

Warszawa, 28 listopada 2022 r.

Dr hab. Katarzyna Bień-Barkowska, prof. SGH
Kolegium Analiz Ekonomicznych
Instytut Ekonometrii
Szkoła Główna Handlowa w Warszawie
ul. Madalińskiego 6/8
02-513 Warszawa

**Recenzja rozprawy doktorskiej mgr. Jakuba Michańkowa pt.
„Metody uczenia głębokiego w prognozowaniu
finansowych szeregów czasowych”**

Przedmiotem recenzji jest rozprawa doktorska mgr. Jakuba Michańkowa pod tytułem „Metody uczenia głębokiego w prognozowaniu finansowych szeregów czasowych”, napisana pod opieką naukową prof. Uniwersytetu Ekonomicznego w Krakowie dr. hab. Janusza Morajdy oraz pełniącego funkcję promotora pomocniczego pracy dr. Łukasza Kwiatkowskiego. Recenzję przedkładam na prośbę prof. dr. hab. Stanisława Popka, Dyrektora Szkoły Doktorskiej UEK (pismo z dn. 18 lipca 2022 r. na podstawie decyzji Rady Dyscypliny Ekonomii i Finansów Uniwersytetu Ekonomicznego w Krakowie z dn. 11 lipca 2022 r.).

Ocena celu rozprawy

Tematyka dysertacji dobrze wpisuje się w dyscyplinę naukową ekonomia i finanse. Z powodu zastosowania w niej zaawansowanych narzędzi ilościowych – w tym zarówno klasycznych modeli ekonometrycznych, jak i rozwinięć sieci neuronowych – przedłożona dysertacja ma charakter interdyscyplinarny i nowatorski. Wpisuje się w obszar zarówno ekonometrii finansowej, jak i nauki o finansach oraz metodach uczenia maszynowego. Celem rozprawy jest „opracowanie modeli predykcyjnych, opartych na metodach uczenia

głębokiego oraz wskazanie możliwości i potencjalnych kierunków ich wykorzystania, a także przeprowadzenie analizy ich efektywności, rozumianej jako trafność predykcji w prognozowaniu szeregów czasowych (logarytmicznych) stóp zwrotu aktywów finansowych oraz ich zmienności”. Ten ogólny cel rozprawy jest doprecyzowany za pomocą aż ośmiu celów szczegółowych oraz siedmiu hipotez badawczych. We Wstępie dysertacji Autor bardzo dobrze uzasadnia podjętą tematykę rozprawy, motywując ją zauważalną luką badawczą w zakresie możliwej synergii nowatorskich metod uczenia głębokiego i klasycznych metod prognozowania finansowych szeregów czasowych stóp zwrotu.

Uważam, że znaczenie naukowe rozprawy doktorskiej i wkład Autora w rozwój metod ekonometrii finansowej w Polsce jest bezsprzeczny i zauważalny. Badania ilościowe prezentowane w rozprawie „budują most” pomiędzy pozornie zupełnie odrębnymi podejściami do analizy danych finansowych – klasycznej ekonometrii bazującej na wnioskowaniu statystycznym oraz uczeniem maszynowym. Są ważnym uzupełnieniem dotychczasowych wyników badań empirycznych w zakresie prognozowania fluktuacji stóp zwrotu. Dysertacja ma wartość zarówno naukową jak i użyteczną, prezentuje bowiem możliwości odpowiedniego „skorygowania” prognoz generowanych przez klasyczne modele prognostyczne stóp zwrotu za pomocą metod sieci neuronowych uczenia głębokiego. W okresie rosnącej mocy obliczeniowej komputerów, coraz bardziej rozległych i różnorodnych źródeł informacji mających potencjalną wartość prognostyczną, a – z drugiej strony – w czasie rosnącej niepewności na rynkach finansowych, takie niestandardowe kombinacje różnych narzędzi ilościowych skutkują poprawą prognoz ryzyka i mają dużą wartość praktyczną dla zarządzających portfelami aktywów finansowych.

