

Łódź, 18.08.2024 r.

Dr hab. Aleksandra Baszczyńska
Uniwersytet Łódzki
Wydział Ekonomiczno-Socjologiczny
Katedra Metod Statystycznych

**Recenzja poprawionej rozprawy doktorskiej mgr. Pawła Zdzisława Lańducha
pt. „Wykorzystanie technik imputacyjnych w szacowaniu informacji
wynikowych oraz w analizie struktury danych w statystyce przedsiębiorstw”**

przygotowanej w Szkole Głównej Handlowej w Warszawie
pod kierunkiem naukowym dra hab. Andrzeja Młodaka

1. Podstawa recenzji:

- Pismo Rady Naukowej dyscypliny ekonomia i finanse z dnia 03.07.2024 r. informujące o ponownej ocenie poprawionej rozprawy doktorskiej mgr. Pawła Zdzisława Lańducha nt.: „Wykorzystanie technik imputacyjnych w szacowaniu informacji wynikowych oraz w analizie struktury danych w statystyce przedsiębiorstw”.
- Dokument z dnia 2.06.2023 r. informujący, że Rada Naukowa dyscypliny ekonomia i finanse uchwałą nr 721 z dnia 14.06.2023 r. powołała mnie na recenzenta rozprawy doktorskiej mgr. Pawła Zdzisława Lańducha nt.: „Wykorzystanie technik imputacyjnych w szacowaniu informacji wynikowych oraz w analizie struktury danych w statystyce przedsiębiorstw” w dziedzinie nauk ekonomicznych w dyscyplinie ekonomia,
- Ustawa z dnia 14.03.2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. z 2017 r., poz. 1789 z późn. zm.).



2. Ocena wyboru problemu badawczego

W rozprawie doktorskiej zawarte są wyniki podjętej próby rozwiązania problemu badawczego dotyczącego zastosowania technik imputacyjnych w analizach z zakresu statystyki przedsiębiorstw. Temat rozprawy jest niezwykle ważny ze względu na znaczenie w praktyce aplikacyjnej szeroko pojętej statystyki publicznej, gdzie braki danych są problemem występującym stosunkowo często. Jednocześnie zagadnienie to jest interesujące z naukowego punktu widzenia. Badanie własności poszczególnych procedur prowadzi do określenia techniki najlepszej z określonego punktu widzenia. Problem badawczy zaprezentowany w rozprawie doktorskiej oznacza więc nie tylko wybór odpowiednich technik imputacyjnych w określonej sytuacji badawczej z zakresu statystyki publicznej ale również obejmuje ocenę ich własności, w tym wybór i analizę otrzymanych wyników badania metodami symulacyjnymi własności technik imputacyjnych.

3. Ocena celów badawczych i hipotez rozprawy doktorskiej

Głównym celem rozprawy jest: „ocena efektywności i jakości (tzn. oszacowanie ich dokładności i obciążenia) obciążonych estymatorów imputacji masowej bazującej na integracji próby losowej i nielosowej, w badaniu symulacyjnym”. W poprawionej rozprawie doktorskiej cel badawczy związany jest zatem z zaplanowaniem i przeprowadzeniem analiz z wykorzystaniem metod symulacyjnych. Cel rozprawy sformułowany w sposób syntetyczny przez Doktoranta niestety nie wskazuje, że w analizach będą wykorzystywane dane ze statystyki przedsiębiorstw GUSu.

Jednakże zaprezentowane dalej w rozprawie hipotezy (główna i dodatkowe) już ten element związany z wykorzystaniem w analizach wyników badań statystyki publicznej jest uwzględniony, co w pełni podkreśla istotny aspekt badań Doktoranta. I jest w zgodzie z tytułem rozprawy doktorskiej.

