zmienność. Wynika to z faktu, że procesy cen są procesami dyskretnymi, zaś głównym założeniem leżącym u podstaw tych estymatorów jest to, że mamy do czynienia z procesem z czasem ciągłym.

Jako próbę rozwiązania tego problemu zaproponował on metodę skalowania estymatora bazującego na prawdziwym zakresie zmian (ang. *True Range TR*), według propozycji Wildera [1978]:

$$TR_t = \max \left\{ H_t - L_t, |C_{t-1} - H_t|, |C_{t-1} - L_t| \right\}. \quad (2.16)$$

Propozycja Fiszedera przyjmuje następującą postać:

$$STR_t = \frac{a}{b} TR, \quad (2.17)$$

$$a = \sqrt{\frac{1}{t} \sum_{i=1}^t r_i^2}, b = \sqrt{\frac{1}{t} \sum_{i=1}^t TR_i^2}, \quad (2.18)$$

gdzie czynniki skalujące a i b wynikają z porównania średnich kwadratów stóp zwrotu i zmienności oszacowanej za pomocą estymatora.

W równaniach czynników skalujących, zadanych w 2.18, zamiast kwadratów można stosować wartości bezwzględne, jednak może to prowadzić do zniżonych prognoz. Jak sugeruje Autor powyższych prac źródłowych, skalowanie 2.17 jest konieczne aby można było wykorzystać STR jako oszacowanie zmienności, które następnie wykorzystane jest w zaproponowanym przez Fiszедера [2005] modelu GARCH-TR.

Wilder [1978] zaproponował także uśrednioną wersję estymatora prawdziwego zakresu cen - (ang. *Averaged True Range* - ATR).

$$ATR = \frac{1}{n} \sum_{t=1}^n TR_t, \quad (2.19)$$

gdzie n to długość okna średniej ruchomej, zazwyczaj ustalany na 14 lub 21 dni.

Dodatkowo stosowanym zabiegiem jest tutaj obliczenie średniej z n okresów wstecz aby uzyskać uśredniony szacunek zmienności. To podejście zakłada jednak stałość wariancji w tym okresie, co z formalnego punktu widzenia nie jest słuszne. Wygładzone szacunki zmienności są o wiele bardziej zbliżone do prognoz uzyskanych przez modele GARCH. Fiszeder [2020] zauważa jednak, że nie należy dążyć do dopasowania zmienności w ten sposób, ponieważ prognozy te są jedynie pewnym przybliżeniem procesu generującego dane finansowe.

Bazując na przeglądzie literatury można stwierdzić, że w praktyce wśród publikacji związanych z badaniami nad zmiennością w celach inwestycyjnych, a także w literaturze dotyczącej uczenia maszynowego, najczęściej stosuje się ujęcie historyczne (HV) lub kwadraty stóp zwrotu (zob. Kim i Won [2018], Kristjanpoller i Minutolo [2018]). Z kolei w literaturze ekonometrycznej zaleca się stosowanie zmienności zrealizowanej (RV) lub estymatorów bazujących na zakresach cen (zob. Fiszeder [2009, 2020]).

Odrębnym aspektem skalowania, stosowanym przy obliczaniu zmienności w dłuższym horyzoncie, jest wspomniane także wcześniej mnożenie estymatora zmienności (wariancji) przez liczbę okresów (dni sesyjnych) w ciągu roku - N - zazwyczaj ustalane na 252, uzyskując w ten sposób zmienność w skali rocznej (zannualizowaną).

2.3.4 Zmienność implikowana

Odrębne podejście do estymacji zmienności polega na obliczaniu jej wartości na podstawie cen opcji, uzyskując w ten sposób warunkowe odchylenie standardowe dla danego instrumentu bazowego. Wyznaczane w ten sposób odchylenie standardowe nazywane jest zmiennością implikowaną (zob. Schmalensee i Trippi [1978], Beckers [1981]). Przykładowo, zakładając, że ceny opcji są realizacją procesu Browna, zmienność dla opcji europejskich wyznaczyć można modelem Blacka i Scholesa (Black i Scholes [1973]) lub jego uogólnieniami. Jednak także ten sposób obliczania zmienności nie jest pozbawiony wad. Podstawowy model Blacka-Scholesa zakłada stałość zmienności. Dodatkowo, uzyskane w ten sposób oceny zmienności zawierają premię za ryzyko (przez co zmienność wyznaczana w ten sposób jest wyższa niż ta uzyskana przy pomocy modeli GARCH), specyfikacja modelu użytego do wyceny opcji może być wadliwa, a także występować może brak równowagi rynkowej lub efekt uśmiechu zmienności (ang. *volatility smile*). Mimo to, zmienność implikowana cieszy się dużą popularnością, przykładowo jest ona podstawą do obliczania wartości dla indeksów zmienności takich jak VIX (ang. *Chicago Board Options Exchange's Volatility Index*), na podstawie indeksu S&P 500, czy indeksu VDAX, bazującego na niemieckim indeksie DAX (zob. Fiszeder [2009]).

2.4 Modele klasy ARMA-GARCH

Jednym z podstawowych narzędzi ekonometrycznych wykorzystywanych w tej pracy są modele ARMA-GARCH. Specyfikacja tych modeli umożliwia opis procesu stóp zwrotu jednocześnie za pomocą warunkowej wartości oczekiwanej i warunkowej wariancji. Należą one do jednych z najpopularniejszych metod stosowanych w finansach empirycznych, stanowią także podstawę w badaniach nad rozwojem nowych narzędzi. Wyniki uzyskane przez modele ARMA-GARCH stanowiąc będą główny punkt odniesienia w analizie porównawczej z modelami hybrydowymi i sieciowymi modelami prognozowania probabilistycznego.

2.4.1 Definicja procesu ARMA-GARCH

Szeregi logarytmicznych stóp zwrotu można opisać za pomocą procesu ze zmienną w czasie warunkową wartością oczekiwaną μ_t i wariancją warunkową σ_t^2 :

$$r_t | \Psi_{t-1} \sim D(\mu_t, \sigma_t^2), \quad (2.20)$$

gdzie Ψ_{t-1} oznacza całą przeszłość procesu (r_t) do chwili $t - 1$ włącznie, zaś D oznacza zadany rozkład prawdopodobieństwa.

W takim przypadku do opisu tego procesu można wykorzystać model ARMA-GARCH, w którym średnia warunkowa opisana jest za pomocą struktury ARMA, natomiast dynamikę warunkowej zmienności opisuje proces GARCH. Proces ARMA-GARCH można zdefiniować następująco:

$$r_t = a_0 + \sum_{i=1}^p a_i r_{t-i} + \varepsilon_t - \sum_{i=1}^q b_i \varepsilon_{t-i}, \quad (2.21)$$

$$\varepsilon_t = z_t \sigma_t, z_t \sim N(0, 1), \quad (2.22)$$

$$\sigma^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2. \quad (2.23)$$

W kolejnych punktach zostaną dokładniej omówione struktury procesów opisujących dynamikę średniej warunkowej oraz składnika losowego, a także metoda wykorzystywana do estymacji łącznego modelu ARMA-GARCH.

2.4.2 Procesy ARMA

Podstawą modelowania średniej warunkowej w dalszych rozważaniach jest proces ARMA(k, m). Specyfikację tego procesu można opisać za pomocą wzoru (zob. Tsay [2010], Doman i Doman [2009]):

$$r_t = a_0 + \sum_{i=1}^k a_i r_{t-i} + \varepsilon_t - \sum_{j=1}^m b_j \varepsilon_{t-j}, \quad (2.24)$$

gdzie k i m są nieujemnymi liczbami całkowitymi oznaczającymi maksymalne rzędy opóźnień, natomiast (ε_t) jest (z reguły gaussowskim) białym szumem. Zmienna losowa (ε_t) zwana jest składnikiem losowym, wstrząsem, szokiem lub innowacją w okresie t .

Dwie główne wyrażenia sumy przedstawionej w równaniu (2.24) opisują: proces autoregresyjny (ang. *Autoregressive process* - AR) oraz proces średniej ruchomej (ang. *Moving Average process* - MA), które zostaną szerzej opisane poniżej. Połączenie tych dwóch składowych jest szczególnie pomocne przy opisie zależności długotrwałych z wykorzystaniem mniejszej liczby parametrów (zob. Doman i Doman [2009]).

Do wskazania optymalnych wartości rzędów opóźnień k i m w modelach ARMA(k, m) najczęściej stosuje się kryterium Akaike (ang. *Akaike Information Criterion* - AIC) lub kryterium Schwarza (ang. *Schwarz Criterion* - SC), zwane również kryterium Bayesowskim (ang. *Bayesian Information Criterion* - BIC).

Równania dla tych kryteriów przedstawiają się następująco (za Doman i Doman [2009]):

$$AIC = -2 \frac{\ln L(\hat{\theta})}{T} + 2 \frac{a}{T}, \quad (2.25)$$

$$BIC = -2 \frac{\ln L(\hat{\theta})}{T} + 2 \frac{a \ln T}{T}, \quad (2.26)$$

gdzie $L(\hat{\theta})$ oznacza wartość funkcji wiarygodności oszacowanego wektora parametrów $\hat{\theta}$ dla obserwacji (r_1, \dots, r_T) , natomiast a oznacza liczbę wszystkich parametrów modelu.¹ Warto zauważyć, że kryterium BIC ma tendencję do preferowania modeli o mniejszej liczbie parametrów. Obok przedstawionych powyżej kryteriów proponowane są także inne, takie jak: kryterium informacyjne Hannana-Quinna (ang. *Hannan-Quinn Information Criterion*, zob. Hannan i Quinn [1979]) lub kryterium Shibaty [1976].

Doman i Doman [2004] oraz Tsay [2010] zwracają uwagę, że funkcje autokorelacji i cząstkowej autokorelacji² w przypadku modeli ARMA raczej nie dają użytecznych informacji na temat rzędów opóźnień. W tym przypadku należy estymować określoną liczbę modeli, dla wybranych wartości K i M , gdzie $0 \leq k \leq K$ i $0 \leq m \leq M$. Wybierane są zazwyczaj te modele, dla których wartości AIC lub BIC są najniższe.

W niektórych przypadkach, podczas wyboru rzędów opóźnień modeli ARMA(k, m), przykładowo przy użyciu opisanych powyżej kryteriów informacyjnych, okazuje się, że optymalną wartością parametru m jest zero. Wartość r_t opisana jest wówczas poprzez liniową kombinację opóźnionych wartości r_{t-i} , co odpowiada procesowi autoregresji AR(p)³:

$$r_t = a_0 + \sum_{i=1}^m a_i r_{t-i} + \varepsilon_t, \quad (2.27)$$

gdzie ε_t oznacza ścisły biały szum z zerową średnią i stałą wariancją⁴ σ^2 .

Gdy mamy do czynienia z procesem AR(1): $r_t = a_0 + a_1 r_{t-1} + \varepsilon_t$, można dowieść następujących własności (zob. Doman i Doman [2009]):

¹Estymację przy pomocy metody największej wiarygodności, opisano w punkcie 2.4.5, w kontekście modeli ARMA-GARCH.

²Opisane wzorami 2.5 oraz 2.28 - 2.30.

³Zdążają się też przypadki, w których wartości optymalne dla obu rzędów opóźnień są równe zero, co oznacza, że logarytm ceny można opisywać za pomocą błędzenia losowego.

⁴W przypadku założenia o zmienności wariancji warunkowej, proces stóp zwrotu możemy modelować wykorzystując model AR-GARCH.

- warunkowa wartość oczekiwana zmiennej r_t wynosi $E(r_t|r_{t-1}) = a_0 + a_1r_{t-1}$,
- warunkowa wariancja $Var(r_t|r_{t-1}) = Var(\varepsilon_t) = \sigma^2$,
- proces (r_t) jest kowariancyjnie stacjonarny w tedy i tylko wtedy gdy $-1 < a_1 < 1$ (zachodzi tu też zależność odwrotna),
- funkcja ACF przyjmuje postać $\rho_l = a_1\rho_{l-1}$ dla $l > 0$ i $\rho_l = a_1^l$ dla $l \geq 0$.

W celu wyboru rzędu modelu AR(p), obok wspomnianych powyżej kryteriów informacyjnych, można posłużyć się także funkcją autokorelacji cząstkowej (ang. *Partial AutoCorrelation Function* - PACF). Definiuje się ją korzystając ze współczynników autokorelacji (za Doman i Doman [2009]):

$$a_{1,1} = \rho_1, \quad (2.28)$$

$$a_{2,2} = (\rho_2 - \rho_1^2)(1 - \rho_1^2), \quad (2.29)$$

$$a_{s,s} = \frac{\rho_s - \sum_{j=1}^{s-1} a_{s-1,j}\rho_{s-j}}{1 - \sum_{j=1}^{s-1} a_{s-1,j}\rho_j}, \text{ dla } s = 3, 4, 5, \dots \quad (2.30)$$

gdzie $a_{s,j} = a_{s-1,j} - a_{s,s}a_{s-1,s-j}$ dla $j = 1, 2, 3, \dots, s-1$. Współczynniki autokorelacji cząstkowej oszacować można estymując za pomocą metody najmniejszych kwadratów równania AR kolejnych rzędów.

W przypadkach, w których wartość r_t jest wyłącznie przez liniową kombinację bieżącego oraz przeszłych zaburzeń losowych ε stosuje się model średniej ruchomej (ang. *Moving Average*) - MA(m). Model ten można zapisać w postaci:

$$r_t = c_0 - \sum_{j=1}^m b_j\varepsilon_{t-j}, \quad (2.31)$$

gdzie c_0 oznacza pewną stałą. „Czyste” modele średniej ruchomej rzadko są wykorzystywane w empirycznych analizach danych finansowych.

Modele ARMA cieszą się wśród badaczy dużą popularnością, przez długi czas były jednym z podstawowych narzędzi służących do opisu stóp zwrotu. Najczęściej stosowane są modele o niskich rzędach opóźnień, głównie ze względu na mniejszą liczbę estymowanych parametrów. Więcej na temat tych modeli, ich właściwości, sposobów doboru liczby opóźnień oraz metod prognozowania z ich wykorzystaniem znaleźć można w licznych opracowaniach (zob. Mondal i in. [2014], Junior i in. [2014], Kocak [2017]).

Warto zauważyć, że procesy AR oraz ARMA, pomimo swoich własności, nie są jednak wystarczające do opisu szeregów stóp zwrotu, i należy je rozbudować o składnik losowy (ε_t), w którym wariancja warunkowa będzie zmienna w czasie.

2.4.3 Opis struktury GARCH

Jako jedno z głównych narzędzi wykorzystywanych w celu modelowania składnika losowego ze zmienną w czasie wariancją warunkową, w badaniach i prognozowaniu szeregów czasowych stosuje się uogólnione modele autoregresyjnej warunkowej heteroskedastyczności (ang. *Generalized Autoregressive Conditional Heteroskedasticity* - GARCH). Wykorzystywane są one przede wszystkim w modelowaniu zmienności w zagadnieniach finansów empirycznych takich jak analiza ryzyka czy budowa portfela inwestycyjnego. Rodzina modeli GARCH jest bardzo szeroka, od czasu pierwszych specyfikacji modelu wprowadzono liczne uogólnienia i modyfikacje, które mają na celu dokładniejszy opis modelowanych danych, z uwzględnieniem ich dodatkowych własności. W tej części pracy opisane zostały wybrane, najpopularniejsze specyfikacje modeli klasy GARCH, które zostały także wykorzystane w części empirycznej.

Pojęcie heteroskedastyczności warunkowej zaproponowane zostało przez R. F. Engle'a [1982]. Zaproponowany przez niego model ARCH cieszył się dużą popularnością głównie w badaniach makroekonomicznych, wykorzystujących dane o niskiej częstotliwości (miesięcznej lub kwartalnej). Poniżej przedstawiona została definicja procesu ARCH, wprowadzająca do metodyki modelowania zmienności warunkowej.

Proces stochastyczny opisujący logarytmiczne stopy zwrotu można przedstawić równaniem (za Doman i Doman [2009]):

$$r_t = \mu_t + \varepsilon_t, \quad (2.32)$$

gdzie $\mu_t = E(r_t | \Psi_{t-1})$ jest warunkową wartością oczekiwaną.

Proces (ε_t) jest procesem ARCH(q) gdy zachodzi:

$$\varepsilon_t = z_t \sigma_t, z_t \sim N(0, 1), \quad (2.33)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2, \quad (2.34)$$

przy ograniczeniach $\alpha_0 > 0$, $\alpha_i \geq 0$ dla $i = 1, \dots, q$, gdzie σ_t^2 oznacza warunkową względem przeszłości wariancję procesu (ε_t) .

Uogólniona postać modelu - GARCH

Zaproponowany przez Engla model rozszerzyli w swoich badaniach jednocześnie T. Bollerslev [1986] i S.Taylor [1986] proponując model GARCH. Okazał się on bardziej adekwatny do pracy z szeregami o wyższej częstotliwości, częściej wykorzystywanych w badaniach na gruncie finansowym.

Zgodnie z definicją, proces (ε_t) jest procesem GARCH(p,q) gdy określony jest on równaniem (2.33), natomiast równanie wariancji warunkowej przyjmuje postać:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (2.35)$$

przy ograniczeniach $\alpha_0 > 0$, $\alpha_i \geq 0$ dla $i = 1, \dots, q$, oraz $\beta_i \geq 0$ dla $i = 1, \dots, p$.

W badaniach empirycznych zwykle przyjmuje się $p = q = 1$. Model w postaci GARCH(1,1) jest więc jednym z najczęściej spotykanych w literaturze. Do własności empirycznych procesu stochastycznego (ε_t) opisanego procesem GARCH(1,1) można zaliczyć brak autokorelacji stóp zwrotu dla tego procesu, co wskazuje na to, że wymagane jest osobne modelowanie warunkowej wartości oczekiwanej np. za pomocą procesu ARMA.

Funkcja autokorelacji procesu (ε_t^2) ma postać (za Kwiatkowski [2008]):

$$\rho_{\varepsilon_t^2}(l) = \begin{cases} \frac{\alpha_1(1-\beta_1^2-\alpha_1\beta_1)}{1-\beta_1^2-2\alpha_1\beta_1}, & \text{dla } l = 1, \\ (\alpha_1 + \beta_1)\rho_{\varepsilon_t^2}(l-1), & \text{dla } l < 1. \end{cases} \quad (2.36)$$

Taki charakter funkcji autokorelacji oznacza pewną persystencję w kwadratach procesu GARCH(1,1), co odpowiada własnościom empirycznym stóp zwrotu i pozwala wychwycić grupowanie zmienności. Suma wartości parametrów α_1 i β_1 odpowiada za tempo wygasania funkcji autokorelacji oraz wpływ przeszłych wartości zmienności na obserwacje w chwili obecnej, który jest większy im suma ta jest bliższa jedności.

Z kolei, jeżeli zachodzi warunek $\alpha_1 + \beta_1 < 1$ (co jest jednocześnie warunkiem kowariancyjnej stacjonarności rozważanego procesu), to istnieje skończona wariancja bezwarunkowa zadana wzorem:

$$Var(\varepsilon_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}. \quad (2.37)$$

Co więcej, w przypadku procesu GARCH(1,1) warunek kowariancyjnej stacjonarności jest silniejszy niż warunek ścisłej stacjonarności $E(\ln(\alpha_1\varepsilon_t^2 + \beta_1)) < 0$ (zob. także Nelson [1990], Osiewalski [2001]).

Z równania (2.33) wynika, że rozkład warunkowy ε_t jest rozkładem normalnym:

$$p(\varepsilon_t | \Psi_{t-1}, \theta) = f_N^{(1)}(\varepsilon_t; 0, \sigma_t^2). \quad (2.38)$$

Mimo to, można pokazać, że rozkład bezwarunkowy posiada kurtozę wyższą od 3, co oznacza, że ogony rozkładu brzegowego ε_t są grubsze niż te w rozkładzie normalnym.

Szersze opracowanie własności procesów GARCH przedstawione zostało także w pracach Piontek [2002], Pipień [2006], Kwiatkowski [2008], Doman i Doman [2009] oraz Fiszeder [2009].

Typy rozkładów warunkowych

Jednym z podstawowych rozkładów prawdopodobieństwa, stosowanych w modelach GARCH jako rozkład zmiennej losowej, jest rozkład normalny, który w pełni opisany może być za pomocą dwóch parametrów: średniej μ (parametr położenia, ang. *location*) oraz wariancji σ^2 (parametr skali, ang. *scale*), co dla zmiennej losowej X można zapisać jako:

$$X \sim N(\mu, \sigma^2). \quad (2.39)$$

Funkcję gęstości rozkładu normalnego zapisać możemy jako:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \quad (2.40)$$

Rozkład normalny ma zerowy współczynnik skośności, kurtozę równą trzy, a kurtozę nadwyżkową (ang. *excess kurtosis*) równą zero. Pomimo tego, że już przy warunkowej normalności kurtoza rozkładu brzegowego jest wyższa niż w rozkładzie normalnym, często w dalszym ciągu rozkład ten okazuje się niewystarczający. Z tego względu stosowane są także inne rozkłady warunkowe, lepiej wychytujące obserwacje odstające, w szczególności rozkład t-Studenta oraz skośny rozkład t-Studenta.

Rozkład t-Studenta po raz pierwszy do modeli GARCH zastosowany został przez Bollersleva [1987]. Do oznaczenia rozkładu t-Studenta stosujemy poniższy zapis:

$$X \sim St(\mu, \sigma^2, \nu), \quad (2.41)$$

gdzie μ oznacza parametr niecentralności (modalną), $\sigma > 0$ parametr skali (odwrotność precyzji) zaś $\nu > 0$ jest liczbą stopni swobody. Funkcję gęstości tego rozkładu można zapisać jako:

$$f_{St}(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\sigma^2\nu\pi}\Gamma(\frac{\nu}{2})} \left(1 + \frac{(x-\mu)^2}{\nu\sigma^2}\right)^{-\left(\frac{\nu+1}{2}\right)}, \quad (2.42)$$

gdzie Γ oznacza funkcję Gamma. Rozkład t-Studenta ma zerowy parametr skośności przy $\nu > 3$, nadwyżkową kurtozę równą $\frac{6}{(\nu-4)}$ dla $\nu > 4$ i wariancję $Var(X) = \frac{\nu}{\nu-2}\sigma^2$ dla $\nu > 2$.⁵ Wraz ze wzrostem liczby stopni swobody, rozkład ten przyjmuje postać zbliżoną do rozkładu normalnego.

Przy warunkowo symetrycznym rozkładzie stóp zwrotu, otrzymujemy również symetryczny rozkład brzegowy, podczas gdy z empirycznych własności stóp zwrotu (asymetria) wynika, że bardziej adekwatny do ich opisu jest rozkład skośny. W tym celu stosuje się skośny rozkład t-studenta, który uzyskać można za pomocą metody zaproponowanej przez Fernandez i Steela [1998]. Do oznaczenia skośnego rozkładu t-Studenta można zastosować zapis:

$$X \sim sSt(\mu, \sigma^2, \nu, \xi), \quad (2.43)$$

gdzie $\xi > 0$ jest parametrem skośności (równym 1 dla symetrycznych rozkładów). Funkcję gęstości zmiennej losowej można zapisać jako:

$$f_{sSt}(x|\xi) = \frac{2}{\xi + \xi^{-1}} \left[f_{St}(\xi x) H_-(-x) + f_{St}(\xi^{-1}x) H_+(x) \right], \quad (2.44)$$

gdzie $H()$ oznacza funkcję Heaviside'a:

$$H(x) = \begin{cases} 0 & \text{dla } x < 0 \\ 1 & \text{dla } x \geq 0. \end{cases} \quad (2.45)$$

W przypadku, gdy parametr skośności $\xi \in (0, 1)$ mamy do czynienia z asymetrią lewostronną, natomiast gdy $\xi > 1$ z asymetrią prawostronną. Jak pokazuje Pipień [2000], warunkową wartość oczekiwaną i warunkową wariancję wyznaczyć można w następujący sposób:

$$E(X) = \phi\sigma, \quad (2.46)$$

$$Var(X) = (\gamma - \phi^2)\sigma^2, \quad (2.47)$$

⁵W pakiecie rugarch zaimplementowany jest standaryzowany rozkład t-Studenta.

gdzie:

$$\phi = \frac{\left(\xi^2 - \frac{1}{\xi^2}\right) 2\nu \Gamma\left(\frac{1}{2}(\nu + 1)\right)}{\left(\xi + \frac{1}{\xi}\right)(\nu - 1) \Gamma\left(\frac{1}{2}\nu\right) \sqrt{\pi\nu}}, \quad (2.48)$$

$$\gamma = \frac{\left(\xi^3 - \frac{1}{\xi^3}\right)}{\left(\xi - \frac{1}{\xi}\right)} \frac{\nu}{\nu - 2}. \quad (2.49)$$

Typy rozkładów opisane powyżej rozważane będą w części empirycznej tej pracy. Niekiedy stosowane są także inne typy rozkładów, wśród których wymienić można uogólniony rozkład błędu (ang. *Generalized Error Distribution* - GED) czy uogólniony rozkład hiperboliczny (ang. *Generalized Hyperbolic Distribution* - GHD, zob. Chorro i in. [2012], Dritsaki [2019]).

2.4.4 Uogólnienia i modyfikacje podstawowej specyfikacji procesu GARCH

Modele GARCH przez długi czas były jednym z najczęściej wykorzystywanych narzędzi w modelowaniu szeregów jednowymiarowych. Z czasem doczekały się jednak licznych rozwinięć. Poszczególne uogólnienia modelu GARCH mają na celu przede wszystkim poprawę opisu poszczególnych własności szeregów czasowych, a co za tym idzie, poprawienie jakości prognoz uzyskanych za pomocą modeli.

Jedną z często obserwowanych zależności, charakterystycznych dla rynków rozwiniętych, jest asymetryczny wpływ dodatnich i ujemnych stóp zwrotu z przeszłości na bieżący poziom wariancji warunkowej (zob. Fiszeder [2009], za którym podane są także dalsze wzory). W takich przypadkach często stosowane są uogólnione specyfikacje podstawowego modelu GARCH, takie jak progowy model GARCH (ang. *Threshold GARCH* - T-GARCH, Rabemananjara i Zakoian [1993]), wykładniczy model GARCH (ang. *Exponential GARCH* - EGARCH, Nelson [1991]) oraz GJR-GARCH (Glosten i in. [1993]). W szczególności, najczęściej wykorzystywany jest ten ostatni, w którym równanie wariancji warunkowej dla procesu GJR-GARCH(p,q) definiowane jest następująco:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^q \omega_i I_{t-i} \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (2.50)$$

gdzie

$$\begin{cases} I_{t-1} = 1, & \text{gd}y \ \varepsilon_{t-1} \leq 0, \\ I_{t-1} = 0, & \text{gd}y \ \varepsilon_{t-1} > 0, \end{cases} \quad (2.51)$$

oraz $\alpha_0 > 0$, $\alpha_i \geq 0$, $\alpha_i + \omega_i \geq 0$, dla $i = 1, \dots, q$, zaś $\beta_j > 0$ dla $j = 1, \dots, p$

Doman i Doman [2009] zwracają uwagę, że taka specyfikacja modelu umożliwia wychwycenie efektu dźwigni, poprzez oszacowanie parametru ω , którego wartość wpływa na poziom asymetrii i różnicę w uwzględnianiu zwrotów dodatnich i ujemnych na poziom zmienności (zob. także Piontek [2002]).

Wśród specyfikacji modelu z nieliniowym równaniem wariancji warunkowej często stosowany jest także model EGARCH, w którym równanie zmienności zadane jest w następujący sposób:

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^q \alpha_i \left\{ \theta z_{t-i} + \gamma \left[|z_{t-i}| - E(|z_{t-i}|) \right] \right\} + \sum_{j=1}^p \beta_j \ln \sigma_{t-j}^2, \quad (2.52)$$

przy założeniu $\alpha_1 \equiv 1$. W modelu EGARCH wariancja warunkowa przyjmuje postać logarytmiczną, dzięki czemu nie jest wymagane wprowadzanie dodatkowych ograniczeń na parametry w celu zapewnienia dodatniej wartości wariancji, jednak taka postać modelu może prowadzić do znacznych przeszacowań prognoz zmienności (zob. Fiszeder [2009]).

Postać modelu EGARCH zależy od przyjętego rozkładu warunkowego zmiennej z_{t-i} . Odchylenie zmiennej z_{t-i} od jej wartości oczekiwanej $E|z_{t-i}|$ wpływa na zmianę wariancji w zależności od kierunku i siły tego odchylenia. Przykładowo, wartość oczekiwana dla rozkładu normalnego wynosi $E|z_{t-i}| = \sqrt{\frac{2}{\pi}}$. Innymi często przyjmowanymi rozkładami dla tego modelu są także skośny rozkład t-Studenta, oraz rozkład GED, dla których postaci wartości oczekiwanej przedstawione są we wspomnianych publikacjach Piontek [2002] oraz Doman i Doman [2009].

Kolejną rozważaną w tej pracy specyfikacją jest model APARCH (ang. *Asymmetric Power ARCH model*), zaproponowany przez Dinga i in. [1993]. Równanie wariancji warunkowej w tym modelu przyjmuje postać:

$$\sigma^\delta = \alpha_0 + \sum_{i=1}^q \alpha_i \left[|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i} \right]^\delta + \sum_{j=1}^p \beta_j \sigma_{t-j}^\delta, \quad (2.53)$$

przy następujących restrykcjach nałożonych na parametry: $\delta > 0$, $-1 < \gamma_i < 1$, $i = 1, \dots, q$.

Dobierając odpowiednio parametry δ , β_j i γ_i model uzyskuje postać zbliżoną do opisanych wcześniej modeli, takich jak ARCH, GARCH i GJR-GARCH i innych (więcej w Ding [2011] oraz Ding i in. [1993]). Parametr γ_i określa tutaj wpływ efektu dźwigni - wartość dodatnia oznacza wpływ informacji pozytywnych na zmienność, natomiast parametr δ określa tzw. efekt Taylora, opisujący różnicę w autokorelacji zwrotów bezwzględnych i kwadratowych.

Dla odpowiednich wartości parametrów δ , β_j i γ_i możliwe jest uzyskanie szczególnych przypadków modelu APARCH, co jest jedną z głównych jego zalet (zob. także Doman i Doman [2009]):

- gdy $\delta = 2$, $\beta_j = 0$ ($j = 1, \dots, p$), $\gamma_i = 0$ ($i = 1, \dots, q$), model APARCH redukuje się do modelu ARCH,
- gdy $\delta = 2$, $\gamma_i = 0$ ($i = 1, \dots, q$), model APARCH redukuje się do modelu GARCH,
- gdy $\delta = 2$, model APARCH redukuje się do modelu GJR-GARCH,
- gdy $\delta = 1$, model APARCH redukuje się do modelu TARCH (zob. Zakoian [1994]),
- gdy $\beta_j = 0$ ($j = 1, \dots, p$), $\gamma_i = 0$ ($i = 1, \dots, q$), model APARCH redukuje się do nieliniowego modelu ARCH (ang. *Non-Linear ARCH* - NARCH, zob. Higgins i Bera [1992]),
- gdy $\delta = \infty$, model APARCH redukuje się do logarytmicznego modelu ARCH (ang. *Logarithmic ARCH* - Log-ARCH, zob. Higgins i Bera [1992]).

Obok omówionych powyżej specyfikacji, istnieje jeszcze także wiele innych modyfikacji podstawowego modelu GARCH, które szerzej opisane zostały m.in. w pracy Teräsvirta [2009].

Do zarzutów stawianych wobec modeli klasy GARCH można zaliczyć fakt, że ich specyfikacja posiada słabe podstawy teoretyczne (zob. Fiszeder [2009]). W podstawowych postaciach modelu napływające na bieżąco na rynek nowe informacje brane są pod uwagę tylko w równaniach stopy zwrotu, natomiast nie są uwzględniane w równaniach zmienności, która jest funkcją tylko i wyłącznie przeszłości. Można też zauważyć, że modele nie wyjaśniają w żaden sposób przyczyn zmienności, a służą jedynie jako narzędzie do opisu przebiegu jej procesu. Mimo to, modele tej klasy wciąż są jednym z najczęściej wykorzystywanych narzędzi służących do szacowania i prognozowania zmienności.

2.4.5 Estymacja modeli klasy ARMA-GARCH

Jednym z podstawowych narzędzi stosowanych do estymacji parametrów modeli ARMA-GARCH jest metoda największej wiarygodności (MNW). Metoda ta polega na wyznaczeniu analitycznej postaci funkcji wiarygodności, czyli funkcji wektora parametrów modelu dla ustalonego wektora obserwacji, a następnie maksymalizacji

względem wektora parametrów θ . W ogólności, dla obserwacji r_1, \dots, r_t logarytm funkcji wiarygodności przyjmuje postać (za Doman i Doman [2009]):

$$\ln L(\theta) = \sum_{t=1}^T \ln f(r_t | \Psi_{t-1}; \theta), \quad (2.54)$$

gdzie $f(r_t | \Psi_{t-1}; \theta)$ oznacza wartość funkcji gęstości rozkładu warunkowego dla stopy zwrotu r_t w zaobserwowanej wartości tejże stopy zwrotu. Dla modeli ARMA-GARCH, funkcja przyjmuje postać:

$$f(r_t | \Psi_{t-1}; \theta) = \frac{1}{\sigma(\theta)} g\left(\frac{r_t - E(r_t | \Psi_{t-1}; \theta)}{\sigma(\theta)}\right), \quad (2.55)$$

gdzie g jest funkcją gęstości innowacji z_t . Estymator MNW wektora nieznanych parametrów θ stanowi taki wektor $\hat{\theta}$, dla którego funkcja $L(\theta)$ przyjmuje wartość maksymalną:

$$\hat{\theta} = \operatorname{argmax}_{\theta \in \Theta} L(\theta) = \operatorname{argmax}_{\theta \in \Theta} \ln L(\theta). \quad (2.56)$$

Autorzy zwracają także uwagę, że jednym z problemów związanych ze stosowaniem MNW jest sytuacja, w której przyjęta funkcja gęstości rozkładu (standaryzowanego składnika losowego z_t) nie odpowiada nieznanemu prawdziwemu rozkładowi, a więc nie ma gwarancji, że estymator $\hat{\theta}$ będzie zgodny. W takim wypadku, często stosowane jest podejście nazywane metodą quasi-największej wiarygodności (ang. *Quasi Maximum Likelihood Estimator/Estimation* - QMLE) (zob. Doman i Doman [2009]).

Warto również zauważyć, że w modelach ARMA-GARCH jednocześnie estymowane są równania średniej i zmienności, co jest lepszym rozwiązaniem niż estymowanie procesów składowych osobno, ponieważ biorąc pod uwagę obie struktury jednocześnie, estymatory uzyskane za pomocą MNK dla modelu ARMA-GARCH są efektywniejsze, w stosunku do estymatorów uzyskanych poprzez oddzielne estymowanie procesów ARMA i GARCH⁶. Z drugiej strony, w celu zapewnienia kowariancyjnej stacjonarności dla procesu ARMA-GARCH wystarczy koniunkcja warunku kowariancyjnej stacjonarności osobno dla struktury ARMA i tego dla struktury GARCH.

Estymacja parametrów z wykorzystaniem przedstawionych powyżej metod przypisuje tym parametrom pewne określone wartości punktowe. Odmiernym podejściem jest wnioskowanie bayesowskie, w którym brana pod uwagę jest

⁶W przypadku estymowania samej struktury ARMA, estymatory uzyskane w ten sposób są zgodne, asymptotycznie nieobciążone, ale nieefektywne.

niepewność dotycząca parametrów, opisana rozkładem prawdopodobieństwa, a nie tylko wartością punktową (zob. Osiewalski [2001], Pipień [2006], Bauwens i in. [2000]).

W celu wyznaczenia wartości rzędów opóźnień dla modeli GARCH (lub ARMA-GARCH) stosuje się te same kryteria informacyjne, co w przypadku modeli ARMA – zwykle kryterium informacyjne Akaike lub kryterium Schwarza. Kryteria te przedstawione zostały w punkcie 2.4.2

2.4.6 Alternatywne specyfikacje modeli zmienności

Innym, często wykorzystywanym podejściem do modelowania zmienności stóp zwrotu są modele oparte na zmienności stochastycznej (ang. *Stochastic Volatility - SV*), zaproponowane przez P.K. Clarka [1973] oraz S. Taylora [1986]. Modele tej klasy wywodzą się z matematyki finansowej. Jednym z założeń modeli SV, odróżniających je od modeli klasy GARCH jest, że to bieżąca wartości wariancji warunkowej nie jest tylko funkcją przeszłości, ale oddziałuje na nią także dodatkowe, równoczesne zaburzenie losowe. Podstawową postać procesu SV opisać można równaniami (zob. Piontek [2002], Pajor [2003]):

$$\varepsilon_t = \sigma_t z_t, \quad (2.57)$$

$$\ln \sigma_t^2 = \phi_0 + \phi \ln \sigma_{t-1}^2 + \sigma_\eta \eta_t, \quad (2.58)$$

gdzie, z_t, η_t mają standaryzowane rozkłady normalny ($N(0, 1)$) i są niezależne. Z równania (2.58) wynika, że σ_t^2 zależy także od innowacji η_t , które określać mogą zaburzenia intensywności napływu informacji, natomiast szoki ε_t są związane z treścią tych informacji (zob. Doman i Doman [2009]). Modele SV, dzięki wprowadzeniu dodatkowego źródła losowości η_t , są w praktyce bardziej elastyczne i łatwiej dopasowują się do danych, natomiast ze względu na trudniejszą estymację wciąż są rzadziej stosowane niż modele GARCH. Szerokie opracowanie modeli zmienności stochastycznej przedstawione zostało w pracach A. Pajor ([2003], [2010]).

W ekonometrii finansowej prowadzone są także badania nad modelami hybrydowymi, łączącymi możliwości modeli GARCH i SV, zarówno jedno-, jak i wielowymiarowych (zob. Osiewalski i in. [2004], Osiewalski i Pajor [2018]). Jak wykazują Pajor i Wróblewska [2017] oraz Wróblewska i Pajor [2019] stosowanie (wielowymiarowych) modeli hybrydowych jest szczególnie opłacalne w kontekście prognozowania, gdyż prognozy uzyskane przez modele hybrydowe są lepsze niż te uzyskane przez modele GARCH czy SV.

2.4.7 Wielowymiarowe uogólnienia modeli ARMA-GARCH i ARMA-SV-GARCH

Opisane powyżej narzędzia służą do modelowania procesów jednowymiarowych, w związku z czym ich zastosowanie jest ograniczone przy takich zagadnieniach wielowymiarowych jak analiza portfelowa. W związku zaproponowane zostały metody pozwalające na modelowanie warunkowych kowariancji, zmienności całego portfela lub jego poszczególnych składowych oraz ich ryzyka. Jedną z najczęściej wykorzystywanych w tym celu klas modeli są wielowymiarowe procesy GARCH (ang. *Multivariate GARCH - MGARCH*).

Ponieważ modele te nie są przedmiotem tej pracy, a ich formalne przedstawienie wymagało by szerokiego wstępu, poniżej wymienione zostaną jedynie najpopularniejsze postaci wielowymiarowych modeli GARCH, wraz z odniesieniami do literatury. Zaznaczyć należy natomiast, że modele te cieszą się obecnie dużą popularnością wśród badaczy, wiąż prowadzone są liczne badania poświęcone tej tematyce, co objawia się dużą ilością nowych publikacji w tym temacie (zob. Fiszeder [2009], Pajor [2010], Francq i Zakoian [2019]).

Jedną z ogólniejszych postaci wielowymiarowych modeli GARCH jest model VECM (alternatywnie oznaczany przez VEC) zaproponowany przez D. F. Krafta i R. F. Engla w 1983 roku (Engle i Kraft [1983]). Jednak ze względu na dużą ilość parametrów i trudności z formułowaniem warunków dodatniej określoności macierzy kowariancji warunkowej (H_t) model ten rzadko znajduje zastosowanie w praktyce. Problem ilości parametrów poprawiony został przez Bollersleva [1988] poprzez sformułowanie diagonalnego modelu VECM (ang. *Diagonal VECM - DVECM*), który jednak wciąż nie gwarantuje dodatniej określoności macierzy. Kolejną propozycją był model BEKK, zaproponowany przez Baba, Engle'a, Krafta i Kronera w 1990 roku (zob. Engle i Kroner [1995]). Przedstawiona przez nich postać modelu zapewnia dodatniość macierzy warunkowych kowariancji, jednak wciąż cechuje się relatywnie wysoką liczbą parametrów. Engle [2002] oraz Tse i Tsui [2000] zaproponowali model opisujący dynamiczne korelacje warunkowe (ang. *Dynamic Conditional Correlation - DCC*). W modelu tym zakłada się, że korelacje warunkowe mogą zmieniać się w czasie. Modele te opisują korelacje warunkowe bezpośrednio, posiadają mniej parametrów i umożliwiają estymację parametrów w dwóch krokach: w pierwszym kroku dokonywana jest estymacja dowolnej specyfikacji jednowymiarowych modeli GARCH, a w drugim wyznacza się korelacje warunkowe na podstawie reszt. Model DCC jest jednym z najpopularniejszych stosowanych obecnie wielowymiarowych modeli zmienności.

Badania nad nowymi postaciami modeli wielowymiarowych wciąż są silnie rozwijane. Proponowane w tym nurcie są także wielowymiarowe modele MSV, modele hybrydowe MSV-MGARCH oraz ich uogólnienia (zob. Osiewalski i in. [2004], Silvennoinen i Teräsvirta [2009], Pajor i Osiewalski [2010], Osiewalski i Pajor [2018]).

2.5 Prognozowanie stóp zwrotu i ich zmienności w ramach modeli klasy ARMA-GARCH

Prognozowanie stóp zwrotu i ich zmienności jest jednym z głównych przedmiotów tej pracy. Stosując w tym celu modele ARMA-GARCH można uzyskać prognozy zarówno warunkowej wartości oczekiwanej jak i warunkowej wariancji. Modele te można wykorzystywać do otrzymania prognoz jednookresowych, jak i długookresowych. Ponieważ w części empirycznej pracy skupiać się będziemy wyłącznie na prognozach jednookresowych, właśnie temu podejściu poświęcony będzie poniższy podrozdział.⁷

2.5.1 Predykcja punktowa

Prognozowanie punktowe stóp zwrotu

Prognoza punktowa stóp zwrotu w modelach ARMA-GARCH przebiega tak samo, jak w zwykłych modelach ARMA, w których, jak wskazano wcześniej:

$$r_t = a_0 + \sum_{i=1}^k a_i r_{t-i} + \varepsilon_t - \sum_{j=1}^m b_j \varepsilon_{t-j}, \quad (2.59)$$

Optymalną prognozą punktową na jeden okres do przodu w przypadku tego modelu, jest:

$$r_{t+1|t} = E(r_{t+1} | \Psi_t). \quad (2.60)$$

Przyjmując określony rozkład warunkowy, wartość oczekiwana, może przyjmować różną postać:

⁷Prognozy długookresowe dla modeli ARMA oraz GARCH szczegółowo opisane są w pracy Doman i Doman [2009]

$$r_{t+1|t} = E(r_{t+1}|\Psi_t) = \begin{cases} \mu_{t+1} & \text{gdy } z_t \sim N(0, 1) \\ \mu_{t+1} & \text{gdy } z_t \sim St(0, 1, \nu), \nu > 1 \\ \mu_{t+1} + \phi\sigma & \text{gdy } z_t \sim sSt(0, 1, \nu, \xi), \nu > 1 \end{cases} \quad (2.61)$$

gdzie ϕ zadane jest wzorem (2.48).

Prognozowanie punktowe zmienności

Punktowa prognoza zmienności w modelach ARMA-GARCH jest uzyskiwana w taki sam sposób, jak w modelach GARCH, ponieważ $Var(r_{t+1}|\Psi_t) = Var(\varepsilon_{t+1}|\Psi_t)$. Przyjmuje ona następującą postać:

$$\sigma_{t+1|t}^2 = E(\sigma_{t+1}^2|\psi_t) = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t+1-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t+1-j}^2 = \sigma_{t+1}^2. \quad (2.62)$$

Prognoza wariancji warunkowej dla określonych rozkładów warunkowych przyjmuje zatem następujące postaci:

$$Var(r_{t+1}|\Psi_t) = \begin{cases} \sigma_{t+1}^2 & \text{gdy } z_t \sim N(0, 1) \\ \frac{\nu}{\nu-2} \sigma_{t+1}^2 & \text{gdy } z_t \sim St(0, 1, \nu), \nu > 2 \\ (\gamma - \phi^2) \sigma_{t+1}^2 & \text{gdy } z_t \sim sSt(0, 1, \nu, \xi), \nu > 2 \end{cases} \quad (2.63)$$

gdzie ϕ zadane jest wzorem (2.48), zaś γ wzorem (2.49).

Zauważyć należy, że wraz z wydłużeniem horyzontu maleje w tempie wykładniczym wpływ zmienności σ_{t+1}^2 na prognozy. W przypadku, gdy mamy do czynienia z procesem kowariancyjnie stacjonarnym, granicą wygasania będzie wariancją bezwarunkową zmiennej ε_t . Efekt powrotu zmienności do średniej jest jedną z własności takich procesów.

Prognozy zmienności w omawianych uprzednio uogólnieniach modelu GARCH wyznacza się w sposób analogiczny do przedstawionego powyżej. W przypadku modelu EGARCH, ze względu na logarytmiczną postać równania zmienności, konieczne jest dokonanie transformacji odwrotnej poprzez funkcję eksponencjalną, co może powodować zawyżone wyniki.

2.5.2 Mierniki trafności prognoz punktowych

Istnieje szereg miar i kryteriów, które można wykorzystać w celu oceny trafności prognoz punktowych zarówno stóp zwrotu, jak i zmienności. Zaprezentowane poniżej

miary (za Fiszeder [2009]), stosowane w dalszej części pracy, należą przede wszystkim do grupy miar symetrycznych (w których taki sam wpływ na ocenę jakości mają prognozy przeszacowanie, jak i niedoszacowane).

Do podstawowych mierników dokładności prognoz należą błąd średni (ang. *Mean Error* - ME) oraz względny błąd średni (ang. *Relative Mean Error* - RME) informujące badacza o względnym obciążeniu prognozy. W praktyce częściej stosowane są jednak takie mierniki jak średni błąd absolutny (ang. *Mean Absolute Error* - MAE), błąd średniokwadratowy (ang. *Mean Squared Error* - MSE) oraz jego pierwiastek (ang. *Root Mean Squared Error* - RMSE). Kryteria MSE oraz MAE skorygowane o heteroskedastyczność określane są odpowiednio jako HMSE (ang. *Heteroskedasticity Adjusted Mean Square Error*) oraz HMAE (ang. *Heteroskedasticity Adjusted Mean Absolute Error*). Chcąc przejść na wielkości relatywne błędów stosuje się średni absolutny błąd procentowy (ang. *Mean Absolute Percentage Error* - MAPE) oraz pierwiastek procentowego błędu średniokwadratowego (ang. *Root Mean Squared Percentage Error* - RMSPE) - które są tożsame z miarami HMAE i HRMSE. W przypadku występowania dużej liczby wartości skrajnych można stosować także logarymiczną funkcję straty (ang. *Logarithmic Loss* - LL), na którą tego typu obserwacje mają mniejszy wpływ. Postaci wymienionych powyżej mierników przedstawione zostały w tabeli 2.1.

Tabela 2.1: Miary trafności prognoz punktowych.

Miernik	Wzór
ME	$\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)$
RME	$ME / \frac{1}{n} \sum_{t=1}^n Y_t$
MAE	$\frac{1}{n} \sum_{t=1}^n Y_t - \hat{Y}_t $
MSE	$\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2$
RMSE	$\sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2}$
HMSE	$\frac{1}{n} \sum_{t=1}^n \left(\frac{Y_t - \hat{Y}_t}{Y_t} \right)^2$
MAPE, HMAE	$\frac{1}{n} \sum_{t=1}^n \left \frac{Y_t - \hat{Y}_t}{Y_t} \right $
RMSPE, HRMSE	$\sqrt{\frac{1}{n} \sum_{t=1}^n \left(\frac{Y_t - \hat{Y}_t}{Y_t} \right)^2}$
LL	$\frac{1}{n} \sum_{t=1}^n \left(\ln(Y_t) - \ln(\hat{Y}_t) \right)^2$

Uwaga: Y_t to wartość badanej zmiennej (przykładowo stopy zwrotu lub zmienności) w chwili t , natomiast \hat{Y}_t to wartość prognozy

Źródło: Opracowanie własne na podstawie Fiszeder [2009].

Dodatkową miarą, którą można wykorzystać do porównania trafności prognoz zmienności jest współczynnik determinacji R^2 , opisany wzorem

$$R^2 = \frac{\sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \geq 0, \quad (2.64)$$

gdzie Y_t jest wartością zaobserwowaną, \hat{Y}_t jest wartością prognozowaną, zaś \bar{Y} jest średnią arytmetyczną wartości zmiennej.

Przedstawione powyżej miary wykazują jednak pewne wady w praktycznym zastosowaniu i mogą prowadzić do mylnych wniosków (zob. Hyndman i Koehler

[2006]). Przykładowo, w przypadku funkcji MSE kwadraty wartości odstających mają większy wpływ na ocenę błędu niż obserwacje o niższych (co do modułu) wartościach. Miary MAPE i HMAE nie mogą być stosowane gdy w szeregu obserwacji zrealizowanych występuje zero (problem też dotyczy miar skorygowanych o heteroskedastyczność, jak HMSE). Wykazują też tendencję do silniejszego penalizowania błędów prognozy, gdy $Y_t < \hat{Y}_t$, co przy porównaniu metod prognostycznych skutkuje wyborem modelu dającego zaniżone prognozy (zob. Kim i Kim [2016], Tofallis [2014]). Współczynnik R^2 również nie penalizuje prognoz obciążonych, przez co mogą charakteryzować się one wyższą wartością współczynnika w stosunku do prognoz nieobciążonych (Fiszeder [2009]).

Ze względu na wspomniane problemy poszczególnych miar oceny prognoz, w wynikach i wnioskach z badań zaprezentowanych w dalszej części pracy, wzięte pod uwagę zostaną także dodatkowe aspekty umożliwiające szersze spojrzenie na uzyskane wyniki, takie jak strategie inwestycyjne zbudowane na podstawie punktowych prognoz zmienności, czy oszacowania poziomu ryzyka z wykorzystaniem prognoz zmienności.

Ponieważ nawet znaczne różnice pomiędzy modelami w ocenie prognoz uzyskane przez miary trafności mogą okazać się nieistotne statystycznie zaleca się dodatkowo stosowanie testów badających istotność różnicy wartości miernika uzyskanych przez dwa różne modele. W części empirycznej pracy wykorzystany zostanie test Diebolda-Mariano [1995] w wersji zmodyfikowanej przez Harveya i in. [1997]. Hipoteza zerowa testu zakłada, że wartości danego miernika uzyskane w wybranych dwóch modelach są jednakowe, natomiast hipoteza alternatywa zakłada konkretny kierunek nierówności pomiędzy modelami lub nierówność w obydwie strony. Dokładny opis procedury testowej można znaleźć we wspomnianych wyżej opracowaniach.

2.5.3 Ocena trafności predykcji punktowej stóp zwrotu w kontekście strategii inwestycyjnych

Zdolności predyktywne badanych modeli ocenione zostaną dodatkowo w kontekście strategii inwestycyjnych. Taka płaszczyzna oceny charakteryzuje się konkretnym i wymiernym aspektem trafności prognoz punktowych.

Podstawowe miary trafności prognoz nie są w stanie jednoznacznie określić, w jaki sposób zachowa się strategia inwestycyjna budowana na podstawie prognozowanych stóp zwrotu. Do oceny jakości algorytmicznych strategii inwestycyjnych (ang. *Algorithmic Investment Strategies* - AIS), wykorzystać można szereg miar wydajności

strategii (ang. *performance metrics*).⁸ Wykaz wybranych miar, zaproponowanych przez Ślepaczuk i in. [2012], oraz Kość i in. [2019] przedstawiony został w tabeli 2.2. Kluczowe przy ocenie strategii inwestycyjnych są nie tylko zannualizowane wskaźniki skumulowanego zwrotu (ang. *Annualized Return Compounded* - ARC), ale przede wszystkim wskaźniki informacyjne (ang. *Information Ratio* - IR), które oprócz zwrotu biorą pod uwagę także zannualizowane odchylenie standardowe (ang. *Annualized Standard Deviation* - ASD), maksymalną procentową wielkość obsunięcia kapitału (ang. *Maximum Drawdown* - MD) oraz długość straty z nią powiązaną (ang. *Maximum Loss Duration* - MLD), stanowiącą liczbę dni pomiędzy lokalnymi maksimumami krzywej kapitałowej. Wynika to między innymi z faktu, że na strategiach charakteryzujących się niewielkimi wskaźnikami MD i MLD, można przykładowo stosować efekt dźwigni, co znacznie wpływa na poprawę uzyskiwanych zwrotów.

Tabela 2.2: Mierniki oceny strategii inwestycyjnej

Miernik	Oznaczenie	Wzór
Skumulowany zwrot roczny	ARC	$\left(\left(\frac{P_T}{P_0} \right)^{N/T} - 1 \right) \cdot 100\%$
Roczne odchylenie standardowe	ASD	$\sqrt{\frac{N}{T} \sum_{i=0}^T (R_{t-i} - \bar{R})^2} \cdot 100\%$
Maksymalne obsunięcie kapitału	MD	$\max_{\tau \in [0, T]} \left(\max_{t \in [0, \tau]} (R_{t, T} - R_{t, \tau}) \right) \cdot 100\%$
Maksymalna długość straty	MLD	$\max \frac{m_j - m_i}{N}$
Wskaźnik informacyjny	IR*	$\frac{ARC}{ASD}$
Zmodyfikowany wskaźnik informacyjny	IR**	$IR^* \cdot ARC \cdot \frac{\text{sign}(ARC)}{MD}$
Zagregowany wskaźnik informacyjny	IR***	$\frac{ARC^3}{ASD \cdot MD \cdot MLD}$

Uwaga: P_T to końcowa cena instrumentu finansowego dla okresu $0, \dots, T$, N oznacza liczbę okresów sesyjnych w roku dla danej częstotliwości, zaś m_j, m_i to liczby dni wskazujące kolejne lokalne maksima krzywej kapitałowej.

Źródło: Ślepaczuk i in. [2012], oraz Kość i in. [2019].

Dla wskaźników ARC oraz IR preferowane są wartości wyższe.⁹ Natomiast, w przypadku wskaźników ASD, MD, MLD, preferowane są wartości niższe, które wskazują na niższe ryzyko związane z zastosowaniem badanej strategii.

Mierniki dokładności prognoz przedstawione w tabeli 2.1 często nie sprawdzają

⁸Wydańność strategii rozumiana jest tutaj jako miara jej opłacalności czy też dochodowości.

⁹Ujemne wartości IR są nieinterpretowalne.

się dobrze w roli funkcji straty¹⁰, której minimalizacja miałaby korzystnie wpłynąć na zyskowność strategii inwestycyjnych (zob. Michańków i in. [2022]). Przykładowo, funkcje MSE lub MAE mogą wskazywać bardzo małe wartości błędu, nie biorąc jednak pod uwagę kierunku przewidywanego zwrotu. Z drugiej, strony istnieje grupa miar, jak przykładowo miara MDA (ang. *Mean Directional Accuracy*), które biorą pod uwagę jedynie kierunek zmian, nie zwracając uwagi na wartość oczekiwanej korzyści lub straty. W odpowiedzi na ten problem autorzy wspomnianych badań zaproponowali funkcję MADL (ang. *Mean Absolute Directional Loss*), zadaną poniższym wzorem:

$$MADL = \frac{1}{n} \sum_{t=1}^n (-1) \cdot \text{sign}(r_t \cdot \hat{r}_{t|t-1}) \cdot |r_t|, \quad (2.65)$$

gdzie funkcja $\text{sign}(r_t \cdot \hat{r}_{t|t-1})$ zwraca znak iloczynu (wartość 1 jest zwracana gdy znak obu czynników się zgadza a -1 w sytuacji niezgodności znaku). Dzięki takiej specyfikacji, niższe wartości funkcji MADL oznaczają bezpośrednio lepsze wyniki uzyskane przez strategię bazującą na prognozowaniu kierunku zmian.

2.5.4 Wyznaczanie i ocena trafności prognoz probabilistycznych

Przedstawione w powyższych podrozdziałach metody służyły w dużej mierze do uzyskiwania szacunków i prognoz punktowych. Nieco odmiennym, a zarazem ogólniejszym podejściem jest prognozowanie całych rozkładów prawdopodobieństwa, które przedstawiają całokształt niepewności związanej z kształtowaniem się danego zjawiska w przyszłości, co wydaje się być bardziej naturalnym i kompleksowym podejściem do problematyki związanej z prognozowaniem. W ostatnich latach podejście to zyskuje na znaczeniu i coraz częściej wykorzystywane jest w pracach badawczych. Gneiting i Katzfuss [2014] zwracają uwagę na pewną zmianę paradygmatu w podejściu do prognozowania, polegającą właśnie na odchodzeniu od prognoz punktowych w stronę prognoz probabilistycznych, dzięki czemu możliwe jest lepsze uchwycenie niepewności związanej z prognozami, co z kolei może wpłynąć na poprawę podejmowanych decyzji. Z tego względu, także w tej pracy zasadnym wydaje się zastosowanie takiego podejścia. W poniższym podrozdziale opisane zostały podstawowe metody prognozowania probabilistycznego finansowych szeregów czasowych oraz sposoby oceny dokładności tych prognoz.

W tym miejscu można wspomnieć o statystyce bayesowskiej, na gruncie której to

¹⁰Rozumianej jako funkcję realizacji procesu i predyktora, reprezentującą ilościowo skutki błędu predykcji (za Fiszeder [2009]).

właśnie probabilistyczne podejście do wyznaczania prognozy — w postaci tzw. rozkładu predyktywnego — jest najbardziej naturalne. W dodatku, bayesowski rozkład predyktywny w sposób w pełni formalny ujmuje w sobie niepewność związaną z estymacją nieznanymi parametrów modelu statystycznego, jak i tę związaną z wyborem poprawnej specyfikacji modelowej (w kontekście porównywania alternatywnych modeli bayesowskich). Szersze informacje na ten temat przedstawione zostały w pracach Osiewalski [2001], Osiewalski i in. [2004].

Jednookresowe prognozy probabilistyczne z wykorzystaniem modeli ARMA-GARCH

Rozważmy szereg czasowy stóp zwrotu (r_t), modelowany za pomocą procesu ARMA-GARCH, na który składają się następujące równania:

$$r_t = \mu_t + \varepsilon_t, \quad (2.66)$$

$$\varepsilon_t = \sigma_t z_t, \quad (2.67)$$

$$(z_t) \sim iid(\eta), \quad (2.68)$$

gdzie proces (σ_t^2) ma specyfikację zadaną zgodnie z wybranym procesem klasy GARCH i jest procesem Ψ_{t-1} -mierzalnym, czyli (σ_t^2) jest funkcją tylko i wyłącznie przeszłości, zaś D jest zadanym typem rozkładu prawdopodobieństwa, o parametrach η .

Z powyższego wynika, że dla każdego $t = 1, \dots, T$ (w próbie), warunkowy rozkład bieżącej stopy zwrotu pod względem przeszłości, przy określonych wartościach parametrów, θ):

$$p(r_t | \Psi_{t-1}; \theta) = f_D^{(1)}(r_t; \theta), \quad (2.69)$$

gdzie θ zawiera w sobie parametry z równań określających μ_t , σ_t oraz rozkładu składnika losowego z_t . W szczególności, w zależności od przyjętego rozkładu warunkowego, rozkład (2.69) może przyjąć postać:

$$p(r_t | \Psi_{t-1}; \theta) = \begin{cases} f_N^{(1)}(r_t; \mu_t, \sigma_t^2), & \text{gdy } D \equiv N(0, 1) \\ f_{St}^{(1)}(r_t; \mu_t, \sigma_t^2, \nu), & \text{gdy } D \equiv St(0, 1, \nu) \\ f_{sSt}^{(1)}(r_t; \mu_t, \sigma_t^2, \nu, \xi), & \text{gdy } D \equiv sSt(0, 1, \nu, \xi) \end{cases} \quad (2.70)$$

Z 2.71 wynika, że rozkład przyszłej stopy zwrotu, r_{t+1} , pod warunkiem informacji dostępnej do chwili t wyłącznie (tj. jednookresowa prognoza probabilistyczna) przyjmuje

postać:

$$p(r_{t+1}|\Psi_t; \theta) = \begin{cases} f_N^{(1)}(r_{t+1}; \mu_{t+1}, \sigma_{t+1}^2), & \text{gdy } D \equiv N(0, 1) \\ f_{St}^{(1)}(r_{t+1}; \mu_{t+1}, \sigma_{t+1}^2, \nu), & \text{gdy } D \equiv St(0, 1, \nu) \\ f_{sSt}^{(1)}(r_{t+1}; \mu_{t+1}, \sigma_{t+1}^2, \nu, \xi), & \text{gdy } D \equiv sSt(0, 1, \nu, \xi) \end{cases} \quad (2.71)$$

Czym jest μ_{t+1} oraz σ_{t+1} zostało wyjaśnione podczas omawiania prognoz punktowych stóp zwrotu i zmienności w punkcie .

Ocena trafności prognoz probabilistycznych

Prognozy probabilistyczne wymagają stosowania odpowiednich narzędzi pozwalających na ocenę ich trafności. Poniżej przedstawiono wybrane metody oceny trafności prognoz rozkładów prawdopodobieństwa, które zastosowane zostały także w części empirycznej prezentowanej w punkcie 5.6.1.

Główną klasę narzędzi oceny prognoz probabilistycznych stanowią tzw. *proper scoring rules*.¹¹ Gneiting i in. [2007] zwracają uwagę, że miary te powinny prawidłowo oceniać zarówno kalibrację jak i ostrość (ang. *sharpness*) prognozowanego rozkładu, przy czym przez ostrość rozumiana jest tu koncentracja prognozowanego rozkładu – preferowane są prognozy o bardziej skoncentrowanym rozkładzie prawdopodobieństwa i tym samym o wyższej ostrości. Ocenic ją można zarówno na podstawie analizy graficznej, jak i numerycznej szerokości przedziałów predykcji. Oceny te są więc pewnymi wskaźnikami błędu prognozy, który powinien być minimalizowany. Są one właściwe (ang. *proper*), w przypadku kiedy przyjmują najmniejsze wartości gdy zachodzi równość $F = G$, gdzie F oznacza dystrybuantę rozkładu prognozy, a G — dystrybuantę prawdziwego rozkładu zmiennej w okresie prognozy.

Jedną z najpopularniejszych miar wykorzystywanych w ocenie trafności predykcji probabilistycznych jest miernik LPS (ang. *Log Predictive Score*, zob. Good [1952], Bernardo [1979], za Gneiting i in. [2007]), którą można opisać wzorem:

$$LPS = -\ln p(r_{t+1} | \Psi_t; \theta), \quad (2.72)$$

gdzie Ψ_t oznacza dane wykorzystane do wyznaczenia rozkładu predyktywnego w okresie t , zaś r_{t+1} jest realizacją zmiennej w okresie $t + 1$. W przypadku rozważania całej trajektorii prognoz wygasłych, oblicza się albo skumulowaną wartość LPS (ang. *Cummulative LPS* - CLPS) albo wartość średnią (ang. *Average LPS* -ALPS), zadane

¹¹Według najlepszej wiedzy autora nie istnieje polski przekład terminu „*proper scoring rules*”, wobec czego, postanowiono zachować tu jego oryginalne, angielskie brzmienie.

wzorami:

$$CLPS = - \sum_{t=1}^n \ln p(r_{t+1} | \Psi_t; \theta), \quad (2.73)$$

$$ALPS = \frac{1}{n} \cdot CLPS. \quad (2.74)$$

Do popularnych miar trafności prognoz probabilistycznych należy także miara CRPS (ang. *Continuous Ranked Probability Score*), która zaproponowana została jako alternatywa dla LPS (zob. Gneiting i Raftery [2007] i Gneiting i in. [2007]). Miarę CRPS dla obserwacji x można przedstawić za pomocą wzoru

$$CRPS(F, x) = \int_{-\infty}^{\infty} (F(y) - \mathbb{1}(y \geq x))^2 dy, \quad (2.75)$$

gdzie $\mathbb{1}$ jest funkcją charakterystyczną zbioru.

Gneiting i Raftery [2007] wykazują, że można to zapisać także jako:

$$CRPS(F, x) = E_F |X - x| - \frac{1}{2} E_F |X - X'|, \quad (2.76)$$

gdzie X oraz X' są niezależnymi kopiami zmiennej losowej o dystrybucji F i skończonym pierwszym momencie. Korzystając z takiego zapisu, niższe wartości CRPS oznaczają lepsze wyniki. Podobnie jak w przypadku miary LPS, dla ciągu prognoz wygasłych z reguły oblicza się średnią wartość CRPS.

Główne zalety tej miary wynikają z faktu, że lepiej oceniane są wartości z rozkładu predyktywnego, które są bliższe (choć niekoniecznie równe) wartościom zaobserwowanym. W stosunku do miary LPS, miara ta jest też mniej wrażliwa na obserwacje znajdujące się w ogonach rozkładu prognozy (zob. Clark i Ravazzolo [2015]). Miara CRPS jest też pewnym uogólnieniem miary MAE, wykorzystywanym w przypadkach gdy mamy do czynienia z prognozami w postaci rozkładów prawdopodobieństwa, a nie prognozami punktowymi.

Jako komplementarne podejście stosowane jest także kryterium PIT (ang. *Probability Integral Transform*), zaproponowane przez Dawida [1984] oraz Diebolda i in. [1998] jako narzędzie stosowane do oceny kalibracji rozkładów predyktywnych:

$$PIT = F(r_{t+1}). \quad (2.77)$$

PIT określa zgodność rozkładu prognozowanego z domniemanym prawdziwym rozkładem przyszłej wartości prognozowanego zjawiska G . W przypadku gdy rozkład prognozy F jest zgodny z rozkładem G , wtedy wartość PIT_{t+1} ma rozkład jednostajny,

tj. $PIT \sim U(0, 1)$).¹² Co więcej, dla ciągu prognoz probabilistycznych, wartości PIT powinny stanowić ciąg niezależnych zmiennych losowych (o tym samym, jednostajnym rozkładzie prawdopodobieństwa).

Tematyka dotycząca miar CRPS oraz PIT, a także wielu innych metod oceny trafności prognoz probabilistycznych należących do *proper scoring rules*, została szerzej przedstawiona w Gneiting i in. [2007], Gneiting i Raftery [2007]. W praktyce w dalszym ciągu jednak najpopularniejszymi sposobami oceny trafności prognoz probabilistycznych są LPS oraz CRPS.

2.6 Ocena zdolności predykcyjnych modeli w kontekście szacowania ryzyka kapitałowego

Jedną z podstawowych dziedzin, w których wykorzystywane są metody oparte na modelowaniu zmienności jest pomiar ryzyka na rynkach finansowych. Istnieje wiele definicji i klasyfikacji ryzyka, z których poniżej przedstawione zostały jedynie wybrane, związane z tematyką rynków finansowych. Knight [1921], (za Piontek [2002]) przedstawił popularną definicję ryzyka, stosując rozróżnienie między pojęciami ryzyka i niepewności. Niepewności określa jako odchylenie możliwość odchylenia od stanu oczekiwanego, która nie jest mierzalna poprzez zastosowanie rachunku prawdopodobieństwa. Z kolei ryzyko definiuje jako „*niepewność mierzalną, w której odpowiednim przyszłym stanom natury można przypisać określone prawdopodobieństwa*”. Piontek [2002] podaje także kilka definicji ryzyka finansowego, według których jest ono „*prawdopodobieństwem utraty przez organizację posiadanych zasobów finansowych oraz prawdopodobieństwem utraty środków finansowych, które są już w organizacji, oraz nieosiągnięcia spodziewanych zysków*” (definicja wg. Bizon-Górecka [2000]). Z kolei według Joriona 2006 ryzyko finansowe można określić jako „*ryzyko, które wiąże się z możliwością poniesienia strat na rynkach finansowych*”.

W dalszych rozważaniach skupiać się będziemy na ryzyku rynkowym, które wyróżniane jest jako jeden z podstawowych rodzajów ryzyka finansowego, obok przykładowo ryzyka kredytowego czy bankowego (zob. Piontek [2002]). Ogólnie, ryzyko rynkowe rozumiane jest najczęściej jako ryzyko wynikające ze zmian cen rynkowych (Małecka [2016]). Autorka przytacza kilka definicji ryzyka, funkcjonujących obecnie w literaturze. Ryzyko rynkowe może więc być rozumiane jako: „*ryzyko strat na pozycjach bilansowych i pozabilansowych, wynikających ze*

¹²Jednym z testów stosowanych w celu oceny zgodności rozkładów jest test Andersona-Darlinga, szerzej opisany w pracy Anderson i Darling [1954].

zmian cen rynkowych” (definicja wg Komitetu Bazylejskiego, zob. Basel Committee on Banking Supervision [2005]); „ryzyko straty w wyniku zmiany wartości aktywów będących przedmiotem obrotu i znajdujących się w posiadaniu przedsiębiorstwa” (definicja wg Tarczyńskiego i Mojsiewicza, zob. Tarczyński i Mojsiewicz [2001]); ryzyko ceny jako możliwość „wystąpienia niekorzystnych zmian ceny rynkowej w czasie, gdy bank zajmuje spekulacyjną lub związaną z obsługą klienta pozycję netto w towarach, stopach procentowych czy też zmienności implikowanej w opcjach” (w terminologii bankowej, zob. Riehl [2001]). W ramach ryzyka rynkowego wyróżnić można dodatkowo: ryzyko cen akcji, ryzyko kursu walutowego, ryzyko cen towarów, ryzyko stopy procentowej (zob. Jajuga [1999], Piontek [2002]). Opisane w poniższym podrozdziale miary ryzyka są istotnym narzędziem wykorzystywanym zarówno przez instytucje finansowe, jak i inwestorów.

