

Warszawa, 28.05.2026

**dr hab. Barbara Kowalczyk, prof. SGH**

Zakład Statystyki Matematycznej

Instytut Ekonometrii

Kolegium Analiz Ekonomicznych

Szkoła Główna Handlowa w Warszawie

**Recenzja rozprawy doktorskiej**  
**magistra Krzysztofa Szymoniaka-Książka**  
**pt. „Estymatory złożone dla stałokosztowego schematu sekwencyjnego”**

Promotor: dr hab. inż. Wojciech Gamrot, prof. UE

Dziedzina: nauki społeczne, dyscyplina: ekonomia i finanse

Przedstawiona do recenzji rozprawa doktorska podejmuje problem estymacji w populacjach skończonych w warunkach niejednorodności kosztów badania poszczególnych jednostek populacji. W takiej sytuacji klasyczne schematy losowania są nieefektywne z punktu widzenia kosztów realizacji badania. W swojej pracy doktorant przyjmuje, że jednostki populacji mogą różnić się nie tylko istotnością informacyjną, lecz także kosztem ich obserwacji, który może wynikać m.in. z ich dostępności, lokalizacji geograficznej czy indywidualnej charakterystyki. Podejście takie ma bardzo duże znaczenie praktyczne, umożliwia prowadzenie badań na dużych i zróżnicowanych populacjach zapewniając dotrzymanie założeń budżetowych, a także umożliwiając bardziej efektywne wykorzystanie dostępnych środków finansowych i czasu. Wprowadzenie komponentu kosztowego do procesu doboru próby generuje szereg problemów natury teoretycznej, a schematy uwzględniające heterogeniczność kosztów charakteryzują się dużym stopniem złożoności, w związku z czym wiele problemów

teoretycznych pozostaje dla tych schematów nierozwiązanych. Oceniając **wybór tematu pracy**, stwierdzam, iż jest on właściwy, ważny z punktu widzenia teorii oraz użyteczny w praktyce.

Doktorant poprawnie określił **lukę badawczą** w przedstawionym obszarze. Należałoby się jednak odnieść do użytego przy jej sformułowaniu określenia (str. 4): „wykorzystanie informacji o cechach dodatkowych” i doprecyzować je. W swojej pracy doktorant wykorzystuje jedną cechę dodatkową. Przypadek wielowymiarowy nie jest rozpatrywany. W pełni podzielam pogląd doktoranta, iż problem niejednorodności kosztów jest zazwyczaj zaniewany w literaturze, mimo iż w wielu sytuacjach praktycznych koszty badania różnych jednostek populacji mają właśnie heterogeniczny charakter. W tym miejscu należy też podkreślić, iż słowo „stałokosztowy” użyte w tytule pracy, odnosi się do całkowitego kosztu badania, a nie indywidualnych kosztów badania pojedynczych jednostek populacji.

Przejdę teraz do omówienia **struktury rozprawy doktorskiej**. Łączna objętość przedstawionej do recenzji pracy wynosi 150 stron, z czego zasadnicza część pracy zajmuje 131 stron, bibliografia niecałe 6 stron, a spis rysunków 10 stron. Główna część pracy obejmuje 6 rozdziałów poprzedzonych wstępem oraz zwieńczonych zakończeniem. Pierwsze dwa rozdziały rozprawy doktorskiej mają charakter wprowadzający do rozważanego zagadnienia, kluczowe wyniki pracy znajdują się w rozdziałach trzecim, czwartym oraz piątym, natomiast porównanie przedstawionych wcześniej estymatorów złożonych dla stałokosztowego sekwencyjnego schematu losowania zawarte zostało w rozdziale szóstym. Struktura rozprawy doktorskiej jest klasyczna oraz zgodna z przyjętymi standardami.

Pod względem **redakcyjno-technicznym** praca napisana jest dość dobrze. Wszystkie rysunki opisane są niezwykle szczegółowo, co przy tak dużej liczbie przeprowadzonych symulacji i parametrów wziętych pod uwagę było uzasadnione koniecznością zachowania precyzji opisu. Tytuły rysunków mogą wydawać się niezwykle długie, jednak finalnie sprzyjają klarowności przekazu oraz ułatwiają odczytanie wyników licznych symulacji. Wzory napisane są dość starannie. Mam jedynie uwagi do wzoru (5.2), ale mają one raczej charakter merytoryczny. Przedstawię je zatem później.

