

Warszawa, 2 lipca 2023 r.

Prof. dr hab. Tomasz Panek
Instytut Statystyki i Demografii
Szkoła Główna Handlowa w Warszawie

Recenzja rozprawy doktorskiej mgr Adama Juszcza
pt. „Zastosowanie danych skrapowanych i skanowanych w pomiarze inflacji”

1. Podstawa recenzji

Podstawę recenzji stanowi pismo Pana dr hab. Rafała Matery, prof. UŁ, Przewodniczącego Komisji Uniwersytetu Łódzkiego do spraw stopni naukowych w dyscyplinie ekonomia i finanse, z dnia 15.05.2023 r. o powołaniu mnie przez Komisję na recenzenta rozprawy doktorskiej mgr Adama Juszcza w dziedzinie nauk społecznych w dyscyplinie ekonomia i finanse. Podstawę prawną recenzji stanowi art. 179 Przepisów wprowadzających ustawę – Prawo o szkolnictwie wyższym i nauce z dnia 3 lipca 2018 r. (Dz.U. 2018 r. poz. 1669) oraz art. 187 ust. 1 i 2 z ustawy z dnia 20 lipca 2018 r (Dz. U. z 2022 r. poz. 574).

2. Ocena wyboru problematyki badawczej

Problematykę badawczą podjętą w rozprawie należy uznać za bardzo trafną i cenną, zarówno od strony teoretycznej jak i aplikacyjnej, a zarazem niezwykle aktualną. Wskaźnik cen towarów i usług konsumpcyjnych (CPI) jest kluczowym narzędziem wykorzystywanym w wielu aspektach polityki gospodarczej i społecznej. Banki centralne wykorzystują go jako główny wskaźnik inflacji, który służy do monitorowania cen i wspomaga podejmowanie decyzji dotyczących wysokości stóp procentowych. Kontrolowanie inflacji jest jednym z podstawowych celów polityki monetarnej. Wiele krajów regularnie aktualizuje swoje systemy podatkowe (np. progi podatkowe) na podstawie CPI, aby zapewnić, że podatki nie rosną



taryfy często są regulowane i indeksowane przy wykorzystaniu CPI. Wskaźnik CPI jest często używany do porównań cen i siły nabywczej dochodów gospodarstw domowych między krajami. Wskaźnik inflacji CPI jest często stosowany do indeksacji płac oraz świadczeń społecznych, takich jak np. emerytury. Szereg umów o pracę i systemów emerytalnych zawiera klauzule indeksacyjne, które wiążą wzrost płac lub świadczeń emerytalnych z CPI. Dzięki temu pracownicy i emeryci są chronieni przed utratą siły nabywczej ich dochodów w wyniku inflacji. Wskaźniki CPI są często używane w negocjacjach płacowych między pracownikami a pracodawcami. Wreszcie inwestorzy mogą wykorzystywać wskaźniki CPI do oceny realnej stopy wzrostu z inwestycji, uwzględniając wpływ na nią inflacji. Podsumowując, pomiar inflacji ma kluczowe znaczenie dla wielu aspektów gospodarki i społeczeństwa, a wskaźnik CPI jest niezbędnym narzędziem dla decydentów, pracodawców, pracowników i inwestorów.

Nowe technologie umożliwiają wykorzystanie do pomiaru inflacji dodatkowych metod pozyskiwania danych do jej pomiaru, jak np. dane skanowane i dane skrapowane, znacznie usprawniających sam proces pomiaru jak i jej precyzję. Dane te ponadto szybciej i lepiej odzwierciedlają zmiany w strukturze konsumpcji oraz zmiany w zachowaniach konsumentów – m.in. coraz więcej transakcji odbywa się online, zmiany w preferencjach dotyczących produktów czy wpływ promocji na decyzje konsumentów lepiej odzwierciedlają dane skanowane i skrapowane.

W świetle powyższych uwag wybór tematyki rozprawy przez mgr Adama Juszcza jest w pełni uzasadniony. Jednocześnie uważam, że przedstawione w niej wyniki badań wnoszą nowe elementy poznawcze, zarówno o charakterze teoretycznym jak i aplikacyjnym, w zakresie analizowanego problemu badawczego.

