

Prof. dr hab. Krzysztof Jajuga
Uniwersytet Ekonomiczny we Wrocławiu

Recenzja pracy doktorskiej Magistra Przemysława Jaśko pt.
„Zastosowanie metod data mining w strategii arbitrażu statystycznego na rynku kapitałowym”

Przedstawiana recenzja jest przygotowana z uwzględnieniem przepisów Ustawy z dnia 14 marca 2003 o stopniach naukowych i tytule naukowym oraz stopniach i tytule w zakresie sztuki.

Recenzowana rozprawa doktorska dotyczy tematyki powiązania szeroko rozumianych metod matematycznych i zagadnień rynku kapitałowego (stanowiącego część rynku finansowego). Takie powiązanie nie budzi wątpliwości. Główny argument wynika z faktu, iż dynamiczny rozwój istotnej części teorii i praktyki rynku finansowego stał się możliwy dzięki włączeniu (a potem rozwinięciu) wielu metod ilościowych (np. ekonometrycznych i statystycznych) do zagadnień rynku finansowego. Wystarczy tu wspomnieć czołowych profesorów, zwłaszcza laureatów nagrody pamięci Nobla z nauk ekonomicznych (potocznie zwanej Nagrodą Nobla z ekonomii). Są to następujący profesorowie (w nawiasie lata uzyskania nagrody): Harry Markowitz (1990), William Sharpe (1990), Merton Miller (1990), Franco Modigliani (1985), James Tobin (1981), Robert Engle (2003), Robert Merton (1997), Myron Scholes (1997), Eugene Fama (2017), Robert Shiller (2017), oprócz tego Fischer Black.

Wybór tematyki rozprawy nie budzi wątpliwości. Jej tekst każe ją rozpatrywać w kategoriach interdyscyplinarnych, jako połączenie dyscypliny ekonomia i finanse (dziedzina nauki społeczne) i dyscypliny matematyka (dziedzina nauki ścisłe i przyrodnicze). Formalnie postępowanie jest w dyscyplinie ekonomia i finanse. Taką dyscyplinę również wskazałbym jako dominującą w kontekście treści rozprawy doktorskiej. Są, moim zdaniem, dwa podstawowe argumenty w tym zakresie. Po pierwsze, najważniejsze osiągnięcia pracy są w zakresie wykorzystania metod ekonometrycznych i statystycznych w analizie zjawisk finansowych. Po drugie, empiryczna strona pracy dotyczy danych z polskiego rynku kapitałowego.

Rozprawa doktorska Przemysława Jaśko liczy rekordową liczbę stron (755 !), jest podzielona na 8 rozdziałów, z tego rozdział 1 to wprowadzenie, a rozdział 8 to zakończenie, ponadto

zawiera dodatki, zawierające m.in. cząstkowe wyniki. Jest też spis literatury, 413 pozycji, oceniam to pozytywnie.

Omawiam teraz strukturę pracy, dokonując przy tej okazji ogólnej oceny jej poszczególnych rozdziałów.

Rozdział 1 jest pomyślany jako wprowadzenie, jednak jego treść jest szersza, gdyż zawiera również opis pewnych metod, które zwyczajowo zamieszcza się w dalszej części rozprawy doktorskiej.

Podrozdział 1.2 zawiera przegląd metod ilościowych (wbrew temu co pisze Autor, nie są to jedynie metody data mining, lecz znacznie szerszy zbiór), które mają zastosowanie w analizie szeregów czasowych). Na pewno plusem tych rozważań jest to, że Autor wykazuje się ogólną znajomością wielu metod. Jednak ten opis jest właściwie „jak leci”, przedstawiona klasyfikacja nie jest rozłączna i brak jest krytycznego porównania tych metod. Wydaje się też, że właściwie ten opis nie jest specjalnie potrzebny z punktu widzenia głównych rozważań pracy.