Jajuga [1999] (za Fiszeder [2009], Piontek [2002]) wyróżnia trzy podstawowe sposoby konstruowania miar ryzyka. Pierwsza grupa opiera się na głównych miarach zmienności - wariancji i odchyleniu standardowym. Najczęściej stosowanym narzędziem są tutaj modele klasy GARCH. Do drugiej grupy zalicza się miary wrażliwości, wykorzystujące współczynniki beta. Miary te wykorzystywane są najczęściej w optymalizacji ryzyka oraz przy zabezpieczaniu portfela akcji. Trzecią grupę stanowią miary zagrożenia. Podstawowym narzędziem jest tutaj miara wartości zagrożonej VaR, która w ciągu ostatnich lat stała się jedną z najpopularniejszych miar ryzyka, oraz miara oczekiwanego niedoboru - ES (zob. Acerbi i in. [2001]). Do zalet miary VaR zaliczyć można możliwość bezpośredniego wskazania rozmiaru potencjalnych strat z określonym prawdopodobieństwem, możliwość szacowania ryzyka dla całych portfeli, oraz zastosowania w ocenie ryzyka kredytowego (zob. Małecka [2016]). W praktyce, jako poziomy tolerancji VaR zwykle przyjmuje się 5%, 2,5% lub 1% (Fiszeder [2009]), przy czym w ostatnich regulacjach Komitetu Bazylejskiego (zob. Basel Committee on Banking Supervision [2017]) zalecane jest ustalanie ich na poziomie 1%. Obok wartości zagrożonej, jako zalecana miara ryzyka, coraz częściej wskazywana jest także miara szacowanego niedoboru (zob. Małecka [2016]), stanowiąca oczekiwaną wartość straty przekraczającą poziom wartości zagrożonej.

2.6.1 Wartość zagrożona - VaR

Wartość zagrożoną (inaczej: wartość narażoną na ryzyko) zdefiniować można (za Fiszeder [2009]) jako oczekiwaną stratę wartości rynkowej instrumentu finansowego lub portfela instrumentów, taką, że prawdopodobieństwo jej osiągnięcia lub przekroczenia w danym przedziale czasowym jest równe zadanemu

prawdopodobieństwu (poziomowi tolerancji). Z kolei Małecka [2016] wskazuje, że VaR należy do grupy tzw. kwantylowych miar ryzyka, których definicje oparte są na pojęciach dolnego i górnego kwantyla zmiennej losowej X obrazującej stratę (stosując ogólną konwencję notacyjną wykorzystywaną we wskazanych pracach), oznaczanych, odpowiednio, jako $q_\alpha(X)$ i $q^\alpha(X)$, $\alpha \in (0, 1)$ (zob. także Embrechts i in. [1997])¹³:

$$q_\alpha(X) = \inf\{x \in \mathbb{R} : P(X \leq x) \geq \alpha\}, \quad (2.78)$$

$$q^\alpha(X) = \inf\{x \in \mathbb{R} : P(X \leq x) > \alpha\}. \quad (2.79)$$

Małecka [2016] oraz Artzner i in. [1999] definiują VaR na poziomie tolerancji $\alpha \in (0, 1)$ zmiennej losowej X jako

$$VaR_\alpha(X) = -q^\alpha(X), \quad (2.80)$$

co można zapisać też jako

$$VaR_\alpha(X) = \inf\{x \in \mathbb{R} : P(-X > x) \leq \alpha\}. \quad (2.81)$$

W ramach tematyki poruszanej w tej pracy, zmienną losową X należy utożsamić z prostą lub logarytmiczną stopą zwrotu. Przykładowo, w celu wyznaczenia wartości zagrożonej dla prostej stopy zwrotu (w próbie) można skorzystać z wzoru:

$$VaR_\alpha(R_t) = -q^\alpha(R_t). \quad (2.82)$$

Ponieważ przedmiotem tej pracy są stopy zwrotu w postaci logarytmicznej, w dalszych rozważaniach stosowana będzie właśnie postać logarytmiczna, pomimo, że bardziej zasadne (od strony praktycznej) jest wyznaczanie wartości zagrożonej dla prostych stóp zwrotu. Warto też wziąć pod uwagę, że w obu podejściach mogą zachodzić znaczne rozbieżności (zob. Osiewalski i Pajor [2010]).

W celu uzyskania prognoz wartości zagrożonej, oznaczanych w dalszym ciągu jako $VaR_{t+1}^{l/s}(\alpha)$ określających prognozę na jeden okres do przodu dla pozycji długiej (oznaczanej indeksem górnym l) lub krótkiej (oznaczanej indeksem górnym s), można skorzystać z wzorów:

¹³W oryginalnych pracach do oznaczenia rzędu kwantyla stosowany jest symbol p . Jednak z uwagi na konwencję typowo stosowaną w opracowaniach z zakresu ekonometrii finansowej w dalszej części pracy stosowane będzie oznaczenie α .

$$VaR_{t+1}^l(\alpha) : Pr\{r_{t+1} \leq VaR_{t+1}^l(\alpha) | \Psi_t\} = \alpha, \quad (2.83)$$

$$VaR_{t+1}^s(\alpha) : Pr\{r_{t+1} \geq VaR_{t+1}^s(\alpha) | \Psi_t\} = \alpha, \quad (2.84)$$

gdzie Ψ_t oznacza informację dostępną do chwili t włącznie. Wobec (2.83) i (2.84) VaR stanowi odpowiedni kwantyl w rozkładzie prognozowanej stopy zwrotu w okresie $t+1$, pod warunkiem wszystkich informacji dostępnych do chwili t włącznie.

Można zauważyć (za Doman i Doman [2009]), że $VaR_{t+1}^s(\alpha)$ dla instrumentu o procesie zwrotu (r_t) pokrywa się z $VaR_{t+1}^l(\alpha)$ dla instrumentu o procesie zwrotu $(-r_t)$. Z tego względu w części empirycznej pracy rozważane będą jedynie prognozy uzyskane dla pozycji długich. W związku z tym, dla uproszczenia notacji, pomijany będzie indeks górny l .¹⁴

W celu obliczenia wartości zagrożonej zdefiniowanej wzorem (2.83) można skorzystać z poniższego zapisu, który można wykorzystać w szczególności przy okazji korzystania z modeli ARMA-GARCH:

$$VaR_{t+1}(\alpha) = -r_{t+1|t} - \sigma_{t+1}q_\alpha^z, \quad (2.85)$$

gdzie $r_{t+1|t}$ oznacza prognozę punktową stopy zwrotu uzyskaną na jeden okres do przodu, σ_{t+1} – prognozę warunkowego odchylenia standardowego, zaś q_α^z oznacza α -kwantyl zmiennej losowej z_t .

Doman i Doman [2009] (s. 202) zwracają uwagę, że w powyższym sposobie wyznaczania wartości zagrożonej bardzo duże znaczenie ma jakość prognoz zmienności, która zależy od jakości przyjętego modelu zmienności. Jest to jeden z powodów, dla których to podejście – oparte na prognozach zmienności uzyskanych za pomocą modeli GARCH – stosowane będzie w dalszej części pracy do oceny prognoz zmienności, wyznaczanych właśnie w ramach modeli GARCH, jak i modeli hybrydowych, a także do oceny prognoz probabilistycznych.

2.6.2 Oczekiwany niedobór - ES

Jako uzupełnienie miary wartości zagrożonej, Acerbi i in. [2001] zaproponowali miarę ES (ang. *Expected Shortfall*), należąca do tak zwanych miar koherentnych.¹⁵ Miara ES

¹⁴Z definicji VaR^l zadanej wzorem 2.83 wynika, że jest to wartość ujemna. W literaturze można się spotkać także z odmienną konstrukcją, zgodnie z którą VaR^l powinien być wyrażony jako wartość dodatnia

¹⁵Własności miar koherentnych omówione zostały poniżej.

oznacza oczekiwaną wartość straty, pod warunkiem, że strata ta przekracza poziom wartości zagrożonej. Acerbi i in. [2001] (za Małecka [2016]), definiują ES jako

$$ES_\alpha(X) = -\frac{1}{\alpha} \left(E(X \mathbb{1}_{X \leq q^\alpha(X)}) - q^\alpha(X)(P(X \leq q^\alpha(X)) - \alpha) \right), \quad (2.86)$$

gdzie $\mathbb{1}$ oznacza funkcję charakterystyczną zbioru. Powyższe równanie można też zapisać jako:

$$ES_\alpha(X) = -\frac{1}{\alpha} \int_0^\alpha q_u(X) du. \quad (2.87)$$

Zbliżoną miarą jest warunkowa wartość zagrożona (ang. *Conditional Value-at-Risk* - CVaR) zaproponowana przez Pfluga [2000] oraz Rockafellara i Uryaseva [2000] (za Małecka [2016]):

$$CVaR_\alpha(X) = \inf \left\{ \frac{E((X - s)^-)}{\alpha} - s : s \in \mathbb{R} \right\}, \quad (2.88)$$

gdzie α to założony poziom tolerancji.

Acerbi i Tasche [2002] wykazali, że miary ES i CVaR są równe:

$$CVaR_\alpha(X) = ES_\alpha(X). \quad (2.89)$$

Podobnie jak w przypadku wyznaczania wartości zagrożonej, do prognozowania oczekiwanego niedoboru można także wykorzystać podejście parametryczne, wykorzystujące prognozy uzyskane z modeli ARMA-GARCH. W takim podejściu, prognoza na jeden okres do przodu, dla pozycji długiej uzyskana może być w następujący sposób (za McNeil i Frey [2000], Osińska i Fałdziński [2007]):

$$ES_{t+1}^l(\alpha) = E(r_{t+1} | r_{t+1} < VaR_{t+1}^l(\alpha)) = r_{t+1|t} + \sigma_{t+1} E(z_t | z_t < q_\alpha^z). \quad (2.90)$$

Miara oczekiwanego niedoboru (wraz z wartością zagrożoną) stanowi obecnie fundament pomiaru ryzyka. Spełnia też podstawowe właściwości miar koherentnych, do których należą: monotoniczność, niezmienniczość, dodatnia jednorodność oraz subaddytywność. Można także zauważyć, że pomimo iż miara VaR nie jest w ogólności miarą subaddytywną, to może taką miarą być, przy spełnionych pewnych warunkach (zob. DDanielsson [2011]). Szersze informacje na temat miar koherentnych i ich własności znaleźć można w Artzner i in. [1999], Dhaene i in. [2008], Doman i Doman [2009] oraz Małecka [2016].

2.6.3 Testy weryfikacyjne modeli ryzyka

Ze względu na rosnącą popularność miar wartości zagrożonej i oczekiwanego niedoboru w szacowaniu ryzyka, bardzo istotne jest odpowiednie zweryfikowanie uzyskanych przez nie wyników. Ponieważ miary te nie są bezpośrednio obserwowalne, konieczne jest stosowanie skonstruowanych specjalnie w tym celu testów statystycznych, których podstawą jest zwykle analiza przekroczeń. Istnieją także inne grupy testów modeli ryzyka, których szczegółowe omówienia znaleźć można w pracy Małeckiej [2016].

Ilość przekroczeń wartości zagrożonej oraz szacowanego niedoboru można testować na danych historycznych (ang. *backtesting*), porównując wyznaczone przez model wartości VaR i ES ze zrealizowanymi zwrotami. Istnieje kilka opracowanych w tym celu testów statystycznych, z których najpopularniejszy jest test Kupca (Kupiec [1995]), należący do klasycznych testów rozkładu bezwarunkowego frakcji przekroczeń VaR (lub inaczej testów bezwarunkowego pokrycia — ang. *unconditional coverage*). Hipoteza zerowa (H_0) testu zakłada, że prawdopodobieństwo przekroczeń jest równe przyjętemu poziomowi tolerancji α . Testowana wstecznie jest liczba przekroczeń VaR w stosunku do ilości obserwacji dla których wyznaczono wartości prognoz. W przypadku, kiedy występuje zbyt duża liczba przekroczeń, model niedoszacowuje ryzyko (mówimy wówczas, że prognozy ryzyka są zbyt liberalne), natomiast w sytuacji przeciwnej, liczba przekroczeń jest zbyt mała, a model przeszacowuje ryzyko (mówimy, że prognozy ryzyka są zbyt konserwatywne). Test może być stosowany zarówno dla pozycji długich, jak i krótkich, a hipoteza zerowa jest odrzucana zarówno gdy liczba przekroczeń jest zbyt mała, jak i zbyt duża. Statystyka testu ma postać (za Doman i Doman [2009]):

$$LR = 2 \left[\ln \left(\left(\frac{N_p}{n} \right)^{N_p} \left(1 - \frac{N_p}{n} \right)^{n-N_p} \right) - \ln \left(\alpha^{N_p} (1 - \alpha)^{n-N_p} \right) \right], \quad (2.91)$$

gdzie N_p oznacza w tym przypadku liczbę przekroczeń VaR, zaś n to liczba obserwacji. Jeżeli hipoteza zerowa jest prawdziwa, statystyka testowa ma asymptotyczny rozkład χ^2 o jednym stopniu swobody.

Test Kupca jednakże nie bierze pod uwagę możliwej zależności przekroczeń w czasie, czyli sytuacji kiedy przekroczenia występują seriami w krótkich odstępach, a nie są rozłożone równomiernie. W tym celu zastosować można, równie popularny, test Christoffersena [1998], należący z kolei do grupy testów warunkowego rozkładu liczby przekroczeń VaR. Kryterium testowe, nazywane własnością warunkowego pokrycia (ang. *conditional coverage*), sprawdza zarówno częstotliwość przekroczeń, jak i ich niezależność. Hipoteza zerowa (łącznego) testu zakłada, że pokrycie

warunkowe i bezwarunkowe jest równe przyjętemu poziomowi tolerancji α . Natomiast hipoteza alternatywna zakłada, że mogą występować zależności o charakterze jednorodnego łańcucha Markowa.

Obok wspomnianych testów Kupca i Christoffersena wśród sposobów testowania wstecznej wartości zagrożonej stosowane są także inne podejścia, jak przykładowo zaproponowany przez Engla i Manganelliego dynamiczny test kwantylowy (ang. *Dynamic Quantile Test*, zob. Engle i Manganelli [2004], Dumitrescu i in. [2012]), czy też przy pomocy pewnych funkcji strat, na podstawie których otrzymujemy informację odnośnie różnie definiowanych strat ekonomicznych/finansowych poniesionych z tytułu przekroczeń VaR (zob. Osiewalski i Pajor [2010]).

Testowanie wsteczne oczekiwanego niedoboru jest zdecydowanie trudniejsze niż w przypadku VaR, przez co ilość metod stosowanych w tym celu jest ograniczona (szerszy opis stosowanych metod przedstawiono w pracy Małecka [2016]). Jako jeden z popularnych testów stosowanych do oceny jakości prognoz ES uzyskanych przez model, wykorzystać można test McNeila i Fraya [2000], należący do grupy testów nieparametrycznych. Test mierzy średnią różnic pomiędzy wartościami stóp zwrotu, które przekroczyły poziom VaR, a oszacowaniami oczekiwanego niedoboru, która powinna wynosić zero (H_0) dla prawidłowo określonego modelu ryzyka. W przypadku, gdy rozważane jest podejście oparte na warunkowych resztach procesu przekroczeń VaR, statystyka testowa ma postać (za Małecka [2016]):

$$U_t = \frac{r_t - (ES_t^l(\alpha))}{\hat{\sigma}_t}, \quad (2.92)$$

dla $r_t < VaR_t^l(\alpha)$, gdzie $ES_t^l(\alpha)$ jest oszacowaną wartością ES przy założonym poziomie tolerancji α , dla pozycji długiej.

W teście dodatkowo generowana jest próba bootstrapowa, co pozwala uniknąć silnych założeń dotyczących rozkładu bazowego reszt oczekiwanego niedoboru. Wynik testu zależy jednak od wyników uzyskanych przez model VaR, ponieważ pod uwagę brane są jedynie te obserwacje, dla których odnotowano przekroczenie poziomu oszacowanego przez ten model.

Rozdział 3

Sieci neuronowe uczenia głębokiego – koncepcja i wykorzystanie w budowie modeli predykcyjnych

Analiza modeli wykorzystujących sieci neuronowe uczenia głębokiego stanowi główny cel niniejszej rozprawy. W poniższym podrozdziale przedstawiono teoretyczne wprowadzenie do zagadnień związanych z budową sieci neuronowych, a także praktyczne aspekty związane procesem ich uczenia i optymalizacji poprzez dostrajanie hiperparametrów. Informacje zawarte w tym rozdziale przedstawione zostały w sposób umożliwiający intuicyjne zrozumienie koncepcji związanych z sieciami uczenia głębokiego. Bardziej formalny opis przedstawionej poniżej problematyki przedstawiony został m.in. w opracowaniach Murphy [2012], Shalev-Shwartz i Ben-David [2014] oraz Deisenroth i in. [2020].

3.1 Wybrane architektury sieci neuronowych

W problematyce analizy danych dużą popularnością w ciągu ostatnich lat cieszą się rozwiązania wykorzystujące metody sztucznej inteligencji, a w szczególności sztuczne sieci neuronowe oraz uczenie maszynowe. Narzędzia te znajdują szerokie zastosowanie w różnych dziedzinach nauki i technologii. Dzięki coraz większej ilości dostępnych danych i łatwego dostępu do sprzętu komputerowego o wysokiej mocy obliczeniowej, stosowanie tego typu technik staje się coraz bardziej powszechne. Obszar zastosowania sztucznej inteligencji jest obecnie bardzo szeroki, jako najważniejsze przykłady wymienić można m. in. rozpoznawanie obrazów oraz pisma ręcznego, przetwarzanie mowy i budowa oprogramowania automatycznego tłumaczenia, produkcję pojazdów autonomicznych oraz pracę z dużymi zestawami

danych.

Ze względu na szeroką gamę struktur i typów sieci neuronowych, które dynamicznie rozwijane są od lat 50-tych dwudziestego wieku, w części teoretycznej i praktycznej pracy wybrane zostały trzy podstawowe architektury. Jako architekturę podstawową zaprezentowano wielowarstwowe sieci perceptronowe, opisane szerzej w punkcie 3.1.1, wykorzystaną także w części empirycznej jako narzędzie należące do klasycznych sieci uczenia maszynowego.

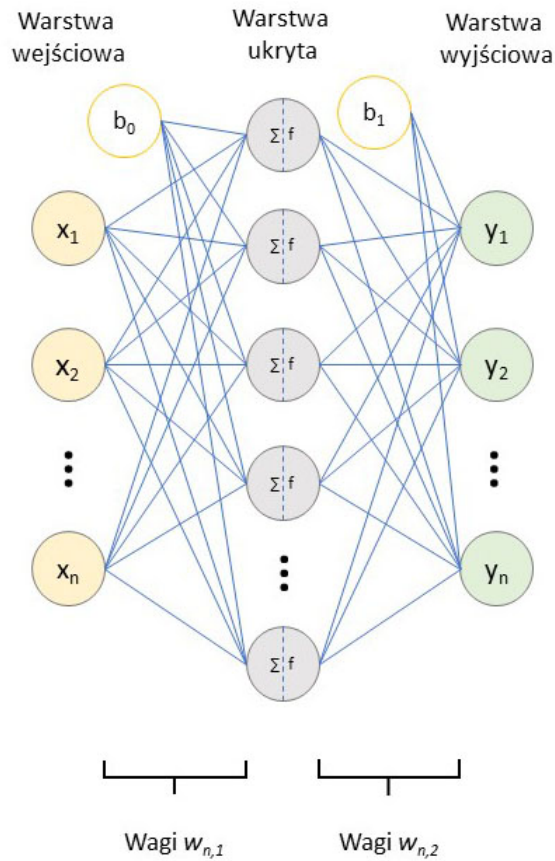
W ciągu ostatniej dekady w ramach metod opartych na sztucznej inteligencji szczególne uznanie zyskują sieci neuronowe uczenia głębokiego, które stanowią pewne rozwinięcie klasycznych sieci wielowarstwowych. Bengio [2009] i Schmidhuber [2015] definiują sieci uczenia głębokiego jako sztuczne sieci neuronowe z kilkoma warstwami ukrytymi - czyli warstwami pomiędzy warstwą wejściową i wyjściową sieci. Spośród architektur sieci uczenia głębokiego w badaniach wybrane zostały dwa typy sieci: rekurencyjne sieci LSTM oraz sieci konwolucyjne, omówione w punktach 3.1.2 oraz 3.1.3.

Zaznaczyć jednak należy, że pojęcie uczenia głębokiego nie stanowi odmiennej klasyfikacji sieci, a jedynie rozwinięcie metod klasycznego uczenia maszynowego. Szerzej odnosi się ono także do dynamicznie rozwijanego nurtu w badaniach naukowych nad sieciami neuronowymi i ich wdrożeniami praktycznymi. Obok nowych rozwiązań związanych z rozwojem struktur sieci, w ramach tego podejścia silnie rozwijane są także nowe narzędzia i metody wykorzystywane podczas uczenia sieci, czy też w procesie doboru hiperparametrów sieci. Tematyce tej poświęcone zostały podrozdziały 3.2 oraz 3.3.

3.1.1 Wielowarstwowe sieci MLP jako podstawowa forma sieci neuronowych

Jedną z podstawowych architektur sieci neuronowych są sieci wielowarstwowe - składające się z kilku warstw z gęsto połączonymi neuronami (ang. *Densely Connected Layers* lub *Fully Connected Layers*). Oznacza to, że neurony w danej warstwie są połączone ze wszystkimi neuronami w warstwie kolejnej (i poprzedniej). Jako jeden z podtypów tej sieci można uznać sieci perceptronowe¹ MLP (ang. *Multi Layer Perceptron*). Klasyczne sieci tego typu składały się najczęściej tylko z trzech warstw: wejściowej, jednej warstwy ukrytej i wyjściowej (zob. Géron [2019]).

¹Pojęcie perceptronu po raz pierwszy zaproponowane zostało w pracy Rosenblatt [1957].



Rysunek 3.1: Struktura sieci MLP.

Źródło: Opracowanie własne.

Rysunek 3.1 przedstawia strukturę podstawowej sieci MLP. Jako x_n oznaczone są wejścia sieci, $w_{n,l}$ oznaczają wagi połączeń pomiędzy neuronami sieci w warstwie l , b oznacza wartość progową czyli inaczej tzw. *bias* (najczęściej ustawiany na wartość 1 lub 0), Σ jest sumą ważoną wejść poszczególnych neuronów, f ich wybraną funkcją aktywacji, natomiast y_n oznaczają wyjścia sieci. Podstawowe perceptrony składały się z tzw. neuronów McCullocha-Pittsa (ang. *McCulloch-Pitts neuron*), nazywanych również neuronami progowymi (ang. *threshold neuron*, zob. Géron [2019]). Wartość na wyjściu takiego neuronu, o , obliczana jest następujący sposób:

$$s = b_0 + \sum_{i=0}^n x_i w_i, \quad (3.1)$$

$$o = f(s) = \begin{cases} 1 & \text{jeśli } s > 0 \\ 0 & \text{jeśli } s \leq 0 \end{cases} \quad (3.2)$$

Wartość wyjściowa neuronu jest więc ewaluowana przez funkcję aktywacji f , której wartość zależy od sumy ważonej wejść x oraz wartości wag. Pierwotnie funkcje

aktywacji miały postać progową (przybierały postać funkcji Heavisidea lub funkcji signum), w późniejszym czasie zaczęto jednak stosować także inne, liniowe oraz nieliniowe, postaci funkcji aktywacji (zob. punkt 3.3.3). Charakter połączeń między neuronami w sieciach tego typu zakłada jednokierunkowy przepływ sygnału (od warstwy wejściowej do wyjściowej).² Z czasem, rosnąca moc obliczeniowa komputerów, pozwoliła na dodawanie kolejnych warstw ukrytych, co przyczyniło się do rozwoju koncepcji uczenia głębokiego. Warstwy gęsto połączone (ang. *densely connected*) są obecnie również często stosowane jako uzupełnienie architektur sieci rekurencyjnych oraz konwolucyjnych, opisanych w kolejnych rozdziałach.

3.1.2 Rekurencyjne sieci LSTM i GRU

W pracy z danymi przedstawionymi w postaci szeregów czasowych często wykorzystywanym typem sieci są sieci rekurencyjne, charakteryzujące się występowaniem połączeń wstecznych. Ze względu na ich konstrukcję, sieci tego typu bardzo dobrze radzą sobie z modelowaniem danych sekwencyjnych. W związku z tym sieci rekurencyjne często znajdują zastosowanie m.in. w zadaniach związanych z przetwarzaniem języka naturalnego: rozpoznawaniem pisma ręcznego, automatycznymi tłumaczeniami i rozpoznawaniem mowy. Koncepcje sieci rekurencyjnych po raz pierwszy w literaturze pojawiły się w pracach Hopfielda [1982] (jako sieci Hopfielda), Rumelharta i in. [1986] oraz Jordana [1986].

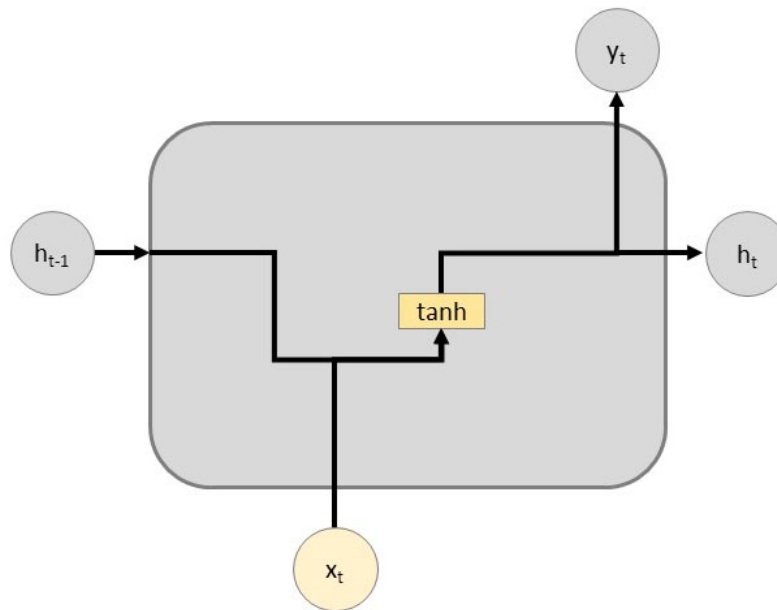
W sieciach rekurencyjnych neurony mają możliwość modelowania relacji przeszłych o dowolnym odstępem czasowym. Sygnał wejściowy w bieżącym momencie t zależy od obliczeń z okresów $t - 1$ i wcześniejszych. Można więc powiedzieć, że posiadają pewną pamięć, w której przechowywane są informacje o poprzednich obliczeniach, dzięki której mogą zachowywać informacje o parametrach (wagach) na całej długości sekwencji. Wzór 3.3 opisuje rekurencyjne połączenia między neuronami (zob. Goodfellow i in. [2016]):

$$h_t = f(h_{t-1}, x_t; \omega). \quad (3.3)$$

gdzie h_t oznacza tzw. ukryty stan pamięci sieci (ang. *hidden state*), x_t jest wektorem wejściowym sieci, a ω jest wektorem parametrów sieci, na który składają się wagi, aktywacje i wartości progowe (inaczej obciążenia, ang. *bias*) sieci. Schemat pojedynczej

²Sieć MLP z jedną warstwą ukrytą wykorzystywana jest w części empirycznej pracy jako przykład klasycznej sieci uczenia maszynowego, choć w jej architekturze wykorzystane zostały także liczne usprawnienia powstałe w wyniku rozwoju prac nad uczeniem głębokim, takie jak nowe typy funkcji aktywacji oraz sposoby regularyzacji.

komórki podstawowej sieci rekurencyjnej przedstawiony został na rysunku 3.2.



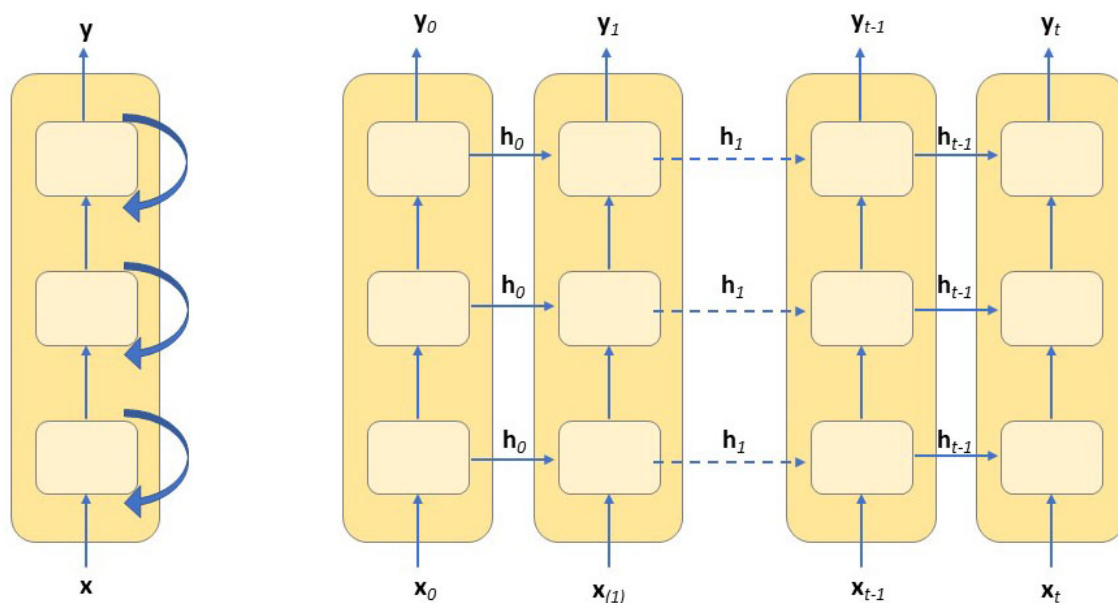
Rysunek 3.2: Struktura podstawowej komórki w sieci rekurencyjnej.

Źródło: Opracowanie własne.

Wektory wejściowe, x_t , oraz wyjściowe, y_t , w sieciach rekurencyjnych mogą być różnej długości, co pozwala modelować zależności typu sekwencja-wektor, wektor-sekwencja lub sekwencja-sekwencja, szczególnie w sieciach typu enkoder-dekoder (zob. podrozdział 3.1.4).

Sieci rekurencyjne w postaci głębokiej składają się najczęściej z kilku nałożonych na siebie warstw rekurencyjnych (zob. Hihi i Bengio [1995], Pascanu i in. [2014], Géron [2019]). Struktura takiej postaci sieci przedstawiona została na rys. 3.3.

Obok klasycznych sieci rekurencyjnych szczególną popularnością w ostatnich latach cieszą się ich rozwinięcia, w postaci sieci LSTM oraz GRU, przedstawione w kolejnych podrozdziałach.

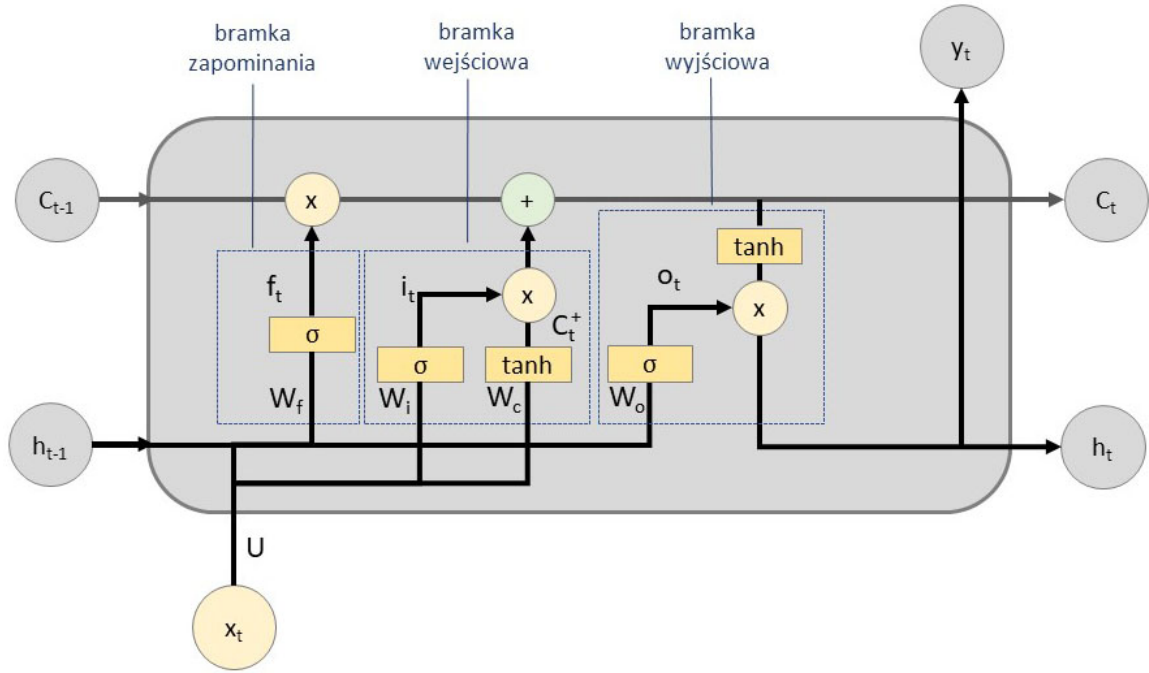


Rysunek 3.3: Struktura wielowarstwowych (głębokich) sieci rekurencyjnych.
 Źródło: Opracowanie własne na podstawie Géron [2019].

Sieci LSTM

Podczas pracy z sieciami rekurencyjnymi zauważono, że dla obserwacji z okresów odległych w czasie informacje zawarte w pamięci z tych okresów stopniowo zanikają. W literaturze problem ten określany jest jako zanikający gradient (ang. *vanishing gradient*, zob. Hochreiter i in. [2001] oraz punkt 3.2.2). W celu jego rozwiązania Hochreiter i Schmidhuber [1997] zaproponowali sieci LSTM (ang. *Long Short-term Memory*). Wprowadzona w nich została dodatkowa jednostka stanu (ang. *cell state*), oznaczona jako C_t , zależna w sposób liniowy od stanu poprzedniego, C_{t-1} . Dzięki takiemu rozwiązaniu udało się wprowadzić do sieci pamięć długoterminową (ang. *long-term memory*), w której informacje o wcześniejszych obliczeniach nie zanikają, nawet przy długich rekurencjach.

Na rys. 3.4 przedstawiona została struktura pojedynczej komórki sieci LSTM. Wektor sygnału wejściowego w sieci oznaczony został jako x_t , natomiast jako h_t oznaczony został stan krótkoterminowy (ang. *short-term*), zwany także stanem ukrytym (ang. *hidden state*). Komórka składa się także z trzech typów kontrolerów bramkowych, decydujące o tym które informacje należy zachować w pamięci długoterminowej, a które mają zostać z niej usunięte. W poszczególnych bramkach stosowane są sigmoidalne funkcje aktywacji (zob. punkt 3.3.3), określające czy dana bramka powinna zostać otwarta czy zamknięta.



Rysunek 3.4: Pełna struktura komórki sieci LSTM.
 Źródło: Opracowanie własne na podstawie Matsumoto [2020].

W pierwszej kolejności, informacje w postaci wektorów x_t i h_{t-1} przechodzą przez bramkę zapominania (ang. *forget gate*), w której przekształcane są z wykorzystaniem następujących wzorów:

$$f_t = \sigma(U_f x_t + W_f h_{t-1} + b_f), \quad (3.4)$$

$$C'_t = f_t \circ C_{t-1}, \quad (3.5)$$

gdzie b, U, W oznaczają kolejno obciążenia, wagi wejściowe i wagi rekurencyjne komórki sieci, natomiast C'_t definiuje obecny stan komórki.

Bramka zapominania, za pomocą funkcji sigmoidalnej, kontroluje ile informacji, pochodzących z bieżącego sygnału wejściowego x_t oraz pamięci krótkoterminowej h_{t-1} , powinno zostać odrzuconych z pamięci długoterminowej. Wynikiem tej funkcji jest liczba z zakresu pomiędzy 0 a 1 – wartości bliższe zera oznaczają, że większa część informacji powinna zostać odrzucona, a bliższe 1 wskazują, że większa część informacji powinna zostać przekazana dalej, do długoterminowego stanu komórki C'_t .

Kolejną bramką, przez którą przechodzą informacje w strukturze komórki sieci LSTM, jest bramka wejściowa (ang. *input gate*). Podobnie jak w poprzednim kroku, do tej bramki także przekazywane są informacje z wektorów h_t (stan krótkoterminowy) i x_t (sygnał wejściowy z bieżącego okresu). Przy pomocy funkcji

sigmoidalnej (wzór 3.6), bramka ta decyduje o istotności informacji które mają zostać przekazane do uaktualnionego stanu komórki C_t . Drugą funkcją aktywacji występującą w tej bramce jest funkcja tangensu hiperbolicznego, zadana wzorem (3.7), która przyjmuje te same informacje z wektorów x_t i h_{t-1} i przekształca je do wyniku w postaci liczby z przedziału między -1 a 1. W kolejnym kroku (wzór 3.8) wartości wyjściowe funkcji sigmoidalnej i tangens przekazywane są do aktualnego stanu komórki:

$$i_t = \sigma(U_i x_t + W_i h_{t-1} + b_i), \quad (3.6)$$

$$C_t^+ = \tanh(U_c x_t + W_c h_{t-1} + b_c), \quad (3.7)$$

$$C_t = C_t' + i_t \circ C_t^+. \quad (3.8)$$

W ostatnim kroku, bramka wyjściowa (ang. *output gate*) decyduje jakie informacje powinny znaleźć się w wektorze wyjściowym komórki, h_t . W tym celu informacje z wektorów h_{t-1} i x_t przechodzą przez sigmoidalną funkcję aktywacji (wzór 3.9) a następnie mnożone są przez funkcję tangens aktualnego stanu komórki (wzór 3.10):

$$o_t = \sigma(U_o x_t + W_o h_{t-1} + b_o), \quad (3.9)$$

$$h_t = o_t \circ \tanh(C_t). \quad (3.10)$$

Utworzony w ten sposób wektor wyjściowy h_t przekazywany jest do kolejnej komórki w strukturze sieci. Ponieważ w sieciach LSTM funkcje aktywacji stosowane są tylko do poszczególnych bramek, a nie całej komórki stanu, jak w przypadku tradycyjnych sieci rekurencyjnych, nie pojawia się tutaj problem znikającego gradientu.

W sieciach rekurencyjnych, w tym w sieciach LSTM, poza podstawową funkcją aktywacji stosowana jest także dodatkowo tzw. rekurencyjna funkcja aktywacji. Podczas gdy podstawowa funkcja aktywacji odpowiada za transformację danych w stanie krótkoterminowym (ukrytym, h_t) i długoterminowym (C_t), aktywacja rekurencyjna stosowana jest wewnątrz komórki pamięci, do transformacji przepływu informacji wewnątrz poszczególnych bramek. W sieciach LSTM domyślnie jako funkcję aktywacji stosuje się funkcję tangensu hiperbolicznego, natomiast do aktywacji rekurencyjnej stosowana jest „twarda” funkcja sigmoidalna (ang. *hard sigmoid*) (zob. Courbariaux i in. [2016]).

Architekturę sieci LSTM można dodatkowo rozróżnić na bezstanową (ang. *stateless*) oraz stanową (ang. *statefull*, zob. Géron [2019]). Rozróżnienie to dotyczy

długości okresu przez jaki przechowywane są informacje w pamięci długoterminowej. W architekturze stanowej, istnieje możliwość dokładnego kontrolowania kiedy (np. co ile epok) nastąpić ma reset długoterminowego stanu sieci. Natomiast w architekturze bezstanowej, pamięć ta resetowana jest dla każdej nowej partii wsadowej (ang. *batch*). Kontrolowanie stanu długoterminowego sieci wciąż jest jednak możliwe, poprzez zwiększenie długości partii wsadowej (np. w przypadku gdy jest ona równa całemu dostępnemu zbiorowi informacji, stan pamięci długoterminowej nie będzie resetowany). Takie podejście może mieć jednak wpływ na czas uczenia sieci i zwiększenie wykorzystania dostępnych zasobów, przede wszystkim pamięci komputera.

Sieci rekurencyjne mogą być wykorzystywane także jako sieci dwukierunkowe (ang. *bidirectional networks*). Rozwiązanie to, zaproponowane przez Schustera i Paliwala [1997], dobrze sprawdza się w sytuacjach, kiedy obecna obserwacja może zależeć także od obserwacji kolejnych. W takich przypadkach sieć przetwarza sekwencje nie tylko do początku do końca, ale także wstecznie, łącząc i osadzając obie interpretacje w stanie długoterminowym. Zastosowanie to szczególnie przydatne jest w rozpoznawaniu mowy i pisma ręcznego, gdy do przewidzenia kolejnego słowa w zdaniu przydatny jest cały kontekst wokół tego słowa, a nie tylko wyrazy poprzedzające. Sieci tego typu znalazły także zastosowanie w prognozowaniu szeregów czasowych, kiedy struktura dwukierunkowa może zostać wykorzystana w procesie uczenia sieci na zbiorze treningowym (zob. Lu i in. [2021], Park i Ryu [2021]).

Sieci GRU

Architektura sieci rekurencyjnych GRU (ang. *Gated Recurrent Unit*), zaproponowana przez Cho i in. [2014], stanowi pewne uproszczenie sieci LSTM. W sieciach tego typu decyzję o odrzuceniu lub uaktualnieniu stanu pamięci podejmuje pojedynczy mechanizm, oznaczony dalej jako z_t , w przeciwieństwie do trzech osobnych bramek używanych w sieciach LSTM. Aktualizacja stanu pamięci w sieciach GRU przedstawiona jest wzorami 3.11 - 3.14 (za Goodfellow i in. [2016]):

$$z_t = \sigma(W_z x_t + U_z h_{t-1} + b_z), \quad (3.11)$$

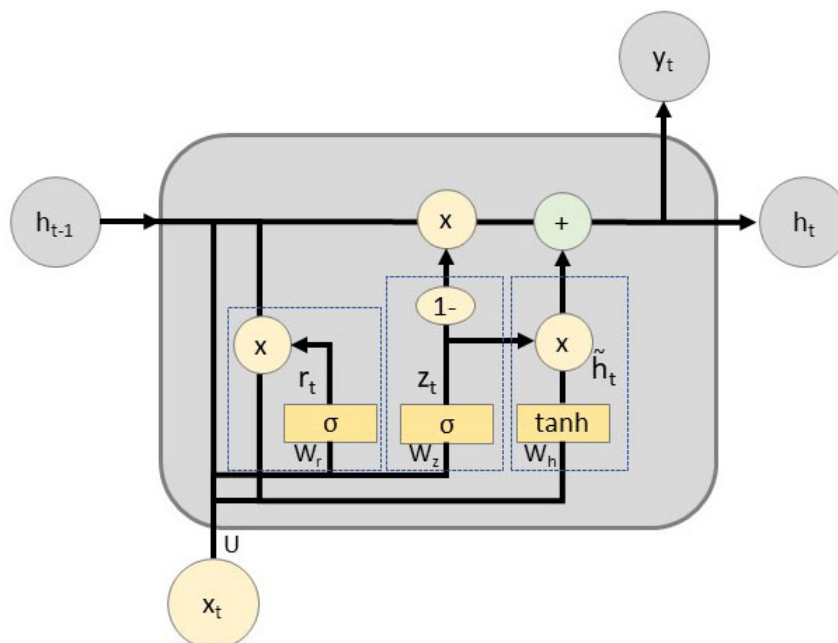
$$r_t = \sigma(W_r x_t + U_r o_{t-1} + b_r), \quad (3.12)$$

$$\tilde{h}_t = \tanh_h(W_h x_t + U_h (r_t \odot h_{t-1}) + b_h), \quad (3.13)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t, \quad (3.14)$$

gdzie z_t oraz r_t oznaczają funkcje aktualizacji stanu pamięci oraz funkcję zapominania, W oraz U to matryce parametrów, b jest wektorem obciążeń, a σ i \tanh to funkcje aktywacji (sigmoidalna oraz tangensa hiperbolicznego).

Struktura komórki w sieciach GRU przedstawiona została na rysunku 3.5.



Rysunek 3.5: Pełna struktura sieci GRU.

Źródło: Opracowanie własne.

Sieci GRU wykazują zbliżoną efektywności do sieci LSTM (zob. Greff i in. [2017]), jednak ze względu na ich prostszą budowę obliczenia zajmują często mniej czasu, przez co sieci tego typu mogą stanowić dobry wybór w badaniach wymagających dużych zasobów i mocy obliczeniowej. Spośród trzech podstawowych rodzajów sieci rekurencyjnych, sieci LSTM są wykorzystywane najczęściej i uważa się że dają najlepsze rezultaty (zob. Goodfellow i in. [2016]). Z tego względu to właśnie ta architektura sieci rekurencyjnych użyta zostanie w badaniach opisanych w części empirycznej.

3.1.3 Sieci konwolucyjne

Innym rodzajem sieci uczenia głębokiego, który znalazł szerokie zastosowanie w praktyce, są sieci konwolucyjne (ang. *Convolutional Neural Networks* - CNN), nazywane również sieciami splotowymi. Funkcja konwolucji, której idea bazuje na badaniach neurobiologicznych dotyczących kory wzrokowej, wykorzystywana była już wcześniej w matematyce i informatyce. Jedną z pierwszych prób wykorzystania tej

koncepcji w sieciach neuronowych była propozycja *neocognitronu* (zob. Fukushima [1980]). Kolejną ważną pracą w tym zakresie była publikacja Lecun i in. [1998], w której autorzy zaproponowali architekturę sieci LeNet-5, stanowiącą podstawy współczesnych sieci konwolucyjnych.

Głównym obszarem zastosowania sieci konwolucyjnych było początkowo rozpoznawanie obrazów, jednak okazało się, że ze względu na bardzo wysoką wydajność tego typu sieci, można je sukcesywnie stosować także w innych obszarach, w tym w modelowaniu szeregów czasowych.

Goodfellow i in. [2016] podstawową operację konwolucji opisują wzorem:

$$s(t) = \int x(a)w(t-a)da. \quad (3.15)$$

gdzie $x()$ oznacza funkcję sygnału wejściowego, zaś funkcja $w()$ oznacza filtr (ang. *kernel*) określający wagi połączeń neuronu, a wartości wyjściowe określane są jako mapa cech (ang. *feature map*).

W formie dyskretnej, równanie przyjmuje postać opisaną wzorem:

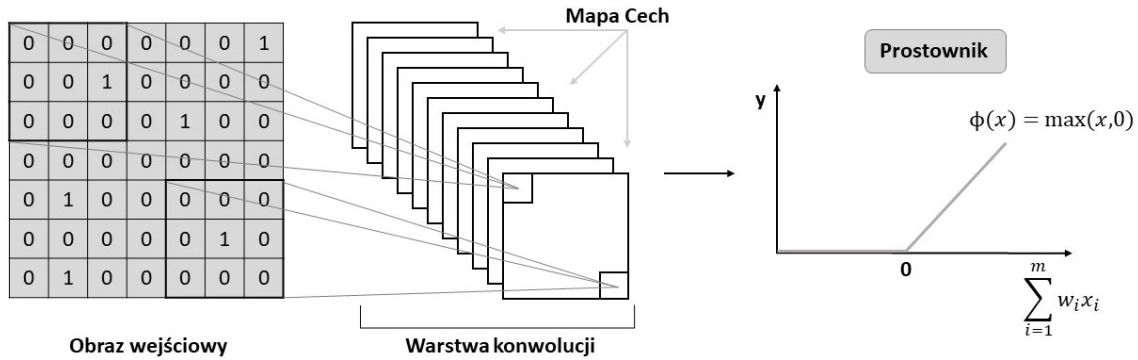
$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a). \quad (3.16)$$

W przypadku gdy sygnałem wejściowym jest dwu-wymiarowy obraz I , oraz gdy korzystamy z dwu-wymiarowego filtra K , funkcję tą można opisać wzorem:

$$S(i, j) = (K * I)(i, j) = \sum_m \sum_n I(i-m, j-n)K(m, n). \quad (3.17)$$

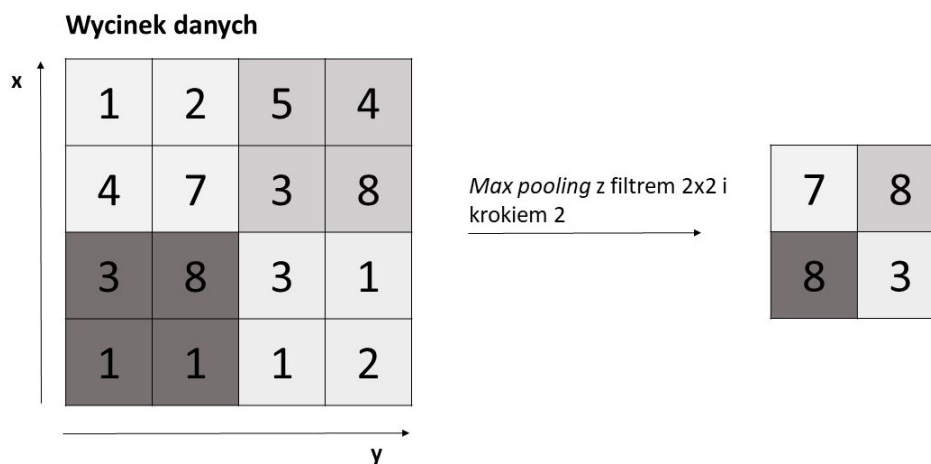
Przy projektowaniu warstwy konwolucyjnej sieci, należy ustalić przede wszystkim za ilość stosowanych filtrów, ich rozmiar, oraz wielkość kroku przesunięcia (ang. *stride*). Konieczne jest także ustalenie formy funkcji aktywacji. Przybiera ona najczęściej postać funkcji ReLU (ang. *Rectified Linear Unit*), jednak niekiedy stosowane są także inne typy funkcji takie jak *softmax* czy funkcja sigmoidalna (zob. Glorot i in. [2010], Ramachandran i in. [2017]). Zastosowanie tych funkcji ma na celu zwiększenie nieliniowości cech (lub zmniejszenie liniowości, która mogła powstać we wcześniejszej fazie konwolucji).

Głównym celem warstwy konwolucyjnej, stanowiącej najczęściej pierwszą warstwę w strukturze sieci CNN, jest redukcja ilości danych. Sieć zachowuje przy tym jedynie najważniejsze informacje tworząc w ten sposób mapę cech (ang. *feature map*, zob. Goodfellow i in. [2016]). Dzięki temu dalsze obliczenia wymagają mniejszej mocy obliczeniowej i zasobów pamięciowych, co skraca czas potrzebny na trenowanie sieci.



Rysunek 3.6: Operacja konwolucji i funkcja aktywacji w sieci CNN.
 Źródło: Opracowanie własne na podstawie superdatascience.com [2018].

Drugą warstwą najczęściej występującą w strukturze sieci konwolucyjnych jest tak zwana warstwa łącząca (ang. *pooling layer*). Warstwa ta wykorzystuje funkcję która zastępuje dane w określonej lokalizacji pewną statystyką danych z pobliskich obszarów. W ramach parametrów tej warstwy należy ustalić rozmiar pola recepcyjnego, wartość kroku (czyli wartość, o jaką to pole będzie przesuwane) oraz sposób wypełnienia zerami (ang. *zero padding*). W warstwie łączącej wybrać należy też sposób w jaki dane z poszczególnych obszarów będą redukowane. Najczęściej stosowanym podejściem jest tzw. *max pooling* (zob. Zhou i Chellappa [1988]), przedstawiony na rys. 3.7, w ramach którego, z określonego pola danych pobierane i przekazywane dalej są jedynie maksymalne wartości. Można tu stosować także inne statystyki, przykładowo pobierając z sąsiadujących obszarów sumę (*sum pooling*), średnią (*average pooling*) lub normę (*norm pooling*).



Rysunek 3.7: Funkcja *max pooling*.
 Źródło: Opracowanie własne na podstawie commons.wikimedia.com.

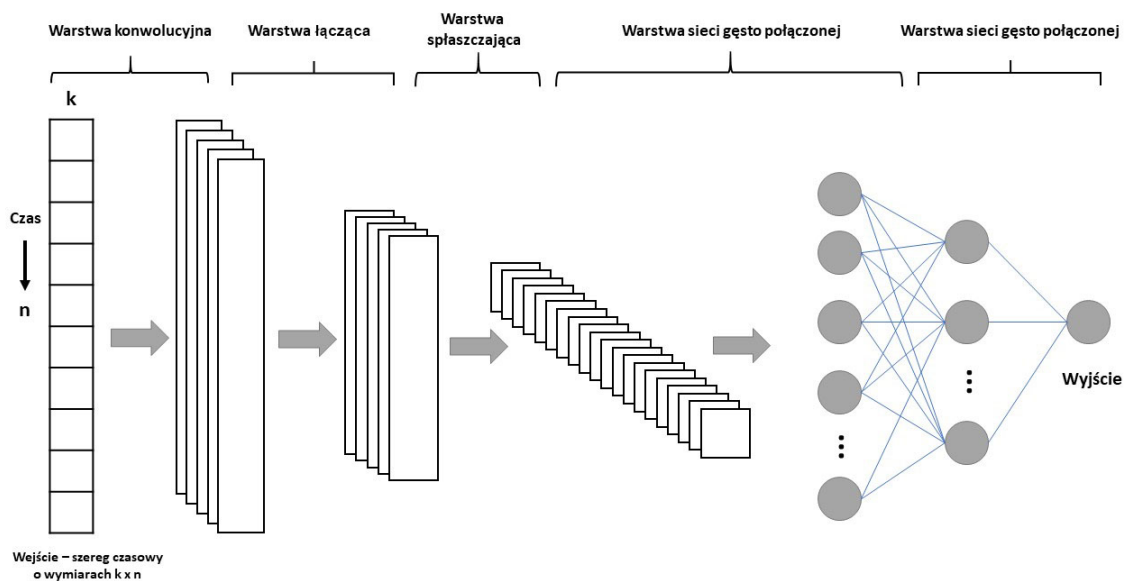
Podstawowym zadaniem warstwy łączącej jest jest dalsza redukcja ilości informacji

poprzez próbkowanie obrazu wejściowego (zob. Boureau i in. [2010], Ciresan i in. [2011]).

W kolejnych warstwach sieci konwulucyjnej często stosuje się także warstwę spłaszczającą (zob. Géron [2019]). Dane w tej warstwie przekształcone zostają do postaci wektorowej. W ten sposób łatwiej jest je przekazać do kolejnych warstw sieci, które z kolei najczęściej przybierają postać warstw gęsto połączonych (ang. *densely/fully connected layer*), w których neurony połączone są z wszystkimi neuronami z poprzedniej warstwy. Celem sieci zawartych w końcowych warstwach jest przekształcenie uzyskanej mapy cech w atrybuty, które pomocne są podczas końcowej klasyfikacji lub predykcji.

Jednowymiarowe sieci konwulucyjne i sieci TCN

Aby wykorzystać sieci CNN do pracy z szeregami czasowymi, warstwa konwulucji musi mieć postać jednowymiarową, w przeciwieństwie do klasycznej postaci dwuwymiarowej, stosowanej najczęściej gdy sygnałem wejściowym jest dwuwymiarowy obraz (zob. Duerr i in. [2020], Géron [2019]). Przykładowa struktura sieci konwulucyjnej, stosowanej do przetwarzania jednowymiarowych sekwencji przedstawiona została na rys. 3.8.



Rysunek 3.8: Struktura jednowymiarowej sieci CNN.

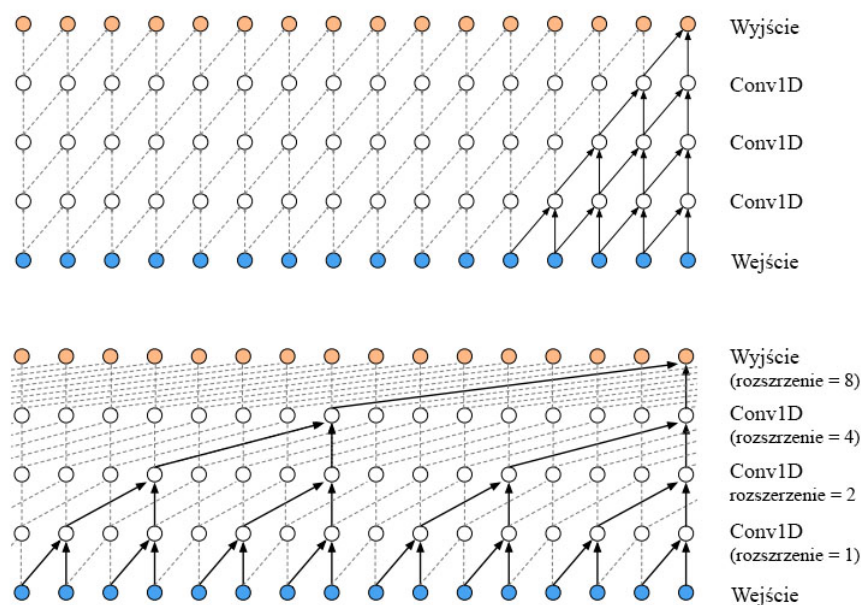
Źródło: Opracowanie własne.

W przypadku gdy dane wejściowe mają strukturę sekwencyjną, filtr w jednowymiarowej warstwie konwulucyjnej porusza się tylko w jednym kierunku — z góry na dół, wykorzystując tylko okresy przeszłe. Taka konfiguracja nazywana jest konwulucją przyczynową (ang. *causal*, zob. Duerr i in. [2020]) i może być wymuszona

poprzez odpowiednie ustawienie wartości parametru *padding* jak i odpowiednie przygotowanie sekwencji wejściowych i oczekiwanych wartości wyjściowych sieci (zob. Géron [2019]).

Początkowo idea wykorzystania sieci konwolucyjnych do przetwarzania sekwencji bazowała na koncepcji sieci typu *time-delay* (zob. Waibel i in. [1989], Lang i in. [1990]). Funkcja konwolucji pozwalała sieci na współdzielenie parametrów w kolejnych okresach czasu, jednak w podstawowej postaci sieci takie zależności są płytkie, ponieważ brana jest pod uwagę tylko niewielką liczbę obserwacji sąsiadujących wektora wejściowego (zob. Goodfellow i in. [2016]).

Podejście to rozwinięte zostało przez van der Oorda i in. [2016], w proponowanej przez nich sieci WaveNet. Architektura tej sieci składa się z wielu jednowymiarowych warstw konwolucyjnych i stosowana jest przez Autorów do generowania dźwięku. Interesującym rozwiązaniem wprowadzonym w architekturze tej sieci było zastosowanie rozszerzonych (ang. *dilated*) połączeń pomiędzy warstwami poszczególnych warstw, stopniowo zwiększając odstępów w poszczególnych warstwach. Rozwiązanie to wprowadziło możliwość obserwowania zależności długoterminowych, podobnie jak jest to możliwe w przypadku sieci rekurencyjnych, choć stosując inny mechanizm działania. Porównanie klasycznych warstw konwolucyjnych z warstwami konwolucji rozszerzonej przedstawione zostało na rys. 3.9.



Rysunek 3.9: Porównanie sieci konwolucyjnych.

Uwaga: Górna figura przedstawia sieć złożoną z wielu warstw konwolucyjnych, na dole przedstawiona została sieć z konwolucjami rozszerzonymi.

Źródło: Opracowanie własne na podstawie van den Oord i in. [2016].

Bai i in. [2018] zaproponowali architekturę sieci TCN (ang. *Temporal Convolutional Networks*), bazującą na podobnych rozwiązaniach jak sieci WaveNet. Podstawową cechą architektury sieci TCN jest wykorzystanie przyczynowych jednowymiarowych warstw konwolucyjnych przetwarzających dane wyłącznie z okresu t i wcześniejszych.³ Co więcej, warstwy ukryte w sieciach TCN mają taką samą liczbę neuronów wejściowych i wyjściowych.

Podstawową operacją wykorzystywaną w sieciach tego typu jest funkcja rozszerzonej konwolucji (ang. *dilated convolution*), która umożliwia wykładnicze zwiększanie pola recepcyjnego. Formalnie można zapisać ją jako (zob. Bai i in. [2018]):

$$F(s) = (x *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i}, \quad (3.18)$$

gdzie $x \in \mathbb{R}^n$ jest wektorem wejściowym, $f : \{0, \dots, k-1\} \rightarrow \mathbb{R}$ jest filtrem o rozmiarze k , d oznacza współczynnik rozszerzenia, natomiast $s-d \cdot i$ określa kierunek w przeszłości. W zaprezentowanych we wspomnianej pracy badaniach empirycznych wykazano, że sieci TCN lepiej radzą sobie z prognozowaniem danych sekwencyjnych, niż sieci LSTM.

Stosowanie sieci CNN w modelowaniu i prognozowaniu finansowych szeregów czasowych wciąż jest rozwijane, najczęściej wykorzystywane są w tym celu jednowymiarowe postaci sieci, lub modele łączące sieci CNN z sieciami rekurencyjnymi LSTM. Jako przykłady takich badań podać można prace Jin i in. [2020], Livieris i in. [2020], Lu i in. [2021], Tang i in. [2021], Mehtab i Sen [2022].

3.1.4 Inne typy sieci głębokich

Obok przedstawionych powyżej struktur sieci, w ramach uczenia głębokiego rozwijana jest szeroka gama innych typów sieci, często mających korzenie w proponowanych już wcześniej rozwiązaniach. W poniższym punkcie przedstawione zostały narzędzia i kierunki badań, które cieszą się obecnie dużą popularnością wśród badaczy.

Sieci generatywne i uczenie niedazorowane

Przedstawione powyżej typy sieci służą przede wszystkim do uczenia nadzorowanego (zob. punkt 3.2.1), czyli przypadków w których docelowe sygnały wyjściowe sieci (klasa lub wartość zrealizowana) są określone. Istnieją także inne typy sieci, wykorzystywane do rozwiązywania problemów uczenia nienadzorowanego

³Istnieją jednak wyjątki od tej zasady w postaci tzw. sieci *Non-causal TCN*, które wykorzystują dane z okresów przyszłych, jednak mają one ograniczone zastosowanie.

i uczenia ze wzmocnieniem (ang. *reinforcement learning*). Jednymi z podstawach typów sieci uczenia nienadzorowanego są ograniczone maszyny Boltzmana (ang. *Restricted Boltzman Machines*, zob. Smolensky [1986]). Ich zastosowanie w uczeniu głębokim spopularyzowane zostało głównie dzięki pracą Hinton [2006, 2009, 2012] w tzw. głębokich sieciach przekonań (ang. *Deep Belief Networks* - DBN), których architektura stanowi najczęściej złożenie prostszych postaci sieci takich jak maszyny Boltzmana czy autoenkodery, oraz w zaproponowanych przez Salakhutdinova i Hinton [2009] głębokich maszynach Boltzmana (ang. *Deep Boltzman Machines*).

Sieci typu GAN (*Generative Adversarial Network*)

Do klasy sieci typu GAN zalicza się sieci umożliwiające generowanie nowych danych (przykładowo obrazów, dźwięku, tekstu i innych). Prace nad sieciami generatywnymi są obecnie jednym z głównych kierunków badań dotyczących rozwoju narzędzi uczenia głębokiego. Wciąż powstają nowe rozwiązania, często wykorzystujące „klasyczne” sieci uczenia głębokiego (jak w przypadku konwolucyjnych sieci generatywnych). Przykłady takich modeli i ich zastosowań przedstawione zostały w pracach Goodfellow i in. [2014], Salimans i in. [2016], Creswell i in. [2018] oraz Yoon i in. [2019].

Autoenkodery

W zadaniach uczenia nienadzorowanego często wykorzystywane są także sieci typu autoenkoder (zob. Hinton i Zemel [1993]). Sieci tego typu stosowane są przede wszystkim do redukcji wymiarowości poprzez uczenie odpowiedniej reprezentacji danych. Proces ten nazywany jest często uczeniem reprezentacji (ang. *representation learning*) lub uczeniem cech (ang. *feature learning*).

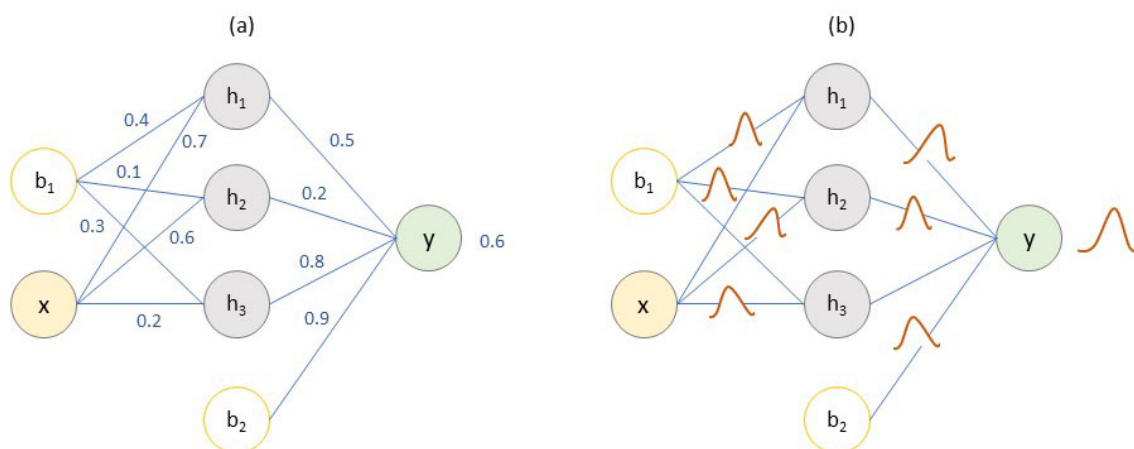
Sieci wykorzystywane w przetwarzaniu języka naturalnego

Jednym z wiodących obszarów zastosowań sieci uczenia głębokiego, obok rozpoznawania obrazów, są zadania związane z przetwarzaniem języka naturalnego, takie jak rozpoznawanie mowy, zautomatyzowane tłumaczenie i analiza sentymentu. Do takich narzędzi zaliczyć można sieci typu sekwencja do sekwencji (ang. *sequence-to-sequence*, zob. Sutskever i in. [2014]), lub enkoder-dekoder (zob. Cho i in. [2014]). Długość sekwencji wyjściowej w tych sieciach może być zmienna, i nie zależy od długości sekwencji wejściowej, co pozwala na stosowanie ich np. przy tłumaczeniu tekstu lub automatycznym odpowiadaniu na pytania. Sieci tego typu najczęściej wykorzystują warstwy rekurencyjne (RNN, LSTM lub GRU). Warstwy kodujące mają za zadanie mapowanie sekwencji wejściowej do reprezentacji wektorowej o

ustalonym lub zmiennym wymiarze (zwanej również kontekstem), która następnie wykorzystywana jest jako wejście do dekodera, generującego sekwencję wyjściową (zob. Alshemali i Kalita [2020]).

Sieci bayesowskie

W prognozowaniu probabilistycznym, coraz większą popularnością cieszą się sieci bayesowskie (ang. *Bayesian Neural Networks* - BNN), zaproponowane po raz pierwszy w pracy Neal [1996]. Wartości wag i wyjść poszczególnych neuronów w sieciach bayesowskich przyjmują postać całych rozkładów prawdopodobieństwa, podczas gdy w innych typach sieci są to najczęściej określone wartości punktowe (zob. rys. 3.10).⁴ Zdaniem sieci BNN jest znalezienie takich rozkładów tych parametrów, które najlepiej pasują do danych uczących. Celem sieci jest określenie niepewności w taki sposób, aby modele mogły przypisywać mniejsze poziomy ufności błędnym prognozom. Źródłami niepewności mogą być zarówno słabej jakości dane jak i nieprawidłowo dobrane przez model parametry. Jako metody wykorzystywane do estymacji rozkładów najczęściej wykorzystywane są metody Monte Carlo łańcuchów markowa (MCMC) oraz wnioskowanie wariacyjne (ang. *variational inference*).



Rysunek 3.10: Oceny parametrów w klasycznych sieciach neuronowych oraz w sieciach bayesowskich.

Uwaga: Figura (a) przedstawia schemat podstawowej sieci neuronowej z punktowymi ocenami parametrów, (b) przedstawia schemat sieci bayesowskiej, w której parametry określają rozkłady prawdopodobieństw.

Źródło: Opracowanie własne na podstawie Géron [2019].

Rozwiązania bazujące na wnioskowaniu bayesowskim nie są przedmiotem niniejszej pracy, jednak w kontekście prognozowania finansowych szeregów czasowych i rozkładów

⁴Podobne podejście stosowane jest także w probabilistycznych sieciach neuronowych (ang. *Probabilistic Neural Networks* - PNN, zob. Mohebbi i in. [2020]).

prawdopodobieństw z pewnością zasługują na szerszą uwagę. Problematyka szerzej opisana została w Duerr i in. [2020], Chang [2021] oraz Jospin i in. [2022].

Inne typy sieci uczenia głębokiego

Spośród pozostałych narzędzi rozwijanych w ramach uczenia głębokiego wymień w tym miejscu należy także:

- VAE - Variational Autoencoder (Kingma i Welling [2014], Kingma i Welling [2019], Rezende i in. [2014])
- Representation Learning (Bengio i in. [2014]), Transfer Learning (Bozinovski [2020]),
- Echo State Networks i Liquid State Machines (Jaeger i Haas [2004], Maass i in. [2002])
- Sieci Hopfielda (Hopfield [1982]), Maszyny Helmholtza (Dayan i in. [1995]),
- Transormery (Vaswani i in. [2017], Wolf i in. [2020]).

3.2 Podstawowe zagadnienia związane z uczeniem sieci

3.2.1 Sposoby uczenia sieci neuronowych

Proces uczenia sieci głębokich, podobnie jak w przypadku klasycznego uczenia maszynowego, przeprowadza się najczęściej na dwa sposoby (zob. Géron [2019]). Uczenie nadzorowane, z nauczycielem, (ang. *supervised learning*) wymaga danych w postaci par obiektu wejściowego (np. wektor obserwacji) i wyjściowego (oznaczone wartości pożądane do których sieć jest uczona). Na tej podstawie można określić błąd pomiędzy wartościami pożądanymi, a rzeczywistymi wygenerowanymi przez sieć i minimalizować go w procesie uczenia. Do przykładów zastosowania uczenia nadzorowanego zaliczyć można takie problemy jak regresja czy klasyfikacja. Podejście to wykorzystywane jest także w algorytmach bazujących na drzewach decyzyjnych, lasach losowych i maszynach wektorów nośnych (ang. *Support Vector Machines - SVM*, zob. Cortes i Vapnik [1995]). Jest to także główny typ uczenia wykorzystywany w badaniach przedstawionych w tej pracy.

Drugim typem uczenia jest uczenie nienadzorowane (ang. *unsupervised learning*, zob. Hinton i Sejnowski [1999]). Głównym zadaniem tego typu procesów jest

odnajdywanie nowych wzorców, bez wcześniej określonych prawidłowych wartości wyjściowych sieci, czyli np. poprawnych prognoz liczbowych lub etykiet klas przynależności danego obiektu. Uczenie nienadzorowane najczęściej stosuje się podczas analizy skupień, wykrywania anomalii, redukcji wymiarowości poprzez analizę składowych oraz przy odkrywaniu asocjacji.

