

Poznań, 21.04.2021

dr hab. Agata Kliber, prof. UEP
Katedra Matematyki Stosowanej
Uniwersytet Ekonomiczny w Poznaniu
Al. Niepodległości 10
61-875 Poznań

**Recenzja rozprawy doktorskiej
pana magistra Jakuba Rybackiego**

p.t. *Essays on ineffectiveness of central bankers, commercial forecasters and international financial institutions – evidence from macroeconomic forecasting and policymaking*

(dziedzina: nauki ekonomiczne, dyscyplina: ekonomia)

**napisanej w Kolegium Analiz Ekonomicznych Szkoły Głównej Handlowej, pod
kierunkiem dr. hab. prof. SGH Dobromiła Serwy**

1. Podstawa formalno-prawna sporządzenia recenzji

Podstawę formalno-prawną sporządzenia recenzji rozprawy doktorskiej pana magistra Jakuba Rybackiego pt. *Essays on ineffectiveness of central bankers, commercial forecasters and international financial institutions – evidence from macroeconomic forecasting and policymaking* stanowi uchwała nr 142 z dnia 17 lutego 2021 roku Rady naukowej Szkoły Głównej Handlowej dyscypliny ekonomia i finanse, sygnowana pismem nr K:589864 z dnia 23 lutego 2021 roku.

2. Cel, układ i struktura pracy

Celem, jaki Doktorant sformułował w Streszczeniu rozprawy, jest: *pokazanie, że działaniom banków centralnych towarzyszyły niewielkie korzyści i duże koszty społeczne* (pierwsza część rozprawy) oraz *pokazanie, że prognozy analityków komercyjnych dotyczące dynamiki wzrostu PKB oraz stanu finansów publicznych zawierają liczne mankamenty* (druga część rozprawy).

W celu realizacji tych celów, pan magister Rybacki sformułował następujące hipotezy:

1. Polityka komunikacyjna banku centralnego ma niski wpływ na oczekiwania odnośnie stóp procentowych w małych gospodarkach otwartych.
2. Utrzymanie łagodnej polityki pieniężnej w strefie euro spowodowało pojawienie się baniek spekulacyjnych na głównych rynkach nieruchomości.

3. Pojawienie się niepożądanych efektów polityki pieniężnej i brak realizacji celów inflacyjnych banków są wynikiem błędów popełnianych przez departamenty badań ekonomicznych.
4. Prognozy analityków komercyjnych w Polsce charakteryzują się nieefektywnością, na przykład zawierają systematyczne obciążenia.
5. Instytucje międzynarodowe stosują podwójne standardy w ocenie finansów publicznych krajów UE – szacunki dla niektórych krajów mogą mieć systematyczne błędy.

Rozprawa składa się z cyklu ośmiu rozdziałów, które stanowią opublikowane przez Autora artykuły, poprzedzonych pięciostronicowym wstępem i zakończonych dwustronicowym zakończeniem. Do rozprawy zostało dołączone jednostronicowe streszczenie.

3. Ocena wartości merytorycznej pracy

Trzon rozprawy stanowi cykl ośmiu artykułów, z których cztery zostały opublikowane, a trzy były przyjęte do druku w momencie złożenia rozprawy. Artykuły zostały wydane bądź przyjęte do druku w czasopismach polskich, ale o zasięgu międzynarodowym, indeksowanych w bazach Scopus lub ERICH +, w tym w Banku i Kredycie (dwa artykuły), w czasopiśmie Central European Economic Journal (dwa artykuły), Gospodarka Narodowa (jeden artykuł). Jeden z artykułów został opublikowany w czasopiśmie Economic Research in Finance, które wprawdzie nie ma przyznanych punktów na liście ministerialnej, ale powiązane jest z międzynarodową konferencją ERFIN, która charakteryzuje się wysokim poziomem merytorycznym. Jeden artykuł został włączony do monografii *Recent trends in the real estate markets and its analysis* (wydawnictwo: SGH).

3.1. Część pierwsza rozprawy

Część pierwszą rozprawy rozpoczyna rozdział: *“Does Forward Guidance Matter in Small Open Economies? Examples from Europe”*, będący przedrukiem artykułu z czasopisma *Econometric Research in Finance* z 2018 roku. Celem artykułu było zbadanie efektywności narzędzia *forward guidance* stosowanego w Szwecji, Norwegii i Czechach. Narzędzie to jest jedną z form polityki informacyjnej dotyczącej stóp procentowych stosowaną przez banki centralne i polega na publikowaniu ścieżki stóp procentowych w horyzoncie czasowym przekraczającym najbliższe posiedzenia organu decyzyjnego (za: NBP, *Raport o inflacji*, 2013). W omawianym artykule Autor postawił sobie za cel *zbadanie maksymalnego horyzontu prognozy, w którym polityka forward guidance, a konkretnie – projekcja przyszłej ścieżki stóp*

procentowych przez bank narodowy, wywiera większy wpływ na oczekiwania rynkowe dotyczące kształtowania się trzymiesięcznej międzybankowej stopy procentowej, niż polityka Europejskiego Banku Centralnego (dalej: EBC). Oczekiwania rynkowe dotyczące krajowej stopy IBOR 3M zostały wyprowadzone z modelu Nelsona-Siegela oszacowanego na podstawie stóp międzybankowych oraz instrumentów *swap* i skonfrontowane z oczekiwaniami rynkowymi dotyczącymi kształtowania się stopy EURIBOR 3M (również wyprowadzonymi z modelu Nelsona-Siegela) oraz projekcjami ścieżki stóp procentowych publikowanymi przez banki narodowe. Jakość prognoz była porównywana za pomocą wielkości błędu RMSE oraz z wykorzystaniem testu Diebolda-Mariano (1995). Badanie obejmowało lata 2007-2018. Autor konkluduje, że do 2012 roku polityka *forward guidance* nie wywierała wpływu na oczekiwania, natomiast po 2021 roku można zaobserwować jej istotny wpływ na prognozy w krótkim horyzoncie (na prognozy o horyzoncie równym i przekraczającym sześć kwartałów, istotny wpływ, według Autora, wywierała polityka monetarna strefy euro).

