

**ANALIZA CZASÓW TRWANIA POMIĘDZY
ZMIANAMI KIERUNKU CEN AKCJI
— WPŁYW UWZGLĘDNIENIA WEWNĄTRZDZIENNEJ
SEZONOWOŚCI NA RANKING MODELI ACD¹**

JUSTYNA BIAŁKOWSKA²

e-mail: justyna_bialkowska@dell.com

MATEUSZ PIPIEŃ

Uniwersytet Ekonomiczny w Krakowie, Katedra Ekonometrii i Badań Operacyjnych

e-mail: eepipien@cyf-kr.edu.pl

ABSTRACT

J. Białkowska, M. Pipień. *Analysis of duration between changes in direction of trend for the price processes — the impact of seasonal adjustment to explanatory power of a class of ACD models*. Folia Oeconomica Cracoviensia 2015, 56: 35–80.

The main purpose of the paper was to investigate the properties of a particular type of price duration process, analysed on the field of modelling transaction data. We considered duration between changes in direction of trend for the price processes and consider the problem of stability of the relative explanatory power of a class of ACD models with respect to the seasonality adjustment.

We report model ranking on the basis of information criteria and discuss its sensitivity with respect to the seasonality adjustment procedure elaborated on the basis of nonparametric Nadaraya and Watson regression.

STRESZCZENIE

Przedmiotem badań są dane o wysokiej częstotliwości opisujące kształtowanie się czasu trwania zmiany kierunku cen akcji. Do analizy odstępów czasu pomiędzy dwoma kolejnymi zdarzeniami transakcyjnymi, polegającymi na zmianie trendu cenowego akcji, wykorzystuje się modele warunkowego czasu trwania (ang. Autoregressive Conditional Duration, ACD). Jednym z celów niniejszej pracy jest zbudowanie rankingu modeli ACD ze względu na ich jakość dopasowania.

¹ Badania finansowane przez Narodowe Centrum Nauki w ramach grantu OPUS, numer grantu DEC-2013/09/B/H54/01945.

² Studentka Wydziału Zarządzania Uniwersytetu Ekonomicznego w Krakowie.

W artykule zbadano wpływ procedury odsezonowania na zmianę pozycji modeli w rankingu. W celu uwzględnienia efektu wewnątrzdziennej sezonowości zastosowano nieparametryczną regresję Nadaraya i Watsona z funkcją jądrową normalną. Aby potwierdzić poprawność zbudowanych rankingów dla danych odsezonowanych i tych bez efektu cykliczności, przeprowadzono weryfikację statystyczną przy pomocy testów t-Studenta oraz testu ilorazu wiarygodności.

KEY WORDS — SŁOWA KLUCZOWE

intraday seasonality, ACD models, modelling tick data

wewnątrzdzienne sezonowość, modele ACD, modelowanie danych transakcyjnych

1. WPROWADZENIE

Od ostatnich czterdziestu lat obserwuje się dynamiczny rozwój ekonometrii finansowej spowodowany przede wszystkim postępem technologicznym i zwiększającą się dostępnością danych pochodzących z rynków finansowych. Dzięki automatyzacji systemów transakcyjnych zbiory danych, opisujące proces transakcyjny, stały się bardziej szczegółowe i jednocześnie łatwo osiągalne. Klasyczne podejście do badań w ekonometrii zakładało analizę danych o regularnych odstępach, np. szeregów obserwacji dziennych lub godzinowych, wykorzystywanych współcześnie w ekonometrii finansowej. Jednak powszechnie dostępne stały się bazy danych zawierające dane o wysokiej częstotliwości, czy tzw. dane tikowe. Dane te charakteryzują się nieregularnym rozmieszczeniem na osi czasu, odpowiadającym chwilom zajścia danego zdarzenia w procesie transakcyjnym. Mikrostruktura rynku określa zbiór cech charakterystycznych i mechanizmów rynku finansowego decydujących o tym w jaki sposób i na jakich warunkach zawierane są transakcje. Badanie mikrostruktury ma na celu uwzględnienie rzeczywistych warunków panujących na danej giełdzie, gdyż zastosowana architektura na rynku finansowym wpływa na własności pozyskiwanych danych, zwłaszcza danych o wysokiej częstotliwości. Do analizy danych tikowych stosuje się modele w czasie transakcyjnym.

Przedmiotem badań w artykule są dane o wysokiej częstotliwości, które opisują kształtowanie się czasu trwania zmiany kierunku cen akcji. Na uwagę zasługuje fakt, że analizy takiego zbioru danych rzadko występują w literaturze przedmiotu. Do analizy odstępów czasu pomiędzy dwoma kolejnymi zdarzeniami transakcyjnymi, polegającymi na zmianie trendu cenowego akcji, wykorzystane będą modele warunkowego czasu trwania (ang. *Autoregressive Conditional Duration*). Modele ACD są obecnie podstawowym narzędziem ekonometrycznym wykorzystywanym do analizy intensywności transakcyjnej oraz badania efektów mikrostruktury rynku. Jednym z celów niniejszej pracy jest zbudowanie rankingu modeli ACD ze względu na ich jakość dopasowania do danych. Badania mają na celu wskazanie która specyfikacja modelu jest

najlepsza oraz jaki rozkład warunkowy należy wybrać. Kolejnym celem jest zbadanie wpływu odsezonowania danych na potencjalną zmianę pozycji modeli w rankingu. Do eliminacji efektu wewnątrzdziennej sezonowości zastosowana zostanie nieparametryczna regresja Nadaraya i Watsona z funkcją jądrową normalną. Aby potwierdzić poprawność zbudowanych rankingów dla danych odsezonowanych i tych bez efektu cykliczności, przeprowadzono weryfikację statystyczną przy pomocy testów *t*-Studenta oraz testu ilorazu wiarygodności.

W artykule opisano pojęcie mikrostruktury rynku jako podstawę badań nad naturą dynamiki danych transakcyjnych. Omówiono w tej kwestii stosowane mechanizmy na rynkach finansowych a w szczególności te, które zastosowano na Giełdzie Papierów Wartościowych w Warszawie. W dalszej części omówiono własności danych o wysokiej częstotliwości oraz metody pomiaru intensywności procesów transakcyjnych. W rozdziale 6 zawarto definicje modeli ACD i ich typy, sposoby estymacji parametrów zastosowanych modeli, opis kryteriów informacyjnych, na podstawie których budowane są rankingi, oraz zarys metody estymacji jądrowej — regresji Nadaraya–Watsona wykorzystanej do eliminacji efektu sezonowości danych. Zawarte w części empirycznej porównanie rankingów rozważanych modeli ukaże czy przeprowadzanie odsezonowania danych transakcyjnych jest konieczne, by poprawnie wnioskować o czasie trwania kierunku zmiany ceny.

2. MIKROSTRUKTURA RYNKU JAKO PODSTAWA BADAŃ NAD NATURĄ DYNAMIKI DANYCH TRANSAKCYJNYCH

Pojęcie mikrostruktury rynku związane jest z dziedziną nauk ekonomicznych zajmujących się procesami transakcyjnymi oraz ich wynikami, przy formalnie ustalonych regułach handlowych; O'Hara (1995). Mikrostrukturę giełdy można zatem zdefiniować jako zespół charakterystyk oraz mechanizmów, które wpływają na proces kształtowania się cen instrumentów finansowych, a także oddziałują na ustalenie warunków i czasu trwania zawarcia transakcji. Badanie mikrostruktury rynku stanowi obecnie dział empirycznych finansów analizujący proces transakcyjny, w którym nieobserwowalne potrzeby inwestorów odzwierciedlane są w cenach transakcyjnych i wolumenie; zob. Madhavan (2000). Klasyczne modele matematyczne opisujące formowanie się cen na rynkach finansowych nie uwzględniają szeregu czynników związanych ze strukturą organizacyjną lub techniczną, czy z obowiązującymi zasadami obrotu dóbr. Takie silne uproszczenia prowadzą do zniekształconego obrazu rynku, za czym idzie podejmowanie nietrafnych decyzji przez inwestorów. Analiza mikrostruktury ma na celu uwzględnienie rzeczywistego funkcjonowania rynku, pokrywa zatem wszystkie te zagadnienia i zjawiska, które są zanedbywane w modelach matematycznych.

Głównymi aspektami mikrostruktury rynku są przede wszystkim: jego płynność, głębokość i efektywność — Doman (2011). Płynność rynku określana jest jako natychmiastowa możliwość kupna lub sprzedaży przez gracza prawie każdej liczby akcji, przy czym mała liczba instrumentów może być zawsze sprzedana lub kupiona po cenie bardzo zbliżonej do aktualnej ceny rynkowej, natomiast duży pakiet akcji może być w długim okresie kupowany lub sprzedawany po cenach, których średnia jest zbliżona do ceny rynkowej. Głębokość rynku wiąże się z jego zdolnością do absorbowania dużych zleceń bez znaczącego wpływu na cenę. Efektywność rynku jest najczęściej analizowana pod kątem informacyjnym — rynek jest efektywny, gdy nowa informacja na temat danej spółki jest szybko i w sposób racjonalny całkowicie odzwierciedlana w cenach jej akcji. Rozwiązania dotyczące handlu stosowane na danym rynku wpływają na wymienione własności, a co za tym idzie na pozyskiwane z niego dane o wysokiej częstotliwości, dlatego też stanowią podstawę badań mikrostruktury rynku. W analizie procesu transakcyjnego należy również brać pod uwagę efekty mikrostruktury, takie jak niesynchroniczny handel czy asymetria informacji. Niesynchroniczny handel związany jest z tym, że transakcje czy zmiany cen instrumentów finansowych nie dokonują się w regularnych odstępach czasu, a także w tym samym momencie istnieje możliwość zawierania transakcji na tym samym instrumencie po różnych cenach. Asymetria informacji na rynku ma miejsce, ponieważ niektórzy inwestorzy oprócz publicznych informacji, które można pozyskać z obserwacji sytuacji na giełdzie, mogą mieć dostęp do prywatnych wiadomości o spółce, które dają im strategiczną przewagę nad resztą graczy.

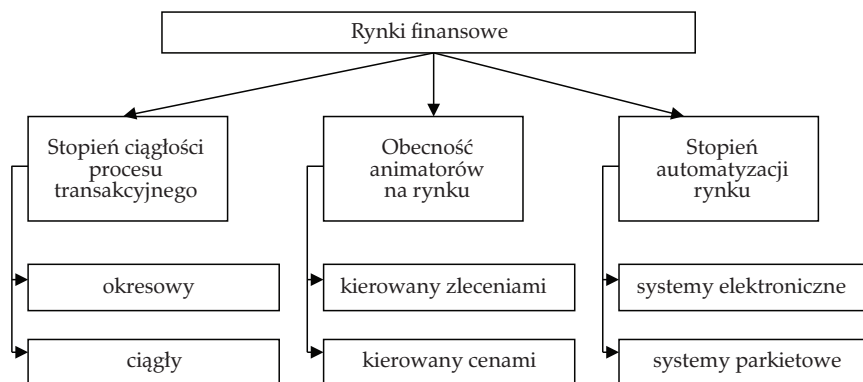
Mechanizmy, które zachodzą pomiędzy określonymi charakterystykami procesu wymiany dóbr determinowane są formalnymi regułami handlowymi oraz pewnymi czynnikami instytucjonalnymi. Zapoznanie się z stosowanymi rozwiązaniami na rynkach finansowych jest kluczowe w poprawnej analizie procesu transakcyjnego.

3. ARCHITEKTURA RYNKÓW FINANSOWYCH — PRZEGLĄD ROZWIĄZAŃ

Rynki finansowe można podzielić ze względu na panujące na nich reguły zawierania transakcji. Jedną z najpopularniejszych klasyfikacji dzieli je według następujących trzech kryteriów: stopień ciągłości procesu transakcyjnego, obecność animatorów na rynku oraz stopień automatyzacji i przejrzystość rynku; Madhavan (2000). Obrazuje to Rysunek 1.

Stopień ciągłości procesu zawierania transakcji wskazuje formalne ustalenia co do okresu, w którym obrót instrumentami finansowymi jest dozwolony. Istnieją dwa rodzaje rynków: okresowe i ciągłe. W rynkach okresowych, zwanych

także wywoławczymi, zawarcie transakcji odbywa się tylko w ściśle określonym momencie. Prowadzona jest na nich aukcja jednej ceny, której pierwowzorem była tzw. aukcja warlasowska. Schemat aukcji warlasowskiej jest następujący: w pierwszej fazie prowadzący aukcję zbiera od inwestorów oferty kupna i sprzedazy na dany instrument finansowy i publicznie ogłasza pierwszą cenę, po której możliwe jest zawarcie transakcji. W następnej fazie inwestorzy znając wycenę aukcjoneru dokonują rewizji ofert i ponownie składają zlecenia. Proces ten trwa do momentu, gdy już żaden inwestor nie będzie chciał rewidować swojej oferty — czyli ustalona zostanie cena równowagi, przy której popyt zrówna się podażą. Po ustaleniu tej ceny zawarcie transakcji staje się możliwe. Równowaga na rynku jest wynikiem przeprowadzania serii cząstkowych aukcji. Obecnie mechanizm aukcji warlasowskiej nie ma częstego zastosowania na rynkach finansowych, jednak procedury ustalania ceny otwarcia lub zamknięcia na niektórych giełdach wykorzystują w pewnym zakresie opisany schemat. Aukcje jednej ceny efektywnie agregują niejednorodną informację wprowadzaną przez uczestników procesu transakcyjnego, dlatego też stosowane są gdy występuje niepewność, co do prawdziwej wartości aktywa. Rynki wywoławcze uznawane są za systemy, które redukują koszty i szum informacyjny dzięki czemu działają w sposób stabilizujący ceny; zob. Doman (2011).



Źródło: opracowanie własne na podstawie Madhavan (2000).

Rysunek 1. Klasyfikacja rynków finansowych

Na rynkach ciągłych zawieranie transakcji możliwe jest nieprzerwanie w sposób ciągły, w godzinach otwarcia danego rynku. System taki sprawdza się w przypadku instrumentów finansowych o znacznej płynności, gdy częstotliwość dokonanych na nich transakcji jest wysoka. Aby system notowań ciągłych funkcjonował sprawnie potrzebna jest duża liczba aktywnych w tym samym momencie inwestorów, co skutkuje ograniczeniem wysokich skoków w cenach instrumentów. Przykładem rynku ciągłego jest otwarty całą dobę

międzybankowy rynek walutowy. Charakterystyczne dla tego systemu jest to, że małe transakcje zawierane są w nim automatycznie, natomiast ceny dużych pakietów instrumentów finansowych są negocjowane telefonicznie. Niewątpliwą zaletą rynków ciągłych jest fakt, że zlecenia realizowane są natychmiast, kiedy jest to możliwe, oraz, że proces cenotwórczy może być obserwowany przez graczy na bieżąco. Współcześnie rynki na świecie łączą te dwa sposoby zawierania transakcji. Często dla systemów notowań ciągłych stosuje się aukcję jednej ceny po to, by ustalić cenę otwarcia lub zamknięcia rynku, a także w przypadku, gdy rynek wznawia pracę po okresie zawieszenia albo znacznego spadku aktywności.

Kryterium obecności animatorów na rynku dzieli systemy transakcyjne na dwa rodzaje: rynki kierowane cenami i rynki kierowane zleceniami. Na rynkach kierowanych cenami, zwanych również rynkami dealerskimi, główną rolę odgrywa *market maker* — specjalista, którego podstawową funkcją jest „tworzenie” rynku instrumentów finansowych. *Market maker* realizuje tę funkcję przede wszystkim poprzez kwotowanie cen kupna (*bid*) i sprzedaży (*ask*), po których inwestorzy będą realizować transakcje. Nazywany jest on także dostawcą płynności, gdyż celem jego działalności jest utrzymywanie jak najwyższego poziomu płynności instrumentów, czyli zapewnienie możliwości ciągłego zawierania transakcji. Charakterystyczne dla rynku dealerskiego jest to, że gracze handlują tylko z *market makerem* po ustalonych przez niego cenach — nie mają możliwości dokonywania transakcji bezpośrednio między sobą. *Market maker* zawiera transakcje na własny rachunek, naraża się zatem na ryzyko zapasu, powstałego między momentem kupna i sprzedaży instrumentu finansowego i związanego z ewentualnym spadkiem wartości danego aktywa. Rekompensatą i zarazem głównym źródłem dochodu *market makera* jest tzw. *bid-ask spread*, czyli różnica pomiędzy kwotowanymi cenami kupna i sprzedaży. Rynki kierowane cenami wyróżniają się tym, że w razie nagłego zapotrzebowania na płynność klienci nie narażają się na ryzyko niezrealizowania transakcji — Bień (2005).