Obok tych dwóch podstawowych typów stosowane są także inne sposoby uczenia sieci, takie jak uczenie częściowo nadzorowane, w których sieci otrzymują zarówno dane oznaczone jak i nieoznaczone, wykorzystywane przykładowo w sieciach DBN (ang. *Deep Belief Networks*), oraz uczenie wzmocnione (ang. *reinforced learning*, zob. Kaelbling i in. [1996]) realizowane na podstawie interakcji z określonym środowiskiem, do którego zaliczyć można przykładowo algorytmy stosowane do gry w szachy lub Go rozwijane m.in. przez zespół *DeepMind* (zob. Silver i in. [2017]), które przyczyniły się do znacznej popularyzacji uczenia głębokiego pokonując w rozgrywkach mistrzów świata obu dyscyplin.

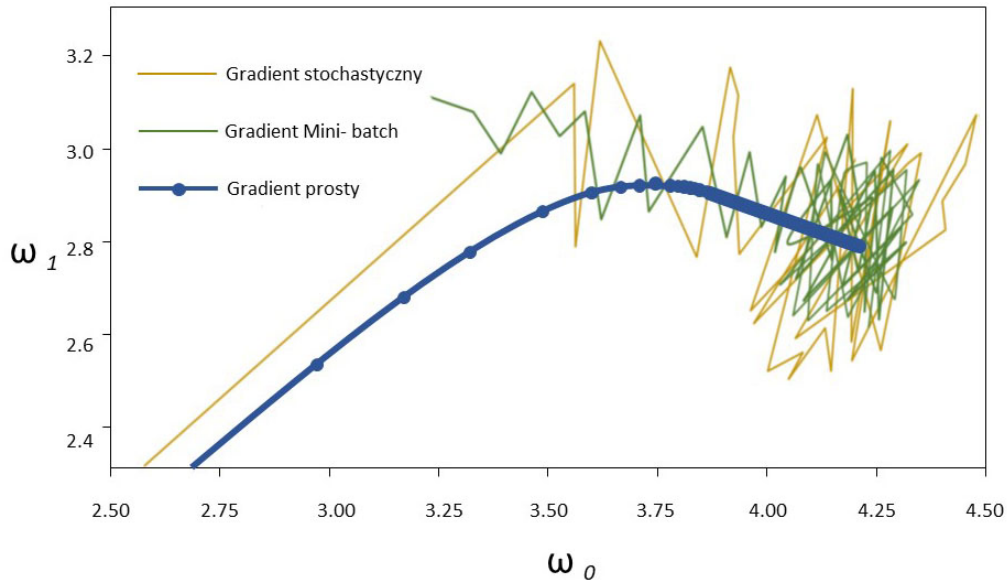
3.2.2 Algorytmy oparte na metodzie gradientu

Podstawową metodą używaną podczas minimalizacji błędu sieci, określonego przez wybraną funkcję straty, jest algorytm oparty na metodzie gradientu prostego (ang. *gradient descent*, zob. Lemaréchal [2012]). Podczas pracy algorytmu wagi poszczególnych neuronów modyfikowane są na podstawie obliczonych pochodnych zadanej funkcji straty. Wagi te modyfikowane są podczas każdej iteracji algorytmu, aż do momentu odnalezienia minimum funkcji straty. Obecnie w uczeniu głębokim stosuje się przede wszystkim metodę gradientu stochastycznego (ang. *Stochastic Gradient Descent* - SGD, zob. Bottou [1998]), w której w każdym kroku uczenia wykorzystywany jest tylko wycinek dostępnych danych, zwany wsadem lub partią wsadową (ang. *minibatch* lub *batch*) (zob. Goodfellow i in. [2016]). Dzięki temu algorytm działa znacznie szybciej i wykorzystuje mniej zasobów obliczeniowych. Niektórzy badacze dokonują dodatkowego rozróżnienia między algorytmem gradientu stochastycznego, a algorytmem gradientu Mini-Batch (zob. Géron [2019]). Długość partii wsadowej (ang. *batch size*) w poszczególnych metodach określana jest następująco:

- dla metody gradientu prostego (*Batch Gradient Descent*) *batch size* = cały zbiór uczący,
- dla metody gradientu stochastycznego (*Stochastic Gradient Descent*) *batch size* = 1,

- dla metody gradientu Mini-Batch (*Mini-Batch Gradient Descent*) $1 < \text{batch size} < \text{cały zbiór uczący}$.

Różnice w przebiegu algorytmów, w zależności od przyjętej wielkości partii wsadowej, przedstawione zostały na rysunku 3.11.



Rysunek 3.11: Wizualizacja przebiegu poszczególnych algorytmów w przestrzeni parametrów ω .

Źródło: Opracowanie własne na podstawie Géron [2019].

Modyfikację wag w procesie uczenia algorytmami opartym na optymalizacji gradientowej można przedstawić wzorem (zob. Goodfellow i in. [2016]):

$$\omega_{t+1} = \omega_t - \eta \nabla J(\omega_t), \quad (3.19)$$

gdzie ω jest wektorem parametrów (wag, obciążeń i aktywacji), η to ustalony współczynnik uczenia (ang. *learning rate*) a $J(\omega)$ to wybrana funkcja straty, natomiast t jest tutaj krokiem algorytmu. Gradient $\nabla J(\omega_t)$ w uczeniu maszynowym obliczane jest z wykorzystaniem algorytmu wstecznej propagacji (ang. *backpropagation*). W przypadku algorytmu SGD wartość η jest stała.

Jedną z głównych zalet uczenia głębokiego jest fakt, że wykorzystywany w procesie uczenia optymalizator SGD nie musi znaleźć minimum globalnego funkcji. Minima lokalne najczęściej w zupełności wystarczają do uzyskania bardzo dobrej jakości prognoz (dobrze generalizują, zob. Choromanska i in. [2015], Duerr i in. [2020]).

Algorytm wstecznej propagacji błędu

Algorytm wstecznej propagacji błędu (Rumelhart i in. [1986], Werbos [1988]) jest jedną z podstawowych metod służących do optymalizacji gradientowej podczas uczenia sieci. Umożliwia on modyfikację wag neuronów w procesie uczenia sieci na zbiorze treningowym, poprzez propagację błędu wyniku otrzymanego na wyjściu sieci w stosunku do wyniku oczekiwanego, dostarczonego wraz z danymi uczącymi. W ramach tej metody, sygnały w sieci przepływają w dwóch kierunkach: do przodu (ang. *forward propagation*), kiedy dokonywana jest predykcja, oraz wstecz (ang. *back propagation*), kiedy aktualizowane są wagi neuronów w poszczególnych warstwach. Wagi dla każdego neuronu modyfikowane są w taki sposób aby wartość błędu była zmniejszana. Z czasem powstały liczne usprawnienia podstawowego algorytmu, z których dużą popularnością cieszył się adaptacyjny algorytm wstecznej propagacji błędu z bezwładnością (ang. *momentum*) (zob. Qian [1999]). Funkcję aktualizacji parametrów w tym algorytmie można zapisać następująco:

$$\Delta\omega_t = -\eta\nabla J(\omega_t) + \beta\Delta\omega_{t-1}, \quad (3.20)$$

gdzie β jest stałym współczynnikiem bezwładności procesu uczenia.

Algorytm ADAM

ADAM (ang. *Adaptive Moment Estimation*) zaproponowany przez Kingma i Ba [2017] jest kolejnym rozszerzeniem podstawowego algorytmu SGD. Powstał jako połączenie koncepcji w proponowanych wcześniej metodach RMSProp (ang. *Root Mean Square Propagation*) oraz ADAGrad (ang. *Adaptive Gradient Algorithm*) (zob. Hinton i in. [2012b], Duchi i in. [2011]). Główną zaletą tych metod jest dynamiczne dostosowywanie parametru η , czyli prędkości uczenia sieci (ang. *learning rate*). Optymizator ADAM wykorzystuje w tym celu pierwsze i drugie momenty gradientu, tworząc wykładnicze średnie ruchome (zob. Kingma i Ba [2017], Hansen [2019]):

$$\omega_{t+1} = \omega_t - \frac{\eta \cdot \hat{m}_t}{\sqrt{\hat{v}_t + \epsilon}}, \quad (3.21)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad (3.22)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t}, \quad (3.23)$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad (3.24)$$

$$v_t = \beta_1 v_{t-1} + (1 - \beta_2) g_t^2, \quad (3.25)$$

gdzie g oznacza gradient danej partii wsadowej $g = \nabla J(\theta_{t,i})$, m, v to wykładnicze współczynniki zanikania, β to współczynniki bezwładności algorytmu, zazwyczaj ustawiane na wartości $0,9 - 0,999$, natomiast ϵ jest niewielką wartością (zazwyczaj 10^{-8}) stosowaną w celu uniknięcia dzielenia przez zero.

Powstały też liczne dodatkowe rozszerzenia algorytmu Adam, takie jak Adamax (Kingma i Ba [2017]), Nadam (Dozat [2016]) czy ND-Adam (Zhang i in. [2018c]). Pomimo dużej popularności algorytm ten nie jest jednak pozbawiony wad, wśród których najczęściej wskazuje się problemu z generalizacją (Wilson i in. [2018]).

Problem zanikającego gradientu

Jednym z problemów pojawiającym się podczas uczenia sieci algorytmami SGD jest zanikający gradient (ang. *vanishing gradient*) (zob. Hochreiter i in. [2001]). Polega on na tym, że podczas procesu uczenia gradient funkcji straty staje się coraz mniejszy wraz ze zbliżaniem się przepływu informacji do warstwy początkowych sieci (bliższych warstwy wejściowej). Z tego powodu wagi neuronów w tych warstwach, przy bardzo małym gradiencie nie są aktualizowane, przez co proces uczenia nigdy nie jest zbliżony do dobrego rozwiązania. Odwrotnym problemem, choć występującym równie często, jest gradient eksplodujący (ang. *exploding gradient*, zob. Géron [2019], Goodfellow i in. [2016]), w którym gradient przyjmuje bardzo duże wartości, przez zmiany wielkości wag sieci są bardzo duże, a uczenie staje się niestabilne. Oba problemy najczęściej pojawiają się przy modelowaniu zależności długoterminowych. Badacze wciąż sugerują nowe rozwiązania tych problemów (zob. He i in. [2015], Noel i in. [2021]), powstają nowe typy sieci które lepiej mają radzić sobie z tymi problemami, stosowane są nowe typy inicjalizacji wag oraz nowe funkcje aktywacji. Więcej informacji na ten temat przedstawionych zostało w kolejnych podrozdziałach.

3.2.3 Generalizacja i regularyzacja

Zdolność sieci do modelowania zupełnie nowych zestawów danych (danych ze zbioru testowego, lub z poza próby uczącej) nazywana jest generalizacją (zob. Géron [2019]). Błąd funkcji straty sieci na zbiorze testowym nazywa się błędem generalizacji (ang. *generalization error*) natomiast uzyskany na danych ze zbioru uczącego, błędem uczenia (ang. *training error*). W celu poprawienia generalizacji sieci stosuje się najczęściej przedstawione poniżej metody oparte na tzw. regularyzacji, przez którą rozumie się takie zmiany wprowadzane do algorytmu uczenia, które poprawiają jego możliwości

generalizacji jednocześnie nie wpływając negatywnie na błąd uczenia (Goodfellow i in. [2016]).

Formalnie, jako regulację w ujęciu sieciowym rozumieć możemy pewną modyfikację funkcji straty poprzez wprowadzenie parametr regularyzacji $\Omega(w)$:

$$\bar{J}(\omega) = J(\omega) + \alpha\Omega(w), \quad (3.26)$$

gdzie $\alpha \in [0, \infty]$ jest hiperparametrem ustalającym wagę parametru regularyzacji.

Obok przedstawionych poniżej form regularyzacji, także opisany w podrozdziale 3.3 proces dostrajania hiperparametrów stosowany jest najczęściej właśnie w celu poprawy możliwości generalizacji sieci oraz uniknięcia problemu przetrenowania.

Dropout

Bardzo istotną formą regularyzacji jest *dropout*, zaproponowany przez Hinton i in. [2012a] oraz Srivastava i in. [2014], który określa ile neuronów z poszczególnych warstw zostanie losowo pominiętych w procesie uczenia sieci.⁵ *Dropout* jest jednym z najpopularniejszych i najbardziej efektywnych sposobów regularyzacji (zob. Géron [2019]). Każdy neuron w sieci, z wyjątkiem neuronów warstwy wyjściowej, ma ustalone prawdopodobieństwo p (określane przy pomocy parametru *dropout rate*), że zostanie pominięty w bieżącym kroku w procesie uczenia, poprzez ustawienie wartości wejścia takiego neuronu na zero.

Założmy, że w sieci o liczbie L warstw ukrytych, $l \in 1, \dots, L$ określa indeks kolejnych warstw, a $l = 0$ jest warstwą wejściową sieci, $z^{(l)}$ oraz $y^{(l)}$ są wektorami wejściowymi i wyjściowymi warstwy l , $W^{(l)}$ oraz $b^{(l)}$ są wektorami wag i obciążeń warstwy l , natomiast f oznacza wybraną funkcję aktywacji. $r^{(l)}$ jest wektorem niezależnych zmiennych losowych z rozkładem Bernoulliego, które z prawdopodobieństwem p wynoszą 1, a z $1 - p$ wynoszą 0. W sieciach jednokierunkowych (typu *feed-forward*) wyjście opisanej w ten sposób warstwy możemy zapisać jako (zob. Srivastava i in. [2014]):

$$\tilde{y}_l = r_l \odot y^{(l)}, \quad (3.27)$$

$$z_{l+1} = W_{l+1}\tilde{y}_l + b_{l+1}, \quad (3.28)$$

$$y_{l+1} = f(z_{l+1}), \quad (3.29)$$

⁵Termin ten niekiedy jest tłumaczony w polskiej literaturze jako „porzucanie” (zob. Sawka [2022]), jednak w dalszej części pracy stosowane będzie jego oryginalna, anglojęzyczna nazwa.

gdzie \odot oznacza iloczyn Hadamarda.

Neurony pomijane są tylko w pojedynczym kroku uczenia, i zostają włączone ponownie w kolejnych krokach. W procesie testowania sieci, wagi poszczególnych neuronów nie są wyłączane, natomiast są skalowane według formuły $W_l^{test} = pW_l$. Wartość hiperparametru *dropout rate* zazwyczaj zwiększana jest w przypadkach dochodzi do przeuczenia (ang. *overfitting*), czyli sytuacji w których sieć wykazuje dużą efektywność na zbiorze uczącym, natomiast uzyskuje słabe wyniki na zbiorach walidacyjnym i testowym. Hiperparametr ten najczęściej ustalany jest na wartości z zakresu 0,001. do 0,5, przy czym wyższe wartości stosuje się częściej w sieciach konwolucyjnych. Ponieważ odpowiednie ustalenie tej wartości może mieć bardzo duży wpływ na uczenie sieci, często brana jest ona pod uwagę jako jedna z optymalizowanych zmiennych w procesie dostrajania hiperparametrów. Metoda regularyzacji wykorzystująca *dropout* opisana została szerzej w Srivastava i in. [2014] oraz Warde-Farley i in. [2014].

Dodatkowym sposobem regularyzacji, stosowanym w sieciach rekurencyjnych, jest *dropout* rekurencyjny (Moon i in. [2015], Gal i Ghahramani [2016] oraz Semeniuta i in. [2016]). Podczas gdy zwykły *dropout* wyłącza połączenia pomiędzy poszczególnymi warstwami sieci, *dropout* rekurencyjny wyłącza połączenia pomiędzy zależnościami rekurencyjnymi w czasie (przykładowo, pomiędzy wektorami ukrytymi h_{t-1} a h_t) podczas obliczania wartości bramek i aktualizacji w bieżącym kroku uczenia. Podobnie jak zwykły *dropout*, pełni on także funkcję regularyzacyjną i zapobiega przetrenowaniu sieci.

Regularyzacja l_1 i l_2

Innym typem regularyzacji wykorzystywanym w sieciach są metody nakładające ograniczenia na parametry modelu sieciowego, przede wszystkim z wykorzystaniem tzw. regularyzacji l_1 oraz l_2 (zob. Goodfellow i in. [2016]).

Regularyzacja l_1 , nazywana również regresją LASSO (ang. *Least Absolute Shrinkage and Selection Operator*) definiowana jest wzorem:

$$\bar{J}(\omega) = J(\omega) + \alpha \sum_i |\omega_i|, \quad (3.30)$$

gdzie ω oznacza w tym przypadku wagi sieci (obciążenia, b , nie są regularyzowane), natomiast α jest ustalonym hiperparametrem regularyzacji. Wartość regularyzacji zależy więc tylko od $sign(\omega_i)$ ⁶ i wpływa na gradient ze stałą wartością,

W regularyzacji tego typu niektóre parametry mogą przyjmować wartości zerowe

⁶Funkcja signum opisana została w rozdziale 2

i w takim wypadku są usuwane, co powoduje że znalezione rozwiązanie jest bardziej rzadkie (ang. *more sparse*). Własność ta wykorzystywana jest często w procesie wyboru zmiennych wejściowych sieci (ang. *feature selection*), gdy dzięki regularyzacji $l1$ można wskazać, które cechy (zmiennne wejściowe) mogą zostać usunięte, jeśli wartości ich wag wyniosą zero (zob. Géron [2019]).

Regularyzacja l_2 określana jest również jako regresja grzbietowa (ang. *ridge regression*), *weight decay* lub regularyzacja Tichonova, i przyjmuje postać:

$$\bar{J}(\omega) = J(\omega) + \frac{1}{2}\alpha \sum_i^n \omega_i^2. \quad (3.31)$$

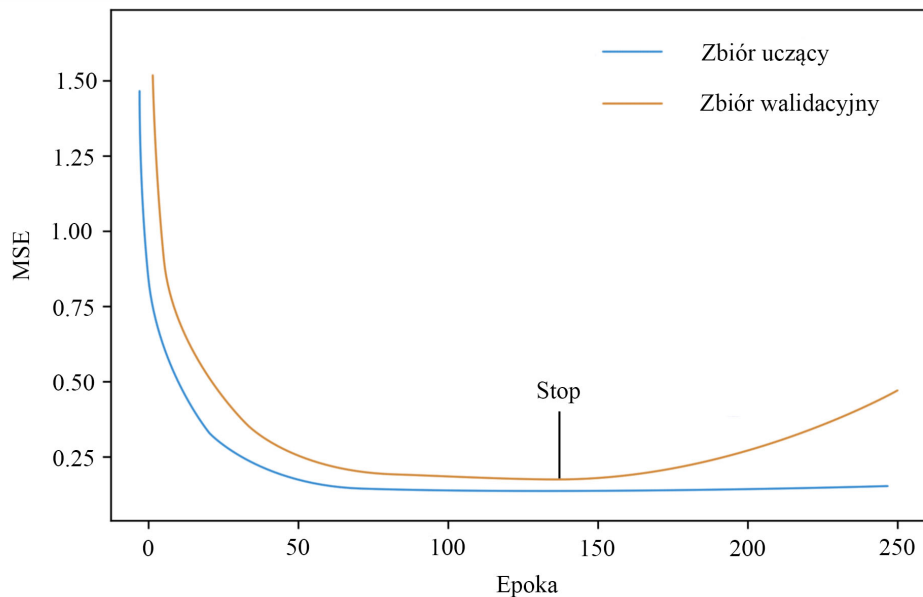
Wprowadzenie tego typu regularyzacji powoduje, że wektor wag w każdym kroku algorytmu optymalizacji gradientu jest pomniejszany o stały współczynnik (zob. Goodfellow i in. [2016]). Podobnie jak w przypadku dropoutu, regularyzacje $l1$ i $l2$ wykorzystywane są tylko w procesie uczenia sieci.

Funkcje wczesnego zatrzymania i punktu kontrolnego modelu

Pewną formą regularyzacji są też dodatkowe funkcje wywołania zwrotnego (ang. *callback function*), stosowane podczas procesu uczenia. Jedną z takich funkcji jest funkcja wczesnego zatrzymania (ang. *early stopping*, (zob. Yao i in. [2007])), umożliwiająca zatrzymanie procesu uczenia w najbardziej optymalnym momencie. Rys. 3.16 przedstawia krzywe kształtowania się wartości funkcji straty (którą przykładowo może być funkcja MSE) w procesie uczenia sieci na zbiorze uczącym i walidacyjnym. Wartość funkcji straty maleje z każdym krokiem uczenia, jednak w pewnych przypadkach, np. przy wystąpieniu efektu przeuczenia sieci, może dojść do sytuacji, gdy wartość ta maleje na zbiorze uczącym ale zaczyna rosnąć na zbiorze walidacyjnym. Moment w którym obie krzywe uzyskują najmniejszą wartość może być interpretowany jako optymalny wynik, który może uzyskać sieć, i w którym powinno nastąpić zatrzymanie uczenia (zob. Prechelt [2012]). W praktyce, ponieważ linie te nie zawsze są gładkie, algorytm pozwala sieci uczyć się jeszcze przez kilka epok, po czym następuje powrót do zapisanego wcześniej punktu, z którego pochodzą optymalne (finalne) są parametry modelu.

Na podobnej zasadzie działa funkcja punktu kontrolnego modelu (ang. *model checkpoint*, zob. Géron [2019]). W każdej epoce (kroku uczenia) sprawdzana jest wartość funkcji straty dla ustalonego zbioru (uczącego lub walidacyjnego). Jeżeli wartość ta jest mniejsza niż w epoce poprzedniej to model zapisywany jest jako najlepszy, natomiast nie jest zapisywany jeśli wartość ta jest wyższa. Zapisany model następnie może zostać wykorzystany jako optymalny model do testowania sieci.

Różnica pomiędzy funkcją wczesnego zatrzymania a funkcją punktu kontrolnego polega na tym, że w pierwszym przypadku proces uczenia jest całkowicie zatrzymywany, a w przypadku punktu kontrolnego sieć uczy się dalej (do czasu osiągnięcia ustalonej wcześniej liczby epok). Funkcja wczesnego zatrzymania pozwala więc na skrócenie czasu uczenia sieci, jednak w przypadku gdy krzywe uczenia są bardzo nierówne, optymalny wynik może zostać uzyskany znacznie później. W takich przypadkach lepiej sprawdza się funkcja punktu kontrolnego.



Rysunek 3.12: Funkcja wczesnego zatrzymania.

Źródło: Opracowanie własne.

Normalizacja wsadowa

Ioffe i Szegedy [2015] jako formę regularyzacji zaproponowali metodę zwaną normalizacją wsadową (ang. *batch normalization*). Pozwala ona na reparametryzację modelu poprzez standaryzację i przeskalowanie wejść względem wartości średniej i wariancji dla każdej partii wsadowej (*minibatch*) danych w procesie uczenia. Metoda ta jest pomocna zarówno w regularyzacji jak i przy problemach związanych z zanikającym (lub eksplodującym) gradientem, przeważnie w sieciach o dużej liczbie warstw ukrytych.

3.2.4 Podział zbioru danych w procesie uczenia i testowania sieci

Podczas budowy modeli bazujących na sieciach neuronowych, istotne jest także odpowiednie przygotowanie danych. Jednym z podstawowych kroków, które należy

w tym procesie wykonać, jest podział danych na zbiór uczący i testowy. Podział taki pozwala uniknąć problemu nadmiernego przeuczenia sieci (ang. *overfitting*), czyli sytuacji, w której sieć dobrze radzi sobie na danych, z którymi zetknęła się wcześniej w zbiorze uczącym natomiast wykazuje brak efektywności dla nowych przypadków zawartych w zbiorze testowym (zob. Goodfellow i in. [2016]). W praktyce, dane dzielone są na zbiór uczący i testowy najczęściej w proporcjach 70:30 lub 80:20. Podczas procesu uczenia sieć wykorzystuje tylko dane dostępne w zbiorze uczącym. Przykładowo, jeżeli dany szereg czasowy obejmuje okres dziesięciu lat, do uczenia sieci wykorzystywanych jest tylko siedem pierwszych lat, a trzy kolejne wykorzystywane są do testowania sieci.

W celu kontrolowania zdolności generalizacyjnych sieci, jeszcze na etapie jej uczenia, często stosowaną metodą jest dodatkowe wydzielenie ze zbioru uczącego tzw. zbioru walidacyjnego (ang. *validation set*), zawierającego ok 15-30% obserwacji (zob. Xu i Goodacre [2018]). Zbiór walidacyjny ma na celu dodatkowe sprawdzanie parametrów (wag) sieci już w trakcie procesu uczenia. Podczas każdego kroku w tym procesie, oprócz dokonywania predykcji na podstawie danych uczących, dodatkowo taka sama operacja wykonywana jest na danych walidacyjnych. Sieć przekazuje informacje o błędzie zarówno z predykcji na danych uczących jak i walidacyjnych, dzięki czemu lepiej ocenić można stopień ewentualnego przetrenowania sieci i poprawić ewentualne problemy poprzez zmianę hiperparametrów.



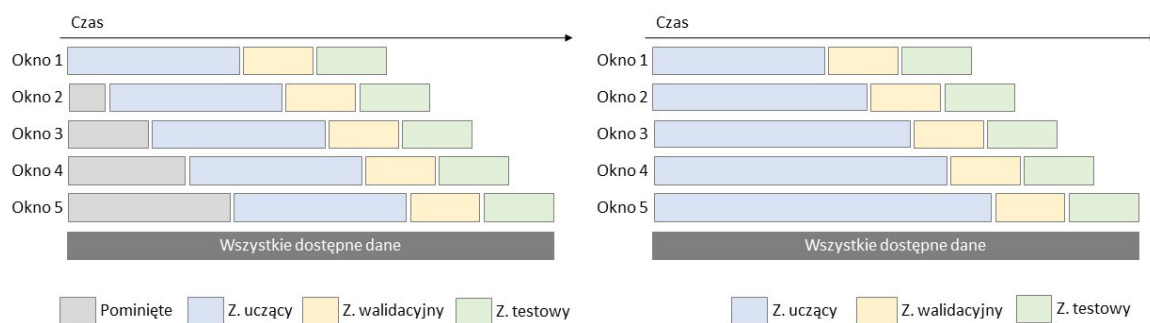
Rysunek 3.13: Podział danych na zbiory uczący, testowy oraz walidacyjny.

Źródło: Opracowanie własne.

Po zakończeniu procesu uczenia, działanie sieci sprawdzane jest na zupełnie

nowym zestawie danych, czyli zbiorze testowym, na którym uzyskiwane przez sieć wyniki są zazwyczaj nieco słabsze. Bardzo istotną kwestią jest, aby wyniki na zbiorach testowych nie były brane pod uwagę podczas dostrajania hiperparametrów sieci. Można więc założyć, że zbiór testowy nie jest dostępny (nie jest znany) w czasie uczenia sieci i podczas dostrajania hiperparametrów.

Jednym ze sposobów zapewniających poprawną ocenę testowanego modelu sieciowego jest zastosowanie schematu rolowanego (przesuwanego) okna (zob. Goodfellow i in. [2016]Bergmeir i Benítez [2012]). W przypadku szeregów czasowych, podejście to polega na przesuwaniu zbiorów uczącego, walidacyjnego i testowego o określoną wartość (najczęściej równą długości zbioru testowego), z zachowaniem kolejności (zob. rys. 3.14). Po przesunięciu okna, model jest uczony i testowany ponownie na nowych zbiorach danych. Stosowanie takiego podejścia umożliwia dokładniejszą weryfikację efektywności modelu na różnych zestawach danych. W niektórych przypadkach stosować można także schemat rekursywny (określany także jako schemat okna rozszerzanego, ang. *expanding window*), w której punkt początkowy zbioru uczącego jest stały, zmienia się tylko jego długość oraz punkt początkowy zbioru testowego (zob. Schnaubelt [2019]). Pozwala to na zweryfikowanie czy długość okna treningowego (ilość danych w zbiorze uczącym) wpłynie na poprawę wyników.



Rysunek 3.14: Wizualizacja podziału danych przy zastosowaniu okna rolowanego oraz okna rozszerzanego.

Źródło: Opracowanie własne.

W pracy z danymi w formie szeregów czasowych, w procesie uczenia sieć dobiera wagi tak, aby wykorzystując sekwencje wejściowe (ang. *sample* lub *sequence*) jak najlepiej prognozować kolejne wartości pożądane (ang. *target*, zob. Brownlee [2018]). Zbiór poszczególnych par sekwencji wejściowych i wartości pożądanych stanowi z kolei paczkę wsadową. Jednokrotna prezentacja wszystkich przypadków uczących (partii wsadowych), stanowiących zbiór uczący nazywana jest epoką (zob. Goodfellow

i in. [2016]). W procesie uczenia sieć przechodzi przez zbiór uczący nawet do kilku tysięcy razy (długość procesu uczenia ustalana jest na kilka tysięcy epok), przy czym po każdej epoce zwracana jest uzyskana wartość funkcji straty, tworząc w ten sposób krzywą uczenia (por. rys. 3.12).

Zarówno sekwencje wejściowe jak i wartości pożąpane, pobierane są ze zbioru uczącego, w związku z czym sieć podczas uczenia może w uzyskać trafność predykcji zbliżoną do poziomu 100%, co jednak nie oznacza, że tak samo dobrze będzie radzić sobie z zupełnie nowymi obserwacjami (pochodzącymi ze zbioru walidacyjnego lub testowego).

3.2.5 Przygotowanie danych

Duży wpływ na efektywność sieci ma także odpowiedni dobór i przygotowanie danych. Wspomniany wcześniej podział danych na zbiory uczące i testowe musi być dokonany w taki sposób, aby w jednym i drugim zbiorze znalazły się zbliżone reprezentacje danych. Jeżeli zbiór uczący nie będzie reprezentatywny (będzie znacząco różny) od zbioru testowego, sieć nie będzie w stanie dobrze wychwycić zależności zachodzących na danych testowych. Ważne jest więc, aby wszystkie trzy typy zbiorów (uczący, walidacyjny i testowy) były zbliżone jeśli chodzi o reprezentację danych (zob. Goodfellow i in. [2016]).

Inną istotną czynnością jest odpowiedni dobór zestawów cech (ang. *feature selection*) - czyli zestawu zmiennych wejściowych do sieci (zob. Cai i in. [2018]). Przykładowo, w finansowych szeregach czasowych mogą to być takie zmienne jak stopa zwrotu, cena, estymowana zmienność lub wolumen. Wszystkie cztery zmienne mogą być użyte jako dane wejściowe do sieci, jednak należy rozważyć (lub zbadać) czy wszystkie mają istotny wpływ przy prognozowaniu przykładowo stóp zwrotu lub zmienności. Często okazuje się, że dodatkowe zmienne wprowadzają niepotrzebne lub redundantne informacje zwiększając szum, co może wpłynąć negatywnie na proces uczenia oraz predykcji dokonywanych przez sieć (zob. Géron [2019]). Upewnić się trzeba także czy dane nie posiadają duplikatów lub brakujących wartości.

Innym często stosowanym zabiegiem podczas przygotowywania danych jest transformacja danych (ang. *feature scaling*) polegająca najczęściej na standaryzacji lub normalizacji (zob. Han i in. [2011], Brownlee [2018]). Zabiegi te stosowane są najczęściej w przypadkach gdy zmienne wejściowe mają różne skale wartości (przykładowo stopy zwrotu i wolumen). Pomocne są także w przypadkach gdy w danych występują wartości odstające. Normalizację i standaryzację stosuje się tylko dla danych wejściowych sieci, z kolei oczekiwane wartości (np. zrealizowana stopa zwrotu) powinny zawierać wartości oryginalne, gdyż w praktyce podczas eksploatacji

modelu są to wartości nieznane.

Odpowiednie przygotowanie danych może mieć bardzo duży wpływ na jakość uzyskanych prognoz. Podobnie jak w przypadku dostrajania hiperparametrów proces najczęściej jest dość czasochłonny, jednak nie powinien być pomijany.

3.3 Dostrajanie hiperparametrów sieci

Istotnym elementem, który należy brać pod uwagę podczas projektowania struktury sieci neuronowych, a w szczególności sieci uczenia głębokiego, jest odpowiedni dobór hiperparametrów. W zależności od typu sieci możemy mieć do czynienia z różnymi zestawami parametrów, które należy ustalić. Doboru hiperparametrów można dokonywać heurystycznie, na podstawie teoretycznych przesłanek i wartości przekazanych w literaturze, lub eksperymentalnie, poprzez dokładne „dostrajanie” hiperparametrów (ang. *hyperparameter tuning*) podczas procesu uczenia sieci. Odpowiedni wybór hiperparametrów sieci może mieć duży wpływ na jakość uzyskiwanych wyników, dlatego jest to proces któremu należy poświęcić dużo uwagi. Ponieważ liczba kombinacji wartości poszczególnych hiperparametrów może być duża, a sprawdzenie wydajności sieci wymaga przynajmniej częściowego jej trenowania, jest to także proces bardzo czasochłonny. Poniżej opisane zostały najważniejsze parametry, zarówno ogólne, występujące w prawie wszystkich strukturach sieci, jak i bardziej szczegółowe, typowe dla poszczególnych rodzajów sieci.

3.3.1 Liczba warstw i neuronów sieci

Do podstawowych parametrów, które najczęściej należy ustalić na początku projektowania struktury sieci należą liczba warstw sieci, oraz liczba neuronów w poszczególnych warstwach. W sieciach neuronowych uczenia głębokiego mówi się najczęściej o liczbie warstw większej niż trzy (wliczając w to warstwę wejściową i wyjściową), w praktyce najczęściej jest ich mniej niż 10. Z kolei liczba neuronów może zależeć od funkcji jaką pełni dana warstwa. Ilość neuronów w warstwach wejściowych i wyjściowych najczęściej uzależniona jest od wymiaru danych wejściowych i danych oczekiwanych na wyjściu modelu. W przypadku jednowymiarowych szeregów czasowych, warstwy wejściowe i wyjściowe mają tylko jeden neuron (przykładowo, gdy na wejściu sieć otrzymuje informacje o wartości w okresie t , a na wyjściu przekazuje wartość prognozowaną w okresie $t + 1$). Liczba neuronów w warstwach ukrytych może być ustalana dowolnie. Zwiększenie liczby neuronów i warstw ukrytych, może prowadzić do poprawy wydajności modelu, jednak

zazwyczaj wiąże się ze znacznym wydłużeniem czasu potrzebnego na wykonanie obliczeń oraz większym użyciem zasobów pamięciowych (zob. Goodfellow i in. [2016]).

3.3.2 Optymalizatory i inicjalizatory

Optymalizator sieci (ang. *optimizer*) informuje sieć w jaki sposób aktualizować parametry takie jak wagi, wykorzystując przy tym parametr prędkości uczenia *learning rate*. Obecnie najczęściej stosowanym optymalizatorem jest ADAM i jego odmiany, takie jak algorytm SGD, RMSProp czy Adagrad, opisane w punkcie 3.2.2.

Inicjalizator (ang. *initializer*) z kolei ustala jakie powinny być początkowe wartości wag parametrów sieci, co ma bezpośredni wpływ na czas procesu uczenia - dobrane odpowiednio wagi początkowe znacznie skrócą ten proces.

Niektóre parametry sieci, w szczególności początkowe wartości wag neuronów, muszą zostać zainicjalizowane (zob. Skorski i in. [2020]). Istotne jest tutaj aby wagi neuronów różniły się między sobą (w przeciwnym razie ich wartości najczęściej zmieniały by się jednakowo). Wartości wag najczęściej losowane są z rozkładu normalnego lub jednostajnego, natomiast wartości parametrów *bias* najczęściej ustawiane są domyślnie na 0 lub 1.

Glorot i Bengio [2010] zaproponowali nowy sposób inicjalizacji wag, mający przede wszystkim pomóc przy problemie zanikającego i eksplodującego gradientu, określony równaniem:

$$W \sim N\left(0, \frac{1}{n_j + n_{j+1}}\right), \quad (3.32)$$

lub dla rozkładu jednostajnego:

$$W \sim U\left(-\sqrt{\frac{6}{n_j + n_{j+1}}}, \sqrt{\frac{6}{n_j + n_{j+1}}}\right), \quad (3.33)$$

gdzie n_j oznacza liczbę wejść w warstwie j . He i in (2015) zauważyli jednak, że w przypadku neuronów z nieliniową funkcją aktywacji ReLU ten typ inicjalizacji nie jest optymalny. W zamian zaproponowali rozwiązanie, nazywane inicjalizacją He, w którym wariancja wag neuronów pomnożona jest dodatkowo przez 2, zgodnie z równaniem:

$$W \sim N\left(0, \frac{2}{n_j}\right). \quad (3.34)$$

3.3.3 Funkcje aktywacji

Jednym z najważniejszych parametrów każdej sieci jest funkcja aktywacji, której wartości stanowią sygnały wyjściowe danego neuronu. Odpowiada ona z to w jaki sposób neurony sieci przetwarzają dane wejściowe. Zdecydowana większość funkcji aktywacji to funkcje nieliniowe. Różne warstwy sieci mogą posiadać różne funkcje aktywacji. Przykładowo, w warstwie wyjściowej sieci typ funkcji zazwyczaj uzależniony jest od tego w jakiej formie dane mają być przekazane na wyjściu sieci (czyli np. od tego co prognozujemy). W warstwach ukrytych najczęściej wybierane są funkcje nieliniowe, takie jak ReLU (ang. *Rectified Linear Unit*), funkcja sigmoidalna czy funkcja tangensu hiperbolicznego.

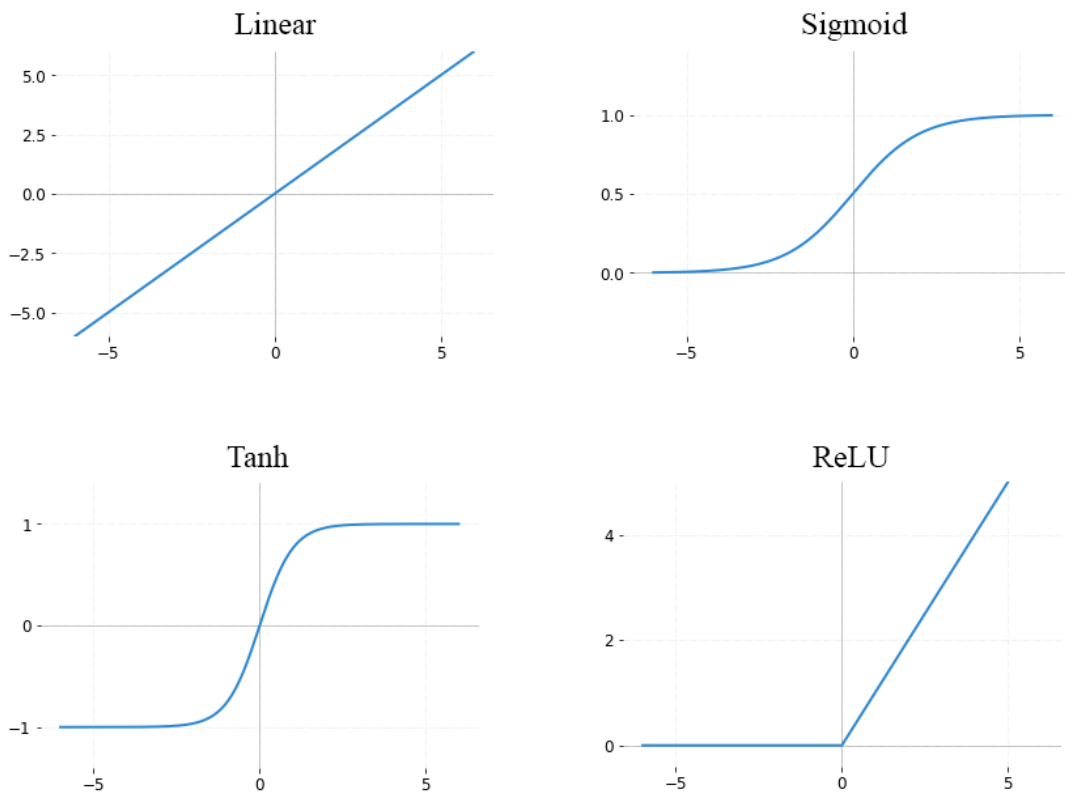
Tabela 3.1 oraz rysunek 3.16 przedstawiają wzory i przebieg podstawowych funkcji aktywacji, najczęściej używanych w uczeniu głębokim. Dokładniejszy przegląd funkcji aktywacji znaleźć można w pracach Nwankpa i in. [2018], oraz w Szandała [2021].

Jak wspomniano, duże znaczenie ma także wybór prawidłowej funkcji aktywacji dla neuronów wyjściowych. Jej forma odpowiada w tym przypadku za dokładną postać danych jaką otrzymamy na wyjściu sieci. Przykładowo dla typowych problemów regresyjnych, używa się najczęściej liniowej funkcji aktywacyjnej, która nie wprowadza już żadnych transformacji do otrzymanego wyniku. W problemach klasyfikacyjnych używana jest najczęściej funkcja *softmax* (w przypadkach kiedy zwracane jest prawdopodobieństwo przynależności do konkretnej klasy) lub sigmoidalna (w przypadkach gdy wyniki nie są rozkładem prawdopodobieństwa i nie muszą się sumować do 1). W zależności od potrzeb mogą tu też być stosowane inne funkcje, jak np. funkcja ReLU w przypadku gdy ujemne wartości chcemy zastąpić zerem.

Tabela 3.1: Podstawowe funkcje aktywacji.

Funkcja aktywacji	Wzór
Linear	$f(x) = x$
Sigmoid	$f(x) = \sigma(x) = \frac{1}{1+e^{-x}}$
Tanh	$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
ReLU	$f(x) = \begin{cases} 0, & \text{dla } x < 0. \\ x, & \text{dla } x \geq 0. \end{cases}$
Softmax	$f(\vec{x}) = \frac{e^{x_i}}{\sum_{j=1}^K e^{x_j}}$

Uwaga: K oznacza liczbę klas klasyfikatora, e^{x_i} i e^{x_j} to funkcje wykładnicze wektorów wejściowego i wyjściowego



Rysunek 3.15: Przebieg funkcji aktywacji.

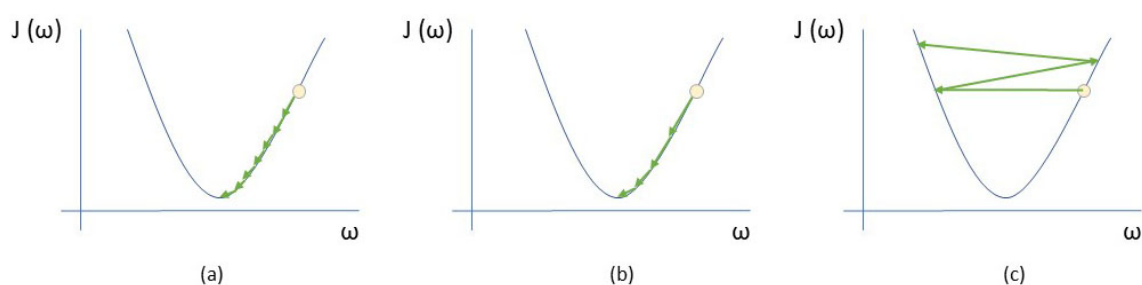
Źródło: opracowanie własne

3.3.4 Pozostałe hiperparametry

Wybór i ilość hiperparametrów, których wartości należy ustalić zależy w dużej mierze od specyfikacji problemu oraz wybranej architektury sieci. Poniżej przedstawione zostały pozostałe hiperparametry, których optymalizację podjęto także w części empirycznej pracy.

Funkcja straty rozumiana jako hiperparametr, jest wybraną funkcją optymalizowaną przez sieć. Najbardziej typowe funkcje straty wykorzystywane w sieciach uczenia głębokiego to MSE oraz funkcja entropii krzyżowej (ang. *cross-entropy*), z których wybrane dokładniej opisane zostały w rozdz. 2.⁷ Dobór odpowiedniej funkcji straty stanowi jeden z podstawowych kroków podczas konstruowania sieci, dlatego należy zwrócić szczególną uwagę na wybór odpowiedniego typu funkcji. Niektóre problemy wymagają także modyfikacji lub stworzenia własnych funkcji straty w celu zoptymalizowania działania sieci w kontekście danego problemu (zob. Michańków i in. [2022]).

Innym ważnym elementem jest **współczynnik uczenia** (ang. *learning rate*), który odpowiada za tempo uczenia się sieci (zob. Plagianakos i in. [2001], Liu i in. [2021]). Parametr ten kontroluje, w jaki sposób wagi neuronów zmieniane są przez algorytm uczenia gradientowego podczas optymalizacji funkcji straty (zob. wzór 3.19). Mniejsza wartość tego parametru powoduje wydłużenie procesu trenowania, jednak ustalenie tej wartości zbyt nisko może skutkować zakończeniem trenowania w minimum lokalnym lub punkcie siodłowym (ang. *saddle point*). Z kolei ustawienie zbyt wysokiej wartości może doprowadzić do pominięcia minimum globalnego (zob. rys.).



Rysunek 3.16: Optymalizacja funkcji straty w zależności o wartości współczynnika uczenia.

Uwaga: Figura (a) przedstawia sytuację, w której wartość jest zbyt niska, figura (b): wartość współczynnika jest optymalna, figura (c): wartość współczynnika jest zbyt duża. Źródło:

Opracowanie własne, na podstawie Jordan [2018].

⁷Funkcja entropii krzyżowej wykorzystywana jest przede wszystkim w problemach klasyfikacyjnych.

Parametr *batch size* określa ilość danych (rozmiar partii wsadowej), które brane są pod uwagę podczas kolejnych przejść przez dane zbioru uczącego sieci (epok) podczas fazy uczenia. Optymalna wielkość tego hiperparametru zależy przede wszystkim od architektury dostępnego sprzętu komputerowego, może jednak także wpływać na zdolności generalizacyjne modelu (zob. Goodfellow i in. [2016]). W sieciach bezstanowych sieciach LSTM, poprzez zmianę wartości tego hiperparametru można kontrolować długość pamięci długoterminowej.

W modelowaniu danych sekwencyjnych ustalana jest także **długość sekwencji** (określana również jako długość wektora wejściowego, *sample, instance lub feature vector*) czyli przykładowo liczba obserwacji które brane są pod uwagę przy dokonywaniu pojedynczej prognozy (zob. Chollet i Allaire [2018]). Poszczególne partie wsadowe danych złożone są z określonej liczby takich sekwencji. Długość sekwencji ustalana jest indywidualnie, w zależności od specyfikacji danego problemu.

Dodatkowo, poszczególne typy sieci mają swoje własne specjalistyczne hiperparametry, które trzeba ustalić przed procesem uczenia. Przykładowo w sieciach konwolucyjnych bardzo istotną rolę odgrywają takie parametry jak liczba filtrów i ich rozmiar. Parametry szczegółowe dla tych typów sieci dokładniej opisane są w punktach dotyczących poszczególnych architektur sieci: 3.1.2 i 3.1.3 oraz w podrozdziale 5.3.

3.3.5 Automatyzacja procesu dostrajania

Proces dostrajania hiperparametrów można częściowo zautomatyzować, korzystając z rozwiązań algorytmicznych. Jednym z popularnych narzędzi jest KerasTuner, zaproponowany przez O'Malleya i in. [2019]. Narzędzie to pozwala na zautomatyzowanie wyboru najlepszych wartości dla poszczególnych hiperparametrów lub ich kombinacji. Wyszukiwanie wartości dokonywane może być losowo (poprzez sprawdzenie wszystkich wartości i kombinacje parametrów) lub w sposób algorytmiczny, np. przy pomocy algorytmu Hyperband (Li i in. [2018]). Algorytm ten działa w sposób turniejowy, testując dużą liczbę zestawów hiperparametrów na kilku początkowych epokach, przenosząc następnie do kolejnego etapu tylko najlepiej spisujące się kombinacje lub hiperparametry. Takie podejście pozwala na znaczne skrócenie czasu potrzebnego na znalezienie optymalnych ustawień wartości hiperparametrów.

3.4 Sieci uczenia głębokiego w prognozowaniu szeregów czasowych

Poniżej, na podstawie wybranych publikacji omówiono badania, których autorzy wykorzystywali podstawowe typy sieci neuronowych, w szczególności sieci rekurencyjnych i konwolucyjnych, w modelowaniu szeregów czasowych instrumentów finansowych.

Chen i in. [2015] zaproponowali model wykorzystujący sieci LSTM do prognozowania stóp zwrotu spółek notowanych na giełdach SSE (ang. *Shanghai Stock Exchange*) i SZSE (ang. *Shenzhen Stock Exchange*), na danych z lat od 1990 do 2015 roku. Badania wykazały, że metody z wykorzystującą sieci LSTM przynosiły znacząco lepsze rezultaty w porównaniu z predykcją losową. W podobnych badaniach, również z wykorzystaniem sieci LSTM, Zhang i in. [2018b] stworzyli model wieloczynnikowy, służący do prognozowania stóp zwrotu indeksu giełdowego CSI300. Model został następnie wykorzystany do podejmowania decyzji klasyfikacyjnych, dotyczących wyboru poszczególnych akcji podczas tworzenia strategii inwestycyjnej. Rezultaty badań na danych spoza próby wskazywały na trafność predykcji nieznacznie powyżej 50%.

Fischer i Krauss [2018] wykorzystali modele uczenia głębokiego bazujące na sieciach LSTM do prognozowania kierunku zmian cen akcji spółek indeksu S&P 500, stosując w tym celu dane dotyczące dziennych zwrotów z lat od 1992 r. do 2015 r. Przeprowadzono także analizę porównawczą z modelami wykorzystującymi regresję logistyczną, sieciami DNN (ang. *Deep Neural Networks*) i las losowy (ang. *random forest*). Wyniki badań wykazały, że sieci LSTM osiągają najlepsze wyniki spośród badanych narzędzi.

Chong i in. [2017] zbadali możliwości predykcyjne algorytmów stworzonych na podstawie rekurencyjnych sieci uczenia głębokiego. Jako modelowany zbiór danych użyte zostały pięciominutowe obserwacje 38 spółek koreańskiej giełdy KOSPI, w postaci logarytmicznych stóp zwrotu, pochodzące z okresu od stycznia 2010 r. do grudnia 2014 r. Badania wykazały, że wykorzystanie metod uczenia głębokiego przyczyniło się do poprawienia możliwości predykcyjnych badanych modeli.

Sieci uczenia głębokiego LSTM wykorzystane zostały także do stworzenia strategii opartej na sentymencie rynkowym w pracy Jiahong Li i in. [2017]. Badacze wykorzystali naiwny klasyfikator bayesowski w celu oceny nastroju inwestorów na podstawie tekstów pobranych ze źródeł internetowych. Wyniki następnie połączone zostały z danymi giełdowymi, pochodzącymi z notowań spółek indeksu CSI300, i użyte jako dane wejściowe w modelu wykorzystującym sieć LSTM. Zaproponowany

przez badaczy model dawał wyniki trafności predykcji znacząco lepsze niż pozostałe testowane przez autorów metody, takie jak modele bazujące na maszynach wektorów nośnych.

Dixon i in. [2017] z powodzeniem wykorzystali głębokie sieci neuronowe DNN do budowy strategii algorytmicznego handlu towarami i kontraktami walutowymi. W badaniach wykorzystane zostały dane dotyczące 43 różnych towarów i walut z lat 1991 do 2014. Uzyskane wyniki wykazały, że wykorzystanie sieci głębokich może znacząco poprawić efektywność strategii inwestycyjnych.

Di Persio i Honchar [2016] porównali modele predycyjne oparte na różnych typach sieci neuronowych w prognozowaniu zmian kierunkowych indeksu giełdowego S&P 500. Zbadane zostały klasyczne sieci MLP (ang. *Multi-layer Perceptron*) a także sieci uczenia głębokiego CNN oraz LSTM. Spośród porównywanych typów sieci nieznacznie lepsze rezultaty uzyskały sieci konwolucyjne. Badacze zwracają także uwagę na bardzo dobre wyniki uzyskiwane przez modele łączące kilka typów sieci (ang. *ensemble models*).

Hansson [2017] wykorzystał modele bazujące na sieciach LSTM do porównania możliwości predykcyjnych sieci tego typu na rynkach o różnej efektywności. Wykorzystał do tego celu dane z lat od 2009 r. do 2017 r. pochodzące z indeksów giełdowych USA (S&P500), Brazylii (Bovespa) oraz Szwecji (OMX). Rezultaty, przedstawione w badaniu wykazały, że sieci LSTM charakteryzują się trafnością na poziomie 51-52% przy prognozowaniu kierunku zmian i najlepiej spisywały się w prognozowaniu rynków mniej efektywnych (OMX).

Cao i Wang [2019] analizowali modele bazujące na sieciach CNN służących do predykcji wartości pięciu indeksów giełdowych. Zaproponowali także autorskie rozwiązanie w postaci hybrydowego modelu CNN-SVM oraz przeprowadzili badania wpływu zmiany wartości poszczególnych hiperparametrów na uzyskiwane rezultaty. Badania wykazały, że modele CNN-SVM oraz CNN cechują się wyższą trafnością prognoz w porównaniu z modelami SVM i BP (ang. *back-propagation*).

Kijewski i Ślepaczuk [2020] wykorzystali modele bazujące na sieciach LSTM do prognozowania cen indeksu S&P 500 z okresu od 2000 r. do 2020 r. Uzyskane wyniki porównali z modelami ARIMA oraz klasycznymi strategiami inwestycyjnymi. W badaniach przeprowadzona została także dokładna analiza wrażliwości, w celu zbadania wpływu zmian poszczególnych parametrów modeli (w tym hiperparametrów sieci LSTM) na uzyskiwane przez nie prognozy.

W przedstawionym w powyższym podrozdziale zestawieniu badań, zauważyć można, że sieci uczenia głębokiego wykazują wysoką efektywność podczas rozwiązywania problemów związanych z predykcją stóp zwrotów różnych aktywów finansowych. Szczególnie często wykorzystywane są w tym celu sieci

rekurencyjne LSTM, a także sieci konwolucyjne. Te dwa rodzaje sieci posłużą jako główne składowe modeli hybrydowych, które omówione zostaną w kolejnym rozdziale.

Rozdział 4

Predykcyjne modele hybrydowe łączące metody ekonometryczne i techniki uczenia głębokiego

Autorskie rozwiązania zaproponowane w ramach tej pracy opierają się przede wszystkim na połączeniu metod ekonometrycznych z sieciami uczenia głębokiego, co zdaniem autora przyczynić się może do poprawy prognoz uzyskiwanych z wykorzystaniem tych narzędzi. W ramach tego rozdziału, przeanalizowane zostały wybrane modele hybrydowe proponowane w literaturze, a następnie przedstawione propozycje autorskich modeli hybrydowych służących do uzyskiwania punktowych prognoz zmienności, a także sieciowe modele umożliwiające prognozowanie parametrów całych rozkładów prawdopodobieństwa przyszłych stóp zwrotu.

4.1 Przegląd i analiza możliwości predykcyjnych modeli hybrydowych

W pierwszej kolejności przedstawione zostały wybrane publikacje dotyczące badań nad modelami hybrydowymi, wykorzystywanymi do punktowego prognozowania zmienności. Prace te analizowane są zarówno pod kątem struktur proponowanych modeli, jak i uzyskanych z ich wykorzystaniem wyników.

Monfared i Enke [2014] wykorzystali modele klasy GARCH w połączeniu z sieciami neuronowymi do prognozowania zmienności indeksu giełdowego NASDAQ Composite. Jako dane wejściowe użyte zostały prognozy zmienności dziesięciu indeksów (dotyczących poszczególnych sektorów gospodarki, wchodzących w skład NASDAQ), uzyskane przez model GJR-GARCH. Dane w postaci

dziesięciowymiarowego wektora przekazane zostały kolejno do sieci FFBP (ang. *Feed Forward with Back Propagation*), GR (ang. *Generalized Regression*) oraz RBF (ang. *Radial Basis Function*). Dane źródłowe, pochodzące z okresu od 1997 do 2011 roku, podzielone zostały na cztery okresy testowe, w ramach których oceniane były poszczególne modele. Do ocen prognoz wykorzystana została miara MSE. Z pośród badanych typów modeli, najlepsze rezultaty w trzech okresach testowych (1997-2000, 2002-2003 i 2005-2008) uzyskał model wykorzystujący sieć RBF. W ostatnim okresie, 2010-2011, najlepiej spisywał się model wykorzystujący sieć GR. Ogólnie, we wszystkich okresach modele hybrydowe dawały lepsze wyniki niż model GJR-GARCH, przy czym modele te najlepsze wyniki uzyskiwały w okresach charakteryzujących się wysoką zmiennością.

W badaniach nad modelami hybrydowymi, Kim i Won [2018] jako wartości wejściowych do sieci LSTM oraz sieci DFN (ang. *Deep Feedforward Network*) użyli oszacowań poszczególnych parametrów modeli GARCH, EGARCH oraz EWMA (oznaczanych dalej kolejno jako G,E,W). Dodatkowo na wejściu do sieci wykorzystane zostały także logarytmiczne stopy zwrotu indeksu KOSPI, stopy procentowe, oraz ceny surowców złota i ropy. Stworzone zostały w ten sposób modele hybrydowe G/E/W - DFN/LSTM. Proponowane przez autorów modele wykorzystane zostały do prognozowania zmienności koreańskiego indeksu giełdowego KOSPI 200. Jednookresowe prognozy zmienności uzyskiwane były na podstawie sekwencji wejściowej o długości 22 dni. Stosując miary MAE, MSE, HMAE i HMSE badacze wykazali, że najlepsze wyniki uzyskał model GEW-LSTM, łączący sieci LSTM jednocześnie z trzema modelami: GARCH, EGARCH oraz EWMA. W kolejnej części badań modele przeanalizowane zostały także pod kątem różnych długości sekwencji wejściowych (7, 15 i 22 dni) oraz długości prognoz (1, 14, 21 dni). Wartość błędu malała wraz z wydłużaniem długości sekwencji wejściowej i rosła przy wydłużaniu okresu prognozowanego.

Podobne badania, dotyczące prognozowania zmienności cen surowca miedzi, przeprowadzili Hu i in. [2020]. W tym celu Autorzy stworzyli modele hybrydowe łączące klasyczne sieci neuronowe, sieci rekurencyjne oraz model GARCH(1,1). Spośród sieci rekurencyjnych zbadane zostały sieci LSTM oraz dwukierunkowe sieci BLSTM (ang. *Bidirectional LSTM*). Te ostatnie charakteryzują się wykorzystaniem dodatkowej warstwy wstecznej, która przetwarza informacje w przeciwnym kierunku. Jako dane wejściowe wykorzystane zostały 21 wskaźniki przedstawiające ceny podstawowych metali, wybrane kursy walut indeksy giełdowe i futures, oraz wskaźnik inflacji, z lat 2008 do 2018. Dane te wraz z prognozami zmienności uzyskanymi przy pomocy modelu GARCH, służyły jako sygnały wejściowe do sieci LSTM-ANN. Do

sieci przekazywane były dane roczne (252 dni), a prognozowana była zmienność na okres 10-dniowy oraz 40-dniowy. W badaniach porównane zostały wyniki sześciu typów modeli: ANN, LSTM-ANN, BLSTM-ANN, GARCH-ANN, GARCH-LSTM-ANN, oraz GARCH-BLSTM-ANN. Do oceny jakości prognoz wykorzystane zostały w tym przypadku miary MSE, MAE, RMSE i MAPE. Na podstawie wyników badań Autorzy stwierdzili, że zastosowanie modelu GARCH w warstwie wejściowej sieci znacznie poprawia prognozy modelu w porównaniu z modelami składającymi się wyłącznie z sieci neuronowych. Zauważono także, że zastosowanie sieci LSTM oraz BLSTM przynosi lepsze wyniki w porównaniu z klasycznymi sieciami ANN. Z pośród wszystkich badanych modeli, najlepsze wyniki uzyskał model GARCH-LSTM-ANN.

Kristjanpoller i Hernández [2017] zastosowali podobne podejście w badaniach dotyczących prognozowania zmienności cen metali szlachetnych. W strukturze badanych przez nich modeli hybrydowych, jako dane wejściowe wykorzystane zostały prognozy uzyskane przez modele ARCH, GARCH, NARCH, TARARCH oraz APARCH. Prognozy te, wraz z innymi zmiennymi wejściowymi, użyte zostały jako wejścia do sieci ANN(l, n), gdzie l oznacza ilość warstw sieci, zaś n to liczba neuronów w poszczególnych warstwach. Badania przeprowadzone były z wykorzystaniem danych z okresu od 1999 roku do 2014 roku. Jako dodatkowe zmienne objaśniające wykorzystane zostały: stopy zwrotu z indeksów SZSE, FTSE i SBSE, kursy walut USD-EUR, USD-YEN oraz ceny ropy. Jako funkcja straty użyta została miara HMSE, natomiast do testowania efektywności modeli — miara MCS (ang. *Model Confidence Set*). Analizując przedstawione przez Autorów wyniki, można zauważyć, że oceny błędów prognoz ulegały poprawie w zależności od ilości warstw zastosowanych w sieci ANN — najlepsze rezultaty uzyskiwały sieci o architekturze składającej się z 3-6 warstw. Na jakość prognoz wpływał także dobór odpowiedniej podklasy modelu GARCH oraz ilości dodatkowych zmiennych użytych jako sygnały wejściowe do sieci.

Liu i So [2020] wykorzystali modele hybrydowe do prognozowania zmienności cen indeksu S&P500. Badacze skupili się na modelu wykorzystującym prognozy uzyskanych przez model GARCH oraz sieć LSTM. Celem badań było przede wszystkim znalezienie optymalnej architektury sieci. Jako dane wykorzystane zostały dzienne stopy zwrotu z lat 2000-2020, natomiast do oceny trafności prognoz zastosowana została miara MAE. W badaniach przetestowane zostały poszczególne architektury sieci: LSTM(32), LSTM(32)+dense(16), LSTM(32)+dense(16-8), LSTM(32)+Dense(16-8-4), LSTM(32)+Dense(32-16-8), LSTM(32)+Dense(64-32-16-8), gdzie w nawiasach podane zostały liczby neuronów na

poszczególnych warstwach. Spośród testowanych architektur najlepsze wyniki uzyskała sieć LSTM(32)+Dense(32-16-8). Autorzy zwrócili także uwagę na fakt, że sieci o większej liczbie warstw, charakteryzujące się większą liczbą parametrów, są bardziej podatne na przeuczenie.

Na podstawie opisanych powyżej struktur modeli hybrydowych, można wnioskować, że w proponowanych w literaturze modelach hybrydowych wykorzystywanych do prognozowania zmienności spośród specyfikacji modeli GARCH zwykle wybierane są modele asymetryczne, takie jak EGARCH oraz GJR-GARCH. Prognozy wykorzystane z użyciem tych modeli są następnie wykorzystywane jako sygnały wejściowe w sieciach neuronowych w celu uzyskania ostatecznych prognoz zmienności. Przedstawione wyniki badań pokazują, że modele hybrydowe można z powodzeniem wykorzystywać do prognozowania zmienności instrumentów finansowych, a uzyskane w ten sposób prognozy najczęściej są lepsze w porównaniu z wynikami klasycznych modeli GARCH.

W tym miejscu odnieść można się także do literatury dotyczącej modeli wykorzystujących sieci uczenia głębokiego w prognozowaniu parametrów rozkładu prawdopodobieństwa. W tym zakresie wskazać należy przede wszystkim pracę Chen i in. [2020] oraz monografię 2020. Ponieważ publikacje te nie dotyczą jednak prognozowania finansowych szeregów czasowych, prezentowane w nich wyniki nie będą tutaj dokładnie omawiane. Zaznaczyć jednak należy, że przedstawione w nich koncepcje stanowią pewne podstawy, na których zaproponowane zostały autorskie rozwiązania sieciowych modeli probabilistycznych, omówione w punkcie 4.2.2.

4.2 Propozycja nowych modeli

Bazując na przedstawionych w powyższym podrozdziale rozważaniach, można stwierdzić, że sieci uczenia głębokiego, zarówno te z grupy sieci rekurencyjnych, jak i konwolucyjnych, można z powodzeniem stosować jako dodatkowe uzupełnienie modeli zmienności, takich jak modele klasy GARCH, w prognozowaniu zmienności oraz przy szacowaniu ryzyka. Na tej podstawie, w dalszej części pracy przedstawione zostały propozycje rozwiązań, łączących metody ekonometryczne oraz sieci neuronowe w modele hybrydowe.

Główną zaletą modeli hybrydowych wykorzystujących sieci uczenia głębokiego jest przede wszystkim ich elastyczność. Sieci pozwalają na stosowanie jako zmiennych wejściowych dowolnych danych, w tym przypadku wygasłych prognoz zmienności uzyskanych przez dowolne modele i estymatory. Duże możliwości daje także wybór odpowiedniej wartości pożądanej sieci. Taka elastyczność sprawia, że tworzyć można

w zasadzie dowolne połączenia metod statystycznych lub ekonometrycznych z modelami sieciowymi. W podobny sposób można tutaj wykorzystać także prognozy uzyskane przez inne modele zmienności, takie jak modele SV, oraz przeprowadzić analizę wielowymiarową. Dodatkowo, większość prezentowanych w literaturze badań wskazuje na zauważalną poprawę prognoz zmienności w przypadku modeli hybrydowych.

Do słabości modeli prezentowanych w publikacjach przedstawionych w podrozdziale 4.1 zaliczyć można: wykorzystywanie tylko jednej specyfikacji GARCH, porównanie z zmiennością historyczną jako estymatorem zmienności, skupianie się wyłącznie na jednym typie rozkładu warunkowego (często nawet nie wspomniano jaki rozkład został wybrany), badanie zmienności tylko na jednym szeregu, nie stosowanie ruchomego okna. Dodatkowo, w zasadzie w żadnych spośród wymienionych badań dotyczących modelowania finansowych szeregów czasowych nie próbowano prognozować (zmiennych w czasie) parametrów całych rozkładów warunkowych.

W poniższych punktach zaprezentowane zostały autorskie rozwiązania, a w szczególności modele oparte na architekturze ARMA-GARCH-LSTM, wykorzystywane do punktowych prognoz zmienności, oraz modele NN-D stosowane do przede wszystkim do prognozowania probabilistycznego.¹

4.2.1 Hybrydowy model punktowych prognoz zmienności

Pierwszy spośród proponowanych modeli hybrydowych służy do uzyskiwania jednookresowych punktowych prognoz zmienności. Składa się on z trzech zasadniczych komponentów, wykorzystujących kolejno metody ekonometryczne, sieci uczenia głębokiego oraz estymator zmienności GKYZ.

W pierwszej części modelu wykorzystywane są struktury ARMA-GARCH do uzyskania jednookresowych prognoz zmienności. Opóźnienia części AR oraz MA w tej strukturze ustalane są na podstawie kryteriów informacyjnych AIC oraz BIC (zob. punkt 2.4.2) dla każdego z badanych aktywów finansowych. Z kolei do modelowania warunkowej wariancji wykorzystywane są następujące specyfikacje: GARCH(1,1), EGARCH(1,1), GJR-GARCH(1,1) oraz APARCH(1,1) o zadanych rozkładach warunkowych: normalnym, t-Studenta i skośnym t-Studenta (zob. tabela 4.1).

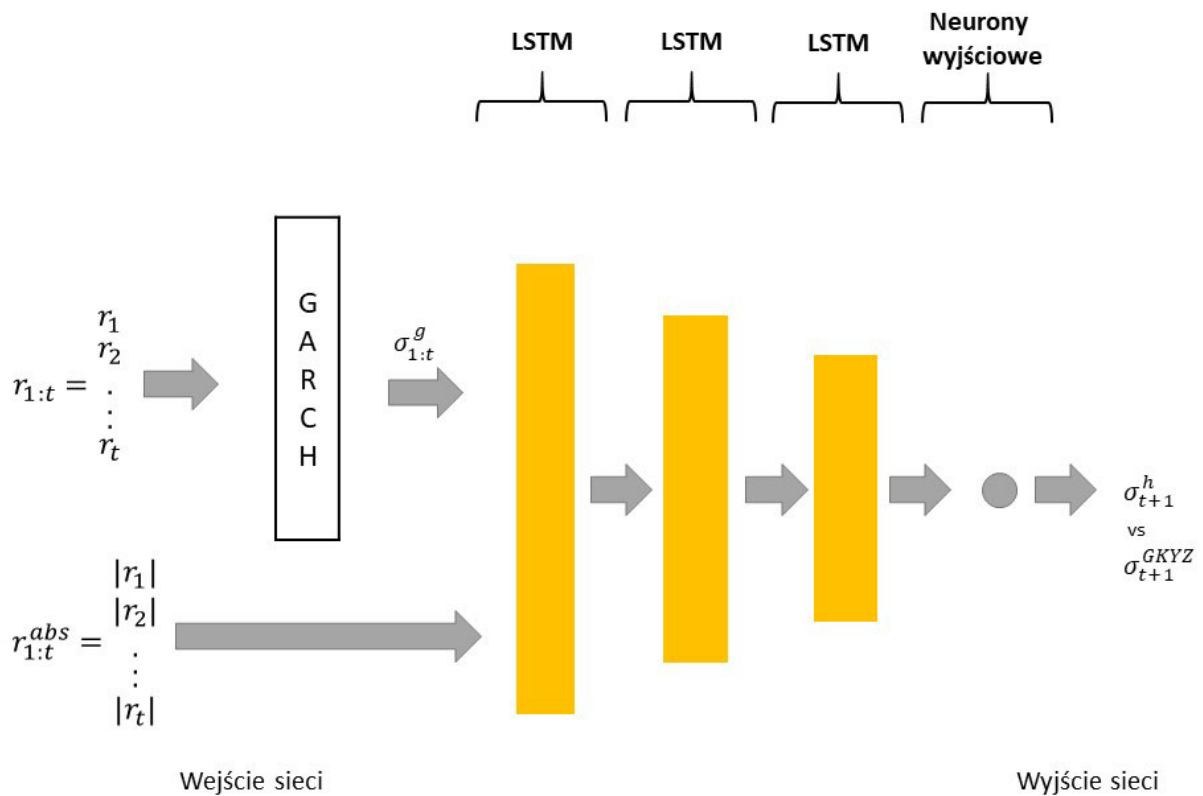
¹Choć ich struktura umożliwia także prognozowanie punktowe.

Tabela 4.1: Wybrane specyfikacje modeli klasy GARCH i oraz użyte typy rozkładów warunkowych.

Specyfikacje modelowe	Rozkłady warunkowe
GARCH (1,1)	normalny t-Studenta Skośny t-Studenta
EGARCH (1,1)	
GJR-GARCH (1,1)	
APARCH (1,1)	

Źródło: Opracowanie własne.