3. Ocena celów badawczych i hipotez rozprawy

Głównym celem recenzowanej rozprawy było przedstawienie zalet i ograniczeń związanych z wykorzystaniem nowych źródeł danych, takich jak dane skanowane i skrapowane, w pomiarze inflacji. Autor dąży w rozprawie do zbadania jak nowe technologie i metody pozyskiwania danych mogą wpływać na zmniejszenie obciążenia i zwiększenia efektywności pomiaru inflacji. Cel główny dysertacji został osiągnięty poprzez realizację celów szczegółowych dotyczących oceny przydatności różnych metod mających na celu zmniejszenie obciążenia pomiaru inflacji poprzez wykorzystanie danych skanowanych i danych skrapowanych, a mianowicie:

- metod rozszerzania okna czasowego dla wyznaczania indeksów multilateralnych,
- metod filtracji danych i eliminacji obserwacji nietypowych,

- metod agregacji danych ze względu na homogeniczne podgrupy produktów i outlety danej sieci handlowej.

Ponadto realizacji celu głównego rozprawy służyło sprawdzenie możliwości zastosowania różnych formuł indeksowych w ramach wspomnianych alternatywnych źródeł danych, poprzez porównanie ich zalet i wad.

Główny cel pracy został sformułowany w sposób prawidłowy i nie budzi zastrzeżeń, aczkolwiek sposób prezentacji celów szczegółowych nie jest w pełni przejrzysty. Realizacja postawionych celów badawczych opierała się na własnych badaniach empirycznych Autora wykorzystujących:

- dane skrapowane pozyskane przez habilitanta z jednego z najpopularniejszych sklepów internetowych specjalizujących się w sprzedaży odzieży i obuwia,
- dwóch zbiorów danych skanowanych zawartych w pakiecie *PriceIndices* i zbiorze danych dołączonych do publikacji Międzynarodowego Funduszu Walutowego (*CPI Manual*),
- siedmiu wygenerowanych przez Autora, sztucznych zbiorów danych.

Obliczenia zostały przeprowadzone w języku R z wykorzystaniem szeregu pakietów statystycznych.

Za pewien mankament należy uznać brak postawienia hipotez badawczych, które zostałyby zweryfikowane w rozprawie. Chociaż w ramach różnych koncepcji metodologicznych prezentowane są odmienne stanowiska (brak jednego standardu) osobiście uważam, że postawienie hipotez badawczych jest celowe, gdyż ułatwia to późniejsze prowadzenie dyskusji i wyciąganie wniosków.

4. Ocena formalna rozprawy

Rozprawa doktorska mgr Adama Juszcza liczy 219 stron, z czego 173 strony stanowi część zasadnicza, a pozostałe strony to aneks zawierający szczegółowe wyniki przeprowadzonych przez Autora analiz oraz spis tabel i wykresów. Dysertacja rozpoczyna się od krótkiego wstępu, w którym Autor wprowadza czytelnika w tematykę pracy, przedstawiając w sposób syntetyczny historię pieniądza i pojęcie inflacji. W dalszej części Doktorant wskazuje na znaczenie pomiaru inflacji dla prowadzenia polityki gospodarczej i społecznej, a następnie prezentuje cele rozprawy oraz źródła danych wykorzystywanych w rozprawie do realizacji tych celów. Zakończenie wstępu w sposób syntetyczny omawia zawartość poszczególnych rozdziałów.

rozprawie do realizacji tych celów. Zakończenie wstępu w sposób syntetyczny omawia zawartość poszczególnych rozdziałów.

Pierwsze trzy rozdziały rozprawy mają charakter teoretyczny. Pierwszy jej rozdział poświęcony jest tradycyjnym metodom pomiaru inflacji. Początek rozdziału zawiera krótki rys historyczny rozwoju badań nad inflacją. Następnie Autor prezentuje historię powstania i charakterystykę podstawowych miar inflacji, takich jak CPI i HICP. W dalszej części rozdziału omawia źródła danych o cenach i poziomie konsumpcji wykorzystywane w pomiarze inflacji. Na zakończenie tego rozdziału Autor przedstawia główne źródła obciążeń pomiaru inflacji oraz wpływ pandemii COVID-19 na zmianę struktury konsumpcji i pomiar inflacji.

Rozdział drugi rozprawy skupia się na danych skrapowanych jako alternatywnym źródle danych w pomiarze inflacji. Autor omawia w nim definicję i genezę pozyskiwania danych skrapowanych oraz metody ich pozyskiwania. W kolejnych częściach rozdziału omówiono zalety stosowania danych skrapowanych w pomiarze inflacji, a także ograniczenia aplikacyjne i wyzwania metodyczne związane z ich stosowaniem. Na zakończenie tego rozdziału przeprowadzono analizę różnych formuł indeksów stosowanych do szacunku dynamiki cen konsumpcyjnych, pod kątem ich zastosowania dla danych skanowanych.