Rynki kierowane zleceniami różnią się od dealerskich przede wszystkim tym, że nie występuje na nich figura *market makera*. Mogą jednak być obecni animatorzy rynku, którzy spełniają funkcję upłynniania rynku, czyli zobowiązują się do sprzedaży instrumentów, na które jest popyt, oraz kupna sprzedawanych przez inwestorów pozycji. Oferty składane przez graczy kojarzone są bezpośrednio, następnie zostają one wysyłane za pośrednictwem domów maklerskich do realizacji. Ceny, po których dokonywane są transakcje, ustalane zostają w chwili pojawienia się oferty albo z pewnym opóźnieniem, zależnie od sytuacji i typu zlecenia.

Na rynku tym istnieją różne rodzaje zleceń, ponieważ inwestorzy mogą kłaść nacisk na rozmaite aspekty realizacji transakcji. Można zatem wyróżnić: zlecenia z limitem ceny oraz zlecenia bez limitu ceny.

Zlecenia z limitem ceny charakteryzują się tym, że inwestor określa cenę, po jakiej jest gotowy zrealizować transakcję. Gdy chce kupić dany instrument finansowy precyzuje ile jest w stanie najwięcej za niego dać, natomiast w przypadku sprzedaży — wyznacza minimalną cenę jaką chce uzyskać za ten instrument. Kształtują one zatem *bid-ask spread*. Zlecenie jest realizowane w sytuacji gdy zaistnieje zgodność cen przeciwstawnych zleceń. Oznacza to, że inwestor jest pewny ceny, jednak nie będzie wiedział kiedy dokładnie jego zlecenie zostanie zrealizowane. Zlecenie z limitem uzyskuje status oczekującego, dopóki nie pojawi się przeciwstawna oferta — jest niejako „zamrożone”, gdyż jego charakterystyki nie będą dostosowywać się do napływającej na rynek informacji. W przypadku wystąpienia paru ofert z tym samym limitem ceny wykorzystywany jest priorytet czasu, czyli zlecenie, które pojawiło się najwcześniej, realizowane jest najpierw.

Zlecenia bez limitu ceny składane są przede wszystkim, gdy inwestorowi zależy na szybkiej realizacji transakcji. Transakcje zawierane są natychmiast, jednak z racji tego, że cena nie jest w nich sprecyzowana, gracze narażeni są na ryzyko kursu transakcji. Inwestor musi mieć świadomość, że po wysłaniu oferty na rynek cena realizacji transakcji może się zmienić na niekorzyść tuż przed sfinalizowaniem zlecenia. Można wyróżnić dwa rodzaje zleceń bez limitu ceny: zlecenia po cenie rynkowej i zlecenia po każdej cenie. Zlecenie po cenie rynkowej dokonuje się bezpośrednio po jego złożeniu, z najlepszą oczekującą ofertą, która jest do niego przeciwstawna — a zatem po najlepszej cenie rynkowej. Jeśli zlecenie nie jest realizowane w całości, pozostała część przechodzi automatycznie na zlecenie z limitem ceny, która odpowiada cenie poprzedniej transakcji. Z kolei zlecenie po każdej cenie realizowane jest po kursie, który aktualnie obowiązuje na rynku. Zawsze musi być ono zrealizowane w całości — Bień (2005).

Ostatni wymiar dotyczący rozważanych rozwiązań instytucjonalnych na rynku obejmuje jego stopień zautomatyzowania. Główną funkcją rynku zawsze było umożliwienie i ułatwienie jego uczestnikom spotkania się i dokonania transakcji, jednak na przestrzeni lat metody prowadzące do sfinalizowania wymiany handlowej ulegały znacznym zmianom. Obecnie na rynkach finansowych wyróżnić można systemy zautomatyzowane oraz systemy parkietowe. Rynki z elektronicznym zawieraniem transakcji charakteryzują się tym, że inwestorzy nie kontaktują się ze sobą, a proces transakcyjny przeprowadzany jest przez systemy komputerowe realizujące napływające zlecenia. Aktualnie prawie wszystkie rynki korzystają z tego rozwiązania dzięki jego licznym zaletom. Jeśli jest to możliwe, zlecenia realizowane są natychmiastowo, a obrót instrumentami finansowymi jest efektywniejszy, gdyż rośnie wydajność realizowanych transakcji i maleją ich koszty, przy jednoczesnej minimalizacji ryzyka błędów operacyjnych popełnianych przez człowieka. Systemy w pełni zautomatyzowane

wane występują przede wszystkim na rynkach kierowanych zleceniami, choć na rynkach dealerskich także wprowadza się rozwiązania elektroniczne.

Systemy parkietowe natomiast cechują się tym, że zawarcie transakcji następuje za pośrednictwem maklera, który jest fizycznie obecny na parkiecie giełdy. Maklerzy negocjują jak najlepsze warunki realizacji transakcji dla swoich klientów (inwestorów).

Dzięki automatyzacji wymiany handlowej i sieciom teleinformatycznym, inwestorzy z całego świata mają możliwość zdalnego dostępu do informacji dotyczących procesu transakcyjnego i cenowego na danym rynku. Szczegółowość tych danych określa stopień przejrzystości rynku. Przejrzystość rynku oznacza, że gracze mają dostęp do informacji o przebiegu handlu na giełdzie; Doman (2011). Informacje te mogą dotyczyć danych, które są dostępne przed zawarciem transakcji, a więc aktualnej sytuacji na rynku. Są to na przykład: oferowane ceny i wolumeny zleceń, tożsamość inwestorów, wiadomości o notowanych spółkach. Drugi typ informacji, które gracz może przyswoić dotyczy już przeprowadzonych transakcji — wielkość i kierunek zrealizowanych zleceń czy personalia inwestorów; zob. Madhavan (2000). Często uważa się rynek za przejrzysty, gdy można obserwować strumień zleceń, czyli wielkość, kierunek i moment pojawiania się zleceń, typ zlecenia, czas jego realizacji. Ważnym aspektem jest także informacja dla kogo udostępniane są te dane i w jakim zakresie. Przejrzystość rynku ma duże znaczenie, gdyż wpływa na strategie inwestycyjne graczy, a co za tym idzie — na proces kształtowania się cen. Rynki kierowane cenami mają stosunkowo niską przejrzystość, natomiast kierowane zleceniami uznaje się za przejrzyste zarówno w zakresie informacji dostępnych przed- jak i po zawarciu transakcji.

Przejrzystość rynku daje inwestorom możliwość obserwacji sytuacji na rynku, jednak dostęp do informacji jest niesymetryczny. Istnieje naturalny podział inwestorów na: poinformowanych i płynnościowych; Doman (2011). Decyzje inwestorów poinformowanych opierają się na ocenie fundamentalnej wartości instrumentu, na podstawie prawidłowo wyciągniętych wniosków z informacji prywatnych lub publicznych. Inwestorzy płynnościowi natomiast nie mają dostępu do informacji lub nie potrafią jej wykorzystać, albo muszą zawierać transakcje z przyczyn płynnościowych. Poinformowani inwestorzy są bardzo ważni na rynku — dzięki ich aktywności cena instrumentu finansowego staje się informacyjna. Gdy oceniają jego cenę jako niedoszacowaną, to dokonują kupna tego instrumentu, natomiast gdy uznają ją za przeszacowaną, to go sprzedają. W ten sposób popychają cenę danego instrumentu do jej wartości fundamentalnej. Niepoinformowani podejmują decyzje, które umożliwiają realizację zleceń złożonych przez poinformowanych i w ten sposób wspierają płynność. Zakłada się, że niepoinformowani inwestorzy wchodzą na rynek w sposób losowy³. Poinformowani zaś tylko wtedy, gdy pojawi się

³ Poddany rozkładowi Poissona.

prywatna wiedza, sygnał. Specjaliści rynkowi powoli przyswajają prywatną informację, poprzez obserwację przepływu zleceń i zgodnie z tym dostosowują ceny. Poinformowani inwestorzy będą dokonywać transakcji tak długo jak ich informacja ma jakąś wartość (zanim specjaliści się jej nauczą). Stąd można zaobserwować grupowanie się intensywności transakcji po zdarzeniu pojawienia się informacji, gdyż wtedy zwiększa się liczba inwestorów poinformowanych; zob. Engle, Russel (1998).

4. MECHANIZMY TRANSAKCYJNE STOSOWANE NA GPW W WARSZAWIE

Giełda Papierów Wartościowych w Warszawie S.A. (GPW) rozpoczęła swą działalność 16 kwietnia 1991 roku jako jednoosobowa spółka Skarbu Państwa. Obecnie jest ona największą krajową giełdą instrumentów finansowych w Europie Środkowej i Wschodniej, uznawana jest także za jedną z najszybciej rozwijających się giełd w Europie. Przedmiotem jej funkcjonowania jest organizowanie oraz promowanie publicznego obrotu papierami wartościowymi, a także upowszechnianie informacji z rynku. Giełda zapewnić ma zatem koncentrację w jednym miejscu i czasie ofert kupujących i sprzedających w celu wyznaczenia kursu i realizacji transakcji. *Misją GPW jest kontynuacja budowy polskiego rynku kapitałowego i wzmocnienie polskiej gospodarki przez zapewnienie, aby GPW była mechanizmem służącym w jak największym stopniu mobilizacji kapitału dla przedsiębiorstw. Jednocześnie GPW powinna osiągać wskaźniki ekonomiczne satysfakcjonujące akcjonariuszy GPW i pozwalające jej plasować się lepiej na tle innych podobnych giełd, działających na rynkach wschodzących* (raportroczny2012.gpw.pl, odczyt z dn. 14.05.2014 r.). Tabela 1. ukazuje, jak w przeciągu kilku lat zmieniały się podstawowe charakterystyki giełdy.

Tabela 1

Podstawowe statystyki GPW

Rok	Kapitalizacja spółek krajowych (mln zł)	Liczba spółek	Obroty akcjami (mln zł)
2013	593 464,45	450	256 147,00
2012	523 390,23	438	202 880,00
2011	446 151,22	426	268 139,00
2010	542 646,12	400	234 288,00
2009	421 178,00	379	175 943,00

Źródło: opracowanie własne na podstawie www.gpw.pl (odczyt z dn. 14.05.2014 r.).

Podstawy prawne GPW obejmują 5 dokumentów: kodeks spółek handlowych, ustawa o obrocie instrumentami finansowymi, ustawa o ofercie publicznej, warunkach wprowadzania instrumentów finansowych do zorganizowanego systemu obrotu oraz o spółkach publicznych, statut Giełdy Papierów Wartościowych w Warszawie S.A. oraz Regulamin Giełdy Papierów Wartościowych w Warszawie S.A.

GPW prowadzi rynek regulowany dla akcji i instrumentów pochodnych, alternatywny rynek akcji dla spółek wzrostowych — NewConnect, rynek przeznaczony dla emitentów obligacji korporacyjnych i komunalnych — Catalyst, a także rynek energii. Z dniem 23 czerwca 2008 roku na GPW wprowadzono Program Wspierania Płynności, który powstał z myślą o spółkach cechujących się niską płynnością, a jego celem jest poprawa płynności obrotu akcjami oraz polepszenie komunikacji z inwestorami danej spółki. 9 listopada 2010 r. GPW stała się spółką publiczną notowaną na Giełdzie Papierów Wartościowych w Warszawie.

GPW jest zautomatyzowanym rynkiem kierowanym zleceniami — kursy poszczególnych instrumentów finansowych są ustalane na podstawie zleceń kupujących i sprzedających. Oznacza to, że w celu ustalenia ceny instrumentu sporządza się zestawienie zleceń zawierających dyspozycje kupna i sprzedaży. Kojarzenia tych zleceń dokonuje się według ściśle określonych na giełdzie zasad, zaś realizacja transakcji odbywa się w trakcie sesji giełdowych. Obrót instrumentami finansowymi odbywa się w systemie notowań jednolitych i ciągłych.

W systemie ciągłym notowane są akcje spółek o najwyższej płynności, jak i również: obligacje, opcje, certyfikaty inwestycyjne, kontrakty terminowe, jednostki indeksowe, ETF-y i warranty. Harmonogram sesji giełdowej w tym systemie przedstawia Tabela 2⁴.

W systemie kursu jednolitego dokonuje się obrotu akcjami, których płynność jest stosunkowo niska. Notowania te bazują na procedurze tzw. fixingu — jest to określenie ceny papieru wartościowego w oparciu o zlecenia, które zostały złożone przed rozpoczęciem notowań. Tabela 3. prezentuje obowiązujący harmonogram sesji giełdowej notowań jednolitych.

Poprzednim systemem komputerowym, w którym prowadzone były notowania na GPW, był system transakcyjny WARSET wprowadzony w listopadzie 2000 roku. Był on modyfikacją i adaptacją do polskich warunków systemu NSC giełdy paryskiej. Zapewniał pełną automatyzację procesu transakcyjnego, a także zdalny dostęp do systemu obrotu instrumentami finansowymi oraz informacji o sytuacji na rynku dla inwestorów. Od 15 kwietnia 2013 roku warszawska giełda funkcjonuje w oparciu o nowy system transakcyjny UTP (*Universal Trading Platform*), stworzonego przez NYSE Technologies. Tę samą platformę

⁴ Zamieszczono harmonogram sesji z roku 2009, ponieważ analiza empiryczna niniejszej pracy została przeprowadzona na danych z tego roku.

Tabela 2

Harmonogram sesji giełdowej na rynku kasowym w systemie notowań ciągłych

Rok 2009		Rok 2014	
8.00–9.00	Przyjmowanie zleceń na otwarcie	8.30–9.00	Faza przed otwarciem
9.00	Faza otwarcia (określanie kursu na otwarciu)	9.00	Faza otwarcia (określanie kursu na otwarciu)
9.00–16.10	Faza notowań ciągłych	9.00–16.50	Faza notowań ciągłych
16.10–16.20	Przyjmowanie zleceń na zamknięcie	16.50–17.00	Faza przed zamknięciem
16.20	Faza zamknięcia (określenie kursu na zamknięciu)	17.00	Faza zamknięcia (określenie kursu na zamknięciu)
16.20–16.30	Faza dogrywki	17.00–17.05	Faza dogrywki

Źródło: opracowanie własne na podstawie www.gpw.pl (odczyt z dn. 14.05.2014 r.).

Tabela 3

Harmonogram sesji giełdowej w systemie notowań jednolitych

Rok 2014	
8.30–11.00	Faza przez otwarciem
11.00	Faza otwarcia (określanie kursu jednolitego)
11.00–11.30	Faza dogrywki
11.30–15.00	Faza przez otwarciem
15.00	Faza otwarcia (określanie kursu jednolitego)
15.00–15.30	Faza dogrywki
15.30–17.05	Faza przez otwarciem

Źródło: opracowanie własne na podstawie www.gpw.pl (odczyt z dn. 14.05.2014 r.).

obrotu stosują giełdy grupy NYSE Euronext z siedzibami w Nowym Jorku, Paryżu, Lizbonie, Amsterdamie i Brukseli, jak również wiele rynków giełdowych. W porównaniu z WARSETEM, UTP jest znacznie szybszy, bardziej wydajny, skalowalny i ma większą przepustowość, co w praktyce oznacza możliwość obsługi kilkadziesiąt razy większej liczby zleceń w danej jednostce czasu. Podczas gdy w poprzednim systemie możliwe było zrealizowanie 850 zleceń w ciągu sekundy, UTP daje możliwość przeprowadzenia około 20 tysięcy zleceń na sekundę. UTP otwiera polski rynek kapitałowy na nowe kategorie graczy —

w tym globalne firmy inwestycyjne, które stosują techniki handlu algorytmicznego. Dzięki wprowadzeniu UTP znacznie wzrosło zróżnicowanie możliwości inwestycyjnych, które oferuje GPW. Uruchamiane są kolejne segmenty rynku instrumentów pochodnych i nowe klasy aktywów, które wcześniej nie mogły być wprowadzane na rynek przez ograniczenia technologiczne poprzedniego system (www.gpw.pl, odczyt z dn. 14.05.2014 r.).