Podrozdział 1.3 dotyczy arbitrażu statystycznego. Uważam, że ten fragment stanowi pewien wkład Autora, poprzez dokładne formalne sprecyzowanie co przez ten termin rozumie, a następnie na tym gruncie przedstawienie formalnego opisu procesu transakcyjnego. Podrozdział 1.4 zawiera motywację, cel pracy i pytania badawcze (jest 7 pytań teoretycznych i 5 pytań empirycznych)

Podrozdział 1.5 jest dość długi, zawiera opis struktury pracy. Jednak z punktu widzenia dyscypliny ekonomia i finanse opis ten jest bardzo hermetyczny, zbyt sformalizowany, bez odniesień i interpretacji finansowych. Podrozdział 1.6 skrótowo przedstawia rodzaje danych empirycznych wykorzystanych w pracy. Oprócz tego zawiera niedługi fragment, m.in. wzory (1.6.1) i (1.6.2) i powiązania między nimi, są to bardzo elementarne kwestie ze studiów licencjackich na kierunku finanse, nie było potrzeby umieszczania tego w pracy doktorskiej.

Rozdział 2 dotyczy zagadnienia arbitrażu statystycznego w kontekście formalnego ujęcia procesów losowych.

W podrozdziałach 2.1 i 2.2 znalazła się pewna część bardzo elementarnych rozważań z probabilistyki, procesów stochastycznych, które są powszechnie znane z dydaktyki, takich jak: zmienna losowa, rozkład, dystrybuanta, wektor losowy, wartość oczekiwana, rozmiar próby. Właściwie bez szkody dla rozprawy można było usunąć te fragmenty.

Podrozdział 2.3 omawia podstawowe kwestie z zakresu finansów. Te pojęcia są w ekonomii i finansach doskonale znane, Doktorant niepotrzebnie komplikuje to przez formalizację matematyczną, nawet wprowadzając ujęcie z topologii. Również nie było potrzeby przypominania podstawowych kwestii z obszaru testowania hipotez statystycznych (str. 65).

Podrozdział 2.4 podaje autorskie określenia arbitrażu klasycznego i arbitrażu statystycznego, co można uznać za wkład. Szkoda, że oprócz czysto formalnych określeń nie znalazły się tam finansowe interpretacje oraz (nawet generyczne) przykłady.

W podrozdziale 2.5 przedstawiony jest na podstawie literatury test wykorzystywany w dalszej części pracy. Podrozdział 2.6 zawiera formalne ujęcie konstruowania arbitrażu statystycznego dla pewnej klasy procesów. Jest to ważny fragment, chociaż jego znaczenie można było wzmocnić poprzez podanie interpretacji i odniesień do dyscypliny ekonomia i finanse.

Rozdział 3 jest bardzo obszerny, liczy około 140 stron. Zawiera rozważania z obszaru, jak to określa Autor, losowych układów dynamicznych. Moim zdaniem, ten rozdział mógł się nie pojawić w pracy przedstawianej w dyscyplinie ekonomia i finanse. Na początku (str. 81-93) są kwestie znane z ekonometrii dynamicznej i ekonometrii. Jednak zasadnicza część rozdziału są to zagadnienia czysto matematyczne, często z obszaru topologii. Autor nie podaje kontekstu (czy też jakiegokolwiek interpretacji) na gruncie finansów. Sądzę, że właściwym rozwiązaniem mogłoby być opublikowanie rozważań tego rozdziału w innej publikacji (publikacjach), niekoniecznie z dyscypliny ekonomia i finanse.

Kolejne rozdziały mają inny charakter i to właśnie ta część rozprawy jest istotna dla przedstawianej rozprawy doktorskiej.

Rozdziały 4 i 5 przedstawiają modele ekonometryczne i statystyczne, które mogą stanowić podstawę dla arbitrażu statystycznego, a ściślej jednego rodzaju, mianowicie arbitrażu pair trading. Autor słusznie koncentruje się na zagadnieniu kointegracji. Przy tym rozdział 4 dotyczy modeli z parametrami stałymi w czasie, a rozdział 5 dotyczy modeli z parametrami zmiennymi w czasie. Autor słusznie koncentruje się na zagadnieniach wnioskowania statystycznego, co jest potrzebne w empirycznej części pracy. Doktorant wykazuje głęboką znajomość w zakresie tych modeli oraz bardzo dużą wnikliwość i rzetelność w swoich dociekaniach naukowych. Uważam te dwa rozdziały za wkład Doktoranta.

