

2017 -10- 30

Wrocław, dnia 25 października 2017

Dr hab. Dariusz Grech, prof. UWr
Instytut Fizyki Teoretycznej
Wydział Fizyki i Astronomii
Uniwersytet Wrocławski
Pl. Maksa Borna 9
50-204 Wrocław

**Recenzja rozprawy doktorskiej mgr. Mateusza Denysa pt.:
„Analysis of interevent times by methods of statistical physics”**

Ogólna charakterystyka rozprawy

Przedstawiona mi do recenzji rozprawa doktorska została napisana w języku angielskim w postaci mini - monografii w formie ustawowo dopuszczonej dla tego typu rozpraw. Składa się z sześciu rozdziałów, bardzo obszernego i wyczerpującego spisu literatury obejmującego 215 pozycji oraz dodatku z wyprowadzeniem podstawowej dla pracy doktorskiej formuły (3.12) przedstawiającej ogólny wzór na rozkład gęstości prawdopodobieństwa czasów oczekiwania między kolejnymi zdarzeniami progowymi (ang. *interevent times*).

Praca doktorska liczy w sumie ponad 130 stron i napisana jest starannie, dobrym językiem. Ilość i rodzaj przedstawianych wykresów w celu poparcia prezentowanych wyników badań jest wystarczająca, choć w wielu wypadkach dobre dopasowanie proponowanych zależności funkcyjnych do danych empirycznych nie jest poparte odpowiednimi testami statystycznymi co należy uznać za pewien mankament w pracy, nie wpływający jednak decydująco na ostateczne główne wyniki uzyskane i prezentowane w doktoracie.

Głównym celem rozprawy jest analiza teoretyczna i numeryczna rozkładów przerw czasowych (ang. *interevent times*) w szeregach czasowych między przyrostami szeregu przewyższającymi co do wartości pewien ustalony próg ilościowy Q . Innymi słowy badane są odstępów między kolejnymi zdarzeniami ekstremalnymi definiowanymi jako osiągnięcie lub przekroczenie (co do wartości bezwzględnej) ustalonej wielkości progu tego zdarzenia ekstremalnego (np. w przypadku finansowych szeregów czasowych takim zdarzeniem ekstremalnym jest przekroczenie przez zwrot ustalonej wartości maksymalnej lub minimalnej podanej w %). Badania tych zjawisk nasiliły się i zyskały znaczne zainteresowanie po serii prac w latach 2008-2014 opublikowanych przez Mikhaila Bogacheva i Armina Bunde (2008, 2009) oraz później przez Josefa Ludeschera, Constantino Tsallisa i Armina Bunde w latach 2011 i 2014. Autorzy ci, należący do światowej czołówki aktywnych obecnie ekonofizyków, odkryli ciekawą zależność empiryczną dla danych giełdowych pokazując, że rozkład czasów między zdarzeniami progowymi ma charakter uniwersalny – nie zależy on bowiem ani od rodzaju papieru wartościowego ani od częstości próbkowania danych w szeregu czasowym zwrotów (innymi słowy brak jest istotnej zależności od przesunięcia czasowego – ang. *time-lag*, dla którego zwrot waloru jest liczony). Rozkład tych przerw czasowych zależy w istocie tylko od wartości progu Q , którego kolejne przekroczenia przez proces stochastyczny są analizowane. Oryginalnie pierwszy opis rozkładu czasów między zdarzeniami progowymi opierał się na znakomitym dopasowaniu przez funkcję q -wykładniczego zaniku (ang. q – *exponent function*) i stąd też brał się udział Constantino Tsallisa w tych pracach, który jest



bardzo przywiązany do wprowadzonego przez siebie do fizyki statystycznej w latach '80 ub. wieku opisu opartego na tzw. q -deformacjach (por. q -entropia Tsallisa).

W swojej pracy doktorskiej Pan mgr Denys rozszerza powyższy opis stawiając pytania o przyczyny tej uniwersalności i podejmuje zakończoną sukcesem próbę jej opisu zarówno analityczną (w oparciu o znaną w fizyce koncepcję błędzenia losowego w czasie ciągłym CTRW – ang. *Continuous Time Random Walk*, wprowadzoną do opisu dyfuzji anomalnej jeszcze w latach '60 ub. wieku przez Montrolla i Weissa) jak i numeryczną opierającą się na własnym modelu agentów na sieci kwadratowej.