Przejdę w tym miejscu do **merytorycznej** oceny pracy. Zacznę od krótkiego opisu poszczególnych rozdziałów, przedstawiając jednocześnie pewne uwagi krytyczne i polemiczne odnoszące się do poszczególnych rozdziałów. W dalszej części recenzji przystąpię do omówienia wartości poznawczej rozprawy doktorskiej, jej mocnych stron i watorów.

W **rozdziale pierwszym** doktorant wprowadził do problemu niejednorodności kosztów tj. do sytuacji, w której koszty badania różnych elementów populacji różnią się między sobą. Autor przedstawił w tym rozdziale trzy przykłady populacji z heterogenicznymi kosztami badania jednostek. W pierwszym przykładzie koszty wygenerowano z rozkładu jednostajnego, w drugim z rozkładu logarytmiczno-normalnego, w trzecim założono, że są one proporcjonalne do powierzchni rozważanych gospodarstw. W każdym z przypadków doktorant zobrazował najmniejsze i największe możliwe względne koszty badania. Wykazał ryzyko znacznego przekroczenia budżetu bądź jego niewykorzystania, gdy schemat doboru próby nie uwzględnia heterogenicznego kosztu badania jednostek.

W rozdziale pierwszym zabrakło mi jedynie nieco szerszego odniesienia się do praktyki badań reprezentacyjnych, zwłaszcza poprzez przywołanie liczniejszych i bardziej szczegółowych przykładów.

**Rozdział drugi** stanowi wprowadzenie teoretyczne do dalszej części pracy. Autor przedstawił w nim podstawowe definicje i twierdzenia dotyczące próbkowania oraz estymacji dla populacji skończonych, a także sposób wyznaczania przybliżonej wartości obciążenia oraz przybliżonej wartości błędu średniokwadratowego estymatora średniej w populacji w oparciu o procedurę linearyzacji Taylora. W tym rozdziale doktorant zaprezentował też stałokosztowy schemat Pathaka, należący do rodziny schematów sekwencyjnych, dla którego jednostki dobierane są do próby aż do momentu wyczerpania budżetu.

W moim odczuciu w rozdziale drugim schemat Pathaka mógłby być omówiony nieco dokładniej. Doceniłabym szersze omówienie schematów sekwencyjnych i głębsze osadzenie schematu Pathaka w tej rodzinie rozkładów.

W rozdziale trzecim autor przedstawił estymator ilorazowy średniej w populacji przy zastosowaniu sekwencyjnego stałokosztowego schematu losowania próby Pathaka. Mimo, iż estymatory ilorazowe są ogólnie znane i przebadane w teorii metody reprezentacyjnej, brakuje teorii i analiz dla schematu losowania Pathaka. Przyczyniają się do tego następujące względy: prawdopodobieństwa inkluzji, nawet te pierwszego rzędu, nie są wyrażone znanymi wzorami analitycznymi w tym schemacie losowania, liczebność próby jest zmienną losową, a sam sekwencyjny schemat losowania z regułą stopu uwzględniającą budżet jest analitycznie złożony. Doktorant wyznaczył wzory na przybliżoną wartość błędu średniokwadratowego oraz przybliżoną wartość obciążenia dla estymatora ilorazowego w schemacie Pathaka, korzystając z klasycznej linearyzacji Taylora. Dalej w tym rozdziale autor zaproponował estymator błędu średniokwadratowego i estymator obciążenia estymatora ilorazowego, a następnie na podstawie szeregu symulacji przebadał ich własności.