W drugiej części modelu, prognozy uzyskane przy pomocy wymienionych powyżej modeli wykorzystane są jako dane wejściowe do rekurencyjnej sieci LSTM. Zastosowanie komponentu sieciowego ma za zadanie poprawę prognoz zmienności poprzez modelowanie ewentualnych dodatkowych zależności nieliniowych. Z kolei na wyjściu sieci, w celu oceny stopnia zgodności generowanych prognoz z wartościami pożądanymi, stosowany jest jeden z estymatorów bazujących na zakresie cen — estymator Garmana-Klassa zmodyfikowany o skoki między cenami otwarcia i zamknięcia (zob. punkt 2.3.2 oraz Yang i Zhang [2000]). Struktura proponowanego modelu przedstawiona została na rysunku 4.1.



Rysunek 4.1: Schemat modelu hybrydowego ARMA-GARCH-LSTM.

Uwaga: $\sigma_{1:t}^g$ oznacza wygasłe prognozy zmienności, uzyskane z wykorzystaniem modelu ARMA-GARCH, σ_{t+1}^h prognozę zmienności z modelu hybrydowego, natomiast σ_{t+1}^{GKYZ} jest wartością pożądaną, czyli zmiennością uzyskaną przy pomocy estymatora GKYZ.

Źródło: Opracowanie własne.

Specyfikacje modelu sieciowego (struktura sieci i wartości poszczególnych hiperparametrów) oraz informacje dotyczące danych wejściowych (wybrane aktywa finansowe oraz sposób podziału danych wejściowych) zostały dokładnie przedstawione w podrozdziałach 5.3 oraz 5.5.

Zaproponowana konstrukcja modelu hybrydowego umożliwia wykorzystanie różnych specyfikacji modeli klasy GARCH, także o różnych typach rozkładów warunkowych. Elastyczność takiej architektury umożliwia także stosowanie innych typów modeli zmienności (jak przykładowo modele SV). Dodatkowo, poprzez zmianę wartości poświadczonych sieci możliwe jest także użycie dowolnego estymatora zmienności, do którego komponent sieciowy modelu jest uczony. Możliwa jest także zmiana typu wykorzystywanej sieci neuronowej, przykładowo na jednowymiarową sieć konwolucyjną lub rekurencyjną sieć GRU.

Ponieważ, jak przedstawiono w poprzednim podrozdziale, sama koncepcja łączenia modeli GARCH z sieciami neuronowymi nie jest nowa, należy wskazać elementy, które w proponowanym podejściu zostały usprawnione. Pierwszą rzeczą, na którą warto zwrócić uwagę w dotychczas publikowanych badaniach jest fakt, że jako

estymator służący do wskazania wartości pożądanych sieci najczęściej wykorzystywany jest estymator bazujący na zmienności historycznej, HV (zob. np. Kim i Won [2018], Kristjanpoller i Minutolo [2018]). W przypadku proponowanego w niniejszej dysertacji modelu hybrydowego korzystamy z estymatora GKYZ – znacznie efektywniejszego w stosunku do oszacowań HV. Decyzja o użyciu tego estymatora podyktowana jest przede wszystkim tym, że jest on stosunkowo prosty w konstrukcji, a przy tym nie zakłada stałości zmienności w czasie (jak ma to miejsce w przypadku estymatora HV). Nie wymaga też stosowania danych o częstotliwości wyższej niż dzienna (tak jak estymatory zmienności zrealizowanej, RV), dzięki czemu jest mniej narażony na efekty mikrostruktury rynku. Łatwy dostęp do historycznych notowań o częstotliwości dziennej pozwala także na testowanie wsteczne modeli na relatywnie długim okresie czasu i dla dużej liczby aktywów. Fiszeder [2020] podkreśla także, że estymatory bazujące na zakresie cen coraz częściej stosowane są jako miara służąca do oceny modeli zmienności.

Kolejna zmiana dotyczy danych wejściowych użytych w komponencie sieciowym. Zamiast stosowania dodatkowych zmiennych wejściowych do sieci, takich jak np. oceny poszczególnych parametrów modelu GARCH (jak w pracy Kim i Won [2018]), rozważone zostały prognozy uzyskane z różnych specyfikacji modeli GARCH wykorzystujących dodatkowo różne typy warunkowych rozkładów prawdopodobieństwa, co zdaniem autora w większy sposób może przyczynić się do poprawy uzyskiwanych prognoz. W niektórych badaniach (zob. np. Hu i in. [2020]), jako dane wejściowe do sieci używane były również dane z estymatorów zmienności. W tym przypadku jednak zdecydowano się na użycie estymatora tylko jako wartości pożądanych sieci, nie wykorzystując go jako zmiennej wejściowej. W ten sposób wyniki uzyskane za pomocą modeli hybrydowych można bezpośrednio porównać z wynikami otrzymanymi w modelach GARCH, które służyły jako dane wejściowe do sieci, aby sprawdzić, czy element sieciowy faktycznie poprawia jakość uzyskanych prognoz².

Proponowany model hybrydowy, poprzez wprowadzenie wskazanych powyżej modyfikacji, powinien lepiej uwzględniać specyficzne własności danych finansowych, co w rezultacie może przyczynić się do poprawy uzyskiwanych z jego wykorzystaniem prognoz. Wyniki punktowych prognoz zmienności uzyskane przez poszczególne specyfikacje model hybrydowego przedstawione zostały w punkcie 5.5.1, natomiast wyniki dotyczące prognoz ryzyka w punkcie 5.5.2. Analiza porównawcza modeli

²We wspomnianych badaniach sprawdzana była sytuacja odwrotna, mianowicie sprawdzano, czy dodanie prognoz uzyskanych przy pomocy modeli GARCH jako danych wejściowych do sieci wpłynie pozytywnie na prognozy modelu sieciowego.

hybrydowych z wybranymi specyfikacjami modeli klasy GARCH przedstawiona została w punkcie 6.1.2.

4.2.2 Model prognoz probabilistycznych

Jako drugi typ modelu zaproponowany został model służący do predykcji probabilistycznych. Pozwala on na prognozowanie całych rozkładów prawdopodobieństwa przyszłych stóp zwrotu, poprzez modelowanie poszczególnych parametrów zadanego typu rozkładu. Dzięki temu wyniki uzyskane przez taki model można bezpośrednio wykorzystać np. w prognozowaniu ryzyka kapitałowego, bez konieczności wykorzystywania dodatkowych parametrów (średniej, liczby stopni swobody czy skośności) z modeli GARCH, jak miało to miejsce w przypadku modeli hybrydowych, omawianych w poprzednim punkcie. Specyfikacja modelu wykorzystująca wyłącznie zadaną sieć neuronową do uzyskania prognoz całego rozkładu prawdopodobieństwa pozwala więc na lepszą (bardziej bezpośrednią) ocenę możliwości predykcyjnych modeli opartych wyłącznie na sieciach neuronowych. Poniżej przedstawiony został sposób konstrukcji sieciowego modelu probabilistycznego, wychodząc od założeń ekonometrycznych.

Niech x_t oznacza wektor danych wejściowych do sieci, zawierający przeszłe informacje o stopach zwrotu i oszacowania zmienności (zob. rysunek 4.2), wykorzystywany do modelowania bieżącej wartości r_t . Proponowany model można zapisać za pomocą następujących równań:

$$r_t = \mu(x_t) + \varepsilon_t, \quad (4.1)$$

$$\varepsilon_t = \sigma(x_t)z_t, (z_t|x_t) \sim iid(\eta(x_t)), \quad (4.2)$$

$$\sigma_t^2 = \sigma^2(x_t), \quad (4.3)$$

gdzie D oznacza zadany z góry typ rozkładu prawdopodobieństwa o parametrach $\eta(x_t)$:

$$f_D(z_t; x_t) = \begin{cases} f_N^{(1)}(z_t; 0, 1), & \text{gdy } D \equiv N \\ f_{St}^{(1)}(z_t; 0, 1, \nu(x_t)), & \text{gdy } D \equiv St \\ f_{sSt}^{(1)}(z_t; 0, 1, \nu(x_t), \xi(x_t)), & \text{gdy } D \equiv sSt \end{cases} \quad (4.4)$$

Funkcje $\mu(x_t)$, $\sigma^2(x_t)$, $\nu(x_t)$ oraz $\xi(x_t)$ są Ψ_{t-1} -mieralne, czyli są pewnymi funkcjami tylko i wyłącznie przeszłości, na które nałożone są restrykcje w postaci: $\sigma^2(x_t) > 0$, $\nu(x_t) > 0$, $\xi(x_t) > 0$. Zależności zadane tymi funkcjami nie są definiowane

w żaden formalny sposób, co świadczy o wysokim poziomie ogólności proponowanego modelu.

Równania 4.1-4.3 definiują (wysoco) nieliniowy model klasy ARMA-GARCH, w którym dodatkowo zmienne w czasie są parametry ν_t oraz ξ_t . Ze względu na taką specyfikację, do wyznaczenia parametrów poszczególnych rozkładów konieczne jest użycie sieci neuronowej.

Parametry modelu sieciowego, ω , wyznaczać można poprzez minimalizację wartości przeciwnej do logarytmu funkcji wiarygodności (ang. *Negative Log-Likelihood* - NLL, zob. Duerr i in. [2020]), w postaci:

$$NLL(\omega) = - \sum_{t=1}^n \ln f_D(r_t; \omega). \quad (4.5)$$

Zgodnie z powyższym, w wyniku minimalizacji funkcji NLL, otrzymujemy:

$$\hat{\omega} = \operatorname{argmin}_{\omega} NLL(\omega). \quad (4.6)$$

Łatwo pokazać, że postaci funkcji NLL dla poszczególnych rozważanych typów rozkładu prawdopodobieństwa (o funkcjach gęstości zadanych wzorami 2.40, 2.42 oraz 2.44), są następujące:

– przy założeniu $D \equiv N$:

$$NLL(\omega) = \frac{n}{2} \ln(2\pi) + \frac{1}{2} \sum_{t=1}^n \ln \sigma_t^2 - \sum_{t=1}^n \frac{(r_t - \mu_t)^2}{2\sigma_t^2}, \quad (4.7)$$

– przy założeniu $D \equiv St$:

$$\begin{aligned} NLL(\omega) = & - \sum_{t=1}^n \ln \left(\frac{\Gamma(\frac{\nu_t+1}{2})}{\Gamma(\frac{\nu_t}{2})} \right) + \frac{1}{2} \sum_{t=1}^n \ln(\nu_t \pi) \\ & + \frac{1}{2} \sum_{t=1}^n \ln \sigma_t^2 + \sum_{t=1}^n \frac{\nu_t + 1}{2} \ln \left(1 + \frac{(r_t - \mu_t)^2}{\nu_t \sigma_t^2} \right), \end{aligned} \quad (4.8)$$

– przy założeniu $D \equiv sSt$:

$$\begin{aligned} NLL(\omega) = & - \sum_{t=1}^n \ln \left(\frac{2}{\xi_t + \xi_t^{-1}} \right) + \sum_{t=1}^n \ln \sigma_t \\ & - \ln \left[f_{St}(\xi_t z_t; 0, 1, \nu_t) H_-(-z_t) + f_{St}(\xi_t^{-1} z_t; 0, 1, \nu_t) H_+(z_t) \right]. \end{aligned} \quad (4.9)$$

Wartości parametrów ω są w istocie estymowane metodą największej

wiarygodności (zob. Duerr i in. [2020]), jednak z uwagi na bardzo ogólną postać modelu, nie dają się udowodnić własności estymatora. Podobnie, nie dają się wyprowadzić warunki ścisłej czy kowariancyjnej stacjonarności procesu zdefiniowanego za pomocą równań (4.1 - 4.1). Z uwagi na wysoce nieliniową postać modelu, „estymacja” jego parametrów (uczenie sieci) przeprowadzana może być tylko za pomocą sieci neuronowych.

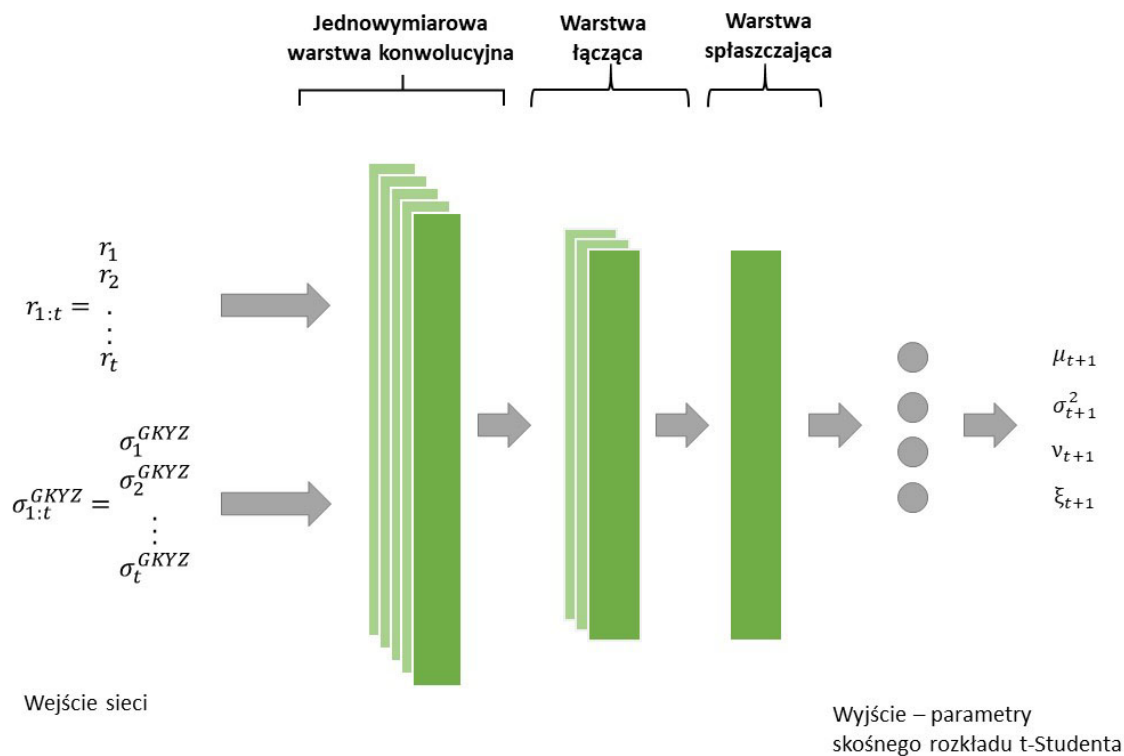
W modelu probabilistycznym prognozy jednookresowe przyjmują następującą postać:

$$p(r_{t+1}|\Psi_t; \omega) = \begin{cases} f_N^{(1)}(r_{t+1}; \mu_{t+1}, \sigma_{t+1}^2), & \text{gdy } D \equiv N \\ f_{St}^{(1)}(r_{t+1}; \mu_{t+1}, \sigma_{t+1}^2, \nu_{t+1}), & \text{gdy } D \equiv St \\ f_{sSt}^{(1)}(r_{t+1}; \mu_{t+1}, \sigma_{t+1}^2, \nu_{t+1}, \xi_{t+1}), & \text{gdy } D \equiv sSt \end{cases} \quad (4.10)$$

Typ sieci w modelu probabilistycznym nie musi być z góry określony, można skorzystać z dowolnego typu sieci dostosowanego do pracy z danymi w postaci finansowych szeregów czasowych. W ramach niniejszej rozprawy zastosowane zostały w tym wypadku jednowymiarowe sieci konwolucyjne (CNN) lub sieci rekurencyjne typu LSTM.

Zadaniem modelu NN-D jest prognozowanie całych rozkładów prawdopodobieństwa (o określonym z góry typie) poprzez prognozowanie wartości parametrów tych rozkładów. Za pomocą opisanych powyżej funkcji straty jesteśmy w stanie prognozować wartości parametrów w chwili $t + 1$, dla wybranych rozkładów (normalnego, t-Studenta i skośnego t-Studenta).³ Dla rozkładu normalnego prognozowane są dwa parametry: μ_t oraz σ_t . Dla rozkładu t-Studenta dodatkowo prognozowana jest wartość parametru ν_t , oznaczającego liczbę stopni swobody, natomiast dla skośnego rozkładu t-Studenta dodatkowym, czwartym parametrem jest ξ_t odpowiedzialny za modelowanie skośności tego rozkładu. Aby uzyskać takie wyniki, na wyjściu sieci określona musi być także odpowiednia ilość neuronów: dwa dla rozkładu normalnego (ponieważ prognozowane są wartości dwóch parametrów), trzy dla rozkładu t-Studenta, i cztery neurony dla skośnego rozkładu t-Studenta. Struktura modelu, dla jednowymiarowej sieci konwolucyjnej, prognozującego cztery parametry skośnego rozkładu T-studenta przedstawiona została na rysunku 4.2.

³W nazwie modelu NN oznaczać będzie sieć neuronową, w tym przypadku sieć CNN lub LSTM, zaś D oznaczać będzie typ konkretny rozkładu.



Rysunek 4.2: Model probabilistyczny CNN-SSTD, z czterema neuronami wyjściowymi zwracającymi wartości parametrów skośnego rozkładu t-Studenta.

Źródło: Opracowanie własne.

Zwróćmy uwagę, że w hybrydowym modelu stosowanym do punktowego prognozowania zmienności (zob. punkt 4.2.1), jako dane wejściowe do sieci służyły prognozy uzyskane za pomocą specyfikacji modeli GARCH. Z kolei w przypadku modelu probabilistycznego, na wejściu sieci, obok szeregu stóp zwrotu, wykorzystane zostały oszacowania zmienności uzyskane z wykorzystaniem estymatora GKYZ. Wynika to przede wszystkim z faktu, że w przypadku modelu stosowanego do prognozowania punktowego wartości pożądane sieci musiały odpowiadać ustalonemu „faktycznemu” poziomowi zmienności, w tym przypadku były to dane z estymatora. W przypadku modelu probabilistycznego takie podejście nie jest konieczne — jako wartości pożądane sieci wykorzystywane są jedynie zrealizowane stopy zwrotu, przez co dane z estymatora mogą być wykorzystane także na wejściu sieci. Oczywiście, ze względu na elastyczność sieci, estymator można było także wykorzystać na wejściu sieci w modelu punktowym, a prognozy uzyskane za pomocą modeli GARCH jako wejścia w modelu probabilistycznym, jednak spowodowało by to powstanie bardzo dużej liczby podklas modelu do oceny w badaniach empirycznych.

Do podstawowych zalet zaproponowanego tu modelu NN-D zaliczyć należy przede wszystkim możliwość prognozowania całego zadanego rozkładu prawdopodobieństwa poprzez prognozowanie parametrów tego rozkładu, wyłącznie przy użyciu sieci

neuronowych. Co prawda podejście to nie jest nowe, jednak stosowane było najczęściej do prognozowania parametrów rozkładu normalnego (jak w przypadku pracy Duerr i in. [2020], w której autorzy prezentują podobną koncepcję na przykładzie danych o charakterze medycznym, lub Chen i in. [2020], gdzie wykorzystano dane sprzedażowe). Proponowany model skonstruowany został z myślą o dodatkowych typach rozkładu, często wykorzystywanych w modelowaniu finansowym, takich jak rozkład t-Studenta i skośny rozkład t-Studenta. W tym celu wyznaczone zostały postaci funkcji straty (NLL) dla wspomnianych rozkładów. O ile wśród gotowych narzędzi programistycznych dostępne są pakiety umożliwiające prognozowanie parametrów rozkładu normalnego i t-Studenta (zob. Dillon i in. [2017]), to wyznaczenie i implementacja funkcji NLL dla skośnego rozkładu t-Studenta jest elementem nowości. Niewiele jest też badań wykorzystujących podejście sieciowe do prognozowania rozkładów prawdopodobieństwa dla finansowych szeregów czasowych. Także na tym gruncie proponowany model jest rozwiązaniem nowatorskim.

Wykorzystanie w modelu komponentu sieci neuronowej pozwala na modelowanie wysoce nieliniowych zależności. Dodatkowo, umożliwia także elastyczne szacowanie i prognozowanie zmiennych w czasie parametrów liczby stopni swobody, ν_t , oraz skośności, ξ . Takie rozwiązanie może przyczynić się do poprawy jakości prognoz wartości tych parametrów, w stosunku do oszacowań uzyskanych przy pomocy modeli kasy GARCH z warunkowym rozkładem prawdopodobieństwa, gdzie wartości tych parametrów są stałe w czasie.

Prognozy uzyskane z wykorzystaniem probabilistycznych modeli sieciowych można także wykorzystać w celu oszacowania poziomów wartości zagrożonej oraz szacowanego niedoboru, analogicznie do tego jak zostało to opisane w podrozdziale 2.6. W tym kontekście, zaletą podejścia sieciowego, w stosunku do przedstawionych w punkcie 4.2.1 modeli hybrydowych ARMA-GARCH-LSTM, jest niezależność modelu od wyników uzyskanych przez modele GARCH. Ponieważ rezultatem prognoz są parametry całego rozkładu prawdopodobieństwa, a nie tylko punktowe oceny zmienności, podczas szacowania VaR oraz ES nie jest konieczne korzystanie z prognoz wartości oczekiwanej (oraz ewentualnie liczby stopni swobody i skośności), uzyskanych wcześniej przy pomocy modeli klasy GARCH.

Zaznaczyć należy także, że struktura opisywanych w tym punkcie modeli, pozwala również na uzyskanie prognoz punktowych stóp zwrotu ($r_{t+1|t}$) i zmienności ($\sigma_{t+1|t}$) jednak nie będzie to przedmiotem rozważań w niniejszej pracy.

Minusem przedstawionej powyżej struktury modelu jest konieczność założenia danego typu rozkładu warunkowego. Ponieważ rozkład nie jest znany, lepszym

podejściem byłoby takie, które nie wymagałoby zakładania z góry typu rozkładu. Takie rozwiązanie dopuszczają przykładowo modele umożliwiające prognozowanie poszczególnych kwantyli rozkładu (zob. Chen i in. [2020]). Naturalnym wydaje się także zastosowanie w tym miejscu sieci bayesowskich (zob. podrozdział 3.1.4), jednak wykracza to poza tematykę poruszaną w tej pracy.

Modele prognoz probabilistycznych ocenione zostały przy pomocy mierników prognoz rozkładów prawdopodobieństwa, opisanych w punkcie 2.5.4, natomiast wyniki empiryczne zaprezentowano w punkcie 5.6.1. Wyniki dotyczące szacowania ryzyka, uzyskane przez modele prognoz probabilistycznych, przedstawione zostały w punkcie 5.6.2, natomiast porównanie wyników modeli prognoz probabilistycznych w zestawieniu z modelami punktowymi i GARCH pod względem szacowania ryzyka przedstawione zostało w punkcie 6.1.2.

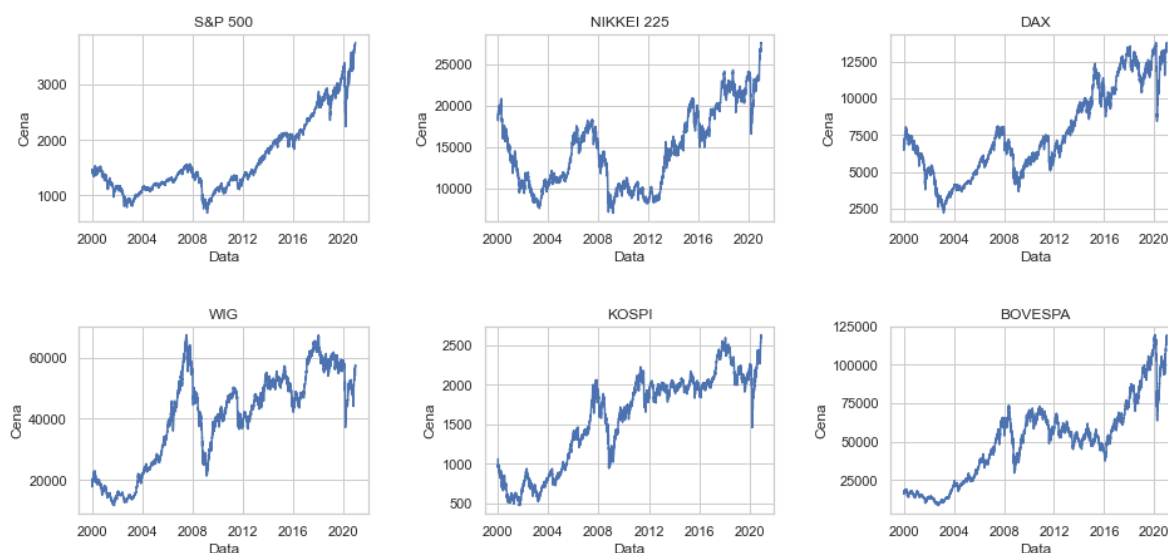
Rozdział 5

Empiryczna ewaluacja modeli predykcyjnych

Szeroko zakrojone badania empiryczne stanowią istotny element niniejszej rozprawy, pozwalający na dokładną ocenę możliwości predykcyjnych analizowanych modeli. Szczegółowa prezentacja rezultatów badań przedstawiona została w poniższym rozdziale, zaś w rozdziale 6 przeprowadzona została analiza porównawcza poszczególnych specyfikacji modelowych. Wyniki omówione zostały także w odniesieniu do celów i hipotez badawczych.

5.1 Charakterystyka analizowanych zbiorów danych

Podstawę badań empirycznych przeprowadzonych w niniejszej pracy stanowią dane dotyczące wybranych indeksów giełdowych, reprezentujących rynki o różnych stopniach rozwoju gospodarczego. Podział na rynki rozwinięte i wschodzące dokonany został na podstawie klasyfikacji MSCI (ang. *Morgan Stanley Capital International* zob. MSCI [2021]). W wyborze poszczególnych kierowano się także położeniem geopolitycznym. Spośród rynków rozwiniętych wybrane zostały indeksy giełdowe: S&P 500 (Stany Zjednoczone Ameryki), DAX (Niemcy) oraz NIKKEI 225 (Japonia). Natomiast z rynków wschodzących wybrane zostały indeksy: Bovespa (Brazylia), WIG (Polska), oraz KOSPI (Korea Południowa). W badaniach wykorzystane zostałyienne notowania poszczególnych indeksów, z okresu od 3 stycznia 2000 roku do 31 grudnia 2021 roku, pochodzące z serwisu finansowego Stooq (zob. stooq.com). Wykresy dziennych notowań wspomnianych indeksów giełdowych przedstawione zostały na rysunku 5.1.

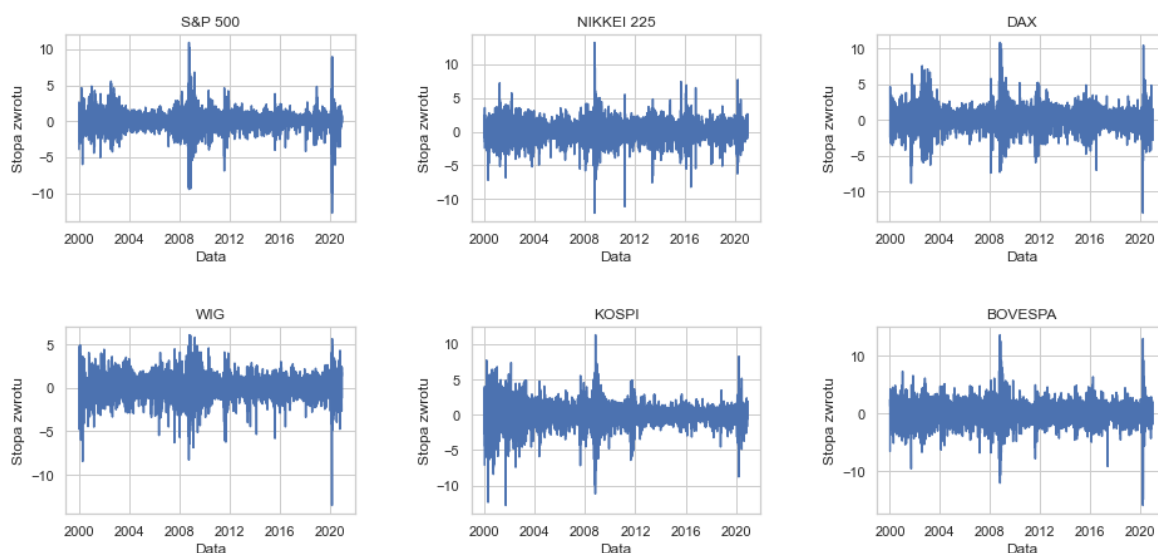


Rysunek 5.1: Dzielne notowania wybranych indeksów giełdowych w okresie od 3.01.2000 do końca 31.12.2021 roku.

Źródło: Opracowanie własne.

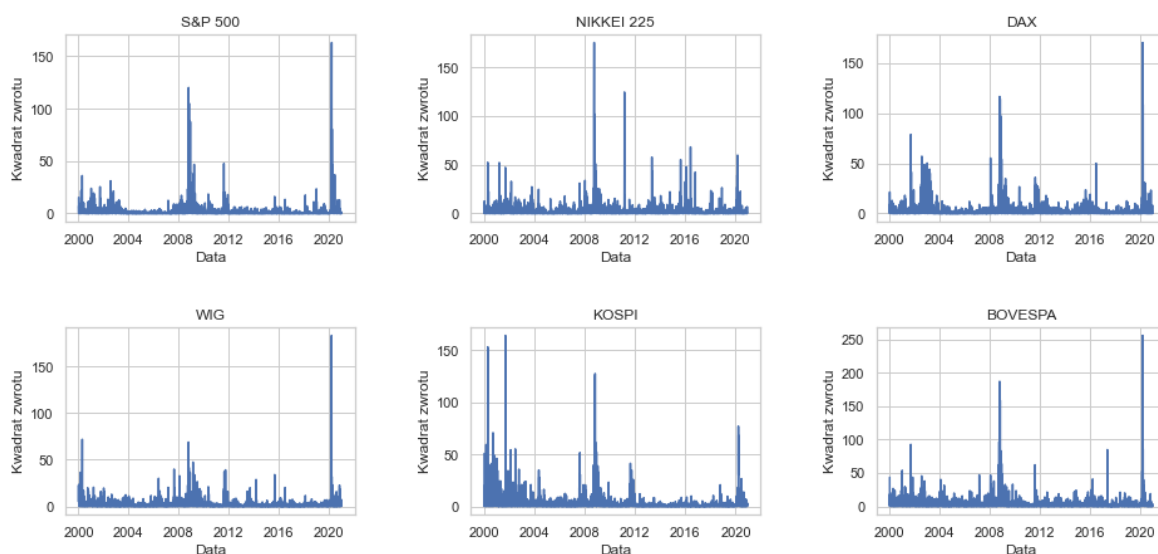
Na wykresach daje się zauważyć trend wzrostowy dla notowań wszystkich indeksów, ale także duże spadki notowań w latach 2008 (związane z globalnym kryzysem finansowym) oraz w roku 2020 (w związku z początkiem pandemii COVID-19), przy czym zauważyć można, że załamanie notowań trwało znacznie dłużej w przypadku kryzysu globalnego, co szczególnie widoczne jest na przykładzie indeksu NIKKEI.

Wartości indeksów przekształcone zostały do postaci logarytmicznych stóp zwrotu, wyrażonych w punktach procentowych i w takiej postaci wykorzystywane były jako dane wejściowe do poszczególnych modeli. Kształtowanie się szeregów stóp zwrotu przedstawione zostało na rysunku 5.2. W zaprezentowanych na rysunku szeregach zwrócić można uwagę na występowanie obserwacji odstających (nietypowych) zauważyć można także zgrupowania zmienności, widoczne szczególnie w okresach kryzysowych. Okresy o zwiększonej zmienności lepiej widoczne są na rysunku 5.3, przedstawiającym kwadraty stóp zwrotu.



Rysunek 5.2: Logarymiczne stopy zwrotu wybranych indeksów giełdowych, wyrażone w punktach procentowych.

Źródło: Opracowanie własne.

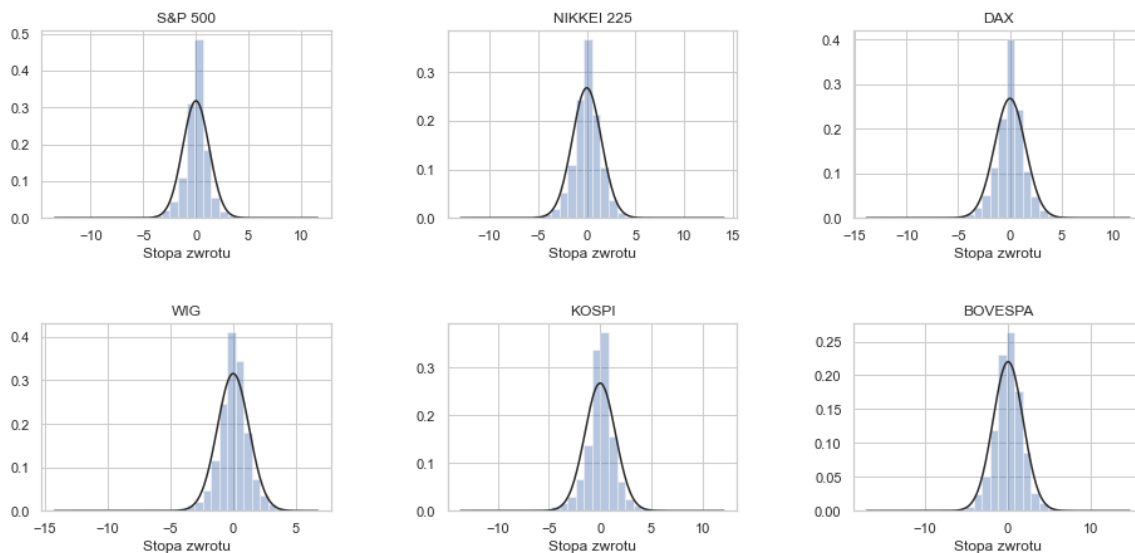


Rysunek 5.3: Kwadraty stóp zwrotu wybranych indeksów giełdowych.

Źródło: Opracowanie własne.

Wspomniane powyżej okresy o podwyższonej zmienności w latach 2008-2009 oraz 2020-2021 zauważyć można na wszystkich wykresach, przy czym najniższe wahania w okresie 2008-2009 zanotował indeks WIG. W przypadku indeksu KOSPI, zauważyć można także okres dużej zmienności w latach 2000-2004, związany prawdopodobnie z azjatyckim kryzysem finansowym. Z kolei załamanie rynków związane z początkiem pandemii w 2020 roku charakteryzuje się bardzo dużymi wahaniami występującymi w krótkim okresie czasu. W tym przypadku, rynki azjatyckie (NIKKEI oraz KOSPI) wykazują niższe wahania niż dla pozostałych indeksów.

Rysunek 5.4 przedstawia rozkłady empiryczne stóp zwrotu poszczególnych indeksów, z dopasowanym rozkładem normalnym. Na wykresach można zauważyć, że rozkłady empiryczne charakteryzują się występowaniem ilości odstających oraz znacznie większą koncentracją wokół średniej.



Rysunek 5.4: Rozkłady wybranych instrumentów finansowych z naniesionym rozkładem normalnym.
Źródło: Opracowanie własne.

Dokładny opis statystyczny danych przedstawiony został w tabeli 5.1, przedstawionej poniżej. Zaprezentowane w niej charakterystyki opisowe rozważanych szeregów czasowych potwierdzają właściwości zaobserwowane na podstawie analizy wykresowej. Dla wszystkich szeregów wartości średniej zbliżone są do zera, natomiast wartości minimalne oraz maksymalne są na poziomie 10-13%, co wskazuje na istnienie wartości znacząco odstających od średniej. Ujemny współczynnik skośności wskazuje z kolei na lewostronną asymetrię rozkładu. Wysoki współczynnik kurtozy świadczy natomiast o leptokurtyczności i grubych ogonach rozkładów, zaś wysokie wartości współczynników zmienności wskazują na ogólnie wysoki poziom zmienności badanych szeregów.¹ Na tej podstawie można stwierdzić, że rozważane szeregi czasowe przejawiają typowe własności stóp zwrotu (zob. punkt 2.2.3).

¹Współczynnik zmienności obliczono jako iloraz odchylenia standardowego i średniej, bez mnożenia wyniku przez czynnik 100.

Tabela 5.1: Opis statystyczny danych.

Charakterystyka	S&P	N225	DAX	WIG	KOSPI	IBOV
Ilość obs.	5283	5148	5329	5256	5163	5196
Średnia	0,0179	0,0071	0,0133	0,0209	0,0173	0,0375
Odch. std.	1,2554	1,4895	1,4909	1,2651	1,4938	1,8124
Wsp. zmien.	70,1340	209,7887	112,0977	60,5311	86,3468	48,3306
Min	-12,7652	-12,1110	-13,0549	-13,5265	-12,8047	-15,9938
Maks	10,9572	13,2346	10,7975	6,0834	11,2844	13,6782
Skośność	-0,3931	-0,3736	-0,1643	-0,5923	-0,5710	-0,3732
Kurtoza	10,9462	6,3532	5,7466	5,8565	7,0621	6,4722

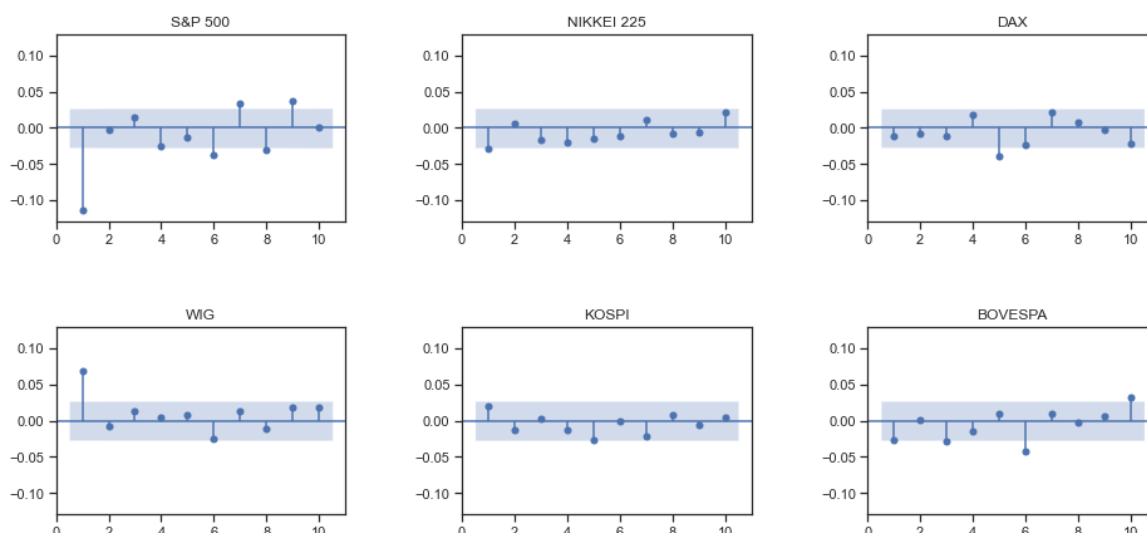
Źródło: Opracowanie własne.

Zaprezentowane powyżej charakterystyki opisowe rozważanych szeregów czasowych potwierdzają właściwości zaobserwowane na podstawie analizy wykresowej. Dla wszystkich szeregów wartości średniej zbliżone są do zera, natomiast wartości minimalne oraz maksymalne są na poziomie 10-13%, co wskazuje na istnienie wartości znacząco odstających od średniej. Ujemny współczynnik skośności wskazuje z kolei na lewostronną asymetrię rozkładu. Wysoki współczynnik kurtozy świadczy natomiast o leptokurtyczności i grubych ogonach rozkładów, zaś wysokie wartości współczynników zmienności wskazują na ogólnie wysoki poziom zmienności badanych szeregów.² Na tej podstawie można stwierdzić, że rozważane szeregi czasowe przejawiają typowe własności stóp zwrotu (zob. punkt 2.2.3).

W kolejnym kroku analizie poddane zostaną funkcje ACF oraz PACF dla stóp zwrotu oraz ich kwadratów. Wykresy tych funkcji dla stóp zwrotu badanych indeksów przedstawione zostały kolejno na rysunkach 5.5 i . Liczba opóźnień ograniczona została w tym przypadku do dziesięciu, natomiast poziom istotności przekroczeń ustalony został na 5%.

Analizując funkcję ACF można przede wszystkim zaobserwować istotne pierwsze opóźnienia dla indeksów S&P 500 oraz WIG. Dodatkowo, w przypadku indeksu S&P 500 można zauważyć nieznaczne zwiększenie istotności przy opóźnieniach rzędów 6, 7, 8 oraz 9, podobnie dla indeksu DAX na piątym opóźnieniu i dla indeksu BOVESPA na opóźnieniach rzędów 1, 3, 6 oraz 9.

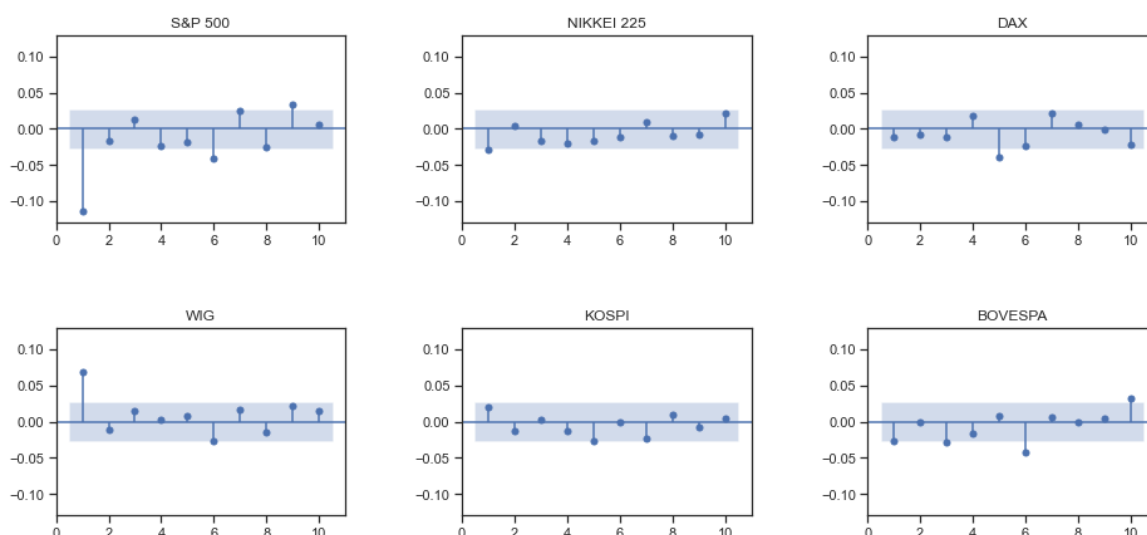
²Współczynnik zmienności obliczono jako iloraz odchylenia standardowego i średniej, bez mnożenia wyniku przez czynnik 100.



Rysunek 5.5: ACF dla stóp zwrotu poszczególnych indeksów.

Źródło: Opracowanie własne.

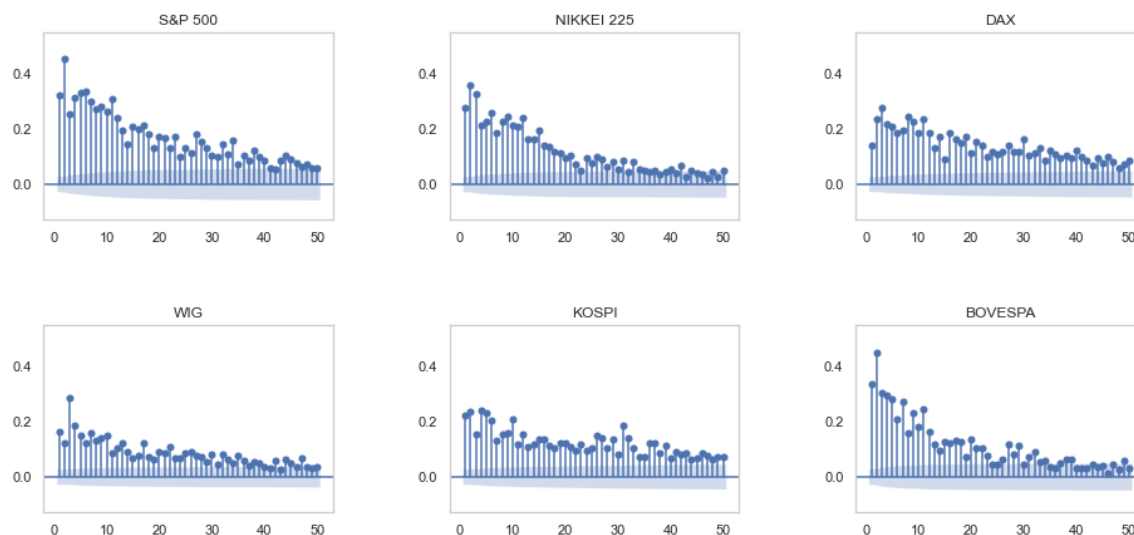
W sposób bardzo zbliżony do wykresów ACF, prezentują się także przebiegi funkcji autokorelacji cząstkowych (PACF, zob. rys. 5.6). Według tradycyjnego podejścia Boxa i Jenkinsa do ustalenia rzędów opóźnień modeli $ARMA(k, m)$ na podstawie wykresów ACF i PACF wydaje się, że do opisu stóp zwrotu indeksów S&P 500 oraz WIG należałoby ustalić $k = 1, m = 0$, zaś dla pozostałych indeksów $k = m = 0$.



Rysunek 5.6: PACF dla stóp zwrotu poszczególnych indeksów.

Źródło: Opracowanie własne.

Przebiegi funkcji ACF i PACF przeanalizować można także dla kwadratów oraz wartości bezwzględnych stóp zwrotu. Wykresy dla kwadratów przedstawione zostały na rysunkach 5.7 oraz 5.8. Liczba opóźnień w tym przypadku ustalona została na 50.



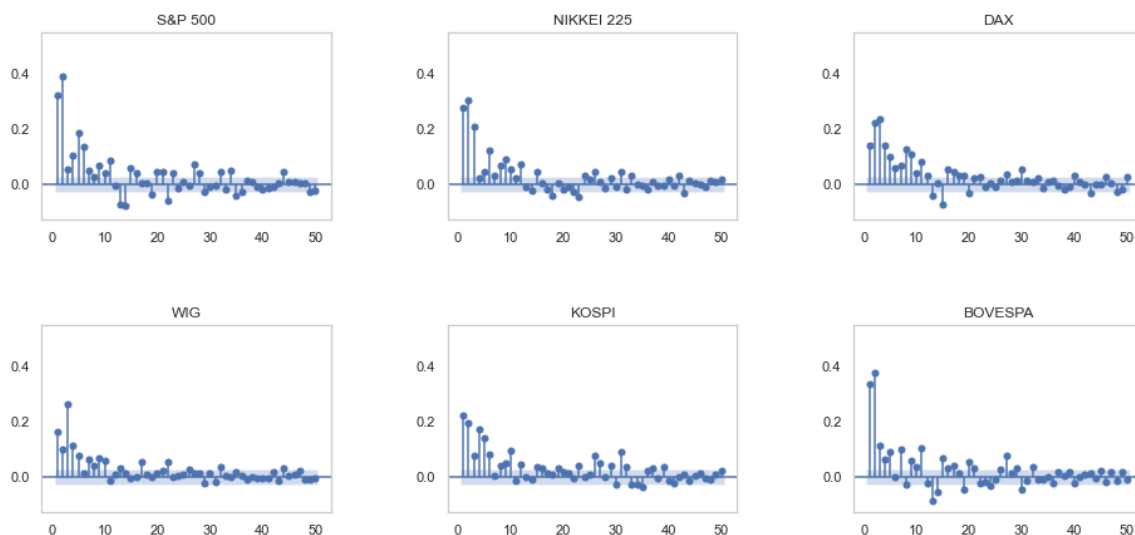
Rysunek 5.7: ACF dla kwadratów stóp zwrotu.

Źródło: Opracowanie własne.

Przebiegi ACF dla kwadratów stóp zwrotu wskazują na istotne, stopniowo wygasające autokorelacje. W przypadku indeksów S&P 500, NIKKEI oraz BOVESPA początkowo wyższa istotność opóźnień wygasa w okolicach 20-30 opóźnień. Dla indeksów DAX, WIG oraz KOSPI początkowa istotność opóźnień jest niższa, a wygasanie znacznie wolniejsze. Niemniej jednak, dla wszystkich badanych indeksów zaobserwować można silne, powolnie wygasające autokorelacje dla kwadratów stóp zwrotu, co może świadczyć o pewnej persystencji zmienności.

Z kolei na podstawie funkcji PACF dla kwadratów stóp zwrotu, zaprezentowanej poniżej, zaobserwować można znacznie szybsze wygasanie istotności opóźnień w przypadku indeksów S&P 500, NIKKEI oraz BOVESPA w okolicach piątego opóźnień, natomiast dla indeksów DAX, WIG oraz KOSPI w okolicach 8-10 opóźnień, co potwierdza występowanie autokorelacji dla tych szeregów.

Funkcje ACF oraz PACF dodatkowo przeanalizowane zostały dla wartości bezwzględnych stóp zwrotu oraz dla oszacowań zmienności uzyskanych z wykorzystaniem estymatora GKYZ. Wyniki zamieszczone zostały w Aneksie (kolejno na rysunkach A.1, A.2 oraz A.3 i A.4). W przypadku wartości bezwzględnych oraz estymatora GKYZ wyniki są bardzo zbliżone do omówionych powyżej wyników dla kwadratów stóp zwrotu, z tym wyjątkiem, że wygasanie w przypadku funkcji ACF jest nieco wolniejsze.

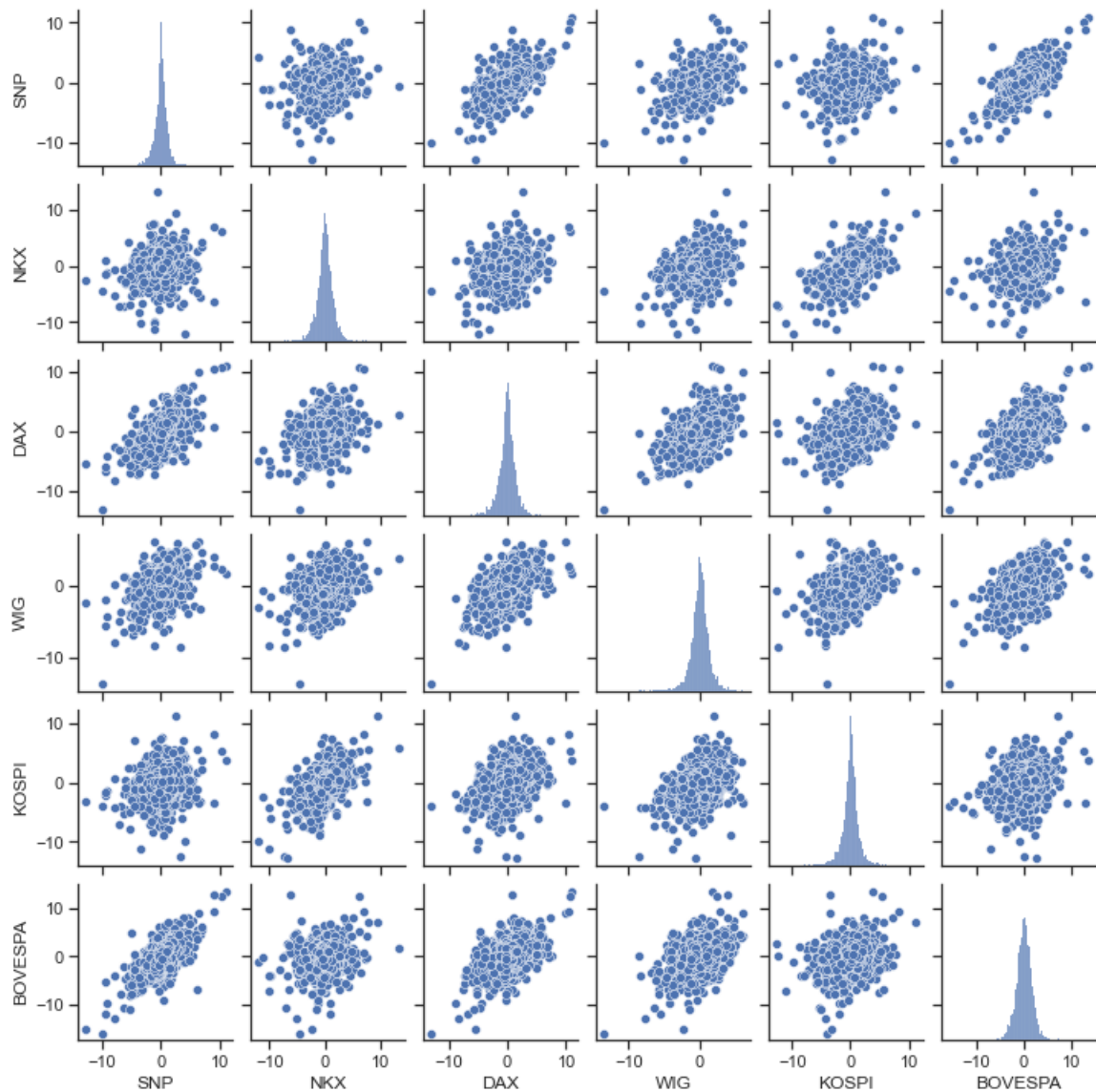


Rysunek 5.8: PACF dla kwadratów stóp zwrotu.

Źródło: Opracowanie własne.

W kolejnym kroku zbadano korelacje pomiędzy szeregami stóp zwrotu badanych indeksów. W pierwszej kolejności, na rysunku 5.9 wyniki zaprezentowane zostały w postaci wykresów punktowych dla par poszczególnych aktywów.

Wartości współczynników korelacji Pearsona, również dla par poszczególnych indeksów, zawarto w tabeli 5.2. Analizując przedstawione wyniki, można przede wszystkim zauważyć, że wszystkie pary charakteryzują się korelacją dodatnią. Indeksy S&P 500 i BOVESPA, oraz S&P 500 i DAX wykazują najsilniejszą korelację (współczynniki na poziomie odpowiednio 0,6186 oraz 0,6137). Indeksy NIKKEI i KOSPI wykazują dodatnią korelację na poziomie 0,5787, natomiast indeksy WIG oraz DAX na poziomie 0,5343. Takie wartości współczynników wskazują na korelację zauważalną, jednak nie szczególnie silną. Zależności pomiędzy tymi parami indeksów można zaobserwować także na przedstawionych wykresach.



Rysunek 5.9: Wykresy punktowe dla par poszczególnych indeksów.

Źródło: Opracowanie własne.

Tabela 5.2: Wartości współczynników korelacji Pearsona

	S&P	N225	DAX	WIG	KOSPI	IBOV
S&P	1,0000	0,1490	0,6137	0,3748	0,1828	0,6186
N225	0,1490	1,0000	0,2826	0,3099	0,5787	0,1643
DAX	0,6137	0,2826	1,0000	0,5343	0,3108	0,4708
WIG	0,3748	0,3099	0,5343	1,0000	0,3592	0,3675
KOSPI	0,1828	0,5787	0,3108	0,3592	1,0000	0,2125
IBOV	0,6186	0,1643	0,4708	0,3675	0,2125	1,0000

Źródło: Opracowanie własne.

Analizując przedstawione wyniki, można przede wszystkim zauważyć, że wszystkie pary charakteryzują się korelacją dodatnią. Indeksy S&P 500 i BOVESPA, oraz S&P 500 i DAX wykazują najsilniejszą korelację (współczynniki na poziomie odpowiednio 0,6186 oraz 0,6137). Indeksy NIKKEI i KOSPI wykazują dodatnią korelację na poziomie 0,5787, natomiast indeksy WIG oraz DAX na poziomie 0,5343. Takie wartości współczynników wskazują na korelację zauważalną, jednak nie szczególnie silną. Zależności pomiędzy tymi parami indeksów można zaobserwować także na przedstawionych wykresach.

Na tej podstawie zauważyć można, że silną korelacją charakteryzują się indeksy krajów pochodzących z tego samego obszaru geograficznego (S&P 500 i BOVESPA, DAX i WIG, NIKKEI i KOSPI). Można zatem wnioskować, że duży wpływ na korelację pomiędzy indeksami poszczególnych krajów ma ich położenie geopolityczne, natomiast w mniejszym stopniu klasyfikacja rynków ze względu na rozwój gospodarczy – w tym wypadku co prawda silną korelację wykazują indeksy S&P 500 i DAX, jednak niską korelacją charakteryzują się indeksy S&P 500 i NIKKEI (wartość współczynnika 0,1490) oraz DAX i NIKKEI (0,2826). Niska korelacja (na poziomie 0,2 - 0,36) występuje także w przypadkach par indeksów z rynków wschodzących.

5.2 Budowa oprogramowania i algorytmów testujących

W celu budowy oprogramowania wykorzystanego przy implementacji metod ekonometrycznych i sieci uczenia głębokiego wykorzystane zostały języki Python w wersji 3.7.10 oraz R w wersji 4.1.0, wraz z dostępnymi wybranymi pakietami oraz bibliotekami.

W środowisku Python wykorzystane zostały przede wszystkim biblioteki:

- TensorFlow (wersje 2.5-2.8) oraz Tensorflow Probability (wersja 0.16.0) - podstawowe biblioteki implementujące mechanizmy uczenia głębokiego,
- Keras (wersja 2.4) - wysokopoziomowe API (ang. *Application Programming Interface*) dla bibliotek TensorFlow,
- KerasTuner - biblioteka stosowana do dostrajania hiperparametrów sieci,
- Numpy oraz Pandas - biblioteki wykorzystywane do przetwarzania danych,
- Statsmodels oraz SciPy - biblioteki wykorzystywane do modelowania i obliczeń statystycznych,

- Matplotlib oraz Seaborn - biblioteki wykorzystywane do opracowania wykresów.

Z kolei w środowisku programistycznym R wykorzystane zostały biblioteki:

- rugarch - biblioteka umożliwiająca estymację i prognozowanie z wykorzystaniem jednowymiarowych modeli ARMA-GARCH,
- tidyverse oraz xts - biblioteki wykorzystywane do przetwarzania danych³,
- TTR oraz FinTS - biblioteki stosowane do pracy z danymi finansowymi
- tensorflow oraz keras - pakiety umożliwiające korzystanie z tych bibliotek w środowisku R (wymagają środowiska Python),
- scoringRules oraz scoringutils - pakiety umożliwiające ocenę trafności prognoz probabilistycznych.

Obliczenia wykonywane były na komputerze wyposażonym w podzespoły:

- procesor: AMD Ryzen 7 3700X 3,6GHz,
- pamięć: 16GB RAM,
- karta graficzna: NVIDIA GeForce RTX 2060, wyposażona w rdzenie tensorowe.

Taka konfiguracja maszyny umożliwiała stosunkowo szybkie wykonywanie obliczeń, bez konieczności wprowadzania ograniczeń dotyczących danych wejściowych w sieciach neuronowych (np. wielkości partii wsadowej). Rdzenie tensorowe umożliwiały dodatkowo wykorzystanie procesora graficznego do przyspieszenia obliczeń w procesie uczenia sieci. Czas obliczeń dla modeli sieciowych dla pełnego zbioru danych (ok. 5000 obserwacji) wynosił w granicach od 5-30 minut, w zależności od struktury sieci, wartości hiperparametrów oraz liczby epok. W przypadku modeli ARMA-GARCH czas potrzebny na estymację oraz prognozowanie wynosił około 5 minut dla pojedynczej specyfikacji modelu.

W ramach badań przygotowano zostały autorskie kody, w których zaimplementowano wyżej wymienione biblioteki w celu rozwiązania zadanych problemów badawczych oraz prezentacji wyników. Zaprogramowane zostały narzędzia umożliwiające uczenie, estymację oraz prognozowanie przy pomocy sieci uczenia głębokiego modeli ekonometrycznych oraz modeli hybrydowych.

W kontekście prognozowania punktowego stóp zwrotu, zaproponowano autorskie algorytmy testujące modele sieci neuronowych, umożliwiające podział danych na okna

³Tidyverse jest zbiorem bibliotek (zob. tidyverse.org)

treningowe, walidacyjne oraz testowe (z wykorzystaniem okna rolowanego), oraz ocenę uzyskanych prognoz w ramach testu wstecznego, przy pomocy mierników trafności prognoz. Algorytmy te służyły także do budowy prostych strategii inwestycyjnych, bazujących na uzyskanych prognozach, oraz ich ocenę z wykorzystaniem mierników oceny strategii.

Zaimplementowane zostały także narzędzia umożliwiające budowę i wykorzystanie modeli hybrydowych ARMA-GARCH-LSTM, w celu uzyskania punktowych prognoz zmienności. Algorytmy te umożliwiały estymację i prognozowanie z wykorzystaniem 12 specyfikacji modeli ARMA-GARCH, połączenie ich w 12 kolejnych modelach hybrydowych oraz ocenę prognoz uzyskanych za pomocą tych modeli z wykorzystaniem mierników oceny prognoz, a także w kontekście szacowania wartości zagrożonej i oczekiwanego niedoboru.

Następnie zaprogramowane zostały zaproponowane przez autora algorytmy testujące dla modeli prognoz probabilistycznych NN-D opartych na sieciach uczenia głębokiego. Jako dodatek do algorytmów opisanych powyżej (służących do uczenia sieci, oceny prognoz z wykorzystaniem odpowiednich miar oraz oceny prognoz ryzyka), w celu umożliwienia poprawnego prognozowania rozkładów prawdopodobieństwa zaimplementowane zostały także funkcje straty (zob. wzory 4.7, 4.8, 4.9), umożliwiające prognozowanie parametrów rozkładów normalnego, t-Studenta oraz skośnego rozkładu t-Studenta.

W celu usprawnienia procesu dostrajania hiperparametrów, przygotowane zostały autorskie kody umożliwiające częściową automatyzację tego procesu (z wykorzystaniem gotowych bibliotek). Dodatkowo przygotowane zostały kody umożliwiające zaprezentowanie wyników prowadzonych badań w formie wykresów graficznych. Opisane powyżej algorytmy i narzędzia tworzone były w dużej mierze od podstaw, z myślą o ich skalowalności oraz możliwości późniejszego wykorzystania w kolejnych badaniach.

5.3 Optymalizacja hiperparametrów modeli sieciowych

Kluczowym aspektem podczas projektowania architektury każdej sieci neuronowej jest odpowiedni dobór hiperparametrów. Proces dostrajania hiperparametrów może mieć duże znaczenie na wyniki uzyskiwane przez sieć, dlatego warto poświęcić mu dużo uwagi podczas przygotowywania struktury sieci. Najczęściej stosowane są dwa podejścia: heurystyczne, w którym dobór poszczególnych parametrów sieci odbywa się na podstawie wartości podanych w badaniach opublikowanych w literaturze naukowej,

lub poprzez proces dostrajania poszczególnych hiperparametrów (ang. *hyperparameters tuning*).

Dobór hiperparametrów poprzez ich dostrajanie jest procesem czasochłonnym, głównie ze względu na ilość hiperparametrów oraz zakres wartości, które mogą one przyjmować. Pomocne w tym procesie mogą okazać się zautomatyzowane algorytmy pozwalające na dobór optymalnych wartości w sposób znacznie szybszy niż manualnie. Do podstawowych narzędzi stosowanych w tym celu należy biblioteka KerasTuner (zob. rozdział 3). Jednym z proponowanych w zakresie tego narzędzia algorytmów, jest tzw. algorytm *hyperband*. Sposób działania algorytmu w szczególności dobrze nadaje się do sytuacji, w których zoptymalizowane muszą być liczne konfiguracje wartości hiperparametrów. Działa on na zasadzie turniejowej – wartości parametrów sprawdzane są w zestawach (ang. *brackets*), z których na niewielkiej liczbie epok testowane są optymalne wartości hiperparametrów. Wartości te przekazywane są do kolejnego etapu, w którym dostrajanie jest powtarzane dla większej ilości epok. Proces ten można powtarzać dla dowolnego zestawu hiperparametrów (nie tylko jednego), co pozwala na jednoczesne sprawdzenie w jaki sposób pary (lub większe ilości) hiperparametrów wpływają na poprawienie wyników uzyskiwanych przez sieć.

Ponieważ nawet z pozoru niewielkie zmiany w wartościach mogą znacząco przyczynić się do poprawy lub pogorszenia wyników, warto sprawdzić wpływ zmian wartości dla możliwie szerokiego zakresu hiperparametrów. Sprawdzanie zmian wartości powinno odbywać jedynie na części dostępnych danych uczących, a także bez sprawdzania zmian na zbiorze testowym, aby uniknąć problemu wychodzenia w przyszłość (ang. *look ahead bias*). Z tego względu, w przypadku stosowania metod walidacji krzyżowej (takich jak schemat rolowany lub rekursywny), zmian dokonywać należy jedynie na pojedynczym wybranym oknie. Ze względu na to, że jest to proces czasochłonny, wielu badaczy pomija go stosując wartości parametrów wybrane heurystycznie. W tych badaniach zdecydowano się jednak na przeprowadzenie bardzo dokładnego dostrajania, w celu uzyskania możliwie najlepszych wyników. Proces ten był jednym z najbardziej czasochłonnych zadań praktycznych wykonanych w ramach przedstawionych w tu badań. Proces dostrajania hiperparametrów przeprowadzony został po części „ręcznie” (poprzez samodzielne zmiany wartości poszczególnych parametrów w kodzie), a częściowo z wykorzystaniem narzędzia KerasTuner. Tabela 5.3 przedstawia zestawienie dostrajanych parametrów, przedziały testowanych wartości, wartości wybrane jako optymalne, oraz sposób wyboru tych wartości. Dostrajanie hiperparametrów wykonano w pierwszej kolejności dla sieci wykorzystywanych do uzyskania punktowych prognoz stóp zwrotu. Wartości uzyskane

w tym procesie stanowiły punkt wyjściowy także dla modeli hybrydowych i modeli prognozowania probabilistycznego, jednak wobec zmiany struktury modelu (i prognozowanych danych), wymagały dalszych usprawnień.

Tabela 5.3: Dostrajanie hiperparametrów dla sieci wykorzystywanych w prognozowaniu punktowym stóp zwrotu.

Hiperparametr	Wartości testowane	Wartość wybrana	Narzędzie
Ilość warstw ukrytych	1-5	3	Ręcznie
Ilość neuronów w warstwie	8-700	512/256/128	KerasTuner
<i>Dropout</i>	0-0,5	0,0002	KerasTuner
Regularyzacja l_2	0-0,5	0,00001	KerasTuner
Optymalizator	Adam/RMSProp/SGD	Adam	Ręcznie
Prędkość uczenia	0,0001-0,5	0,0015	KerasTuner
Długość sekwencji	1-200	10	Ręcznie
Długość okna	252-Exp./21-1008	2016/756	Ręcznie
Rozmiar partii wsadowej	1-Exp.	756	Ręcznie
Liczba epok	10-100	200 (MC)	Ręcznie
<i>Kernel size</i> *	1-5	2	Ręcznie
Liczba filtrów*	10-512	256	KerasTuner
<i>Pool size</i> *	1-5	2	Ręcznie

Uwaga: Oprócz optymalizatora Adam przetestowane zostały także jego wariacje. Exp. oznacza zbiór okna rozszerzanego, zaś MC oznacza zatrzymanie uczenia przy wykorzystaniu punktu kontrolnego modelu. * - dotyczy sieci CNN.

Źródło: Opracowanie własne.

Dodatkowe zmiany, które wprowadzone zostały w strukturze sieci LSTM wykorzystanej w modelu hybrydowym, służącym do punkowego prognozowania zmienności:

- liczba neuronów: 128/512/128 (KerasTuner),
- l_2 : 0,00002,
- prędkość uczenia: 0,005,
- długość sekwencji: 3,
- długość zbiorów: Exp.(od 1008)/504,
- rozmiar partii wsadowej: Exp.

- epoki: 150(MC).

Z kolei w modelu prognozowania probabilistycznego zmienione zostały następujące hiperparametry architektury sieci LSTM oraz CNN:

- liczba neuronów: 128/64/32,
- l2: 0,002,
- *dropout*: 0,02,
- prędkość uczenia: 0,002,
- długość sekwencji: 3,
- długość zbiorów: Exp.(od 1008)/504,
- rozmiar partii wsadowej: 128,
- epoki: 300(MC).

Warto podkreślić, że finalnie wybrane wartości hiperparametrów mogą nie być wartościami optymalnymi. Wynika to przede wszystkim z faktu, że zmiany tych wartości sprawdzane powinny być w różnych konfiguracjach dla zestawów hiperparametrów (osobno, parami, trójkami itd.). Jednak ze względu na dużą ilość kombinacji i zakresu wartości, przetestowanie wszystkich kombinacji nie byłoby możliwe w ramach tych badań. Problem ten częściowo rozwiązuje stosowanie wspomnianego algorytmu KerasTuner, który sprawdza możliwie dużą ilość poszczególnych kombinacji, jednak nawet jego wykorzystanie do analizy wszystkich kombinacji nie byłoby czasowo możliwe. Ponieważ wyniki uzyskane przez sieci, przedstawione w poniższych podrozdziałach, są w dużym stopniu zależne od wartości hiperparametrów, należy założyć, że mogą one nie być jednoznacznie optymalne i istnieje możliwość ich poprawy.

5.4 Prognozowanie punktowe stóp zwrotu

W poniższym podrozdziale przedstawione zostały wyniki prognoz punktowych stóp zwrotu, uzyskanych przez modele oparte na wybranych typach sieci neuronowych: klasyczne sieci MLP, oraz sieci uczenia głębokiego typu LSTM oraz CNN. Wyniki zaprezentowane zostały dla wszystkich badanych indeksów, przedstawionych w podrozdziale 5.1, i dotyczą ocen trafności prognoz punktowych

z wykorzystaniem poszczególnych mierników, a także w kontekście strategii inwestycyjnych.

W tej części badań jako dane wejściowe do sieci, a także jako jej wartości pożądane, wykorzystane zostały dane w postaci dziennych logarytmicznych stóp zwrotu, opisane w podrozdziale 5.1. W przypadku modeli wykorzystujących jako funkcję straty miernik MSE, dane zostały znormalizowane do zakresu $[-1; 1]$, natomiast w przypadku, gdy jako funkcja straty stosowana była miara MADL, normalizacja nie była stosowana, przede wszystkim ze względu na sposób konstrukcji tej miary.⁴

W procesie uczenia sieci stosowane było podejście walidacji krzyżowej, z wykorzystaniem okna rolowanego. Dane podzielone zostały na zbiory uczące, o długości 2056 obserwacji (8 lat), z których dodatkowo wydzielane były zbiory walidacyjne o długości 665 obserwacji, co stanowi 33% danych w zbiorze uczącym. Długość zbioru testowego, i zarazem okna rolowanego wynosiła w tym przypadku 756 obserwacji (3 lata). Łączna ilość prognoz dla pojedynczego indeksu wyniosła około 3267 obserwacji, co odpowiada okresowi od 3 stycznia 2008 roku do 31 grudnia 2020 roku. Wartości poszczególnych hiperparametrów sieci przedstawione zostały w tabeli 5.3.⁵

W pierwszej kolejności, w tabeli 5.4, przedstawione zostały wartości poszczególnych metryk (MSE, MAE, MAPE, MADE oraz trafność procentowa) dla trzech typów sieci, uczonych z wykorzystaniem miernika MSE jako funkcji straty (zob. punkt 2.5.2).