Poziom przejrzystości na Giełdzie Papierów Wartościowych w Warszawie jest wysoki, zarówno w przypadku transparentności przed transakcją jak i po. Członkowie i inwestorzy mają pełny dostęp do książki zleceń. System transakcyjny daje możliwość pozyskania szczegółowych informacji o obrocie instrumentów finansowych. Zachowana jest jednak anonimowość dostarczycieli płynności i graczy na giełdzie; zob. Doman (2011).

5. METODY POMIARU INTENSYWNOŚCI PROCESÓW TRANSAKCYJNYCH

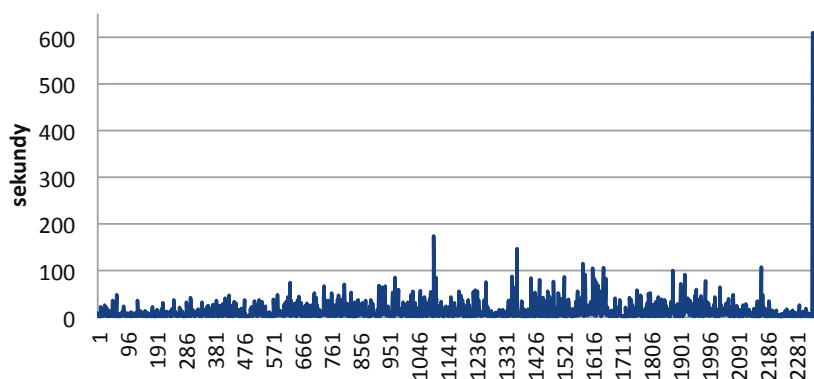
Modelowanie procesu napływu informacji na rynek odgrywa kluczową rolę w inwestowaniu na rynkach finansowych, gdyż daje możliwość prognozowania wszelkich wielkości charakteryzujących badane instrumenty finansowe. Klasyczne podejście w ekonometrii finansowej obejmuje modelowanie stóp zwrotu danego instrumentu, które odpowiadają równym odstępom czasu. W rzeczywistości częstotliwość kolejnych kwotowanych cen nie jest stała. Przedział czasu pomiędzy kolejnymi zmianami cen instrumentów finansowych, lub pomiędzy transakcjami, dostarcza istotnych informacji dotyczących intensywności procesu transakcyjnego.

Dzięki automatyzacji systemów transakcyjnych na giełdach papierów wartościowych powszechnie dostępne stały się zbiory danych, które dokładnie opisują handel walorami. Rejestrowane są wszystkie transakcje, w miarę ich pojawiania się, ze wszelkimi charakterystykami opisującymi dane zdarzenie finansowe. Takie dane nazywane są danymi o wysokiej częstotliwości (ang. *high frequency data*), danymi tikowymi (ang. *tick-by-tick data*) lub danymi transakcyjnymi (ang. *transaction data*) — Bień (2005). Dane tikowe obejmują czas dokonania transakcji, cenę, wolumen, cenach *bid* i *ask*, czy też szczegółowe informacje o zleceniach. Przy wyborze odpowiedniego narzędzia do analizy ekonometrycznej danych o wysokiej częstotliwości, należy pamiętać o ich pewnych charakterystycznych cechach jakimi są: nierównomierne rozmieszczenie obserwacji na osi czasu, występowanie wielu transakcji w tym samym momencie, dyskretne zmiany cen transakcyjnych, odbicia *bid-ask*, grupowanie się intensywności transakcyjnych i występowanie śróddziennej sezonowości; Huptas (2013).

Konsekwencją niesynchronicznego handlu i zarazem najważniejszą cechą danych o wysokiej częstotliwości jest losowe rozłożenie kolejnych obserwa-

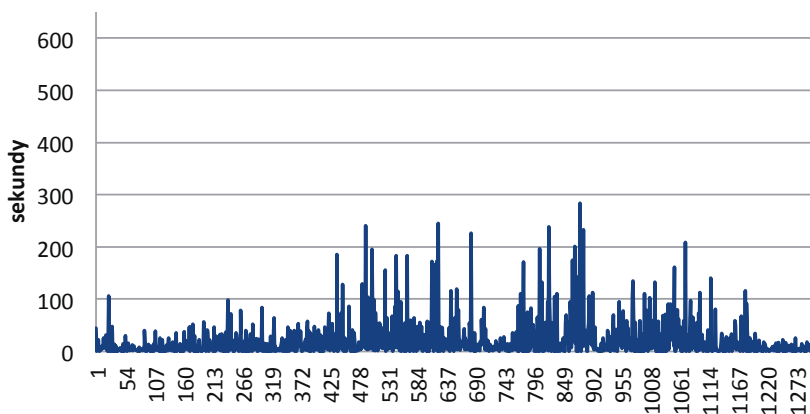
cji w czasie, a więc nierówne odstępy między nimi. Nieregularne obserwacje można zamienić na równoodległe w czasie (np. godzinowe, 15-minutowe, minutowe), co umożliwiać będzie stosowanie klasycznych metod modelowania szeregów czasowych. Jednak wiąże się to najczęściej z utratą informacji jaką niesie za sobą proces transakcyjny, także wybór sposobu transformacji danych oraz przedziału czasu, w którym mają być one zagregowane, jest kwestią problematyczną. Istnieje również ryzyko błędnego wnioskowania, ponieważ efektem analizy danych tikowych w ustalonych odstępach czasu mogą być pozorne korelacje krzyżowe, a także autokorelacje zwrotów. Aby uwzględnić informacje jakie niewątpliwie niosą za sobą czasy trwania pomiędzy kolejnymi obserwacjami, należy korzystać z tzw. modeli w czasie transakcyjnym. Odstępy między wybranymi zdarzeniami traktowane są jako zmienne losowe, natomiast ciąg tych zmiennych jako proces stochastyczny. W badaniach mikrostruktury rynków finansowych rozważane są następujące czasy trwania — Doman (2011), transakcyjne czasy trwania (ang. *trade duration*), cenowe czasy trwania (ang. *price duration*), wolumenowe czasy trwania (ang. *volume duration*) oraz czasy trwania kierunku zmiany ceny (ang. *directional change duration*).

Transakcyjny czas trwania to czas oczekiwania na kolejną transakcję. Jest jedną z metod pomiaru intensywności procesu transakcyjnego. W teorii mikrostruktury wykorzystuje się te czasy trwania jako zmienne zastępcze dla pojawienia się nowej informacji na rynku finansowym. Często też dane te są źródłem informacji na temat intensywności popytu na płynność. Odstępy pomiędzy transakcjami są najczęstszym przedmiotem badań empirycznych danych o wysokiej częstotliwości. Poniższy wykres przedstawia przykładowy szereg transakcyjnych czasów trwania dla akcji spółki PKO BP z dnia 4 maja 2009 roku.



Źródło: opracowanie własne.

Wykres 1. Transakcyjne czasy trwania w przypadku akcji PKO BP z dnia 4 maja 2009 roku

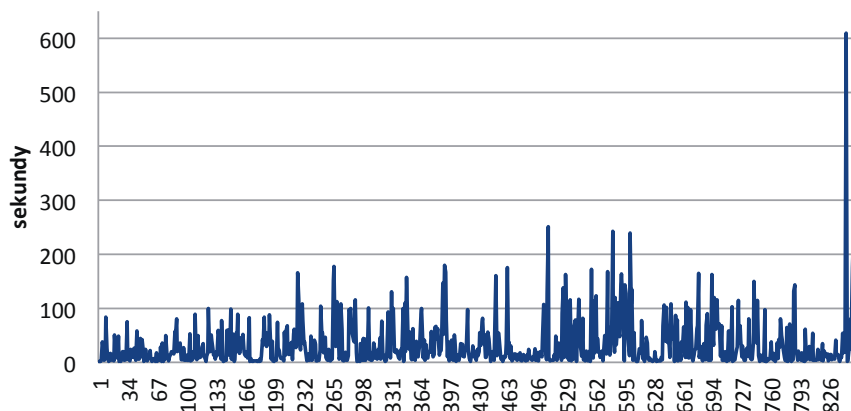


Źródło: opracowanie własne.

Wykres 2. Cenowe czasy trwania w przypadku akcji PKO BP z dnia 4 maja 2009 roku

Cenowy czas trwania definiowany jest jako odstęp czasu potrzebny do zaobserwowania zmiany ceny instrumentu finansowego o pewną przyjętą arbitralnie wartość. Wartość ma być wielokrotnością tzw. tiku cenowego i określa o ile cena musi spaść (lub wzrosnąć) by zarejestrować jej zmianę. Cenowe czasy trwania związane są z płynnością danego instrumentu, a także z intensywnością napływu informacji na rynek. Często, a w szczególności na rynkach kierowanych cenami, zamiast cen transakcyjnych rozważa się tzw. *midquotes*, czyli średnią arytmetyczną cen kupna i sprzedaży w chwili zawarcia transakcji. Przykładowy szereg cenowych czasów trwania został przedstawiony na powyższym wykresie (Wykres 2). Arbitralnie przyjętą wartością jest 1 grosz.

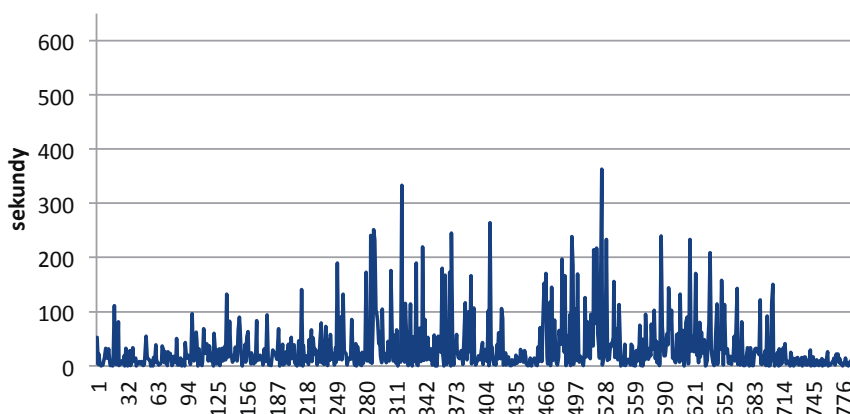
Wolumenowy czas trwania to czas potrzebny do zrealizowania w kolejnych transakcjach określonej wartości wolumenu, która została ustalona arbitralnie. Można zauważyć, że ta definicja nie określa czasu trwania tak precyzyjnie jak poprzednie — np. gdy założona wielkość wolumenu zostanie przekroczona w pojedynczej transakcji, powstaje problem czy nadwyżkę zaliczać na poczet kolejnego czasu trwania czy nie. Podobnie jak transakcyjne i cenowe czasy trwania, wolumenowe czasy oczekiwania są źródłem informacji o intensywności popytu na płynność. Traktowane są także jako zmienna zastępcza dla tzw. intensywności wolumenu. Ponadto wykorzystuje się je do analizy płynności oraz głębokości rynku. Można również badać czas trwania wolumenu nadwyżkowego (ang. *excess volume duration*), czyli odstęp czasu potrzebnego do zrealizowania nadwyżki wolumenu ofert sprzedaży lub ofert kupna. Nadwyżka ta może być wskaźnikiem pojawienia się nowej informacji na rynku, która do tej pory nie została zdyskontowana w cenach. Wykres poniżej ukazuje przykładowy szereg wolumenowego czasu trwania, gdzie ustaloną wielkością był średni poziom wolumenu w badanym dniu.



Źródło: opracowanie własne.

Wykres 3. Wolumenowe czasy trwania w przypadku akcji PKO BP z dnia 4 maja 2009 roku

Czas trwania kierunku zmiany ceny określa się jako odległość w czasie pomiędzy kolejnymi dwoma ekstremami lokalnymi procesu cenowego. Jest to zatem oczekiwanie na zakończenie ciągu kolejnych wzrostów (spadków) ceny, a więc zmiany trendu. Analogicznie jak w przypadku cenowych czasów trwania można badać zmiany ceny, która przekracza pewną wartość, lub rozważać ceny *mid*. Przykładowy szereg omawianych czasów trwania przedstawiony został poniżej.



Źródło: opracowanie własne.

Wykres 4. Czasy trwania kierunku zmiany ceny w przypadku akcji PKO BP z dnia 4 maja 2009 roku

Podstawowe charakterystyki opisowe zostały zestawione w Tabeli 4.

Tabela 4

Charakterystyki opisowe przedstawionych czasów trwania

Czasy trwania (w sekundach)	liczba	średnia	mediana	modalna	minimum	maksimum	wariancja (s^2)	odchylenie standardowe
transakcyjny	855	31,57	16	3	1	609	1853,35	43,05
cenowy	1314	20,09	8	0	0	609	1328,26	36,45
wolumenowy	2369	11,40	6	1	1	609	360,37	18,98
kierunku zmiany ceny	785	32,86	16	0	0	363	2222,87	47,15

Źródło: opracowanie własne.

Ze względu na to, że do przedstawienia przykładowych czasów trwania zdecydowano się wykorzystać proces transakcyjny dotyczący jednej akcji z tego samego dnia, można lepiej zauważyć jak różnią się charakterystyki badanych szeregów. W niniejszej pracy przedmiotem analiz będzie szereg czasów trwania zmiany kierunku ceny, który rzadko występuje w badaniach empirycznych w literaturze przedmiotu.

Powszechnym zjawiskiem w przypadku danych o wysokiej częstotliwości jest występowanie wielu transakcji w tej samej sekundzie. Także transakcje te mogą być zawierane na tym samym instrumencie w tej samej chwili, ale po różnych cenach. Dzieje się tak, ponieważ w okresach bardzo natężonego handlu na rynkach jedna sekunda okazuje się być już zbyt długim przedziałem czasu by zarejestrować wszystkie zmiany procesu transakcyjnego. Sytuacja ta jest najbardziej widoczna na rynkach kierowanych zleceniami. Pojawianie się wielu transakcji w tym samym momencie stanowi wyzwanie dla badacza, gdyż staje on między innymi przed problemem ustalenia jednej ceny dla danej sekundy, a więc sposobem agregacji danych. Również występowanie wielu zerowych czasów trwania może sprawiać trudności w zastosowaniu określonych metod modelowania. Wystąpienie wielokrotnych transakcji w tej samej sekundzie niesie za sobą istotną informację o tempie procesu transakcyjnego, jednak już ilość zawartych transakcji w danej sekundzie nie jest kluczowa dla wnioskania; zob. Huptas (2013).

Kolejną cechą charakterystyczną danych tikowych jest ich dyskretność. Ceny instrumentów finansowych zmieniają się w sposób skokowy. Wartością minimalną o jaką zmienia się cena jest tzw. tik cenowy, można zatem zauważyć,

że ceny walorów są wielokrotnościami pojedynczego tiku cenowego. Najczęściej zmiana ceny z transakcji na transakcję następuje o małą liczbę tików. Na Giełdzie Papierów Wartościowych w Warszawie wielkość tiku cenowego jest zależna od wartości akcji. Warto nadmienić, że wartość tiku silnie wpływa na funkcjonowanie rynku finansowego. Wraz ze zmniejszeniem tiku cenowego zmniejsza się *bid-ask spread*, co z kolei prowadzi do spłykania rynku; Bień (2005).

Następna cecha danych transakcyjnych związana jest z *bid-ask spreadem* oraz występowaniem na rynku animatorów, którzy podtrzymują jego płynność. Ustalają oni różne ceny kupna i sprzedaży instrumentu finansowego z czego wynika wielkość *bid-ask spreadu*. Konsekwencją występowania *spreadu* może być pozorna ujemna autokorelacja pierwszego rzędu zmian kursu transakcyjnego, którą można zaobserwować nawet w wypadku gdy fundamentalna wartość instrumentu nie ulega zmianie. Zjawisko to nazywane jest odbiciem *bid-ask* i powoduje ciągłą fluktuację cen transakcyjnych pomiędzy cenami *bid* i *ask* w zależności od zleceń. Można zatem wyciągnąć wniosek, że obserwowane zmiany cen na rynku nie muszą świadczyć o pojawieniu się nowej informacji na rynku; zob. Doman (2006).