Charakterystyka i ocena zawartości rozprawy

Dysertacja liczy 232 strony, jej ogólna struktura jest klarowna i dobrze przemyślana. Autor wyodrębnił w rozprawie Wstęp, sześć rozdziałów o odrębnej zawartości merytorycznej oraz Zakończenie podsumowujące najważniejsze wnioski.

Pierwszy rozdział ma charakter ogólny i w bardzo syntetyczny sposób przedstawia *status quo* wiedzy w głównych nurtach ekonometrii finansowej i w zakresie sieci neuronowych, głównie w obszarze prognozowania finansowych szeregów czasowych stóp zwrotu. Ten rozdział jest bardzo krótki (niepełnie 9 stron), a zaprezentowany w nim stan badań w zakresie ekonometrii finansowej Autor zamieścił na zaledwie dwóch stronach, co w mojej opinii jest zbyt skrótowe i pomija istotne współczesne kierunki badań w tym

obszarze. Można było wspomnieć o modelowaniu stóp zwrotu w kontekście teorii wartości ekstremalnych, o analizie zdarzeń, o specyfice badań mikrostruktury rynku lub o różnorodnych parametryzacjach rozkładów wielowymiarowych za pomocą funkcji copula. Zakładam jednak, że intencją Autora w tej początkowej części rozprawy jest uwypuklenie kwestii dotyczących funkcjonowania rynków akcji oraz podejść do prognozowania warunkowej średniej i wariancji stóp zwrotu.

W drugim rozdziale Autor dokonuje syntetycznego porównania ekonometrycznych modeli klasy ARMA-GARCH, jako standardowego narzędzia opisu wahań stóp zwrotu w kontekście ich zmiennych w czasie warunkowych momentów: wartości oczekiwanej oraz wariancji. W celu umotywowania znaczenia modeli, ich prezentację słusznie poprzedził zestawieniem charakterystycznych teoretycznych i empirycznych własności stóp zwrotu. Nie mam generalnych zastrzeżeń do struktury i zawartości merytorycznej rozdziału. Modele zostały poprawnie scharakteryzowane, z uwzględnieniem wszystkich kluczowych zalet i ewentualnych wad. Autor rozpoczyna prezentację od standardowego modelu GARCH z rozkładem normalnym, przechodząc do różnorodnych uogólnień w zakresie rozkładów zakładanych w odniesieniu do składnika losowego oraz modyfikacji postaci funkcyjnej warunkowej wariancji stóp zwrotu (GJR-GARCH, EGARCH i APARCH). W drugiej części rozdziału Autor zajmuje się m.in. zagadnieniami prognozowania stóp zwrotu w oparciu o miary prognoz *ex post* oraz w kontekście oceny zyskowności strategii inwestycyjnej. Rozdział wieńczy prezentacja głównych miar ryzyka, czyli wartości zagrożonej (VaR) oraz oczekiwanego niedoboru (ES), będących przedmiotem prognozowania w empirycznej części pracy.

W kolejnym, trzecim rozdziale analizowane są sieci neuronowe uczenia głębokiego. Ponieważ funkcjonowanie sieci, zarówno w kontekście teoretycznym, jak i w praktyce, jest kluczowym elementem dysertacji doktorskiej, Autor poświęca tym zagadnieniom bardzo wiele uwagi, analizując różne rodzaje architektury sieci, aspekty ich uczenia się i optymalizacji ich działania poprzez zmiany wartości hiperparametrów. Rozdział dobrze tłumaczy te kwestie, jednak z ekonometrycznego punktu widzenia zabrakło mi, szczególnie na początku, próby powiązania prezentowanych metod uczenia maszynowego z metodami modelowania ekonometrycznego. Przedstawienie modelu prostej regresji jako pewnej sieci neuronowej pozwoliłoby na bardziej intuicyjne zrozumienie działania tej drugiej i przyczyniłoby się do zachowania większej ciągłości wywodu w kontekście wcześniejszych oznaczeń i wprowadzonej terminologii. Przykładem może być pierwsza część rozdziału,