4. Ocena metod badawczych

W procesie przygotowania rozprawy doktorskiej wykorzystane są przede wszystkim dwie podstawowe metody badawcze: analiza literatury przedmiotu oraz metody statystyczne, w szczególności metody symulacyjne. Analiza literatury przedmiotu dokonana jest właściwie, uwzględniono najważniejsze pozycje literaturowe w języku polskim oraz w języku angielskim. Są to i pozycje książkowe i artykuły w czasopiśmie, jak również dokumenty międzynarodowych instytucji statystycznych. Doktorant umiejętnie dobierał literaturę i prezentował najważniejsze osiągnięcia zawarte w omawianej książce czy artykule. Jest to o tyle ważne, że część empiryczna pracy odwołuje się do wyników badań już przedstawionych w literaturze przedmiotu.

Metody symulacyjne, których wyniki zastosowania są zaprezentowane w rozprawie doktorskiej są adekwatne do celu badań prowadzonych przez Doktoranta. Techniki symulacyjne są zastosowane we właściwy sposób. Jak również właściwie przeprowadzona jest interpelacja wyników tych badań. Część empiryczna pracy zawiera dwa ważne dokonania Doktoranta – prezentuje właściwy dobór metody badawczej w analizach statystyki przedsiębiorstw, jak również prezentuje wyniki analiz symulacyjnych. Doktorant swobodnie porusza się i w zakresie metod statystyki ekonomicznej i statystyki matematycznej.

5. Ocena układu pracy i zawartości poszczególnych części

Praca liczy 161 stron. Zawiera stronę tytułową, spis treści, streszczenie w języku polskim, streszczenie w języku angielskim, wprowadzenie, pięć rozdziałów, zakończenie, spis literatury, wykaz tablic, wykaz rysunków i wykresów oraz aneks.

Układ pracy jest czytelny i nie budzi większych zastrzeżeń. Zawartości poszczególnych części pracy są zgodne z tytułami tych części i cechują się wysokim poziomem merytorycznym.

Uwagi i wątpliwości związane z układem pracy, zawartością poszczególnych części i uwagi formalne i redakcyjne są podane poniżej.

Uwagi ogólne do całej pracy:

- Brak jest ujednolicenia w strukturze pracy – niestety tylko dwa pierwsze rozdziały posiadają podsumowanie. Nie każdy rozdział ma pierwszy podrozdział określony jako wstęp.
- W pracy stosowane są określenia „cecha” i „zmienna”. Brak jest informacji czy stosowanie tych określeń zamiennie jest celowe czy Doktorant rozróżnia te pojęcia.
- Brak jest ujednolicenia powołań na literaturę (zapis współautorstwa, zapis stron). Ponadto istnieją powołania na literaturę w pracy a brak ich jest w spisie literatury.
- Podsumowania w rozdziale 1 i 2 zawierają raczej informacje co do dalszych kierunków badań. Bardzo cenne są to uwagi ale nie stanowią podsumowania informacji zawartych w rozdziale.

Wprowadzenie zawiera uwagi wstępne związane ze statystyką przedsiębiorstw i procesem imputacji danych. Zostały sformułowane zarówno cel pracy, jak i hipotezy badawcze. Jest to właściwe wprowadzenie do tematyki rozprawy doktorskiej. Uwagi są następujące:

- Tablica 1 nie jest dokładnie taka sama jak w pracy Dehnel i Kowalewski (2015).
- Nie zawsze zachowany jest ten sam czas gramatyczny w opisie badania.
- Braki danych są policzalne stąd określenie „ilość braków danych” nie jest właściwe.

Rozdział 1 dotyczy badań statystycznych przedsiębiorstw. Omówiony został model procesowy funkcjonujący w statystyce publicznej. Podane są również cechy badań przedsiębiorstw. Struktura tej części pracy jest prawidłowa, chociaż naturalnym wydaje się połączenie podrozdziałów 1.4 i 1.5 (uwzględniając chociażby długość podrozdziału 1.5).