Dla danych o wysokiej częstotliwości charakterystyczne jest grupowanie się intensywności transakcyjnych. Nowa informacja napływa na rynek w ciągu dnia z różnym natężeniem, co przekłada się na zróżnicowaną intensywność procesu transakcyjnego. Można zatem zauważyć okresy podwyższonej aktywności inwestorów na rynku, jak i okresy, w których liczba zawieranych transakcji jest niska. W związku z tym występuje grupowanie się czasów trwania — po okresach długich czasów trwania następują okresy krótkiego oczekiwania na badaną zmianę (i na odwrót) — Huptas (2013).

Bardzo typowym zjawiskiem, które obserwuje się w szeregach danych tiko- wych jest występowanie silnych efektów cyklicznych w ciągu dnia. Związane jest to z pewną ustaloną dzienną rutyną graczy na giełdzie. Większe natężenie aktywności transakcyjnej ma miejsce tuż po otwarciu rynku, ponieważ inwestorzy wykorzystują wtedy informacje, które napłynęły do nich w ciągu nocy. Wraz z przyswajaniem tych wiadomości aktywność transakcyjna spada osiągając swoje minimum we wczesnych godzinach popołudniowych (tzw. efekt lunchu). Wraz ze zbliżaniem się do końca sesji natężenie zawieranych transakcji wzrasta, ponieważ inwestorzy dokonują korekty swoich pozycji, bądź je zamykają. Zwiększona aktywność transakcyjna graczy na początku i na końcu sesji uwidacznia się w krótszych czasach oczekiwania na kolejną transakcję, natomiast wyraźnie niższa aktywność w czasie lunchu skutkuje tym, że odstępy między transakcjami są dłuższe. Zatem wykres odwzorowujący śróddzienną sezonowość będzie miał kształt odwróconej litery „U”. Podobnie jest w przypadku cenowych oraz wolumenowych czasów trwania — Huptas (2013). Należy mieć na uwadze tę niezwykle istotną cechę danych o wysokiej częstotliwości i przed przystąpieniem do analiz zaleca się wyeliminowanie czyn-

nika periodycznego, tak by nie fałszował on wniosków. W niniejszej pracy do zlikwidowania czynnika wewnątrzdziennej sezonowości zastosowano regresję Nadaraya–Watsona, która zostanie opisana w części 3.3.

6. MODELE ACD W BADANIACH DYNAMIKI CZASÓW TRWANIA

Dane o wysokiej częstotliwości charakteryzują się nieregularnością występowania kolejnych zdarzeń transakcyjnych. Obserwacje o nierównych odstępach można zamienić na dane regularnie pojawiające się na osi czasu, jednak wiąże się to z wieloma niedogodnościami. Przede wszystkim badacz musi być świadomy, że takiej transformacji danych towarzyszy przeważnie utrata informacji. Wybór sposobu przekształcenia danych także jest problematyczny, a otrzymany regularny szereg jest najczęściej zbudowany z przeważającej liczby zer, co skutecznie utrudnia jego modelowanie; zob. Doman (2011). Stosowanie modeli w czasie transakcyjnym jest lepszym rozwiązaniem, gdyż nieregularność danych, uwzględniana w tych modelach, niesie ze sobą istotne informacje.

Czas oczekiwania na kolejne zdarzenie traktuje się jako zmienną losową o ustalonym rozkładzie prawdopodobieństwa, natomiast ciąg tych zmiennych jest procesem stochastycznym. W związku z tym, że dane te nie napływają regularnie uznaje się taki szereg za proces punktowy. Procesem punktowym nazywamy szczególny rodzaj procesu stochastycznego, w którym punkty generowane są losowo na osi czasu; Pacuar (2006). Niech $t \in [0; +\infty)$ będzie zmienną opisującą czas, natomiast $\{t_0, t_1, \dots, t_n, \dots\}$ to ciąg chwil, w których nastąpiły dane zdarzenia procesu transakcyjnego, gdzie $0 = t_0 \leq t_1 \leq \dots \leq t_n \leq \dots$. Zatem taki szereg momentów $t = t_0, t_1, \dots, t_n, \dots$ jest procesem punktowym. Mówimy o prostym procesie punktowym, gdy spełniony jest warunek, że $t_i < t_{i+1}$ dla $i=1, 2, \dots$, czyli momenty nie mogą zachodzić jednocześnie. Do opisu napływania kolejnych zdarzeń stosuje się te procesy punktowe, w których realizacja zdarzeń zależy od chwil, w których występowały zdarzenia poprzedzające; Pacuar (2006). Oznacza to, że proces taki może być sformułowany jako proces warunkowej intensywności transakcji. Podejście to występuje w przypadku rodziny modeli autoregresyjnego warunkowego czasu trwania wprowadzonych przez Engle'a i Russela–Engle, Russel (1998) — które zostaną szczegółowo opisane w niniejszym rozdziale. Klasyczny model Engle'a i Russela został w ciągu lat wielokrotnie modyfikowany — ze względu na różny rozkład składnika losowego lub różną postać funkcyjną modelu; Bień (2005). Wybrane podejścia wykorzystane w niniejszej pracy w analizie empirycznej zostaną omówione w części 2.1.

Dynamikę oczekiwanego czasu trwania można opisać modelem przedstawionym w 1998 roku przez Engle'a i Russela zwanym modelem autoregresyjnego warunkowego czasu trwania, ACD (ang. *Autoregressive Conditional Duration*).

Rozważmy $x_i = t_i - t_{i-1}$, co oznacza odstęp czasu pomiędzy dwoma kolejnymi zdarzeniami transakcyjnymi, które zaistniały w chwili t_i oraz t_{i-1} . Ponadto, niech \mathcal{F}_{i-1} oznacza zbiór informacji dostępnej do momentu t_{i-1} włącznie. Obserwowany czas trwania (*duration*) wyrażony jest w postaci następującego iloczynu; zob. Engle, Russel (1998):

$$x_i = \Psi_i \cdot \varepsilon_i,$$

gdzie:

Ψ_i — warunkowa wartość oczekiwana czasu trwania,

$\varepsilon_i > 0$ — składnik losowy.

Zakłada się, że ε_i to ciąg nieskorelowanych zmiennych losowych na dodatniej półosi rzeczywistej o jednakowym rozkładzie prawdopodobieństwa. Charakteryzują się one funkcją gęstości postaci $f_\varepsilon(\varepsilon_i)$ o wartości oczekiwanej $E(\varepsilon_i) = 1$ oraz stałej wariancji $Var(\varepsilon_i) = \sigma_\varepsilon^2$, co można zapisać $\varepsilon_i \sim i.i.d(1, \sigma_\varepsilon^2)$. Zakładamy, że warunkowa wartość Ψ_i jest poprawnie zdefiniowana, w związku z czym prawdziwe jest założenie, że składniki ε_i są stochastycznie niezależne od Ψ_i , czyli $\varepsilon_i \perp \Psi_i$; zob. Huptas (2013). Skoro $E(\varepsilon_i) = 1$, to Ψ_i można przedstawić następująco:

$$\Psi_i = E(x_i | \mathcal{F}_{i-1}).$$

Z kolei warunkowa wariancja czasu trwania jest postaci:

$$Var(x_i | \mathcal{F}_{i-1}) = \Psi_i^2 \varepsilon_i^2.$$

Podstawowa specyfikacja modelu ACD bazuje na liniowej parametryzacji dynamiki oczekiwanego czasu trwania i jest ona postaci — Engle, Russel (1998):

$$\Psi_i = \omega + \sum_{j=1}^p \alpha_j x_{i-j} + \sum_{j=1}^q \beta_j \Psi_{i-j},$$

gdzie:

$\omega > 0$, $\alpha_j \geq 0$, $\beta_j \geq 0$ — parametry modelu dla każdego $j = 1, 2, \dots$,

p, q — rzędy opóźnień odpowiednio dla x_i oraz Ψ_i .

Jest to klasyczny model określany jako ACD(p,q). Narzucone restrykcje na parametry są wystarczające dla nieujemności procesu czasu trwania. Proces ten jest kowariancyjnie stacjonarny, gdy spełniony jest warunek, że $\sum_{j=1}^p \alpha_j + \sum_{j=1}^q \beta_j < 1$. Wówczas istnieje analityczna postać bezwarunkowej wartości oczekiwanej:

$$E(x_i) = \frac{\omega}{1 - (\sum_{j=1}^p \alpha_j + \sum_{j=1}^q \beta_j)}.$$

W badaniach empirycznych często wykorzystywana jest uproszczona specyfikacja, czyli model ACD(1,1). Postać modelu przedstawia się następująco:

$$\begin{aligned} x_i &= \Psi_i \varepsilon_i, \\ \Psi_i &= \omega + \alpha x_{i-1} + \beta \Psi_{i-1}, \end{aligned}$$

gdzie:

$\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$ — parametry modelu.

W niniejszej pracy zostały zastosowane modele jedynie o pojedynczych opóźnieniach, w związku z tym indeksy przy parametrach modelu będą pomijane. Bezwarunkowa wartość oczekiwana procesu w takim przypadku ma postać:

$$E(x_i) = \frac{\omega}{1 - (\alpha + \beta)} \quad \text{gdy } \alpha + \beta < 1.$$

Natomiast jego bezwarunkowa wariancja wygląda następująco:

$$Var(x_i) = [E(x_i)]^2 \sigma_\varepsilon^2 \frac{1 - 2\alpha\beta - \beta^2}{1 - (\alpha + \beta)^2 - \alpha^2 \sigma_\varepsilon^2} \quad \text{gdy } (\alpha + \beta)^2 - \alpha^2 \sigma_\varepsilon^2 < 1. \quad (1)$$

Restrykcja na istnienie bezwarunkowej wariancji procesu jest jednocześnie warunkiem kowariancyjności procesu ACD(1,1). Z formuły (1) można wyprowadzić wzór na kwadrat współczynnika zmienności dla procesu, poprzez podzielenie obu stron przez $[E(x_i)]^2$:

$$V_x^2 = \frac{Var(x_i)}{[E(x_i)]^2} = \sigma_\varepsilon^2 \cdot \frac{1 - 2\alpha\beta - \beta^2}{1 - (\alpha + \beta)^2 - \alpha^2 \sigma_\varepsilon^2}. \quad (2)$$

Z formuły (2) wynika, że zawsze gdy parametr α jest większy od 0, współczynnik zmienności V_x jest większy od odchylenia standardowego składnika innowacji σ_ε ($\sigma_\varepsilon = \sqrt{\sigma_\varepsilon^2}$). Modele ACD charakteryzują się tym, że odchylenie standardowe składnika losowego σ_ε równa się współczynnikowi zmienności ε_i (gdyż $E(\varepsilon_i) = 1$, zatem $V_\varepsilon = \frac{\sigma_\varepsilon}{E(\varepsilon_i)} = \sigma_\varepsilon$). W związku z tym, prawdziwe jest stwierdzenie, że kiedy $\alpha > 0$, czas trwania x_i cechuje się większym rozproszeniem od składnika ε_i . Również wraz ze wzrostem wartości parametru α dyspersja zmiennej x_i rosnąć będzie w stosunku do dyspersji zmiennej ε_i . Kiedy odchylenie standardowe σ_ε będzie większe lub równe 1, współczynnik zmienności czasu trwania będzie zawsze większy od 1. Zjawisko to nazywane jest nadmiernym rozproszeniem zmiennej (*overdispersion*) i jest własnością bardzo często spotykaną w analizie empirycznej czasów trwania — zob. Huptas (2013).

W praktyce, jeżeli w modelowanym procesie pojawiają się bardzo długie lub bardzo krótkie czasy trwania, model ACD(p, q) ma tendencję do przeszaco-

wywania warunkowej wartości oczekiwanej. Ponadto, narzucone restrykcje na dodatniość parametrów modelu mogą sprawiać problemy natury numerycznej. Dlatego powstało wiele modyfikacji równania opisującego warunkową wartość oczekiwaną. Jedną z najwcześniejszych rozszerzeń specyfikacji modelu ACD była wprowadzona przez Bauwensa i Giota — zob. Bauwens, Giot (2000) — logarytmiczna wersja modelu ACD. Czas trwania opisywany jest tutaj poprzez następujące relacje:

$$\begin{aligned} x_i &= e^{\psi_i} \cdot \varepsilon_i \\ \Psi_i &= \exp(\psi_i) \quad \text{zatem } \psi_i = \ln \Psi_i \end{aligned} \quad (3)$$

gdzie błędy $\varepsilon_i \sim i.i.d(\mu, \sigma_\varepsilon^2)$, tak więc:

$$E(x_i | \mathcal{F}_{i-1}) = \mu \cdot \exp(\psi_i).$$

Zakłada się również, że dla modeli ACD stosunek $\frac{x_i}{E(x_i | \mathcal{F}_{i-1})}$ jest *i.i.d.*, a zatem $\frac{x_i}{\Psi_i}$ także jest *i.i.d.* — Engle, Russel (1998). Do ukazania zależności, która zgodnie z empirią skutkuje skupianiem się czasów oczekiwania, Bauwens i Giot wprowadzili dwie specyfikacje modelu log-ACD różniące się postacią funkcyjną składnika ε_i . Pierwszy model log-ACD₁(p, q) dany jest następującym wzorem (Bauwens, Giot (2008)):

$$\begin{aligned} \phi_i &= \omega + \sum_{j=1}^p \alpha_j \ln \varepsilon_{i-j} + \sum_{j=1}^q \beta_j \phi_{i-j} = \\ &= \omega + \sum_{j=1}^p \alpha_j \ln x_{i-j} + \sum_{j=1}^q (\beta_j - \alpha_j) \phi_{i-j}. \end{aligned}$$

Druga wersja modelu log-ACD₂(p, q) ma postać:

$$\begin{aligned} \phi_i &= \omega + \sum_{j=1}^p \alpha_j \varepsilon_{i-j} + \sum_{j=1}^q \beta_j \phi_{i-j} = \\ &= \omega + \sum_{j=1}^p \alpha_j \left(\frac{x_{i-j}}{\Psi_i} \right) + \sum_{j=1}^q \beta_j \phi_{i-j}. \end{aligned}$$

Dla obu wersji logarytmicznych modeli ACD warunkiem ścisłej stacjonarności procesu jest nierówność: $|\beta_j| < 1$ — Huptas (2013). W niniejszej pracy wykorzystywana jest wyłącznie specyfikacja log-ACD₁(1,1), którą można zapisać jako następujące równanie:

$$\ln \Psi_i = \omega + \alpha \ln \varepsilon_{i-1} + \beta \ln \Psi_{i-1}.$$

Proces ten jest stacjonarny gdy $|\beta| < 1$.

Wadą logarytmicznej specyfikacji modeli ACD jest przede wszystkim fakt, że nie istnieją ogólne analityczne postaci bezwarunkowych momentów procesu. Z drugiej strony, niewątpliwą zaletą modeli log-ACD jest to, że nie wymagają one nakładania restrykcji na α , β i ω , gdyż budowa równania (3) zapewnia dodatniość parametrów. Ponadto, innowacje ε_{i-1} , które wiążą się ze zbyt długim lub zbyt krótkim czasem trwania, oddziałują na warunkową wartość oczekiwaną w sposób nieliniowy. Zastosowanie logarytmicznych modeli ACD w celu testowania efektów mikrostruktury rynku jest korzystniejsze niż wykorzystanie klasycznych modeli warunkowego czasu trwania; zob. Bauwens, Giot (2008).

Rozszerzenia modeli ACD rozwijały się także w kierunku zmian typu rozkładu składnika innowacji ε_i , jako że wyniki i jakość estymacji zależą od wybranej funkcji gęstości składników losowych. W tej rodzinie modeli, możliwym jest zastosowanie dla zaburzeń jedynie rozkładów prawdopodobieństwa określonych na zbiorze liczb dodatnich rzeczywistych.

Najprostszym i zarazem najpopularniejszym rozkładem wykorzystanym do opisu zmiennej losowej ε_i jest rozkład wykładniczy. Model taki nosi nazwę EACD(p, q); Engle, Russel (1998). Funkcja gęstości rozkładu wykładniczego dana jest wzorem:

$$f_{\varepsilon}(\varepsilon_i) = \lambda \exp(-\lambda\varepsilon_i),$$

gdzie:

λ — parametr, $\varepsilon_i > 0$.