Szczegółowa analiza rozprawy

Rozprawę można podzielić na dwie zasadnicze części. W części pierwszej (rozdziały 2 i 3) prezentowany jest teoretyczny opis CTRW i wyprowadzony stąd, mający bardzo ogólny charakter, rozkład czasów między zdarzeniami progowymi (równanie 3.12). Tak naprawdę jednolity zapis związku zawiera w sobie co najmniej 4 różne ilościowo przypadki – dwa dla różnych relacji między czasem oczekiwania na wydarzenie progowe $\Delta_Q t$ a czasem relaksacji $\tau_Q^\pm$ rozpadające się dodatkowo na podprzypadek, gdy tworzy się klastery volatylity „zacieśniających się zdarzeń” i gdy wydarzenia progowe stają się coraz rzadsze (zwane w rozprawie *expanding hierarchy*), tj., gdy osiąganie progu Q staje się coraz rzadsze.

W tym miejscu muszę przyznać, że nie jest dla mnie jasne jak te dwa różne przypadki zostały wplecione w jednolite studia nad rozkładem $\Delta_Q t$ dla danych rzeczywistych? Czy to oznacza, że przedstawione wykresy w pracy zawierają już dane podzielone na okresy różnych zachowań, którym przypisano inne parametry relaksacji? Z samych tabel i wykresów przedstawionych w części pierwszej to raczej nie wynika.

Następnie porównuje się uzyskany rozkład teoretyczny z danymi empirycznymi, które były podstawą wspomnianych pionierskich prac Bogacheva, Bunde, Ludeschera i Tsallisa. Doktorant pokazuje, że wprowadzony przez niego i wyprowadzony teoretycznie rozkład czasów między zdarzeniami progowymi jest konkurencyjny w stosunku do prac poprzedników dając równie dobre dopasowanie do danych, a dodatkowo mając silne umocowanie (uzasadnienie) natury fizycznej (dowód oparty o CTRW) i matematycznej (wykorzystanie twierdzenia o wartościach ekstremalnych w skumulowanych rozkładach podprogowych Fishera-Tippetta-Gnedenko). Takiego umocowania rozkład q -wykładniczy oczywiście nie ma i w pracach wspomnianych autorów przyjęto go bez uzasadnienia jedynie jako wynik czysto empiryczny.

W drugiej części pracy doktorskiej w rozdziale 5. Doktorant prezentuje swój własny model sprytnych agentów (ang. *cunning agents*) oparty na strukturze spinowej modelu Potts'a. Model w rozprawie jest znacznym udoskonaleniem modelu Sieczki i Hołysta z 2008 roku. W prezentowanej przez Doktoranta modyfikacji modelu Sieczki i Hołysta bardzo dobrze odtwarza się rozkład czasów między zdarzeniami progowymi wyprowadzony na początku rozprawy. Należy podkreślić, że uzyskana zgodność zachowania się indeksu giełdowego NASDAQ (bo w oparciu o te dane prowadzone jest tu porównanie) zarówno z modelem teoretycznym jak i danymi empirycznymi jest zdumiewająca. Jest to pierwszy według mojej wiedzy model agentów oparty na głębokich przesłankach fizycznych tak doskonale symulujący autentyczne zjawiska na giełdzie włącznie ze świetnym odtworzeniem obserwowanych *interevent times* oraz dodatkowo pozostający w pełnej zgodzie z wyprowadzonym w pierwszej części rozprawy wzorem analitycznym opartym na czysto fizycznych przesłankach. Odkryta wcześniej uniwersalność pozwala sądzić, że uzyskana zgodność będzie miała miejsce też i dla innych danych giełdowych. Czy również dla danych spoza giełdy? Częściowa twierdząca odpowiedź pojawia się w rozprawie w rozdziale 3 na

rys. 3.10 ukazującym zgodność przewidywań modelu z rozkładem przerw czasowych między wstrząsami geotektonicznymi.

Znaczna część pracy doktorskiej (rozdział 4) omawia własności różnych modeli agentów, które zdobyły jak dotąd dość dużą popularność w literaturze przedmiotu (ekonomii, ekonofizyce) jak modele: Kim–Markowitza, Levy–Salomona, Luxa–Marchesi, Conta–Bouchaud, czy modele spinowe lori, Bornholdta i Gontisa. W większości (poza modelem Gontisa, który jako jedyny w tej grupie twórców modeli oddziałujących agentów badał rozkład czasów między zdarzeniami progowymi) modele te nie mają bezpośredniego związku z doktoratem, ale ich omówienie w jednym miejscu ma charakter ciekawego przeglądu mającego niewątpliwie walor dydaktyczny (szkoleniowy) i może być wartościowe dla czytelnika rozprawy. Słusznie przy tym Doktorant zauważa, że modelowanie numeryczne czasów między zdarzeniami progowymi jest bardzo młode w literaturze, bo poruszone jak dotąd tylko w publikacjach: V. Gontis, S. Havlin, A. Kononovicius, B. Podobnik, H. E. Stanley, *Physica A* 426 (2016) 1091 i w publikacji samego Doktoranta – szkoda, że tylko w wersji arXiv: (M. Denys, T. Gubiec, R. Kutner, arXiv: 1411.1689 2014).