W wynikach zaprezentowanych w tabeli zauważyć można, że obie sieci uczenia głębokiego spisują się nieznacznie lepiej od klasycznych sieci MLP pod względem niemalże wszystkich mierników trafności. Pod względem miernika MSE, który wykorzystywany był jako funkcja straty, sieci LSTM uzyskały najlepsze wyniki dla czterech spośród sześciu badanych aktywów, co daje najlepszy wynik. Także pod względem trafności prognoz nieznacznie lepiej spisywały się sieci LSTM. Z kolei dla miary MADL, nieznacznie lepiej spisywały się sieci CNN, które najniższe wartości uzyskały dla trzech spośród badanych indeksów.

Dodatkowo zauważyć można, że wyniki nie są jednoznaczne. Jedynie w przypadku miar MSE oraz MAE wyniki są zbliżone (z wyjątkiem indeksu KOSPI), głównie ze względu na bardzo zbliżoną konstrukcję tych mierników. W pozostałych przypadkach niskie wyniki dla jednego miernika, przykładowo MSE, nie przekładają się na niskie wyniki miary MADL, czy też wysoką trafność procentową. Warto mieć ten fakt na

⁴Dane normalizowane byłby według wzoru $r'_t = a + \frac{(r_t - \min(r_t))(b-a)}{\max(r_t) - \min(r_t)}$, gdzie, w tym przypadku, $a = -1$, $b = 1$.

⁵Ilość dostępnych danych dla wybranych aktywów była różna, różna jest też ilość dni sesyjnych w ciągu roku. Ponieważ użyte zostały wszystkie dostępne dane, a długości okien są stałe, powstały niewielkie różnice w ilości prognozowanych obserwacji.

uwadze podczas wyboru funkcji straty, która optymalizowana jest podczas procesu uczenia sieci. Z tego względu, w dalszej części badań, sieci uczone będą także wykorzystując jako funkcję straty miarę MADL.

Tabela 5.4: Prognozowanie punktowe stóp zwrotu: mierniki błędów dla sieci trenowanych z wykorzystaniem MSE

Indeks	MSE	MAE	MADL	TRAFNOŚĆ
Sieci MLP				
S&P	1,8384	0,8378	-0,0238	50,902%
NKX	2,3771	1,0530	0,0141	49,904%
DAX	2,1450	0,9959	0,0141	51,071%
WIG	1,5936	0,8711	-0,0103	50,586%
KOSPI	1,5671	0,8189	0,0114	49,539%
BVP	3,2845	1,2520	0,0036	50,062%
Sieci CNN				
S&P	1,7904	0,8269	-0,0527	52,8007%
NKX	2,3747	1,0519	-0,0091	50,766%
DAX	2,1385	0,9910	0,0008	50,558%
WIG	1,5804	0,8656	0,0196	49,506%
KOSPI	1,5630	0,8133	-0,0390	52,430%
BVP	3,2488	1,2443	-0,0428	50,911%
Sieci LSTM				
S&P	1,8046	0,8275	-0,0381	52,372%
NKX	2,3569	1,0465	-0,0275	51,628%
DAX	2,1725	0,9882	-0,0327	51,946%
WIG	1,5763	0,8629	0,2510	50,185%
KOSPI	1,5541	0,8145	0,0177	49,857%
BVP	3,2427	1,2411	-0,0199	51,006%

Uwaga: Pogrubione zostały wartości najniższe dla mierników MSE oraz MADL, oraz wartości najwyższe dla trafności.

Źródło: Opracowanie własne.

Podsumowując powyższe rezultaty, spośród sieci badanych w kontekście oceny punktowej za pomocą wybranych miar trafności prognoz, najlepsze wyniki uzyskały sieci LSTM, generując nieznacznie lepsze wyniki w porównaniu z siecią CNN. Sieci MLP w tym przypadku uzyskały najslabsze rezultaty. Dokładniejsza analiza

porównawcza wyników uzyskanych przez modele bazujące na poszczególnych typach sieci przedstawiona została w rozdziale 6. Z kolei w poniższych podrozdziałach prognozy uzyskane przez poszczególne sieci ocenione zostaną pod kątem możliwości ich wykorzystania przy tworzeniu strategii inwestycyjnych.

5.4.1 Wyniki dla strategii wykorzystującej sieci MLP

W pierwszej kolejności, w tabeli 5.5, przedstawione zostały wyniki prognoz uzyskanych przez sieć MLP, w kontekście strategii inwestycyjnych. W tabeli zestawione zostały wyniki dla strategii pasywnej (ang. *Buy and Hold* - B&H), strategii opartych na prognozach stóp zwrotu uzyskanych przez sieci, gdzie prognozowany kierunek zmiany oznaczał sygnał kupna lub sprzedaży (ang. *Long-Short* - LS), oraz strategii, w której wykorzystane zostały jedynie sygnały kupna, a ujemny kierunek prognozy oznaczał wyjście z pozycji (ang. *Long Only* - LO).⁶ Do oceny efektywności strategii inwestycyjnych zastosowano mierniki przedstawione w punkcie 2.5.3.

Na podstawie zaprezentowanych wyników można zauważyć, że pod względem skumulowanego zwrotu wynik lepszy od strategii pasywnej udało się uzyskać jedynie w przypadku indeksu WIG dla strategii LO, natomiast w pozostałych przypadkach nie udało się uzyskać lepszych rezultatów. Zauważyć można także, że strategię LO cechują się mniejszą zmiennością (mają mniejsze wartości mierników aSD, MD oraz MLD). Dzięki temu, w przypadku indeksu S&P 500, wskaźniki IR są wyższe niż przypadku strategii pasywnej, mimo niższego skumulowanego zwrotu (aRC).

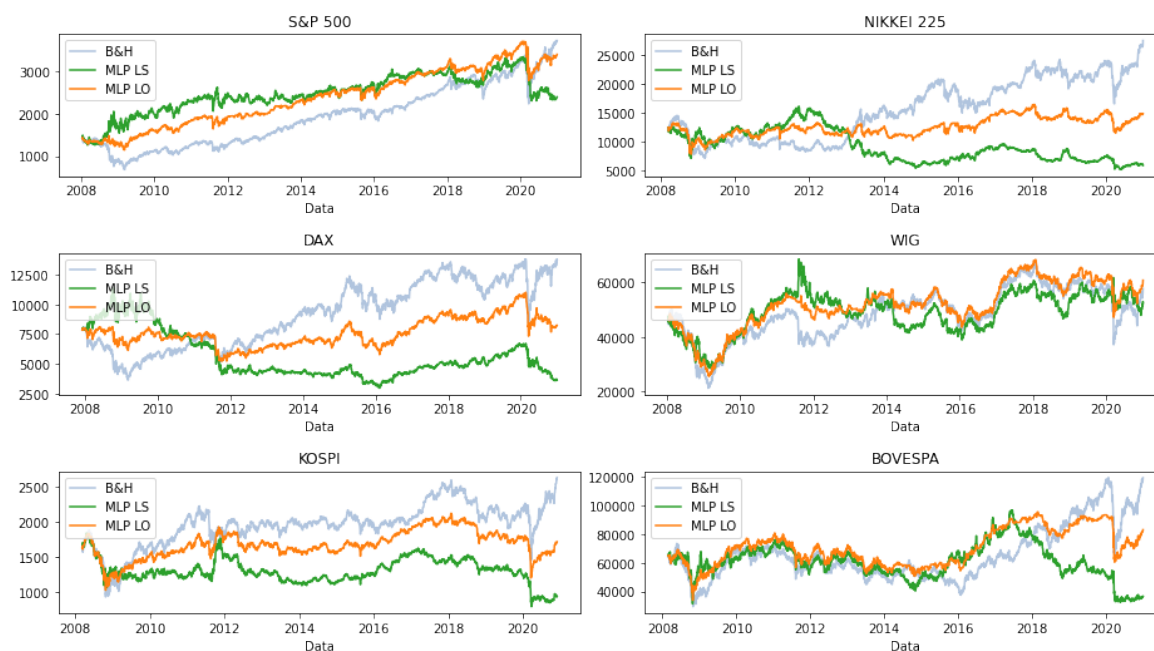
Rysunek 5.10 przedstawia zestawianie wizualnego przebiegu poszczególnych strategii dla wszystkich rozważanych aktywów. Zaobserwować na nim można, że dla indeksów S&P 500, WIG oraz BOVESPA, strategię LS i LO przez większość czasu utrzymywały się na podobnym (lub wyższym jak przypadku S&P 500) poziomie do strategii pasywnej, natomiast w przypadku pozostałych indeksów — przynosiły straty, przy czym krzywa kapitałowa dla strategii LO jest w większości przypadków nieznacznie wyżej niż dla strategii LS.

⁶Użycie strategii wykorzystującej tylko sygnały kupna znajduje uzasadnienie jedynie w przypadku aktywów charakteryzujących się stałym długoterminowym wzrostem, tak jak ma to miejsce w przypadku większości indeksów giełdowych.

Tabela 5.5: Prognozowanie punktowe stóp zwrotu: wyniki strategii dla sieci MLP trenowanych na MSE

Strategia	aRC	aSD	MD	MLD	IR*	IR**	IR***
Indeks S&P500							
B&H	7,74	21,12	52,58	4,30	0,37	0,05	0
LS	4,13	21,13	35,18	3,04	0,20	0,02	0
LO	6,98	15,69	24,91	1,17	0,45	0,12	0,01
Indeks NIKKEI 225							
B&H	6,62	24,33	51,31	4,78	0,27	0,04	0
LS	-5,92	24,34	68,52	9,00	-0,24	-0,02	0
LO	1,42	18,46	41,01	3,63	0,08	0,00	0
Indeks DAX							
B&H	4,34	22,97	54,6	5,43	0,19	0,02	0
LS	-5,65	22,98	73,8	12,28	-0,25	-0,02	0
LO	0,35	17,36	35,57	6,46	0,02	0,00	0
Indeks WIG							
B&H	1,74	19,74	57,62	3,13	0,09	0	0
LS	0,99	19,74	43,21	9,29	0,05	0	0
LO	2,15	15,36	47,38	2,90	0,14	0,01	0
Indeks KOSPI							
B&H	3,86	19,67	50,30	5,88	0,20	0,02	0
LS	-4,32	19,67	56,63	12,37	-0,22	-0,02	0
LO	0,38	15,75	43,78	5,08	0,02	0,00	0
Indeks BOVESPA							
B&H	4,94	28,47	59,96	9,15	0,17	0,01	0
LS	-4,43	28,47	66,26	5,12	-0,16	-0,01	0
LO	1,96	21,24	51,03	5,75	0,09	0,00	0

Źródło: Opracowanie własne.



Rysunek 5.10: Krzywe kapitałowe dla sieci MLP trenowanych na MSE.

Źródło: Opracowanie własne.

5.4.2 Wyniki dla strategii wykorzystującej sieci CNN

W tej części przedstawione zostały wyniki strategii zbudowanych z wykorzystaniem prognoz uzyskanych przy pomocy sieci CNN. Wskaźniki oceny strategii inwestycyjnych przedstawione zostały w tabeli 5.6. Oznaczenia poszczególnych strategii oraz wykorzystanych mierniki są analogiczne do tych przedstawionych w poprzednim podrozdziale, dotyczącym wyników uzyskanych z wykorzystaniem sieci MLP.

W przypadku sieci CNN, analizując wskaźnik skumulowanego zwrotu aRC, wyniki lepsze niż strategia pasywna udało się uzyskać dla trzech indeksów: S&P 500, KOSPI oraz BOVESPA (zarówno dla strategii LS oraz LO). W pozostałych przypadkach wartości skumulowanego zwrotu były niższe, dodatkowo, w przypadku indeksu WIG, strategię bazującą na modelu sieciowym przynosiły straty, podobnie jak w przypadku strategii LS dla indeksu DAX.

Tabela 5.6: Prognozowanie punktowe stóp zwrotu: wyniki strategii dla sieci CNN trenowanych na MSE

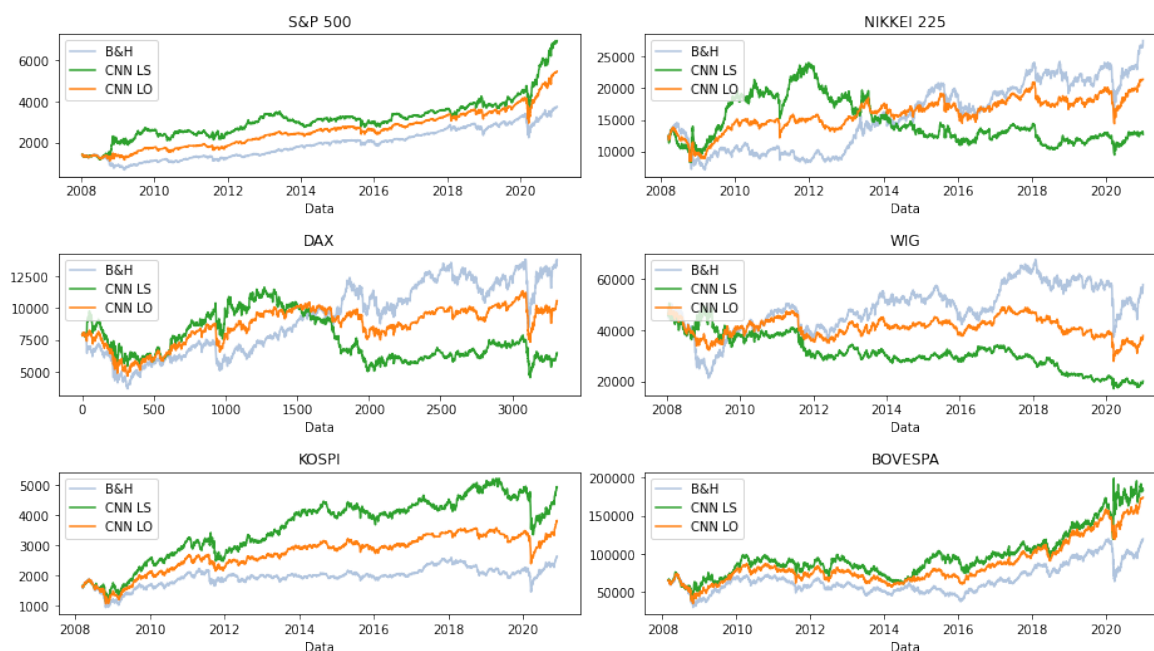
Strategia	aRC	aSD	MD	MLD	IR*	IR**	IR***
Indeks S&P500							
B&H	7,74	21,12	52,58	4,30	0,37	0,05	0,00
LS	13,01	21,11	28,52	4,08	0,62	0,28	0,01
LO	10,95	18,34	29,34	1,1	0,60	0,22	0,02
Indeks NIKKEI 225							
B&H	6,62	24,33	51,31	4,78	0,27	0,04	0
LS	0,15	24,34	60,91	8,75	0,01	0,00	0
LO	4,48	19,30	39,31	2,74	0,23	0,03	0
Indeks DAX							
B&H	4,34	22,97	54,6	5,43	0,19	0,02	0
LS	-1,47	22,98	60,87	8,10	-0,06	0,00	0
LO	2,26	18,92	43,48	5,40	0,12	0,01	0
Indeks WIG							
B&H	1,74	19,74	57,62	3,13	0,09	0	0
LS	-6,29	19,74	66,18	12,83	-0,32	-0,03	0
LO	-1,55	15,01	43,61	9,20	-0,10	0	0
Indeks KOSPI							
B&H	3,86	19,67	50,30	5,88	0,20	0,02	0
LS	9,20	19,66	39,44	2,48	0,47	0,11	0
LO	6,99	17,16	43,13	1,85	0,41	0,07	0
Indeks BOVESPA							
B&H	4,94	28,47	59,96	9,15	0,17	0,01	0
LS	8,56	28,46	47,96	2,52	0,30	0,05	0
LO	8,11	23,52	52,84	6,03	0,34	0,05	0

Źródło: Opracowanie własne.

Dla indeksów na których strategie przynosiły zyski większe od strategii pasywnej (S&P 500, KOSPI, BOVESPA), nieznacznie lepiej spisywały się strategię LS. Dla pozostałych trzech indeksów lepsze wyniki uzyskano stosując strategię LO. Także w przypadku tego modelu, zauważyć można, że strategia LO cechuje się niższymi

wskaźnikami zmienności, w szczególności odchylenia standardowego (σ) oraz maksymalnej długości straty (MLD), choć w tym wypadku nie zawsze przekłada się to na wyższe wartości wskaźników IR, co może być spowodowane większymi wartościami wskaźnika maksymalnego obsunięcia kapitału (MD).

Wizualizacja wyników w postaci krzywej kapitałowej, w zestawieniu dla wszystkich aktywów, zaprezentowana została na rysunku 5.11.



Rysunek 5.11: Krzywe kapitałowe dla sieci CNN trenowanych na MSE.

Źródło: Opracowanie własne.

W przypadku indeksów S&P 500 oraz BOVESPA, krzywe kapitałowe strategii LS i LO mają zbliżoną trajektorię do strategii pasywnej, przy czym ich poziom jest nieznacznie wyższy, a silniejszy wzrost zaobserwować można w roku 2020 (po załamaniu związanym z początkiem pandemii COVID-19). Dla indeksu KOSPI krzywe kapitałowe strategii LS i LO przewyższają strategię pasywną niemal w całym testowanym okresie, natomiast w przypadku indeksów NIKKEI, DAX i WIG załamanie tych strategii następuje w okolicach lat 2012-2014.

5.4.3 Wyniki dla strategii wykorzystującej sieci LSTM

W kolejnym kroku sporządzone zostały strategie bazujące na prognozach otrzymanych z modelu wykorzystującego sieci LSTM. Oceny strategii uzyskane przy pomocy tego modelu przedstawione zostały w tabeli 5.7, z zachowaniem tych samych oznaczeń poszczególnych strategii oraz mierników.

Tabela 5.7: Prognozowanie punktowe stóp zwrotu: wyniki strategii dla sieci LSTM trenowanych na MSE

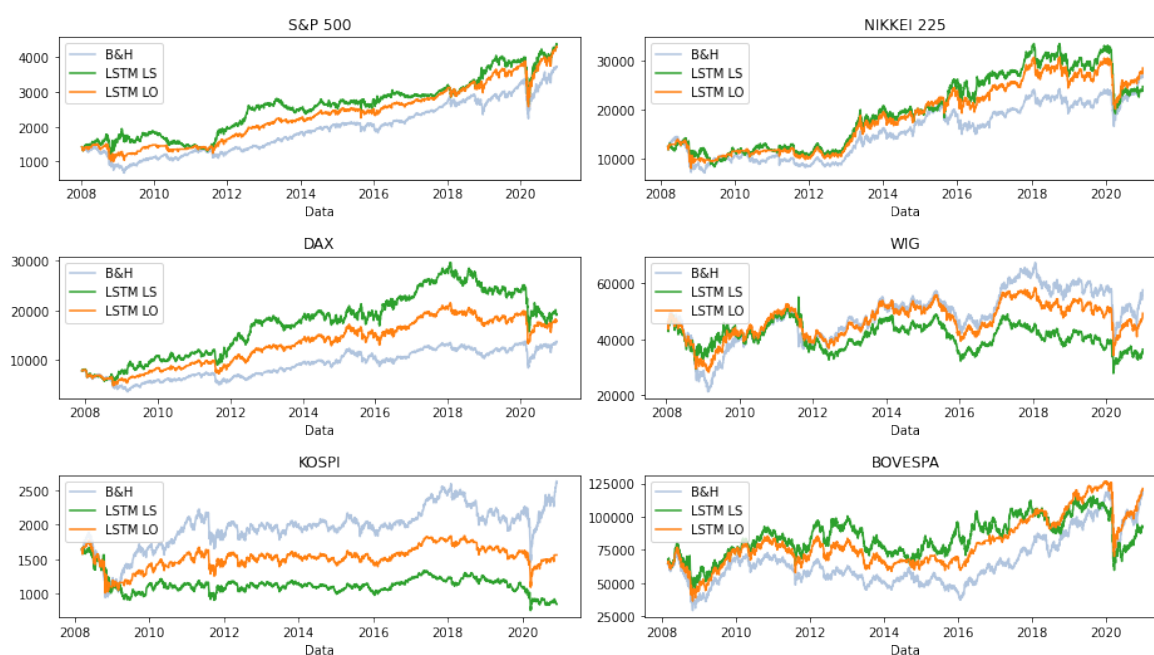
Strategia	aRC	aSD	MD	MLD	IR*	IR**	IR***
Indeks S&P500							
B&H	7,74	21,12	52,58	4,30	0,37	0,05	0
LS	9,07	21,12	34,50	2,75	0,43	0,11	0
LO	8,92	18,70	34,03	3,05	0,48	0,13	0
Indeks NIKKEI 225							
B&H	6,62	24,33	51,31	4,78	0,27	0,04	0
LS	5,68	24,34	42,65	4,28	0,23	0,03	0
LO	6,89	21,28	42,76	4,56	0,32	0,05	0
Indeks DAX							
B&H	4,34	22,97	54,60	5,43	0,19	0,02	0
LS	7,04	22,97	47,11	2,92	0,31	0,05	0
LO	6,47	19,49	39,89	2,93	0,33	0,05	0
Indeks WIG							
B&H	1,74	19,74	57,62	3,13	0,09	0	0
LS	-1,83	19,74	49,59	9,29	-0,09	0	0
LO	0,50	16,67	42,64	2,90	0,03	0	0
Indeks KOSPI							
B&H	3,86	19,67	50,30	5,88	0,20	0,02	0
LS	-5,20	19,67	55,29	12,44	-0,26	-0,02	0
LO	-0,41	17,68	43,12	8,83	-0,02	0,00	0
Indeks BOVESPA							
B&H	4,94	28,47	59,96	9,15	0,17	0,01	0
LS	2,87	28,47	48,37	3,40	0,10	0,01	0
LO	5,07	24,24	52,48	4,56	0,21	0,02	0

Źródło: Opracowanie własne.

Dla sieci LSTM, lepsze wartości wskaźnika skumulowanego zwrotu (aRC), w stosunku do strategii pasywnej, model uzyskał dla czterech indeksów: S&P 500 (strategie LS oraz LO), NIKKEI 225 (strategia LO), DAX (strategie LS oraz LO) a także BOVESPA (strategia LO). W przypadku indeksu KOSPI obie strategie

przyniosły straty (ujemna wartość wskaźnika aRC), podobnie jak strategia LS dla indeksu WIG. Także w przypadku tego modelu, dla większości aktywów zauważyć można, że strategię LO cechują się mniejszym ryzykiem (niższe wartości mierników aSD, MD), oraz, co za tym idzie, większymi wartościami współczynników IR – dla indeksów S&P 500 oraz DAX wskaźniki IR dla strategii LO są wyższe w stosunku do strategii LS, mimo niższego skumulowanego zwrotu.

Wizualizacja wyników uzyskanych przy pomocy modelu LSTM dla wszystkich aktywów przedstawiona została na rysunku 5.12.



Rysunek 5.12: Krzywe kapitałowe dla sieci LSTM trenowanych na MSE.

Źródło: Opracowanie własne.

W przypadku sieci LSTM krzywe kapitałowe strategii LS oraz LO mają bardzo zbliżony przebieg do strategii pasywnej. Pewne różnice zaobserwować można dla indeksu DAX, gdzie krzywe charakteryzują się wyższym poziomem od strategii pasywnej, przy czym nieco lepsze wyniki uzyskuje strategia LS. Dla indeksu KOSPI sytuacja jest odwrotna – krzywe kapitałowe strategii LS i LO mają niższy poziom niż krzywa strategii pasywnej, a poziom krzywej LS jest w tym przypadku niższy.

Podsumowując wyniki dla wszystkich rozważanych modeli, zauważyć należy, że sieci LSTM uzyskały lepsze wyniki niż strategia pasywna dla czterech aktywów, a więc najlepiej spośród analizowanych do tej pory modeli. Strategii bazujące na prognozach uzyskanych przez model wykorzystujący sieć CNN, w stosunku do strategii pasywnej uzyskały lepsze wyniki dla trzech aktywów, natomiast w przypadku sieci MLP jedynie dla indeksu WIG strategia LO uzyskała lepszy wynik niż strategia pasywna.

Porównując wyniki strategii LS i LO, dla sieci MLP strategia LO uzyskała lepsze

wyniki od strategii LS we wszystkich przypadkach. Dla sieci CNN i LSTM wyniki uzyskane przez obie strategie są zbliżone, natomiast strategia LO charakteryzuje się znacznie niższymi wartościami wskaźników aSD, MD oraz MLD, a co za tym idzie, najczęściej także lepszymi wskaźnikami IR. Świadczy to o niższym poziomie zmienności strategii LO, co może być spowodowane mniejszym marginesem błędu, który uzyskiwać może strategia bazująca jedynie na sygnałach kupna.

Sieci LSTM, ze względu na to, że do tej pory przynosiły najlepsze rezultaty, dodatkowo wykorzystane zostały do budowy modelu, w którym sieć uczona jest przy użyciu funkcji straty minimalizującej miarę MADL (zob. rozdział 2 oraz Michańków i in. [2022]). Wyniki dla tego modelu zaprezentowane zostały w tabeli 5.8.

Dla sieci LSTM trenowanych na z wykorzystaniem funkcji straty MADL, lepsze wyniki udało się uzyskać dla indeksów S&P 500, DAX oraz BOVESPA, przy czym wyniki dla WIG i KOSPI były zbliżone do strategii pasywnej, w szczególności dla strategii LO. W przypadku indeksu KOSPI strategia LS przyniosła stratę, uzyskując wartość wskaźnika skumulowanego zwrotu (aRC) na poziomie -6,73 oraz dodatnią wartość miary MADL wynoszącą 0,0199. Zaobserwować można także, że również w tym przypadku wyniki dla strategii LO cechowały się znacznie mniejszym ryzykiem w przypadku wszystkich indeksów, oraz większymi wartościami IR. Wizualizacja wyników dla tego modelu w postaci krzywej kapitałowej przedstawiona została na rysunku 5.13.

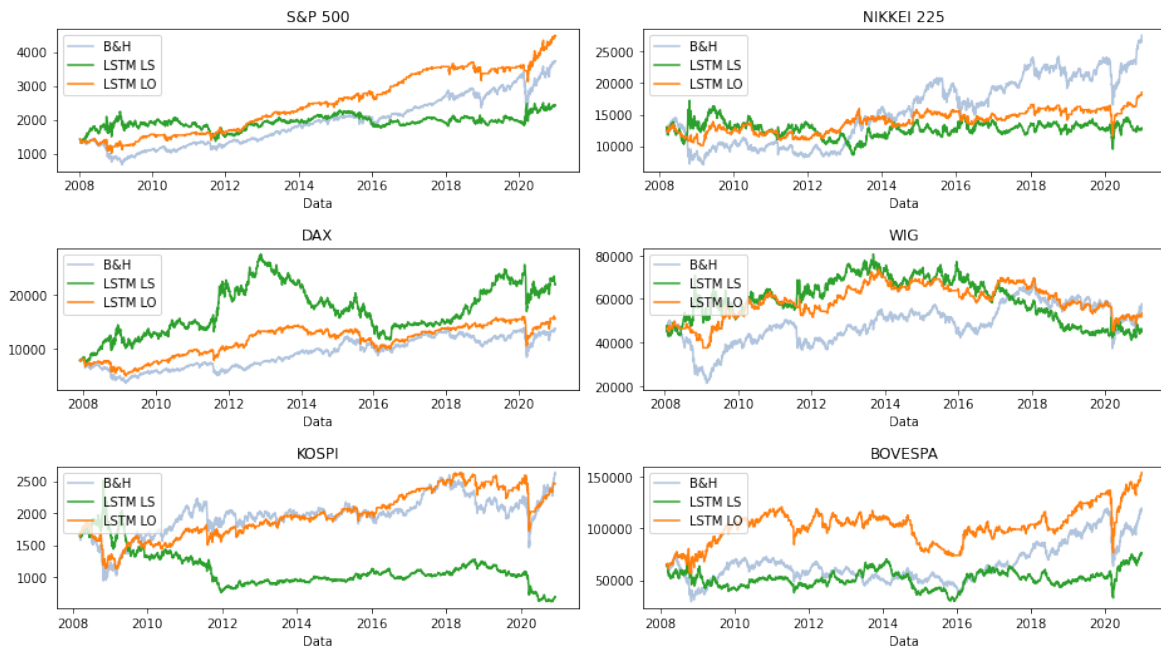
Na podstawie analizy przebiegu krzywych kapitałowych dla strategii LS i LO, zauważyć można, że ich przebieg w większości przypadków nie odbiegał znacznie od strategii pasywnej. Wyjątki stanowią tutaj: strategia LS dla indeksu DAX, gdzie w latach 2012-2015 zaobserwować można znaczny wzrost, a następnie powrót do poprzedniego poziomu, oraz strategia LS dla indeksu KOSPI, dla której odnotowano wyraźny spadek w latach 2010-2012.

Dla modelu sieci LSTM trenowanej z funkcją straty w postaci MADL, strategii LO cechują się znacznie lepszymi wynikami i znacznie mniejszą zmiennością niż strategii LS, które są bardziej niestabilne. Ze względu na niskie wskaźniki ekspozycji na ryzyko i większe wartości IR, można założyć, że uzasadnione byłoby zastosowanie dźwigni dla tego modelu, co przyczyniłoby się do poprawy wyników pod względem zwrotów dla poszczególnych aktywów.

Tabela 5.8: Prognozowanie punktowe stóp zwrotu: wyniki strategii dla sieci LSTM trenowanych na MADL

Strategia	aRC	aSD	MD	MLD	IR*	IR**	IR***	MADL
Indeks S&P500								
B&H	7,74	21,12	52,58	4,30	0,37	0,05	0,00	
LS	4,23	21,13	38,90	5,96	0,20	0,02	0	-0,03534
LO	9,26	15,69	29,69	1,55	0,59	0,18	0,01	-0,0402
Indeks NIKKEI 225								
B&H	6,62	24,33	51,31	4,78	0,27	0,04	0	
LS	0,29	24,34	50,26	11,81	0,01	0,00	0	-0,0152
LO	3,27	15,81	30,00	4,43	0,21	0,02	0	-0,0175
Indeks DAX								
B&H	4,34	22,97	54,6	5,43	0,19	0,02	0	
LS	8,06	22,97	57,47	8,11	0,35	0,05	0	-0,0391
LO	5,31	17,79	38,45	4,59	0,30	0,04	0	-0,0267
Indeks WIG								
B&H	1,74	19,74	57,62	3,13	0,09	0	0	
LS	0,00	19,74	48,91	7,24	0,00	0	0	-0,00084
LO	1,14	12,80	38,85	7,09	0,09	0	0	-0,00739
Indeks KOSPI								
B&H	3,86	19,67	50,30	5,88	0,20	0,02	0	
LS	-6,73	19,67	75,57	11,87	-0,34	-0,03	0	0,0199
LO	3,29	14,55	38,97	4,62	0,23	0,02	0	-0,0170
Indeks BOVESPA								
B&H	4,94	28,47	59,96	9,15	0,17	0,01	0	
LS	1,32	28,47	57,92	6,45	0,05	0,00	0	-0,0202
LO	7,08	22,18	44,08	7,80	0,32	0,05	0	-0,0367

Źródło: Opracowanie własne.



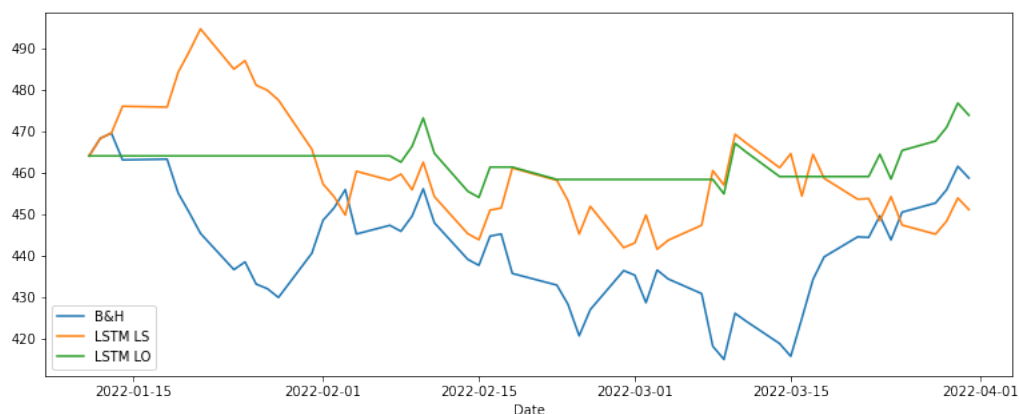
Rysunek 5.13: Krzywe kapitałowe dla sieci LSTM trenowanych na MADL.

Źródło: Opracowanie własne.

Testowanie modelu LSTM w czasie rzeczywistym

Przedstawione do tej pory wyniki dotyczyły testów wstecznych oceniających skuteczność poszczególnych modeli na danych z wybranych indeksów giełdowych. W celu uzupełnienia oceny stosowanych narzędzi przeprowadzona została także ewaluacja na danych pozyskiwanych w czasie rzeczywistym. W tym przypadku ocenie poddany został model LSTM minimalizujący funkcję MADL, stosowany do budowy strategii bazującej tylko na sygnałach kupna. Badanie przeprowadzone zostało na danych pochodzących z funduszu inwestycyjnego *SPDR S&P 500 trust - SPY*, bazującego na indeksie S&P 500. Okres testowy dotyczył w tym wypadku dni pomiędzy 10 stycznia 2021 r. a 31 marca 2022 r. Model uczony był na danych z okresu od 8 stycznia 2014 r. do 9 stycznia 2022 r. (2016 obserwacji). Struktura sieci, wartości hiperparametrów oraz podział sposób wydzielenia danych do zbioru uczącego pozostał bez zmian w stosunku do opisywanego powyżej. Zmianie uległ natomiast sposób ewaluacji modelu. Zamiast oceny modelu na zbiorze testowym, model wykorzystany został do prognozowania wartości oczekiwanej stopy zwrotu na koniec każdego dnia sesyjnego, bazując na danych z ostatnich 10 dni sesyjnych. Wygenerowane na tej podstawie sygnały służyły do otwarcia pozycji długich (jeżeli prognozowana stopa zwrotu była dodatnia) lub zamknięcia pozycji (jeżeli prognozowana stopa zwrotu była ujemna lub zerowa). Model nie był reestymowany w okresie testowym, a ewaluacja modelu nastąpiła po zakończeniu badania. Uzyskane

wyniki przedstawione zostały w tabeli 5.9 oraz na rysunku 5.14.



Rysunek 5.14: Krzywe kapitałowe dla sieci LSTM trenowanych na MADL — pobieranie danych w czasie rzeczywistym.

Źródło: Opracowanie własne.

Tabela 5.9: Prognozowanie punktowe stóp zwrotu: wyniki strategii dla sieci LSTM trenowanych na MADL — pobieranie danych w czasie rzeczywistym.

Strategia	aRC	aSD	MD	MLD	IR*	IR**	IR***	MADL
SPY								
B&H	-5,09 (-1,15)	22,22 (5,02)	11,63 (2,63)	0,21 (0,04)	-0,23 (0,05)	-0,10 (0,02)	-0,02 (0,004)	
LS	-11,97(-2,70)	22,21 (5,02)	10,75 (2,43)	0,19 (0,04)	-0,54 (-0,12)	-0,6(-0,13)	-0,38(-0,08)	0,06701
LO	9,86 (2,23)	11,95 (2,70)	4,05 (0,91)	0,13 (0,02)	0,83 (0,18)	2,01 (0,45)	1,56 (0,35)	-0,00962

Uwaga: Współczynniki podane są w skali rocznej (252 dni), zaś w nawiasach – w skali 57 dni.

Źródło: Opracowanie własne.

Analizując wyniki badań przeprowadzonych na danych pobieranych w czasie rzeczywistym można zauważyć, że najlepsze wyniki uzyskała strategia bazująca tylko na pozycjach długich, która jako jedyna przynosiła w tym wypadku zyski. Charakteryzowała się także znacznie lepszymi wskaźnikami aSD, MD, MLD oraz IR w stosunku do strategii pasywnej. Obserwując krzywą kapitałową przedstawioną na rysunku 5.14 zaobserwować można, że strategia ta przez większość czasu miała pozycję zamkniętą, głównie w okresach charakteryzujących się dużymi spadkami, co zabezpieczało przed dużymi obsunięciami kapitału. Otwarcie pozycji długiej w końcowym okresie, pozwoliło na wzbicie się ponad poziom zerowy, przez co finalnie wynik strategii był dobry.

5.5 Prognozowanie zmienności i ryzyka

W poniższym podrozdziale zaprezentowane zostaną wyniki uzyskane przez modele punktowych prognoz zmienności stóp zwrotu, rozumianej jako warunkowe odchylenie

standardowe. W tym zakresie rozważone zostały modele ekonometryczne klasy ARMA-GARCH, z wybranymi specyfikacjami struktury GARCH (GARCH, EGARCH, GJR-GARCH oraz APARCH) i różnymi typami rozkładu warunkowego (rozkład normalny, t-Studenta oraz skośny rozkład t-Studenta), a także modele hybrydowe, łączące wymienione specyfikacje z rekurencyjnymi sieciami LSTM.

W badaniach dotyczących prognozowania zmienności, jako podstawowe dane wykorzystane zostały procentowe logarytmiczne stopy zwrotu. Na ich podstawie dokonano estymacji poziomów zmienności z wykorzystaniem estymatora GKYZ (zob. punkt 2.3.2) a także estymacji i prognoz w ramach wspomnianych modeli ARMA-GARCH. Prognozy zmienności uzyskane przy pomocy tych modeli posłużyły, wraz z wartościami bezwzględnymi stóp zwrotu, jako dane wejściowe do komponentu sieciowego w modelach hybrydowych. Struktura tych modeli opisana została szerzej w punkcie 4.2.1.

Modele ARMA-GARCH estymowane były w ramach schematu rolowanego o długości okna równej 504 obserwacji (2 lata). Długość szeregu, który następnie przekazany był jako dane wejściowe do sieci ustalona została na 3495 obserwacji (w przybliżeniu 14 lat).⁷ Komponent sieciowy modelu hybrydowego uczony był z wykorzystaniem okna rozszerzanego⁸ (zob. rys. 3.14). Początkowa długość zbioru uczącego wynosiła 1008 obserwacji (4 lata), natomiast długość zbioru testowego wynosiła 504 obserwacje, i o taką długość powiększany był także zbiór uczący w każdej iteracji. Z danych w zbiorze uczącym wydzielane dodatkowo były zbiory walidacyjne, stanowiące 33% danych dostępnych w zbiorze uczącym.

Dla rozważanych danych przeprowadzone zostały analizy wskazujące optymalne wartości opóźnień struktury autoregresyjnej dla modeli ARMA-GARCH. W tym celu wykorzystano kryterium informacyjne Schwarza (BIC). Dla dwóch indeksów: S&P500 oraz BOVESPA optymalne wartości wyniosły (1,0), natomiast dla pozostałych indeksów wartości te wyniosły (0,0).⁹

Struktura i wartości poszczególnych hiperparametrów sieci LSTM w modelu hybrydowych ustalona została w wyniku procesu dostrajania hiperparametrów, którego wyniki zaprezentowane zostały w podrozdziale (5.3). Model uczony był

⁷W przypadku modeli hybrydowych ilość obserwacji była taka sama dla wszystkich badanych aktywów, co umożliwiło bezpośrednie porównanie liczby przekroczeń wartości zagrożonej. Zakres dat natomiast mógł się w tym przypadku nieznacznie różnić, ze względu na różne ilości dni sesyjnych w ciągu roku.

⁸Badania wstępne wykazały w tym zakresie, że zastosowanie okna rozszerzanego przyniosło w tym wypadku lepsze rezultaty niż podejście oparte na schemacie rolowanym.

⁹W dalszej części dla poprawienia czytelności przedrostek ARMA w nazwach modeli będzie najczęściej pomijany.

z wykorzystaniem funkcji straty minimalizującej miernik błędu MSE.

5.5.1 Ocena trafności prognoz punkowych zmienności

W pierwszej kolejności zaprezentowane zostało zestawienie poszczególnych badanych modeli, pod względem zgodności prognoz z poziomem zmienności uzyskanym przy pomocy estymatora GKYZ, z wykorzystaniem miernika oceny prognoz MSE. Wyniki przedstawione zostały w tabeli 5.10. Tabela zawiera wyniki zarówno dla poszczególnych specyfikacji modeli GARCH z różnymi badanymi typami rozkładów warunkowych, jak i modeli hybrydowych wykorzystujących te specyfikacje w połączeniu z sieciami LSTM (zob. rozdział 4).

Tabela 5.10: Oceny trafności prognoz zmienności — wartości MSE.

Indeks/Model	G(N)	G(N)-LSTM	G(STD)	G(STD)-LSTM	G(SSTD)	G(SSTD)-LSTM
S&P	0,3285	0,3281	0,3251	0,3240	0,3353	0,3315
NIKKEI	0,7190	0,6296	0,7156	0,6413	0,7232	0,6445
DAX	1,0697	0,7849	1,0517	0,7771	1,0753	0,7910
WIG	0,5017	0,4807	0,4898	0,4648	0,4919	0,4899
KOSPI	0,7378	0,5530	0,7188	0,5574	0,7388	0,5594
BOVESPA	0,7934	0,6666	0,7988	0,6707	0,8114	0,6784
	E(N)	E(N)-LSTM	E(STD)	E(STD)-LSTM	E(SSTD)	E(SSTD)-LSTM
S&P	0,3576	0,3519	0,3517	0,3449	0,3633	0,3605
NIKKEI	0,7293	0,6421	0,7186	0,6293	0,7181	0,6454
DAX	1,2148	0,8233	1,1960	0,8740	1,1949	0,8505
WIG	0,5110	0,4883	0,5081	0,4868	0,5032	0,4993
KOSPI	0,7896	0,5862	0,7751	0,5751	0,7710	0,5626
BOVESPA	0,8487	0,6504	0,8181	0,6579	0,8024	0,6735
	GJR(N)	GJR(N)-LSTM	GJR(STD)	GJR(STD)-LSTM	GJR(SSTD)	GJR(SSTD)-LSTM
S&P	0,3026	0,2915	0,3012	0,2809	0,3063	0,2946
NIKKEI	0,7043	0,6496	0,6901	0,6344	0,6962	0,6277
DAX	1,0180	0,7593	0,9778	0,7556	0,9763	0,7446
WIG	0,4802	0,4320	0,4727	0,4331	0,4729606	0,4336
KOSPI	0,7034	0,5403	0,6913	0,5533	0,7065	0,5603
BOVESPA	0,7888	0,6736	0,7942	0,6760	0,8232	0,6914
	AP(N)	AP(N)-LSTM	AP(STD)	AP(STD)-LSTM	AP(SSTD)	AP(SSTD)-LSTM
S&P	0,3423	0,3345	0,3242661	0,3180	0,3265	0,3162
NIKKEI	0,7258	0,6386	0,7014102	0,6341	0,7013	0,6347
DAX	1,0802	0,7755	1,03333	0,7596	1,0486	0,7690
WIG	0,4952	0,4778	0,4825994	0,4723	0,5063	0,4659
KOSPI	0,7496	0,5318	0,7233	0,5512	0,7324	0,5664
BOVESPA	0,8060	0,6740	0,8054	0,6695	0,8800	0,7010

Uwaga: W tabeli zastosowane zostały następujące oznaczenia: G – GARCH, E – EGARCH, GJR – GJR-GARCH, AP – APARCH, N – rozkład normalny, STD – rozkład t-Studenta, SSTD - skośny rozkład t-Studenta. Pogrubione zostały najniższe wartości miernika MSE dla poszczególnych indeksów.

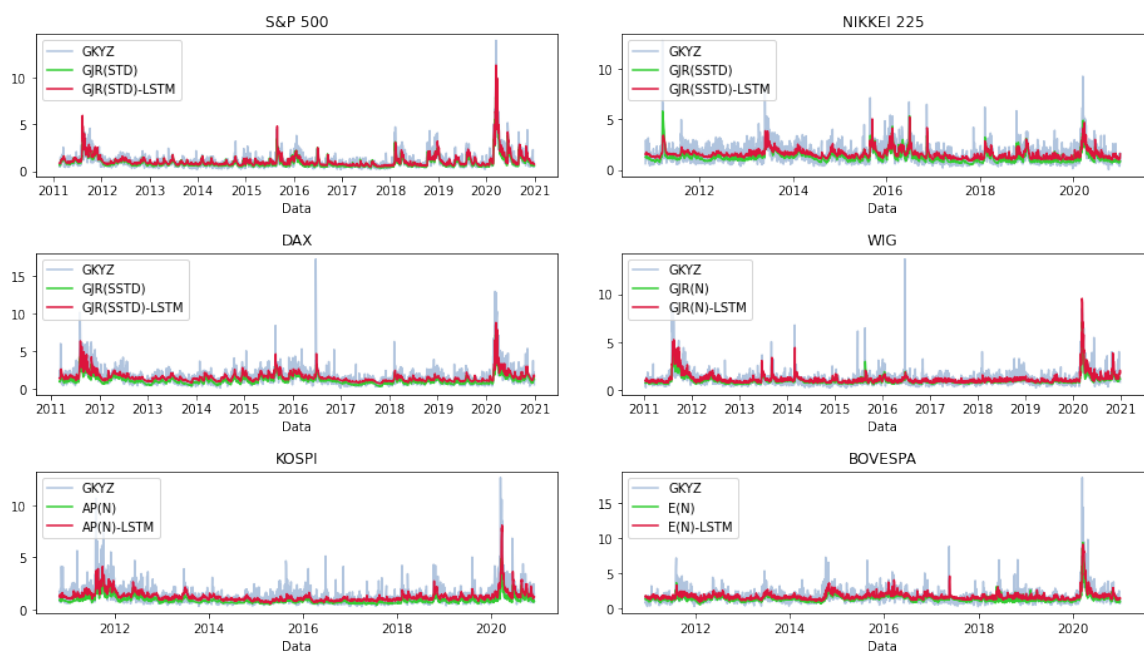
Źródło: Opracowanie własne.

Na podstawie zaprezentowanych wyników zauważyć można, że modele hybrydowe, w porównaniu z „czystymi” modelami klasy GARCH, przyczyniły się do poprawy prognoz zmienności w stosunku do estymatora GKYZ w zasadzie w każdym przypadku. W tym zestawieniu najlepsze wyniki uzyskały modele GJR-GARCH-LSTM, które uzyskały najniższe wartości miernika MSE dla czterech aktywów: S&P 500 (z rozkładem normalnym), NIKKEI 225 (z rozkładem t-Studenta), DAX (ze skośnym rozkładem t-Studenta) oraz WIG (z rozkładem t-Studenta). Dla indeksu KOSPI najlepszy wynik uzyskał model APARCH-LSTM z rozkładem normalnym, a dla BOVESPA — model EGARCH-LSTM, również z rozkładem normalnym. Wyniki te pokrywają się zatem z opublikowanymi do tej pory badaniami (zob. 4.1), które wykazywały, że łączenie modeli klasy GARCH z sieciami neuronowymi przyczynia się do poprawy punktowych prognoz zmienności (przy danej metodzie oceny zmienności oraz przy wybranej funkcji straty).

Porównując przekrojowo wyniki pod kątem wykorzystanych specyfikacji modeli GARCH, najniższe wartości współczynnika MSE uzyskały modele wykorzystujące specyfikację GJR-GARCH, dla zdecydowanej większości badanych typów rozkładów oraz indeksów. Jeśli zaś chodzi o badane typy rozkładów, można zauważyć, że modele wykorzystujące rozkład t-Studenta oraz skośny rozkład t-Studenta wykazują niższe wartości miernika MSE niż modele wykorzystujące rozkład normalny.

Wizualne przedstawienie prognoz zmienności dla wszystkich sześciu aktywów zaprezentowane zostało na rysunku 5.15. Wybrane zostały tutaj modele hybrydowe, które uzyskały najlepsze wyniki dla poszczególnych aktywów, wraz z odpowiednimi specyfikacjami bazowych modeli GARCH o tym samym rozkładzie.

Analiza graficzna pokazuje, że modele hybrydowe (oznaczone czerwonym kolorem) nieznacznie lepiej dopasowywały poziom (średnią) prognoz zmienności do danych uzyskanych z wykorzystaniem estymatora (kolor szary), w stosunku do (nieznacznie zaniżonych) oszacowań zmienności uzyskanych w modelach GARCH (kolor zielony). Największa różnica w poziomach zmienności występuje w przypadku indeksu NIKKEI 225; nieco mniejsza, choć wciąż widoczna, w przypadku indeksów KOSPI oraz DAX. Pozostałe indeksy są pod tym względem bardziej zbliżone. Na wykresach zaobserwować można także, że oszacowania zmienności uzyskane za pomocą estymatora GKYZ odznaczają się większą wahliwością oraz występowaniem gwałtownych skoków zmienności (co wynika z faktu, że zmienność ta szacowania jest dla każdego z dni osobno, bez uwzględnienia persystencji zmienności jak ma to miejsce w przypadku modeli GARCH czy stosowania technik wygładzania, jak w przypadku estymatora zmienności historycznej).



Rysunek 5.15: Prognozy zmienności dla poszczególnych aktywów.

Źródło: Opracowanie własne.

Tabela 5.11: Współczynniki korelacji Pearsona pomiędzy najlepszymi specyfikacjami modeli GARCH i GARCH-LSTM dla poszczególnych aktywów

Index	Specyfikacje	Wsp. korelacji
S&P 500	GJR(STD)	0,9519
	GJR(STD)-LSTM	
NIKKEI 225	GJR(SSTD)	0,9064
	GJR(SSTD)-LSTM	
DAX	GJR(SSTD)	0,9517
	GJR(SSTD)-LSTM	
WIG	GJR(N)	0,9444
	GJR(N)-LSTM	
KOSPI	AP(N)	0,9301
	AP(N)-LSTM	
BOVESPA	E(N)	0,9058
	E(N)-LSTM	

Źródło: Opracowanie własne.

Tabela 5.11 prezentuje współczynniki korelacji pomiędzy oszacowaniami zmienności uzyskanymi dla najlepszych specyfikacji modelowych. Zauważyć w niej

można, że najwyższe wartości współczynnika korelacji występują w przypadku indeksów S&P 500 oraz DAX, zaś najniższe dla indeksów BOVESPA i NIKKEI 225, co odpowiada także różnicą w wartościach oceny prognoz z wykorzystaniem miernika MSE, przedstawionych w tabeli 5.10.

5.5.2 Ocena trafności prognoz ryzyka uzyskanych przy pomocy modeli hybrydowych

Aby lepiej ocenić badane modele, punktowe prognozy zmienności uzyskane za pomocą modeli GARCH oraz modeli hybrydowych zastosowane zostały także do pomiaru ryzyka kapitałowego z wykorzystaniem wartości zagrożonej oraz oczekiwanego niedoboru. Wartość zagrożona szacowana była dla pozycji dłuższej, z wykorzystaniem wzoru 2.83 (zob. punkt 2.6.1). W przypadku modeli hybrydowych, oceny parametrów μ_t (oraz ewentualnie parametrów ν i ξ) pochodziły z wartości oszacowanych przez modele ARMA-GARCH, natomiast wartości parametru σ_t zastąpione zostały przez prognozy zmienności uzyskane z wykorzystaniem podejścia hybrydowego. Tego typu zabieg jest uzasadniony z tego względu, że to właśnie prognozy zmienności mają największy wpływ na oszacowany poziom wartości zagrożonej, natomiast prognozy wartości oczekiwanej oraz parametry rozkładu odgrywają w tym przypadku znacznie mniejszą rolę¹⁰. Wartości oczekiwanego niedoboru obliczane były analogicznie, na podstawie wzoru 2.90 (zob. punkt 2.6.2).

W pierwszej kolejności, w tabeli 5.12 przedstawione zostały wyniki procentowych przekroczeń VaR dla poszczególnych modeli, w zestawieniu dla wszystkich aktywów, dla 5% i 1% poziomu tolerancji.

¹⁰Wyniki bazujące na podejściu, w którym wszystkie te parametry pochodzą z sieciowych modeli prognozowania probabilistycznego przedstawione zostały w kolejnym podrozdziale.

Tabela 5.12: Procentowe udziały przekroczeń oszacowań VaR(0,05)/VaR(0,01) w modelach ARMA-GARCH i hybrydowych ARMA-GARCH-LSTM

Indeks/Model	G(N)	G(N)-LSTM	G(STD)	G(STD)-LSTM	G(SSTD)	G(SSTD)-LSTM
S&P	6,35/2,73	<u>4,34/1,65</u>	7/1,81	5,11*/1,13	6,35/1,49	<u>4,22/0,97*</u>
NKX	<u>5,47*/2,33</u>	<u>3,22/1,05*</u>	6,15/1,81	<u>3,62/0,72*</u>	<u>5,39*/1,49</u>	3,1/0,44
DAX	7/2,61	<u>2,73/0,68*</u>	7,4/1,61	3,14/0,36	6,88/1,49	2,69/0,28
WIG	<u>5,71/1,85</u>	<u>3,94/1,13*</u>	<u>6,03/1,33*</u>	<u>4,3*/0,84*</u>	<u>5,67/1,17*</u>	<u>3,62/0,76*</u>
KOSPI	<u>6,71/2,53</u>	<u>2,57/0,64</u>	7,24/1,89	3,18/0,44	6,27/1,45*	2,37/0,2
BVP	<u>5,35*/1,45*</u>	<u>2,73/0,72*</u>	<u>5,83*/1,17*</u>	2,77/0,52	<u>5,51*/1,21*</u>	<u>2,69/0,64*</u>
	E(N)	E(N)-LSTM	E(STD)	E(STD)-LSTM	E(SSTD)	E(SSTD)-LSTM
S&P	6,55/2,57	<u>4,42/1,45</u>	7,2/1,85	<u>4,74/1,33</u>	6,23/0,88*	3,94/0,88
NKX	<u>5,59*/2,13</u>	<u>2,85/0,8*</u>	5,95*/1,73	3,34/0,52	<u>5,43*/1,13*</u>	3,06/0,4
DAX	<u>6,39/2,29</u>	<u>2,69/0,72*</u>	6,88/2,05	3,1/0,6*	6,03/1,53	2,41/0,36
WIG	6,03/1,97	<u>3,34/1,13*</u>	6,35/1,65	3,98/0,97*	<u>5,79*/1,25*</u>	<u>3,46/0,64*</u>
KOSPI	6,88/2,73	<u>2,65/0,68*</u>	7,16/2,05	2,98/0,6*	<u>6,35/1,17*</u>	2,13/0,32
BVP	<u>5,23*/1,69</u>	<u>2,41/0,56*</u>	4,95*/0,76*	3,02/0,44	<u>4,78*/0,64*</u>	3,1/0,48
	GJR(N)	GJR(N)-LSTM	GJR(STD)	GJR(STD)-LSTM	GJR(SSTD)	GJR(SSTD)-LSTM
S&P	5,99*/2,37	3,94/1,21	6,88/1,49*	<u>4,5*/0,97</u>	<u>5,63*/1,05*</u>	3,22/0,48
NKX	<u>5,63*/2,33</u>	<u>3,06/0,84*</u>	<u>5,83*/1,73</u>	<u>3,42/0,68*</u>	<u>5,39*/1,49*</u>	2,98/0,44
DAX	6,51/2,25	<u>2,49/0,76*</u>	6,88/1,65	2,69/0,48	<u>6,07/1,33*</u>	2,41/0,4
WIG	<u>5,47*/1,81</u>	<u>3,58/1,09*</u>	5,91/1,29*	3,94/0,84*	5,39*/1,21*	<u>3,5/0,64*</u>
KOSPI	<u>6,63/2,29</u>	<u>2,45/0,76*</u>	7,04/1,85	2,69/0,52	<u>6,23/1,21*</u>	2,45/0,4
BVP	<u>5,15*/1,45*</u>	2,65/0,56*	<u>5,43*/1,01*</u>	2,69/0,52	<u>5,71*/1,13*</u>	2,77/0,4
	AP(N)	AP(N)-LSTM	AP(STD)	AP(STD)-LSTM	AP(SSTD)	AP(SSTD)-LSTM
S&P	6,67/2,85	<u>4,46*/1,65</u>	7,08/1,65	<u>4,62/1,17</u>	5,95/1,09*	3,98/0,72*
NKX	<u>5,43*/2,01</u>	<u>2,98/0,92*</u>	<u>5,63*/1,57</u>	3,26/0,56	4,91*/1,09*	3,02/0,44
DAX	6,51/2,37	<u>2,53/0,84*</u>	7,04/1,89	2,69/0,52	6,07/1,57	2,29/0,4
WIG VaR	<u>5,71*/1,93</u>	<u>3,82/1,17*</u>	<u>5,83/1,33*</u>	<u>3,86/0,92*</u>	<u>5,47*/1,25*</u>	<u>3,42/0,8*</u>
KOSPI VaR	6,92/2,65	2,9/0,8*	7,2/1,85	2,98/0,52	6,15/1,25*	2,61/0,4
BVP VaR	<u>5,79*/1,65</u>	<u>2,69/0,64*</u>	<u>5,75*/1,25*</u>	3,18/0,48	<u>5,55*/1,33*</u>	3,26/0,6*

Uwaga: Pogrubione zostały wyniki najbardziej zbliżone do przyjętego poziomu tolerancji, podkreślone te, które uzyskały także poprawny wynik testu Kupca, zaś gwiazdką oznaczone zostały te, które uzyskały poprawny wynik testu Christoffersena. Oczekiwana liczba przekroczeń wynosi

124/24, ilość prognoz 2487.

Źródło: Opracowanie własne.

Analizując wyniki przedstawione w powyższej tabeli, pod względem procentowego udziału przekroczeń VaR najlepsze wyniki uzyskał model APARCH-LSTM z rozkładem normalnym, który dla 1% poziomu tolerancji najlepsze rezultaty uzyskał dla trzech indeksów: NIKKEI 225 (0,92%), DAX (0,84%) oraz KOSPI (0,8%). Z kolei model APARCH ze skośnym rozkładem t-Studenta najlepiej spisywał się na 5% poziomie tolerancji dla indeksów NIKKEI 225 (4,91%) oraz KOSPI (6,15%). Ogólnie zauważyć można, że modele wykorzystujące specyfikację APARCH spisywały się w tym zestawieniu najlepiej. Wyniki ilości procentowych przekroczeń VaR sugerują

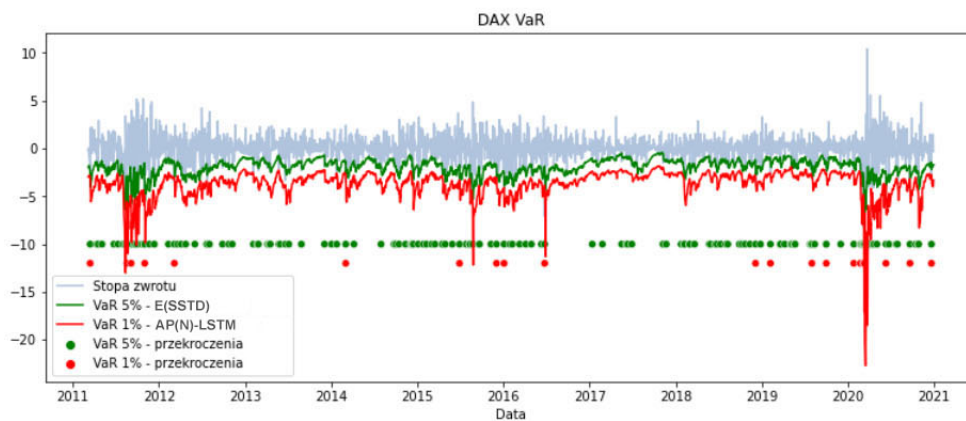
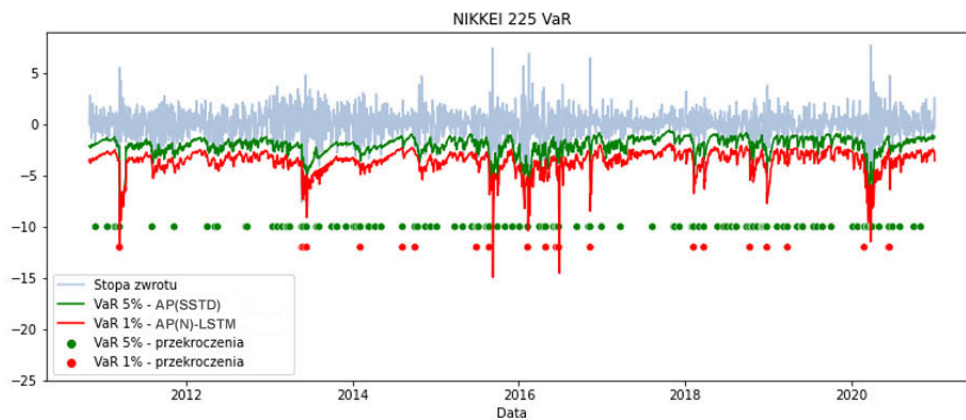
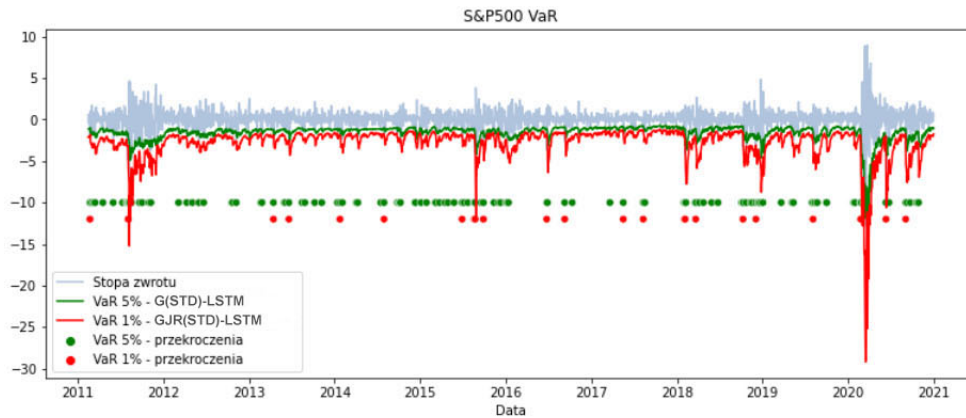
dotatkowo, że nieznacznie lepiej spisują się modele wykorzystujące rozkłady t-Studenta i skośny t-Studenta (w pięciu przypadkach dla każdego rozkładu) niż modele wykorzystujące rozkład normalny (w trzech przypadkach). Dodatkowa tabela, A.1, przedstawiająca liczbę przekroczeń poziomu VaR zamieszczona została w Aneksie.

Wstępne porównanie wyników uzyskanych przez modele hybrydowe z modelami klasy GARCH wskazuje pod względem wartości najbliższych przyjętym poziomom tolerancji (wartości pogrubione), pokazuje, że modele hybrydowe lepiej prognozują wartość zagrożoną dla 1% poziomu tolerancji, zaś „czyste” modele GARCH — dla poziomu 5%. Szczegółowa analiza porównawcza wybranych specyfikacji modelowych przedstawiona zostanie w rozdziale 6.

Przekroczenia VaR dla poszczególnych modeli przeanalizowane zostały dodatkowo z wykorzystaniem testu Kupca, badającego poprawność ilości przekroczeń poziomu VaR w stosunku do przyjętego poziomu tolerancji, oraz testu Christoffersena, sprawdzającego dodatkowo niezależność przekroczeń w czasie (zob. punkt 2.6.3). Prawidłowe wyniki dla tych testów zaznaczone zostały w tabeli 5.12 odpowiednio podkreśleniem oraz gwiazdką. Analizując wyniki w tym kontekście przede wszystkim zauważyć zauważyć można, że dla 1% poziomu tolerancji najlepsze wyniki przekrojowo uzyskały modele hybrydowe wykorzystujące specyfikacje GARCH z rozkładem normalnym. Bardzo dobre wyniki, dla obu poziomów tolerancji, uzyskały także „czyste” modele klasy GARCH ze skośnym rozkładem t-Studenta.

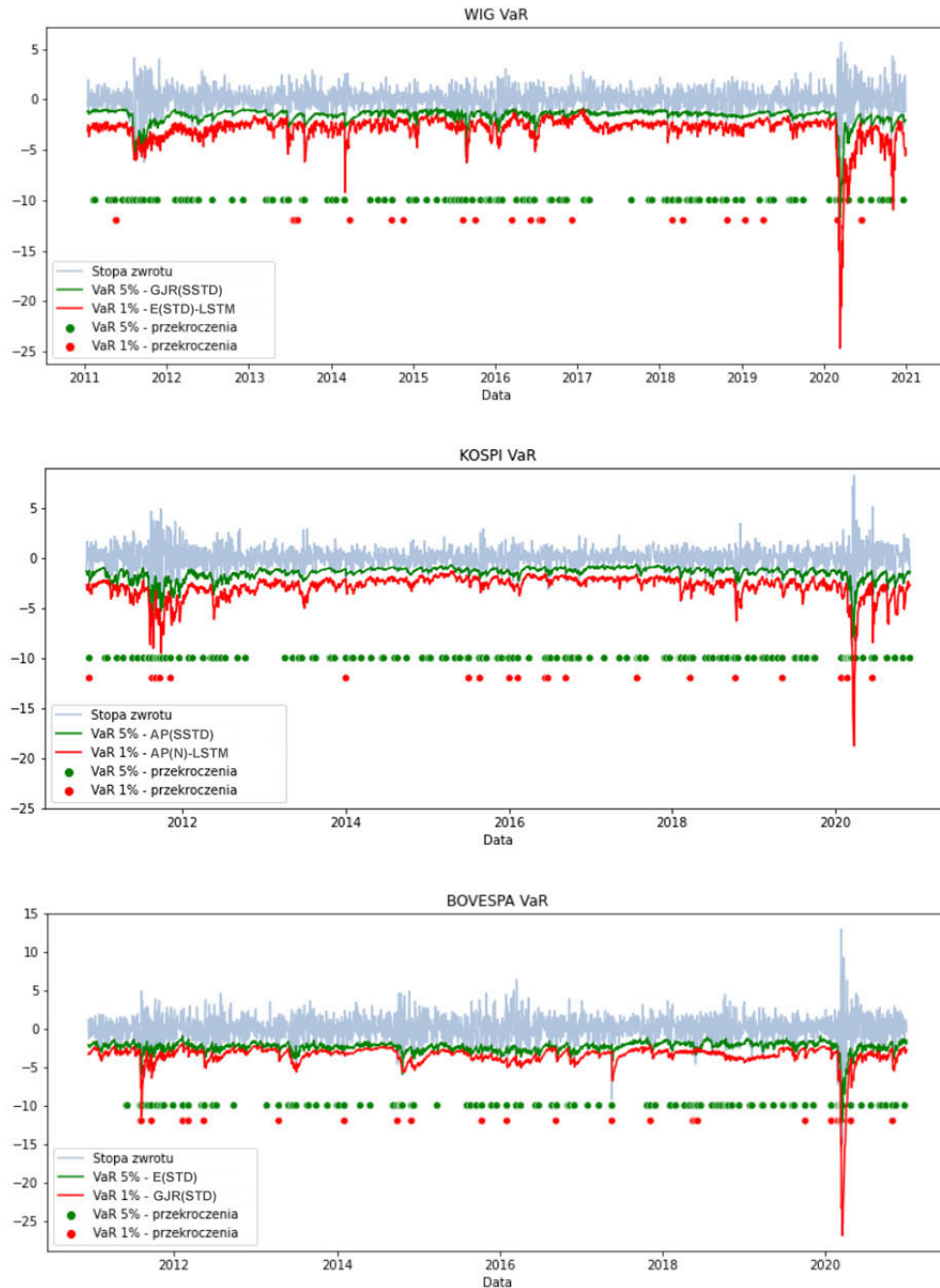
Wyniki oszacowań ryzyka w kolejnym kroku poddane zostały analizie wykresowej. Na rysunkach 5.16 i 5.17 przedstawione zostały oszacowania VaR oraz ich przekroczenia dla modeli najlepiej spisujących się pod względem liczby przekroczeń dla danego indeksu, zarówno dla poziomu tolerancji 1%, jak i 5%.

Na prezentowanych wykresach, zielonym kolorem zaznaczony jest prognozowany poziom VaR dla 5% poziomu tolerancji, natomiast kolorem czerwonym — dla 1% poziomu tolerancji. Punktowo zaznaczone zostały natomiast przekroczenia VaR. Dla wszystkich badanych aktywów zauważyć można pojawiające się na wykresach zgrupowania przekroczeń, widoczne dla obu poziomów tolerancji. Zgrupowania te widoczne są zarówno w okresach podwyższonej, jak i niskiej zmienności. Takie wyniki sugerować by mogły, że przekroczenia nie są niezależne w czasie.



Rysunek 5.16: Prognozy wartości zagrożonej dla indeksów S&P 500, NIKKEI oraz DAX, uzyskane za pomocą wybranych modeli klasy GARCH oraz modeli hybrydowych GARCH-LSTM.

Źródło: Opracowanie własne.



Rysunek 5.17: Prognozy wartości zagrożonej dla indeksów WIG, KOSPI oraz BOVESPA, uzyskane za pomocą wybranych modeli klasy GARCH oraz modeli hybrydowych GARCH-LSTM.

Źródło: Opracowanie własne.

Szczegółowe wyniki dla wszystkich analizowanych modeli przedstawione zostały w tabelach A.2 - A.25, zamieszczonych w Aneksie. Tabele te zawierają informacje dotyczące ocen prognoz zmienności za pomocą miar MSE, MAE, HMSE i R^2 , wyniki testu Diebolda-Mariano na istotność różnic trafności prognoz, ilości przekroczeń VaR, wyniki testów Kupca oraz Christoffersena dla obu przyjętych poziomów tolerancji, oraz wyniki testu McNeila i Fraya (zob. punkt 2.6.3) dotyczące poprawności prognoz

oczekiwanego niedoboru. Ogólnie, zaprezentowane we wspomnianych tabelach wyniki charakteryzują się dużym rozrzutem. Z tego względu, szczegółowej analizie porównawczej przedstawionej w punkcie 6.1.2, poddane zostały wyniki uzyskane przez najlepsze specyfikacje modelowe.