Mam kilka uwag do tej części pracy. Pierwsza jest natury czysto technicznej: Dlaczego odwołując się do modelu Nelsona-Siegela w artykule pierwszym, autor cytuje pracę Diebolda i Lee z 2006 roku, zamiast oryginalnego artykułu z 1987 roku? Ten ostatni w ogóle nie został uwzględniony w literaturze. Drugi komentarz dotyczy samego badania wpływu polityki banku centralnego na oczekiwania. Oczekiwania wyprowadzane są z modelu struktury terminowej, który może być oszacowany różnymi metodami. W tym badaniu Autor użył metody Nelsona-Siegela. Dlaczego tej? Czy Autor testował wrażliwość wyników na zmianę metody uzyskiwania struktury terminowej (Nelsona-Siegela-Swenssona, *cubic splines*, czy zupełnie innej klasy modeli)? W takim przypadku najprawdopodobniej uzyskamy inne oszacowania oczekiwań. Dodatkowo, często prognozy uzyskane z modeli struktury terminowej dekomponuje się na „czyste” oczekiwania oraz na premię terminową – czy Autor rozważał zastosowanie tej dekompozycji i zweryfikowanie wniosków? Wydaje mi się, że dobrym uzupełnieniem artykułu byłoby też policzenie trafności prognoz uzyskanych za pomocą modelu NS i zbadanie, czy wyniki uzyskane przez Autora zmieniają się wraz ze zmianą trafności tych prognoz.

Ostatecznie – w swoim artykule Autor porównuje oczekiwania rynkowe dotyczące krajowych stóp IBOR z oczekiwaniami rynkowymi dotyczącymi stóp EURIBOR 3M, wyprowadzonymi z tego samego modelu. Dlaczego te ostatnie miałyby odzwierciedlać politykę EBC? Analogicznie można byłoby przecież założyć, że rynkowe oczekiwania dotyczące stóp IBOR odzwierciedlają politykę krajowych banków centralnych. Ostatecznie, czy nie jest naturalne, że te oczekiwania, wyprowadzone na podstawie tego samego modelu,

będą ze sobą ściślej powiązane, niż ze ścieżką stóp procentowych publikowaną przez banki centralne?

Rozdział drugi nosi tytuł *Forward guidance and the private forecast disagreement – case of Poland* i jest przedrukiem artykułu opublikowanego w 2019 roku w czasopiśmie Bank i Kredyt. Poświęcony jest on analizie efektywności polityki *forward guidance* prowadzonej (z przerwami) przez Narodowy Bank Polski w latach 2013-2019. Jego celem było zweryfikowanie, czy wprowadzenie tego narzędzia istotnie obniżyło niepewność związaną z kształtowaniem się stóp procentowych. Autor wykazał, że wprowadzona w 2013 roku polityka *forward guidance* przyczyniła się do obniżenia ryzyka stopy procentowej w horyzoncie jednego i dwóch lat. Jej porzucenie w 2014 roku spowodowało wzrost ryzyka, natomiast powrót do tego narzędzia w 2017 roku spowodował powolne zmniejszenie niepewności. Wpływ na prognozy o długim horyzoncie można było obserwować dopiero po roku. Autor dokładnie opisuje metodę badawczą, poddając ją konstruktywnej krytyce (kontrowersyjne przyjęcie założenia o rozkładzie normalnym). Na podstawie wniosków z badania, pan mgr Rybacki formułuje interesujące wnioski: inaczej niż w krajach rozwiniętych, zastosowanie narzędzia *forward guidance* okazało się skuteczne, przyczyniając się do obniżenia niepewności związanej z prognozą stóp procentowych. Dodatkowo, wpływ *forward guidance* okazał się zróżnicowany w okresie prezydentury panów: Marka Belki i Adama Glapińskiego.

Artykuł ten oceniam wysoko – poruszane w nim problemy mają wymiar zarówno praktyczny, jak i empiryczny. Badanie zaplanowane zostało starannie i celowo. Pewnym mankamentem jest brak statystyk opisowych badanych szeregów, brak informacji o częstotliwości wykorzystywanych danych i brak ewentualnych testów stacjonarności uprawniających badacza do wykorzystania regresji liniowej.

Rozdział trzeci jest przedrukiem artykułu w jego wersji roboczej z bazy MPRA, zaakceptowanego do druku w monografii pt. *Recent trends in the real estate markets and its analysis*. Koresponduje on z **hipotezą 2**: *Utrzymanie łagodnej polityki pieniężnej w strefie euro spowodowało pojawienie się baniek spekulacyjnych na głównych rynkach nieruchomości*. Celem artykułu było zbadanie, czy polityka Europejskiego Banku Centralnego wywierała wpływ na ceny mieszkań w pięciu miastach europejskich: Amsterdamie, Frankfurtie, Londynie, Paryżu i Mediolanie (nawiasem mówiąc, Frankfurt nie jest stolicą, w związku z czym cel badania przedstawiony w *Introduction* powinien być poprawiony).

Ten rozdział jest napisany dużo mniej starannie niż poprzednie dwa. Przede wszystkim, jest bardzo chaotyczny i sprawia wrażenie, jakby Autor zaczął pisać go „od środka”. Zamiast zacząć od krótkiego wyjaśnienia modelu teoretycznego, który jest podstawą analizy, na

początku opisuje niekonwencjonalne instrumenty polityki monetarnej EBC stosowane w latach 2008-2009, model teoretyczny opisuje dopiero w podrozdziale 4, a z kolei podrozdział piąty zawiera kontynuację opisu modelu i opis danych. Autor nie wyjaśnia nigdzie w artykule, co rozumie pod pojęciem *discretionary bias* i czy pojęcie to stosowane jest wymiennie z *time-inconsistency bias*. Brakuje też wyjaśnienia, dlaczego Autor uważa, że polityka banków centralnych miałaby mieć wpływ na to, czy ludzi jest stać na mieszkania? Czym jest *headline index*, o którym Autor wspomina na stronie 7, a który zdaje się być istotny z punktu widzenia badania? Czy ma on coś wspólnego z indeksem UBS (rozumiem, że jest to jeden z indeksów opracowywanych przez bank inwestycyjny UBS, o których Autor wspomina na stronie 1)? W artykule brakuje podrozdziału z dokładnym opisem danych i statystykami opisowymi. Ostatecznie, wyniki oszacowania modelu VARX dla Frankfurtu (Tabela 2) budzą wątpliwości natury formalnej. Wydaje mi się, że oszacowany model VARX może być niestacjonarny: współczynnik α_{11} w równaniu dla ΔBbl wynosi 1.16, co sugeruje że sam proces jest niestacjonarny, a zatem modelowanie tego szeregu za pomocą VAR nie jest uprawnione. Nie jest do końca jasne, w jakim celu Autor prezentuje w podrozdziale 7 wyniki wyszukań frazy: *central bank independence*.