Rozdział 6 ma charakter empiryczny. Badaniu podlega 21 szeregów czasowych (ceny akcji 20 spółek indeksu WIG20 oraz wartości indeksu WIG20) w okresie niecałych czterech lat (w sumie 643 obserwacje). Autor stosuje podejście dwuetapowe. Najpierw analizuje charakterystyki szeregów czasowych za pomocą metod eksploracyjnych, w tym

podstawowych statystyk opisowych. Jest to bardzo kompleksowa analiza. Autor bada następujące właściwości:

- deterministyczny czy stochastyczny (tu akurat jest potwierdzenie stochastyczności, o czym wiadomo od dawna w odniesieniu do finansowych szeregów czasowych);
- stacjonarność czy niestacjonarność;
- powrót do średniej;
- autokorelacja.

Kolejny bardzo istotny fragment tego rozdziału to przeprowadzenie testów kointegracji liniowej i nieliniowej w odniesieniu do różnych kombinacji spółek (począwszy od 2-elementowych po 21-elementową). Stanowi to podstawę do zweryfikowania, które kombinacje spółek mogłyby być potencjalnie zastosowane w strategiach arbitrażowych typu pair trading. Rozważania są bardzo wnikliwe i pogłębione i świadczą o dużej umiejętności Autora w zakresie prowadzenia badań ekonometrycznych.

Wyniki testów kointegracji liniowej znajdują zastosowanie w modelu CCEM (podrozdział 6.3), zaś wyniki testów kointegracji nieliniowej (Breitunga) znajdują zastosowanie w klasie modeli bayesowskich TVP-VECM-SV (punkty od 6.5.2 do 6.5.7). Bardzo wysoko oceniam te rozważania. Ważne są wnioski, mogące mieć znaczenie empiryczne (choć dotyczą jedynie kointegracji) w tabeli 6.1.30 w punkcie 6.5.8.

Z kolei podrozdział 6.7 odbiega charakterem od całości tego rozdziału i raczej stanowi implementację programów komputerowych (10 rodzajów kodów). Myślę, że bez szkody dla całości pracy można było pominąć ten podrozdział.

Rozdział 7 jest podsumowaniem badań empirycznych rozdziału szóstego. Jest to krótki rozdział, ale istotny, gdyż wskazuje na istnienie kointegracji (bądź nie), co może stanowić przesłankę do przeprowadzenia arbitrażu. Szkoda, że Autor nie wykorzystał tych wyników do pokazania możliwych rzeczywistych transakcji arbitrażowych, co znacznie uwiarygodniłoby otrzymane wyniki.

Rozdział 8 to zakończenie, w którym Autor wskazuje na realizację celu i pytań badawczych. Oceniam to pozytywnie, z tym, że z punktu widzenia dyscypliny ekonomia i finanse istotne znaczenie mogą mieć pytania badawcze PE 1 – PE 4. Inne pytania sam Autor zakwalifikował jako matematyczne, a więc przynależące do dyscypliny matematyka.

Moja ocena rozprawy doktorskiej uwzględnia formalne wymagania ustawowe, jakie stawiane są rozprawom doktorskim, mianowicie:

- rozprawa doktorska powinna stanowić oryginalne rozwiązanie problemu naukowego;

- rozprawa doktorska powinna wykazywać ogólną wiedzę teoretyczną kandydata w danej dyscyplinie naukowej;
- rozprawa doktorska powinna wykazywać umiejętność samodzielnego prowadzenia pracy naukowej.

Nie ulega mojej najmniejszej wątpliwości, że w przypadku rozprawy doktorskiej Pana Magistra Przemysława Jaśko wszystkie powyższe trzy warunki są spełnione.

Po pierwsze, odpowiedź na pytanie o oryginalne rozwiązanie zagadnienia naukowego w rozprawie doktorskiej jest jednoznacznie pozytywna. Co prawda, nigdzie w pracy, Doktorant nie określa wprost co jest (Jego zdaniem) tym zagadnieniem naukowym. Niemniej jednak, lektura rozprawy pozwala za określenie tego zagadnienia naukowego. Moim zdaniem, jest to: „sformułowanie formalnych podstaw dla określenia warunków przeprowadzenia jednej klasy transakcji arbitrażowych, mianowicie tych mających u podstaw tzw. pair trading”.