5.6 Prognozowanie rozkładów prawdopodobieństw

W poniższym podrozdziale zaprezentowane zostaną wyniki uzyskane z wykorzystaniem sieciowych modeli prognoz probabilistycznych, opisanych w punkcie 4.2.2. Analizie poddane zostały modele bazujące na jednowymiarowych sieciach CNN oraz sieciach rekurencyjnych LSTM. Struktura i hiperparametry zastosowanych sieci przedstawione zostały dokładniej w podrozdziale 5.3. Modele wykorzystane zostały do prognozowania parametrów wybranych rozkładów warunkowych: rozkładu normalnego, t-Studenta oraz skośnego rozkładu t-Studenta. W pierwszej kolejności zaprezentowane zostaną oceny wykorzystujące kryteria oceny prognoz probabilistycznych: LPS, CRPS oraz PIT (zob. punkt 2.5.4). W dalszej części modele te ocenione zostały w kontekście szacowania ryzyka kapitałowego za pomocą wartości zagrożonej.

Szczegółowe wyniki, dla wszystkich badanych modeli i instrumentów finansowych, dodatkowo przedstawione zostały w Aneksie, w tabelach A.27 - A.38.

5.6.1 Ocena trafności prognoz probabilistycznych

W tabeli 5.13 przedstawiono uzyskane wartości mierników LSP oraz CRPS, a także wartości p-value testu Andersona-Darlinga jednostajności rozkładu PIT (zob. punkt 2.5.4), którego hipoteza zerowa zakłada, że rozkład skalibrowany jest prawidłowo.

Na podstawie przedstawionych w tabeli wyników, można zauważyć, że dla większości aktywów najlepsze wyniki uzyskały sieci LSTM dla skośnego rozkładu t-Studenta. W przypadku miernika LPS, dla indeksów DAX, WIG oraz BOVESPA lepsze wyniki uzyskały sieci LSTM z rozkładem t-Studenta, natomiast pod względem miernika CRPS sieci LSTM z rozkładem normalnym w przypadku indeksu BOVESPA.

Tabela 5.13: Oceny prognoz uzyskanych za pomocą sieciowych modeli prognozowania probabilistycznego

Kryterium/Model	CNN-N	CNN-STD	CNN-SSTD	LSTM-N	LSTM-STD	LSTM-SSTD
Indeks S&P 500						
LPS	1,2820	1,2416	1,2220	1,2632	1,2104	1,1933
CRPS	0,5229	0,5248	0,5197	0,5146	0,5137	0,5094
PIT p-value	2,41e-07	1,55e-05	0,1144	2,41e-07	0,0057	0,0309
Indeks NIKKEI						
LPS	1,6136	1,5990	1,5865	1,6124	1,5870	1,5854
CRPS	0,6914	0,6995	0,6963	0,6892	0,6894	0,6874
PIT p-value	2,41e-07	4,41e-05	0,0476	2,41e-07	2,41e-07	2,41e-07
Indeks DAX						
LPS	1,5800	1,5598	1,5708	1,5540	1,5178	1,5229
CRPS	0,6804	0,6892	0,6901	0,6651	0,6663	0,6645
PIT p-value	2,41e-07	2,41e-07	2,41e-07	2,41e-07	2,41e-07	0,3536
Indeks WIG						
LPS	1,4520	1,3925	1,4055	1,4265	1,3400	1,3475
CRPS	0,5616	0,5703	0,5716	0,5605	0,5553	0,5553
PIT p-value	2,41e-07	2,41e-07	2,41e-07	2,41e-07	0,0023	0,0087
Indeks KOSPI						
LPS	1,3349	1,3147	1,3172	1,3240	1,2961	1,2847
CRPS	0,5285	0,5297	0,5302	0,5246	0,5201	0,5165
PIT p-value	2,41e-07	2,41e-07	2,41e-07	2,41e-07	4,70e-07	5,08e-06
Indeks BOVESPA						
LPS	1,7740	1,7829	1,7875	1,7776	1,7751	1,7851
CRPS	0,8268	0,8369	0,8349	0,8222	0,8269	0,8335
PIT p-value	2,41e-07	0,1866	0,0069	2,41e-07	0,0038	0,0056

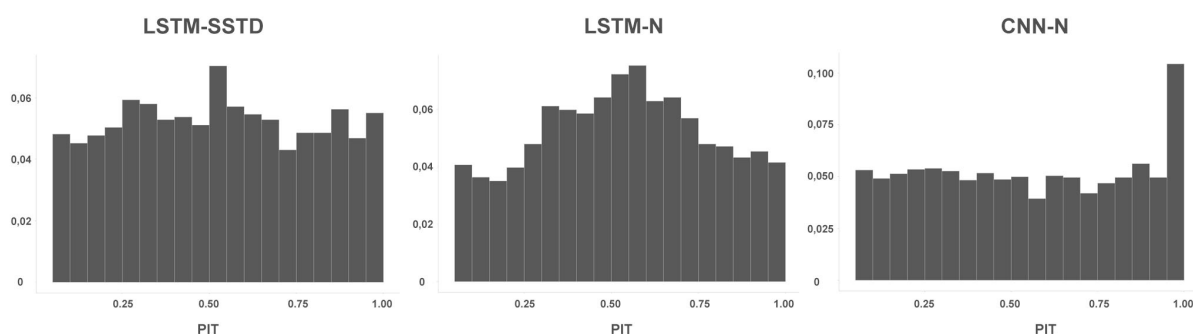
Uwaga: W tabeli pogrubione zostały najniższe wartości kryteriów LPS oraz CRPS oraz najwyższe wartości p-value dla testu Andersona-Darlinga.

Źródło: Opracowanie własne.

Analizując wyniki poprawności kalibracji rozkładu, zauważyć można, że w zdecydowanej większości przypadków wartości p-value były niskie, co wskazuje na niepoprawną kalibrację. W przypadku indeksu S&P 500, jedynie sieci CNN ze skośnym rozkładem t-Studenta uzyskały poprawny wynik testu. Model ten uzyskał najlepszy wynik pod tym względem także w przypadku indeksu NIKKEI, wartość p-value była w tym przypadku nieznacznie niższa niż 0,1. Dla indeksu DAX

poprawny wartość p-value wyższą niż 0,1 uzyskał model bazujący na sieciach LSTM ze skośnym rozkładem t-Studenta. Model ten uzyskał też najwyższą wartość p-value w przypadku indeksów WIG oraz KOSPI, w dalszym ciągu wartość ta jest niższa niż 0,01, co sugeruje niedostateczną kalibrację rozkładów prognozy. Dla indeksu BOVESPA najlepszy i zarazem poprawny wynik testu uzyskały sieci CNN z rozkładem t-Studenta.

Na rysunku 5.18 przedstawione zostały rozkłady wartości PIT, dla trzech wybranych specyfikacji modelowych: LSTM-SSTD, LSTM-N oraz CNN-N, dotyczące indeksu DAX.



Rysunek 5.18: Wykres rozkładów wartości PIT, indeks DAX.

Źródło: Opracowanie własne.

Obserwując wykresy przedstawione na powyższym rysunku można zauważyć, że model LSTM-SSTD, który uzyskał najlepszy wynik testu Andersona-Darlinga, charakteryzuje się najlepszym poziomem dopasowania — wykres wartości PIT jest najbardziej zbliżony do rozkładu jednostajnego. W przypadku modelu LSTM-N zaobserwować można nadmierne rozproszenie wartości prognozowanych — relatywnie mało obserwacji, w stosunku do tendencji centralnej, realizuje się w ogonach rozkładów predykcyjnych. Natomiast dla modelu CNN-N zaobserwować można, że zbyt duża ilość obserwacji realizuje się w górnych kwantylach rozkładów, co może świadczyć o tym, że rozkłady predykcyjne były w tym przypadku systematycznie zaniżone.

5.6.2 Ocena trafności prognoz ryzyka uzyskanych przy pomocy modeli prognozowania probabilistycznego

Wyniki uzyskane przez sieciowe modele prognozowania probabilistycznego poddane zostały analizie także w kontekście szacowania ryzyka kapitałowego. W pierwszej kolejności przedstawione zostały wyniki procentowego udziału przekroczeń VaR dla wszystkich badanych indeksów giełdowych. Zestawienie wyników

przedstawiono w tabeli 5.14.

Tabela 5.14: Procentowe udziały przekroczeń oszacowań VaR(0,05)/VaR(0,01) w modelach prognozowania probabilistycznego.

Indeks/Model	CNN-N	CNN-STD	CNN-SSTD
S&P	<u>4,50</u> /1,56*	<u>4,58</u> /0,92*	<u>5,30</u> /0,88*
NKX	<u>4,30</u> */1,48	4,02*/ <u>0,92</u> *	3,53/0,40
DAX	6,35/2,53	5,42* / 1,12*	<u>4,14*</u> /0,80*
WIG	<u>5,07</u> /1,97	<u>4,58*</u> / 1,12	<u>3,69</u> /0,80*
KOSPI	<u>5,70</u> /2,09	3,86/ <u>0,56</u> *	2,29/0,32
BVP	<u>5,38*</u> / 1,20*	3,98/ <u>0,72</u> *	2,61/0,28
	LSTM-N	LSTM-STD	LSTM-SSTD
S&P	4,86* /1,97	<u>5,34</u> / 1,01*	<u>4,50</u> /0,84*
NKX	<u>4,46*</u> /1,88	4,58* / 1,01*	3,81/0,64
DAX	<u>5,87</u> /1,97	6,59/ <u>1,16*</u>	<u>5,66</u> / 0,88*
WIG	<u>4,82</u> /1,84	4,95 /1,20	<u>4,26</u> /0,76*
KOSPI	5,42* /1,93	<u>4,46*</u> / 0,84*	3,37/0,44
BVP	5,18* /1,24*	3,65/ <u>0,68*</u>	2,41/0,32

Uwaga: Pogrubione zostały wyniki najbardziej zbliżone do przyjętego poziomu tolerancji, podkreślone te, które uzyskały także poprawny wynik testu Kupca, zaś gwiazdką oznaczone zostały te, które uzyskały poprawny wynik testu Christoffersena. Oczekiwana liczba przekroczeń wynosi 124/24, ilość prognoz 2487.

Źródło: Opracowanie własne.

Analizując przedstawione w powyższej tabeli wyniki, można stwierdzić, że dla indeksu S&P 500 najlepsze wyniki pod względem ilości przekroczeń (wartości pogrubione) uzyskały sieci LSTM, przy czym dla 5% poziomu tolerancji był to model z rozkładem normalnym, natomiast dla poziomu 1% — model z rozkładem t-Studenta. W przypadku indeksu NIKKEI, dla obu poziomów tolerancji najlepsze wyniki uzyskały sieci LSTM z rozkładem t-Studenta. Z kolei w przypadku indeksu DAX dla 5% poziomu tolerancji najlepszy wynik uzyskały sieci CNN z rozkładem t-Studenta, natomiast dla poziomu 1% były to sieci LSTM ze skośnym rozkładem T-studenta oraz, ponownie, sieci CNN z rozkładem t-Studenta. Dla indeksu WIG, dla 5% poziomu tolerancji najlepiej spisywał się model wykorzystujące sieci LSTM z rozkładem t-Studenta, zaś dla 1% poziomu tolerancji sieci CNN z rozkładem t-Studenta. Dla indeksów KOSPI oraz BOVESPA, dla 5% poziomu tolerancji najlepsze wyniki uzyskał model LSTM z rozkładem normalnym, natomiast dla

poziomu 1% odpowiednio model LSTM z rozkładem t-Studenta oraz CNN z rozkładem normalnym.

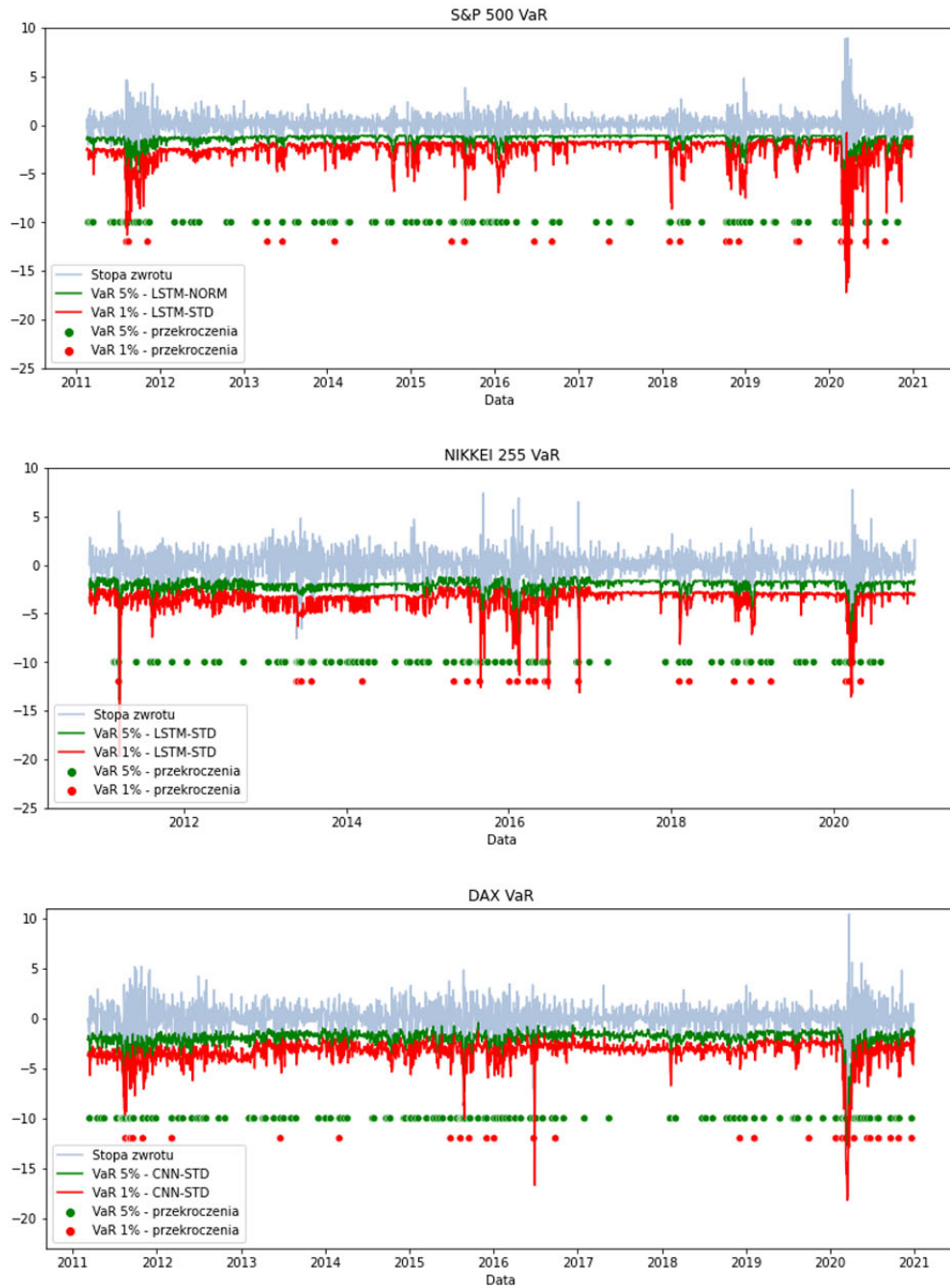
Podsumowując wyniki dotyczące procentowych udziałów przekroczeń VaR można wskazać, że sieci LSTM przynosiły nieco lepsze rezultaty, uzyskując najlepsze wyniki w dziewięciu na dwanaście przypadków, natomiast sieci CNN — w czterech. W przypadku porównania badanych rozkładów prawdopodobieństwa można stwierdzić, że modele oparte na rozkładzie t-Studenta uzyskały najlepsze rezultaty w ośmiu przypadkach, rozkładzie normalnym w czterech przypadkach, zaś oparte na skośnym rozkładzie t-Studenta w jednym przypadku.

Odnosząc się do rezultatów testów Kupca i Christoffersena (dla których poprawne wyniki oznaczone zostały odpowiednio podkreśleniem oraz gwiazdką), zauważyć można, że sieciowe modele prognozowania probabilistycznego ogólnie przynoszą dobre rezultaty. Uwagę zwracają przede wszystkim dobre rezultaty modelu LSTM z rozkładem normalnym dla 5% poziomu tolerancji oraz modeli LSTM i CNN z rozkładem t-Studenta dla 1% poziomu tolerancji. Ogólnie najlepsze wyniki pod kątem liczby przekroczeń oraz poprawnych wyników testów wstecznych uzyskał dla obu poziomów tolerancji uzyskał model LSTM z rozkładem t-Studenta.

W kolejnej części wyniki uzyskane przez najlepsze modele poddane zostały analizie graficznej.¹¹ Rysunki 5.19 i 5.20 przedstawiają wykresy prognozowanych poziomów wartości zagrożonej dla poszczególnych indeksów (oznaczone linią ciągłą) oraz momenty przekroczeń tych poziomów, zaznaczone punktowo.

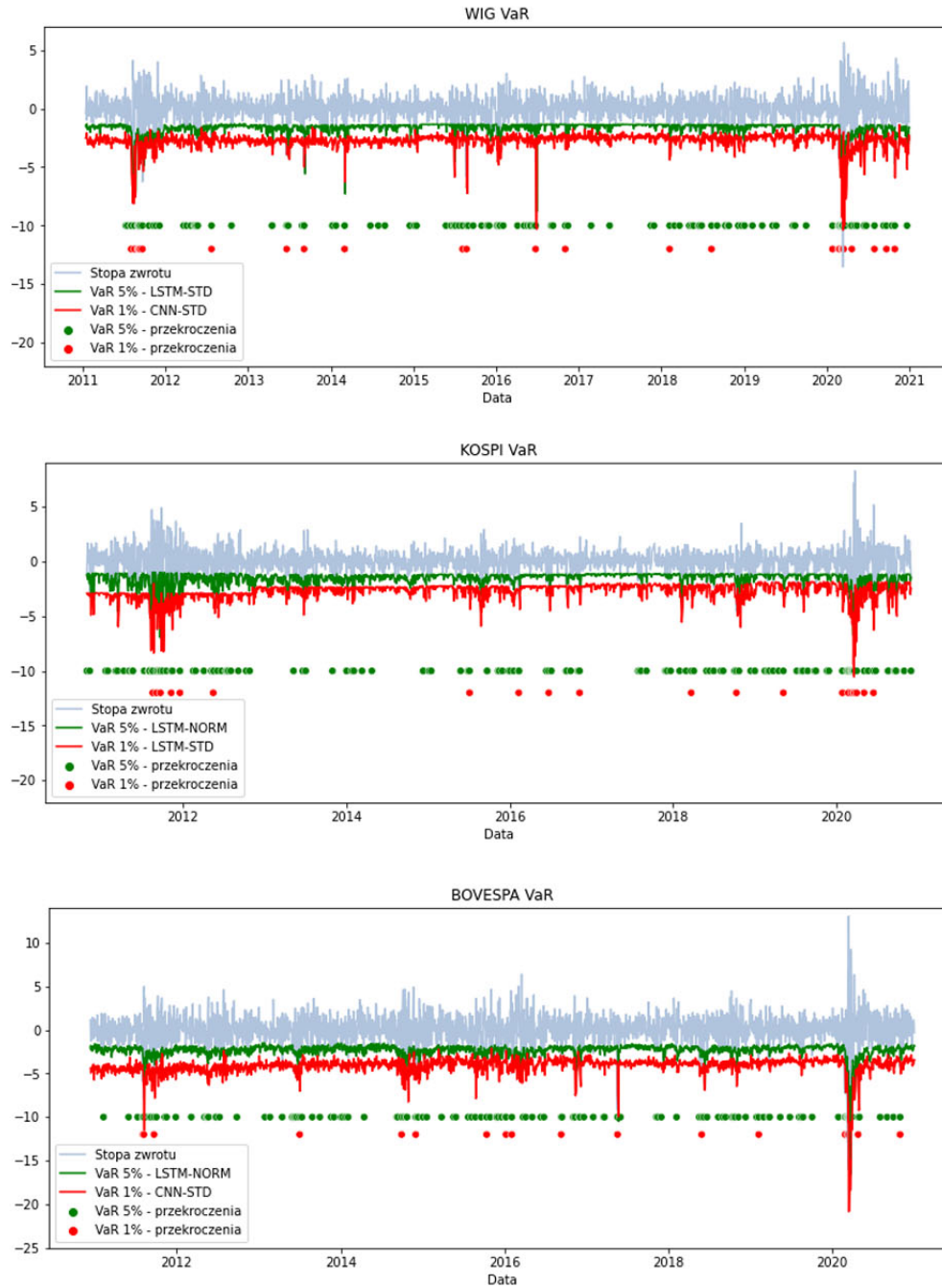
Analogicznie do wykresów przedstawiających oszacowania VaR dla modeli hybrydowych, zielonym kolorem oznaczone zostały prognozy dla przyjętego 5% poziomu tolerancji, a kolorem czerwonym dla poziomu 1%. Także w tym wypadku zauważyć można, że przekroczenia prognozowanego poziomu VaR są zgrupowane i mają miejsce najczęściej w okresach wysokiej zmienności, co — podobnie, jak w przypadku modeli hybrydowych — świadczyć mogło by o tym że przekroczenia te nie są niezależne. Dokładne wyniki testu Kupca i testu Christoffersena dla VaR oraz testu McNeila i Freya dla ES przedstawione zostały w tabelach A.27 - A.38 zamieszczonych w Aneksie. Porównanie wyników uzyskanych przez modele prognozowania probabilistycznego z modelami hybrydowymi m.in. pod względem procentowego udziału przekroczeń VaR przedstawione zostało w punkcie 6.1.3 gdzie zestawiono najlepsze specyfikacje modelowe.

¹¹W przypadku indeksu WIG, sieci CNN z rozkładem t-Studenta i LSTM ze skośnym rozkładem t-Studenta uzyskały bardzo zbliżone wyniki. W analizie graficznej przedstawiono wyniki tylko dla sieci CNN, ponieważ model ten uzyskał lepsze wyniki testów wstecznych.



Rysunek 5.19: Prognozy wartości zagrożonej z wykorzystaniem modeli prognozowania probabilistycznego dla indeksów S&P 500, NIKKEI oraz DAX.

Źródło: Opracowanie własne.



Rysunek 5.20: Prognozy wartości zagrożonej z wykorzystaniem modeli prognozowania probabilistycznego dla indeksów WIG, KOSPI oraz BOVESPA.

Źródło: Opracowanie własne.

Rozdział 6

Szczegółowa analiza porównawcza wyników empirycznych uzyskanych za pomocą wybranych modeli

6.1 Porównanie uzyskanych wyników

W poniższym podrozdziale dokonana została analiza porównawcza wyników uzyskanych z wykorzystaniem wybranych badanych specyfikacji modelowych — tych, które uzyskały najlepsze wyniki dla badanych aktywów w poszczególnych scenariuszach badawczych. Porównanie to ma na celu przede wszystkim wskazanie obszarów, w których poszczególne specyfikacje modelowe sprawdzają się najlepiej.

6.1.1 Modele punktowych prognoz stóp zwrotu

W pierwszym punkcie porównane zostaną wybrane modele wykorzystane do uzyskania punktowych prognoz stóp zwrotu, które uzyskały najlepsze wyniki w kontekście wybranych miar błędu (MSE i MADL) oraz ocen strategii inwestycyjnych.

W podrozdziale 5.4 omówione zostały wyniki ewaluacji modeli opartych na sieciach neuronowych za pomocą poszczególnych metryk błędu (zob. tabela 5.4). Pod względem procentowej trafności prognoz oraz miernika MSE (który wykorzystywany był także jako funkcja straty w procesie uczenia sieci), najlepsze wyniki uzyskały sieci LSTM, uzyskując najniższe wartości MSE dla czterech z sześciu badanych indeksów, i najwyższe wartości procentowej trafności predykcji dla trzech indeksów. Z kolei sieci CNN lepsze wyniki uzyskały biorąc pod uwagę wartości błędu mierzonego przez funkcję MADL. Najslabiej w tym zestawieniu wypadły sieci MLP, które tylko nieco

lepszy wynik od pozostałych sieci pod względem funkcji MADL i trafności predykcji uzyskały dla indeksu WIG.

W kolejnej części modele prognoz punktowych stóp zwrotu porównane będą pod względem poszczególnych wskaźników oceny strategii inwestycyjnych, przy czym do zestawienia wybrane zostaną jedynie najlepsze strategie. Tabela 6.1 uwzględnia także wyniki modelu wykorzystującego sieć LSTM uczoną z wykorzystaniem funkcji straty zadanej jako MADL. Porównanie wyników, dla wszystkich badanych aktywów, zaprezentowane zostało w tabeli 6.1. W celu określenia najlepszych modeli należy zwrócić przede wszystkim uwagę na dwa wskaźniki: roczny skumulowany zwrot (aRC) oraz wskaźniki informacyjne (IR), głównie ze względu na to, że wskaźniki IR biorą pod uwagę także pozostałe wskaźniki dotyczące zmienności i ryzyka związanego z daną strategią (aSD, MD oraz MLD).

Porównując wyniki przedstawione w poniższej tabeli można zauważyć, że najlepsze wyniki w pod względem wskaźnika rocznego skumulowanego zwrotu (aRC) uzyskały sieci CNN, uzyskując najwyższe wartości tego wskaźnika dla trzech indeksów: S&P 500, KOSPI oraz BOVESPA. Sieci LSTM najlepiej wypadły dla indeksów NIKKEI oraz DAX (wynik ten został dodatkowo poprawiony w modelu LSTM trenowanym na funkcji MADL), natomiast dla indeksu WIG najlepszy wynik uzyskała sieć MLP.

Biorąc pod uwagę wskaźnik informacyjny IR^* , sieci LSTM uzyskały najlepsze wyniki dla indeksów NIKKEI, DAX, oraz BOVESPA, z tym że w tym przypadku były to sieci używające miernika MADL jako funkcji straty. Zbliżone wyniki uzyskały sieci CNN, które najlepiej spisywały się na indeksach S&P500 oraz KOSPI. Dla indeksu WIG najlepszy wynik uzyskała ponownie sieć MLP. Modele wykazują także podobne wyniki pod względem miernika IR^{**} – w tym przypadku sieci CNN dodatkowo uzyskały wynik na tym samym poziomie (0,05) co sieci LSTM dla indeksu BOVESPA.

Analizując wyniki sieci LSTM uczonej z wykorzystaniem miernika MADL jako funkcji straty można zauważyć, że cechuje się ona bardzo dobrymi wynikami ogólnymi. Dla wszystkich badanych aktywów strategie wykorzystujące ten model przynosiły zyski, często na poziomie zbliżonym do pozostałych badanych sieci. Modele te cechują się także najlepszymi ogólnymi wskaźnikami aSD oraz MD, co wskazuje, że charakteryzują się najniższym ryzykiem, dzięki czemu możliwe byłoby zastosowanie w tym przypadku dźwigni w celu uzyskania większych zysków. Dobre wyniki w kontekście wskaźników aSD oraz MD uzyskały także modele wykorzystujące sieci MLP, mimo że nie przełożyło się to w tym wypadku na pozostałe wskaźniki. Z kolei dla wskaźnika MLD, najlepiej spisywały się sieci CNN.

Tabela 6.1: Porównanie modeli punktowych prognoz stóp zwrotu, pod względem wskaźników oceny strategii inwestycyjnych.

Indeks/Strategia	aRC	aSD	MD	MLD	IR*	IR**	IR***
Sieci MLP (MSE)							
S&P500 LO	6,98	15,69	24,91	1,17	0,45	0,12	0,01
NIKKEI LO	1,42	18,46	41,01	3,63	0,08	0,00	0
DAX LO	0,35	17,36	35,57	6,46	0,02	0,00	0
WIG LO	2,15	15,36	47,38	2,90	0,14	0,01	0
KOSPI LO	0,38	15,75	43,78	5,08	0,02	0,00	0
BOVESPA LO	1,96	21,24	51,03	5,75	0,09	0,00	0
Sieci CNN (MSE)							
S&P500 LS	13,01	21,11	28,52	4,08	0,62	0,28	0,01
S&P500 LO	10,95	18,34	29,34	1,1	0,60	0,22	0,02
NIKKEI LO	4,48	19,30	39,31	2,74	0,23	0,03	0
DAX LO	2,26	18,92	43,48	5,40	0,12	0,01	0
WIG LO	-1,55	15,01	43,61	9,20	-0,10	0	0
KOSPI LS	9,20	19,66	39,44	2,48	0,47	0,11	0
BOVESPA LS	8,56	28,46	47,96	2,52	0,30	0,05	0
Sieci LSTM (MSE)							
S&P500 LS	9,07	21,12	34,50	2,75	0,43	0,11	0
S&P500 LO	8,92	18,70	34,03	3,05	0,48	0,13	0
NIKKEI LO	6,89	21,28	42,76	4,56	0,32	0,05	0
DAX LS	7,04*	22,97	47,11	2,92	0,31	0,05	0
DAX LO	6,47	19,49	39,89	2,93	0,33*	0,05	0
WIG LO	0,50	16,67	42,64	2,90	0,03	0	0
KOSPI LO	-0,41	17,68	43,12	8,83	-0,02	0,00	0
BOVESPA LO	5,07	24,24	52,48	4,56	0,21	0,02	0
Sieci LSTM (MADL)							
S&P500 LO	9,26	15,69	29,69	1,55	0,59	0,18	0,01
NIKKEI LO	3,27	15,81	30,00	4,43	0,21	0,02	0
DAX LS	8,06	22,97	57,47	8,11	0,35	0,05	0
WIG LO	1,14	12,80	38,85	7,09	0,09	0	0
KOSPI LO	3,29	14,55	38,97	4,62	0,23	0,02	0
BOVESPA LO	7,08	22,18	44,08	7,80	0,32	0,05	0

Uwaga: Pogrubiono najlepsze rezultaty uzyskane dla określonej strategii inwestycyjnej, tj. dla wskaźników aRC oraz IR pogrubione zostały najwyższe wartości dla poszczególnych indeksów, zaś dla wskaźników aSD, MD oraz MLD — wartości najniższe. Gwiazdką oznaczone zostały najlepsze wyniki spośród sieci wykorzystujących MSE jako funkcję straty.

Źródło: Opracowanie własne.

Na podstawie wyników przedstawionych w tabeli 6.1 nie sposób jednoznacznie wskazać, który typ sieci jest najlepszy pod względem punktowego prognozowania stóp zwrotu. Można natomiast zauważyć, że sieci uczenia głębokiego (LSTM oraz CNN) spisywały się lepiej niż klasyczne sieci MLP. Biorąc jednak dodatkowo pod uwagę wyniki miar błędu (MSE, MAE oraz trafności prognoz) z tabeli 5.4 sieci LSTM można wyróżnić jako spisujące się nieznacznie lepiej od sieci CNN.

Porównując sposoby konturowania strategii inwestycyjnych, należy zwrócić uwagę, że tabeli w znacznie większości znalazły się strategie wykorzystujące wyłącznie sygnały kupna (LO). Wynikać to może z faktu, że w tych strategiach możliwość popełnienia błędu jest ograniczona, jak i z tego, że badane instrumenty finansowe w postaci indeksów giełdowych charakteryzują się trendem wzrostowym, co także może mieć pozytywny wpływ na wyniki strategii bazujących na pozycjach długich.

6.1.2 Modele hybrydowe punktowych prognoz zmienności

W poniższym punkcie przedstawione zostały zestawienia szczegółowych wyników dla modeli wykorzystanych uprzednio w celu uzyskania punktowych prognoz zmienności – modeli ARMA-GARCH oraz hybrydowych modeli ARMA-GARCH-LSTM. Spośród wszystkich rozważanych specyfikacji modelowych do porównań wybrane zostały te, które uzyskały najniższe wartości błędu MSE oraz odpowiednich liczb przekroczeń VaR.¹ Zestawienia przygotowane zostały w tabelach osobno dla każdego z aktywów. W tabelach zawarte zostały dodatkowo także mierniki błędów MAE, HMSE i R^2 , a także wyniki (wartości p-value) testów Kupca i Christoffersena dla oszacowań wartości zagrożonej, oraz testu McNeila i Fraya dla oczekiwanego niedoboru (zob. rozdział 4). Testy przeprowadzone zostały zarówno dla 5%, jak i 1% poziomu tolerancji.

Indeks S&P 500

W tabeli 6.2 przedstawiono wyniki dla modeli, które uzyskały najlepsze rezultaty na danych pochodzących z indeksu S&P 500. Do zestawienia wybrane zostały modele GJR-GARCH(SSTD)-LSTM, GARCH(STD)-LSTM oraz GARCH(SSTD)-LSTM.

¹Szczegółowe wyniki dla wszystkich badanych specyfikacji modelowych przedstawione zostały w tabelach A.2 - A.25 zamieszczonych w Aneksie.

Tabela 6.2: Zestawienie modeli, indeks S&P 500, najlepsze modele pod względem MSE i ilości przekroczeń VaR

Mierniki/Model	GJR(STD)-LSTM	G(STD)-LSTM	G(SSTD)-LSTM
MSE	0,2810	0,3240	0,3315
MAE	0,3632	0,3707	0,3783
HMSE	0,3274	0,3472	0,3768
R^2	0,7677	0,5822	0,5668
Przek. VaR	112/ 24	127 /28	105/ 24
Przek. VaR (%)	4,5/ 0,97	5,11 /1,13	4,22/ 0,97
Kupiec 5%	0,2481	0,8080	0,0677
Kupiec 1%	0,8600	0,5363	0,8600
Christoff. 5%	0,1259	0,2230	0,0105*
Christoff. 1%	0,0051*	0,0107*	0,0052
ES bootstrap 5%	0,6861	0,5438	0,7727
ES próbkowy 5%	0,7225	0,5342	0,8536
ES bootstrap 1%	0,4358	0,4240	0,6902
ES próbkowy 1%	0,3510	0,4240	0,7638

Uwaga: Pogrubione zostały najniższe wartości MSE, oraz ilość przekroczeń VaR najbardziej zbliżona do przyjętego poziomu tolerancji. * oznaczone zostały wyniki testów, w których hipoteza zerowa została odrzucona, mimo akceptowalnej ilości przekroczeń.

Źródło: Opracowanie własne.

W przypadku indeksu S&P 500, model, który uzyskał najlepsze wyniki pod względem podstawowej miary błędu MSE, tj. GJR-GARCH(STD)-LSTM, osiągnął także najlepsze wyniki dla pozostałych wymienionych mierników błędu prognoz. Pod względem liczby przekroczeń VaR dla 1% poziomu tolerancji jest na tym samym poziomie, co model GARCH(SSTD)-LSTM. Natomiast dla 5% poziomu tolerancji najlepsze wyniki uzyskał model GARCH(STD)-LSTM. Przedstawione w tabeli modele mają także akceptowalne wyniki testów VaR i ES (wartości p-value). Wyjątki stanowią tutaj modele GJR-GARCH(STD)-LSTM oraz GARCH(STD)-LSTM, które uzyskały niską wartość p-value testu Christoffersena dla 1% poziomu tolerancji, oraz model GARCH(SSTD)-LSTM także pod względem testu Christoffersena dla 5% poziomu tolerancji.

Zauważyć można także, że dla rozważanego tu indeksu, w zestawieniu najlepszych modeli znalazły się tylko modele hybrydowe, w dwóch przypadkach bazujące na skośnym rozkładzie t-Studenta, a w jednym — na rozkładzie t-Studenta.

Indeks NIKKEI 225

Wyniki dla najlepszych modeli dla indeksu NIKKEI 225 przedstawione zostały w tabeli 6.3. W zestawieniu uwzględniono modele GJR-GARCH(SSTD)-LSTM, APARCH(SSTD) oraz APARCH(N)-LSTM.

Tabela 6.3: Zestawienie modeli, indeks NIKKEI 225, najlepsze modele pod względem MSE i ilości przekroczeń VaR

Mierniki/Model	GJR(SSTD)-LSTM	AP(SSTD)	AP(N)-LSTM
MSE	0,6277	0,7013	0,6386
MAE	0,5669	0,5645	0,5659
HMSE	4,4495	2,7125	4,6458
R^2	0,1699	0,1998	0,1568
Przek. VaR	74/11	122/27	74/ 23
Przek. VaR (%)	2,98/0,44	4,91/1,09	2,98/ 0,92
Kupiec 5%	5,89e-07*	0,8283	5,89e-07*
Kupiec 1%	0,0017*	0,6712	0,7027
Christoff. 5%	2,03e-06*	0,8924	2,03e-06*
Christoff. 1%	0,0008*	0,5333	0,4229
ES próbkowy 5%	—	0,2384	—
ES próbkowy 5%	—	0,1476	—
ES bootstrap 1%	—	0,2315	0,0398
ES próbkowy 1%	—	0,1445	0,0185*

Uwaga: Pogrubione zostały najniższe wartości MSE, oraz ilość przekroczeń VaR najbardziej zbliżona do przyjętego poziomu tolerancji. * oznaczone zostały wyniki testów, w których hipoteza zerowa została odrzucona, mimo akceptowalnej ilości przekroczeń. W tabeli pominięte zostały wyniki testu dotyczącego ES, w przypadku gdy wyniki obu testów na liczbę przekroczeń VaR były niewłaściwe.

Źródło: Opracowanie własne.

Pod względem miary MSE najlepszy wynik uzyskał model GJR-GARCH(SSTD)-LSTM. Wynik ten jednak nie przekłada się na dobre rezultaty pod względem pozostałych kryteriów, zarówno w odniesieniu do innych mierników błędu, jak i prognoz ryzyka, gdzie lepsze wyniki uzyskały pozostałe modele. Liczba przekroczeń VaR w przypadku tego modelu była znacznie niższa niż założona, co odbiło się także na wynikach testów.

Pod względem liczby przekroczeń VaR dla 5% poziomu tolerancji najlepszy wynik dla tego indeksu uzyskał model APARCH(SSTD). Wyniki testów wstecznych dla tego modelu są także we wszystkich wypadkach poprawne. Model ten uzyskał także najniższą wartość mierników HMSE, MAE i najniższą miernika R^2 , oraz poprawne

wyniki testu wstecznego McNeila i Fraya (ES), dla obydwu rozważanych poziomów tolerancji. Dla 1% VaR najlepszy wynik uzyskał model APARCH(N)-LSTM. Nie sprawdził się on jednak dla poziomu 5%, gdzie otrzymał wyniki znacznie zaniżone. Także w przypadku testu wstecznego dla 1% ES wartości p-value były zbyt niskie.

Indeks DAX

Dla indeksu DAX szczegółowe wyniki przedstawione zostały w tabeli 6.4, zestawiającej w tym przypadku modele GJR-GARCH(SSTD)-LSTM, EGARCH(SSTD) oraz APARCH(N)-LSTM.

Tabela 6.4: Zestawienie modeli, indeks DAX, najlepsze modele pod względem MSE i ilości przekroczeń VaR

Mierniki/Model	GJR(SSTD)-LSTM	E(SSTD)	AP(N)-LSTM
MSE	0,7446	1,1950	0,7755
MAE	0,5396	0,6437	0,5483
HMSE	0,3278	0,1699	0,3300
R^2	0,3765	0,2423	0,3679
Przek. VaR	60/10	150/38	63/21
Przek. VaR (%)	2,69/0,48	6,03/1,53	2,53/ 0,84
Kupiec 5%	5,51e-11*	0,02207*	5,19e-10*
Kupiec 1%	0,0007*	0,01408*	0,4229
Christoff. 5%	4,19e-10*	0,0582	2,46e-09*
Christoff. 1%	0,0029*	0,0432*	0,6065
ES bootstrap 5%	—	0,2602	—
ES próbkowy 5%	—	0,1686	—
ES bootstrap 1%	—	—	0,1535
ES próbkowy 1%	—	—	0,0874

Uwaga: Pogrubione zostały najniższe wartości MSE, oraz ilość przekroczeń VaR najbardziej zbliżona do przyjętego poziomu tolerancji. * oznaczone zostały wyniki testów, w których hipoteza zerowa została odrzucona, mimo akceptowalnej ilości przekroczeń. W tabeli pominięte zostały wyniki testu dotyczącego ES, w przypadku gdy wyniki obu testów na liczbę przekroczeń VaR były niewłaściwe.

Źródło: Opracowanie własne.

W kontekście mierników błędów prognoz najlepsze wyniki uzyskał model GJR-GARCH(SSTD)-LSTM, z wyjątkiem miary HMSE. Prognozy uzyskane z wykorzystaniem tego modelu dawały jednak zdecydowanie zawyżone oszacowania ryzyka, co skutkuje zbyt niską ilością przekroczeń VaR w stosunku do założonej. Dla 5% poziomu tolerancji najlepsze prognozy uzyskał model EGARCH(SSTD), także

pod kątem testu Christoffersena dla VaR oraz testu McNeila i Freya, choć w przypadku testu Kupca hipoteza zerowa została odrzucona. Model ten uzyskał także najniższą wartość dla miernika HMSE. Z kolei dla 1% poziomu tolerancji najlepszy wynik uzyskał model APARCH(N)-LSTM, uzyskując także wysokie wartości p-value dla testów wstecznych zarówno dla wartości zagrożonej, jak i oczekiwanego niedoboru.

Indeks WIG

Zestawienie dla indeksu WIG przedstawione zostało w tabeli 6.5.

Tabela 6.5: Zestawienie modeli, indeks WIG, najlepsze modele pod względem MSE i ilości przekroczeń VaR

Mierniki/Model	GJR(N)-LSTM	GJR(SSTD)	E(STD)-LSTM
MSE	0,4320	0,4730	0,4868
MAE	0,3878	0,3873	0,4054
HMSE	0,2457	0,1476	0,2541
R^2	0,2663	0,2537	0,2004
Przek. VaR	89/27	134/30	99/ 24
Przek. VaR (%)	3,58/1,09	5,39/1,21	3,98/ 0,97
Kupiec 5%	0,0006*	0,3803	0,0157*
Kupiec 1%	0,6720	0,3166	0,8600
Christoff. 5%	0,0010*	0,2588	0,0091
Christoff. 1%	0,6797	0,4198	0,4796*
ES bootstrap 5%	—	0,1012	—
ES próbkowy 5%	—	0,0376	—
ES bootstrap 1%	0,0010	0,0352	0,1890
ES próbkowy 1%	0,0001*	0,0088*	0,1107

Uwaga: Pogrubione zostały najniższe wartości MSE, oraz ilość przekroczeń VaR najbardziej zbliżona do przyjętego poziomu tolerancji. * oznaczone zostały wyniki testów, w których hipoteza zerowa została odrzucona, mimo akceptowalnej ilości przekroczeń. W tabeli pominięte zostały wyniki testu dotyczącego ES, w przypadku gdy wyniki obu testów na liczbę przekroczeń VaR były niewłaściwe.

Źródło: Opracowanie własne.

W przypadku tego indeksu, jako najlepsze wybrane zostały modele GJR-GARCH(N)-LSTM, GJR-GARCH(SSTD) oraz EGARCH(STD)-LSTM. Pierwszy z tych modeli uzyskał najniższe wyniki mierników błędów prognoz, z wyjątkiem HMSE i MAE. Prognozy poziomu wartości zagrożonej uzyskane z wykorzystaniem tego modelu były zbyt wysokie dla 5% poziomu tolerancji VaR (co

skutkuje zbyt niską liczbą przekroczeń), natomiast na poziomie 1% były one poprawne. W przypadku testu McNeila i Freya dla 1% poziomu tolerancji wyniki p-value były bardzo niskie. Model GJR-GARCH(SSTD) uzyskał najlepsze wyniki jeśli chodzi o poziom VaR na poziomie 5%, przy czym dla poziomu 1% też były one poprawne. Wyniki testu dla ES dla tego modelu były poprawne dla 5% poziomu tolerancji, ale niskie dla poziomu 1%. Model uzyskał też najniższe miary błędu HMSE oraz MAE. Dla poziomu 1%, zarówno dla VaR, jak i ES najlepszy wynik uzyskał model EGARCH(SSTD)-LSTM, jednak zbyt wysoko oszacował VaR dla 5% poziomu tolerancji.

Indeks KOSPI

Dla indeksu KOSPI zestawienie przedstawione zostało w tabeli 6.6, zawierającej szczegółowe dane o wynikach modeli APARCH(N)-LSTM oraz APARCH(SSTD).

Tabela 6.6: Zestawienie modeli, indeks KOSPI, najlepsze modele pod względem MSE i ilości przekroczeń VaR

Mierniki/Model	AP(N)-LSTM	AP(SSTD)
MSE	0,5318	0,7324
MAE	0,4403	0,5024
HMSE	0,2372	0,1407
R^2	0,3248	0,3991
Przek. VaR	72/20	153/31
Przek. VaR (%)	2,9/ 0,8	6,15/1,25
Kupiec 5%	1,87e-07*	0,0108*
Kupiec 1%	0,3096	0,2342
Christoff. 5%	5,99e-07*	0,0122*
Christoff. 1%	0,5074	0,3331
ES bootstrap 5%	—	—
ES próbkowy 5%	—	—
ES bootstrap 1%	0,0377	0,5111
ES próbkowy 1%	0,0124*	0,4713

Uwaga: Pogrubione zostały najniższe wartości MSE, oraz ilość przekroczeń VaR najbardziej zbliżona do przyjętego poziomu tolerancji. * oznaczone zostały wyniki testów, w których hipoteza zerowa została odrzucona, mimo akceptowalnej ilości przekroczeń. W tabeli pominięte zostały wyniki testu dotyczącego ES, w przypadku gdy wyniki obu testów na liczbę przekroczeń VaR były niewłaściwe.

Źródło: Opracowanie własne.

Pierwszy z tych modeli uzyskał najniższe miary błędu MSE i MAE oraz najlepszy

wynik dla liczby przekroczeń VaR na poziomie 1%, poparty także wynikami testów wstecznych. Na poziomie 5% model ten prognozował jednak zawyżony poziom VaR, w rezultacie uzyskując niskie frakcje przekroczeń. Także w tescie McNeila i Freya dla poziomu 1% model ten uzyskał zbyt niskie wyniki. W tym zestawieniu nieco lepiej wypadł model APARCH(SSTD), warto zauważyć jednak, że wartości p-value dla testów wstecznych dla tego modelu są niskie dla poziomu 5%, co oznacza, że pomimo najlepszych prognoz poziomu wartości zagrożonej spośród wszystkich testowanych modeli, poziomy te wciąż są w tym przypadku zaniżone. Dla 1% poziomu tolerancji, wyniki testów wstecznych VaR i ES dla tego modelu są natomiast prawidłowe.

Indeks BOVESPA

Dla indeksu BOVESPA, do zestawienia zaprezentowanego w tabeli 6.7, wybrane zostały modele EGARCH(N)-LSTM, EGARCH(STD) oraz GJR-GARCH(STD).

Tabela 6.7: Zestawienie modeli, indeks BOVESPA, najlepsze modele pod względem MSE i ilości przekroczeń VaR

Mierniki/Model	E(N)-LSTM	E(STD)	GJR(STD)
MSE	0,6504	0,8182	0,7942
MAE	0,5364	0,5532	0,5654
HMSE	0,2506	0,1730	0,1536
R^2	0,5072	0,3992	0,4550
Przek. VaR	60/14	123/19	135/ 25
Przek. VaR (%)	3,58/1,09	4,95/0,76	5,43/ 1,01
Kupiec 5%	5,51e-11*	0,9020	0,3335
Kupiec 1%	0,0170*	0,2171	0,9791
Christoff. 5%	1,05e-10*	0,9226	0,3950
Christoff. 1%	0,0535	0,4033	0,7754
ES bootstrap 5%	—	0,7938	0,2229
ES próbkowy 5%	—	0,8796	0,1365
ES bootstrap 1%	0,0232	0,6882	0,1027
ES próbkowy 1%	0,0108*	0,7377	0,0438*

Uwaga: Pogrubione zostały najniższe wartości MSE, oraz ilość przekroczeń VaR najbardziej zbliżona do przyjętego poziomu tolerancji. * oznaczone zostały wyniki testów, w których hipoteza zerowa została odrzucona, mimo akceptowalnej ilości przekroczeń. W tabeli pominięte zostały wyniki testu dotyczącego ES, w przypadku gdy wyniki obu testów na liczbę przekroczeń VaR były niewłaściwe.

Źródło: Opracowanie własne.

Pierwszy z wymienionych modeli uzyskał najniższe wyniki mierników błędu

z wyjątkiem HMSE, jednak prognozy poziomu wartości zagrożonej uzyskane przez ten model były zawyżone dla 5% poziomu tolerancji. Dla 5% poziomu tolerancji najlepsze prognozy VaR uzyskał model EGARCH(STD). Także dla poziomu 1% model ten prawidłowo prognozował poziom VaR, jednak lepsze wyniki pod tym względem uzyskał model GJR-GARCH(STD), uzyskując także najniższą wartość HMSE. Oba te modele uzyskały prawidłowe wyniki testów wstecznych, można także wnioskować, że poprawnie szacują także poziom oczekiwanego niedoboru.

Podsumowanie

Na podstawie przedstawionych w tym punkcie rezultatów, dotyczących wszystkich indeksów, można wnioskować, że modele hybrydowe lepiej spisywały się w szacowaniu ryzyka na 1% poziomie tolerancji. Świadczyć to może o tym, że uzyskane za pomocą tych modeli oszacowania VaR są zdecydowanie bardziej konserwatywne (prognozowany jest wyższy poziom VaR), niż w przypadku modeli klasy GARCH.²

6.1.3 Modele prognozowania probabilistycznego

Poniższy punkt dotyczy rezultatów uzyskiwanych przez sieciowe modele prognozowania probabilistycznego. W pierwszej kolejności, porównując wyniki zaprezentowane w tabelach 5.13 oraz 5.14, należy zwrócić uwagę na fakt, że dobre wyniki uzyskane w kontekście mierników ocen trafności prognoz nie zawsze przekładają się na poprawne prognozy poziomu wartości zagrożonej. O ile w pierwszym przypadku zdecydowanie lepiej spisywały się modele LSTM ze skośnym rozkładem t-Studenta, to nie znajduje to jednak odzwierciedlenia w wynikach tego modelu w kontekście VaR. Podobna sytuacja miała miejsce w przypadku: modeli hybrydowych (gdzie niskie oceny miernika MSE nie szły w parze z poprawą oszacowań ryzyka), oraz modeli stosowanych do punktowych prognoz zmienności (w których wyniki strategii inwestycyjnych często nie były jednoznaczne z niskimi ocenami dokładności prognoz punktowych). Także w przypadku modeli hybrydowych i probabilistycznych świadczyć to może o konieczności wprowadzenia innej funkcji straty, stosowanej w celu poprawy prognoz wartości zagrożonej.

²Do takich rezultatów przyczynić się mógł wysoki poziom ocen zmienności wynikających z estymatora GKYZ.

Porównanie modeli w kontekście prognozowania rozkładów prawdopodobieństwa

W tabeli 6.8 przedstawione zostało porównanie wybranych sieciowych modeli prognozowania probabilistycznego z modelami klasy AR-GARCH. Spośród badanych modeli w zestawieniu wykorzystane zostały specyfikacje wykorzystujące skośny rozkład t-Studenta — modele sieciowe bazujące na architekturze CNN i LSTM porównane zostały z modelami GJR-GARCH i APARCH. Do porównania prognoz probabilistycznych zastosowane zostały mierniki LPS, CRPS, oraz test Andersona-Darlinga (p-value) dla poprawności kalibracji modelu z wykorzystaniem kryterium PIT.

Analizując przedstawione w tabeli wyniki można zauważyć, że zestawione w niej specyfikacje uzyskały zbliżone wyniki. Modele APARCH i GJR-GARCH cechują się częściej niższymi wartościami CRPS — w przypadku modelu APARCH miało to miejsce dla indeksów S&P 500, NIKKEI oraz DAX, zaś dla modelu GJR-GARCH w przypadku indeksów WIG, KOSPI oraz BOVESPA. Z kolei modele LSTM wykazują niższe wartości miernika LPS, co miało miejsce w przypadku czterech indeksów: NIKKEI, DAX, WIG oraz BOVESPA. Lepsze wyniki modeli sieciowych pod względem LPS są mogą wynikać z faktu, że kryterium to jest zbliżone w konstrukcji do funkcji straty minimalizowanych przez sieć.³

Biorąc pod uwagę wartości p-value testu Andersona-Darlinga, najlepsze wyniki uzyskały sieci CNN — miało to miejsce w przypadku indeksów S&P 500, NIKKEI i BOVESPA. W przypadku dwóch indeksów — WIG oraz KOSPI, wyniki lepsze od pozostałych uzyskał model GJR(SSTD). Należy jednak zwrócić uwagę na problemy z osiągnięciem prawidłowej kalibracji prognoz rozkładów w przeważającej liczbie rozważanych przypadków modeli i instrumentów. Jedynie w sześciu przypadkach możemy stwierdzić, że była ona prawidłowa.

³Podobna sytuacja miała miejsce także w przypadku modeli hybrydowych, które poprawiały wyniki oceny prognoz przy pomocy miernika MSE, wykorzystywanego jako funkcja straty komponentu sieciowego.

Tabela 6.8: Porównanie modeli prognozowania probabilistycznego z modelami klasy AR-GARCH pod kątem oceny trafności prognoz

Kryteria/Model	CNN-SSTD	LSTM-SSTD	GJR(SSTD)	APARCH(SSTD)
Indeks S&P 500				
LPS	1,2220	1,1934	1,2321	1,1700
CRPS	0,5197	0,5095	0,5017	0,5012
PIT p-value	0,0309	1,56e-05	2,41e-07	2,41e-07
Indeks NIKKEI				
LPS	1,5865	1,5854	1,5900	1,5921
CRPS	0,6963	0,6874	0,6831	0,6826
PIT p-value	0,0476	2,41e-07	2,41e-07	2,41e-07
Indeks DAX				
LPS	1,5708	1,5230	1,5316	1,5364
CRPS	0,6901	0,6646	0,6607	0,6606
PIT p-value	2,41e-07	0,3537	2,41e-07	2,41e-07
Indeks WIG				
LPS	1,4055	1,3475	1,3600	1,3582
CRPS	0,5716	0,5554	0,5516	0,5519
PIT p-value	2,41e-07	0,0087	0,7779	2,41e-07
Indeks KOSPI				
LPS	1,3172	1,2848	1,2616	1,2643
CRPS	0,5302	0,5165	0,5146	0,5147
PIT p-value	2,41e-07	5,09e-06	0,1953	0,1606
Indeks BOVESPA				
LPS	1,7875	1,7851	1,8361	1,7992
CRPS	0,8349	0,8335	0,8229	0,8233
PIT p-value	0,0069	0,0056	2,41e-07	2,41e-07

Uwaga: Pogrubione zostały najniższe wartości LPS i CRPS, oraz najwyższe wartości p-value.

Źródło: Opracowanie własne.

Porównanie modeli prognozowania probabilistycznego i modeli punktowych

W kolejnym kroku porównane zostały rezultaty uzyskane przez modele prognozowania probabilistycznego z wynikami grupy modeli hybrydowych oraz

AR-GARCH, stosowanych do punktowego prognozowania zmienności, w kontekście szacowania ryzyka. W tym celu, w tabeli 6.9, przedstawiono zestawienie najlepszych specyfikacji modelowych dla poszczególnych aktywów pod kątem procentowego udziału przekroczeń wartości zagrożonej.

Tabela 6.9: Porównanie modeli prognozowania probabilistycznego z modelami hybrydowymi oraz GARCH — procentowy udział przekroczeń VaR dla wszystkich aktywów.

Indeks/Model	NN-D	% Przekr.	Model hybrydowy	% Przekr.
S&P VaR 5%	LSTM-N	4,86	G(STD)-LSTM	5,11
NKX VaR 5%	LSTM-STD	4,58	AP(SSTD)	4,91
DAX VaR 5%	CNN-STD	5,42	E(SSTD)	6,03
WIG VaR 5%	LSTM-STD	4,95	GJR(SSTD)	5,39
KOSPI VaR 5	LSTM-N	5,42	AP(SSTD)	6,15
BVP VaR 5%	LSTM-N	5,18	E(STD)	4,95
S&P VaR 1%	LSTM-STD	1,01	G(SSTD)-LSTM/GJR(STD)-LSTM	0,97
NKX VaR 1%	LSTM-STD	1,01	AP(N)-LSTM	0,92
DAX VaR 1%	CNN-STD/LSTM-SSTD-	1,12/0,88	AP(N)-LSTM	0,84
WIG VaR 1%	CNN-STD	1,12	E(STD)-LSTM	0,97
KOSPI VaR 1%	LSTM-STD	0,84	AP(N)-LSTM	0,80
BVP VaR 1%	CNN-N	1,20	GJR(STD)	1,01

Źródło: Opracowanie własne.

Analizując procentowe udziały przekroczeń VaR, nie można jednoznacznie wskazać grupy modeli, które uzyskały znacząco lepsze wyniki od pozostałych. Zarówno modele prognozowania probabilistycznego jak i modele hybrydowe oraz GARCH uzyskują zbliżone wyniki. Pewne różnice zauważyć można w przypadku indeksów DAX oraz KOSPI przy 5% poziomie tolerancji, gdzie udział przekroczeń dla modeli odpowiednio EGARCH i APARCH ze skośnym rozkładem t-Studenta wynosi powyżej 6%, natomiast w przypadku modeli prognoz probabilistycznych nie występują tak duże odchylenia. Podobny wniosek można wysnuć porównując wyniki z tabel 5.14 oraz 5.12, omawianych w poprzednim rozdziale. W grupie modeli hybrydowych i GARCH uzyskane wartości frakcji przekroczeń charakteryzują się większymi odstępstwami od przyjętych poziomów tolerancji, niż w przypadku grupy modeli prognozowania probabilistycznego.

6.2 Podsumowanie wyników

W ramach tego podrozdziału ogólne wyniki prezentowane w rozdziale 5 oraz ich analiza porównawcza przeprowadzona w punkcie 6.1.1 podsumowane zostaną

w odniesieniu do sformułowanych we wstępie pracy poszczególnych celów oraz hipotez badawczych.

6.2.1 Realizacja celów badawczych

Podsumowując wyniki przedstawione w poprzednim podrozdziale można wnioskować, że główny cel pracy został osiągnięty. W pracy pokazano, że modele bazujące na sieciach uczenia głębokiego można z powodzeniem stosować w prognozowaniu finansowych szeregów czasowych. Modele te wykorzystane zostały do uzyskania prognoz na trzech różnych poziomach - punktowych prognoz stóp zwrotu, punktowych prognoz zmienności (za pomocą modeli hybrydowych), oraz do prognozowania całych rozkładów prawdopodobieństwa. We wszystkich trzech przypadkach rozważane modele sieciowe uzyskały zadowalające, choć nie bezbłędne rezultaty.

W kolejnych rozdziałach pracy zrealizowane zostały także poszczególne cele cząstkowe. Podrozdział 5.4 oraz punkt 6.1.1 odnoszą się do celów dotyczących efektywności modeli w kontekście prognoz punktowych stóp zwrotu, ich oceny poprzez zastosowanie uzyskanych prognoz do budowy strategii inwestycyjnych, a także sprawdzenia wyników w zależności od poziomu rozwinięcia badanych rynków. Punkt 4.2.1 oraz podrozdział 5.5 dotyczą celów badawczych związanych z konstrukcją i oceną modeli hybrydowych stosowanych do uzyskiwania punktowych prognoz zmienności oraz ich ocenę w kontekście szacowania ryzyka kapitałowego. Z kolei w punkcie 4.2.2 oraz podrozdziale 5.6 zrealizowane zostały cele związane z budową sieciowych modeli prognozowania probabilistycznego. Porównanie efektywności modeli ekonometrycznych z modelami wykorzystującymi sieci uczenia głębokiego przedstawione zostało w punktach 6.1.2 oraz 6.1.3. Realizacja wspomnianych celów badawczych możliwa była poprzez stworzenie autorskich kodów, co dokładniej omówione zostało w podrozdziale 5.2.

6.2.2 Odniesienie do hipotez badawczych

W poniższej części pracy zweryfikowane zostaną poszczególne hipotezy badawcze, które sformułowane zostały na początku badań i przedstawione we wstępie do niniejszej dysertacji. W pierwszej kolejności omówione zostaną hipotezy szczegółowe, a następnie hipoteza główna, która jest niejako ich podsumowaniem.

Pierwsza z postawionych hipotez szczegółowych głosiła, że modele oparte na uczeniu głębokim generują lepsze efekty predykcji w porównaniu z modelami wykorzystującymi klasyczne metody sztucznej inteligencji, w kontekście punktowych prognoz stóp zwrotu.

Wyniki przedstawione w podrozdziale 5.4 oraz punkcie 6.1.1, na podstawie których wnioskować można, że badane typy sieci uczenia głębokiego w postaci rekurencyjnych sieci LSTM oraz sieci konwolucyjnych istotnie przyniosły lepsze rezultaty predykcji, w porównaniu z klasycznymi sieciami MLP. Wyniki te nie dają podstaw do odrzucenia omawianej hipotezy.

Należy jednak mieć na uwadze, że rezultaty uzyskane poprzez sieci neuronowe w dużym stopniu zależą od struktury sieci oraz ustalonych hiperparametrów, w związku z czym mogą wyniki uzyskane przez poszczególne specyfikacje modelowe mogą nie być optymalne. Zauważyć także należy, że sieci MLP są tylko jednym z bardzo licznych narzędzi klasycznego uczenia maszynowego, cechujących się często różnymi sposobami konstrukcji, nie należy więc, opierając się na tym jednym przykładzie dokonywać nadmiernie szerokiego uogólnienia.

Kolejna z badanych hipotez głosiła, że prognozy punktowe stóp zwrotu uzyskane przez modele sieciowe, można wykorzystać do budowy skutecznych strategii inwestycyjnych. Wyniki przedstawione w podrozdziale 5.4 oraz punkcie 6.1.1 pokazują, że budowa strategii inwestycyjnych na podstawie prognoz uzyskanych przez modele sieciowe jest możliwa, jednak w większości przypadków nie przynoszą one znacząco lepszych rezultatów od prostej strategii pasywnej. W przypadku sieci LSTM wykorzystujących miarę MADL do budowy strategii bazującej na pozycjach długich, można stwierdzić, że model ten może służyć jako podstawa w konstrukcji narzędzi inwestycyjnych, ponieważ ze względu na niski poziom ekspozycji na ryzyko umożliwia wprowadzenie mechanizmu dźwigni. Poza tym przypadkiem, badane modele nie wskazują na możliwości poprawy rezultatów względem strategii pasywnej. Mając dodatkowo na uwadze, że w procesie inwestowania należy brać pod uwagę także dodatkowe koszty transakcyjne, na podstawie uzyskanych wyników hipotezę tą należy odrzucić.

Na podstawie tych samych wyników (zob. podrozdział 5.4 oraz punkt 6.1.1), odnieść się można także do hipotezy, głoszącej że prognozy stóp zwrotu indeksów notowanych na rynkach wschodzących dają lepsze wyniki w kontekście strategii inwestycyjnych w porównaniu z indeksami z rynków rozwiniętych. Wyniki te nie wykazują znacznych różnic pomiędzy prognozami uzyskanymi na podstawie danych pochodzących z rynków rozwijających się (reprezentowanych przez indeksy WIG, KOSPI oraz BOVESPA) w stosunku do rynków rozwiniętych. W związku z tym, w tym przypadku postawioną hipotezę także należałoby odrzucić.

W kontekście punktowych prognoz zmienności, zbadana została hipoteza głosząca, że połączenie metod ekonometrycznych z metodami uczenia głębokiego w ramach modeli hybrydowych przyczynia się do poprawy efektywności prognoz zmienności.

Wyniki dotyczące punktowych predykcji zmienności omówione zostały w podrozdziale 5.5 oraz punkcie 6.1.2. Na podstawie ocen trafności predykcji wykorzystujących miarę MSE stosowaną jako funkcję straty w modelach hybrydowych, można zauważyć, że wyniki uzyskane przez te modele uległy poprawie we wszystkich badanych przypadkach, co nie daje podstaw do odrzucenia tej hipotezy. Jeśli jednak dokładniej przyjrzeć się pozostałym miernikom oceny trafności prognoz, oraz wynikom testów na istotność różnic miar błędów prognoz, przedstawionych w tabelach A.2 - A.25, można zauważyć, że różnice nie zawsze były istotne, a pozostałe mierniki błędów niekiedy wskazały odmienne rezultaty.

Oceny prognoz punktowych, uzyskane przy pomocy typowych mierników, okazały się niejednoznaczne, niemniej jednak przeprowadzono również badania pod kątem prognozowania poziomów ryzyka kapitałowego. W tym przypadku hipoteza badawcza głosiła, że prognozy zmienności uzyskane przez modele hybrydowe przyczyniają się do poprawy prognoz ryzyka kapitałowego. W tym kontekście, bazując na wynikach zaprezentowanych w punktach 5.5.2 oraz 6.1.2, można zauważyć, że modele hybrydowe przyczyniały się do poprawy wyników przede wszystkim w przypadku prognoz wartości zagrożonej z jednoprocetowym poziomem tolerancji. Natomiast w przypadku poziomu pięcioprocetowego natomiast dawały prognozy zawyżone. Biorąc pod uwagę najnowsze zalecenia Komitetu Bazylejskiego, w których preferowany jest jednoprocetowy poziom VaR oraz modele bardziej konserwatywne, można wnioskować, że uzyskane wyniki nie dają podstaw do odrzucenia przyjętej hipotezy. Dodatkowym atutem analizowanych specyfikacji modelowych (także tych dotyczących prognozowania probabilistycznego) są zadowalające wyniki uzyskane w kontekście prognoz oczekiwanego niedoboru.

Kolejna spośród badanych hipotez głosiła, że sieci neuronowe uczenia głębokiego można wykorzystać jako narzędzie w prognozowaniu całych rozkładów prawdopodobieństwa. W celu zweryfikowania tej hipotezy zaproponowane zostały sieciowe modele prognozowania probabilistycznego umożliwiające prognozowanie parametrów różnych zadanych typów rozkładów predykcyjnych, oraz zdefiniowane zostały stosowane w nich nowe funkcje straty. Wyniki omówione w podrozdziale 5.6, dotyczące oceny jakości prognoz probabilistycznych oraz szacowania ryzyka kapitałowego z wykorzystaniem tych modeli, wskazują, że nie ma podstaw do odrzucenia tej hipotezy.

Sieciowe modele prognoz probabilistycznych wykorzystane zostały także w celu weryfikacji hipotezy głoszącej, że prognozy probabilistyczne uzyskane przez sieci uczenia głębokiego przynoszą lepsze rezultaty w kontekście szacowania ryzyka kapitałowego w porównaniu z modelami klasy ARMA-GARCH. W punkcie 6.1.3

zaprezentowane zostały wyniki porównujące ze sobą skuteczność badanych modeli. Na ich podstawie można wnioskować, że sieciowe modele prognozowania probabilistycznego, mimo że charakteryzują się nieco mniejszymi odstępstwami procentowych udziałów przekroczeń od przyjętych poziomów tolerancji, nie uzyskują jednoznacznie lepszych prognoz od modeli punktowych prognoz zmienności (zarówno modeli hybrydowych, jak ARMA-GARCH). W związku z tym hipotezę tą należałoby odrzucić.

Z uwagi na wyniki związane z badaniem poszczególnych hipotez szczegółowych, można wnioskować, że nie ma również podstaw do odrzucenia głównej hipotezy pracy, głoszącej że modele predykcyjne oparte na metodach uczenia głębokiego, można efektywnie wykorzystywać do prognozowania finansowych szeregów czasowych w postaci logarytmicznych stóp zwrotu. Odrzucone zostały hipotezy dotyczące budowy strategii inwestycyjnych opartych o modele uczenia głębokiego oraz o przewadze sieciowych modeli probabilistycznych nad modelami klasy ARMA-GARCH. Pozostałe zaprezentowane w pracy wyniki świadczą o tym, że badane modele wykorzystujące sieci uczenia głębokiego dobrze sprawdzają się w prognozowaniu zmienności i szacowaniu ryzyka kapitałowego, przede wszystkim jako narzędzia uzupełniające w połączeniu z metodami ekonometrycznymi.

6.3 Zalety i ograniczenia wynikające ze stosowania metod uczenia głębokiego

Jak wykazano w pracy, modele bazujące na sieciach uczenia głębokiego mogą stanowić pomocne narzędzie podczas modelowania i prognozowania instrumentów finansowych. Aby umożliwić pełniejszą ocenę tych modeli, należy wskazać zalety i wady związane z ich zastosowaniem.

Do głównych zalet modeli uczenia głębokiego należy ich elastyczność. Jak wiadomo, modele te stosowane są w bardzo szerokim zakresie, z którego badanie szeregów czasowych jest tylko niewielkim wycinkiem. Ta elastyczność narzędzi uczenia głębokiego umożliwia pracę z po pierwsze, pracę z dowolnymi zestawami danych, a po drugie – znacznie ułatwia łączenie ich z narzędziami bardziej wyspecjalizowanymi w danej dziedzinie, jak w tym przypadku z modelami ekonometrycznymi. Tego typu rozwiązania hybrydowe najczęściej przyczyniają się do poprawy uzyskiwanych wyników oraz otwierają nowe możliwości badawcze. Elastyczność ta pozwala także na stosunkowo prostą możliwość dostosowania konstrukcji modelu w zależności od danego problemu badawczego, jak miało to w przypadku budowy sieciowych modeli prognoz probabilistycznych.

Stosowanie modeli opartych na sieciach neuronowych uczenia głębokiego umożliwia także modelowanie danych w postaci zarówno jednowymiarowej, jak to miało w przypadku tej pracy, jak i wielowymiarowych, przykładowo w celu analizy całego portfela inwestycyjnego. Dobór danych wejściowych do sieci może być w zasadzie zupełnie dowolny, jeśli tylko uznamy, że dodatkowe zmienne wejściowe mogą przyczynić się do poprawy wyników uzyskiwanych przez model. W wielu badaniach, m.in. opisanych w podrozdziałach 1.3 oraz 4.1, jako dane wejściowe stosowane było nawet kilkanaście zmiennych, obejmujących przykładowo różne typy aktywów czy też oszacowania parametrów modeli. Podobnie, ilość wektorów wyjściowych sieci zależy jedynie od ilości neuronów w warstwie wyjściowej i odpowiedniego dopasowania funkcji straty minimalizowanej przez model, co, jak wspomniano, pozwala także na modelowanie danych wielowymiarowych. Także w przypadku zaproponowanego w tej pracy modelu prognozowania probabilistycznego własność ta umożliwiła zmianę ilości prognozowanych parametrów rozkładu.⁴

Dodatkowo, dzięki stosowaniu w sieciach uczenia głębokiego nieliniowych funkcji aktywacji oraz wielu warstw ukrytych możliwe jest wykrywanie skomplikowanych i często wysoce nieliniowych zależności, co pozwala w stosunkowo prosty sposób modelować dane o zróżnicowanej strukturze.