„Dodatkowo, w artykule brakuje numeracji stron, występują liczne błędy językowe (*expect* zamiast *except*, *few* zamiast *a few*, brak przedimków określonych i nieokreślonych, konstrukcja: *may (...) did not generate*, i inne), zdarzają się też niewyjaśnione skróty (np. skrót LSAP pojawia się na stronie 1 w *Introduction*, a wyjaśniony jest dopiero na str. 2 w *Literature Review*, skróty TLTRO, UBS i PSPP nie są wyjaśnione nigdzie). Edycja wzorów jest niestaranna, na przykład zamiast symbolu mnożenia: \cdot , Autor używa gwiazdki: $*$, zamiast napisać: R^2 , pisze: R^2 (Tabela 1). Konkludując – rozdział ten przypomina raczej szkic notatek, a nie artykuł naukowy. Łączy się bardzo luźno z pozostałymi rozdziałami i w mojej opinii w obecnej postaci nie powinien być w ogóle umieszczony w rozprawie.

Rozdział czwarty, kończący pierwszą część pracy, zatytułowany został: *Are Central Banks' Research Teams Fragile Because of Groupthink? Evidence on Inflation Targeting*, który jest przedrukiem artykułu zaakceptowanego do publikacji w *Gospodarce Narodowej*. Jest to interesujące badanie, w którym Autor bada wpływ zróżnicowania grup badawczych analityków banków centralnych na zdolność banków do realizacji założonego celu inflacyjnego. W przeciwieństwie do poprzednich artykułów, ten obejmuje obszerną próbę badawczą, składającą się jednak w przeważającej części z europejskich banków centralnych. Okres badania ograniczony jest do lat 2014-2019. Pan magister Rybacki obrazuje stopień koncentracji wykorzystując dwa mierniki: jeden jest sumą udziałów autorów przypisanych do

danego klastra w całej populacji naukowców publikujących w serii *Working Papers* danego banku centralnego i opiera się na liczbie naukowców przypadających na dany ośrodek badawczy, pomijając liczbę publikacji przypadającą na jednego autora. Drugi indeks koncentruje się na udziale publikacji wydanych przez autorów w całkowitej liczbie publikacji z cyklu *Working Papers* danego banku centralnego. Pan magister Rybacki pokazuje zróżnicowanie pomiędzy grupami analitycznymi różnych banków centralnych, a także – korzystając z modelu regresji liniowej – weryfikuje hipotezę o tym, że małe zróżnicowanie badaczy przyczynia się do odchyień w realizacji celu inflacyjnego. Przy poziomie istotności przekraczającym nieznacznie 5% Autor stwierdza, że nie ma podstaw do odrzucenia hipotezy zerowej. Przyznaje jednak, że badanie należałoby przeprowadzić ponownie na większej próbie.

Przedstawione badanie jest dobrze zaprojektowane, przedstawione są w nim konkretne hipotezy badawcze, które Autor weryfikuje metodami ilościowymi. Zabrakło mi w nim jednak weryfikacji statystycznej liczby otrzymanych klastrów, na przykład za pomocą miernika *silhouette*. Ponadto, oszacowania modelu zostały przedstawione tylko na wykresach. W załączniku pojawia się Tabela 8, w której Autor przedstawia analizę stabilności wyników, natomiast nie wiadomo, co przedstawiają liczby w tabeli. Autor twierdzi, że są to oszacowania parametrów, ale nie zgadza się to z liczbami na wykresie 6. Wydaje się, że są to p -wartości, a nie oszacowania – przynajmniej w przypadku HHI (na wykresie 6 oszacowanie parametru wynosi 5.19, a z tekstu wynika, że p -wartość przewyższa 0.05). Z kolei oszacowanie stałej odczytane z wykresu wynosi 18.3, a podane w tabeli 5.89. Czym są w liczby w nawiasie? Wyglądają, jak przedziały ufności, ale nie obejmują krańcami oszacowanego parametru (wiersz pierwszy).

Dodatkowo, mimo wszystko trudno jest obronić tezę, że małe zróżnicowanie ośrodków badawczych przyczynia się do odchyień w realizacji celu inflacyjnego. Po pierwsze, czy pracownicy tych ośrodków koncentrują swoje badania tylko na realizacji celu inflacyjnego, czy są to publikacje zróżnicowane tematycznie? W wielu publikacjach analityków banków centralnych pojawia się też klauzula, że badania te odzwierciedlają przekonania badaczy, a nie oficjalne stanowisko banku. Wydaje mi się, że teza została w tym badaniu postawiona nieco zbyt śmiało – równie dobrze można było się ograniczyć do zbadania, jak bardzo skoncentrowane są ośrodki badawcze w różnych bankach, co samo w sobie jest bardzo ciekawym zagadnieniem.

3.2. Część druga rozprawy

Druga część rozprawy doktorskiej poświęcona jest weryfikacji jakości prognoz wielkości makroekonomicznych formułowanych przez profesjonalnych prognostów. Otwiera ją **rozdział piąty** zatytułowany *Polish GDP Forecast Errors: A Tale of Inefficiency*, który jest przedrukiem artykułu przyjętego do publikacji w czasopiśmie *Bank i Kredyt*. Celem artykułu było zbadanie własności błędów prognoz PKB formułowanych przez profesjonalnych prognostów w ramach konkursu na najlepszego analityka makroekonomicznego, organizowanego przez dziennik „Rzeczpospolita” i „Parkiet”. Badanie dotyczy okresu 2013-2019. Autor wykazuje, że prognozy formułowane w latach 2016-2019 były systematycznie zaniżane, prognostom w badanym okresie nie udało się skutecznie przewidzieć skutków zmian polityki fiskalnej, a także że progności mieli tendencję do nadmiernego rewidowania swoich prognoz. Dodatkowo, Doktorant wykazał istnienie efektu stadnego w formułowaniu prognoz, polegającego na tym, że analitycy, których prognozy odchyłały się znacznie od mediany, mieli tendencję do rewidowania ich w tym kierunku.