W rozdziale trzecim rozprawy przedstawiono, w sposób analogiczny jak w rozdziale drugim, dane skanowane jako alternatywne źródło danych w pomiarze inflacji. Autor omówił w nim definicję danych skrapowanych i ich genezę, metody ich pozyskiwania oraz ich zalety i wady. W końcowej części rozdziału poddał dyskusji różne formuły indeksów cen, w kontekście ich zastosowania dla danych skanowanych.

Kolejne rozdziały rozprawy mają charakter aplikacyjny. Rozdział czwarty dotyczy praktycznego zastosowania danych skrapowanych do pomiaru dynamiki cen konsumpcyjnych. Autor omawia w nim swoje badanie empiryczne, zrealizowane na podstawie danych skrapowanych dotyczących odzieży i obuwia, pozyskanych od jednego z największych sklepów internetowych. Na początku porównuje wyniki szacunku dynamiki cen uzyskane za pomocą indeksów bilateralnych i multilateralnych. Przeprowadza także analizę wpływu różnych metod rozszerzeń indeksów multilateralnych oraz metod filtrowania danych na wartości wskaźników cen.

Rozdział piąty rozprawy koncentruje się na wynikach analiz dynamiki cen konsumpcyjnych przeprowadzanych przez Autora na danych skanowanych pozyskanych z publikacji *CPI Manual* oraz zawartych w dodatku do pakietu *PriceIndices*. Na początku rozdziału Doktorant szczegółowo omawia charakterystyki danych, które zostały

wykorzystane w analizie. Następnie przechodzi do porównania różnych indeksów cen zastosowanych w analizie dla różnych zbiorów danych, w tym indeksów bilateralnych i multilateralnych. W dalszej części rozdziału Autor ocenia jak różne metody filtrowania danych oraz rozszerzania okna czasowego obliczeń wpływają na wartości indeksów cen. Poddaje także weryfikacji wpływ różnych metod agregacji danych według grup produktów i według outletów na wartości wybranych indeksów cen. Ponadto, dokonuje analizy czasu obliczeń tych indeksów cen przy zastosowaniu poszczególnych sposobów agregacji danych.

Ostatni, szósty rozdział dysertacji skupia się na porównaniu wyników oceny dynamiki cen za pomocą różnych formuł indeksowych, uzyskanych na podstawie siedmiu sztucznie wygenerowanych zbiorów danych skanowanych w pakiecie *PriceIndices*, na podstawie różnych wartości parametrów zmienności i cen produktów ustalonych na bazie rozkładu lognormalnego. Autor przedstawia charakterystykę utworzonych zbiorów, a następnie dokonuje oceny różnic pomiędzy oszacowaniami dynamiki cen uzyskanymi za pomocą różnych indeksów bilateralnych i multilateralnych, uwzględniając zmienność cen i skład koszyka produktów. Dodatkowo przeprowadził analizę wpływu metod rozszerzeń okna czasowego dla indeksów multilateralnych na uzyskane wyniki. W kolejnej części rozdziału Doktorant porównał czasy obliczeń poszczególnych indeksów cen w zależności od wielkości analizowanego zbioru danych. Na zakończenie przeprowadza analizę zmienności wartości indeksów cen ze względu na rotację produktów w próbie, za pomocą metody *jackknife*.

W Podsumowaniu rozprawy Autor dokonał syntezy uzyskanych wyników badań oraz sformułował w oparciu o nie wnioski i rekomendacje. Na wstępie Autor rozprawy krótko przedstawił wkład jaki rozprawa wniosła do literatury przedmiotu badań, zwłaszcza w kontekście oceny przydatności różnych metod do zmniejszenia obciążeń pomiaru inflacji, wykorzystujących dane skanowane i dane skrapowane. W zakończeniu Doktorant sformułował także rekomendacje dotyczące kierunku i zakresu dalszych badań w omawianym w pracy obszarze.

Tytuł rozprawy w pełni odpowiada omawianej w niej problematyce. Struktura pracy odpowiada realizacji celów rozprawy. Praca jest prawidłowo sformatowana, z odpowiednim użyciem tytułów i podtytułów oraz numeracji stron. Przyjęty w rozprawie układ i sposób uporządkowania treści są właściwe i logiczne, ułatwiające czytelnikowi poruszanie się po omawianej problematyce. Rozprawa jest nie tylko dobrze zorganizowana, ale i zawiera wszystkie niezbędne elementy, takie jak wprowadzenie, opis zastosowanych metod, prezentacja i dyskusja uzyskanych wyników oraz wnioski. W pracy wykorzystano liczne

wykresy, które są odpowiednio oznaczone i opisane, co dodatkowo ułatwia zrozumienie przedstawionych danych.