Modele bazujące na sieciach uczenia głębokiego nie są jednak pozbawione wad. Jednym z podstawowych problemów jest odpowiedni dobór hiperparametrów sieci, od których w dużej mierze zależą uzyskiwane wyniki. Ze względu na dużą liczbę hiperparametrów i możliwości dodatkowych zmian w strukturze sieci, jest to najczęściej proces czasochłonny, co więcej nie daje pewności, że wybrane końcowo wartości są optymalne. Dodatkowym problemem jest wybór odpowiedniego okna danych, na których przeprowadzany jest proces dostrajania hiperparametrów. W przypadku finansowych szeregów czasowych, wybór jednego wycinka danych w żaden sposób nie gwarantuje nam, że w przyszłości będziemy mieli do czynienia z danymi o podobnej charakterystyce, w związku z czym hiperparametry, które mogą okazać się optymalne dla tego wycinka, wcale nie muszą się sprawdzić podczas ewaluacji sieci na zbiorze testowym lub w trakcie jej eksploatacji.

Podobny problem dotyczy wyboru minimalizowanej w trakcie uczenia sieci funkcji straty, co również może mieć kluczowy wpływ na wyniki. Ogólnie stosowane w prognozowaniu finansowych szeregów czasowych funkcje straty często okazują się niewystarczające w celu rozwiązania takich problemów jak budowa optymalnych strategii inwestycyjnych czy szacowanie poziomu ryzyka. Łączy się to także często

⁴Parametry te, w określonych przypadkach, traktować można także jako punktowe prognozy stóp zwrotu oraz zmienności, co otwiera nowe możliwości implementacyjne proponowanego modelu.

z koniecznością tworzenia nowych postaci funkcji straty, jak miało to w tym przypadku miejsce przy podjęciu próby prognozowania probabilistycznego, oraz omawianej w kontekście strategii inwestycyjnych funkcji MADL.⁵

Utrudnieniem, które zaobserwowano podczas badań empirycznych dotyczących prognozowania parametrów rozkładu prawdopodobieństwa, były problemy z uzyskaniem zbieżności przez model, szczególnie w przypadku prognozowania parametrów skośnego rozkładu t-Studenta. Spowodowało to konieczność zmiany wielkości paczki wsadowej (na znacznie większą niż początkowo założono) oraz konieczność stosowania schematu rekursywnego, aby zwiększyć ilość dostępnych danych wejściowych. Schemat rekursywny musiał być zastosowany także w przypadku modeli hybrydowych, wykorzystanych w tej pracy do punktowego prognozowania zmienności, w celu poprawy uzyskiwanych wyników.

6.4 Proponowane kierunki dalszych prac badawczych

W poniższym podrozdziale zaproponowane zostały dalsze kierunki badań nad zastosowaniem narzędzi wykorzystujących sieci uczenia głębokiego w prognozowaniu finansowych szeregów czasowych.

Analizy zaprezentowane w tej pracy dotyczyły tylko indeksów giełdowych, które - jako typ - stanowią tylko niewielki wycinek spośród dostępnych instrumentów finansowych. Zatem naturalnym rozszerzeniem zakresu badań byłoby przetestowanie proponowanych modeli na innych typach instrumentów, takich jak przykładowo kursy walut, kryptowaluty, surowce, towary handlowe czy też opcje i kontrakty terminowe. Dodatkowo, metody prezentowane w tej pracy stosować można także w innych dziedzinach, jak choćby prognozowanie cen lub zużycia energii elektrycznej.

Obszar badań poszerzyć można także do analizowania szeregów wielowymiarowych. W tym zakresie sieci uczenia głębokiego wykorzystać można do analizy portfelowej – prób zbudowania strategii inwestycyjnych czy też szacowania ryzyka dla całego portfela. Dodatkowo modele uczenia głębokiego mogłyby być także zbadane w kontekście danych o wysokiej częstotliwości.

Do kluczowych zagadnień, które wymagają dalszych badań jest także usprawnienie procesu dostrajania hiperparametrów oraz kwestie związane z wyborem danych, na

⁵W tym miejscu zwrócić także należy uwagę na pewne problemy, które pojawiły się podczas uczenia sieci z wykorzystaniem funkcji MADL. Ze względu na konstrukcję tej funkcji, słabo sprawdza się w jej przypadku wykorzystanie metod optymalizacyjnych bazujących na gradiencie, przez co uczenie sieci w tym przypadku miało charakter losowy, bardziej zbliżony do metody siatki *grid search*.

których proces dostrajania jest prowadzony. Wciąż trwają prace nad rozwojem narzędzi umożliwiających zautomatyzowanie tego procesu, jednak bazując na doświadczeniu dotyczącym ich stosowania w zaprezentowanych powyżej badaniach, wciąż wymagają one dalszego usprawniania.

Jak wykazano w badaniach przeprowadzonych w tej pracy, ważną kwestią związaną z projektowaniem modeli sieciowych jest dobór odpowiedniej funkcji straty. Proponowane mierniki błędów, o ile sprawdzają się jako narzędzia oceny uzyskanych prognoz, często wymagają doprecyzowania w przypadkach kiedy wykorzystane mają być jako funkcja straty stosowana do rozwiązania konkretnego problemu. Przykładowo, zamiast korzystania z ogólnych mierników błędu prognoz, można skonstruować funkcję straty optymalizującą, przykładowo, wskaźniki informacyjne w kontekście strategii inwestycyjnych lub bezpośrednio poziom prognozowanej wartości zagrożonej podczas szacowania ryzyka kapitałowego.

Sieciowe modele prognoz probabilistycznych, stosowane do prognozowania parametrów zadanego rozkładu prawdopodobieństwa, można w praktyce wykorzystać także do uzyskania prognoz punktowych, jak ma to miejsce przykładowo w przypadku rozkładu normalnego, gdzie wartość parametru μ_{t+1} można traktować jako prognozę punktową stopy zwrotu, a parametru σ_{t+1} — jako punktową prognozę zmienności, pomimo, że w przypadku tych modeli funkcja straty bazuje na zupełnie innej koncepcji, niż w przypadku modeli prognozowania punktowego. Uzasadnione byłoby więc przeprowadzenie badań porównujących wyniki prognoz punktowych uzyskanych za pomocą modeli bazujących na tych dwóch odmiennych podejściach. Z kolei w prognozowaniu całych rozkładów prawdopodobieństwa przyszłych stóp zwrotu Naturalnym kierunkiem naturalnym kierunkiem badań byłoby także wykorzystanie w tym zakresie sieci bayesowskich.

Zaprezentowane w niniejszej pracy modele ewaluowane były przede wszystkim z wykorzystaniem testów wstecznych (z wyjątkiem sieci LSTM, która zastosowana została także do uzyskania prognoz punktowych w czasie rzeczywistym). W celu uzyskania pełniejszego obrazu możliwości prognostycznych proponowanych modeli, należało by je ocenić także w kontekście eksploatacji poprzez budowę zautomatyzowanych algorytmów inwestycyjnych lub szacujących ryzyko w czasie rzeczywistym. Pozwoliłoby to na dokładniejszą ocenę omawianych narzędzi od strony praktycznej.

Zakończenie

Głównym celem pracy była analiza modeli bazujących na sieciach neuronowych uczenia głębokiego w prognozowaniu szeregów czasowych stóp zwrotu wybranych instrumentów finansowych oraz ich zmienności. Cel ten osiągnięty został poprzez szczegółowe przedstawienie zagadnień teoretycznych związanych z poruszaną problematyką omówionych w pierwszej części pracy, w szczególności w rozdziałach 2 i 3, oraz wieloaspektowe badania empiryczne dotyczące możliwości prognostycznych omawianych modeli, zaprezentowane w rozdziałach 5 i 6.

W badaniach przedstawionych w niniejszej rozprawie zastosowane zostało podejście interdyscyplinarne. Szczególny nacisk położony został na połączenie narzędzi ekonometrycznych z metodami z zakresu sztucznej inteligencji. Podjęta została próba syntezy tych metodologii, która zaowocowała propozycjami nowych narzędzi służących do prognozowania finansowych szeregów czasowych, w postaci modeli hybrydowych i sieciowych modeli prognozowania probabilistycznego.

Omawiane w pracy modele przeanalizowane zostały na trzech płaszczyznach. W pierwszej kolejności zbadano możliwości modeli wykorzystujących sieci neuronowe w prognozowaniu punktowym stóp zwrotu. Następnie, uzyskane prognozy wykorzystano do budowy strategii inwestycyjnych. W tym zakresie zaobserwowano przewagę sieci uczenia głębokiego (LSTM oraz CNN) nad klasycznymi sieciami MLP. Zwrócono też uwagę na fakt, że duży wpływ na uzyskiwane wyniki ma także sposób konstrukcji strategii.

W kolejnej części zaproponowane w ramach pracy modele hybrydowe ARMA-GARCH-LSTM posłużyły do punkowego prognozowania zmienności oraz szacowania ryzyka kapitałowego. Wyniki prognoz porównane zostały także z „czystymi” modelami klasy ARMA-GARCH. Początkowa ocena jakości prognoz, bazująca na funkcji MSE, która minimalizowana była przez sieć, wskazywała na poprawę wyników uzyskanych przez specyfikacje rozbudowane o komponent sieciowy. Biorąc jednak pod uwagę dodatkowe mierniki oraz ocenę prognoz ryzyka kapitałowego stwierdzono, że rezultaty uzyskane poprzez poszczególne grupy modeli są zbliżone.

Trzecia część badań empirycznych dotyczyła ewaluacji modeli sieciowych

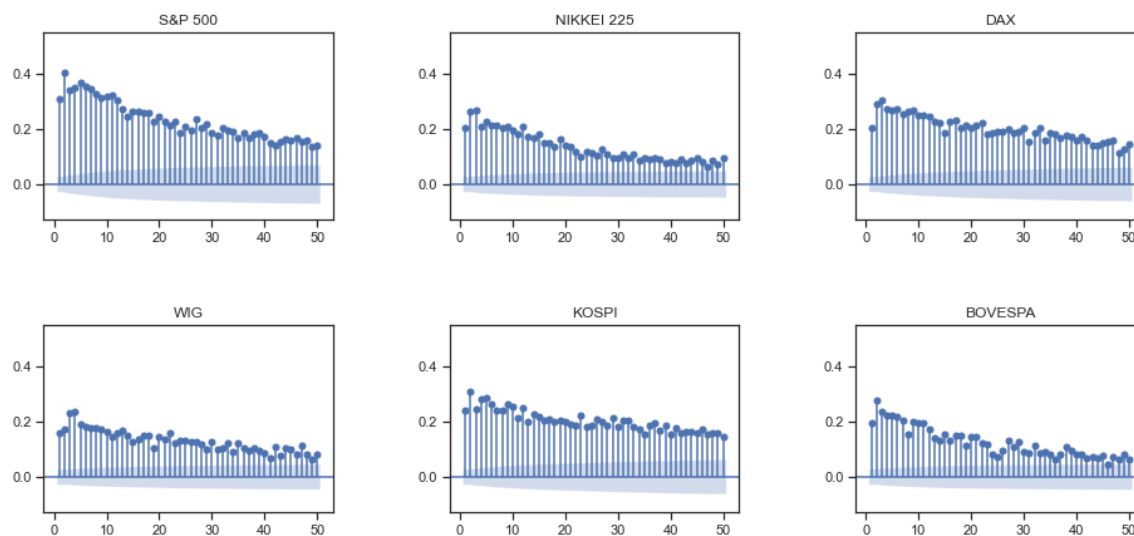
stosowanych do prognozowania parametrów rozkładów prawdopodobieństwa przyszłych stóp zwrotu. W tym celu zastosowane zostały kryteria oceny prognoz probabilistycznych, oraz ponownie, miary trafności prognoz ryzyka kapitałowego. Także w tym zakresie jakość prognoz zależała w dużym stopniu od stosowanego kryterium oceny. Modele sieciowe najlepsze rezultaty uzyskały pod kątem miernika LPS, którego konstrukcja jest najbardziej zbliżona do wykorzystywanych przez sieć funkcji straty.

Wyniki przeprowadzonych badań wskazują na zasadność wykorzystania sieci neuronowych uczenia głębokiego w prognozowaniu finansowych szeregów czasowych, choć nie pokazują jednoznacznej przewagi konkretnej specyfikacji modelowej nad innymi. Ta niejednoznaczność w dużej mierze wynika z kwestii związanych z konstrukcją i wyborem funkcji straty minimalizowanych przez modele sieciowe, co w bezpośredni sposób przekłada się na wyniki uzyskiwane przez dany model. Mając to na uwadze, można jednak stwierdzić, że spośród badanych typów sieci neuronowych, rekurencyjne sieci LSTM regularnie charakteryzowały się dobrymi rezultatami we wszystkich trzech aspektach.

Ze względu na szeroki zakres tematyki związanej z modelowaniem i prognozowaniem finansowych szeregów czasowych, należy wziąć pod uwagę, że metody i narzędzia przedstawione w tej pracy, choć analizowane były w możliwie szerokim zakresie, stanowią jedynie podstawę, na bazie której można dokonywać dalszej ich rozbudowy koncepcyjnej i analizy poprzez badania empiryczne, co pozostawać będzie w obszarze dalszych zainteresowań autora.

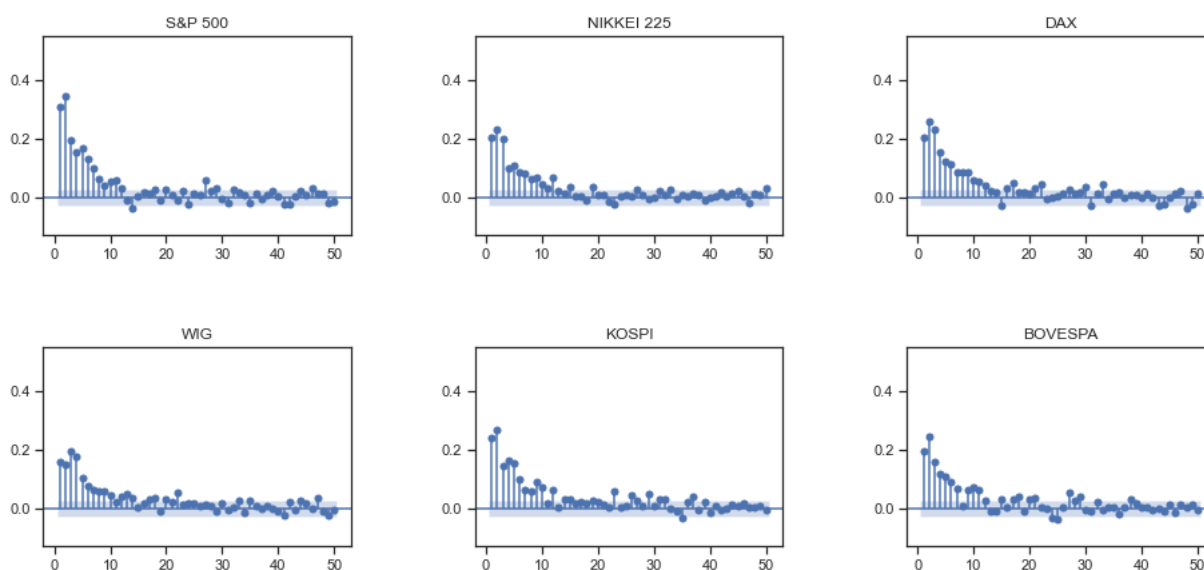
Aneks

Ilustracje



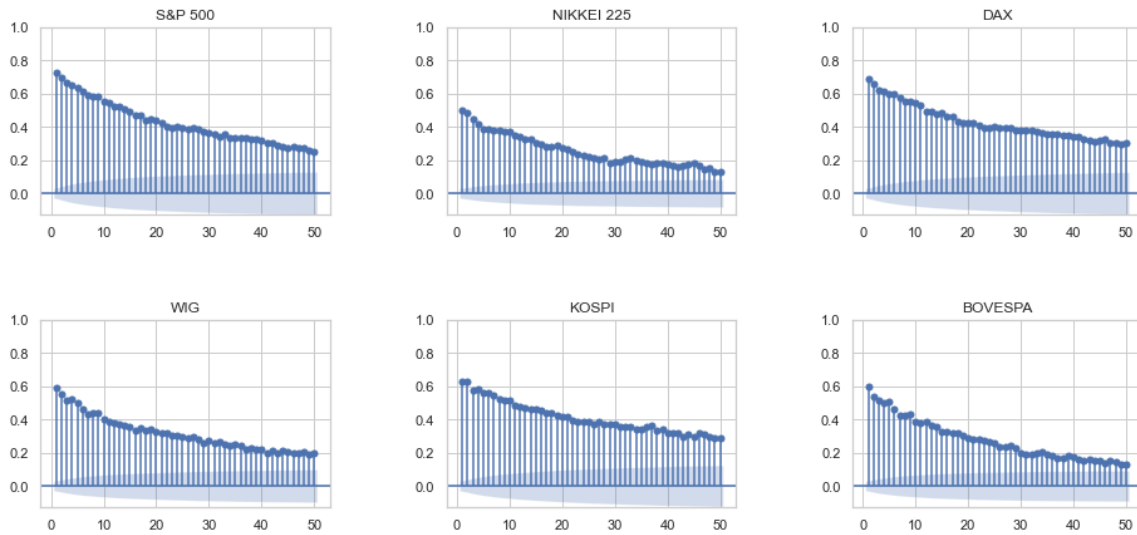
Rysunek A.1: ACF dla wartości bezwzględnych stóp zwrotu

Źródło: Opracowanie własne.



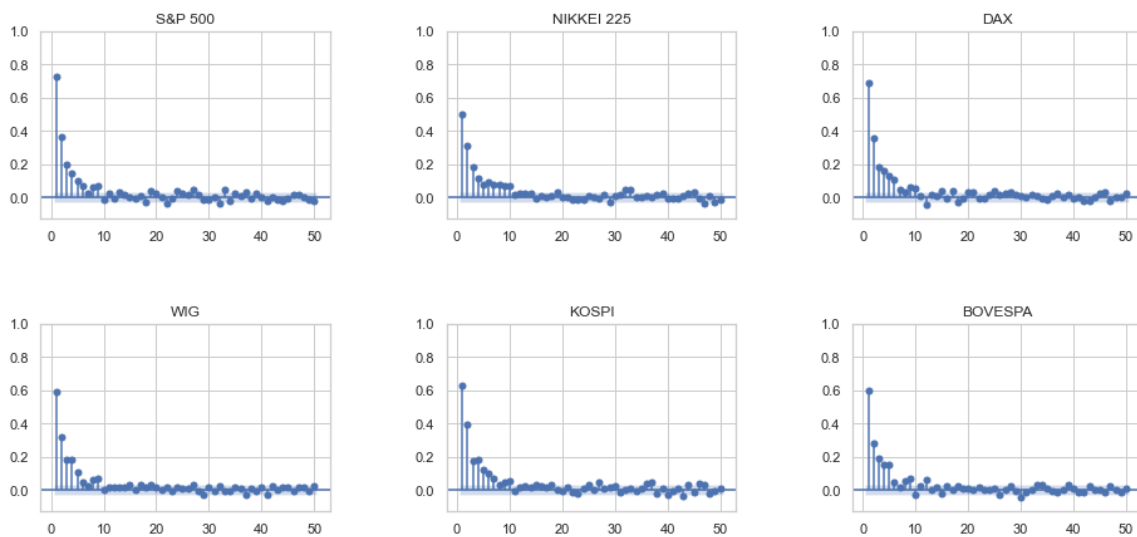
Rysunek A.2: PACF dla wartości bezwzględnych stóp zwrotu

Źródło: Opracowanie własne.



Rysunek A.3: ACF dla oszacowań zmienności uzyskanych za pomocą estymatora GKYZ

Źródło: Opracowanie własne.



Rysunek A.4: PACF dla oszacowań zmienności uzyskanych za pomocą estymatora GKYZ

Źródło: Opracowanie własne.

Wyniki modeli ARMA-GARCH oraz modeli hybrydowych ARMA-GARCH-LSTM

Tabela A.1: Ilość przekroczeń oszacowań VaR(0,05)/VaR(0,01) w modelach ARMA-GARCH i hybrydowych ARMA-GARCH-LSTM

Model/Metrics	G(N)	G(N)-LSTM	G(STD)	G(STD)-LSTM	G(SSTD)	G(SSTD)-LSTM
S&P VaR 5%/1%	158/68 (6,35/2,73)	108/41	174/45	127/28	158/37	105/24
NKX VaR 5%/1%	136/58 (5,47/2,33)	80/26	153/45	90/18	134/37	77/11
WIG VaR 5%/1%	142/46 (5,71/1,85)	98/28	150/33	107/21	141/29	90/19
KOSPI VaR 5%/1%	167/63 (6,71/2,53)	64/16	180/47	79/11	156/36	59/5
BVP VaR 5%/1%	133/36 (5,35/1,45)	68/18	145/29	69/13	137/30	67/16
DAX VaR 5%/1%	174/65 (7/2,61)	68/17	184/40	78/9	171/37	67/7
	E(N)	E(N)-LSTM	E(STD)	E(STD)-LSTM	E(SSTD)	E(SSTD)-LSTM
S&P VaR 5%/1%	163/64 (6,55/2,57)	110/36	179/46	118/33	155/22	98/22
NKX VaR 5%/1%	139/53 (5,59/2,13)	71/20	148/43	83/13	135/28	76/10
WIG VaR 5%/1%	150/49 (6,03/1,97)	83/28	158/41	99/24	144/31	86/16
KOSPI VaR 5%/1%	171/68 (6,88/2,73)	66/17	178/51	74/15	158/29	53/8
BVP VaR 5%/1%	130/42 (5,23/1,69)	60/14	123/19	75/11	119/16	77/12
DAX VaR 5%/1%	159/57 (6,39/2,29)	67/18	171/51	77/15	150/38	60/9
	GJR(N)	GJR(N)-LSTM	GJR(STD)	GJR(STD)-LSTM	GJR(SSTD)	GJR(SSTD)-LSTM
S&P VaR 5%/1%	149/59 (5,99/2,37)	98/30	171/37	112/24	140/26	80/12
NKX VaR 5%/1%	140/58 (5,63/2,33)	76/21	145/43	85/17	134/37	74/11
WIG VaR 5%/1%	136/45 (5,47/1,81)	89/27	147/32	98/21	134/30	87/16
KOSPI VaR 5%/1%	165/57 (6,63/2,29)	61/19	175/46	67/13	155/30	61/10
BVP VaR 5%/1%	128/36 (5,15/1,45)	66/14	135/25	67/13	142/28	69/10
DAX VaR 5%/1%	162/56 (6,51/2,25)	62/19	171/41	67/12	151/33	60/10
	AP(N)	AP(N)-LSTM	AP(STD)	AP(STD)-LSTM	AP(SSTD)	AP(SSTD)-LSTM
S&P VaR 5%/1%	166/71 (6,67/2,85)	111/41	176/41	115/29	148/27	99/18
NKX VaR 5%/1%	135/50 (5,43/2,01)	74/23	140/39	81/14	122/27	75/11
WIG VaR 5%/1%	142/48 (5,71/1,93)	95/29	145/33	96/23	136/31	85/20
KOSPI VaR 5%/1%	172/66 (6,92/2,65)	72/20	179/46	74/13	53/31	65/10
BVP VaR 5%/1%	144/41 (5,79/1,65)	67/16	143/31	79/12	138/33	81/15
DAX VaR 5%/1%	162/59 (6,51/2,37)	63/21	175/47	67/13	151/39	57/10

Uwaga: Oczekiwana liczba przekroczeń wynosi 124/24, ilość prognoz 2487

Źródło: Opracowanie własne.

Tabela A.2: Prognozy zmienności, indeks S&P500, model GARCH i GARCH LSTM

Model/Metrics	G(N)	G(N)-LSTM	G(STD)	G(STD)-LSTM	G(SSTD)	G(SSTD)-LSTM
MSE	0,3285463	0,3280888	0,3251186	0,324028	0,3352583	0,3314607
DM p-value		0,8624		0,8597		0,7724
MAE	0,3652576	0,3712369	0,363694	0,370711	0,3656465	0,3782946
HMSE	0,2156606	0,3453763	0,2237549	0,3471955	0,223676	0,3767537
R^2	0,6341285	0,5696598	0,6186275	0,5822008	0,6143863	0,5668014
Przek. VaR	158/68	108/41	174/45	127/28	158/37	105/24
Przek. VaR (%)	158/68	108/41	174/45	127/28	158/37	105/24
Kupiec 5%	0,002911799(R)	0,1241958	1,559684e-05(R)	0,8080036	0,002911799(R)	0,06766877
Kupiec 1%	7,948087e-13(R)	0,002949403(R)	0,0002689616(R)	0,5363499	0,02262881(R)	0,8600111
Christoff. 5%	0,007515332(R)	0,02411495(R)	6,208225e-05(R)	0,2230072	0,003462283(R)	0,01050691(R)
Christoff. 1%	5,308975e-12(R)	2,53154e-05(R)	0,0006885124(R)	0,01072578(R)	0,02265446(R)	0,005157251(R)
ES bootstrap 5%	0,0001550028	0,0007601218	0,3960251	0,5437578	0,561588	0,7726647
ES próbkowy 5%	3,201136e-07(R)	2,719803e-05(R)	0,3386457	0,534278	0,5917593	0,8535576
ES bootstrap 1%	—	—	—	0,4240474	—	0,6901583
ES próbkowy 1%	—	—	—	0,3757155	—	0,7637902

Źródło: Opracowanie własne.

Tabela A.3: Prognozy zmienności, indeks S&P500, model EGARCH i EGARCH-LSTM

Model/Metrics	E(N)	E(N)-LSTM	E(STD)	E(STD)-LSTM	E(SSTD)	E(SSTD)-LSTM
MSE	0,3575556	0,3518625	0,3517481	0,3448754	0,3632544	0,3605409
DM p-value		0,1617		0,1553		0,1453
MAE	0,3642571	0,3835117	0,3609983	0,37705	0,3635144	0,3819407
HMSE	0,1732228	0,3853719	0,1751933	0,3553598	0,1756505	0,3638078
R^2	0,5729105	0,5488056	0,561027	0,5877998	0,5661337	0,5163117
Przek. VaR	163/64	110/36	179/46	118/33	155/22	98/22
Przek. VaR (%)	163/64	110/36	179/46	118/33	155/22	98/22
Kupiec 5%	0,0006710911(R)	0,1784333	2,227363e-06(R)	0,5558262	0,006522105(R)	0,01194658(R)
Kupiec 1%	4,569256e-11(R)	0,03551858(R)	0,0001401641(R)	0,1186792	0,5552061	0,5552061
Christoff. 5%	0,002998533(R)	0,01632977(R)	1,302748e-05(R)	0,0369363(R)	0,01829748(R)	0,002670763(R)
Christoff. 1%	2,395822e-10(R)	2,545875e-07(R)	0,0001252356(R)	0,000998233(R)	0,355113	0,0001042493(R)
ES bootstrap 5%	3,307416e-05	0,007496304	0,174615	0,2261359	0,6434566	0,5576949
ES próbkowy 5%	2,938924e-07(R)	0,0007393309(R)	0,09533645	0,1474097	0,6735219	0,5641093
ES bootstrap 1%	—	—	—	0,6758302	0,1517443	0,6403693
ES próbkowy 1%	—	—	—	0,70159	0,08271212	0,6664602

Źródło: Opracowanie własne.

Tabela A.4: Prognozy zmienności, indeks S&P500, model GJR-GARCH i GJR-GARCH-LSTM

Model/Metrics	GJR(N)	GJR(N)-LSTM	GJR(STD)	GJR(STD)-LSTM	GJR(SSTD)	GJR(SSTD)-LSTM
MSE	0,302625	0,2914628	0,3012009	0,2809567	0,3063293	0,2946242
DM p-value		0,1478		0,1521		0,1521
MAE	0,3519807	0,3741351	0,3516875	0,3631845	0,3533803	0,3788403
HMSE	0,1807714	0,3842018	0,183153	0,3273964	0,1848005	0,4225433
R^2	0,6068593	0,7384955	0,5981365	0,7677426	0,6015356	0,7731989
Przek. VaR	149/59	98/30	171/37	112/24	140/26	80/12
Przek. VaR (%)						
Kupiec 5%	0,02764251(R)	0,01194658(R)	4,674136e-05(R)	0,2481387	0,1577436	1,337974e-05(R)
Kupiec 1%	5,095557e-09(R)	0,3166331	0,02262881(R)	0,8600111	0,821158	0,003927525(R)
Christoff. 5%	0,08359927	0,002670763(R)	0,0002338338(R)	0,1259191	0,274063	4,477951e-06(R)
Christoff. 1%	3,589397e-08(R)	0,0116681(R)	0,06384671	0,005157251(R)	0,5376108	0,01474698(R)
ES bootstrap 5%	0,0008605812	0,08362556	0,5690876	0,6861423	0,6974109	0,8121736
ES próbkowy 5%	1,545614e-05(R)	0,02975946(R)	0,589491	0,7224967	0,7623357	0,8867095
ES bootstrap 1%	—	0,0310762	0,5853297	0,4358049	0,3789528	—
ES próbkowy 1%	—	0,01149588(R)	0,5632319	0,3510686	0,3088883	—

Źródło: Opracowanie własne.

Tabela A.5: Prognozy zmienności, indeks S&P500, model APARCH i APARCH-LSTM

Model/Metrics	AP(N)	AP(N)-LSTM	AP(STD)	AP(STD)-LSTM	AP(SSTD)	AP(SSTD)-LSTM
MSE	0,3426945	0,3344834	0,3242661	0,318056	0,3264551	0,3162408
DM p-value		0,1849		0,204		0,1616
MAE	0,3606189	0,3699534	0,3544302	0,3627978	0,3531873	0,3610026
HMSE	0,1685517	0,3312288	0,1715618	0,3185091	0,1670245	0,3035204
R^2	0,5456934	0,5692383	0,5582577	0,611704	0,5676243	0,5926861
Przek. VaR	166/71	111/41	176/41	115/29	148/27	99/18
Przek. VaR (%)						
Kupiec 5%	0,0002585233(R)	0,2112618	7,28635e-06(R)	0,3838599	0,0344(R)	0,01574687(R)
Kupiec 1%	3,264056e-14(R)	0,002949403(R)	0,002949403(R)	0,4173806	0,6719996	0,1453207
Christoff. 5%	0,0006112344(R)	0,1036547	4,226065e-05(R)	0,04996012(R)	0,03931851(R)	0,003824839(R)
Christoff. 1%	5,928591e-14(R)	2,410587e-06(R)	0,004948776(R)	0,0114341(R)	0,5333337	0,1031526
ES bootstrap 5%	6,28606e-06	0,002173954	0,242324	0,3242384	0,6982161	0,6809137
ES próbkowy 5%	2,985549e-08(R)	0,000195468(R)	0,1541806	0,2350046	0,7501902	0,7400651
ES bootstrap 1%	—	—	—	0,5986175	0,3858598	0,4302824
ES próbkowy 1%	—	—	—	0,6280052	0,307657	0,3673147

Źródło: Opracowanie własne.

Tabela A.6: Prognozy zmienności, indeks NIKKEI 225, model GARCH i GARCH-LSTM

Model/Metrics	G(N)	G(N)-LSTM	G(STD)	G(STD)-LSTM	G(SSTD)	G(SSTD)-LSTM
MSE	0,7190396	0,6295932	0,7156355	0,6413457	0,7232306	0,6445305
DM p-value		0,0001895		0,4401		0,3759
MAE	0,5676317	0,5713175	0,5689244	0,5740221	0,571226	0,5768095
HMSE	2,850735	4,558046	2,988673	4,457486	2,84862	4,60183
R^2	0,1691152	0,1733029	0,1691473	0,1571037	0,1664218	0,14434
Przek. VaR	136/58	80/26	153/45	90/18	134/37	77/11
Przek. VaR (%)						
Kupiec 5%	0,2906901	1,337974e-05(R)	0,01082282(R)	0,0009088047(R)	0,3803133	2,977075e-06(R)
Kupiec 1%	1,245618e-08(R)	0,821158	0,0002689616(R)	0,1453207	0,02262881(R)	0,001678938(R)
Christoff. 5%	0,5591795	1,255502e-05(R)	0,0381103(R)	0,001483662(R)	0,6504183	5,581325e-06(R)
Christoff. 1%	4,149255e-08(R)	0,5376108	0,0002087952(R)	0,1031526	0,004497022(R)	0,000831569(R)
ES bootstrap 5%	8,991041e-06	0,02218931	0,07341403	0,5331799	0,09595087	0,6353035
ES próbkowy 5%	3,819098e-07(R)	0,006081351(R)	0,02829267(R)	0,4774846	0,03660707(R)	0,6201986
ES bootstrap 1%	—	0,04672565	—	0,4804514	—	—
ES próbkowy 1%	—	0,01821063(R)	—	0,4009626	—	—

Źródło: Opracowanie własne.

Tabela A.7: Prognozy zmienności, indeks NIKKEI 225, model EGARCH i EGARCH-LSTM

Model/Metrics	E(N)	E(N)-LSTM	E(STD)	E(STD)-LSTM	E(SSTD)	E(SSTD)-LSTM
MSE	0,7293213	0,6421487	0,7185632	0,6292921	0,718122	0,6454369
DM p-value		0,5743		0,509		0,5842
MAE	0,5737363	0,571063	0,5684049	0,5657533	0,5669657	0,5696178
HMSE	2,667138	4,72994	2,687264	4,619211	2,616769	4,430363
R^2	0,2060841	0,1326951	0,2125535	0,175777	0,2064342	0,124935
Przek. VaR	139/53	71/20	148/43	83/13	135/28	76/10
Przek. VaR (%)						
Kupiec 5%	0,1855083	1,030394e-07(R)	0,0344(R)	5,371088e-05(R)	0,3335094	1,757875e-06(R)
Kupiec 1%	8,393498e-07(R)	0,3095893	0,0009299509(R)	0,008473333(R)	0,5363499	0,0006566578(R)
Christoff. 5%	0,3750778	5,71622e-07(R)	0,1064974	0,000130801(R)	0,6060121	3,122448e-06(R)
Christoff. 1%	4,039331e-06(R)	0,2143328	0,003996211(R)	0,005037558(R)	0,5077687	0,0002877767(R)
ES bootstrap 5%	0,0003025123	0,04243671	0,02555243	0,4185492	0,1609098	0,6743024
ES próbkowy 5%	1,929293e-05(R)	0,01453429(R)	0,006372951(R)	0,3478262	0,08017253	0,7029006
ES bootstrap 1%	—	0,02279984	—	—	0,2118688	—
ES próbkowy 1%	—	0,0111404(R)	—	—	0,1254408	—

Źródło: Opracowanie własne.

Tabela A.8: Prognozy zmienności, indeks NIKKEI 225, model GJR-GARCH i GJR-GARCH-LSTM

Model/Metrics	GJR(N)	GJR(N)-LSTM	GJR(STD)	GJR(STD)-LSTM	GJR(SSTD)	GJR(SSTD)-LSTM
MSE	0,7042808	0,6496123	0,6900629	0,6344323	0,6961625	0,6277231
DM p-value		0,7503		0,7358		0,6868
MAE	0,5644395	0,5706564	0,5603332	0,5675919	0,5620999	0,5669418
HMSE	2,886685	4,5297	2,958561	4,473554	2,634196	4,449538
R^2	0,197785	0,113094	0,2090007	0,160146	0,2040482	0,1698732
Przek. VaR	140/58	76/21	145/43	85/17	134/37	74/11
Przek. VaR (%)						
Kupiec 5%	0,1577436	1,757875e-06(R)	0,06383829	0,0001276986(R)	0,3803133	5,888947e-07(R)
Kupiec 1%	1,245618e-08(R)	0,4229126	0,0009299509(R)	0,09252608	0,02262881(R)	0,001678938(R)
Christoff. 5%	0,3480243	6,414211e-06(R)	0,1546133	0,0003315829(R)	0,6023921	2,026285e-06(R)
Chrisotff. 1%	7,936179e-08(R)	0,2833833	0,001947491(R)	0,06519099	0,06384671	0,000831569(R)
ES bootstrap 5%	0,0006524609	0,05506156	0,0622781	0,4463583	0,2221808	0,6550945
ES próbkowy 5%	4,228499e-05(R)	0,02204104(R)	0,01948298(R)	0,3895219	0,1306741	0,6805171
ES bootstrap 1%	—	0,03188525	—	0,401404	—	—
ES próbkowy 1%	—	0,01836927(R)	—	0,3184132	—	—

Źródło: Opracowanie własne.

Tabela A.9: Prognozy zmienności, indeks NIKKEI 225, model APARCH i APARCH-LSTM

Model/Metrics	AP(N)	AP(N)-LSTM	AP(STD)	AP(STD)-LSTM	AP(SSTD)	AP(SSTD)-LSTM
MSE	0,725722	0,6385541	0,7014102	0,6340941	0,7013369	0,6347263
DM p-value		0,6452		0,69		0,7182
MAE	0,5717364	0,5658564	0,5641562	0,5677007	0,5644852	0,566896
HMSE	2,788908	4,645786	2,817499	4,70593	2,712554	4,604738
R^2	0,1649554	0,1567614	0,194411	0,1519926	0,1997656	0,1436635
Przek. VaR	135/50	74/23	140/39	81/14	122/27	75/11
Przek. VaR (%)						
Kupiec 5%	0,3335094	5,888947e-07(R)	0,1577436	2,152801e-05(R)	0,8283143	1,024309e-06(R)
Kupiec 1%	8,449136e-06(R)	0,7026887	0,008555221(R)	0,01697476(R)	0,6719996	0,001678938(R)
Christoff. 5%	0,6212114	2,026285e-06(R)	0,3385526	4,861257e-05(R)	0,8924112	3,634332e-06(R)
Chrisotff. 1%	4,934183e-05(R)	0,4228579	0,02833089(R)	0,01076814(R)	0,5333337	0,000831569(R)
ES bootstrap 5%	0,001917887	0,04622795	0,1655853	0,4697999	0,2384024	0,6936294
ES próbkowy 5%	0,0001867613(R)	0,01756195(R)	0,08321602	0,4091589	0,1476266	0,7310164
ES bootstrap 1%	—	0,03978208	—	—	0,2314639	—
ES próbkowy 1%	—	0,01846625(R)	—	—	0,1444633	—

Źródło: Opracowanie własne.

Tabela A.10: Prognozy zmienności, indeks DAX, model GARCH i GARCH-LSTM

Model/Metrics	G(N)	G(N)-LSTM	G(STD)	G(STD)-LSTM	G(SSTD)	G(SSTD)-LSTM
MSE	1,069736	0,7849515	1,051762	0,7771242	1,075352	0,7910381
DM p-value		0,002308		0,001331		0,0007548
MAE	0,6289005	0,5624794	0,6203608	0,5560515	0,6246122	0,5605514
HMSE	0,1854313	0,3559361	0,1831631	0,340455	0,1836664	0,3445947
R^2	0,3513935	0,3170056	0,3469804	0,3242	0,3267928	0,3125383
Przek. VaR	174/65	68/17	184/40	78/9	171/37	67/7
Przek. VaR (%)						
Kupiec 5%	1,559684e-05(R)	1,586138e-08(R)	2,754904e-07(R)	4,976404e-06(R)	4,674136e-05(R)	8,253918e-09(R)
Kupiec 1%	1,697487e-11(R)	0,09252608	0,005080548(R)	0,0002327628(R)	0,02262881(R)	2,073181e-05(R)
Christoff. 5%	7,284972e-06(R)	8,449678e-08(R)	2,363876e-07(R)	2,83897e-05(R)	1,486731e-05(R)	6,066743e-08(R)
Chrisotff. 1%	4,364076e-11(R)	0,2161062	0,01026075(R)	0,001107336(R)	0,04253562(R)	0,0001138994(R)
ES bootstrap 5%	0,0001373883	0,1848147	0,2414165	0,9135146	0,464242	0,9517954
ES próbkowy 5%	8,476633e-06(R)	0,1060256	0,1595951	0,9731608	0,4183047	0,9933471
ES bootstrap 1%	—	0,213894	—	—	—	—
ES próbkowy 1%	—	0,1262211	—	—	—	—

Źródło: Opracowanie własne.

Tabela A.11: Prognozy zmienności, indeks DAX, model EGARCH i EGARCH-LSTM

Model/Metrics	E(N)	E(N)-LSTM	E(STD)	E(STD)-LSTM	E(SSTD)	E(SSTD)-LSTM
MSE	1,214811	0,8233499	1,196003	0,8740355	1,19496	0,8504577
DM p-value		0,0009276		0,001369		0,001283
MAE	0,6577704	0,5625484	0,6427355	0,5718593	0,6437478	0,5654752
HMSE	0,1719085	0,3506878	0,1695586	0,3428184	0,1699276	0,3554711
R^2	0,2859736	0,342667	0,2403574	0,3205843	0,2422677	0,3380977
Przek. VaR	159/57	67/18	171/51	77/15	150/38	60/9
Przek. VaR (%)						
Kupiec 5%	0,002198025(R)	8,253918e-09(R)	4,674136e-05(R)	2,977075e-06(R)	0,02207304(R)	5,51249e-11(R)
Kupiec 1%	2,993902e-08(R)	0,1453207	3,98726e-06(R)	0,03176469(R)	0,0140778(R)	0,0002327628(R)
Christoff. 5%	0,007694525(R)	2,118517e-08(R)	0,0001991965(R)	7,274061e-07(R)	0,05817807	2,357327e-10(R)
Chrisotff. 1%	9,139484e-08(R)	0,3037263	1,676681e-05(R)	0,09102395	0,04315141(R)	0,001107336(R)
ES bootstrap 5%	1,195145e-05	0,1550033	0,0004296552	0,1640598	0,2602961	0,7191898
ES próbkowy 5%	9,584718e-07(R)	0,08640518	4,446945e-05(R)	0,08472213	0,1686346	0,7487983
ES bootstrap 1%	—	0,08146706	—	0,3471822	—	—
ES próbkowy 1%	—	0,04933472(R)	—	0,240651	—	—

Źródło: Opracowanie własne.

Tabela A.12: Prognozy zmienności, indeks DAX, model GJR-GARCH i GJR-GARCH-LSTM

Model/Metrics	GJR(N)	GJR(N)-LSTM	GJR(STD)	GJR(STD)-LSTM	GJR(SSTD)	GJR(SSTD)-LSTM
MSE	1,018062	0,7592761	0,9778356	0,7556117	0,9762771	0,7446205
DM p-value		0,0151		0,01883		0,009107
MAE	0,6204018	0,551187	0,6046727	0,5491992	0,6033362	0,5396463
HMSE	0,1662555	0,3420855	0,1643428	0,3494785	0,1640458	0,3278431
R^2	0,3871644	0,3670471	0,3966749	0,3592803	0,403628	0,3764779
Przek. VaR	162/56	62/19	171/41	67/12	151/33	60/10
Przek. VaR (%)						
Kupiec 5%	0,0009111476(R)	2,499176e-10(R)	4,674136e-05(R)	8,253918e-09(R)	0,0175152(R)	5,51249e-11(R)
Kupiec 1%	7,073541e-08(R)	0,2170907	0,002949403(R)	0,003927525(R)	0,1186792	0,0006566578(R)
Christoff. 5%	0,003015094(R)	1,902654e-09(R)	0,0001991965(R)	4,316764e-08(R)	0,04891764(R)	4,19351e-10(R)
Christoff. 1%	4,784607e-07(R)	0,4032968	0,01122788(R)	0,01474698(R)	0,1898935	0,002895986(R)
ES bootstrap 5%	0,000112717	0,09519169	0,2659333	0,6875935	0,6054211	0,7795719
ES próbkowy 5%	6,030285e-06(R)	0,04682177(R)	0,1829912	0,7082645	0,6473815	0,8633752
ES bootstrap 1%	—	0,1365961	—	—	0,6832069	—
ES próbkowy 1%	—	0,0700981	—	—	0,7042714	—

Źródło: Opracowanie własne.

Tabela A.13: Prognozy zmienności, indeks DAX, model APARCH i APARCH-LSTM

Model/Metrics	AP(N)	AP(N)-LSTM	AP(STD)	AP(STD)-LSTM	AP(SSTD)	AP(SSTD)-LSTM
MSE	1,080154	0,7755289	1,03333	0,7596108	1,048637	0,7690159
DM p-value		0,00331		0,007559		0,005758
MAE	0,6322617	0,5483295	0,616666	0,546451	0,6158596	0,5453936
HMSE	0,163826	0,3299685	0,161454	0,3395394	0,1606082	0,3411371
R^2	0,3900843	0,3679386	0,3932616	0,3684221	0,3722437	0,3481467
Przek. VaR	162/59	63/21	175/47	67/13	151/39	57/10
Przek. VaR (%)						
Kupiec 5%	0,0009111476(R)	5,191794e-10(R)	1,069158e-05(R)	8,253918e-09	0,0175152(R)	5,027312e-12(R)
Kupiec 1%	5,095557e-09(R)	0,4229126	7,156069e-05(R)	0,008473333(R)	0,008555221(R)	0,0006566578(R)
Christoff. 5%	0,002264429(R)	2,459031e-09(R)	5,436756e-05(R)	4,316764e-08(R)	0,03771531(R)	3,757084e-11(R)
Christoff. 1%	3,403486e-08(R)	0,6065243	0,0003738905(R)	0,02919851(R)	0,02833089(R)	0,002895986(R)
ES bootstrap 5%	8,81916e-05	0,06353271	0,1944392	0,5367139	0,4903197	0,6874604
ES próbkowy 5%	5,70966e-06	0,03009668	0,1071196	0,5193546	0,4539598	0,7172347
ES bootstrap 1%	—	0,1535328	—	—	—	—
ES próbkowy 1%	—	0,08743779	—	—	—	—

Źródło: Opracowanie własne.

Tabela A.14: Prognozy zmienności, indeks WIG, model GARCH i GARCH-LSTM

Model/Metrics	G(N)	G(N)-LSTM	G(STD)	G(STD)-LSTM	G(SSTD)	G(SSTD)-LSTM
MSE	0,5016779	0,4807461	0,4898187	0,4648025	0,4919046	0,4899494
DM p-value		0,2081		0,00342		0,3169
MAE	0,3971172	0,4007828	0,3892141	0,3971894	0,3901081	0,404276
HMSE	0,159309	0,2495876	0,1512678	0,2510024	0,1506661	0,2596854
R^2	0,2438276	0,2024343	0,2395174	0,2168488	0,238097	0,1963516
Przek. VaR	142/46	98/28	150/33	107/21	141/29	90/19
Przek. VaR (%)						
Kupiec 5%	0,1119399	0,01194658(R)	0,02207304(R)	0,1023182	0,1332994	0,0009088047(R)
Kupiec 1%	0,0001401641(R)	0,5363499	0,1186792	0,4229126	0,4173806	0,2170907
Christoff. 5%	0,008804667(R)	0,006510383(R)	0,002855768(R)	0,2129693	0,0193541(R)	9,135134e-05(R)
Christoff. 1%	2,704862e-05(R)	0,5077687	0,2252329	0,2833833	0,4646559	0,1531993
ES bootstrap 5%	6,965654e-05	0,003694259	0,02497194	0,3177466	0,04674462	0,2333109
ES próbkowy 5%	3,788183e-06(R)	0,0003317787(R)	0,005185661(R)	0,2363042	0,01537991(R)	0,1491645
ES bootstrap 1%	—	0,0005649516	0,01611454	0,3218875	0,009718938	0,1384656
ES próbkowy 1%	—	3,453853e-05(R)	0,002733272(R)	0,2975999	0,001389077	0,06825786

Źródło: Opracowanie własne.

Tabela A.15: Prognozy zmienności, indeks WIG, model EGARCH i EGARCH-LSTM

Model/Metrics	E(N)	E(N)-LSTM	E(STD)	E(STD)-LSTM	E(SSTD)	E(SSTD)-LSTM
MSE	0,5110341	0,4883024	0,5080994	0,4868426	0,5032369	0,4992776
DM p-value		0,3057		0,2132		0,2438
MAE	0,4009833	0,4030376	0,4020795	0,4053615	0,4008149	0,4059339
HMSE	0,1465773	0,2628319	0,147502	0,2540669	0,1486358	0,2608203
R^2	0,2842135	0,189866	0,2724925	0,2004278	0,2428964	0,1927532
Przek. VaR	150/49	83/28	158/41	99/24	144/31	86/16
Przek. VaR (%)						
Kupiec 5%	0,02207304(R)	5,371088e-05(R)	0,002911799(R)	0,01574687(R)	0,07746676	0,0001934506(R)
Kupiec 1%	1,756379e-05(R)	0,5363499	0,002949403(R)	0,8600111	0,2341585	0,0558152
Christoff. 5%	0,01098066(R)	7,786051e-06(R)	0,005324692(R)	0,009110769(R)	0,05870036	0,0002627241(R)
Christoff. 1%	2,317021e-05(R)	0,5077687	0,004948776(R)	0,4796015	0,08470451	0,03851254
ES bootstrap 5%	0,0001821562	0,001073194	0,02395514	0,1064	0,08013354	0,3480869
ES próbkowy 5%	1,398831e-05(R)	9,564629e-05(R)	0,004593311(R)	0,04680702(R)	0,03123847(R)	0,2800814
ES bootstrap 1%	—	0,0008099588	—	0,1890437	0,05412765	0,1126649
ES próbkowy 1%	—	6,40275e-05(R)	—	0,1106635	0,01295855(R)	0,05013207

Źródło: Opracowanie własne.

Tabela A.16: Prognozy zmienności, indeks WIG, model GJR-GARCH i GJR-GARCH-LSTM

Model/Metrics	GJR(N)	GJR(N)-LSTM	GJR(STD)	GJR(STD)-LSTM	GJR(SSTD)	GJR(SSTD)-LSTM
MSE	0,480196	0,4319645	0,4727176	0,4331499	0,4729606	0,4335862
DM p-value		0,006797		0,009753		0,01336
MAE	0,3904561	0,3877966	0,3870368	0,3887263	0,3873296	0,3894942
HMSE	0,1470687	0,245685	0,1467416	0,2391514	0,1476416	0,243874
R^2	0,2643956	0,2663223	0,260102	0,2653745	0,2536893	0,2634236
Przek. VaR	136/45	89/27	147/32	98/21	134/30	87/16
Przek. VaR (%)						
Kupiec 5%	0,2906901	0,0006277842(R)	0,04254057(R)	0,01194658(R)	0,3803133	0,0002896842(R)
Kupiec 1%	0,0002689616(R)	0,6719996	0,1688211	0,4229126	0,3166331	0,0558152
Christoff. 5%	0,1491261	0,0009832874(R)	0,0263967(R)	0,01368701(R)	0,2588073	0,0007930645(R)
Christoff. 1%	0,001283738(R)	0,6796999	0,2556413	0,6065243	0,4198489	0,1447965
ES bootstrap 5%	0,0001293162	0,01579327	0,04979409	0,2571328	0,1011758	0,4251651
ES próbkowy 5%	1,047157e-05(R)	0,002703568(R)	0,01810339(R)	0,1709857	0,03751224	0,3802523
ES bootstrap 1%	—	0,001002251	0,01049462	0,3490302	0,03524295	0,1406133
ES próbkowy 1%	—	0,000104181(R)	0,001351261(R)	0,3071385	0,008806436(R)	0,08105896

Źródło: Opracowanie własne.

Tabela A.17: Prognozy zmienności, indeks WIG, model APARCH i APARCH-LSTM

Model/Metrics	AP(N)	AP(N)-LSTM	AP(STD)	AP(STD)-LSTM	AP(SSTD)	AP(SSTD)-LSTM
MSE	0,4951572	0,4778275	0,4825994	0,4723306	0,5062653	0,4659261
DM p-value		0,3771		0,2187		0,09545
MAE	0,3978857	0,3957532	0,390865	0,4014725	0,3932025	0,3988304
HMSE	0,150432	0,2455245	0,1467749	0,2646203	0,1752026	0,2662842
R^2	0,2545484	0,2061789	0,2517307	0,2072169	0,162094	0,2169121
Przek. VaR	142/48	95/29	145/33	96/23	136/31	85/20
Przek. VaR (%)						
Kupiec 5%	0,1119399	0,004914388(R)	0,06383829	0,006674846(R)	0,2906901	0,0001276986(R)
Kupiec 1%	3,580577e-05(R)	0,4173806	0,1186792	0,7026887	0,2341585	0,3095893
Christoff. 5%	0,06768427	0,0002796864(R)	0,03123931(R)	0,003193499(R)	0,1491261	0,0001643573(R)
Christoff. 1%	0,0001196566(R)	0,4646559	0,1898935	0,4228579	0,3331484	0,2143328
ES bootstrap 5%	0,0002196592	0,007406649	0,03522117	0,2679577	0,05801245	0,3727497
ES próbkowy 5%	1,353553e-05(R)	0,001012632(R)	0,006961378(R)	0,1756529	0,01795494(R)	0,3007678
ES bootstrap 1%	—	0,00128344	0,005429643	0,2118191	0,01828824	0,1518267
ES próbkowy 1%	—	4,851239e-05(R)	0,0003618305(R)	0,1263483	0,003464845(R)	0,08199591

Źródło: Opracowanie własne.

Tabela A.18: Prognozy zmienności, indeks KOSPI, model GARCH i GARCH-LSTM

Model/Metrics	G(N)	G(N)-LSTM	G(STD)	G(STD)-LSTM	G(SSTD)	G(SSTD)-LSTM
MSE	0,7378322	0,553092	0,7188387	0,5574311	0,7388017	0,559372
DM p-value		0,007025		0,01168		0,008227
MAE	0,5041342	0,4498382	0,4950363	0,450868	0,5003664	0,4527608
HMSE	0,1486261	0,2673304	0,1465225	0,2714647	0,1473925	0,2759781
R^2	0,2754385	0,2566355	0,2819561	0,252989	0,262754	0,249279
Przek. VaR	167/63	64/16	180/47	79/11	156/36	59/5
Przek. VaR (%)						
Kupiec 5%	0,000185844(R)	1,061341e-09(R)	1,483234e-06	8,211886e-06	0,005015865(R)	2,524525e-11(R)
Kupiec 1%	1,211037e-10(R)	0,0558152	7,156069e-05(R)	0,001678938(R)	0,03551858(R)	1,037271e-06(R)
Christoff. 5%	1,42488e-05(R)	5,173137e-09(R)	1,958115e-07	3,207632e-05	0,0003327804(R)	1,913443e-10(R)
Christoff. 1%	8,738547e-10(R)	0,1447965	0,0003738905(R)	0,006843567(R)	0,09169552	6,531212e-06(R)
ES bootstrap 5%	2,999049e-06	0,2047013	0,07023574	0,9739512	0,1828492	0,9879942
ES próbkowy 5%	1,440762e-08(R)	0,1119625	0,02321616(R)	0,9974071	0,1093607	0,9995317
ES bootstrap 1%	—	0,391375	—	—	0,8428028	—
ES próbkowy 1%	—	0,3104891	—	—	0,9125288	—

Źródło: Opracowanie własne.

Tabela A.19: Prognozy zmienności, indeks KOSPI, model EGARCH i EGARCH-LSTM

Model/Metrics	E(N)	E(N)-LSTM	E(STD)	E(STD)-LSTM	E(SSTD)	E(SSTD)-LSTM
MSE	0,7895534	0,5861569	0,7751247	0,5751065	0,7710203	0,5626108
DM p-value		0,01117		0,004848		0,003027
MAE	0,5224552	0,4577732	0,5150387	0,4656652	0,5142949	0,4605595
HMSE	0,1493635	0,2748749	0,1470675	0,2939542	0,1467586	0,293472
R^2	0,3587522	0,2984503	0,3502296	0,3046175	0,3549323	0,3079475
Przek. VaR	171/68	66/17	178/51	74/15	158/29	53/8
Przek. VaR (%)						
Kupiec 5%	4,674136e-05(R)	4,230779e-09(R)	3,325637e-06(R)	5,888947e-07(R)	0,002911799(R)	1,606493e-13(R)
Kupiec 1%	7,948087e-13(R)	0,09252608	3,98726e-06(R)	0,03176469(R)	0,4173806	7,391891e-05(R)
Christoff. 5%	7,013816e-05(R)	1,038133e-09(R)	6,054885e-06(R)	2,026285e-06(R)	0,003215756(R)	1,134759e-12(R)
Christoff. 1%	2,690514e-12(R)	0,2161062	6,654995e-06(R)	0,09102395	0,5111283	0,0003783371(R)
ES bootstrap 5%	1,28053e-05	0,2098873	0,02930189	0,8671703	0,3388093	0,9264903
ES próbkowy 5%	2,839617e-07(R)	0,117655	0,005943415(R)	0,9436366	0,2729836	0,9824801
ES bootstrap 1%	—	0,04334628	—	0,8574863	0,2039617	—
ES próbkowy 1%	—	0,01240531(R)	—	0,9381089	0,1226525	—

Źródło: Opracowanie własne.

Tabela A.20: Prognozy zmienności, indeks KOSPI, model GJR-GARCH i GJR-GARCH-LSTM

Model/Metrics	GJR(N)	GJR(N)-LSTM	GJR(STD)	GJR(STD)-LSTM	GJR(SSTD)	GJR(SSTD)-LSTM
MSE	0,7034035	0,5402718	0,6912846	0,553268	0,7065312	0,5602887
DM p-value		0,02455		0,05489		0,0555
MAE	0,5023142	0,4447278	0,4944834	0,4506149	0,4973971	0,4444016
HMSE	0,1421074	0,2522675	0,1401998	0,2670855	0,1404988	0,241665
R^2	0,4039474	0,3128245	0,3982769	0,2954005	0,3884698	0,2837887
Przek. VaR	165/57	61/19	175/46	67/13	155/30	61/10
Przek. VaR (%)						
Kupiec 5%	0,0003574606(R)	1,183531e-10(R)	1,069158e-05(R)	8,253918e-09(R)	0,006522105(R)	1,183531e-10(R)
Kupiec 1%	2,993902e-08(R)	0,2170907	0,0001401641(R)	0,008473333(R)	0,3166331	0,0006566578(R)
Christoff. 5%	0,0002933049(R)	2,102257e-10(R)	1,477245e-05(R)	4,316764e-08(R)	0,01324942(R)	2,102257e-10(R)
Christoff. 1%	1,814355e-07(R)	0,4032968	0,0003938719(R)	0,02919851(R)	0,4198489	0,002895986(R)
ES bootstrap 5%	0,6929617	0,9603083	0,1267458	0,7745528	0,4490631	0,8761657
ES próbkowy 5%	0,7371838	0,9936208	0,04995367(R)	0,8262102	0,4041286(R)	0,9525879
ES bootstrap 1%	—	0,1595358	—	—	0,6235896	—
ES próbkowy 1%	—	0,07610875	—	—	0,6472259	—

Źródło: Opracowanie własne.

Tabela A.21: Prognozy zmienności, indeks KOSPI, model APARCH i APARCH-LSTM

Model/Metrics	AP(N)	AP(N)-LSTM	AP(STD)	AP(STD)-LSTM	AP(SSTD)	AP(SSTD)-LSTM
MSE	0,7496204	0,5317775	0,7233447	0,5511946	0,7324045	0,5664421
DM p-value		0,00115		0,005486		0,01627
MAE	0,5138654	0,4403128	0,5016307	0,454489	0,5024283	0,453435
HMSE	0,1444482	0,237207	0,1411206	0,2727843	0,1407051	0,2578569
R^2	0,4013762	0,3247839	0,4044557	0,2982872	0,3991118	0,2823087
Przek. VaR	172/66	72/20	179/46	74/13	153/31	65/10
Przek. VaR (%)						
Kupiec 5%	3,26119e-05(R)	1,868141e-07(R)	2,227363e-06(R)	5,888947e-07(R)	0,01082282(R)	2,135601e-09(R)
Kupiec 1%	6,210588e-12(R)	0,3095893	0,0001401641(R)	0,008473333(R)	0,2341585(R)	0,0006566578(R)
Christoff. 5%	2,040205e-05(R)	5,99583e-07(R)	1,697194e-06(R)	2,026285e-06(R)	0,0122336(R)	4,909902e-09(R)
Christoff. 1%	5,364187e-11(R)	0,5074174	0,0007011565(R)	0,02919851(R)	0,3331484	0,002895986(R)
ES bootstrap 5%	4,622223e-05	0,1316408	0,09488302	0,8681284	0,3414234	0,9324648
ES próbkowy 5%	2,071653e-07(R)	0,06027143	0,03768734(R)	0,9394527	0,2840894	0,9833066
ES bootstrap 1%	—	0,03769699	—	—	0,5110893	—
ES próbkowy 1%	—	0,01248621(R)	—	—	0,4713184	—

Źródło: Opracowanie własne.

Tabela A.22: Prognozy zmienności, indeks BVP, model GARCH i GARCH-LSTM

Model/Metrics	G(N)	G(N)-LSTM	G(STD)	G(STD)-LSTM	G(SSTD)	G(SSTD)-LSTM
MSE	0,7934381	0,6665721	0,7988081	0,6707073	0,8114086	0,6783546
DM p-value		0,08275		0,1263		0,09155
MAE	0,560111	0,5333564	0,5577967	0,5324868	0,5613681	0,5337788
HMSE	0,1526598	0,2279903	0,1491898	0,2304943	0,1499011	0,2281416
R^2	0,4379412	0,4460823	0,4162867	0,4415638	0,4118728	0,4420366
Przek. VaR	133/36	68/18	145/29	69/13	137/30	67/16
Przek. VaR (%)						
Kupiec 5%	0,4310731	1,586138e-08(R)	0,06383829	3,003042e-08	0,25182	8,253918e-09(R)
Kupiec 1%	0,03551858(R)	0,1453207	0,4173806	0,008473333(R)	0,3166331	0,0558152
Christoff. 5%	0,7327024	1,708624e-08	0,1205989	2,99768e-08	0,3447194	9,584317e-09(R)
Chrisotff. 1%	0,06463278	0,3037263	0,5111283	.02919851(R)	0,4198489	0,1447965
ES bootstrap 5%	0,0162502	0,1187758	0,3304445	0,4189453	0,246131	0,4448575
ES próbkowy 5%	0,004163439(R)	0,05844181	0,269692	0,3577988	0,1497405	0,3922364
ES bootstrap 1%	0,006502646	0,04925743	0,248665	—	0,1034294	0,3923926
ES próbkowy 1%	0,002624935(R)	0,02356333(R)	0,1716354	—	0,05026162	0,289595

Źródło: Opracowanie własne.

Tabela A.23: Prognozy zmienności, indeks BVP, model EGARCH i EGARCH-LSTM

Model/Metrics	E(N)	E(N)-LSTM	E(STD)	E(STD)-LSTM	E(SSTD)	E(SSTD)-LSTM
MSE	0,8487132	0,6503974	0,8181792	0,657855	0,8023545	0,6735475
DM p-value		0,03847		0,05963		0,05385
MAE	0,5765123	0,5364117	0,5532199	0,5433847	0,5480561	0,5441965
HMSE	0,1519191	0,2506146	0,173009	0,2567203	0,1845385	0,2609665
R^2	0,450027	0,5072175	0,3991935	0,5164304	0,4446621	0,515144
Przek. VaR	130/42	60/14	123/19	75/11	119/16	77/12
Przek. VaR (%)						
Kupiec 5%	0,6057333	5,51249e-11(R)	0,9009813	1,024309e-06(R)	0,6201504	2,977075e-06(R)
Kupiec 1%	0,001674417(R)	0,01697476(R)	0,2170907	0,001678938(R)	0,0558152	0,003927525(R)
Christoff. 5%	0,8291842	1,049834e-10(R)	0,9225836	4,080854e-06(R)	0,6498937	1,057124e-05(R)
Chrisotff. 1%	0,003483071(R)	0,05346227	0,4032968	0,006843567(R)	0,1447965	0,01474698(R)
ES bootstrap 5%	0,002761678	0,2262921	0,7937584	0,9126173	0,9192221	0,9322515
ES próbkowy 5%	0,0003896895	0,1394997	0,8796151	0,976134	0,9788193	0,9856935
ES bootstrap 1%	—	0,02325789	0,6881639	—	0,7878925	—
ES próbkowy 1%	—	0,01083468(R)	0,7376287	—	0,8509674	—

Źródło: Opracowanie własne.

Tabela A.24: Prognozy zmienności, indeks BVP, model GJR-GARCH i GJR-GARCH-LSTM

Model/Metrics	GJR(N)	GJR(N)-LSTM	GJR(STD)	GJR(STD)-LSTM	GJR(SSTD)	GJR(SSTD)-LSTM
MSE	0,7888172	0,6736189	0,7941569	0,6760366	0,8232394	0,6914141
DM p-value		0,08934		0,07432		0,1092
MAE	0,5665861	0,5339551	0,5653967	0,5378323	0,5705334	0,5373571
HMSE	0,1528855	0,2354904	0,1536421	0,2429683	0,1549077	0,2275408
R^2	0,4643912	0,4576045	0,455036	0,440399	0,3623912	0,3844152
Przek. VaR	128/36	66/14	135/25	67/13	142/28	69/10
Przek. VaR (%)						
Kupiec 5%	0,7381654	4,230779e-09(R)	0,3335094	8,253918e-09(R)	0,1119399	3,003042e-08(R)
Kupiec 1%	0,03551858(R)	0,01697476(R)	0,9791164	0,008473333(R)	0,5363499	0,0006566578(R)
Christoff. 5%	0,2482394	5,289744e-09(R)	0,395033	4,89735e-08(R)	0,2019643	1,628031e-07(R)
Christoff. 1%	0,06463278	0,05346227	0,7754523	0,02919851(R)	0,6004122	0,002895986
ES bootstrap 5%	0,01449717	0,1607308	0,222932	0,4686802	0,3506208	0,4887399
ES próbkowy 5%	0,003722228(R)	0,08179994	0,1364645	0,4455389	0,2912791	0,4420894
ES bootstrap 1%	0,04193854	0,008548071	0,1027438	—	0,2010305	0,1732925
ES próbkowy 1%	0,01566763(R)	0,003259883(R)	0,04382025(R)	—	0,1218748	0,1094762

Źródło: Opracowanie własne.

Tabela A.25: Prognozy zmienności, indeks BVP, model APARCH i APARCH-LSTM

Model/Metrics	AP(N)	AP(N)-LSTM	AP(STD)	AP(STD)-LSTM	AP(SSTD)	AP(SSTD)-LSTM
MSE	0,8060332	0,6740019	0,8053575	0,6695474	0,88005	0,7010085
DM p-value		0,1087		0,0696		0,04171
MAE	0,5722127	0,5344938	0,5554061	0,5327831	0,565691	0,5393321
HMSE	0,146249	0,2419149	0,1496342	0,2234588	0,1606514	0,2263562
R^2	0,404476	0,4157662	0,3523447	0,4160728	0,2348076	0,4008941
Przek. VaR	144/41	67/16	143/31	79/12	138/33	81/15
Przek. VaR (%)						
Kupiec 5%	0,07746676	8,253918e-09(R)	0,09341474	8,211886e-06(R)	0,2168065	2,152801e-05(R)
Kupiec 1%	0,002949403(R)	0,0558152	0,2341585	0,003927525(R)	0,1186792	0,03176469(R)
Christoff. 5%	0,2087564	4,89735e-08(R)	0,2352261	1,699626e-05(R)	0,4115865	4,861257e-05(R)
Christoff. 1%	0,00605536(R)	0,1447965	0,3331484	0,01474698(R)	0,1898935	0,09102395
ES bootstrap 5%	0,005872602	0,1979213	0,2344821	0,6529934	0,3395229	0,6089566
ES próbkowy 5%	0,001471848	0,119337	0,140498	0,6682199	0,2588353	0,5965249
ES bootstrap 1%	—	0,09421114	0,1387183	—	0,2645184	—
ES próbkowy 1%	—	0,04267026	0,07048602	—	0,1656818	—

Źródło: Opracowanie własne.

Wyniki modeli prognoz probabilistycznych

Tabela A.26: Ilość przekroczeń oszacowań VaR(0,05)/VaR(0,01) w sieciowych modelach prognozowania probabilistycznego

Model/Metrics	CNN-N	CNN-STD	CNN-SSTD
S&P VaR 5%/1%	112/39	114/23	132/22
NKX VaR 5%/1%	107/37	100/23	88/10
DAX VaR 5%/1%	158/63	135/28	103/20
WIG VaR 5%/1%	126/49	114/28	92/20
KOSPI VaR 5%/1%	142/52	96/14	57/8
BOVESPA VaR 5%/1%	134/30	99/18	65/7
	LSTM-N	LSTM-STD	LSTM-SSTD
S&P VaR 5%/1%	121/49	133/25	112/21
NKX VaR 5%/1%	111/47	114/25	95/16
DAX VaR 5%/1%	146/49	164/29	141/22
WIG VaR 5%/1%	120/46	123/30	106/19
KOSPI VaR 5%/1%	135/48	111/21	84/11
BOVESPA VaR 5%/1%	129/31	91/17	60/8

Uwaga: Oczekiwana liczba przekroczeń wynosi 124/24, ilość prognoz 2487

Źródło: Opracowanie własne.

Tabela A.27: Prognozy probabilistyczne, indeks S&P500, sieci CNN

Model/Metrics	CNN(N)	CNN(STD)	CNN(SSTD)
LPS	1,28204	1,24162	1,222025
CRPS	0,5229748	0,5248868	0,5197192
PIT p-value	2,412545e-07	1,55995e-05	0,1144406
Przek. VaR	112/39	114/23	132/22
Przek. VaR (%)	4,50/1,56	4,58/0,92	5,30/0,88
Kupiec 5%	0,2481387	0,3344264	0,4856905
Kupiec 1%	0,008555221(R)	0,7026887	0,5552061
Christoff. 5%	0,00356562(R)	0,04092647(R)	0,03516364(R)
Chrisotff. 1%	0,0003746328(R)	0,05439649	0,355113
ES bootstrap 5%		0,8487771	0,9775541
ES próbkowy 5%		0,929866	0,9989004
ES bootstrap 1%	6,021294e-05	0,7343305	0,9529538
ES próbkowy 1%	2,237537e-06(R)	0,7875293	0,9917414

Źródło: Opracowanie własne.