Artykuł ten jest dobrze napisany, z precyzyjnie sformułowanym celem i poprawnie dobranymi metodami badawczymi. Modele są poprawnie opisane. Hipotezy postawione w pierwszej części artykułu zostały zweryfikowane na podstawie oszacowań modeli panelowych z efektami stałymi (ang. *fixed effects*), w których zmienną objaśnianą były: błędy prognoz (model 1), wielkość korekty (model 2) i odchylenia prognoz od mediany (model 3). Wybrane metody badawcze są zgodne z tymi, które stosowane są w literaturze przedmiotu (Autor powołuje się na badanie: Pons-Novell, 2003). Badanie zostało dobrze przemyślane, a przeprowadzona dyskusja i wyciągnięte wnioski są spójne z celem badawczym. To, czego brakuje mi w tym badaniu, to pokazanie własności statystycznych badanych szeregów, testów diagnostycznych wykorzystanego modelu panelowego (np. testu Hausmana) oraz wyjaśnienia, jaką metodą było szacowane równanie (4)? Nie jest też jasne, co pokazuje Autor w Tabeli 4 w wartościach w nawiasie. Czy jest to błąd oszacowania i *p*-wartość?

Problematyka jakości prognoz formułowanych przez uczestników konkursu jest kontynuowana w kolejnym, **szóstym rozdziale**. Nosi on tytuł: *Macroeconomic forecasting in Poland: The role of the forecasting competitions* i jest przedrukiem artykułu opublikowanego w 2020 roku w czasopiśmie *Central European Economic Journal*. Jest to interesujące i rzetelnie przeprowadzone badanie. Jego celem było zbadanie, czy bardziej trafne prognozy pewnych zmiennych makroekonomicznych (np. inflacji) zwiększają istotnie szanse uczestników na wygranę konkursu. Dodatkowo, Autor bada, czy istnieją pewne prawidłowości w tym, jacy analitycy zajmują jedno z pierwszych pięciu miejsc.

Badanie zostało przeprowadzone z wykorzystaniem modelu logitowego. Poprzedza je dokładny opis danych i konkursu organizowanego przez „Parkiet”, tak że artykuł może być zrozumiały nawet dla potencjalnego czytelnika spoza Polski. W przeglądzie literatury Autor odwołuje się do pewnych prawidłowości, o których mowa była w poprzednim artykule. Jedną z nich jest dążenie prognostów do „rynkowego konsensusu” (tzw. zachowanie stadne) – przy czym bardziej doświadczeni progności mogą wykazywać skłonność przeciwną, tj. do celowego stawiania prognoz znacznie odbiegających od mediany. Inną prawidłowością jest poruszana w rozdziale czwartym skłonność analityków do stawiania prognoz zgodnych z prognozami środowiska, z którego się wywodzą (np. prognozy analityków z banków komercyjnych często charakteryzują się podobnymi błędami). Autor podaje postać funkcji, za pomocą której liczone są punkty przyznawane poszczególnym prognozom i komentuje jej własności. W szczególności zwraca uwagę na to, że analitycy nie są szczególnie mocno „karani” za błędy prognoz pojawiające się w okresie podwyższonej zmienności lub niepewności. Autor wykazuje, że najmniejsze rozbieżności prognoz występują w przypadku: inflacji, rynku pracy i wskaźników finansowych (kurs EUR/PLN oraz rentowność dziesięcioletnich obligacji rządowych) – w związku z czym wnioskuję, że zasady konkursu Parkietu mogą skłaniać prognostów do koncentrowania się na tych kategoriach w większym stopniu, niż na pozostałych. Dodatkowo, pokazuje też, że istnieje istotnie (w sensie statystycznym) większa szansa, iż progności plasujący się w pierwszej piątce najlepszych w roku poprzedzającym, będą zwycięzcami konkursu w roku bieżącym. Na podkreślenie zasługuje rzetelna analiza dobroci dopasowania modeli przedstawiona w załączniku do artykułu.

Rozdział siódmy to jedyny nierecenzowany artykuł prezentowany w tej rozprawie. Nosi on tytuł: *Fiscal Deficit Forecasts by International Institutions: Evidence for a Double Standard?* Autor kontynuuje w nim temat oceny jakości prognoz. Tym razem nie ogranicza się do rynku polskiego, ale koncentruje się na prognozach wielkości fiskalnych w trzydziestu sześciu państwach. Stawia tezę, iż międzynarodowe instytucje finansowe konstruują nadmiernie optymistyczne prognozy dla krajów mocno zadłużonych (takich jak Hiszpania, Portugalia, Włochy), natomiast nadmiernie pesymistyczne dla krajów, w których rządy postrzegane są jako populistyczne. Rozdział ten koresponduje z **hipotezą piątą**: Instytucje międzynarodowe stosują podwójne standardy w ocenie finansów publicznych krajów UE – szacunki dla niektórych krajów mogą mieć systematyczne błędy.

W celu weryfikacji tej hipotezy, Autor bada prognozy publikowane przez Międzynarodowy Fundusz Walutowy (IMF), Komisję Europejską (EC) i Organizację Współpracy Gospodarczej i Rozwoju (OECD), dotyczące deficytu fiskalnego w wybranych

trzydziestu sześciu gospodarkach obejmujących kraje Unii Europejskiej, kraje członkowskie OECD i wybrane duże gospodarki spoza tej grupy, tj. Chiny i Rosję. Autor pomija Grecję i Cypr, argumentując to faktem otrzymywania przez te państwa pakietów pomocowych (tzw. *bailout*) od innych krajów członkowskich, jak również Irlandię, Islandię i Słowenię, które w danym okresie doświadczyły kryzysów bankowych. Zakres czasowy badania obejmuje lata 2008-2018. Dane wykorzystane przez Autora obejmują wybrane wielkości reprezentujące sytuację fiskalną państwa, np. zobowiązania i pożyczki rządowe roczną dynamikę PKB, deflator PKB, inflację (CPI), konsumpcje sektora prywatnego i publicznego – w zależności od ich dostępności w bazach. Wykorzystane w badaniu prognozy dotyczyły okresu: bieżącego roku (tzw. *nowcast*) oraz następnych trzech lat (IMF) lub dwóch lat (EC, OECD). Badanie przeprowadzono za pomocą modelu panelowego z efektami stałymi, w którym zmienną objaśnianą była różnica między faktyczną realizacją deficytu fiskalnego, a jego prognozą.