Do rozdziału trzeciego mam kilka uwag. W badaniach symulacyjnych w rozdziale trzecim autor wziął pod uwagę trzy populacje, w tym dwie wygenerowane oraz jedną rzeczywistą. Doktorant wziął pod uwagę także różne rozkłady kosztów badania jednostek populacji. Mimo mnogości rozpatrywanych przypadków, zabrakło mi tutaj wzięcia pod uwagę sytuacji, w której zmienna badana miałaby rozkład przeciwnie asymetryczny do rozkładu kosztów obserwacji jednostek populacji, tj. przypadku, w którym zmienna badana miałaby przykładowo rozkład prawostronnie asymetryczny, a rozkład kosztów byłby lewostronnie asymetryczny. W praktyce statystycznej w wielu obszarach życia społecznego i gospodarczego znaleźć można wiele tego typu sytuacji.

Moja kolejna uwaga dotyczy estymatora błędu średniokwadratowego oraz estymatora obciążenia estymatora ilorazowego średniej, danymi odpowiednio wzorami (3.5) oraz (3.7). Autor pisze „Zastępując we wzorze (3.4) nieznane wartości oczekiwane, wariacje oraz kowariancję estymatorami (2.23), (2.25) oraz (2.27) uzyskujemy estymator błędu średniokwadratowego (...)”. Chciałabym zaznaczyć, iż w przypadku stosowania estymatora ilorazowego średnia cechy pomocniczej w populacji nie jest nieznana, musimy ją znać jeśli chcemy stosować estymator ilorazowy. Nieznana jest tylko średnia cechy badanej w populacji, ale nie pomocniczej. Bardzo łatwo wyobrazić sobie zatem inny estymator błędu średniokwadratowego danego wzorem (3.4), w którym zostawiamy

średnią cechy pomocniczej w populacji, którą przecież znamy i nie zastępujemy jej estymatorem na podstawie próby. Wydaje się, iż celowe byłoby rozważenie i porównanie własności tych dwóch różnych estymatorów błędu średniokwadratowego estymatora ilorazowego średniej w losowaniu Pathaka. Analogiczna uwaga dotyczy estymacji obciążenia estymatora ilorazowego, które dane jest wzorem (3.6).

Rozważając estymatory złożone wykorzystujące informacje o cenie pomocniczej, rozważa się najczęściej dwa różne scenariusze. W scenariuszu pierwszym zakłada się znajomość średniej wartości cechy pomocniczej w populacji, przy jednoczesnym braku założenia o znajomości indywidualnych wartości tej cechy dla wszystkich jednostek populacji. W drugim scenariuszu założenie jest mocniejsze, zakłada się mianowicie znajomość nie tylko wartości średniej cechy pomocniczej w populacji, ale też znajomość wartości tej cechy dla każdej jednostki populacji. Doktorant nie wskazuje *explicite*, z którym przypadkiem mamy w pracy do czynienia. Patrząc na definicję (3.1) oraz analogiczne definicje w kolejnych rozdziałach, wydaje się, iż doktorant rozważa jednak przypadek pierwszy. Ale skoro zakładamy jedynie znajomość średniej cechy pomocniczej w całej populacji i nie znamy wartości tej cechy dla wszystkich jednostek populacji, to wydaje się, iż powinno być to uwzględnione w koszcie badania, który w przypadku stosowania estymatora ilorazowego powinien jednak wzrosnąć ze względu na dodatkowe koszty badania cechy dodatkowej w wylosowanych jednostkach. Będzie to miało znaczenie zwłaszcza w porównaniu z estymatorem bezpośrednim. Gdybyśmy z kolei przyjęli scenariusz drugi i założyli znajomość cechy pomocniczej dla wszystkich jednostek populacji, to powstaje pytanie czy we wzorach (3.4) oraz (3.6) musielibyśmy szacować wariancję estymatora średniej cechy pomocniczej, czy nie moglibyśmy wziąć jej dokładnej wartości, a co za tym idzie czy nie dostalibyśmy w ten sposób innych, być może lepszych estymatorów błędu średniokwadratowego oraz obciążenia estymatora ilorazowego średniej. Chciałam tą dyskusją pokazać jak ważne jest dokładne i precyzyjne określanie założeń dotyczących cechy pomocniczej, przy których estymacja się odbywa.