gdzie Autor używa wielu nowych pojęć, takich jak „wejścia/wyjścia sieci”, „wagi połączeń”, „wektory wejściowe sieci”, „sygnały wejściowe” oznaczając je w sposób podobny do tego, zgodnie z którym wcześniej oznaczał zmienne losowe lub parametry modelu, a nie nadając im jakiegokolwiek próby interpretacji w kontekście terminologii wcześniej omawianych modeli ekonometrycznych. Nie twierdzą, że takie powiązanie dwóch koncepcji opisu procesów generujących dane byłoby proste, ale mogłoby przyczynić się do lepszego połączenia obu części pracy. Ważnym elementem rozdziału jest dokonanie przeglądu badań wykorzystujących sieci uczenia głębokiego (LSTM i CNN) w prognozowaniu zmiennych finansowych. Podoba mi się, że Autor zwraca uwagę, że w dotychczasowych publikacjach inni badacze pomijają możliwości sieci neuronowych, takie jak zróżnicowanie zmiennych wejściowych do sieci, na przykład w postaci prognoz generowanych przez różnorodne specyfikacje modeli zmienności.

Wywód prezentowany w rozdziale czwartym pracy dotyczy możliwości połączenia klasycznych modeli ekonometrycznych z metodami uczenia głębokiego w tzw. modele hybrydowe. Istotą tych konstrukcji jest potraktowanie samych stóp zwrotu i ich transformacji, jak również różnorodnych prognoz zmienności generowanych przez różne specyfikacje rodziny modeli GARCH, jako informacji wejściowych do sieci. W następstwie takiego działania sieci możliwe jest wyznaczenie prognoz szukanych parametrów rozkładów stóp zwrotu, a w konsekwencji wyznaczenie całych rozkładów warunkowych i ich kwantyli, czyli wartości zagrożonej. Autor słusznie podkreśla kompleksowość i „elastyczność” takiego podejścia w przypadku predykcji dynamiki stóp zwrotu. Podoba mi się, że proponowane przez siebie autorskie rozwinięcia modeli hybrydowych, na przykład w odniesieniu do skośnego rozkładu t-Studenta, osadza w wynikach dotychczasowych analiz (np. dla rozkładu normalnego).

Rozdział piąty jest najważniejszym elementem rozprawy, bowiem przedstawia oraz poddaje starannej interpretacji oryginalne, autorskie wyniki analiz na kanwie współdziałania parametrycznych modeli ekonometrycznych z metodami uczenia głębokiego. Przedmiotem badania są dzienne logarytmiczne stopy zwrotu z indeksów akcji trzech rynków rozwiniętych, tj. S&P 500, DAX i NIKKEI 225 oraz trzech rynków wschodzących: BOVESPA, WIG i KOSPI. Prezentacja wyników badania ma klarowną strukturę. Na wstępie Autor przedstawia elementy statystyki opisowej stóp zwrotu, akcentując charakterystyczne własności empirycznych rozkładów stóp zwrotu, a także własności dynamiczne (efekt skupiania zmienności). Moim zdaniem nieco niepotrzebnie analizowane

są współzależności pomiędzy różnymi stopami zwrotu, bowiem analiza wielowymiarowa (w formie wielowymiarowych modeli hybrydowych) nie jest kontynuowana w dalszej części rozprawy. Część empiryczna pracy odnosi się do trzech różnych obszarów: (1) prognozowania punktowego stóp zwrotu, (2) prognozowania zmienności i ryzyka oraz (3) prognozowania warunkowych rozkładów stóp zwrotu. W zależności od problemu badawczego Autor słusznie dobiera różne zbiory informacji wejściowych (wartości stóp zwrotu i ich odpowiednie transformacje) oraz architekturę sieci.

Szczegółowe i staranne porównanie wybranych najlepszych specyfikacji stóp zwrotu, głównie w kontekście ich własności prognostycznych, jest również kontynuowane w szóstym rozdziale pracy. Autor dokonuje w nim porównania wyników osiągniętych przez najlepsze specyfikacje modeli.