Opis modelu (3) jest niestaranny, w związku z czym trudno się zorientować, co w zasadzie Autor policzył. Z jednej strony, na str. 9⁵, Doktorant pisze, że zmienna ma_x „describes a vector of macroeconomic assumptions”. Czym są te założenia makroekonomiczne? Czy Autor ma na myśli prognozy? Trzy linijki później, odnosząc się do zmiennej ma_x^3 , stwierdza: *We introduced the third power of macroeconomic assumptions error*. Czym w takim razie jest zmienna ma_x ? Czy są to oczekiwania (jak zdefiniowane?), prognozy, czy błędy? Nie jest też jasne założenie Autora dotyczące dodatniego znaku parametrów a_x oraz a_{2x} : „*better activity, labor market conditions, or higher inflation should result in lower deficits*” (ortografia i interpunkcja Autora). Przecież po lewej stronie równania mamy różnicę: $deficit_t - forecast_{t,h}$, czyli dodatnie znaki parametrów oznaczałyby, że np. im lepsze warunki na rynku pracy, tym większa różnica między realizacją a prognozą?

W równaniu (4) Autor poszerza zbiór zmiennych objaśniających o zmienne opisujące sytuację polityczną kraju oraz o czynniki instytucjonalne. Równanie (4) również opisane jest niestarannie. Nie wiadomo, czym jest zmienna $Event_t$ oraz w jaki sposób otrzymano zmienną $e_{fin,t-1,h}$ - czy jest to po prostu opóźniona zmienna objaśniana, czy błędy zostały uzyskane w inny sposób? Ostatecznie, równanie (4) nie ma postaci panelowej – brakuje indeksu kraju. Jeśli μ ma być efektem stałym, to w modelu powinno się pojawić jako μ_i , natomiast stałej a_0 nie powinno tam w ogóle być. W opisie równania pojawiają się parametry beta, które nie pojawiają się w żadnym z równań. Brak jest szczegółów dotyczących sposobu uzyskania parametrów, tj. jaką metodą został oszacowany model panelowy (4)? Jeśli jest to model z efektami stałymi, a po prawej stronie równania znajduje się opóźniona zmienna objaśniana, to mamy duże

niebezpieczeństwo (w zasadzie pewność) pojawienia się obciążenia Nickela („Biases in Dynamic Models with Fixed Effects”, *Econometrica* vol 49, no 6, ss. 1417-1426, 1981) (ta sama uwaga dotyczy modelu (4) z rozdziału piątego). Czym Autor się kierował wybierając ten konkretny model? Na jakiej podstawie przyjęto, że efekty indywidualne są deterministyczne, a nie losowe? Czy przeprowadzony został test Hausmanna w celu wykluczenia modelu panelowego z efektami losowymi? Czy Autor zweryfikował możliwość zastosowania „zwykłej” regresji, bez efektów stałych (tzw. *pooled regression*)? W części artykułów, na które powołuje się Autor, jest to właśnie preferowana specyfikacja (np. Pina i Venes, 2011). Jaka jest istotność statystyczna efektów stałych? Dlaczego w Tabeli 2 nie podano oszacowań parametrów φ_t ? Czy w ogóle występowały w modelu? Dlaczego w Tabeli 5 wyniki są rozbite na kolumny $T0, T1, T2$? Czy dla każdego horyzontu prognozy był inny panel i osobny model?

Na stronie 10₁₄ Autor pisze też: (...) *forecast errors of government deficits are not randomly distributed*. Co Autor przez to rozumie – że błędy były deterministyczne? Czy chodziło raczej o to, że wykazywały autokorelację (nie były od siebie niezależne)? A może nie były *iid*?

O ile pomysł na artykuł jest ciekawy, to sam tekst napisany jest bardzo niestarannie. Pomijając bardzo nieprecyzyjny opis modelu, w artykule występują liczne błędy językowe i literówki, brak jest numeracji stron, a bibliografia nie została uszeregowana alfabetycznie. Występują błędy w cytowaniu literatury (np. na str. 8 powinno być Brueck & Stephan, 2005 a nie: Brueck & Tilman, 2005). Brakuje standardowo wykonywanych testów poprawności specyfikacji modelu oraz wrażliwości wyników. Tabele dołączone do załącznika nie są dobrze opisane. Czym są liczby w nawiasach w Tabelach 2-4? Czy są to przedziały ufności? Jeśli tak, to jaki był przyjęty poziom ufności? Dlaczego w Tabeli 2 nie ma zmiennej zerojedynkowej *Polish reform of retirement system*, a jest analogiczna zmienna dla Węgier? Dlaczego w Tabeli 3 są po dwie zmienne zerojedynkowe o takiej samej nazwie? Dlaczego w tabelach nie ma oszacowania zmiennej $Event_t$? Wszystko to sprawia, że trudno jest ocenić jakość prezentowanych przez Autora wyników. Autor pisze: *cross-section fixed effects were statistically significant in nearly every equation*. Zgaduję, że odnosi się do wartości w Tabeli 5, w której *p*-wartości w ogóle nie podano. Nie jest powiedziane, jak należy interpretować te wartości w odniesieniu do wyrazów wolnych z pozostałych tabel (2-4) – czy są to różnice w stosunku do tego wyrazu? Współczynniki przy błędach prognoz podniesionych do trzeciej potęgi Autor przedstawia jako 0.00 (***) – rozumiem, że jest to kwestia zaokrąglenia – jednak w takiej sytuacji należałoby przedstawić je w notacji naukowej (np. $5 \cdot 10^{-5}$). Sugerowałabym, aby pod tabelą umieścić postać modelu i napisać wyraźnie, co jest w niej przedstawione. Biorąc

pod uwagę powyższe zastrzeżenia, sugerowałabym przemyślenie badania i wykonanie go od początku – a jeśli jest ono wykonane poprawnie, to jego porządne opisanie.