Wartość oczekiwana oraz wariancja zmiennej wynoszą odpowiednio:

$$\begin{aligned} \mu_{\varepsilon} &= E(\varepsilon_i) = \lambda, \\ \sigma_{\varepsilon}^2 &= \text{Var}(\varepsilon_i) = \lambda^2. \end{aligned}$$

Składniki losowe ε_i mają być niezależnymi zmiennymi losowymi o jednakowym rozkładzie z wartością oczekiwaną równą 1, zatem dla $\varepsilon_i \sim \text{Exp}(\lambda)$ dla $\lambda = 1$, czyli:

$$f_{\varepsilon}(\varepsilon_i) = \exp(-\varepsilon_i).$$

Stosowanie rozkładu wykładniczego może jednak okazać się zbyt restrykcyjne; Huptas (2013). Naturalnym uogólnieniem wykładniczego typu rozkładu ε_i dla modeli ACD będzie rozkład Weibulla, również często używany w analizach czasów trwania. Model ten nazywany jest WACD(p, q). Funkcja gęstości rozkładu Weibulla ma następującą postać:

$$f_{\varepsilon}(\varepsilon_i) = \frac{\gamma}{\lambda} \left(\frac{\varepsilon_i}{\lambda}\right)^{\gamma} \exp\left(-\left(\frac{\varepsilon_i}{\lambda}\right)^{\gamma}\right),$$

gdzie:

λ, γ — parametry. Ze względu na występowanie w tym rozkładzie parametru kształtu γ funkcja gęstości może lepiej dopasowywać się do danych. Odpowiednio, wartość oczekiwana i wariancja wynoszą:

$$\mu_\varepsilon = E(\varepsilon_i) = \lambda \Gamma\left(1 + \frac{1}{\gamma}\right),$$

$$\sigma_\varepsilon^2 = \text{Var}(\varepsilon_i) = \lambda^2 \left[\Gamma\left(1 + \frac{2}{\gamma}\right) - \left(\Gamma\left(1 + \frac{1}{\gamma}\right)\right)^2 \right].$$

Gdzie $\Gamma(\cdot)$ — funkcja gamma. Aby spełnione zostało założenie o niezależności i jednakowym rozkładzie zmiennych ε_i , o wartości oczekiwanej równej 1, należy dokonać reparametryzacji λ :

$$\lambda = \frac{1}{\Gamma\left(1 + \frac{1}{\gamma}\right)}.$$

Po tej modyfikacji gęstość rozkładu Weibulla zmiennej ε_i będzie mieć postać:

$$f_\varepsilon(\varepsilon_i) = \gamma \Gamma\left(1 + \frac{1}{\gamma}\right) \left(\varepsilon_i \Gamma\left(1 + \frac{1}{\gamma}\right)\right)^{\gamma-1} \exp\left(-\left(\varepsilon_i \Gamma\left(1 + \frac{1}{\gamma}\right)\right)^\gamma\right).$$

W niniejszej pracy wykorzystane zostaną dwa przedstawione powyżej typy rozkładów składnika losowego ε_i dla modeli ACD.

Do szacowania parametrów modeli ACD zastosowano Metodę Największej Wiarygodności (MNW). Dla MNW fundamentalną kwestią będzie wybór i przyjęcie danego rozkładu prawdopodobieństwa zmiennej losowej ε_i . Proponowanymi przez literaturę przedmiotu rozkładami dla modeli ACD są między innymi: rozkład wykładniczy, Weibulla, uogólniony gamma, logarymiczno-normalny czy rozkład Burra; Bień (2005). Warunkowy rozkład zmiennej x_i stanowi szczególny przypadek rozkładu składnika losowego $\varepsilon_i = \frac{x_i}{\Psi_i}$ (gdy $\Psi_i = 1$), zatem warunkowa funkcja gęstości x_i ma następującą postać:

$$(x_i | \Psi_i; \theta) = \frac{1}{\Psi_i} f_\varepsilon\left(\frac{x_i}{\Psi_i}\right),$$

gdzie:

θ — wektor estymowanych parametrów. Funkcję wiarygodności dla nieznanymi parametrów modelu konstruuje się jako iloczyn N funkcji gęstości $f(x_i | \mathcal{F}_i)$ dla $i = 1, 2, \dots, N$ (gdzie N — liczebność próby), przy warunkach początkowych x_0, Ψ_0 :

$$L(\theta) = \prod_{i=1}^N f(x_i | \Psi_i; \theta) = \prod_{i=1}^N \frac{1}{\Psi_i} f_\varepsilon\left(\frac{x_i}{\Psi_i}\right).$$

Estymatorami parametrów θ będą takie $\hat{\theta}$, dla których funkcja wiarygodności $L(\theta)$ osiąga maksimum. Do obliczeń wykorzystuje się logarytm naturalny funkcji wiarygodności $l(\theta)$. Taka transformacja nie zmienia położenia maksimum, a znacznie ułatwia proces optymalizacji:

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^N [\ln f(x_i | \Psi_i; \theta)] = \sum_{i=1}^N \left[\ln f_\varepsilon\left(\frac{x_i}{\Psi_i}\right) - \ln(\Psi_i) \right].$$

Estymatory MNW charakteryzują się asymptotycznym rozkładem normalnym. Oznacza to, że przy liczebności N zmierzającej do nieskończoności rozkład estymatora $\hat{\theta}$ będzie rozkładem normalnym z wartością oczekiwaną równą prawdziwej wartości wektora θ i macierzy kowariancji równej $V(\theta)$. Estymator macierzy kowariancji ma następującą postać:

$$\hat{V}(\hat{\theta}) = \left[-\frac{1}{N} \sum_{i=1}^N \frac{\partial^2 \ln f(x_i | \Psi_i; \theta)}{\partial \hat{\theta} \partial \hat{\theta}'} \right]^{-1}.$$

W niniejszej pracy przy budowie modeli warunkowego czasu trwania zastosowane zostały dwa rozkłady warunkowe: rozkład wykładniczy oraz Weibulla. Dla rozkładu wykładniczego warunkowa funkcja gęstości czasu trwania jest postaci:

$$f(x_i | \Psi_i; \theta) = \frac{1}{\Psi_i} \exp\left(-\frac{x_i}{\Psi_i}\right).$$

W związku z czym logarytm funkcji wiarygodności wyraża się wzorem:

$$l(\theta) = - \sum_{i=1}^N \left[\frac{x_i}{\Psi_i} + \ln(\Psi_i) \right].$$

W przypadku, gdy za rozkład ε_i przyjmiemy rozkład Weibulla, warunkowa funkcja gęstości procesu x_i wygląda następująco:

$$f(x_i | \Psi_i; \theta) = \frac{\gamma}{x_i} \left[\frac{x_i \Gamma\left(1 + \frac{1}{\gamma}\right)}{\Psi_i} \right]^\gamma \exp \left\{ - \left[\frac{x_i \Gamma\left(1 + \frac{1}{\gamma}\right)}{\Psi_i} \right]^\gamma \right\}.$$

Zatem logarytm funkcji wiarygodności opisuje się za pomocą wzoru:

$$l(\theta) = \sum_{i=1}^N \left\{ \ln\left(\frac{\gamma}{x_i}\right) + \gamma \ln \left[\frac{x_i \Gamma\left(1 + \frac{1}{\gamma}\right)}{\Psi_i} \right] - \left[\frac{x_i \Gamma\left(1 + \frac{1}{\gamma}\right)}{\Psi_i} \right]^\gamma \right\}.$$

Oszacowane estymatory będą zgodne i asymptotycznie najbardziej efektywne, o ile rozkład składnika losowego ε_i został wybrany poprawnie. Jeśli rozkład ten nie jest właściwy estymatory MNW tracą swoje własności; zob. Bień (2005).

W przypadku estymacji procesu z wykorzystaniem logarytmów funkcji wiarygodności definiuje się tzw. kryteria informacyjne pozwalające na porównywanie modeli. Najbardziej znaną i najczęściej wykorzystywaną miarą w badaniach empirycznych jest kryterium informacyjne Akaike'a AIC (*Akaike Information Criterion*). Jego formuła określona jest następująco — Kufel (2007):

$$AIC = -\frac{2l(\hat{\theta})}{N} - \frac{2k}{N},$$

gdzie:

k to liczba parametrów modelu, zaś N oznacza liczba obserwacji. Model o najniższej wartości kryterium AIC jest uznawany za najlepszy w takim sensie,

że jest on najbardziej dopasowany do danych i zarazem najprostszy. Powyżej przedstawiona miara zostanie wykorzystana do sporządzenia rankingu modeli ACD w dalszej części pracy.

W celu oceny i sprawdzenia poprawności obliczonych modeli, a w związku z tym trafności rankingu, można posłużyć się weryfikacją statystyczną. Z racji tego, że w niniejszej pracy zestawiane są modele ACD o warunkowych rozkładach wykładniczym i Weibulla, zasadnym jest testowanie parametru kształtu γ , który przyjmując za wartość 1 sprowadza rozkład Weibulla do wykładniczego. Testowanie γ można przeprowadzić przy pomocy testu t -Studenta. W tym przypadku hipoteza zerowa wobec alternatywnej będzie przedstawiać się następująco:

$$\begin{aligned} H_0 : \gamma &= 1, \\ H_1 : \gamma &\neq 1. \end{aligned}$$

Sprawdzianem w tym teście jest statystyka postaci:

$$t_\gamma = \frac{\hat{\gamma} - 1}{d(\hat{\gamma})}.$$

W ramach modelu regresji liniowej przy założeniu normalności składników losowych oraz prawdziwości hipotezy zerowej statystyka t ma rozkład t -Studenta o $N-k-1$ stopniach swobody, gdzie k to liczba parametrów (warto zauważyć, że przy dużej liczbie obserwacji rozkład t -Studenta sprowadza się do rozkładu normalnego); zob. Wywiół (2004). W przypadku modeli ACD stosuje się testowanie przybliżone. Z asymptotycznej normalności estymatora MNW wynika przybliżony rozkład t -Studenta dla rozważanej statystyki. Pozwala to na wykorzystanie opisanego testu do weryfikacji istotności parametrów w modelu ACD, który został oszacowany metodą największej wiarygodności.

Model można również ocenić wykorzystując test ilorazu wiarygodności, który służy do porównywania różnych form funkcyjnych modeli. Jest on oparty na różnicy logarytmów wartości funkcji wiarygodności dla dwóch zestawianych modeli. Test ilorazu wiarygodności stosowany jest bardzo często, gdyż daje możliwość porównywania modeli oszacowanych na tej samej próbie. W tym przypadku test ten sprawdzać będzie istotność całego modelu z warunkowym rozkładem Weibulla względem modelu z warunkowym rozkładem wykładniczym, w którym parametr kształtu sprowadza się do 1 (czyli będzie to model z warunkiem narzuconym na γ). Hipoteza zerowa zakładać zatem będzie, że rozkładem warunkowym danego modelu jest rozkład wykładniczy, natomiast hipoteza alternatywna mówić będzie o tym, że jest to rozkład Weibulla. Test przeprowadzany jest za pomocą poniższej statystyki; Wywiół (2004):

$$LR = 2(\ln L_1 - \ln L_2),$$

gdzie $\ln L_1$ to logarytm naturalny funkcji wiarygodności modelu z narzuconym warunkiem (z rozkładem wykładniczym), a $\ln L_2$ oznacza logarytm naturalny

funkcji wiarygodności modelu bez ograniczeń (z rozkładem Weibulla). Statystyka LR ma rozkład chi-kwadrat o liczbie stopni swobody równej liczbie wszystkich parametrów, występujących w obu omawianych modelach.

Przedstawione metody zostaną wykorzystane w empirycznej części pracy, w której za pomocą kryterium informacyjnego AIC zbudowano ranking modeli ACD. Jego zasadność weryfikowana będzie przy użyciu opisanych testów statystycznych. Ranking zostanie utworzony dla danych odsezonowanych oraz dla tych niepoddanych odsezonowaniu, dzięki czemu możliwe będzie porównanie pozycji modeli w rankingu przy uwzględnieniu wewnątrzdziennej cykliczności. W kolejnym podrozdziale omówione zostaną narzędzia wykorzystane do eliminacji sezonowości.

7. REGRESJA NADARAYA I WATSONA JAKO NARZĘDZIE ELIMINACJI SEZONOWOŚCI WEWNĄTRZDZIENNEJ

Wewnątrzdzienne sezonowości jest cechą charakterystyczną danych o wysokiej częstotliwości, w tym czasów trwania. Własność ta zazwyczaj powoduje ich silną dodatnią autokorelację; zob. Bień (2005). Aby wyeliminować zjawisko autokorelacji, należałoby usunąć efekt cykliczności, co jest możliwe dzięki odpowiedniej transformacji danych. Literatura przedmiotu postuluje, by przed właściwym szacowaniem modeli ACD stosownie przekształcić zbiór danych; Engle, Russel (1998). W tym celu dokonuje się dekompozycji czasów trwania na część deterministyczną $\phi(t_i)$ oraz stochastyczną x_i . Dane transformowane są zgodnie z następującą formułą:

$$\hat{x}_i = \frac{x_i}{\phi(t_i)},$$

gdzie \hat{x}_i — czas trwania bez efektu cykliczności. Metodą powszechnie stosowaną do szacowania wewnątrzdziennej funkcji sezonowości jest estymacja jądrowa. Charakteryzuje się ona przede wszystkim tym, że analityczna postać funkcji gęstości czasu trwania nie musi być określona. Niech X będzie zbiorem składającym się z n obserwacji d -wymiarowych, generowanych przez rozkład prawdopodobieństwa $p(x)$. Dla $\tilde{x} \in X$ konstruowane jest O — czyli małe otoczenie, w którym \tilde{x} jest położone w centrum. W przypadku gdy $d=1$, O będzie odcinkiem o długości h . Otoczenie O ma własność — Härdle (1992):

$$\int_0^1 p(x) = P.$$

Prawdopodobieństwo, że l z n obserwacji należy do obszaru O zależy od wielkości n i P . Dla dostatecznie dużego n , l przybliża się następująco:

$$l \approx n \cdot P.$$

Przy bardzo małym h przyjmuje się, że $p(x)$ jest stałe nad obszarem O :

$$P \approx p(x) \cdot h.$$

Z czego wynika, że:

$$p(x) = \frac{1}{n \cdot h}.$$

Formuła ta stanowi przybliżenie wartości funkcji gęstości prawdopodobieństwa w punkcie x . W celu oszacowania estymatora funkcji gęstości dla badanego zbioru danych badacz przesuwając po obserwacjach ruchomym otoczeniem (oknem) i w każdego z nich wylicza $p(x)$. W przypadku jądrowych estymatorów gęstości szerokość otoczenia jest stała, natomiast zmienia się liczba l . Wpływ punktów należących do okna waży się za pomocą funkcji jądrowej. Funkcją jądrową nazywa się jednoczynnikową nieujemną funkcję gęstości prawdopodobieństwa, która spełnia następujące założenia — zob. Heiler (1999):

$$\begin{aligned} \int K(z) dz &= 1, \\ \int zK(z) dz &= 0, \\ \int z^2K(z) dz &= K_2 < \infty. \end{aligned}$$

Zakłada się ponadto, że dla każdego $z \in \mathbb{R}$ funkcja ta jest symetryczna względem zera oraz ma w zerze maksimum globalne. Przykłady często wykorzystywanych funkcji jądrowych przedstawia poniższa tabela:

Tabela 5

Przykłady funkcji jądrowych

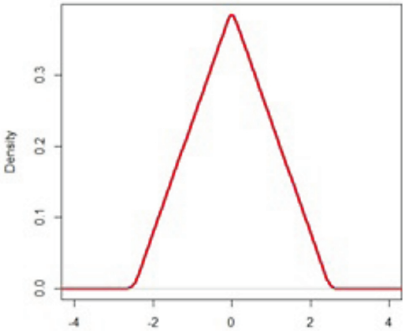
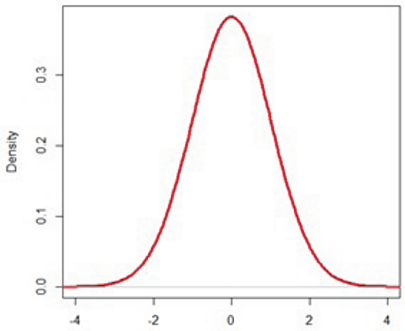
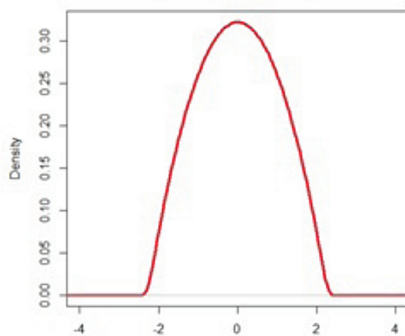
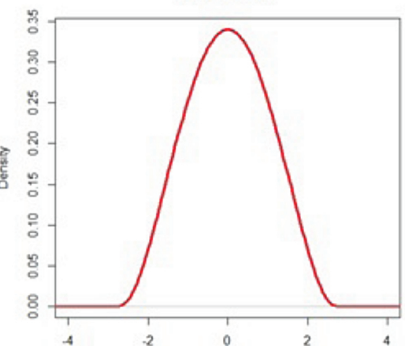
Jądro trójkątne	
$K(z) = 1 - z \mathbb{I}_{ z \leq 1}(z)$	

Tabela 5

<p>Jądro gaussowskie (normalne)</p> $K(z) = (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{z^2}{2}\right) I_{\mathbb{R}}(z)$	<p>jądro normalne</p> 
<p>Jądro Epanechnikova</p> $K(z) = \frac{3}{4}(1 - z^2) I_{ z \leq 1}(z)$	<p>jądro Epanechnikova</p> 
<p>Jądro quartic</p> $K(z) = \frac{15}{16}(1 - z^2)^2 I_{ z \leq 1}(z)$	<p>jądro quartic</p> 

Źródło: opracowanie własne.