Uważam, że cel rozprawy został osiągnięty. Hipotezy postawione we Wstępie zostały poddane poprawnej i szczegółowej weryfikacji. Ogólny wydźwięk prezentowanych wniosków jest taki, że modele uczenia głębokiego w niektórych przypadkach poprawiają jakość modeli stóp zwrotu, jednak trudno jest o jednoznaczną i ogólną konkluzję odnośnie ich przewagi nad innymi metodami i bezkrytyczną rekomendację ich wykorzystania w każdym przypadku. Prezentowane analizy są dojrzałe, Autor sprawnie operuje narzędziami ilościowymi i czytelnie opisuje osiągnięte wnioski. Praca jest napisana dość ładnym językiem, a sposób prowadzenia wywodu naukowego dowodzi rozległej wiedzy Autora. Dotyczy ona zarówno zaawansowanych narzędzi uczenia maszynowego, jak i zagadnień klasycznej ekonometrii finansowej. W pracy dostrzegam umiejętność krytycznego myślenia – przedstawione wnioski badawcze są odpowiednio wyważone, z wyraźnym zaakcentowaniem zarówno zalet nowatorskich narzędzi, jak i związanych z nimi wątpliwości i ograniczeń. Doceniam również ewidentną pracowitość przeprowadzonych analiz, związaną z koniecznością samodzielnego oprogramowania lub dostosowania istniejących kodów w języku R lub Python. *Nota bene*, sugerowałabym rozważenie upublicznienia tych autorskich kodów celem rozpowszechnienia stosowanych metod, szczególnie, że „stworzenie autorskich kodów umożliwiających przeprowadzenie badań z wykorzystaniem sieci uczenia głębokiego” to jeden ze szczegółowych celów niniejszej rozprawy.

Pytania i komentarze

Poniżej przedstawiam pytania i uwagi, które w zasadniczej części mają charakter polemiczny i nie umniejszają mojej pozytywnej oceny dysertacji.

- (1) Ocena trafności prognoz punktowych zmienności na podstawie różnorodnych modeli parametrycznych GARCH oraz modeli hybrydowych dokonywana jest na podstawie wartości błędów średniokwadratowych *ex post* (MSE). Ponieważ „rzeczywista” zmienność nie jest obserwowalna, Autor porównuje wartości prognozowane z pewnym estymatorem zmienności, który bierze pod uwagę obserwowane w dniu prognozy ceny otwarcia i zamknięcia, a także ceny najwyższe i najniższe w tym dniu. Porównanie dużej liczby konkurencyjnych prognoz miałyby większą wartość informacyjną, gdyby Autor pokusił się o ustalenie dla nich swoistego rankingu popartego wnioskowaniem statystycznym. Ponieważ różnice w wielkości błędów prognoz nie wydają się znaczące, to mogłoby się okazać, że część modeli należy uznać za jednakowo dobre. Przykładem takich analiz jest zbudowanie tzw. zbioru ufności dla modeli (MCS, *model confidence set*), zgodnie z procedurą Hansena, Lunde i Nasona (2011). Jestem również ciekawa, czy zdaniem Autora, ranking porównujący modele mógłby mieć inną postać, gdyby do porównania zastosowano inny estymator zmienności (np. zmienność zrealizowaną obliczoną na podstawie śróddziennych stóp zwrotu)?
- (2) Testowanie wsteczne wartości zagrożonej przeprowadzone jest na podstawie testów Kupca i Christoffersena. O ile testy te są standardowym narzędziem weryfikacji, to test Christoffersena krytykowany jest ze względu na fakt, że ściśle bazuje na procesie Markowa pierwszego rzędu. Do bardziej wiarygodnego porównania jakości prognostycznej modeli zabrakło szerszej palety testów statystycznych badających czy przekroczenia VaR nie zależą od szerszego zbioru informacji (takiego jak bardziej odległe w czasie przekroczenia VaR, czy też warunki rynkowe aproksymowane przez wielkość prognoz VaR). Przykładowym narzędziem są w tym kontekście testy Dynamic Quantile (DQ) Engle’a i Manganeli (2004).
- (3) Zaletą modeli uczenia głębokiego jest możliwość wykorzystania dużych zbiorów zmiennych jako „wejść” sieci służącej prognozowaniu miar ryzyka. Ciekawa jestem zdania Autora na temat możliwości rozszerzenia tego zbioru o dodatkowe zmienne

(prócz oszacowań zmienności), które miałyby wartość prognostyczną dla wartości zagrożonej (np. zakresy cen, zmienność implikowana z cen opcji, zmienność zrealizowana)?