Rozprawę zamyka **rozdział ósmy** pt. *Are the European Commission's forecasts of public finances better than those of national governments?* będący przedrukiem artykułu opublikowanego w 2020 roku w czasopiśmie *Central European Economic Journal*, w wydaniu specjalnym: "Perspectives on Fiscal Policy". Celem tego rozdziału było zweryfikowanie, czy prognozy wielkości makroekonomicznych publikowane przez Komisję Europejską są trafniejsze, niż prognozy publikowane przez poszczególne agencje rządowe, ze względu na neutralność tego organu i jego niezależność od lokalnej polityki. Autor nie określił, z jaką hipotezą koresponduje ten rozdział i w jaki sposób ten artykuł przybliży go do realizacji celu pracy. Zakres czasowy badania obejmuje lata 2014-2019. Próba badawcza składa się z krajów EU-27 (z pominięciem Grecji) i Wielkiej Brytanii. Źródłem danych są aneksy do artykułów z serii: *Stability of Convergence Programmes: An Overview and Assessment of the Euro Area Fiscal Stance* z lat 2014 – 2019. Autor przeprowadził badanie panelowe z efektami stałymi, w którym zmienną objaśnianą był kwadrat różnicy między błędami prognoz dla danego wskaźnika, publikowanymi przez dwie instytucje. Na jego podstawie wnioskuje, że prognozy Komisji Europejskiej stawiane na bieżący rok charakteryzują się taką samą trafnością, jak prognozy poszczególnych agencji rządowych. W przypadku prognoz salda budżetu państwa na okres dłuższy niż rok, agencje krajowe osiągały większą trafność, niż Komisja Europejska. Autor zauważa też istotne różnice w jakości prognoz poszczególnych krajów – prognozy Komisji Europejskiej są mniej trafne w przypadku małych gospodarek oraz gospodarek silnie zadłużonych (Hiszpania, Portugalia).

Badanie przedstawione przez Autora jest interesujące, jego cel jest poprawnie i jasno zdefiniowany. Brakuje w nim natomiast uzasadnienia wybranej procedury testowania. Autor twierdzi, że liczba obserwacji dla poszczególnych krajów jest za mała, by można było zastosować naturalny w tym przypadku test Diebolda-Mariano (1995). Dlaczego zatem nie został zastosowany test Morgana-Grangera-Diebolda, który wykazuje bardzo dobre własności dla krótkiej próby (por. Harvey, Leybourne, Newbold, 1997) – przynajmniej jako test stabilności wyników? Dlaczego Autor wybrał taką, a nie inną specyfikację modelu panelowego? Czy przeprowadzony został test Hausmanna w celu wykluczenia modelu panelowego z efektami losowymi (*random effect*)? Czy Autor zweryfikował możliwość zastosowania „zwykłej” regresji, bez efektów stałych (tzw. *pooled regression*)? Na stronie 104 w sekcji 4. Methodology, Autor pisze: "(...) if test is performed on pooled data, one makes a very strong assumption that forecast variance is identical for each country". Czy Autor

przeprowadził test Breuscha-Pagana, żeby wykluczyć na pewno to założenie? Czy wykonane zostały testy diagnostyczne (homoskedastyczność składnika losowego?) W jaki sposób Autor interpretuje (ekonomicznie) parametry: μ_i oraz θ_i ? Efekty stałe okazały się statystycznie istotne, a efekty czasowe nie – dlaczego Autor nie pominął ich w takim razie w modelu? Jaki jest sens umieszczania nieistotnych zmiennych, które zmniejszają tym samym znacznie liczbę stopni swobody? Podsumowując – o ile zagadnienie poruszane w tym artykule jest bardzo ciekawe i warte badania, to z metodycznego punktu widzenia zbyt wiele jest w nim przemilczanych założeń. Być może spowodowane to było narzuconym limitem stron w artykule, niemniej nie stało na przeszkodzie, aby testy diagnostyczne i wspomnianą wyżej dyskusję dołączyć do doktoratu w odpowiednim podrozdziale.

4. Ocena pracy od strony formalnej i redakcyjnej

Stronę formalną i redakcyjną pracy oceniam bardzo nisko. Wprawdzie Autor nazywa we wstępie swoje artykuły rozdziałami, to jednak nie pełnią one takiej funkcji. Strony wstępu i zakończenia nie zostały ponumerowane, co jest tym bardziej uciążliwe, że praca rozpada się w rękach na zbiór luźnych kartek. Brak jest również numeracji stron w rozdziałach: trzecim, czwartym i siódmym. Edycja wzorów jest niestaranna, Autor nagminnie używa symbolu „*” na oznaczenie mnożenia, mimo że w edytorze równań, z którego wydaje się korzystać, jest odpowiedni symbol: \cdot . Rysunki z rozdziału czwartego są nieczytelne (nie widać krawędzi, a polskie znaki na wykresach nie są wyświetlane).

Rozprawa składa się z przedruku ośmiu artykułów, do których Autor dopisał krótki wstęp i zakończenie. Nie wydaje mi się to pomysłem trafionym. W moim odczuciu odbiór pracy dużo bardziej poprawiłby i ułatwił porządną autoreferat, do którego, jako załącznik, Autor mógłby dodać opublikowane prace. Jeśli jednak Autor obstawałby przy formie książkowej, to aby rozprawa sprawiała wrażenie jednej całości, powinien poprzedzić każdą część i każdy rozdział stosownym wstępem opisującym cel przedstawianego badania i jego wkład w rozwiązanie problemu badawczego. Idealnie byłoby, gdyby artykuły były napisane tą samą czcionką i ponumerowane.