Tabela A.28: Prognozy probabilistyczne, indeks S&P500, sieci LSTM

Model/Metrics	LSTM(N)	LSTM(STD)	LSTM(SSTD)
LPS	1,263296	1,210462	1,193376
CRPS	0,5146218	0,5137579	0,5094981
PIT p-value	2,412545e-07	0,005700126	0,03094408
Przek. VaR	121/49	133/25	112/21
Przek. VaR (%)	4,86/1,97	5,34/1,005	4,50/0,84
Kupiec 5%	0,7569064	0,4310731	0,2481387
Kupiec 1%	1,756379e-05(R)	0,9791164	0,4229126
Christoff. 5%	0,4313099	0,01607378(R)	0,0264911(R)
Chrisotff. 1%	1,7719e-08(R)	0,5192243	0,2833833
ES bootstrap 5%	1,266502e-05	0,597149	0,5066612
ES próbkowy 5%	1,520095e-07(R)	0,6093057	0,4703062
ES bootstrap 1%	—	0,7041345	0,7884135
ES próbkowy 1%	—	0,7506336	0,8679756

Źródło: Opracowanie własne.

Tabela A.29: Prognozy probabilistyczne, indeks NIKKEI, sieci CNN

Model/Metrics	CNN(N)	CNN(STD)	CNN(SSTD)
LPS	1,613601	1,599068	1,586503
CRPS	0,6914938	0,6995629	0,6963353
PIT p-value	2,412545e-07	4,41466e-05	0,04761705
Przek. VaR	107/37	100/23	88/10
Przek. VaR (%)	4,30/1,48	4,02/0,92	3,53/0,40
Kupiec 5%	0,1023182	0,02055537(R)	0,000428861(R)
Kupiec 1%	0,02262881	0,7026887	0,0006566578(R)
Christoff. 5%	0,1453825	0,06066346	0,001789705(R)
Chrisotff. 1%	0,004497022	0,4228579	0,0002877767(R)
ES bootstrap 5%	0,002102751	0,3075887	—
ES próbkowy 5%	0,0003844503(R)	0,2251687	—
ES bootstrap 1%	—	0,380783	—
ES próbkowy 1%	—	0,2856043	—

Źródło: Opracowanie własne.

Tabela A.30: Prognozy probabilistyczne, indeks NIKKEI, sieci LSTM

Model/Metrics	LSTM(N)	LSTM(STD)	LSTM(SSTD)
LPS	1,61246	1,587088	1,585482
CRPS	0,6892744	0,6894952	0,6874244
PIT p-value	2,412545e-07	2,412545e-07	2,412545e-07
Przek. VaR	111/47	114/25	95/16
Przek. VaR (%)	4,46/1,88	4,58/1,005	3,81/0,64
Kupiec 5%	0,2112618	0,3344264	0,004914388(R)
Kupiec 1%	7,156069e-05	0,9791164	0,0558152
Christoff. 5%	0,1036547	0,5192243	0,009433387(R)
Chrisotff. 1%	7,318607e-05	0,5192243	0,03851254(R)
ES bootstrap 5%	0,0001054134	0,2559864	—
ES próbkowy 5%	7,944203e-06(R)	0,161735	—
ES bootstrap 1%	—	0,4130939	0,2318004
ES próbkowy 1%	—	0,3101348	0,1424559

Źródło: Opracowanie własne.

Tabela A.31: Prognozy probabilistyczne, indeks DAX, sieci CNN

Model/Metrics	CNN(N)	CNN(STD)	CNN(SSTD)
LPS	1,580043	1,559851	1,570825
CRPS	0,6804201	0,6892444	0,6901093
PIT p-value	2,412545e-07	2,412545e-07	2,412545e-07
Przek. VaR	158/63	135/28	103/20
Przek. VaR (%)	6,35/2,53	5,42/1,12	4,14/0,80
Kupiec 5%	0,002911799	0,1855083	0,04319162(R)
Kupiec 1%	1,211037e-10	0,5363499	0,3095893
Christoff. 5%	0,0001283507	0,1349565	0,05763739
Chrisotff. 1%	9,511005e-10	0,6004122	0,5074174
ES bootstrap 5%	—	0,6737697	0,6691323
ES próbkowy 5%	—	0,7277809	0,7148299
ES bootstrap 1%	—	0,5022836	0,6010707
ES próbkowy 1%	—	0,4382549	0,5626957

Źródło: Opracowanie własne.

Tabela A.32: Prognozy probabilistyczne, indeks DAX, sieci LSTM

Model/Metrics	LSTM(N)	LSTM(STD)	LSTM(SSTD)
LPS	1,55403	1,517858	1,522978
CRPS	0,6651706	0,666342	0,6645636
PIT p-value	2,412545e-07	2,412545e-07	0,3536619
Przek. VaR	146/49	164/29	141/22
Przek. VaR (%)	5,87/1,97	6,59/1,16	5,66/0,88
Kupiec 5%	0,05227703	0,0004912755(R)	0,1332994
Kupiec 1%	1,756379e-05	0,4173806	0,5552061
Christoff. 5%	0,007751175(R)	0,0001170212(R)	0,03889874(R)
Chrisotff. 1%	9,913803e-05	0,4646559	0,355113
ES bootstrap 5%	0,0004913535	—	0,8738948
ES próbkowy 5%	3,112749e-05(R)	—	0,9464557
ES bootstrap 1%	—	0,3795639	—
ES próbkowy 1%	—	0,3082807	—

Źródło: Opracowanie własne.

Tabela A.33: Prognozy probabilistyczne, indeks WIG, sieci CNN

Model/Metrics	CNN(N)	CNN(STD)	CNN(SSTD)
LPS	1,452008	1,392544	1,405504
CRPS	0,5616494	0,570386	0,5716405
PIT p-value	2,412545e-07	2,412545e-07	2,412545e-07
Przek. VaR	126/49	114/28	92/20
Przek. VaR (%)	5,066/1,97	4,58/1,12	3,69/0,80
Kupiec 5%	0,8795866	0,3344264	0,001843078
Kupiec 1%	1,756379e-05(R)	0,5363499	0,3095893
Christoff. 5%	0,007760636(R)	0,04092647	7,177948e-05
Chrisotff. 1%	1,059784e-06(R)	0,01072578(R)	0,2143328
ES bootstrap 5%	2,403899e-05	0,08714938	—
ES próbkowy 5%	2,618736e-06(R)	0,03894271(R)	—
ES bootstrap 1%	—	0,05862529	0,1747533
ES próbkowy 1%	—	0,0240911(R)	0,09648941

Źródło: Opracowanie własne.

Tabela A.34: Prognozy probabilistyczne, indeks WIG, sieci LSTM

Model/Metrics	LSTM(N)	LSTM(STD)	LSTM(SSTD)
LPS	1,426541	1,340019	1,347512
CRPS	0,560521	0,5553697	0,5553525
PIT p-value	2,412545e-07	0,002294637	0,008695127
Przek. VaR	120/46	123/30	106/19
Przek. VaR (%)	4,82/1,84	4,945/1,20	4,26/0,76
Kupiec 5%	0,6873387	0,9009813	0,08357332
Kupiec 1%	0,0001401641	0,3166331	0,2170907
Christoff. 5%	0,000830165	.001584901(R)	0,0001538501(R)
Chrisotff. 1%	0,0001252356	0,0009668114	0,1531993
ES bootstrap 5%	0,000117884	0,2029977	0,4422916
ES próbkowy 5%	7,166383e-06	0,1261955	0,3926888
ES bootstrap 1%	—	0,3234363	0,2413903
ES próbkowy 1%	—	0,2291602	0,1497585

Źródło: Opracowanie własne.

Tabela A.35: Prognozy probabilistyczne, indeks KOSPI, sieci CNN

Model/Metrics	CNN(N)	CNN(STD)	CNN(SSTD)
LPS	1,334934	1,314745	1,317218
CRPS	0,528494	0,5296849	0,5301505
PIT p-value	2,412545e-07	2,412545e-07	2,412545e-07
Przek. VaR	142/52	96/14	57/8
Przek. VaR (%)	5,70/2,09	3,86/0,56	2,29/0,32
Kupiec 5%	0,1119399	0,006674846(R)	5,027312e-12(R)
Kupiec 1%	1,846464e-06(R)	0,01697476	7,391891e-05(R)
Christoff. 5%	0,008804667(R)	0,006769434(R)	3,708023e-11(R)
Chrisotff. 1%	9,733326e-07(R)	0,05346227	0,0003783371(R)
ES bootstrap 5%	0,0002535663	—	—
ES próbkowy 5%	1,132197e-05(R)	—	—
ES bootstrap 1%	—	0,1928189	—
ES próbkowy 1%	—	0,1338842	—

Źródło: Opracowanie własne.

Tabela A.36: Prognozy probabilistyczne, indeks KOSPI, sieci LSTM

Model/Metrics	LSTM(N)	LSTM(STD)	LSTM(SSTD)
LPS	1,324042	1,296143	1,284761
CRPS	0,5246137	0,520129	0,5164842
PIT p-value	2,412545e-07	4,702576e-07	5,087511e-06
Przek. VaR	135/48	111/21	84/11
Przek. VaR (%)	5,42/1,93	4,46/0,84	3,37/0,44
Kupiec 5%	0,3335094	0,2112618	8,331177e-05(R)
Kupiec 1%	3,580577e-05(R)	0,4229126	0,001678938(R)
Christoff. 5%	0,2382491	0,1846839	0,0004326179(R)
Chrisotff. 1%	0,0001947526(R)	0,6065243	0,006843567(R)
ES bootstrap 5%	0,000525885	0,5541596	—
ES próbkowy 5%	3,741374e-05	0,5469214	—
ES bootstrap 1%	—	0,6018079	—
ES próbkowy 1%	—	0,6328942	—

Źródło: Opracowanie własne.

Tabela A.37: Prognozy probabilistyczne, indeks BOVESPA, sieci CNN

Model/Metrics	CNN(N)	CNN(STD)	CNN(SSTD)
LPS	1,774015	1,782918	1,787533
CRPS	0,8267813	0,8369301	0,8349038
PIT p-value	2,412545e-07	0,186626	0,006914349
Przek. VaR	134/30	99/18	65/7
Przek. VaR (%)	5,38/1,20	3,98/0,72	2,61/0,28
Kupiec 5%	0,3803133	0,01574687(R)	2,135601e-09(R)
Kupiec 1%	0,3166331	0,1453207	2,073181e-05(R)
Christoff. 5%	0,6504183	0,0187212(R)	1,378183e-08(R)
Chrisotff. 1%	0,4198489	0,3037263	0,0001138994(R)
ES bootstrap 5%	0,09109641	—	—
ES próbkowy 5%	0,03699279(R)	—	—
ES bootstrap 1%	0,00768195	0,2585884	—
ES próbkowy 1%	0,003277963	0,1794694	—

Źródło: Opracowanie własne.

Tabela A.38: Prognozy probabilistyczne, indeks BOVESPA, sieci LSTM

Model/Metrics	LSTM(N)	LSTM(STD)	LSTM(SSTD)
LPS	1,777575	1,775115	1,785089
CRPS	0,8221921	0,8268515	0,833506
PIT p-value	2,412545e-07	0,003849091	0,005559229
Przek. VaR	129/31	91/17	60/8
Przek. VaR (%)	5,18/1,24	3,65/0,68	2,41/0,32
Kupiec 5%	0,6705888	0,00130125(R)	5,51249e-11(R)
Kupiec 1%	0,2341585	0,09252608	1,049834e-10
Christoff. 5%	0,4559963	0,005579407(R)	1,049834e-10(R)
Chrisotff. 1%	0,3474328	0,2161062	0,0003783371
ES bootstrap 5%	0,01629235	—	—
ES próbkowy 5%	0,004144868	—	—
ES bootstrap 1%	—	0,2206124	—
ES próbkowy 1%	—	0,1296427	—

Źródło: Opracowanie własne.

Bibliografia

- M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, i X. Zheng. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv:1603.04467 [cs]*, 2016.
- C. Acerbi i D. Tasche. On the coherence of Expected Shortfall. *arXiv:cond-mat/0104295*, 2002.
- C. Acerbi, C. Nordio, i C. Sirtori. Expected Shortfall as a Tool for Financial Risk Management. *arXiv:cond-mat/0102304*, 2001.
- C. Aloui i S. Mabrouk. Value-at-risk estimations of energy commodities via long-memory, asymmetry and fat-tailed GARCH models. *Energy Policy*, 38(5):2326–2339, 2010.
- B. Alshemali i J. Kalita. Improving the Reliability of Deep Neural Networks in NLP: A Review. *Knowledge-Based Systems*, 191:105–210, 2020.
- T. G. Andersen i T. Bollerslev. Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review*, 39(4):885–905, 1998.
- T. G. Andersen, T. Bollerslev, F. X. Diebold, i P. Labys. Great Realizations. *RISK*, 13:105–108, 2000.
- T. W. Anderson i D. A. Darling. A Test of Goodness of Fit. *Journal of the American Statistical Association*, 49(268):765–769, 1954.
- D. Ardia i L. F. Hoogerheide. GARCH models for daily stock returns: Impact of estimation frequency on Value-at-Risk and Expected Shortfall forecasts. *Economics Letters*, 123(2):187–190, 2014.

- P. Artzner, F. Delbaen, J.-M. Eber, i D. Heath. Coherent Measures of Risk. *Mathematical Finance*, 9(3):203–228, 1999.
- G. R. Babu. *Financial Markets And Institutions*. Concept Publishing Company, 2006.
- S. Bai, J. Z. Kolter, i V. Koltun. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. *arXiv:1803.01271 [cs]*, 2018.
- D. Bams, G. Blanchard, i T. Lehnert. Volatility measures and Value-at-Risk. *International Journal of Forecasting*, 33(4):848–863, 2017.
- Basel Committee on Banking Supervision. Basel II: International Convergence of Capital Measurement and Capital Standards: A Revised Framework. 2005.
- Basel Committee on Banking Supervision. Basel III: International regulatory framework for banks. 2017.
- L. Bauwens, M. Lubrano, i J. Richard. *Bayesian Inference in Dynamic Econometric Models*. Advanced Texts in Econometrics. OUP Oxford, 2000.
- S. Beckers. Standard deviations implied in option prices as predictors of future stock price variability. *Journal of Banking & Finance*, 5(3):363–381, 1981.
- Y. Bengio. Learning Deep Architectures for AI. *Foundations and Trends® in Machine Learning*, 2(1):1–127, 2009.
- Y. Bengio, A. Courville, i P. Vincent. Representation Learning: A Review and New Perspectives. *arXiv:1206.5538 [cs]*, 2014.
- C. Bergmeir i J. M. Benítez. On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191:192–213, 2012.
- A. Bernal, S. Fok, i R. Pidaparathi. Financial Market Time Series Prediction with Recurrent Neural Networks. *State College: Citeseer*, 2012.
- J. M. Bernardo. Reference Posterior Distributions for Bayesian Inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, 41(2):113–128, 1979.
- K. Bień-Barkowska. *Mikrostruktura rynku: ekonometryczne modelowanie dynamiki procesu transakcyjnego*. Szkoła Główna Handlowa - Oficyna Wydawnicza, 2016.
- J. Bizon-Górecka. źródła ryzyka finansowego w organizacji gospodarczej. W: W. Tarczyński (Red.), *Rynek Kapitałowy - Skuteczne Inwestowanie, Uniwersytet Szczeciński*. Szczecin, 2000.

- B. Będowska-Sójka. Daily Var Forecasts with Realized Volatility and Garch Models. *Argumenta Oeconomica*, (nr 1 (34)):157–173, 2015.
- B. Będowska-Sójka i A. Kliber. Realized Volatility Versus GARCH and Stochastic Volatility Models. The Evidence from the WIG20 Index and the EUR/PLN Foreign Exchange Market. *Przegląd Statystyczny*, 57(4):105–127, 2010.
- F. Black i M. Scholes. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81(3):637–654, 1973.
- Z. Bodie, A. Kane, i A. Marcus. *Essentials of Investments*. McGraw Hill, New York, NY, 2018.
- T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.
- T. Bollerslev. A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *The Review of Economics and Statistics*, 69(3):542–547, 1987.
- T. Bollerslev, R. F. Engle, i J. M. Wooldridge. A Capital Asset Pricing Model with Time-Varying Covariances. *Journal of Political Economy*, 96(1):116–131, 1988.
- L. Bottou. On-line learning and stochastic approximations. W: *In On-line Learning in Neural Networks*, s. 9–42. Cambridge University Press, 1998.
- Y.-L. Boureau, J. Ponce, i Y. Lecun. A Theoretical Analysis of Feature Pooling in Visual Recognition. W: *ICML'10: Proceedings of the 27th International Conference on Machine Learning*, s. 118, 2010.
- S. Bozinovski. Reminder of the First Paper on Transfer Learning in Neural Networks, 1976. *Informatika*, 44(3), 2020.
- J. Brownlee. *Deep Learning for Time Series Forecasting*. Machine Learning Mastery, 2018.
- A. Brusaferrri, M. Matteucci, P. Portolani, i A. Vitali. Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. *Applied Energy*, 250: 1158–1175, 2019.
- J. Brzeszczyński i R. Kelm. *Ekonometryczne modele rynków finansowych: Modele kursów giełdowych i kursów walutowych*. Wig-Press, Warszawa, 2002.

- J. Cai, J. Luo, S. Wang, i S. Yang. Feature selection in machine learning: A new perspective. *Neurocomputing*, 300:70–79, 2018.
- J. Y. Campbell, A. W. Lo, i A. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1997.
- J. Cao i J. Wang. Stock price forecasting model based on modified convolution neural network and financial time series analysis. *International Journal of Communication Systems*, 32(12):e3987, 2019.
- D. T. Chang. Bayesian Neural Networks: Essentials. *arXiv:2106.13594 [cs]*, 2021.
- J. Chen, W. Chen, C. Huang, S. Huang, i A. Chen. Financial Time-Series Data Analysis Using Deep Convolutional Neural Networks. W: *2016 7th International Conference on Cloud Computing and Big Data (CCBD)*, s. 87–92, 2016.
- K. Chen, Y. Zhou, i F. Dai. A LSTM-based method for stock returns prediction: A case study of China stock market. W: *2015 IEEE International Conference on Big Data (Big Data)*, s. 2823–2824, 2015.
- T.-T. Chen i S.-J. Lee. A weighted LS-SVM based learning system for time series forecasting. *Information Sciences*, 299:99–116, 2015.
- Y. Chen, Y. Kang, Y. Chen, i Z. Wang. Probabilistic Forecasting with Temporal Convolutional Neural Network. *arXiv:1906.04397 [cs, stat]*, 2020.
- K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, i Y. Bengio. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *arXiv:1406.1078 [cs, stat]*, 2014.
- F. Chollet i J. Allaire. *Deep Learning with R*. Manning Publications, 2018.
- F. Chollet i in. Keras, 2015. URL: <https://keras.io/>.
- E. Chong, C. Han, i F. C. Park. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Systems with Applications*, 83:187–205, 2017.
- A. Choromanska, M. Henaff, M. Mathieu, G. B. Arous, i Y. LeCun. The Loss Surfaces of Multilayer Networks. *arXiv:1412.0233 [cs]*, 2015.
- C. Chorro, D. Guégan, i F. Ielpo. Option pricing for GARCH-type models with generalized hyperbolic innovations. *Quantitative Finance*, 12(7):1079–1094, 2012.

- P. F. Christoffersen. Evaluating Interval Forecasts. *International Economic Review*, 39 (4):841–862, 1998.
- D. Cireşan, U. Meier, J. Masci, L. M. Gambardella, i J. Schmidhuber. Flexible, High Performance Convolutional Neural Networks for Image Classification. W: *International Joint Conference on Artificial Intelligence IJCAI-2011*, s. 1242, 2011.
- P. K. Clark. A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica*, 41(1):135–155, 1973.
- T. E. Clark i F. Ravazzolo. Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. *Journal of Applied Econometrics*, 30(4): 551–575, 2015.
- R. Cont. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.
- P. H. Cootner. *The Random Character of Stock Market Prices*. M.I.T. Press, University of California, 1964.
- C. Cortes i V. Vapnik. Support-vector networks. *Machine Learning*, 20(3):273–297, 1995.
- F. J. M. Costa. *Forecasting Volatility Using GARCH Models*. Praca magisterska, University of Minho, Minho, 2017.
- M. Courbariaux, Y. Bengio, i J.-P. David. BinaryConnect: Training Deep Neural Networks with binary weights during propagations. *arXiv:1511.00363 [cs]*, 2016.
- A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, i A. A. Bharath. Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, 35(1):53–65, 2018.
- J. Danielsson. *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and Matlab*. John Wiley & Sons, 2011.
- A. P. Dawid. Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach. *Journal of the Royal Statistical Society. Series A (General)*, 147(2):278, 1984.
- P. Dayan, G. E. Hinton, R. M. Neal, i R. S. Zemel. The Helmholtz Machine. *Neural Computation*, 7(5):889–904, 1995.

- M. P. Deisenroth, A. A. Faisal, i C. S. Ong. *Mathematics for Machine Learning*. Cambridge University Press, 2020.
- A. V. Devadoss, T. Antony, i A. Ligorì. Stock Prediction Using Artificial Neural Networks. *International Journal of Web Technology*, 2(2):45–51, 2013.
- J. Dhaene, R. J. A. Laeven, S. Vanduffel, G. Darkiewicz, i M. J. Goovaerts. Can a Coherent Risk Measure Be Too Subadditive? *Journal of Risk and Insurance*, 75(2): 365–386, 2008.
- L. Di Persio i O. Honchar. Artificial neural networks architectures for stock price prediction: Comparisons and applications. *International Journal of Circuits, Systems and Signal Processing*, 10:403–413, 2016.
- F. X. Diebold i R. S. Mariano. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics*, 13(3):253–263, 1995.
- F. X. Diebold, T. A. Gunther, i A. S. Tay. Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39 (4):863–883, 1998.
- J. V. Dillon, I. Langmore, D. Tran, E. Brevdo, S. Vasudevan, D. Moore, B. Patton, A. Alemi, M. Hoffman, i R. A. Saurous. TensorFlow Distributions. *arXiv:1711.10604 [cs, stat]*, 2017.
- D. Ding. Modeling of Market Volatility with APARCH Model, 2011. URL: <https://www.diva-portal.org/smash/get/diva2:417608/FULLTEXT01.pdf>.
- X. Ding, Y. Zhang, T. Liu, i J. Duan. Deep Learning for Event-Driven Stock Prediction. W: *IJCAI'15: Proceedings of the 24th International Conference on Artificial Intelligence*, s. 7, 2015.
- Z. Ding, C. W. J. Granger, i R. F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106, 1993.
- M. Dixon, D. Klabjan, i J. H. Bang. Classification-based financial markets prediction using deep neural networks. *Algorithmic Finance*, 6(3-4):67–77, 2017.
- M. Doman i R. Doman. Ekonometryczne modelowanie dynamiki polskiego rynku finansowego. *Prace Habilitacyjne, Akademia Ekonomiczna w Poznaniu*, (nr 15), 2004.
- M. Doman i R. Doman. *Modelowanie Zmienności i Ryzyka. Metody Ekonometrii Finansowej*. Wolters Kluwer, Kraków, 2009.

- T. Dozat. Incorporating Nesterov Momentum into Adam. *ICLR 2016 workshop submission*, 2016.
- C. Driksaki. Modeling the Volatility of Exchange Rate Currency using GARCH Model. *Economia Internazionale / International Economics*, 72(2):209–230, 2019.
- J. Duchi, E. Hazan, i Y. Singer. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*, 12(61):2121–2159, 2011.
- O. Duerr, B. Sick, i E. Murina. *Probabilistic Deep Learning: With Python, Keras and TensorFlow Probability*. Simon and Schuster, 2020.
- E.-I. Dumitrescu, C. Hurlin, i V. Pham. Backtesting Value-at-Risk: From Dynamic Quantile to Dynamic Binary Tests. *Finance*, 33(1):79–112, 2012.
- C. Dunis, P. Middleton, A. Karathanasopolous, i K. Theofilatos. *Artificial Intelligence in Financial Markets: Cutting Edge Applications for Risk Management, Portfolio Optimization and Economics*. New Developments in Quantitative Trading and Investment. Palgrave Macmillan UK, 2016.
- P. Embrechts, C. Klüppelberg, i T. Mikosch. *Modelling Extremal Events: For Insurance and Finance*. Springer, New York, 1997.
- R. F. Engle. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1007, 1982.
- R. F. Engle. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics*, 20(3):339–350, 2002.
- R. F. Engle i D. Kraft. Autoregressive Conditional Heteroskedasticity in Multiple Time Series. *Department of Economics, UCSD, Working Paper*, 1983.
- R. F. Engle i K. F. Kroner. Multivariate Simultaneous Generalized Arch. *Econometric Theory*, 11(1):122–150, 1995.
- R. F. Engle i S. Manganelli. CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics*, 22(4):367–381, 2004.
- R. F. Engle i A. Patton. What good is a volatility model? *Quantitative Finance*, 1, 2001.

- M. Fałdziński i M. Osińska. Volatility Estimators in Econometric Analysis of Risk Transfer on Capital Markets. *Dynamic Econometric Models*, 16:21–35, 2016.
- E. F. Fama. The Behavior of Stock-Market Prices. *The Journal of Business*, 38(1): 34–105, 1965a.
- E. F. Fama. Random Walks in Stock Market Prices. *Financial Analysts Journal*, 21 (5):55–59, 1965b.
- R. T. Farias Nazário, J. L. e Silva, V. A. Sobreiro, i H. Kimura. A literature review of technical analysis on stock markets. *The Quarterly Review of Economics and Finance*, 66:115–126, 2017.
- C. Fernandez i M. F. J. Steel. On Bayesian Modeling of Fat Tails and Skewness. *Journal of the American Statistical Association*, 93(441):359–371, 1998.
- T. Fischer i C. Krauss. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2): 654–669, 2018.
- P. Fiszeder. Forecasting the volatility of the Polish stock index – WIG20. W: *Forecasting Financial Markets. Theory and Applications*. Łódź, 2005.
- P. Fiszeder. Prognozowanie VaR – zastosowanie wielorównaniowych modeli GARCH, Modelowanie i prognozowanie gospodarki narodowej. *Prace i Materiały Wydziału Zarządzania Uniwersytetu Gdańskiego*, s. 365–376, 2007.
- P. Fiszeder. *Modele klasy Garch w empirycznych badaniach finansowych*. Rozprawa habilitacyjna / Uniwersytet Mikołaja Kopernika. Wydawnictwo Naukowe Uniwersytetu Mikołaja Kopernika, Toruń, 2009.
- P. Fiszeder. *Ceny Minimalne i Maksymalne w Modelowaniu i Prognozowaniu Zmienności Oraz Zależności Na Rynkach Finansowych*. Wydawnictwo Naukowe PWN, Warszawa, 2020.
- T. Fletcher. *Machine Learning for Financial Market Prediction*. Praca doktorska, 2012.
- C. Francq i J.-M. Zakoian. *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons, 2019.
- K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4): 193–202, 1980.

- Y. Gal i Z. Ghahramani. A Theoretically Grounded Application of Dropout in Recurrent Neural Networks. *arXiv:1512.05287 [stat]*, 2016.
- D. García i W. Kristjanpoller. An adaptive forecasting approach for copper price volatility through hybrid and non-hybrid models. *Applied Soft Computing*, 74:466–478, 2019.
- M. B. Garman i M. J. Klass. On the Estimation of Security Price Volatilities from Historical Data. *The Journal of Business*, 53(1):67–78, 1980.
- A. Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O’Reilly Media, 2019.
- A. Ghalanos. Rugarch: Univariate GARCH models, 2020. URL: <https://cran.r-project.org/web/packages/rugarch/rugarch.pdf>.
- X. Glorot i Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research - Proceedings Track*, 9:249–256, 2010.
- X. Glorot, A. Bordes, i Y. Bengio. Deep Sparse Rectifier Neural Networks. W: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15, s. 315–323, 2010.
- L. R. Glosten, R. Jagannathan, i D. E. Runkle. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.
- T. Gneiting i M. Katzfuss. Probabilistic Forecasting. *Annual Review of Statistics and Its Application*, 1(1):125–151, 2014.
- T. Gneiting i A. E. Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.
- T. Gneiting, F. Balabdaoui, i A. E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(2):243–268, 2007.
- I. J. Good. Rational Decisions. *Journal of the Royal Statistical Society: Series B (Methodological)*, 14(1):107–114, 1952.
- I. Goodfellow, Y. Bengio, i A. Courville. *Deep Learning*. The MIT Press, 2016.

- I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, i Y. Bengio. Generative Adversarial Networks. *arXiv:1406.2661 [cs, stat]*, 2014.
- K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, i J. Schmidhuber. LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10):2222–2232, 2017.
- J. Han, J. Pei, i M. Kamber. *Data Mining: Concepts and Techniques*. Elsevier, 2011.
- E. J. Hannan i B. G. Quinn. The Determination of the Order of an Autoregression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(2):190–195, 1979.
- C. Hansen. Optimizers Explained - Adam, Momentum and Stochastic Gradient Descent, 2019. URL: <https://mlfromscratch.com/optimizers-explained/>.
- M. Hansson. *On Stock Return Prediction with LSTM Networks*. Praca magisterska, 2017.
- C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. Fernández del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, i T. E. Oliphant. Array programming with NumPy. *Nature*, 585:357–362, 2020.
- D. Harvey, S. Leybourne, i P. Newbold. Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2):281–291, 1997.
- K. He, X. Zhang, S. Ren, i J. Sun. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]*, 2015.
- M. L. Higgins i A. K. Bera. A Class of Nonlinear ARCH Models. *International Economic Review*, 33(1):137–158, 1992.
- S. Hihi i Y. Bengio. Hierarchical recurrent neural networks for long-term dependencies. W: *Advances in Neural Information Processing Systems*, volume 8. MIT Press, 1995.
- G. E. Hinton. Deep belief networks. *Scholarpedia*, 4(5):5947, 2009.
- G. E. Hinton. A Practical Guide to Training Restricted Boltzmann Machines. W: *Neural Networks: Tricks of the Trade: Second Edition*, Lecture Notes in Computer Science, s. 599–619. Springer, Berlin, Heidelberg, 2012.

- G. E. Hinton i T. J. Sejnowski. *Unsupervised Learning: Foundations of Neural Computation*. MIT Press, 1999.
- G. E. Hinton i R. S. Zemel. Autoencoders, minimum description length and Helmholtz free energy. W: *Proceedings of the 6th International Conference on Neural Information Processing Systems*, NIPS'93, s. 3–10, San Francisco, CA, USA, 1993.
- G. E. Hinton, S. Osindero, i Y.-w. Teh. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7):1527–54, 2006.
- G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, i R. R. Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv:1207.0580 [cs]*, 2012a.
- G. E. Hinton, K. Swersky, i N. Srivastava. Neural Networks for Machine Learning - Lecture 6a - Overview of mini-batch gradient descent, 2012b. URL: https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides lec6.pdf.
- S. Hochreiter i J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 1997.
- S. Hochreiter, Y. Bengio, P. Frasconi, i J. Schmidhuber. Gradient Flow in Recurrent Nets: The Difficulty of Learning Long-Term Dependencies. *A Field Guide to Dynamical Recurrent Networks*, 2001.
- J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982.
- A. Hossain i M. Nasser. Comparison of the finite mixture of ARMA-GARCH, back propagation neural networks and support-vector machines in forecasting financial returns. *Journal of Applied Statistics*, 38(3):533–551, 2011.
- Y. Hu, J. Ni, i L. Wen. A hybrid deep learning approach by integrating LSTM-ANN networks with GARCH model for copper price volatility prediction. *Physica A: Statistical Mechanics and its Applications*, 557:124907, 2020.
- J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing in science & engineering*, 9(3):90–95, 2007.
- R. J. Hyndman i A. B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.

- S. Ioffe i C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv:1502.03167*, 2015.
- H. Jaeger i H. Haas. Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication. *Science*, 304:78–80, 2004.
- K. Jajuga. Miary ryzyka rynkowego - cz. I. *Rynek Terminowy*, (4):67–69, 1999.
- Y. Jeong i S. Lee. Recurrent neural network-adapted nonlinear ARMA-GARCH model with application to S&P 500 index data. *Journal of the Korean Data and Information Science Society*, 30(5):1187–1195, 2019.
- Jiahong Li, Hui Bu, i Junjie Wu. Sentiment-aware stock market prediction: A deep learning method. W: *2017 International Conference on Service Systems and Service Management*, s. 1–6, 2017.
- X. Jin, X. Yu, X. Wang, Y. Bai, T. Su, i J. Kong. Prediction for Time Series with CNN and LSTM. W: R. Wang, Z. Chen, W. Zhang, i Q. Zhu (red.), *Proceedings of the 11th International Conference on Modelling, Identification and Control (ICMIC2019)*, Lecture Notes in Electrical Engineering, s. 631–641, Singapore, 2020. Springer.
- A. Jordan, F. Krüger, i S. Lerch. Evaluating probabilistic forecasts with scoringRules. *arXiv:1709.04743 [cs]*, 2018.
- J. Jordan. Setting the learning rate of your neural network., 2018. URL: <https://www.jeremyjordan.me/nn-learning-rate/>.
- M. I. Jordan. Serial order: A parallel distributed processing approach. Technical report, June 1985-March 1986. Technical Report AD-A-173989/5/XAB; ICS-8604, California University, 1986.
- P. Jorion. *Value at Risk, 3rd Ed.: The New Benchmark for Managing Financial Risk*. McGraw-Hill Education, 2006.
- L. V. Jospin, W. Buntine, F. Boussaid, H. Laga, i M. Bennamoun. Hands-on Bayesian Neural Networks – a Tutorial for Deep Learning Users. *arXiv:2007.06823 [cs, stat]*, 2022.
- P. R. Junior, F. L. R. Salomon, i E. d. O. Pamplona. ARIMA: An Applied Time Series Forecasting Model for the Bovespa Stock Index. *Applied Mathematics*, 05(21):3383, 2014.
- M. Kachniewski, B. Majewski, i P. Wasilewski. *Rynek Kapitałowy i Giełda Papierów Wartościowych*. Fundacja Edukacji i Rynku Kapitałowego, Warszawa, 2008.

- L. P. Kaelbling, M. L. Littman, i A. W. Moore. Reinforcement Learning: A Survey. *Journal of Artificial Intelligence Research*, 4:237–285, 1996.
- M. G. Kendall i A. B. Hill. The Analysis of Economic Time-Series-Part I: Prices. *Journal of the Royal Statistical Society. Series A (General)*, 116(1):11–34, 1953.
- M. Kijewski i R. Ślepaczuk. Predicting prices of S&P500 index using classical methods and recurrent neural networks. Working Paper 2020-27, Faculty of Economic Sciences, University of Warsaw, 2020.
- H. Y. Kim i C. H. Won. Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert Systems with Applications*, 103:25–37, 2018.
- S. Kim i H. Kim. A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3):669–679, 2016.
- D. P. Kingma i J. Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, 2017.
- D. P. Kingma i M. Welling. Auto-Encoding Variational Bayes. *arXiv:1312.6114 [cs, stat]*, 2014.
- D. P. Kingma i M. Welling. An Introduction to Variational Autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019.
- F. H. Knight. *Risk, Uncertainty and Profit*. Houghton Mifflin Company, Boston; New York, 1921.
- C. Kocak. ARMA(p,q) type high order fuzzy time series forecast method based on fuzzy logic relations. *Applied Soft Computing*, 58:92–103, 2017.
- V. Kodogiannis i A. Lolis. Forecasting Financial Time Series using Neural Network and Fuzzy System-based Techniques. *Neural Computing & Applications*, 11(2):90–102, 2002.
- K. Kość, P. Sakowski, i R. Ślepaczuk. Momentum and contrarian effects on the cryptocurrency market. *Physica A: Statistical Mechanics and its Applications*, 523: 691–701, 2019.
- K. Kozdra. Rynek Catalyst w finansowaniu przedsiębiorstw. *Zeszyty Naukowe Państwowej Wyższej Szkoły Zawodowej w Płocku. Nauki Ekonomiczne*, 22:235–246, 2015.

- W. Kristjanpoller i E. Hernández. Volatility of main metals forecasted by a hybrid ANN-GARCH model with regressors. *Expert Systems with Applications*, 84:290–300, 2017.
- W. Kristjanpoller i M. C. Minutolo. A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis. *Expert Systems with Applications*, 109:1–11, 2018.
- J. Kudła. *Instrumenty Finansowe i Ich Zastosowania*. Key Text, Warszawa, 2011.
- P. Kupiec. Techniques for Verifying the Accuracy of Risk Measurement Models. SSRN Scholarly Paper ID 6697, Rochester, NY, 1995.
- A. Kurani, P. Doshi, A. Vakharia, i M. Shah. A Comprehensive Comparative Study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on Stock Forecasting. *Annals of Data Science*, 2021.
- Ł. Kwiatkowski. *Modele Dwuliniowe, GARCH i SV w Modelowaniu Finansowych Szeregów Czasowych*. Praca magisterska, Uniwersytet Ekonomiczny w Krakowie, 2008.
- K. J. Lang, A. H. Waibel, i G. E. Hinton. A time-delay neural network architecture for isolated word recognition. *Neural Networks*, 3(1):23–43, 1990.
- A. G. Laporta, L. Merlo, i L. Petrella. Selection of Value at Risk Models for Energy Commodities. *Energy Economics*, 74:628–643, 2018.
- Y. Lecun, L. Bottou, Y. Bengio, i P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- C. Lemaréchal. Cauchy and the Gradient Method, 2012. URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.398.8884>.
- L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, i A. Talwalkar. Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. *arXiv:1603.06560 [cs, stat]*, 2018.
- L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, i J. Han. On the Variance of the Adaptive Learning Rate and Beyond. *arXiv:1908.03265 [cs, stat]*, 2021.
- W. K. Liu i M. K. P. So. A GARCH Model with Artificial Neural Networks. *Information*, 11(10):489, 2020.

- I. E. Livieris, E. Pintelas, i P. Pintelas. A CNN–LSTM model for gold price time-series forecasting. *Neural Computing and Applications*, 32(23):17351–17360, 2020.
- W. Lu, J. Li, Y. Li, A. Sun, i J. Wang. A CNN-LSTM-Based Model to Forecast Stock Prices. *Complexity*, 2020:e6622927, 2020.
- W. Lu, J. Li, J. Wang, i L. Qin. A CNN-BiLSTM-AM method for stock price prediction. *Neural Computing and Applications*, 33(10):4741–4753, 2021.
- X. Lu, D. Que, i G. Cao. Volatility Forecast Based on the Hybrid Artificial Neural Network and GARCH-type Models. *Procedia Computer Science*, 91:1044–1049, 2016.
- W. Maass, T. Natschläger, i H. Markram. Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations. *Neural Computation*, 14(11):2531–2560, 2002.
- R. Majid. Advances in Statistical Forecasting Methods: An Overview. *Economic Affairs*, 63(4), 2018.
- M. Małecka. *Weryfikacja hipotez w ocenie ryzyka rynkowego*. Wydawnictwo Uniwersytetu Łódzkiego, Łódź, 2016.
- M. Mallikarjuna i R. P. Rao. Evaluation of forecasting methods from selected stock market returns. *Financial Innovation*, 5(1):40, 2019.
- B. Mandelbrot. The Variation of Certain Speculative Prices. *The Journal of Business*, 36, 1963.
- D. Marcek, M. Marcek, i J. Babel. Granular RBF NN Approach and Statistical Methods Applied to Modelling and Forecasting High Frequency Data. *International Journal of Computational Intelligence Systems*, 2, 2009.
- M. Martens. Measuring and forecasting S&P 500 index-futures volatility using high-frequency data. *Journal of Futures Markets*, 22(6):497–518, 2002.
- J. M. Matías, M. Febrero-Bande, W. González-Manteiga, i J. C. Reboredo. Boosting GARCH and neural networks for the prediction of heteroskedastic time series. *Mathematical and Computer Modelling*, 51(3):256–271, 2010.
- F. Matsumoto. Redes Neurais — LSTM, 2020. URL: <https://medium.com/turing-talks/turing-talks-27-modelos-de-predicao-lstm-df85d87ad210>.
- W. McKinney. Data structures for statistical computing in python. W: *Proceedings of the 9th Python in Science Conference*, volume 445, s. 51–56. Austin, TX, 2010.

- A. J. McNeil i R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance*, 7 (3):271–300, 2000.
- S. Mehtab i J. Sen. Analysis and Forecasting of Financial Time Series Using CNN and LSTM-Based Deep Learning Models. W: J. P. Sahoo, A. K. Tripathy, M. Mohanty, K.-C. Li, i A. K. Nayak (red.), *Advances in Distributed Computing and Machine Learning*, Lecture Notes in Networks and Systems, s. 405–423, Singapore, 2022.
- J. Michańków, P. Sakowski, i R. Ślepaczuk. LSTM in Algorithmic Investment Strategies on BTC and S&P500 Index. *Sensors*, 22(3):917, 2022.
- B. Mohebbali, A. Tahmassebi, A. Meyer-Baese, i A. H. Gandomi. Chapter 14 - Probabilistic neural networks: A brief overview of theory, implementation, and application. W: P. Samui, D. Tien Bui, S. Chakraborty, i R. C. Deo (red.), *Handbook of Probabilistic Models*, s. 347–367. Butterworth-Heinemann, 2020.
- P. Mondal, L. Shit, i S. Goswami. Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. *International Journal of Computer Science, Engineering and Applications*, 4:13–29, 2014.
- S. A. Monfared i D. Enke. Volatility Forecasting Using a Hybrid GJR-GARCH Neural Network Model. *Procedia Computer Science*, 36:246–253, 2014.
- T. Moon, H. H. Choi, H. Lee, i I. Song. *RnnDrop: A Novel Dropout for RNNs in ASR*. 2015.
- J. Moraĳda. Neuronowe prognozowanie szeregów czasowych metodą przesuwanego okna danych. *Zeszyty Naukowe Małopolskiej Wyższej Szkoły Ekonomicznej w Tarnowie*, (nr 1(10)):189–199, 2007.
- MSCI. Market classification, 2021. URL: <https://www.msci.com/our-solutions/indexes/market-classification>.
- K. P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.
- K. P. Murphy. *Probabilistic Machine Learning: An Introduction*. MIT Press, 2022.
- R. M. Neal. *Bayesian Learning for Neural Networks*. Lecture Notes in Statistics. Springer New York, New York, NY, 1996.
- D. B. Nelson. Stationarity and Persistence in the GARCH(1,1) Model. *Econometric Theory*, 6(3):318–334, 1990.

- D. B. Nelson. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2):347–370, 1991.
- C. Nilsson. Forecasting Swedish Stock Market Volatility and Value-at-Risk: A Comparison of EWMA and GARCH Models. 2017. URL: <http://lup.lub.lu.se/student-papers/record/8923808>.
- M. M. Noel, A. L. A. Trivedi, i P. Dutta. Growing Cosine Unit: A Novel Oscillatory Activation Function That Can Speedup Training and Reduce Parameters in Convolutional Neural Networks. *arXiv:2108.12943 [cs]*, 2021.
- J. Nowotarski i R. Weron. Recent advances in electricity price forecasting - A review of probabilistic forecasting. *Renewable and Sustainable Energy Reviews*, 81:1548–1568, 2018.
- C. Nwankpa, W. Ijomah, A. Gachagan, i S. Marshall. Activation Functions: Comparison of trends in Practice and Research for Deep Learning. *arXiv:1811.03378 [cs]*, 2018.
- T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, i L. Invernizzi. KerasTuner, 2019. URL: <https://keras.io/keras-tuner/>.
- R. Oomen. Using High Frequency Stock Market Index Data to Calculate, Model & Forecast Realized Return Variance. *SSRN Electronic Journal*, 2001.
- J. Osiewalski. *Ekonometria Bayesowska w Zastosowaniach*. Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków, 2001.
- J. Osiewalski i A. Pajor. Bayesian Value-at-Risk for a Portfolio: Multi- and Univariate Approaches Using MSF-SBEKK Models. *Central European Journal of Economic Modelling and Econometrics*, 2(4):253–277, 2010.
- J. Osiewalski i A. Pajor. A hybrid MSV-MGARCH generalisation of the t-MGARCH model. *Socio-Economic Modelling and Forecasting. The 12th Professor Aleksander Zelias International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*, (1):345–354, 2018.
- J. Osiewalski, A. Pajor, i M. Pipień. Bayesowskie modelowanie i prognozowanie indeksu WIG z wykorzystaniem procesów GARCH i SV. W: A. Zelias (red.), *XX Seminarium Ekonometryczne im. Prof. Zbigniewa Pawłowskiego*, s. 17–39, 2004.

- K. Osiewalski i J. Osiewalski. Long-Run Relationship between Daily Prices on Two Markets: The Bayesian VAR(2)-MSF-SBEKK Model. *Central European Journal of Economic Modelling and Econometrics*, s. 65–83, 2013.
- M. Osińska i M. Fałdziński. Modele GARCH i SV z zastosowaniem teorii wartości ekstremalnych. W: *Dynamiczne Modele Ekonometryczne*, Toruń, 2007. Wydawnictwo Uniwersytetu Mikołaja Kopernika.
- A. Pajor. Procesy zmienności stochastycznej SV w bayesowskiej analizie finansowych szeregów czasowych. *Monografie: prace doktorskie, Akademia Ekonomiczna w Krakowie*, (nr 2), 2003.
- A. Pajor. Wielowymiarowe procesy wariancji stochastycznej w ekonometrii finansowej: ujęcie bayesowskie. *Zeszyty Naukowe, Uniwersytet Ekonomiczny w Krakowie. Seria Specjalna, Monografie*, nr 195, 2010.
- A. Pajor i J. Osiewalski. Bayesian Value-at-Risk for a Portfolio: Multi- and Univariate Approaches using MSF-SBEKK Models. *Central European Journal of Economic Modelling and Econometrics; 2010; No 4; 253-277*, 2010.
- A. Pajor i J. Wróblewska. VEC-MSF models in Bayesian analysis of short- and long-run relationships. *Studies in Nonlinear Dynamics & Econometrics*, 21(3), 2017.
- D. Park i D. Ryu. Forecasting Stock Market Dynamics using Bidirectional Long Short-Term Memory. *Journal for Economic Forecasting*, (2):22–34, 2021.
- M. Parkinson. The Extreme Value Method for Estimating the Variance of the Rate of Return. *The Journal of Business*, 53(1):61–65, 1980.
- R. Pascanu, C. Gulcehre, K. Cho, i Y. Bengio. How to Construct Deep Recurrent Neural Networks. *arXiv:1312.6026 [cs, stat]*, 2014.
- G. C. Pflug. Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk. W: *Probabilistic Constrained Optimization: Methodology and Applications, Nonconvex Optimization and Its Applications*, s. 272–281. Boston, MA, 2000.
- K. Piontek. Zmienność historyczna i implikowana jako prognozy zmienności instrumentów finansowych. *Prace Naukowe / Akademia Ekonomiczna w Katowicach, Modelowanie preferencji a ryzyko '99*. Cz. 1.:325–336, 1999.
- K. Piontek. Heteroskedastyczność szeregu stóp zwrotu a koncepcja pomiaru ryzyka metodą VaR. *Prace Naukowe / Akademia Ekonomiczna w Katowicach*, s. 339–349, 2001.

- K. Piontek. *Modelowanie i Prognozowanie Zmienności Instrumentów Finansowych*. Praca doktorska, Wrocław, 2002.
- K. Piontek i D. Papla. Wykorzystanie wielorównaniowych modeli AR-GARCH w pomiarze ryzyka metodą VaR. *Prace Naukowe Akademii Ekonomicznej we Wrocławiu*, (nr 1088, t. 2), 2005.
- M. Pipień. *Bayesowska analiza finansowych szeregów czasowych z wykorzystaniem modeli GARCH*. Praca doktorska, Kraków, 2000.
- M. Pipień. Wnioskowanie bayesowskie w ekonometrii finansowej. *Zeszyty Naukowe / Akademia Ekonomiczna w Krakowie. Seria Specjalna, Monografie*, (nr 176), 2006.
- V. P. Plagianakos, G. D. Magoulas, i M. N. Vrahatis. Learning Rate Adaptation in Stochastic Gradient Descent. W: N. Hadjisavvas i P. M. Pardalos (red.), *Advances in Convex Analysis and Global Optimization: Honoring the Memory of C. Caratheodory (1873–1950)*, Nonconvex Optimization and Its Applications, s. 433–444. Boston, MA, 2001.
- L. Prechelt. Early Stopping — But When? W: G. Montavon, G. B. Orr, i K.-R. Müller (red.), *Neural Networks: Tricks of the Trade: Second Edition*, Lecture Notes in Computer Science, s. 53–67. Springer, Berlin, Heidelberg, 2012.
- N. Qian. On the momentum term in gradient descent learning algorithms. *Neural Networks*, 12(1):145–151, 1999.
- R. Rabemananjara i J. M. Zakoian. Threshold arch models and asymmetries in volatility. *Journal of Applied Econometrics*, 8(1):31–49, 1993.
- P. Ramachandran, B. Zoph, i Q. V. Le. Searching for Activation Functions. *arXiv:1710.05941 [cs]*, 2017.
- V. Ravi, D. Pradeepkumar, i K. Deb. Financial time series prediction using hybrids of chaos theory, multi-layer perceptron and multi-objective evolutionary algorithms. *Swarm and Evolutionary Computation*, 36:136–149, 2017.
- D. J. Rezende, S. Mohamed, i D. Wierstra. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. *arXiv:1401.4082 [cs, stat]*, 2014.
- H. Riehl. *Zarządzanie Ryzykiem Na Rynku Pieniężnym, Walutowym i Instrumentów Pochodnych*. Warszawski Instytut Bankowości, Warszawa, 2001.
- R. T. Rockafellar i S. Uryasev. Optimization of conditional value-at-risk. *The Journal of Risk*, 2(3):21–41, 2000.

- L. C. G. Rogers i S. E. Satchell. Estimating Variance From High, Low and Closing Prices. *The Annals of Applied Probability*, 1(4):504–512, 1991.
- F. Rosenblatt. *The Perceptron, a Perceiving and Recognizing Automaton Project Para*. Cornell Aeronautical Laboratory, 1957.
- D. E. Rumelhart, G. E. Hinton, i R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.
- R. Salakhutdinov i G. Hinton. Deep Boltzmann Machines. W: *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, s. 448–455. PMLR, 2009.
- T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, i X. Chen. Improved Techniques for Training GANs. *arXiv:1606.03498 [cs]*, 2016.
- D. Salinas, V. Flunkert, J. Gasthaus, i T. Januschowski. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 36(3):1181–1191, 2020.
- K. Sawka. *Uczenie maszynowe z użyciem Scikit-Learn i TensorFlow. Wydanie II*. Helion, 2022.
- R. Schmalensee i R. R. Trippi. Common Stock Volatility Expectations Implied by Option Premia. *The Journal of Finance*, 33(1):129–147, 1978.
- J. Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 61: 85–117, 2015.
- M. Schnaubelt. A comparison of machine learning model validation schemes for non-stationary time series data. Working Paper 11/2019, FAU Discussion Papers in Economics, 2019.
- M. Schuster i K. Paliwal. Bidirectional recurrent neural networks. *Signal Processing, IEEE Transactions on*, 45:2673–2681, 1997.
- S. Semeniuta, A. Severyn, i E. Barth. Recurrent Dropout without Memory Loss. *arXiv:1603.05118 [cs]*, 2016.
- O. B. Sezer, M. U. Gudelek, i A. M. Ozbayoglu. Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. *Applied Soft Computing*, 90:106181, 2020.

- T. B. Shahi, A. Shrestha, A. Neupane, i W. Guo. Stock Price Forecasting with Deep Learning: A Comparative Study. *Mathematics*, 8(9):1441, 2020.
- S. Shalev-Shwartz i S. Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.
- R. Shibata. Selection of the Order of an Autoregressive Model by Akaike's Information Criterion. *Biometrika*, 63(1):117–126, 1976.
- J. Shu i J. E. Zhang. Testing range estimators of historical volatility. *Journal of Futures Markets*, 26(3):297–313, 2006.
- A. Silvennoinen i T. Teräsvirta. Multivariate GARCH Models. W: T. Mikosch, J.-P. Kreiß, R. A. Davis, i T. G. Andersen (red.), *Handbook of Financial Time Series*, s. 201–229. Springer, Berlin, Heidelberg, 2009.
- D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, i D. Hassabis. Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359, 2017.
- M. Skorski, A. Temperoni, i M. Theobald. Revisiting Initialization of Neural Networks. *arXiv:2004.09506 [cs, math, stat]*, 2020.
- R. Ślepaczuk, G. Zakrzewski, i P. Sakowski. Investment Strategies Beating the Market: What Can We Squeeze from the Market? SSRN Scholarly Paper ID 2508647, Social Science Research Network, Rochester, NY, 2012.
- S. Slim, Y. Koubaa, i A. BenSaïda. Value-at-Risk under Lévy GARCH models: Evidence from global stock markets. *Journal of International Financial Markets, Institutions and Money*, 46:30–53, 2017.
- P. Smolensky. Information processing in dynamical systems: Foundations of harmony theory. *Parallel Distributed Process*, 1, 1986.
- N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, i R. Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014.
- Superdatascience.com. Convolutional Neural Networks (CNN), 2018. URL: <https://www.superdatascience.com/blogs/convolutional-neural-networks-cnn-step-1b-relu-layer>.

- I. Sutskever, O. Vinyals, i Q. V. Le. Sequence to Sequence Learning with Neural Networks. *arXiv:1409.3215 [cs]*, 2014.
- T. Szandała. Review and Comparison of Commonly Used Activation Functions for Deep Neural Networks. W: A. K. Bhoi, P. K. Mallick, C.-M. Liu, i V. E. Balas (red.), *Bio-Inspired Neurocomputing*, Studies in Computational Intelligence, s. 203–224. Springer, Singapore, 2021.
- W. Tang, G. Long, L. Liu, T. Zhou, J. Jiang, i M. Blumenstein. Rethinking 1D-CNN for Time Series Classification: A Stronger Baseline. *arXiv:2002.10061 [cs, stat]*, 2021.
- W. Tarczyński i M. Mojsiewicz. *Zarządzanie Ryzykiem*. Polskie Wydawnictwo Ekonomiczne, Warszawa, 2001.
- H.-Z. Tay, K.-H. Ng, Y.-B. Koh, i K.-H. Ng. Model selection based on value-at-risk backtesting approach for GARCH-Type models. *Journal of Industrial & Management Optimization*, 16(4):1635, 2020.
- S. Taylor. *Modelling Financial Time Series*. Wiley, 1986.
- T. Teräsvirta. An Introduction to Univariate GARCH Models. W: T. Mikosch, J.-P. Kreiß, R. A. Davis, i T. G. Andersen (red.), *Handbook of Financial Time Series*, s. 17–42. Springer, Berlin, Heidelberg, 2009.
- A. Timmermann. Forecasting Methods in Finance. *Annual Review of Financial Economics*, 10(1):449–479, 2018.
- C. Tofallis. A Better Measure of Relative Prediction Accuracy for Model Selection and Model Estimation. SSRN Scholarly Paper ID 2635088, Social Science Research Network, Rochester, NY, 2014.
- J. Tong, C. Wu, i L. Lian. Comparison of GARCH Family Models for Shanghai Stock Exchange Index Prediction. *Review of Integrative Business and Economics Research*, 2020.
- J.-F. Toubeau, J. Bottieau, F. Vallée, i Z. De Grève. Deep Learning-Based Multivariate Probabilistic Forecasting for Short-Term Scheduling in Power Markets. *IEEE Transactions on Power Systems*, 34(2):1203–1215, 2019.
- S. Trück i K. Liang. Modelling and Forecasting Volatility in the Gold Market. *International Journal of Banking and Finance*, 2012.
- R. S. Tsay. *Analysis of Financial Time Series*. John Wiley & Sons, Chicago, 2010.

- Y. K. Tse i A. K. C. Tsui. A Multivariate GARCH Model with Time-Varying Correlations. SSRN Scholarly Paper ID 250228, Social Science Research Network, Rochester, NY, 2000.
- A. van den Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, i K. Kavukcuoglu. WaveNet: A Generative Model for Raw Audio. *arXiv:1609.03499 [cs]*, 2016.
- A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, i I. Polosukhin. Attention Is All You Need. *arXiv:1706.03762 [cs.CL]*, 2017.
- C. S. Vui, G. K. Soon, C. K. On, R. Alfred, i P. Anthony. A review of stock market prediction with Artificial neural network (ANN). W: *2013 IEEE International Conference on Control System, Computing and Engineering*, s. 477–482, 2013.
- A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, i K. Lang. Phoneme recognition using time-delay neural networks. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37:328–339, 1989.
- D. Warde-Farley, I. J. Goodfellow, A. Courville, i Y. Bengio. An empirical analysis of dropout in piecewise linear networks. *arXiv:1312.6197 [cs, stat]*, 2014.
- P. J. Werbos. Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, 1(4):339–356, 1988.
- J. W. Wilder. *New Concepts in Technical Trading Systems*. Trend Research, Greensboro, N.C, 1978.
- A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, i B. Recht. The Marginal Value of Adaptive Gradient Methods in Machine Learning. *arXiv:1705.08292 [cs, stat]*, 2018.
- T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, i A. Rush. Transformers: State-of-the-Art Natural Language Processing. W: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, s. 38–45, Online, 2020.
- J. Wróblewska i A. Pajor. One-period joint forecasts of Polish inflation, unemployment and interest rate using Bayesian VEC-MSF models. *Central European Journal of Economic Modelling and Econometrics*, 11(1):23–45, 2019.

- Y. Xu i R. Goodacre. On Splitting Training and Validation Set: A Comparative Study of Cross-Validation, Bootstrap and Systematic Sampling for Estimating the Generalization Performance of Supervised Learning. *Journal of Analysis and Testing*, 2(3):249–262, 2018.
- D. Yang i Q. Zhang. Drift-Independent Volatility Estimation Based on High, Low, Open, and Close Prices. *The Journal of Business*, 73(3):477–492, 2000.
- J. Yang, Y. Li, X. Chen, J. Cao, i K. Jiang. Deep Learning for Stock Selection Based on High Frequency Price-Volume Data. *arXiv:1911.02502 [cs, q-fin]*, 2019.
- Y. Yao, L. Rosasco, i A. Caponnetto. On Early Stopping in Gradient Descent Learning. *Constructive Approximation*, 26(2):289–315, 2007.
- J. Yoon, D. Jarrett, i M. van der Schaar. Time-series Generative Adversarial Networks. W: *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.
- S. Yu i Z. Li. Forecasting Stock Price Index Volatility with LSTM Deep Neural Network. W: M. Tavana i S. Patnaik (red.), *Recent Developments in Data Science and Business Analytics*, Springer Proceedings in Business and Economics, s. 265–272, Cham, 2018.
- J.-M. Zakoian. Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955, 1994.
- Q. Zhang, R. Luo, Y. Yang, i Y. Liu. Benchmarking Deep Sequential Models on Volatility Predictions for Financial Time Series. *arXiv:1811.03711 [cs, q-fin, stat]*, 2018a.
- R. Zhang, C. Huang, W. Zhang, i S. Chen. Multi Factor Stock Selection Model Based on LSTM. *International Journal of Economics and Finance*, 10:36, 2018b.
- Z. Zhang, L. Ma, Z. Li, i C. Wu. Normalized Direction-preserving Adam. *arXiv:1709.04546 [cs, stat]*, 2018c.
- K. Zhou, W. Y. Wang, T. Hu, i C. H. Wu. Comparison of Time Series Forecasting Based on Statistical ARIMA Model and LSTM with Attention Mechanism. *Journal of Physics: Conference Series*, 1631(1):012141, 2020.
- Y. Zhou i R. Chellappa. Computation of optical flow using a neural network. *IEEE 1988 International Conference on Neural Networks*, 1988.

M. Zolfaghari i S. Gholami. A hybrid approach of adaptive wavelet transform, long short-term memory and ARIMA-GARCH family models for the stock index prediction. *Expert Systems with Applications*, 182:115149, 2021.

Spis rysunków

3.1	Struktura sieci MLP.	63
3.2	Struktura podstawowej komórki w sieci rekurencyjnej.	65
3.3	Struktura wielowarstwowych (głębokich) sieci rekurencyjnych.	66
3.4	Pełna struktura komórki sieci LSTM.	67
3.5	Pełna struktura sieci GRU.	70
3.6	Operacja konwolucji i funkcja aktywacji w sieci CNN.	72
3.7	Funkcja <i>max pooling</i>	72
3.8	Struktura jednowymiarowej sieci CNN.	73
3.9	Porównanie sieci konwolucyjnych.	74
3.10	Oceny parametrów w klasycznych sieciach neuronowych oraz w sieciach bayesowskich.	77
3.11	Wizualizacja przebiegu poszczególnych algorytmów w przestrzeni parametrów ω	80
3.12	Funkcja wczesnego zatrzymania.	86
3.13	Podział danych na zbiory uczący, testowy oraz walidacyjny.	87
3.14	Wizualizacja podziału danych przy zastosowaniu okna rolowanego oraz okna rozszerzanego.	88
3.15	Przebieg funkcji aktywacji.	93
3.16	Optymalizacja funkcji straty w zależności o wartości współczynnika uczenia.	94
4.1	Schemat modelu hybrydowego ARMA-GARCH-LSTM.	105
4.2	Model probabilistyczny CNN-SSTD, z czterema neuronami wyjściowymi zwracającymi wartości parametrów skośnego rozkładu t-Studenta.	110
5.1	Dzienne notowania wybranych indeksów giełdowych w okresie od 3.01.2000 do końca 31.12.2021 roku.	114
5.2	Logarytmiczne stopy zwrotu wybranych indeksów giełdowych, wyrażone w punktach procentowych.	115
5.3	Kwadraty stóp zwrotu wybranych indeksów giełdowych.	115

5.4	Rozkłady wybranych instrumentów finansowych z naniesionym rozkładem normalnym.	116
5.5	ACF dla stóp zwrotu poszczególnych indeksów.	118
5.6	PACF dla stóp zwrotu poszczególnych indeksów.	118
5.7	ACF dla kwadratów stóp zwrotu.	119
5.8	PACF dla kwadratów stóp zwrotu.	120
5.9	Wykresy punktowe dla par poszczególnych indeksów.	121
5.10	Krzywe kapitałowe dla sieci MLP trenowanych na MSE.	132
5.11	Krzywe kapitałowe dla sieci CNN trenowanych na MSE.	134
5.12	Krzywe kapitałowe dla sieci LSTM trenowanych na MSE.	136
5.13	Krzywe kapitałowe dla sieci LSTM trenowanych na MADL.	139
5.14	Krzywe kapitałowe dla sieci LSTM trenowanych na MADL — probieranie danych w czasie rzeczywistym.	140
5.15	Prognozy zmienności dla poszczególnych aktywów.	144
5.16	Prognozy wartości zagrożonej dla indeksów S&P 500, NIKKEI oraz DAX, uzyskane za pomocą wybranych modeli klasy GARCH oraz modeli hybrydowych GARCH-LSTM.	148
5.17	Prognozy wartości zagrożonej dla indeksów WIG, KOSPI oraz BOVESPA, uzyskane za pomocą wybranych modeli klasy GARCH oraz modeli hybrydowych GARCH-LSTM.	149
5.18	Wykres rozkładów wartości PIT, indeks DAX.	152
5.19	Prognozy wartości zagrożonej z wykorzystaniem modeli prognozowania probabilistycznego dla indeksów S&P 500, NIKKEI oraz DAX.	155
5.20	Prognozy wartości zagrożonej z wykorzystaniem modeli prognozowania probabilistycznego dla indeksów WIG, KOSPI oraz BOVESPA.	156
A.1	ACF dla wartości bezwzględnych stóp zwrotu	181
A.2	PACF dla wartości bezwzględnych stóp zwrotu	181
A.3	ACF dla oszacowań zmienności uzyskanych za pomocą estymatora GKYZ182	
A.4	PACF dla oszacowań zmienności uzyskanych za pomocą estymatora GKYZ182	

Spis tabel

2.1	Miary trafności prognoz punktowych.	46
2.2	Mierniki oceny strategii inwestycyjnej	48
3.1	Podstawowe funkcje aktywacji.	93
4.1	Wybrane specyfikacje modeli klasy GARCH i oraz użyte typy rozkładów warunkowych.	104
5.1	Opis statystyczny danych.	117
5.2	Wartości współczynników korelacji Pearsona	121
5.3	Dostrajanie hiperparametrów dla sieci wykorzystywanych w prognozowaniu punktowym stóp zwrotu.	126
5.4	Prognozowanie punktowe stóp zwrotu: mierniki błędu dla sieci trenowanych z wykorzystaniem MSE	129
5.5	Prognozowanie punktowe stóp zwrotu: wyniki strategii dla sieci MLP trenowanych na MSE	131
5.6	Prognozowanie punktowe stóp zwrotu: wyniki strategii dla sieci CNN trenowanych na MSE	133
5.7	Prognozowanie punktowe stóp zwrotu: wyniki strategii dla sieci LSTM trenowanych na MSE	135
5.8	Prognozowanie punktowe stóp zwrotu: wyniki strategii dla sieci LSTM trenowanych na MADL	138
5.9	Prognozowanie punktowe stóp zwrotu: wyniki strategii dla sieci LSTM trenowanych na MADL — pobieranie danych w czasie rzeczywistym.	140
5.10	Oceny trafności prognoz zmienności — wartości MSE.	142
5.11	Współczynniki korelacji Pearsona pomiędzy najlepszymi specyfikacjami modeli GARCH i GARCH-LSTM dla poszczególnych aktywów	144
5.12	Procentowe udziały przekroczeń oszacowań $VaR(0,05)/VaR(0,01)$ w modelach ARMA-GARCH i hybrydowych ARMA-GARCH-LSTM	146
5.13	Oceny prognoz uzyskanych za pomocą sieciowych modeli prognozowania probabilistycznego	151

5.14	Procentowe udziały przekroczeń oszacowań $\text{VaR}(0,05)/\text{VaR}(0,01)$ w modelach prognozowania probabilistycznego.	153
6.1	Porównanie modeli punktowych prognoz stóp zwrotu, pod względem wskaźników oceny strategii inwestycyjnych.	159
6.2	Zestawienie modeli, indeks S&P 500, najlepsze modele pod względem MSE i ilości przekroczeń VaR	161
6.3	Zestawienie modeli, indeks NIKKEI 225, najlepsze modele pod względem MSE i ilości przekroczeń VaR	162
6.4	Zestawienie modeli, indeks DAX, najlepsze modele pod względem MSE i ilości przekroczeń VaR	163
6.5	Zestawienie modeli, indeks WIG, najlepsze modele pod względem MSE i ilości przekroczeń VaR	164
6.6	Zestawienie modeli, indeks KOSPI, najlepsze modele pod względem MSE i ilości przekroczeń VaR	165
6.7	Zestawienie modeli, indeks BOVESPA, najlepsze modele pod względem MSE i ilości przekroczeń VaR	166
6.8	Porównanie modeli prognozowania probabilistycznego z modelami klasy AR-GARCH pod kątem oceny trafności prognoz	169
6.9	Porównanie modeli prognozowania probabilistycznego z modelami hybrydowymi oraz GARCH — procentowy udział przekroczeń VaR dla wszystkich aktywów.	170
A.1	Ilość przekroczeń oszacowań $\text{VaR}(0,05)/\text{VaR}(0,01)$ w modelach ARMA-GARCH i hybrydowych ARMA-GARCH-LSTM	183
A.2	Prognozy zmienności, indeks S&P500, model GARCH i GARCH LSTM	184
A.3	Prognozy zmienności, indeks S&P500, model EGARCH i EGARCH-LSTM	184
A.4	Prognozy zmienności, indeks S&P500, model GJR-GARCH i GJR-GARCH-LSTM	185
A.5	Prognozy zmienności, indeks S&P500, model APARCH i APARCH-LSTM	185
A.6	Prognozy zmienności, indeks NIKKEI 225, model GARCH i GARCH-LSTM	186
A.7	Prognozy zmienności, indeks NIKKEI 225, model EGARCH i EGARCH-LSTM	186
A.8	Prognozy zmienności, indeks NIKKEI 225, model GJR-GARCH i GJR-GARCH-LSTM	187
A.9	Prognozy zmienności, indeks NIKKEI 225, model APARCH i APARCH-LSTM	187
A.10	Prognozy zmienności, indeks DAX, model GARCH i GARCH-LSTM .	188

A.11 Prognozy zmienności, indeks DAX, model EGARCH i EGARCH-LSTM	188
A.12 Prognozy zmienności, indeks DAX, model GJR-GARCH i GJR-GARCH-LSTM	189
A.13 Prognozy zmienności, indeks DAX, model APARCH i APARCH-LSTM	189
A.14 Prognozy zmienności, indeks WIG, model GARCH i GARCH-LSTM .	190
A.15 Prognozy zmienności, indeks WIG, model EGARCH i EGARCH-LSTM	190
A.16 Prognozy zmienności, indeks WIG, model GJR-GARCH i GJR-GARCH-LSTM	191
A.17 Prognozy zmienności, indeks WIG, model APARCH i APARCH-LSTM	191
A.18 Prognozy zmienności, indeks KOSPI, model GARCH i GARCH-LSTM	192
A.19 Prognozy zmienności, indeks KOSPI, model EGARCH i EGARCH-LSTM	192
A.20 Prognozy zmienności, indeks KOSPI, model GJR-GARCH i GJR-GARCH-LSTM	193
A.21 Prognozy zmienności, indeks KOSPI, model APARCH i APARCH-LSTM	193
A.22 Prognozy zmienności, indeks BVP, model GARCH i GARCH-LSTM . .	194
A.23 Prognozy zmienności, indeks BVP, model EGARCH i EGARCH-LSTM	194
A.24 Prognozy zmienności, indeks BVP, model GJR-GARCH i GJR-GARCH- LSTM	195
A.25 Prognozy zmienności, indeks BVP, model APARCH i APARCH-LSTM	195
A.26 Ilość przekroczeń oszacowań $\text{VaR}(0,05)/\text{VaR}(0,01)$ w sieciowych modelach prognozowania probabilistycznego	196
A.27 Prognozy probabilistyczne, indeks S&P500, sieci CNN	197
A.28 Prognozy probabilistyczne, indeks S&P500, sieci LSTM	197
A.29 Prognozy probabilistyczne, indeks NIKKEI, sieci CNN	198
A.30 Prognozy probabilistyczne, indeks NIKKEI, sieci LSTM	198
A.31 Prognozy probabilistyczne, indeks DAX, sieci CNN	199
A.32 Prognozy probabilistyczne, indeks DAX, sieci LSTM	199
A.33 Prognozy probabilistyczne, indeks WIG, sieci CNN	200
A.34 Prognozy probabilistyczne, indeks WIG, sieci LSTM	200
A.35 Prognozy probabilistyczne, indeks KOSPI, sieci CNN	201
A.36 Prognozy probabilistyczne, indeks KOSPI, sieci LSTM	201
A.37 Prognozy probabilistyczne, indeks BOVESPA, sieci CNN	202
A.38 Prognozy probabilistyczne, indeks BOVESPA, sieci LSTM	202