Eksploracja tego zagadnienia dopiero się więc zaczyna. Ze względu na znakomity skład autorów pierwszej publikacji daje to uzasadnione nadzieje że problem jest ciekawy i nie pozostanie niezauważony. Tym bardziej należy docenić wkład Doktoranta do tej tematyki na wczesnym etapie jej rozwoju. Rozprawa wpisuje się zatem doskonale w jeden z nurtów badań współczesnej ekonofizyki, a rozdział 5. otwiera nowe pole badań nad specyfikacją warunków odtwarzających rzeczywiste czasy między zdarzeniami progowymi.

Według mnie, rozdział 5. stanowi najciekawszą część rozprawy, jest najlepiej napisany i zawiera najważniejsze wyniki, które wskazują też przyszłe ważne do podjęcia kierunki badań. Zwłaszcza istotne są paragrafy: 5.2.1 – wprowadzenie do modelu szumu o rozkładzie ciężko-ogonowym, 5.2.2 – wprowadzenie zwrotu waloru związanego ze zmianą spinu agenta a nie z samą wartością spinu, co ma uzasadnienie w istniejącym w rzeczywistości wysyceniu możliwości pozyskania kapitału (kupno) lub płytkim rynkiem (brak ofert kupna w przypadku sprzedaży) oraz podrozdział 5.3 ukazujący świetne jednoczesne odtworzenie kilku zjawisk znanych z prawdziwego rynku giełdowego, jak klastrowanie zmienności (ang. *volatility clustering*), grube zanikające potęgowe ogony rozkładów zwrotów, szybko zanikająca funkcja liniowej autokorelacji zwrotów przy zauważalnych jednocześnie potęgowo zanikających autokorelacjach nieliniowych (czyli obecności długiej pamięci w układzie) i znakomite odtworzenie rozkładów czasów międzyzdarzeniowych (rys. 5.12). Dotąd takie połączenie wyników w modelu agentów nikomu się nie udało! Słusznie przy tym wskazuje się w doktoracie, że należałoby jeszcze sprawdzić jakie własności multifraktalne (znane z danych empirycznych) ma proponowany model numeryczny (rzeczywistość jest bowiem multifraktalna) lub jak te własności uzyskać, jeśli ich w modelu jeszcze nie ma. W tym miejscu należałoby też postawić pytanie czy i jak poszczególne „składowe cegiełki” modelu mgr. Denysa, w szczególności postać i wielkość szumu, wartości liczbowe parametrów modelu, w tym wartość progu λ aktywacji agenta (równanie 5.1) wpływają na makroskopowe własności modelu odtwarzające obserwowane w rzeczywistości zjawiska. Dokładniej, czy wpływ poszczególnych „cegiełek składowych” modelu jest niezależny czy ma też charakter bardziej emergentny? Nie wiem czy nie jest to pytanie zbyt trudne na obecnym etapie, ale na pewno warte kontynuacji badań w tym kierunku. Doktorant wyznaczył bowiem pewien nowy standard numerycznych modeli agentów, które w ogóle mają głębszy pozaakademicki sens. Czynnikiem różnicującym jest tu właśnie dodatkowo zgodność z obserwacjami empirycznymi w rzeczywistych dowolnych układach złożonych przewidywanego przez model rozkładu czasów między zdarzeniami progowymi (*interevent times*), mającego jak się wydaje charakter uniwersalny – a więc

nieograniczonego tylko do giełdy papierów wartościowych. Rozkłady takiego prawdopodobieństwa określone są, jak uzasadnia Doktorant, przez zaledwie dwa wolne parametry (p. wzór 3.12).