Sam wstęp i zakończenie pracy zostały napisane bardzo niestarannie od strony językowej, tak że czyta się je ciężko. Brakuje licznych przedimków określonych i nieokreślonych, występują błędy gramatyczne, interpunkcyjne i literówki. Przykłady:

- str. 1¹: jest: *hypothesis*, powinno być: *hypotheses*;
- str. 2¹: Hipoteza 1: użycie rodzajnika *the* przed *economies*, które jeszcze nie zostały zdefiniowane

- str. 2⁶: jest: *was results*, powinno być (zapewne) *resulted from*?
- str. 2⁷: jest: *analyses*, a linijka niżej: *I analyze* – brak konsekwencji w stosowaniu pisowni amerykańskiej i brytyjskiej;
- str. 2₁₃: jest: *are often expressing*, powinno być: *often express*;
- str. 2₃: jest *foundlimited* powinno być: *found limited*.

Nie widzę przeciwwskazań, żeby autoreferat, a nawet wstęp i zakończenie rozprawy składającej się z cyklu artykułów nie mogło być napisane w języku ojczystym Autora. Jeśli jednak z jakichś względów chciałby je napisać w języku angielskim, to wskazane byłoby oddanie tekstu do korekty językowej.

Oprócz błędów językowych, w treści wstępu pojawiają się błędy formalne. Przykładowo, pisząc o badaniu przedstawionym w rozdziale siódmym, powołuje się na hipotezę 3, gdy w rzeczywistości jest to hipoteza 5.

Rozdziały pracy napisane są bardzo nierówno. Widać wyraźnie, że artykuły opublikowane przeszły przez proces korekty (autorskiej i edytorskiej). Najgorzej pod względem edytorskim wypada rozdział siódmy, który nie został nigdzie opublikowany oraz rozdział trzeci, który został zaakceptowany do druku, ale widocznie w momencie składania rozprawy, nie przeszedł jeszcze korekty edytorskiej.

5. Uwagi krytyczne dotyczące celu pracy

Autor podzielił rozprawę na dwie części. Celem pierwszej z nich jest *pokazanie, że działaniom banków centralnych towarzyszyły niewielkie korzyści i duże koszty społeczne* natomiast drugiej: *pokazanie, że prognozy analityków komercyjnych dotyczące dynamiki wzrostu PKB oraz stanu finansów publicznych zawierają liczne mankamenty*. Pomijając już fakt, że sama struktura pracy nie oddaje tego podziału, to nigdzie nie jest wyjaśnione, jak te części się ze sobą łączą w jedną całość? Dlaczego Autor postanowił w swojej pracy realizować te dwa cele, a nie skoncentrować się na jednym z nich? Co jest ich częścią wspólną?

Trudno oprzeć się wrażeniu, że cel rozprawy został sformułowany *ex post*, już po opublikowaniu cyklu artykułów. Samo sformułowanie celu (celów) jest nieprecyzyjne. Cel przyświecający pierwszej części cyklu artykułów został sprecyzowany jako: *pokazanie, że działaniom banków centralnych towarzyszyły niewielkie korzyści i duże koszty społeczne*. Przy takiej konstrukcji zdania czytelnik nie wie, o jakie banki centralne chodzi: wszystkie europejskie? Wszystkie na świecie? W gospodarkach wschodzących? O jakiego typu działania banków chodzi? Co to za korzyści? Z czyjego punktu widzenia? Gospodarki? Inwestorów? Nie wiadomo też, jakiego okresu ma dotyczyć analiza: czy chodzi o cały okres istnienia banków?

O jakiś specyficzny wycinek czasu? Hipotezy, które zostały sformułowane w kontekście tego celu również nie rozwiewają powyższych wątpliwości.

Hipoteza pierwsza brzmi:

Polityka komunikacyjna banku centralnego ma niski wpływ na oczekiwania odnośnie stóp procentowych w małych gospodarkach otwartych.

W celu pojawiła się liczba mnoga (banki centralne) a w hipotezie pierwszej pojedyncza (bank). O który bank chodzi? O wszystkie badane? Czy o jeden specyficzny? Z przedstawionych przez Autora badań dowiadujemy się, że analizował banki centralne Szwecji, Norwegii, Czech (rozdział pierwszy) i Polski (rozdział drugi). Czy taka próba uprawnia rozciąganie wniosków na wszystkie małe gospodarki otwarte? W autoreferacie albo wstępie rozprawy brakuje mi zwięzłego podsumowania, czy uzasadnienia: Dlaczego najpierw analizowane są Szwecja, Norwegia, Czechy, a potem Polska? Co jest specyficznego w pierwszej grupie państw, a co w Polsce? Czy Polska jest pod pewnymi względami specyficzna, czy wyróżniająca na tle innych krajów, czy po prostu chodziło o dostępność danych?

Druga hipoteza brzmi:

Utrzymanie łagodnej polityki pieniężnej w strefie euro spowodowało pojawienie się baniek spekulacyjnych na głównych rynkach nieruchomości.

O ile doprecyzowano zakres geograficzny (strefa euro), to co oznacza sformułowanie: główne rynki nieruchomości? Dodatkowo, okazuje się, że Autor przebadał tylko pięć miast, z czego cztery stołeczne i w określonym przedziale czasowym. W związku z tym pojawia się pytanie, czy jest to wystarczająco duża próba, aby uogólniać uzyskane wnioski?

Ponadto – w autoreferacie (wstępie) brakuje mi uzasadnienia, jak ta hipoteza koresponduje z celem badawczym? Dlaczego w ogóle łagodna polityka pieniężna miałaby przyczyniać się do powstawania baniek spekulacyjnych na rynku nieruchomości? Czy są jakieś podstawy teoretyczne, by tak sądzić, czy raczej podstawą badania były obserwacje empiryczne? Co Autor rozumie przez łagodną politykę pieniężną? Odpowiedzi na te pytania nie znajdujemy w rozdziale trzecim, który koresponduje z tą właśnie hipotezą badawczą.

Hipoteza trzecia budzi najmniej wątpliwości, choć nadal nie wiadomo, czy jest to stwierdzenie ogólne, czy odnosi się do konkretnej grupy banków w konkretnym czasie:

Pojawienie się niepożądanych efektów polityki pieniężnej i brak realizacji celów inflacyjnych banków są wynikiem błędów popełnianych przez departamenty badań ekonomicznych.