Kolejna uwaga odnosi się do przedstawienia schematu Pathaka. Autor w rozdziale trzecim wielokrotnie podkreśla, iż prawdopodobieństwa inkluzji dla schematu Pathaka są niezwykle trudne do obliczenia. Doktorant nie podaje jednak informacji na temat aktualnego stanu wiedzy dotyczącego prób wyznaczania tych prawdopodobieństw.

Doceniłabym, gdyby doktorant przedstawił informację na temat tego czy istnieją jakieś aproksymacje prawdopodobieństw inkluzji pierwszego i drugiego rzędu dla schematu Pathaka, czy były wyznaczane te prawdopodobieństwa symulacyjnie, czy są znane wzory analityczne w pewnych szczególnych przypadkach etc. Tego typu odniesienia do aktualnego stanu wiedzy dotyczącego schematu Pathaka mogłyby się w pracy znaleźć.

**W rozdziale czwartym** Doktorant przeszedł do kolejnego estymatora złożonego, przedstawił estymator iloczynowy przy zastosowaniu stałokosztowego sekwencyjnego schematu losowania, wyznaczył analityczny wzór na przybliżoną wartość błędu średniokwadratowego i analityczny wzór na przybliżoną wartość obciążenia tego estymatora, zaproponował ich estymatory, a następnie przeanalizował dokładność estymacji średniej w populacji za pomocą estymatora iloczynowego z wykorzystaniem schematu losowania Pathaka. Rozdział ten został napisany w sposób bardzo analogiczny do rozdziału trzeciego. Metodyka oraz struktura eksperymentów symulacyjnych zastosowane dla estymatora iloczynowego były w pełni analogiczne do procedury przedstawionej dla estymatora ilorazowego w poprzednim rozdziale.

Moje uwagi sformułowane w odniesieniu do rozdziału trzeciego dla estymatora ilorazowego pozostają zatem aktualne również w odniesieniu do estymatora iloczynowego dla schematu Pathaka. Nie będę ich tu ponownie przytaczać.

**Rozdział piąty** dotyczy kolejnego estymatora złożonego, tj. estymatora regresyjnego. Na początku autor przedstawił estymator regresyjny w różnych ujęciach, przechodząc finalnie do ujęcia z wykorzystaniem sekwencyjnego stałokosztowego schematu Pathaka. Następnie doktorant, wykorzystując jak wcześniej linearyzację Taylora, wyznaczył przybliżoną wartość błędu średniokwadratowego estymatora regresyjnego dla rozważanego stałokosztowego schematu losowania oraz zaproponował estymator błędu średniokwadratowego estymatora regresyjnego. W dalszej części rozdziału piątego autor przedstawił szereg eksperymentów symulacyjnych, opartych zarówno na populacjach wygenerowanych z różnych rozkładów, jak i populacji rzeczywistej. Autor uwzględnił też różne rozkłady kosztów indywidualnych. W symulacjach doktorant wyznaczał symulacyjne wartości błędu średniokwadratowego oraz obciążenia estymatora regresyjnego średniej oraz badał własności estymatora błędu

średniokwadratowego estymatora regresyjnego dla schematu Pathaka w zależności od budżetu badania, współczynnika korelacji pomiędzy cechą badaną i cechą pomocniczą, stosując przy tym przy różne rozkłady cech i kosztów.

Do rozdziału piątego mam dwie uwagi. Przede wszystkim chciałabym zwrócić uwagę na dużą ilość błędów we wzorze (5.2). Doktorant pomieszał w tym wzorze średnią z wartością globalną. Estymatory Horvitz-Thompsona przedstawione zostały dla wartości globalnej cechy badanej i pomocniczej, natomiast za znany parametr populacji wzięto średnią cechy pomocniczej. W rezultacie od średniej w populacji cechy pomocniczej odjęto estymator jej wartości globalnej, natomiast cały estymator nazwano estymatorem wartości średniej w populacji. Błędy są również w estymatorze współczynnika regresji. Pojawił się tam indeks M, który do tej pory odnosił się jedynie do schematu Pathaka, zabrakło też w tym miejscu prawdopodobieństw inkluzji pierwszego rzędu. Na szczęście dla doktoranta wzór (5.2) jest niewykorzystywany w dalszej części pracy i może być potraktowany jedynie jako swojego rodzaju dygresja, gdyż dotyczy on sytuacji estymacji przy znanych prawdopodobieństwach inkluzji pierwszego rzędu. Błędny wzór (5.2) nie ma zatem wpływu na wyniki przedstawione w tym rozdziale, ani na inne wyniki przedstawione w pracy.