Uwagi szczegółowe są następujące:

- Nie zawsze poprawna nazwa Generic Statistical Business Process Model (GSBPM) jest zastosowana.
- Struktura modelu GSBPM jest następująca: 4 poziomy (nie 3), osiem (nie 9) faz procesu statystycznego.
- Błędny zapis źródeł dla rysunku 1 i rysunku 2 (brak jest określenia „opracowanie własne na podstawie....”).
- Graficzna prezentacja modelu odpowiedzi w badaniach przedsiębiorstw nie jest w języku polskim.
- Dyskusyjne jest przypisanie „Opracowania pojęć statystycznych” do fazy 2.
- Niefortunne określenie wskazujące, że próba Poissona jest metodą.
- Błędnie podany rozkład przedsiębiorstw w Polsce w tablicy 2 (podane wartości nie sumują się do 100%).

Rozdział 2 zawiera informacje dotyczące integracji danych. Przedstawione w sposób syntetyczny są odpowiednie definicje, jak i rodzaje integracji danych. Zauważalny jest jednak brak usystematyzowania i dodatkowych wyjaśnień co do określeń: integracja danych na poziomie mikro, integracja danych na poziomie jednostkowym i mikrointegracja. Tytuł rozdziału powinien być sformułowany w sposób szerszy. Obecnie tytuł rozdziału 2 i podrozdziału 2.4 wskazuje na te same zagadnienia. Drobne uwagi są następujące:

- Nie ma potrzeby używać określeń „po pierwsze” i „po drugie” gdy zastosowana jest punktacja.
- Błędnie podany numer tablicy (nie 1 a 3).
- Brak ujednolicenia oznaczeń (kolor szary) w tablicy 3 i 4.
- Niejasne określenie celów parowania statystycznego. Celem parowania nie jest mikro.
- Błąd w zapisie odległości Hamminga.
- Brak ujednolicenia zapisu d_H i d_h .

Rozdział 3 zawiera informacje dotyczące imputacji masowej ze szczególnym uwzględnieniem metod wykorzystywanych w badaniu. Opis metod jest wyczerpujący. Drobne uwagi są następujące:

- Dyskusyjne wprowadzenie oznaczenia „ y ” oraz „ x ” dla zmiennej, jak również wykorzystanie tego samego oznaczenia dla wektora.
- Brak uporządkowania oznaczeń. W tym rozdziale wprowadzony jest symbol „ d ” jako indeks domeny oraz w tym samym rozdziale wagi. Podczas gdy w poprzednim rozdziale symbol ten oznaczał odległość.
- Brak opisu oznaczeń wprowadzonym w tekście (na przykład symbole we wzorze 9).
- Brak ujednolicenia zapisu big data.
- Brak definicji nadpopulacji.
- Brak jednoznaczności w oznaczeniu liczebności populacji i próby.
- Brak opisu symbolu h w równaniu 15.
- Symbol m oznacza co innego niż w poprzednim rozdziale.
- Brak opisu dla symbolu w równaniu 20.

Rozdział 4 zawiera szczegółowo opisane założenia analizy empirycznej. Powołując się na cel badania i hipotezy badawcze omówiono materiał empiryczny wykorzystany w zaprojektowanym badaniu statystycznym. Niestety materiał empiryczny nie został w pełni zaprezentowany – brak jest bowiem informacji w wierszu pierwszym tablicy 7.

Uwagi szczegółowe są następujące:

- Brak jest wyjaśnienia oznaczenia użytego w tablicy 7.
- Nie wiadomo czy materiał empiryczny pochodzi z maja czy lipca 2022 roku.
- Opis pominiętych działów nie jest ujednolicony z informacjami w tablicy 7.
- Brak ujednolicenia określenia jednostek dużych i średnich.
- Dyskusyjne jest sformułowanie dotyczące ciągłości zmiennej „Liczba pracujących”.
- Brak jest informacji dlaczego $h = 1, \dots, 25$, podczas gdy działów jest 26.
- Powołanie na ppkt 1.4 nie jest prawidłowe.
- Brak ujednolicenia zapisu akronimu mechanizmu missing not at random.
- Brak ujednolicenia zapisu akronimu relative mean squared error.
- Wprowadzenie symbolu B dla liczby powtórzeń nie jest właściwe. B oznacza nielosową próbę.