Dodatkowo nie wiemy, czym są według Autora niepożądane efekty polityki pieniężnej. Spójnik „i” sugeruje, że Autor ma na myśli coś jeszcze, poza brakiem realizacji celów inflacyjnych. W jaki sposób ta hipoteza przybliży Autora do realizacji celu pierwszego? W jaki sposób odnosi się do konsekwencji działań banków, korzyści z działań banków, czy kosztów społecznych? Badanie przedstawione w rozdziale czwartym jest interesujące, ale w moim odczuciu zupełnie oderwane od celu pracy.

Hipoteza czwarta brzmi:

Prognozy analityków komercyjnych w Polsce charakteryzują się nieefektywnością, na przykład zawierają systematyczne obciążenia

Wydaje mi się, że skoro pojęcie efektywności jest zdefiniowane w literaturze, a pan magister Rybacki powołuje się na konkretną definicję (Nordhaus, 1987), wystarczyłoby dodać tu autora definicji, zamiast sformułowania: na przykład zawierają systematyczne obciążenia.

Z hipotezą czwartą korespondują dwa rozdziały: piąty i szósty. Są to najbardziej spójne rozdziały z całej pracy. Oba koncentrują się na jednym zagadnieniu i jednym obszarze geograficznym. Natomiast zupełnie niekonsekwentnie, tę część rozprawy Autor rozpoczyna analizą dla Polski, a następnie rozszerza próbę. Wydaje mi się, że lepiej byłoby, wzorem części pierwszej – na początku pokazać szerszą perspektywę, a dopiero potem zawęzić ją do rynku polskiego, podkreślając jego specyfikę – albo wykazując analogie z innymi gospodarkami.

Pewne wątpliwości budzi również **hipoteza piąta:**

Institucje międzynarodowe stosują podwójne standardy w ocenie finansów publicznych krajów UE- szacunki dla niektórych krajów mogą mieć systematyczne błędy

Sugerowałabym doprecyzowanie, że te podwójne standardy i systematyczne błędy są w tym artykule równoważne. Jej weryfikacji poświęcony jest rozdział siódmy. Nie wiadomo natomiast, jaki jest cel rozdziału ósmego. Nie koresponduje on z żadną hipotezą. Nie przybliży też Autora do realizacji celu drugiego pracy, ponieważ nie koncentruje się w nim na prognozach komercyjnych. Swoją drogą, z tego samego powodu hipoteza piąta też nie jest związana z celem drugim.

6. Konkluzja

Zgodnie z zapisem ustawy z dn. 14 marca 2013 roku (art. 13 ust.1) o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (tekst jedn. Dz. U. z 2017, poz. 1789), doktorant w swojej rozprawie doktorskiej powinien wykazać się umiejętnością

samodzielnego prowadzenia badań, wiedzą teoretyczną oraz oryginalnością rozwiązania problemu naukowego. Bezsprzecznie **pan magister Rybacki wykazał się zdolnością do samodzielnego prowadzenia badań**. Świadczy o tym cykl ośmiu artykułów przedstawionych w rozprawie, w przypadku których każdorazowo Doktorant był jedynym autorem. Chciałabym podkreślić, że jest to znaczna liczba publikacji. W każdym z artykułów **wykazał się również wiedzą teoretyczną** na temat badanych przez siebie zagadnień oraz wiedzą z zakresu metod ilościowych. Uwagi krytyczne, zamieszczone w recenzji, nie wpływają na moje przekonanie co do umiejętności prowadzenia badań przez pana magistra Rybackiego, natomiast w głównej mierze dotyczą formy prezentacji jego dokonań, która w wielu przypadkach utrudnia zrozumienie tego, co Autor policzył, albo co chciał przekazać. Chciałabym też zaznaczyć, że bardzo podoba mi się idea tworzenia rozprawy doktorskiej na podstawie artykułów naukowych. Mimo wszystko jednak, w wielu czasopismach (zwłaszcza polskich) występuje wymóg limitu stron, w związku z czym trudno jest w nich zaprezentować pełne badanie, razem z analizą stabilności wyników. Stąd też dużo pytań z mojej strony, odnośnie założeń modeli i ich weryfikacji. Artykuły pana magistra Rybackiego są ciekawe, poruszają wiele interesujących zagadnień z wielu zakresów. Metody ilościowe zastosowane w badaniach są bardzo różnorodne i zazwyczaj odpowiednio dobrane do celu badań.

Niestety, rozprawa przedstawiona w obecnej postaci **uniemożliwia mi ocenę trzeciego aspektu, tj. oryginalności rozwiązania problemu naukowego**, ze względu na to, że problem badawczy nie został w rozprawie jednoznacznie wyszczególniony. W związku z powyższym **wniosuję o uzupełnienie rozprawy o staranny autoreferat**, w którym zostanie jasno przedstawiony problem badawczy oraz precyzyjnie sformułowane cele badawcze lub hipotezy. W autoreferacie Autor powinien przekonująco wyjaśnić, w jaki sposób oba cele badawcze łączą się ze sobą w jeden problem badawczy - o ile zdecyduje się na pozostawienie dwóch celów; a także w jaki sposób poszczególne publikacje prowadziły do wyjaśnienia bądź rozwiązania zdefiniowanego problemu badawczego. Proszę też o wykazanie zidentyfikowanej przez Autora luki badawczej (bądź luk badawczych) i określenie własnego wkładu w rozwój dziedziny. W zależności od postawionego celu, Autor może dojść do wniosku, że nie wszystkie artykuły, składające się na rozprawę w obecnej postaci, muszą być w niej zawarte. W przypadku, gdyby Autor zdecydował się na pisanie autoreferatu w języku angielskim, sugeruję oddanie tekstu do korekty językowej.

Ze swojej strony zachęcam też Autora do zainspirowania się artykułem: Faff, R i in. (2020) *Pitching research for engagement and impact: a simple tool and illustrative examples* <https://doi.org/10.1111/acfi.12704> w celu usystematyzowania celów i hipotez badawczych.

Ostatecznie, jeśli Autor zdecyduje się pozostawić w rozprawie rozdziały 4 i 7, proszę o poprawę opisu wyników (w przypadku rozdziału 7 również opisu modelu).

A. Kliber
Agata Kliber