Po drugie, Doktorant wykazuje dobrą wiedzę teoretyczną w dyscyplinie ekonomia i finanse. Dotyczy to zwłaszcza dwóch ważnych obszarów. Pierwszy z nich to narzędzia ekonometryczne i statystyczne właściwe w analizie rynku kapitałowego (chodzi zwłaszcza o analizę cen i stóp zwrotu inwestycji w akcje spółek). Drugi obszar to formalne podstawy transakcji arbitrażowych, stanowiących jedną z trzech podstawowych grup transakcji dokonywanych na rynku kapitałowym.

Po trzecie, Doktorant wykazuje umiejętność samodzielnego prowadzenia badań naukowych. Z jednej strony, jest to dogłębna analiza literatury przedmiotu. Po drugie, a może przede wszystkim, jest to bardzo precyzyjne, profesjonalne i kompleksowe zastosowanie metod ekonometrycznych i statystycznych, prowadzące do odpowiedzi na pytania badawcze, zwłaszcza te, które są istotne dla dyscypliny ekonomia i finanse.

Wkład Autora, który uznaję za osiągnięcie w dyscyplinie ekonomia i finanse, i który wzmacnia moją pozytywną ocenę pracy, jest następujący:

- przeprowadzenie bardzo głębokich badań empirycznych w odniesieniu do możliwości arbitrażu wynikającego z badań kointegracji;
- wyróżnienie i zbadanie bardzo dużej liczby charakterystyk szeregów czasowych (zwanym przez Autora „metacechami”);
- podanie formalnego ujęcia zagadnienia arbitrażu statystycznego.

Lektura recenzowanej pracy, którą to pracę oceniam bardzo pozytywnie, prowadzi (jak w przypadku każdej pracy naukowej) do sformułowania uwag, niektóre z nich mogą stanowić podstawę do dyskusji.

Uwaga o charakterze strukturalnym.

Moja główna uwaga wynika już z ogólnej oceny poszczególnych rozdziałów przedstawionej powyżej. Dotyczy ona bardzo dużej objętości pracy. Wskazałem już, które części (zwłaszcza w rozdziale 2 i 3) mogły być usunięte bez szkody (a nawet z pożytkiem) dla pracy rozpatrywanej z punktu widzenia dyscypliny ekonomia i finanse. Oczywiście te rozważania to efekt dużej pracy Doktoranta, którą bardzo doceniam. Szkoda, że nie znalazły one miejsca w innej publikacji, w dyscyplinie matematyka.

Uwagi merytoryczne.

1. Istotnym niedostatkim pracy jest brak interpretacji rozważań z punktu widzenia finansów (zwłaszcza rynków finansowych). Matematyczne ujęcie jest kluczowe w rozważaniach z finansów. Należy jednak pamiętać, że nawet (pozornie) bardzo zaawansowane metody (choćby model Blacka-Scholesa-Mertona) są interpretowane bardzo klarownie na gruncie dyscypliny ekonomia i finanse. Brak interpretacji i rozumienia przez użytkownika oznacza, że model jest nieprzydatny dla końcowego użytkownika.
2. Uważam, że w pracy zabrakło weryfikacji skuteczności potencjalnych strategii arbitrażowych na podstawie danych z polskiego rynku, które to strategie spełniałyby warunki wstępne określone przez Autora. Autor sam stwierdza na str. 662, że odczuwa niedosyt z tego powodu.
3. Pewne wątpliwości może budzić tytuł rozprawy z dwóch powodów. Po pierwsze, w pracy są nie tylko metody data mining (Autor stosuje nie tylko metody eksploracyjne lecz również confirmacyjne). Nastąpiło zawężenie tytułu. Po drugie, Autor rozpatruje tylko jeden rodzaj arbitrażu statystycznego i tylko dla bardzo szczególnego przypadku (spółki WIG20) z wąskiego okresu. Tutaj nastąpiło rozszerzenie tytułu, sugerujące na wyrost ogólność pewnej części rozważań.
4. Autor w pracy rozpatruje sporo modeli. Szkoda, że w ogóle nie zajmuje się ryzykiem tych modeli, np. kwestią braku odporności modeli na zmiany parametrów. Jest to jedno z ważnych zagadnień rozpatrywanych w finansach. Modele nieodporne nie mają wielkiej wartości na rynkach finansowych.