Odkryta uniwersalność i jej opis teoretyczny winna zatem zaowocować zastosowaniami interdyscyplinarnymi również poza ekonomią i giełdą papierów wartościowych. Doktorant pokazuje w pracy doktorskiej przykład takiego zastosowania w geofizyce przy analizie czasów między kolejnymi wstrząsami geotektonicznymi o ustalonym progu w skali Richtera (Fig. 3.10). Wydaje się, że wdzięcznym polem zastosowań związanym też z dużą ilością zgromadzonych danych jest (byłaby) np. meteorologia czy fizjologia, gdzie teoretycznie uzyskane zależności można zweryfikować porównując z danymi zjawisk ekstremalnych. Oczywiście praktyczne zastosowania będą tu równie istotne co przy ocenie ryzyka w inżynierii finansowej, o której doktorant też wspomina w rozprawie i daje odpowiedni przykład (rysunek Fig. 3.7). W tym miejscu pragnę wskazać, że nie bardzo zrozumiałe jest jaki konkretnie próg zdarzenia ekstremalnego został przyjęty dla zrobienia dopasowania do danych empirycznych wzoru analitycznego (3.30) na wykresie Fig. 3.10? W podpisie pod rysunkiem widnieje wartość progu od 1,5 do 7,5 w skali Richtera, a z tekstu w rozprawie zdaje się wynikać, że dobór parametrów modelu, tj. $\alpha(Q)$ i $\tau(Q)$ nie zależy od progu Q . Czyżby to była jeszcze jakaś dodatkowa uniwersalność zjawisk sejsmicznych w zakresie lekkich i średnich wstrząsów nie mająca odpowiednika na rynku giełdowym? Myślę, że w trakcie obrony będzie można rzucić więcej światła na tę kwestię.

Nieco mniej zrozumiała w czytaniu (przynajmniej dla mnie) jest natomiast pierwsza część rozprawy. Niewątpliwie wymaga ona dużego skupienia od czytelnika i nie jest napisana w niektórych fragmentach tak przejrzysto jak reszta. To w niej (i w dołączonym Dodatku) wyprowadza się kluczowy wzór (3.12) na rozkład gęstości prawdopodobieństwa przerw czasowych między zdarzeniami progowymi. Przez rozprawę przewijają się w pierwszej części w paragrafach 3.2.1, 3.2.2 i 3.3 aż trzy rozkłady prawdopodobieństwa tj. Weibulla, q -Weibulla i q -wykładniczy. Ich pojawienie się nie jest zbyt jasno sformułowane. O ile rozkład Weibulla jest jednym z rozkładów (a ściślej rozkładów skumulowanych) proponowanych przez twierdzenie o wartościach ekstremalnych (ang. *maximum limit theorem*), to dwa pozostałe już nie wiążą się zupełnie z tym twierdzeniem. Ogólnie w pracy nie jest dość wyraźnie sprecyzowany związek między tymi rozkładami a główną linią logiczną rozprawy. Osobiście dopiero przy drugim czytaniu zrozumiałem, że Doktorantowi chodziło o znalezienie poprzez to twierdzenie ogólnego związku między średnim czasem zdarzeń progowych R_Q a parametrami rozkładów z twierdzenia Fishera-Tippetta-Gnedenki, by potem użyć tego związku do uzyskanego przez siebie niezależnego związku (3.12) na rozkład $\Delta_Q t$, i następnie wyrazić go poprzez parametry rozkładów Weibulla, q -Weibulla i q -exponent, i tym samym uzasadnić dodatkowo występowanie superskalowania, tj. dodatkowego skalowania parametrów rozkładu $\varphi_Q(\Delta_Q t)$ (3.12) czyli α_Q oraz τ_Q poprzez parametry rozkładów z tego właśnie twierdzenia. Sposób przedstawienia był tu do tego stopnia niejasny, że zgubiłem się w momencie, gdy w rozprawie pojawiły się inne (nie-Weibullowskie) rozkłady, których twierdzenie Fishera-Tippetta-Gnedenki nie przewidywało.

Ostatecznie wynik w postaci superskalowania jest nieco dyskusyjny. Gdyby był dokładnie spełniony, czy nie stanowiłby tylko akademickiej ciekawostki? Z rysunku Fig.(3.6) wynika, że w oparciu o 5 punktów na wykresie raczej trudno go jednoznacznie potwierdzić, nadto brakuje tu testu statystycznego który rozstrzygnąłby który z rozkładów (Weibull, q -Weibull, q -exp) najbardziej zbliża nas do hipotezy superskalowania i jak ją interpretować, gdyby została potwierdzona. Hipoteza ta nie jest jednak kluczowa dla dalszych rozważań w rozprawie i odnoszę wrażenie że została zaprezentowana w doktoracie niejako „przy okazji”.

Praca doktorska, choć napisana w postaci monografii zawiera wyniki już opublikowane przez Autora: M. Denys, et al., *Acta Physica Polonica A* 123, 513-514 (2013); *arXiv:1411-1689 v1*, 2014; *Acta Physica Polonica A* 129, 913-916 (2016); *PRE* 94(4), 042305 (2016).