Autor umiejętnie prowadzi tekst rozprawy konsekwentnie dążąc do osiągnięcia celów badawczych postawionych we wstępie rozprawy. Wskazuje to na dobrą organizację Jego warsztatu badawczego. Doktorant wykazał się także bardzo dobrą znajomością literatury przedmiotu.

Pomimo pozytywnej oceny rozprawy pod względem formalnym chciałem wskazać na pewne błędy i niedociągnięcia, a mianowicie:

1. brak wyjaśnienia niektórych pojęć i metod, jak na przykład *price stickiness* (sądzę, że bardziej poprawnym tłumaczeniem tego terminu jest „sztywność cen”), indeksy superlatywne, metoda Jaccarda,
2. brak opisów osi pionowych na wykresach. Opisy osi są kluczowe dla zrozumienia prezentowanych danych, a ich brak może prowadzić do nieporozumień,
3. brak spójności w odniesieniach do autorów badań i publikacji wspomnianych w rozprawie. Autor w tych odniesieniach raz operuje tylko nazwiskiem, a innym razem pierwszą literą imienia i nazwiska,
4. w pracy pojawiają się nieraz tylko nazwy angielskie pojęć i kategorii zamiast ich polskich odpowiedników, na przykład na str. 40 „*requesty*”,
5. nieprecyzyjne opisy nagłówków niektórych tabel, jak na przykład tabela 35. Głównym punktem odniesienia powinno być „średnie odchylenie absolutne...”, a „kawa” itd. są podkategoriami tego punktu,
6. w pracy występują nieliczne niedociągnięcia w opisie symboli we wzorach, na przykład na str. 41, wzór (6), N_{0t} to jest po prostu liczba produktów dostępnych w okresach 0 i t,
7. w pracy pojawiają się nieliczne błędy stylistyczne i gramatyczne, na przykład:
 - str. 32 „Dane pobierane są z użyciem wcześniej stosowanej tabeli mapującej stworzonej w Exelu...”,
 - str. 34 „W przypadku braku możliwości zastosowania API możliwe jest użycie alternatywnych metod skrapowania. Jednym z nich są...”,
 - str. 38 „Zautomatyzowanie procesu pozyskiwania danych pozwala na ich zbieranie ich z częstotliwością...”,
 - str. 42 „... jest średnią geometryczną średnią $T=1$ stosunków indeksów cen...”,

Rozprawa doktorska mgr Adama Juszcza stanowi obszerny i zarazem kompleksowy studyum na temat wykorzystania danych skanowanych i skrapowanych w pomiarze inflacji. Autor wykazał się dogłębną znajomością wiedzy teoretycznej z zakresu badanej problematyki. Dokonał nie tylko szerokiego przeglądu metodologii dotyczącej wykorzystania danych skanowanych i skrapowanych do pomiaru dynamiki cen detalicznych, ale przeprowadził także szereg własnych analiz empirycznych, korzystając z zaawansowanych technik statystycznych i specjalistycznego oprogramowania. W ramach tych analiz dokonał analizy porównawczej kilkudziesięciu formuł indeksowych stosowanych w pomiarze inflacji, w tym także tych, które ostatnio pojawiły się w literaturze przedmiotu, badając różne aspekty ich stosowania w praktyce. Za główne osiągnięcia Doktoranta uważam:

1. Szeroki przegląd metodologii dedykowanej wykorzystaniu danych skanowanych i skrapowanych w pomiarze inflacji.
2. Kompleksowe przedstawienie metodologii pomiaru inflacji w Polsce, ze szczególnych uwzględnieniem źródeł danych wykorzystywanych do tego pomiaru.
3. Wskazanie zalet i wad stosowania danych skanowanych i skrapowanych w pomiarze inflacji.
4. Przeprowadzenie kompleksowych empirycznych analiz porównawczych kilkudziesięciu formuł indeksowych, w tym także najnowszych propozycji w tym zakresie, weryfikując różne praktyczne aspekty ich stosowania do badania dynamiki cen, a mianowicie:
 - wpływ filtrowania danych na końcowe wartości indeksów,
 - wpływ metod agregacji na wartości indeksów cen,
 - wpływ metod rozszerzania okna czasowego na oceny zmian cen dokonywanych za pomocą indeksów multilateralnych,
 - czas obliczeń indeksów.
5. Propozycja zastosowania indeksu GEKS opartego na formule Dutot do pomiaru inflacji na podstawie danych skanowanych i skrapowanych.
6. Zastosowanie metody *jackknife* do empirycznego sprawdzenia wrażliwości indeksów cen na obciążenie wynikające ze zmian wewnątrz próby, co nie było spotykane wcześniej w literaturze.