- (4) Mimo, że Autor korzysta z metod uczenia głębokiego do prognozowania miar zmienności i ryzyka, sama konstrukcja miar ryzyka jest ściśle związana założeniami dotyczącymi parametryzacji rozkładów stóp zwrotu. Tym samym, wartość zagrożona jest w rozprawie definiowana jako odpowiedni kwantyl rozkładu normalnego, rozkładu t-Studenta lub skośnego rozkładu t-Studenta (o odpowiednio sparametryzowanych momentach warunkowych). Nawet jeśli dynamika tychże momentów zostaje poprawnie scharakteryzowana za pomocą metod uczenia sieciowego, nieadekwatne prognozy VaR wynikać mogą z wyboru rodzaju rozkładu. Czy zatem możliwym i racjonalnym rozwiązaniem powyższego problemu mogłoby być uczenie sieci *stricte* pod kątem prognozy VaR (kwantyla warunkowego), gdzie funkcja straty miałaby postać analogiczną do problemu regresji kwantylowej?
- (5) W przypadku modeli punktowych prognoz stóp zwrotu zabrakło mi porównania jakości prognoz generowanych przez sieci neuronowe z prognozami z modeli klasy ARMA oraz z często stosowanym „benchmarkiem”, czyli procesem błędzenia przypadkowego dla logarytmów cen.

Jako drobną uwagę należy wskazać, że niektóre wzory nie zostały całkiem bezbłędnie opisane. Na przykład, we wzorach (2.21)-(2.23) komponent ARMA(p,q) ma taką samą strukturę opóźnień jak komponent GARCH(p,q). *Nota bene*, następnie struktura opóźnień modelu ARMA ma postać ARMA(k,m). W równaniu (2.23) brakuje indeksu dolnego t przy symbolu σ . We wzorze (2.52) zamiast definiowanego symbolu ε (dla składnika losowego), pojawia się niewyjaśniony symbol z . Niekiedy wartości zmiennych losowych przedstawiane są małą literą, niekiedy wielką. Przed ewentualną publikacją rozprawy warto byłoby to ujednolicić.

W pracy zdarzają się drobne niezręczności językowe natury znaczeniowej lub formalnej. Nie podoba mi się np. określenie „prognozowanie instrumentów finansowych”, ponieważ nie jest jasne jakie zmienne Autor ma w tym przypadku na myśli – czy prognozowanie cen, czy stóp zwrotu, czy warunkowej wariancji stóp zwrotu, czy może wielkości obrotu instrumentami finansowymi? Co istotne, od wyboru takiej zmiennej zależy także wybór metody badawczej. Dostyc często w rozprawie zdarzają się literówki, np.

„synergia (...) możliwa do osiągnięcia” (s. 2; powinno być „do osiągnięcia”), „McNeil i Fray” (s. 160, 163; powinno być „McNeil i Frey”), „dane z estymatora” (s. 110, lepiej brzmiałoby „uzyskane na podstawie estymatora”). W lekturze przeszkadzało mi również niekonsekwentne podejście do odmiany cytowanych nazwisk, np. „(...) do dalszego rozwoju dyscypliny przyczyniły się prace Mandelbrot [1963], Fama [1965a], Engle [1982] (...)” (s. 13) i „Rozwiązanie to, zaproponowane przez Schustera i Paliwała [1997] (...)” (s. 69).

Konkluzja

Podsumowując, stwierdzam, że rozprawa doktorska mgr. Jakuba Michańkova „Metody uczenia głębokiego w prognozowaniu finansowych szeregów czasowych” w pełni spełnia kryteria stawiane rozprawom doktorskim określone w art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz. U. z 2022 r. poz. 574). Doktorant podjął się badania oryginalnego i ważnego problemu naukowego, który rozwiązał w sposób rzetelny. W rozprawie doktorskiej wykazał się umiejętnością prowadzenia samodzielnej i pracochłonnej pracy naukowej, wymagającej opanowania nowatorskich narzędzi ilościowych. Niniejszym wnoszę o przyjęcie rozprawy do publicznej obrony.

K. Bień