O publikacji tych wyników dowiadujemy się czytając rozprawę i to dobry sposób informowania co i gdzie się już ukazało w czasopismach naukowych, odnosząc się do poszczególnych referencji w tekście. Mój niedosyt budzi natomiast mało precyzyjne wskazanie w poszczególnych rozdziałach rozprawy jaki faktyczny udział miał doktorant w uzyskaniu poszczególnych wyników. W szczególności dotyczy to wyprowadzenia wzoru (3.12), modyfikacji modelu agentów Sieczki i Hołysta, przeprowadzeniu symulacji i opracowania wyników. W opublikowanych pracach Doktorant figuruje jako pierwszy autor więc można się domyślać, że jego wkład jest kluczowy. Oczekiwałem jednak dokładniejszego wskazania szczegółów, gdyż prace są wieloautorskie a doktorat napisany w oparciu o ich wyniki. Będę wdzięczny za dokładniejsze omówienie tej kwestii przy obronie.

Pozostałe uwagi

Zauważyłem tylko kilka małych usterek edytorskich w rozprawie (wskazuję w kolejności występowania); winno być:

- Δt zamiast t na str. 44 w 4-ej linii paragrafu 3.5.1
- Q zamiast q na rys. 4.8 i w podpisie pod nim (dla utrzymania jednolitości notacji dla wartości progowej i w celu uniknięcia pomyłki z parametrem Tsallisa)
- "power-law tail with slope $(-3/2)$ " zamiast "power-law tail with slope $3/2$ " na str. 84 w 6-ej linii

Użyłbym również sformułowania „inverse proportionality” zamiast pompatycznego „universal Zipf law” na str. 44 w ostatniej linii.

Mam też uwagę natury ogólnej. Brak jest precyzyjnego wskazania jak dokładnie pozyskano dane. Oczywiście Doktorant pisze, że są to dane z poprzednich prac J. Ludeschera i innych. Czy oznacza to jednak że wykorzystał dane od tych autorów, którzy mu je przesłali? Czy też odtworzył je z sieci? A może wykorzystał gotowe rysunki w opublikowanych pracach i na tej podstawie raz jeszcze dopasował do danych z tych wykresów swoje modele (w tym ostatnim wypadku musiała być jakaś niedokładność samego odtworzenia rysunku, o której się nie mówi)?

Podsumowanie i wnioski końcowe

Praca jest ciekawie i w większości przystępnie napisana (może poza wspomnianymi fragmentami części pierwszej) z niewielką ilością usterek, czy niedomówień które nie mają zasadniczego wpływu na końcowe wyniki i główne konkluzje rozprawy. Znakomicie wpisuje się we współczesny nurt badań ekonofizyki. Rozprawa prezentuje szereg nowych rezultatów, z których najciekawsze to analityczny wzór wyprowadzony z teoretycznego modelu CTRW opisujący rozkład przerw czasowych między zdarzeniami progowymi lub ekstremalnymi w dowolnym układzie złożonym. Wzór został uzyskany jedynie przy upraszczającym założeniu, że procesy stochastyczne przyrostów sygnału i upływu czasu między tymi przyrostami są niezależne, czyli w oparciu o „rozdzieloną” wersję CTRW (ang. *decoupled CTRW*). Zawiera bardzo obszerny i wyczerpujący zestaw pozycji w bibliografii. Przedstawiony doktorat jest inspiracją do dalszych badań a z rozwinięcia prezentowanych wyników wskazanych na końcu w rozprawie (np. rozszerzanie zastosowań poza giełdę papierów wartościowych lub głębsza analiza zjawisk ekstremalnych na giełdzie na poziomie zleceń), czy też wspomniana

już analiza wpływu poszczególnych składowych modelu sprytnych agentów na prawidłowe odtwarzanie rozkładu czasów między zdarzeniami progowymi, mogą prowadzić do co najmniej 2-3 dalszych publikacji. Rozprawa postawiła kilka istotnych tez i dość szczegółowo je udowodniła stanowiąc tym samym oryginalne rozwiązanie problemu naukowego. Pewne zauważone potknięcia czy niedomówienia nie dominują nad poprawnością wyciąganych wniosków i uzasadnianych tez.

Dlatego w podsumowaniu moja ocena przedstawionej rozprawy jest bardzo pozytywna. Uważam, że w pełni spełnia ona wymagania ustawowe i zwyczajowe stawiane rozprawom doktorskim i wnoszę o dopuszczenie mgr. Mateusza Denysa do dalszych etapów przewodu doktorskiego.

Dariusz Grech