W rozdziale piątym, w przeciwieństwie do rozdziałów wcześniejszych, autor nie wyznaczył analitycznej postaci przybliżonego obciążenia rozważanego estymatora złożonego. Mogę się domyślać, że nie udało się tej postaci analitycznej obliczyć. Trochę szkoda. Doktorant jednak uwzględnił analizę zachowania obciążenia estymatora regresyjnego w kontekście schematu Pathaka w badaniach symulacyjnych.

W **rozdziale szóstym** doktorant przedstawił porównanie dokładności i precyzji trzech rozważanych estymatorów złożonych oraz estymatora bezpośredniego, który informacji o cenie pomocniczej nie wykorzystuje w przypadku stosowania statokosztowego sekwencyjnego losowania Pathaka. Porównanie zostało dokonane na podstawie szeregu badań symulacyjnych. Autor uwzględnił w tym rozdziale, analogicznie jak w rozdziałach poprzednich, populację generowaną z dwuwymiarowego rozkładu normalnego, dwuwymiarowego rozkładu logarymiczno-normalnego oraz populację rzeczywistą pochodzącą ze spisu rolnego, obejmującą trzy giny z powiatu dąbrowskiego.

Generując wektor kosztów doktorant wziął pod uwagę rozkład jednostajny, rozkład logarytmiczno-normalny oraz koszty proporcjonalne do powierzchni gospodarstw. We wszystkich przypadkach autor uwzględnił różne wartości współczynników korelacji pomiędzy cechą badaną i pomocniczą. Rozdział szósty jest bardzo wartościowy, pozwala na porównanie estymatorów złożonych i analizę zasadności stosowania informacji pomocniczej w przypadku użycia stałokosztowego sekwencyjnego schematu losowania. Co prawda wyniki otrzymane w tym rozdziale są dosyć intuicyjne, jednak nie były wcześniej otrzymane explicite dla schematu Pathaka i niejednorodnych kosztów badania jednostek populacji.

Szkoda trochę, iż autor nie przedstawił procentowego zysku na efektywności wynikającego z zastosowania estymatorów złożonych, żeby pokazać jakiego dokładnie rzędu zysk na efektywności może być osiągalny. Można to co prawda oszacować na podstawie przedstawionych wykresów, jednak w sposób mocno przybliżony.

W tym momencie przejdę do omówienia **walorów merytorycznych i zalet pracy**.

Podjęta przez doktoranta problematyka dotycząca różnych kosztów badania jednostek należy do szczególnie złożonych i wymagających, a jednocześnie cechuje się wysoką wartością poznawczą i bardzo dużym potencjałem wdrożeniowym. W praktycznych zastosowaniach badawczych założenie o stałości kosztów badania poszczególnych jednostek populacji często nie znajduje odzwierciedlenia w rzeczywistości. W wielu obszarach, obejmujących w szczególności nauki ekonomiczne i społeczne, koszty pozyskania informacji mogą bardzo istotnie różnić się pomiędzy jednostkami populacji. Zróżnicowanie to wynika m.in. z odmiennej dostępności danych, różnego stopnia złożoności jednostek badania, konieczności wykorzystania specjalistycznych metod pomiaru czy też zróżnicowanych warunków terenowych i organizacyjnych. Temat ten, mimo iż niezwykle ważny i absolutnie podstawowy z punktu widzenia praktyki badań reprezentacyjnych i badań statystycznych w ogólności, bardzo często pomijany jest w opracowaniach naukowych. Dzieje się tak ze względu na dużą złożoność problemu. Uwzględnienie w schemacie losowania heterogeniczności kosztów oraz pewnej reguły stopu związanej z ograniczonym budżetem, powoduje wiele komplikacji teoretycznych, w szczególności problemy z wyznaczeniem

prawdopodobieństw inkluzji oraz losową liczebnością próby. Bezpośrednie zastosowanie klasycznych wyników teorii metody reprezentacyjnej oraz wynikających z nich wzorów nie jest wtedy możliwe, co prowadzi do konieczności modyfikacji, adaptacji i weryfikacji wcześniejszego aparatu teoretycznego. Tę właśnie pracę wykonał doktorant w swojej rozprawie doktorskiej.