Dla zbiorów pochodzących z przestrzeni wielowymiarowych (czyli gdy $d > 1$) można przyjąć, że ich jądrem jest iloczyn funkcji jądrowych jednowymiarowych:

$$K(x_1, x_2) = K(x_1) \cdot K(x_2).$$

Przy zadanej postaci funkcji jądrowej K formuła na liczbę l przyjmuje postać:

$$l = \sum_{i=1}^n K\left(\frac{x - x_i}{h^d}\right),$$

przy czym dodatni współczynnik h określa się mianem parametru wygładzania lub szerokością pasma. Po podstawieniu powyższej zależności do równania 2.38 na $p(x)$ wzór na estymator gęstości w d -wymiarowej przestrzeni, czyli tzw. estymator Rosenblatta–Parzena przedstawia się następująco; Härdle (1992):

$$p(x) = \frac{1}{n \cdot h^d} \sum_{i=1}^n K\left(\frac{x - x_i}{h^d}\right).$$

Nadaraya i Watson w oparciu o powyższy estymator jądrowy funkcji gęstości, skonstruowali regresję jądrową. Dana niech będzie próba $\{y_i, x_i\}_{i=1}^n$ taka, że $y_i \in \mathbb{R}$ i $x_i \in \mathbb{R}^d$ o łącznej gęstości $f(y, x)$. Do prognozowania wartości zmiennej y potrzebna jest jej warunkowa funkcja gęstości dana wzorem:

$$g(y|x) = \frac{f(y, x)}{f_x(x)}.$$

Na podstawie estymatora jądrowego funkcji gęstości Rosenblatta–Parzena otrzymuje się następujące oszacowanie warunkowej gęstości y :

$$\hat{g}(y|x) = \frac{\frac{1}{h_y} \cdot \sum_{i=1}^n K_y\left(\frac{y - y_i}{h_y}\right) \cdot K_x\left(\frac{x - x_i}{h_x}\right)}{\sum_{i=1}^n K_x\left(\frac{x - x_i}{h_x}\right)}.$$

W wyniku wyliczenia następującej całki: $\int y \cdot \hat{g}(y|x) dy$ uzyskuje się oszacowanie warunkowej wartości oczekiwanej $E(y|x)$, zwane estymatorem Nadaraya i Watsona — Andersen i in. (2009):

$$m(x) = \frac{\sum_{i=1}^n y_i K_x\left(\frac{x - x_i}{h_x}\right)}{\sum_{i=1}^n K_x\left(\frac{x - x_i}{h_x}\right)}.$$

Jest to średnia ważona zmiennej zależnej, gdzie wagi zależą od zmiennych niezależnych w następujący sposób:

$$w_i = \frac{K_x\left(\frac{x - x_i}{h_x}\right)}{\sum_{i=1}^n K_x\left(\frac{x - x_i}{h_x}\right)}.$$

Stosując powyższy nieparametryczny estymator do zagadnienia eliminacji sezonowości w niniejszej pracy, czynnik deterministyczny przyjmować będzie postać:

$$\phi(t_i) = \frac{\sum_{i=1}^n x_i K\left(\frac{t - t_i}{h}\right)}{\sum_{i=1}^n K\left(\frac{t - t_i}{h}\right)},$$

gdzie t to liczba sekund od początku sesji każdego dnia (ewentualnie od północy), natomiast t_i to liczba sekund od początku sesji (ewentualnie od północy) do momentu wystąpienia zmiany kierunku ceny akcji. x_i to czasy trwania odpowiadające momentom t_i . K jest wybraną funkcją jądra przy współczynniku wygładzania h .

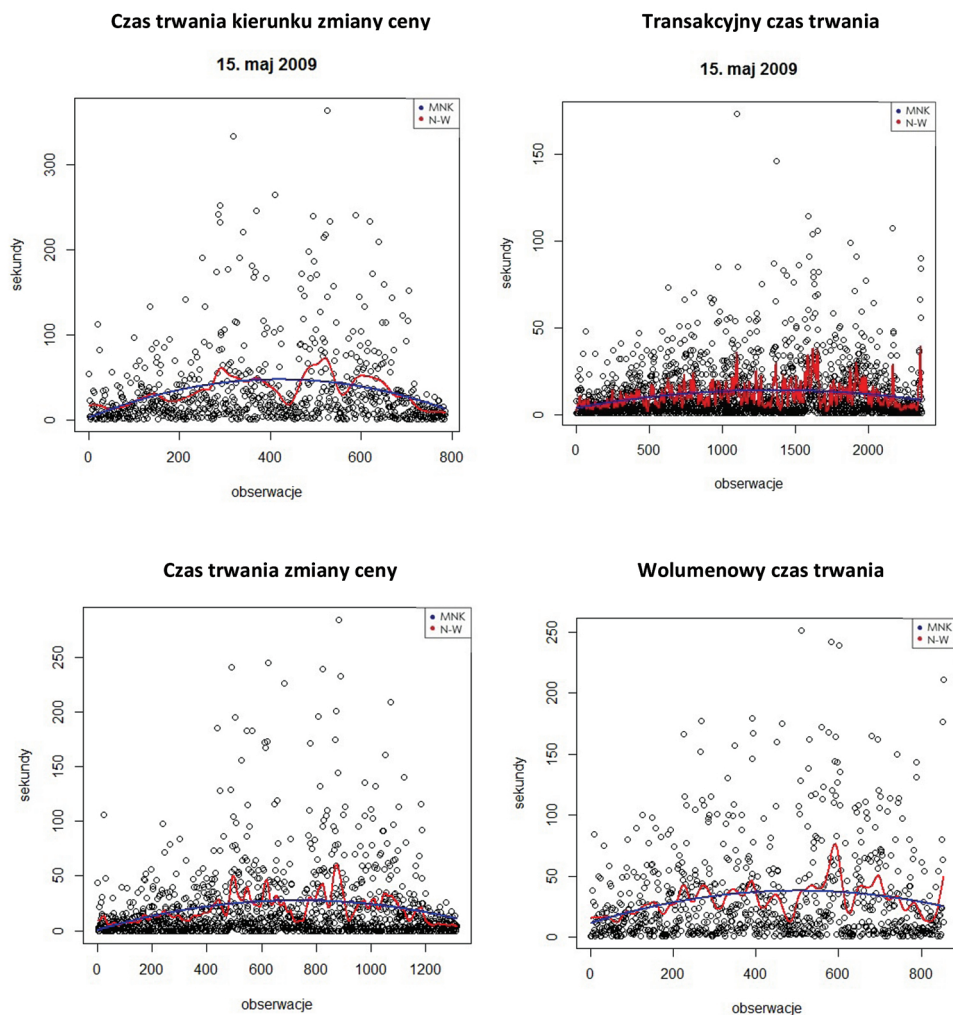
Przykład działania powyższej metody najlepiej zobrazuje zestawienie jej z dowolną regresją parametryczną. Rysunek poniżej (Wykres 5) przedstawia odsezonowanie przykładowych czasów trwania przy pomocy regresji Nadaraya i Watsona oraz regresji liniowej.

Ukazane są na nim wykresy rozrzutu, na których dla każdej obserwacji zaznaczona jest liczba sekund oczekiwania pomiędzy kolejnym zdarzeniem transakcyjnym. Krzywa naniesiona szarym kolorem symbolizuje przebieg regresji Nadaraya–Watsona, natomiast niebieska regresji liniowej estymowanej MNK. Piki krzywych oznaczają, że liczba sekund do następnego zdarzenia była duża, czyli intensywność zjawiska była niska. Natomiast kształtowanie się wykresu na niskim poziomie oznacza wysoką intensywność procesu transakcyjnego. Na przykładzie regresji Nadaraya i Watsona dla cenowego czasu trwania można zauważyć, że przy 400. obserwacji zmiany ceny zaczynają zachodzić coraz rzadziej, ponieważ na wykresie zaczynają pojawiać się piki. Związane jest to zapewne z tzw. efektem lunchu. Na początku i końcu sesji giełdowej wykres kształtuje się na niskich poziomach co jest typowe dla dziennego cyklu transakcyjnego, ponieważ częstość zmian cen w tych okresach jest bardzo duża, a więc czas oczekiwania na kolejną zmianę bardzo mały.

Przytoczone porównania wyraźnie ukazują, że regresja Nadaraya–Watsona lepiej dopasowuje się do danych niż regresja liniowa, a co za tym idzie wierniej oddaje efekt sezonowości wewnątrzdziennej.

Należy zaznaczyć, że wyniki danej estymacji nieparametrycznej zależą będą od wybranej funkcji jądrowej. W powyższym przykładzie użyto jądra normalnego. Wybór jądra jest ważny, ale nie aż tak jak właściwy wybór szerokości pasma.

Decyzja co do parametru wygładzania zależy od wielkości próby. W niniejszej pracy jest on równy $2,78sN^{-0,2}$ (s — odchylenie standardowe z próby, N — liczba obserwacji), natomiast zastosowaną funkcją K jest jądro normalne.



Źródło: opracowanie własne.

Wykres 5. Porównanie regresji Nadaraya i Watsona z regresją liniową

8. WYNIKI EMPIRYCZNE

Obliczenia przeprowadzone zostały przy użyciu pakietu statystycznego *R* oraz *Statistica*, natomiast przygotowanie danych dokonano za pomocą programu *MS Excel*. Dane transakcyjne pobrane zostały ze strony *www.stooq.pl*. Modelowaniu poddano czasy trwania pomiędzy kolejnymi zmianami kierunku cen akcji. Zdecydowano się na analizę czasów oczekiwania na zdarzenia tego typu ze względu na rzadkie występowanie podobnego badania w literaturze przedmiotu. Głównym celem było rozpatrzenie wpływu wewnątrzdziennej sezonowości występującej w rozważanym zbiorze danych. Do opisu kształtowania się czasów trwania kierunku zmiany cen wykorzystano różne typy modeli ACD, które uporządkowano ze względu na jakość dopasowania do danych. Analiza sprowadzała się do monitorowania ewentualnej zmiany rankingu modeli wobec odsezonowania danych metodą estymacji jądrowej.

Badany zbiór danych transakcyjnych zawiera zestawienie cen akcji i wolumenów spółki PKO BP z godziną ich pojawienia się oraz datą, w której dana sesja giełdowa nastąpiła. Przykładowy fragment zbioru danych wygląda następująco:

Tabela 6

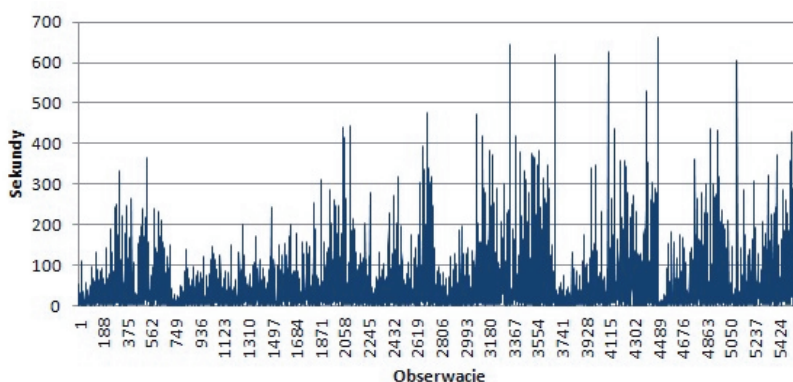
Fragment badanego zbioru danych

Spółka	Data	Godzina	Cena	Wolumen
PKOBP	20090504	09:00:53	27,41	80
PKOBP	20090504	09:00:56	27,45	360
PKOBP	20090504	09:01:18	27,40	100
PKOBP	20090504	09:01:21	27,44	40
PKOBP	20090504	09:01:21	27,40	2688

Źródło: opracowanie własne.

Dane zebrane zostały z dni od 4. do 15. maja 2009 roku, a więc obejmowały 2 tygodnie — 10 dni roboczych. Wybrano spółkę PKO BP, ponieważ w badanym okresie odznaczała się ona najwyższą płynnością spośród spółek portfela WIG20; zob. Huptas (2013). Na podstawie powyższych danych obliczono czas trwania kierunku zmiany ceny, czyli okres pomiędzy kolejnymi zmianami trendu procesu cenowego. Czas ten mierzono z dokładnością do jednej sekundy. Można zatem w rozważanym szeregu znaleźć obserwacje zerowe. Dzieje się tak gdy zdarzenie transakcyjne zajdzie w czasie krótszym niż 1 sekunda, czyli w tym przypadku — jeśli w przeciągu sekundy cena akcji zmieni swoją tendencję z rosnącej na malejącą (lub odwrotnie). Zerowe czasy trwania pociągają za

sobą problemy natury numerycznej, z tego powodu przed estymacją modeli skorygowano szereg dodając do każdego pomiaru 1 sekundę. Do analizy wzięto pod uwagę jedynie obserwacje dla fazy notowań ciągłych. Na GPW notowania ciągłe w 2009 roku miały miejsce od godziny 9:00 do 16:30 w dni robocze. Na potrzeby badania przyjęto „płynne” końce sesji giełdowej. Oznaczało to, że ostatni czas trwania zmiany kierunku ceny w dniu odpowiadał ostatniej pojawiającej się transakcji, a okres pomiędzy końcem jednego dnia a początkiem drugiego wynosi 0. Wykres poniżej obrazuje kształtowanie się poziomów czasów oczekiwania na kolejną zmianę trendu cenowego w badanym okresie. Zbudowany szereg składał się z 5587 obserwacji, które podlegały modelowaniu.



Źródło: opracowanie własne.

Wykres 6. Czas trwania kierunku zmiany ceny w dniach od 4. do 15. maja 2009 roku

W celu przybliżenia wstępnych charakterystyk zbioru danych obliczono statystyki opisowe, które przedstawia Tabela 7. Natomiast kształtowanie się gęstości szeregu ukazano na wykresie poniżej, który przedstawia histogram badanych czasów trwania.