Rozdział 5 zawiera podstawowe wyniki analizy w formie tabelarycznej, jak i opisowej. Interpretacja otrzymanych wyników jest właściwa. Dyskusyjne jest określenie

akceptowanej jakości jeżeli chodzi o dokładność estymatorów. Co oznacza „akceptowalne” w tym wypadku? Uwagi szczegółowe są następujące:

- Niezrozumiałe określenie próby nielosowej.
- Nieprawidłowy opis zawartości kolumn pod tablicą 10. Kolumny nie zawierają metod.
- Błędne powołanie na tablice 29-48.

6. Ocena rozprawy pod względem formalnym i redakcyjnym

Praca niestety zawiera dość dużo błędów redakcyjnych i formalnych. Nie wpływają one zasadniczo na ogólną pozytywną ocenę pracy.

Spis literatury został stworzony nierzetelnie (w tekście występują powołania na literaturę i brak jest ich w spisie, spis nie został utworzony alfabetycznie, występują błędy i brak ujednolicenia w zapisach poszczególnych pozycji).

7. Ocena rozprawy doktorskiej

Rozprawa doktorska charakteryzuje się dobrym poziomem merytorycznym. Część teoretyczna pracy dokonana jest w sposób właściwy wskazując na umiejętności Doktoranta w zakresie doboru właściwych pozycji literaturowych. Prezentacja metod imputacji danych dokonana jest prawidłowo, chociaż czasami widoczny jest brak ujednolicenia zapisu. Należy jednak zwrócić uwagę, że podział na część teoretyczną i praktyczną nie jest jednoznaczny gdyż Doktorant nawet przy opisie istniejących w literaturze przedmioty metod i procedur zwraca uwagę i nawiązuje do swoich badań. Świadczy to między innymi o prawidłowym zaprojektowaniu badania statystycznego. Podczas lektury rozprawy doktorskiej łatwa do wykrycia jest znajomość zarówno metod statystyki opisowej i ekonomicznej, metody reprezentacyjnej, jak i zaawansowanych metod statystyki matematycznej. Badanie statystyczne, którego wyniki przedstawiono w rozprawie wymagało nie tylko znajomości właściwego doboru metod statystycznych ale również znajomości oceny właściwości zastosowanych metod. Metody zastosowane w badaniu to nie tylko metody znane z literatury przedmiotu ale również metoda zaprojektowana i przeprowadzona przez Doktoranta. Świadczy to jednoznacznie o tym, że rozprawa doktorska wykazuje umiejętność samodzielnego prowadzenia pracy naukowej przez Autora.

W ogólnej ocenie rozprawy doktorskiej niestety muszę zaznaczyć, że niektóre z błędów wymienionych przy opisach poszczególnych rozdziałów i błędów formalnych oraz redakcyjnych zostały wykazane były już w pierwszej recenzji i niestety nie zostały poprawione.



Biorąc pod uwagę powyższe stwierdzam, że przedłożona do recenzji rozprawa doktorska mgr. Pawła Zdzisława Lańducha pt. „Wykorzystanie technik imputacyjnych w szacowaniu informacji wynikowych oraz w analizie struktury danych w statystyce przedsiębiorstw” spełnia wymogi stawiane pracom doktorskim określone w Ustawie z dnia 14.03.2003 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. z 2017 r., poz. 1789 z późn. zm.). Rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego i praktycznego oraz wykazuje ogólną wiedzę teoretyczną Doktoranta w dyscyplinie ekonomia i finanse oraz umiejętność samodzielnego prowadzenia pracy naukowej.

W związku z powyższym wnoszę do Rady Naukowej dyscypliny ekonomia i finanse Szkoły Głównej Handlowej w Warszawie o jej przyjęcie i dopuszczenie Doktoranta do publicznej obrony.

Aleksandra Baszczyńska