Wszystkie te przytoczone osiągnięcia wskazują, że Autor dokonał znaczącego wkładu do literatury przedmiotu w dziedzinie zastosowania danych skanowanych i danych skrapowanych w pomiarze inflacji.

Pomimo pozytywnej oceny całości rozprawy, praca zawiera pewne niedociągnięcia, których usunięcie poprawiłoby jej jakość, a mianowicie:

Uwaga ogólna

Opracowanie i przetwarzanie danych skanowanych i skrapowanych stanowi wieloetapową procedurę obejmującą klasyfikację produktów do grup elementarnych, dopasowanie produktów w czasie (*matching*), filtrowanie danych, implementację wybranych formuł obliczania wskaźników cen oraz ich agregację. Autor co prawda omawia w dysertacji te etapy w ramach operowania danymi skanowanymi i skrapowanymi, ale wydaje się, że wcześniej, przed częścią empiryczną, powinny być one w sposób syntetyczny scharakteryzowane.

Wstęp

Należało szerzej pokazać kluczową rolę wskaźników inflacji w prowadzeniu polityki gospodarczej i społecznej (por. Ocena wyboru problematyki badawczej).

Rozdział I

Wśród źródeł obciążeń pomiaru inflacji można wspomnieć jeszcze o obciążeniach wynikającym z sezonowości (*seasonal bias*) z różnic regionalnych (*geographical bias*) oraz z różnic w preferencjach konsumentów (*preference bias*).

Rozdział II

Str. 39. Autor napisał w rozprawie, że metoda CLIP korzysta z analizy skupień (metod grupowania). Nie podał jednocześnie do czego analiza skupień jest wykorzystywana w metodzie CLIP,

Rozdział III

Jako jedną z zalet danych skanowanych Autor podaje możliwość użycia tzw. *unit value* do sprawdzenia poprawności danych z innych źródeł. Mają one o wiele większą zaletę, a mianowicie umożliwiają odejście od operowania cenami produktów reprezentantów oraz reagują na zmiany preferencji zakupowych konsumentów w ramach grup elementarnych.

Rozdział IV

1. Brak oceny wpływu metod agregacji grup produktów na wartości indeksów cen.

2. Autor podaje w rozprawie, że metoda *jackknife* została wykorzystana do walidacji wyników obliczeń wskaźnika CPI dla danych skanowanych, ale pomocne byłoby przedstawienie więcej szczegółów na ten temat. Warto zauważyć, że metoda ta ma pewne ograniczenia. Na przykład może ona nie być skuteczna w przypadku niektórych skomplikowanych charakterystyk lub gdy dane mają określoną strukturę zależności. Habilitant opisuje, jak metoda została implementowana, ale nie dostarcza szczegółów, jak dokładnie są obliczane średnia i odchylenie standardowe.
3. W pracy nie ma rozdziału dedykowanego porównaniu formuł indeksowych na podstawie sztucznych zbiorów danych skrapowanych, podobnie jak to jest zrobione dla danych skanowanych w rozdziale 6. Sugeruję Autorowi przeprowadzenie tego typu analiz w dalszych jego pracach badawczych.

Konkluzja końcowa

Biorąc pod uwagę poszczególne elementy oceny stwierdzam, że recenzowana rozprawa doktorska stanowi oryginalne rozwiązanie problemu naukowego, o walorach poznawczych i aplikacyjnych, w dziedzinie nauk społecznych, w dyscyplinie ekonomia i finanse. Doktorant podjął ważny dla nauki i praktyki temat oraz wykazał się umiejętnością samodzielnego prowadzenia badań naukowych. Autor rozprawy wykazał się jednocześnie bardzo dobrym ogólnym poziomem wiedzy teoretycznej w badanym obszarze oraz opanowaniem warsztatu badawczego. Pracę oceniam bardzo pozytywnie, mimo zgłoszonych do niej uwag, które częściowo mają charakter dyskusyjny.

Podsumowując, stwierdzam, że przedłożona do recenzji rozprawa doktorska mgr Adama Juszcza pt. *Zastosowanie danych skrapowanych i skanowanych w pomiarze inflacji* w pełni spełnia wymogi stawiane pracom doktorskim określone w Ustawie z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce (Dz.U. z dnia poz. 1669 ze zm.). Wnoszę z tego powodu o jej przyjęcie i dopuszczenie mgr Adama Juszcza do publicznej obrony. Biorąc pod uwagę przedstawione w recenzji argumenty wnioskuję również o wyróżnienie pracy.