Średnia liczba sekund jaka mijała między dwoma kolejnymi zmianami kierunku trendu cenowego wynosiła około 45. Połowa obserwacji nie przekraczała poziomu 22 sekund. Najczęstszą wartością w badanym szeregu była 1 sekunda. Maksymalnym zaobserwowanym czasem trwania były 662 sekundy, natomiast minimalnym — 1 sekunda. Wariancja szeregu wyniosła około 3918. Przeciętnie obserwacje różniły się od średniego czasu trwania o około ± 62 sekundy. Współczynnik zmienności na poziomie 1,39 wskazuje, że wartość odchylenia standardowego jest wyższa niż średniej. Obserwujemy zatem zjawisko nadmiernego rozproszenia badanej zmiennej, które jest typowe dla danych o wysokiej częstotliwości. Skośność na poziomie 3,27 mówi o dodatniej asymetrii rozkładu czasów oczekiwania, co również potwierdza histogram. Oznacza to, że większość obserwacji ma wartość niższą niż średnia. Wysoka dodatnia wartość kur-

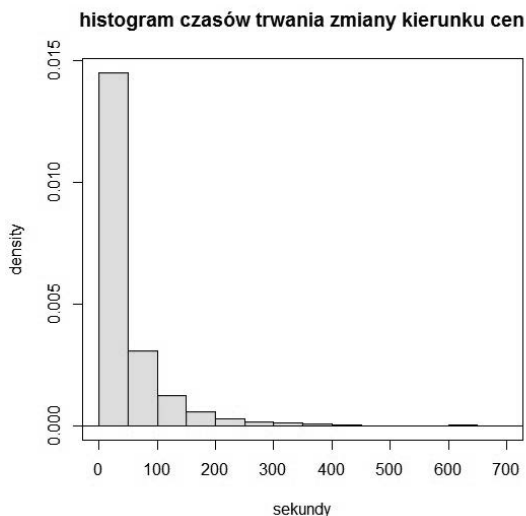
toży (16,1) jest konsekwencją dyskretnego charakteru analizowanych danych. Czasy trwania na zmianę kierunku ceny podlegają także istotnej dodatniej autokorelacji, co ukazuje poniższy autokorelogram.

Tabela 7

Statystyki opisowe badanych czasów trwania w sekundach

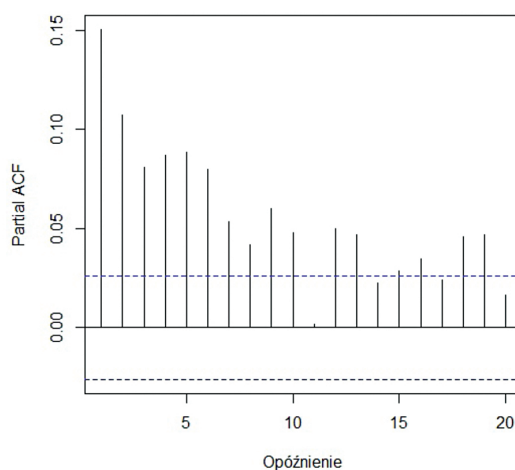
Średnia	45,08
Mediana	22,00
Modalna	1,00
Wariancja	3918,21 [s ²]
Odchylenie standardowe	62,60
Minimum	1,00
Maksimum	662,00
Współczynnik zmienności	1,39
Skośność	3,27 [s ³]
Kurtoza	16,10 [s ⁴]
ACF(1)	0,151

Źródło: opracowanie własne.



Źródło: opracowanie własne.

Wykres 7. Histogram badanych czasów trwania



Źródło: opracowanie własne.

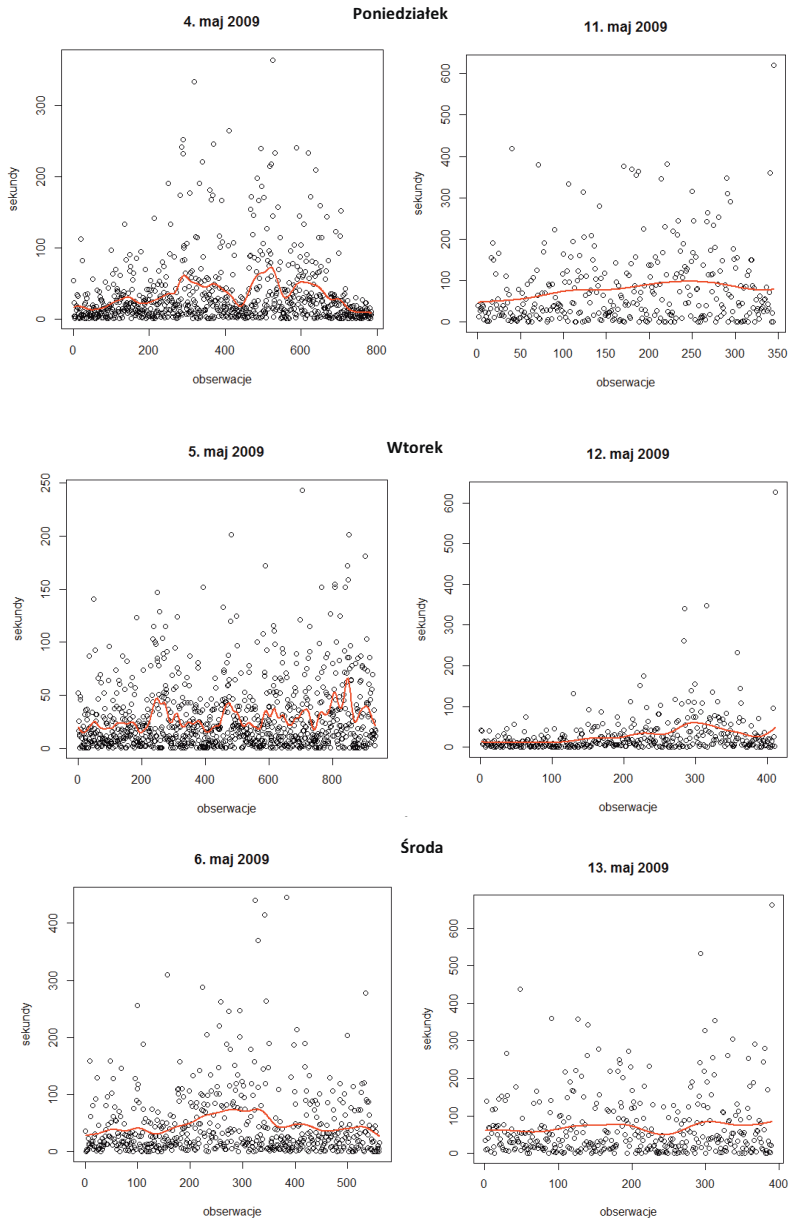
Wykres 8. Funkcja autokorelacji cząstkowej analizowanego czasu trwania

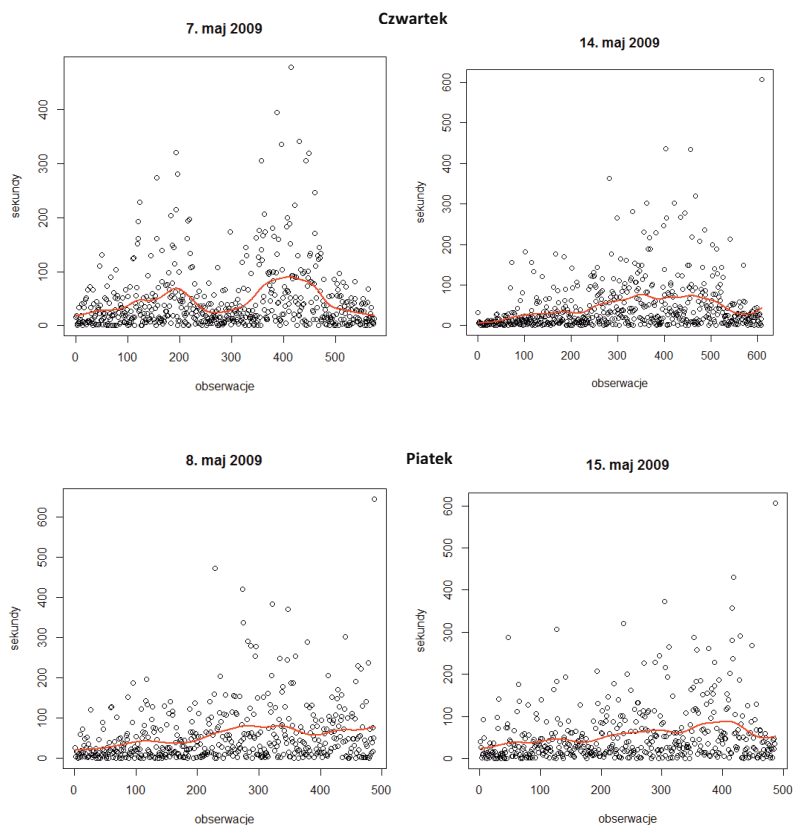
Wykres zbudowany został dla pierwszych 20 opóźnień. Wartość współczynnika autokorelacji pierwszego rzędu wynosi 0,151. Zauważalna jest malejąca tendencja funkcji autokorelacji cząstkowej. Dodatnia autokorelacja może być wynikiem efektu sezonowości wewnątrzdziennej. Uwzględnienie efektu wewnątrzdziennej cykliczności odbyła się przy pomocy regresji nieparametrycznej Nadaraya i Watsona z funkcją jądrową normalną. Metodę tę wykorzystano do określenia czynnika deterministycznego $\phi(t_i)$ dla odsezonowanego czasu trwania. Oszacowanie przeprowadzone zostało dla każdego dnia osobno. Odzworowanie wewnątrzdziennej sezonowości na podstawie wyników estymacji jądrowej przedstawiają wykresy poniżej.

Typowym kształtem odzworowania wewnątrzdziennej cykliczności transakcyjnych czasów trwania jest, opisywany w literaturze, kształt odwróconej litery „U” — Huptas (2013). Świadczy on o wysokiej intensywności zawierania transakcji w godzinach porannych, następnie spowolnieniu procesu transakcyjnego w godzinach około-południowych (tzw. „efekt lunchu”) i znów wysokiemu natężeniu transakcji pod koniec sesji giełdowej. Najbardziej podobny kształt odzworowania sezonowości można zaobserwować na wykresie z dnia 6 maja, jednak łatwo zauważyć, że w przypadku pozostałych badanych dni nie ma tego efektu silnie zarysowanego. Efekt cykliczności można tu uznać za słaby, dlatego badanie czy istotnie wpływa on na wybór najlepszego modelu opisującego czas oczekiwania jest zasadne.

Statystyki opisowe czasów trwania po usunięciu komponentu wewnątrzdziennej cykliczności zestawione zostały w Tabeli 8. Poniżej przedstawiony został również histogram odsezonowanych danych oraz wykres ich funkcji

autokorelacji cząstkowej. Można zauważyć, że dane nie charakteryzują się już istotną dodatnią autokorelacją rzędu 1. (współczynnik autokorelacji na poziomie 0,011). Zatem likwidacja efektu sezonowości wewnątrzdziennej wyeliminowała to zjawisko.





Źródło: opracowanie własne.

Wykres 9. Kształtowanie się wewnątrzdziennej sezonowości
w dniach od 4. do 15. maja 2009 roku

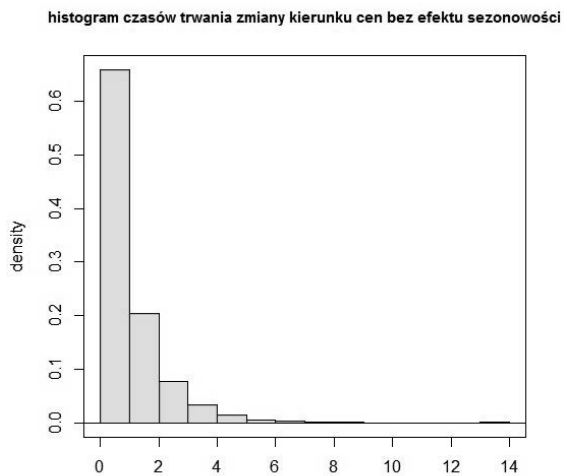
Do opisu czasu oczekiwania na zmianę kierunku ceny zastosowano modele warunkowego czasu trwania ACD — specyfikację klasyczną oraz logarymiczną z rozkładami warunkowymi wykładniczym i Weibulla. Estymacja parametrów przeprowadzona została Metodą Największej Wiarygodności. Modele zbudowano i oszacowano dla czasów trwania z efektem cykliczności wewnątrzdziennej oraz dla danych odsezonowanych. Wyniki przedstawiono poniżej w Tabelach 9 oraz 11. Tabele wynikowe podzielone są na dwie części pod względem postaci rozkładu warunkowego składnika losowego modeli. Dla każdego modelu podano wartości oceny MNW parametrów, ich błędy średnie szacunku oraz wartość *p-value*, czyli empiryczny poziom istotności, przy którym odrzuca się hipotezę zerową mówiącą o zerowaniu się danego parametru. Podano również kryteria informacyjne AIC obliczone dla każdego modelu.

Tabela 8

Statystyki opisowe odsezonowanych czasów trwania

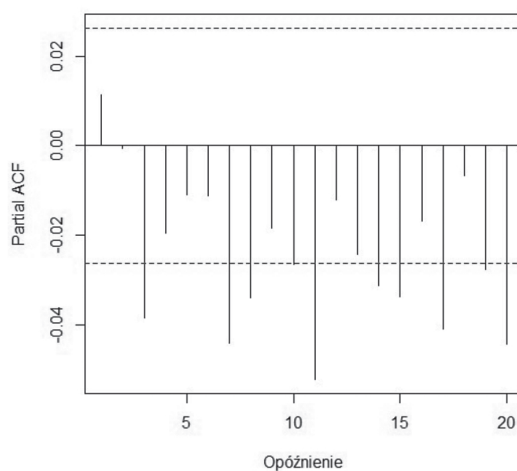
Liczba obserwacji	5587
Średnia	0,982
Mediana	0,609
Wariancja	1,278
Odchylenie standardowe	1,131
Minimum	0,010
Maksimum	13,989
Współczynnik zmienności	1,152
Skośność	2,728
Kurtoza	13,109
ACF(1)	0,011

Źródło: opracowanie własne.



Źródło: opracowanie własne.

Wykres 10. Histogram badanych czasów trwania



Źródło: opracowanie własne.

Wykres 11. Funkcja autokorelacji cząstkowej analizowanego czasu trwania bez efektu sezonowości

Tabela 9

Estymacja MNW parametrów modeli ACD dla danych z wewnątrzdziennej sezonowością

Warunkowy rozkład wykładniczy							
Model ACD				Model log-ACD			
AIC		52803,42		AIC		52758,37	
Parametr	Ocena	Błąd śr. szacunku	p	Parametr	Ocena	Błąd śr. szacunku	p
ω	0,5095	0,10	0,00	ω	0,0743	0,010	0,00
α	0,0628	0,01	0,00	α	0,0502	0,003	0,00
β	0,9268	0,01	0,00	β	0,9398	0,004	0,00
Warunkowy rozkład Weibulla							
Model ACD				Model log-ACD			
AIC		52499,18		AIC		52478,25	
Parametr	Ocena	Błąd śr. szacunku	p	Parametr	Ocena	Błąd śr. szacunku	p
ω	0,5128	0,11	0,00	ω	0,0803	0,010	0,00
α	0,0652	0,01	0,00	α	0,0532	0,004	0,00
β	0,9242	0,01	0,00	β	0,9357	0,005	0,00
γ	0,8466	0,01	0,00	γ	0,8503	0,010	0,00

Źródło: opracowanie własne.

Można zauważyć, że wszystkie parametry w każdym z analizowanych modeli są statystycznie istotne, gdyż wartości p -value są mniejsze od założonego poziomu istotności 0,05. Warunki nieujemności parametrów modeli ACD są spełnione. Dowodem na zakładaną własność nadmiernego rozproszenia danych jest to, że oceny parametrów α modeli klasycznych są większe od 0. Warunek kowariancyjnej stacjonarności dla formuły podstawowej na warunkową wartość oczekiwaną Ψ_i został spełniony zarówno w przypadku modelu z warunkowym rozkładem wykładniczym jak i rozkładem Weibulla, ponieważ suma parametrów $\alpha + \beta < 1$. Także dla każdej oszacowanej specyfikacji log-ACD proces jest stacjonarny, gdyż wartości parametrów β są na moduł mniejsze od 1.

W przypadku podstawowej formuły ACD możliwym jest określenie bezwarunkowej wartości oczekiwanej modelu $E(x)$, ponieważ istnieje jej analityczna postać. Dla modelu ACD z warunkowym rozkładem wykładniczym średni bezwarunkowy czas oczekiwania na kolejną zmianę kierunku cen akcji PKO BP wyniósł około 49 sekund. Natomiast średni bezwarunkowy czas trwania zmiany kierunku cen w przypadku modelu ACD z warunkowym rozkładem Weibulla jest równy około 48 sekund. Rozkład warunkowy rozważanych modeli nie ma wpływu na jakość oszacowania tej własności procesu.