Czytając rozprawę doktorską od razu pojawiła mi się idea wykorzystania wyników doktoranta do estymacji ilorazu wartości globalnych dwóch cech w przypadku zastosowania stałokosztowego sekwencyjnego schematu Pathaka. W naukach ekonomicznych iloraz wartości globalnych dwóch cech pełni szczególnie ważną rolę. Postać taką przyjmują wszelkiego rodzaju wskaźniki efektywności, produktywności, struktury, intensywności etc. Wyniki doktoranta zawarte w rozdziale trzecim dotyczące estymatora ilorazowego średniej w populacji dla stałokosztowego sekwencyjnego schematu Pathaka niemal automatycznie przenoszą się na estymację ilorazu wartości globalnych dwóch cech. Możliwość bezpośredniego zaadaptowania wyników pracy do kolejnego problemu jest swojego rodzaju wartością dodaną pracy.

Za istotną zaletę pracy należy też uznać zaprojektowanie, zaimplementowanie i przeprowadzenie dużej liczby symulacji o wysokim stopniu złożoności. Symulacje te uwzględniały nie tylko różne scenariusze wyjściowych populacji, ale też różne rozkłady kosztów i różne założone budżety badania. Zastosowana ciekawa i różnorodna wizualizacja wyników symulacyjnych jest dodatkowym atutem pracy.

Na podstawie omówionej wcześniej merytorycznej zawartości pracy, odnosząc się do sformułowanych w pracy **celów rozprawy**:

„C1 Zaproponowanie złożonych estymatorów parametrów populacji skończonej wykorzystujących na różne sposoby dostępne obserwacje cech dodatkowych dla prób losowanych stałokosztowym sekwencyjnym schematem Pathaka

C2 Zaproponowanie sposobów oceny własności oszacowań uzyskiwanych za pośrednictwem złożonych estymatorów parametrów populacji

C3 Ocena własności zaproponowanych estymatorów parametrów populacji w wybranych sytuacjach

C4 Porównanie własności zaproponowanych estymatorów parametrów populacji w wybranych sytuacjach”

stwierdzam, iż wyniki analityczne oraz symulacyjne osiągnięte przez doktoranta i przedstawione w rozdziale trzecim, czwartym, piątym i szóstym pozwalają na potwierdzenie, iż cele te zostały zrealizowane.

Ważny aplikacyjnie i trudny temat pracy, uzyskane wyniki teoretyczne oraz symulacyjne świadczą o dobrym warsztacie badawczym doktoranta, posiadanej wiedzy teoretycznej, zdolności przeprowadzania złożonych badań symulacyjnych, a co za tym idzie o **umiejętności samodzielnego prowadzenia badań naukowych.**

### **Wniosek końcowy**

Rozprawa doktorska magistra Krzysztofa Szymoniaka-Książka pt. „Estymatory złożone dla stałokosztowego schematu sekwencyjnego”, pomimo drobnych uwag krytycznych ujętych w recenzji, stanowi w mojej opinii wartościowe opracowanie naukowe z zakresu dyscypliny ekonomia i finanse.

Stwierdzam, że przedłożona do recenzji rozprawa doktorska spełnia przesłanki, o których mowa w art. 187 ustawy z dnia 20 lipca 2018 r. *Prawo o Szkolnictwie Wyższym i Nauce*. **Wnioskuje o dopuszczenie rozprawy doktorskiej magistra Krzysztofa Szymoniaka-Książka do publicznej obrony w dziedzinie nauk społecznych w dyscyplinie ekonomia i finanse.**