5. Nie do końca uzasadniony został okres, z którego pochodzą dane, kończący się na marcu 2020, a więc prawie cztery lata temu, argument pandemii trudno uznać, gdyż efekty pandemii wystąpiły przez ledwie kilka tygodni w 2020.
6. W podrozdziale 6.1.4 Autor przeprowadza skalowanie wielowymiarowe, w ramach tego wyznacza odległości euklidesowe. Ta operacja przeprowadzana jest w odniesieniu do 184 zmiennych – charakterystyk (nazywanych przez Autora „metacechami”). Jednak pewna część tych zmiennych są to zmienne binarne, a więc przedstawiane na skali nominalnej. Wiadomo, że dla zmiennych przedstawianych na skali nominalnej odległość euklidesowa nie jest właściwą miarą odległości (podobieństwa, niepodobieństwa). W sytuacji w pracy zmienne są przedstawiane na różnych skalach, należało sięgnąć np. po GDM.
7. Pewne drobne kwestie:
 - Autor stosuje całą Stratonowicza zamiast powszechnie przyjmowanej w finansach całki Ito, porównanie na str. 634 nie przekonuje dlaczego tak uczynił;
 - definicja 2.1.18 – definicja jest dla zmiennej dyskretnej, a nie ciągłej (w tytule definicji);
 - w literaturze finansowej stosuje się zwrot „wolny od ryzyka”, a nie „pozbawiony ryzyka”;
 - str. 25 – nie jest prawdą, że w literaturze są tylko sądy o charakterze spekulatywnym.

Uwagi konstrukcyjne:

- Doktorant zastosował w pracy niezręczny (i niestandardowy) sposób powoływania źródeł literatury, z reguły poprzez podawanie tylko numeru, zamiast np. nazwiska autora i roku wydania – bardzo to utrudnia lekturę pracy i weryfikację zasadności odniesienia;
- W głównym tekście pracy znalazło się 15 fragmentów, zawierających kod na potrzeby działania programów (R oraz pakiet REDUCE), te fragmenty powinny być w załączniku, gdyż mają charakter czysto techniczny, nie są istotne z punktu widzenia dyscypliny ekonomia i finanse;
- str. 361 – Autor odsyła do załącznika, jeśli chodzi o charakterystyki szeregów czasowych (Autor używa tu nazwy „metacechy”), a to z kolei jest na tyle istotne, że mogło znaleźć się w głównym tekście rozprawy.

Uwagi językowe:

- Autor nadużywa zwrotu „kanoniczna praca” w odniesieniu do mniej znanych pozycji literatury;
- W pracy jest pewna (niewielka) liczba błędów gramatycznych i literowych;

- błędy ortograficzne: str. 22 – jest „zmianą”, ma być „zmianom”; str. 70, 228, 336 – jest „na prawdę”, ma być „naprawdę”;
- w kilku miejscach błąd odmiany, str. 17 – nie: „możliwym jest”, prawidłowo: „możliwe jest”; str. 20 – nie: „zasadnym jest”, prawidłowo: „zasadne jest”, str. 24 – nie: „niezbędnym jest”, prawidłowo: „niezbędne jest”;
- dziwne zwroty: „przestrzeń instancji” (str. 11 i wiele innych), „foliacja”, „postparsyfikacja”, „obsunięcie kapitału”, „derywacja” (str. 118), „zaniedbywalny”;
- nieprawidłowe zwroty: „ruchy cen”, „ruch wartości”, ma być „zmiany cen”, „zmiana wartości”;
- słowo „aktywa” ma w języku polskim tylko liczbę mnogą (błędy na str. 14, 15, 16, 18, 19, 59, 362).

Przedstawione uwagi konstrukcyjne i formalne nie zmieniają mojej pozytywnej oceny rozprawy.

Ostateczna konkluzja

W konkluzji stwierdzam, że rozprawa doktorska Magistra Przemysława Jaśko spełnia wymogi ustawowe stawiane rozprawom doktorskim w dziedzinie nauk społecznych dyscyplina ekonomia i finanse. Wnioskuje przeto do Rady Dyscypliny o dopuszczenie tej rozprawy do obrony.



Wrocław, 2024-04-05.