Zależnie od wybranego rozkładu warunkowego można zauważyć, że dla tych samych specyfikacji wartości ocen MNW się różnią. Dla podstawowego ACD parametry ω mają bardzo podobny poziom, jednak parametr α w modelu z warunkowym rozkładem Weibulla jest wyższy o 3% od jego odpowiednika w modelu z warunkowym rozkładem wykładniczym. Parametr β zaś w modelu z rozkładem Weibulla jest niższy od tego w modelu z rozkładem wykładniczym o około 1%. Największą różnicę można zaobserwować, gdy porównuje się wartości parametrów γ . Należy mieć na uwadze, że w przypadku specyfikacji z warunkowym rozkładem wykładniczym parametr ten sprowadza się do 1, jest on zatem wyższy o ponad 15% od γ oszacowanego dla modelu z warunkowym rozkładem Weibulla. W przypadku modeli log-ACD można zaobserwować większą różnicę w wartościach parametrów. Parametr ω modelu z wykładniczym rozkładem warunkowym jest o 8% niższy niż jego odpowiednik w modelu z rozkładem Weibulla. Również większa zmiana występuje dla α , gdyż parametr ten w modelu z warunkowym rozkładem Weibulla jest o 5% wyższy od α w modelu z warunkowym rozkładem wykładniczym. Parametr β zachowuje się podobnie jak dla specyfikacji klasycznej ACD — różnica jest taka sama. Również sytuacja dla parametru γ jest analogiczna, ponieważ w modelu z rozkładem warunkowym Weibulla jest on niższy o 15% niż jego odpowiednik w modelu z warunkowym rozkładem wykładniczym.

Z zaprezentowanego zestawienia można również odczytać jak dobrze dany model w stosunku do innych dopasowuje się do danych. Mówi o tym wartość

kryteriów informacyjnych AIC. Najlepsza specyfikacja to ta o najniższej wartości AIC — została ona podkreślona w powyższym zestawieniu. Na podstawie wartości kryteriów zbudowano ranking modeli ACD, przedstawiony poniżej.

Tabela 10

Ranking modeli ACD dla danych z wewnątrzdziennej sezonowością

Pozycja	Model	AIC
1.	log-ACD z warunkowym rozkładem Weibulla	52478,25
2.	ACD z warunkowym rozkładem Weibulla	52499,18
3.	log-ACD z warunkowym rozkładem wykładniczym	52758,37
4.	ACD z warunkowym rozkładem wykładniczym	52803,42

Źródło: opracowanie własne.

Modelem, który jest najlepszy do opisu czasu trwania zmiany kierunku cen akcji PKO BP, dla danych niepoddanych eliminacji wewnątrzdziennej sezonowości, jest model log-ACD z warunkowym rozkładem Weibulla.

Pierwsza pozycja w rankingu tej specyfikacji została podparta testami statystycznymi przybliżonym t -Studenta oraz ilorazu wiarygodności. Sprawdzone hipotezę zerową mówiącą o tym, że parametr γ jest równy 1. Statystyki testowe $t = -15,34$ dla podstawowej specyfikacji oraz $t = -14,97$ dla modelu log-ACD znajdują się w obszarze krytycznym, zatem wskazuje one na odrzucenie hipotezy zerowej na poziomie istotności równym 0,05. Oznacza to, że warunkowy rozkład wykładniczy (dla którego $\gamma = 1$) jest mniej preferowany. W teście ilorazu wiarygodności hipoteza zerowa mówiła o tym, że rozkładem składników losowych jest rozkład wykładniczy, natomiast hipoteza alternatywna wskazuje, że innowacje mają rozkład Weibulla. Dla klasycznych modeli ACD statystyka testowa $LR = 306,24$ zaś dla modeli log-ACD $LR = 282,12$. Obie statystyki znajdują się w obszarze krytycznym, dlatego hipoteza zerowa została odrzucona na poziomie istotności 0,05, na korzyść hipotezy alternatywnej mówiącej o tym, że składniki losowe mają warunkowy rozkład Weibulla.

Dla danych bez wewnątrzdziennej cykliczności wyniki estymacji zostały przedstawione poniżej, w Tabeli 11. Tabela wynikowa wskazuje, że większość parametrów jest statystycznie nieistotna na powszechnie przyjmowanych poziomach istotności. Błędy średnie szacunku są na o wiele wyższym poziomie niż w przypadku modelowania danych z efektem sezonowości. Można zauważyć, że warunek nieujemności w przypadku modelu ACD z warunkowym rozkładem wykładniczym nie został spełniony dla parametru β , jednak wynikać to będzie z błędu numerycznego; dla innych parametrów modeli klasycznych założenie to jest spełnione. Oceny parametrów α specyfikacji podstawowych

są dodatnie, tak więc zjawisko nadmiernego rozproszenia danych nadal występuje, nawet po eliminacji efektu cykliczności wewnątrzdziennej. Parametry β każdego oszacowanego modelu są na moduł mniejsze od 1 co wskazuje, że procesy te są stacjonarne.

Tabela 11

Estymacja MNW modeli ACD dla danych bez efektu wewnątrzdziennej sezonowości

Warunkowy rozkład wykładniczy							
Model ACD				Model log-ACD			
AIC		10974,03		AIC		10972,89	
Parametr	Ocena	Błąd śr. szacunku	p	Parametr	Ocena	Błąd śr. szacunku	p
ω	0,9792	0,36	0,0059	ω	-0,0133	0,02	0,444
α	0,0121	0,01	0,3279	α	0,0048	0,01	0,6434
β	-0,0098	0,36	0,9783	β	0,1186	0,47	0,8014
Warunkowy rozkład Weibulla							
Model ACD				Model log-ACD			
AIC		10880,54		AIC		10869,91	
Parametr	Ocena	Błąd śr. szacunku	p	Parametr	Ocena	Błąd śr. szacunku	p
ω	0,9562	0,41	0,0189	ω	-0,0464	0,03	0,1075
α	0,0108	0,01	0,4233	α	-0,0082	0,00	0,0176
β	0,0131	0,42	0,9747	β	-0,9463	0,02	0,0000
γ	0,9066	0,01	0,0000	γ	0,907	0,01	0,0000

Źródło: opracowanie własne.

Na podstawie oszacowanych ocen MNW obliczono bezwarunkowe wartości oczekiwane dla specyfikacji klasycznych ACD. Dla modelu z warunkowym rozkładem wykładniczym bezwarunkowa wartość oczekiwana czasu trwania zmiany kierunku ceny bez efektu sezonowości wyniósł około 0,9814, natomiast dla modelu z warunkowym rozkładem Weibulla $E(x) = 0,9796$.

Można zauważyć, że ze względu na wybrany rozkład warunkowy, parametry w tych samych specyfikacjach modeli różnią się diametralnie. Przykładem może być parametr ω w modelach logarytmicznych — dla modelu z warunkowym rozkładem wykładniczym jest aż 350% niższy od swego odpowiednika

w modelu z warunkowym rozkładem Weibulla. Również znaki stojące przy danych parametrach różnią się w tych samych specyfikacjach. Można na podstawie tej obserwacji wysnuć wniosek, że dla danych odsezonowanych stabilność parametrów jest o wiele niższa niż dla tych z efektem cykliczności wewnątrz-dziennej. Parametry γ dla modeli z warunkowym rozkładem Weibulla (zarówno klasycznych jak i log-ACD) są o 10% niższe od γ w modelach z warunkowym rozkładem wykładniczym.

Najniższą wartość AIC, a zarazem najlepszy model, zaznaczono w powyższej tabeli wyników. Ranking modeli ACD dla danych odsezonowanych przedstawiono poniżej.

Tabela 12

Ranking modeli ACD dla danych bez efektu wewnątrzdziennej sezonowości

Pozycja	Model	AIC
1.	log-ACD z warunkowym rozkładem Weibulla	10869,91
2.	ACD z warunkowym rozkładem Weibulla	10880,54
3.	log-ACD z warunkowym rozkładem wykładniczym	10972,89
4.	ACD z warunkowym rozkładem wykładniczym	10974,03

Źródło: opracowanie własne.

Na podstawie kryterium informacyjnego AIC najlepszym modelem dla danych poddanych odsezonowaniu jest specyfikacja log-ACD z warunkowym rozkładem Weibulla. Jak wcześniej, poprawność rankingu potwierdzają przeprowadzone testy statystyczne. Ponownie w przybliżonym teście *t*-Studenta hipoteza zerowa mówiła o tym, że parametr $\gamma = 1$. Odpowiednio dla specyfikacji podstawowej ACD i logarytmicznej ACD statystyki testowej kształtują się na następujących poziomach: $t = -9,34$ oraz $t = -9,3$. Oznacza to, że na poziomie istotności równej 0,05 znajdują się one w obszarze krytycznym, a więc przy tej istotności hipoteza zerowa została odrzucona. Dla danych odsezonowanych również warunkowy rozkład wykładniczy nie jest preferowany, co odzwierciedlone jest w rankingu. Podobnie jak wcześniej, w teście ilorazu wiarygodności hipoteza zerowa mówi, że rozkładem składników losowych jest rozkład wykładniczy, natomiast hipoteza alternatywna głosi, że rozkładem innowacji jest rozkład Weibulla. Statystyki testowe $LR = 95,47$ dla modeli podstawowych oraz $LR = 104,98$ dla modeli log-ACD znajdują się w obszarze krytycznym przy istotności na poziomie 0,05. Zatem odrzucona została hipoteza zerowa na korzyść alternatywnej, która głosi, że rozkładem warunkowym jest rozkład Weibulla.

Przeprowadzona analiza ukazała, że pomimo eliminacji efektu sezonowości wewnątrzdziennej ranking oszacowanych modeli ACD nie zmienia się. Niezależnie czy estymuje się dane z sezonowością czy bez, najlepszym modelem jest ten najbardziej skomplikowany pod względem specyfikacji i o ogólniejszej postaci rozkładu warunkowego, czyli model log-ACD z warunkowym rozkładem Weibulla. Fakt, że ranking modeli po odsezonowaniu danych się nie zmienia świadczy o tym, że poddawanie szeregu eliminacji cykliczności wewnątrzdziennej jest zbędne. Może to być związane z tym, że efekt sezonowości dla czasów oczekiwania na kolejną zmianę kierunku ceny jest słaby. Badacz nie musi zatem stosować skomplikowanych metod likwidowania efektu sezonowości, jeśli analizuje czas trwania kierunku zmiany ceny.

9. PODSUMOWANIE

Jednym z celów niniejszej pracy był wybór najlepszego modelu ACD opisującego kształtowanie się czasów oczekiwania na kolejną zmianę trendu cenowego akcji PKO BP. Rozważane były tezy, czy rozszerzone modele ACD lepiej modelują badany zbiór danych od podstawowych specyfikacji, jak również czy rozkład warunkowy powinien przyjmować bardziej ogólną postać. Kolejnym celem była analiza wpływu odsezonowania danych metodą estymacji jądrowej na zachowanie się rankingu dopasowania modeli ACD. Dokonano dekompozycji czasów trwania kierunku zmiany ceny na część stochastyczną oraz deterministyczną. Następnie dla każdego badanego dnia osobno oszacowano kształtowanie się funkcji sezonowości wewnątrzdziennej, na podstawie których można było określić jak silny jest efekt cykliczności dla tego typu danych.

Podjęto analizy z wykorzystaniem modeli warunkowego czasu trwania, gdyż pomimo tego, że w literaturze światowej modele te mają już ugruntowaną pozycję, to w Polsce obserwuje się nieliczne prace z zakresu ich zastosowań, zob. Doman (2011), Huptas (2013), Bień (2005). Również decyzja co do badania czasów trwania kierunku zmiany ceny powodowana była tym, że tego typu dane są bardzo rzadko rozważane w literaturze przedmiotu — wiedza na temat ich własności nie jest tak mocno ugruntowana jak na temat właściwości czasów transakcyjnych lub cenowych czasów trwania.

Analiza empiryczna dotyczyła oszacowania modeli warunkowego czasu trwania dla czasów oczekiwania pomiędzy kolejną zmianą kierunku ceny akcji PKO BP w dniach od 4 do 15 maja 2009 roku. Sprawdzone własności badanego szeregu, czy pokrywały się z typowymi własnościami danych transakcyjnych. Można było zaobserwować cechy nadmiernego rozproszenia danych oraz dodatniej autokorelacji spowodowanej występowaniem efektu sezonowości śróddziennej. Zbudowano i oszacowano Metodą Największej Wiarygodności 4 modele dla danych z komponentem sezonowości wewnątrzdziennej: ACD

z warunkowym rozkładem wykładniczym, ACD z warunkowym rozkładem Weibulla, log-ACD z warunkowym rozkładem wykładniczym oraz log-ACD z warunkowym rozkładem Weibulla. Estymację powtórzono dla danych odsezonowanych, po ówczesnej eliminacji efektu cykliczności wewnątrzdziennej przy pomocy regresji Nadaraya i Watsona. W sumie uzyskano zatem 8 modeli. Estymacja wskazała, że dla danych odsezonowanych jednym modelem, którego wszystkie parametry strukturalne były istotne był log-ACD z warunkowym rozkładem Weibulla. Dzięki oszacowaniom MNW obliczono kryteria informacyjne AIC dla każdego badanego modelu. Na podstawie wyznaczonych wartości AIC zbudowano dwa rankingi modeli ACD — odpowiednio dla danych z efektem sezonowości oraz tych odsezonowanych. Najlepszym modelem opisujących oba typy danych okazał się log-ACD z warunkowym rozkładem Weibulla. Potwierdzeniem zajętych pozycji w rankingu były przeprowadzone testy statystyczne, które wykazały wyższość modeli z warunkowymi rozkładami Weibulla. Porównanie rankingów wykazało, że pozycja żadnego z rozważanych modeli nie zmienia się. Oznaczać to będzie, że poddawanie czasów trwania zmiany kierunku ceny zabiegom mającym na celu eliminację sezonowości nie jest konieczne.

Wykazano, że dla badanego szeregu efekt cykliczności wewnątrzdziennej jest na tyle słaby, że odsezonowanie danych nie wpłynie na ranking jakości dopasowania modeli ACD. Jednak może być to prawdą jedynie w przypadku, gdy do eliminacji sezonowości wykorzystana się opisaną regresję Nadaraya i Watsona. Dlatego bardziej wnikliwa analiza własności czasów trwania kierunku zmiany ceny wydaje się być atrakcyjnym polem dalszych badań.

BIBLIOGRAFIA

- Adresen T.G., Davis R.A., Kreiss J.F., Mikosch T. (2009), *Handbook of Financial Time Series*, Springer.
- Bauwens L., Giot P. (2001), *Econometric Modelling of Stock Market Intraday Activity*, Kluwer Academic Publishers, Boston.
- Bauwens L., Giot P. (2000), *The Logarithmic ACD Models: An application to the bid-ask quote process of three NYSE stocks*, *Annales d'Economie et de Statistique* 60.
- Bauwens L., Giot P. (2008), *The Moments of Log-ACD Models*, *Quantative and Qualitative Analysis in Socila Science*, vol. 2.
- Bień K. (2005), *Wybrane modele ekonometryczne finansowych szeregów czasowych o ultrawysokiej częstotliwości*, Szkoła Główna Handlowa w Warszawie.
- Doman M. (2011), *Mikrostruktura giełd papierów wartościowych*, Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu.
- Doman M. (2006), *Modelowanie Mikrostruktury Polskiego Rynku Kapitałowego*, *Zeszyty Naukowe Katedry Uniwersytetu Ekonomicznego w Poznaniu* 78, s. 150–166.
- Engle R.F., Russel J.R. (1998), *Autoregressive Conditional Duration: A new model for irregularly spaced transaction data*, *Econometrica*, vol. 66, no. 5.
- Härdle W. (1992), *Applied Nonparametric Regression*, Cambridge University Press.

- Heiler S. (1999), *A survey on Nonparametric Time Series Analysis*, CoFE Discussion Paper no. 9906, University of Konstanz.
- Huptas R. (2013), *Modele ACD i wnioskowanie bayesowskie w analizie danych transakcyjnych z polskiego rynku akcji*, Uniwersytet Ekonomiczny w Krakowie.
- Kufel T. (2007), *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRET*, Wydawnictwo Naukowe PWN, Warszawa.
- Madhavan A. (2000), *Market Microstructure: A survey*, Journal of Financial Markets 3.
- O'Hara M. (1995), *Market Microstructure Theory*, Blackwell, Cambridge MA.
- Pacurar M. (2006), *Autoregressive Conditional Duration (ACD) Models in Finance: A survey of the Theoretical and Empirical Literature*, School of Business Administration, Dalhousie University.