

Metody Ekonometryczne

Informacje o raporcie zaliczeniowym

Jakub Mućk

1. Kryteria formalne:

Termin oddania: 8 czerwca 2025 r.

Forma raportu zaliczeniowego: raport powinien zostać przesłany mailowo do osoby prowadzącej ćwiczenia. Raport może zostać przygotowany indywidualnie lub w zespołach dwuosobowych.

- W przypadku indywidualnych prac raport powinien zostać przesłany mailem o tytule [SGH] [Metody Ekonometryczne] Raport zaliczeniowy. Imię Nazwisko. Do maila proszę załączyć następujące pliki:

- (a) Plik pdf z raportem,
- (b) Zbiór danych – jeden plik zawierający Państwa zbiór danych w dowolnym formacie,
- (c) Skrypt w pakiecie R, który umożliwi replikację Państwa obliczeń.

Proszę aby powyższe pliki miały nazwę `Imię.Nazwisko.NumerIndeksu.Rozszerzenie`, gdzie `Rozszerzenie` to format odpowiedniego pliku. W nazwach plików proszę nie korzystać z polskich znaków.

- W przypadku prac przygotowywanych w zespołach dwuosobowych raport powinien zostać przesłany mailem o tytule [SGH] [Metody Ekonometryczne] Raport zaliczeniowy. Imię Nazwisko1 oraz Imię Nazwisko2. Do maila proszę załączyć następujące pliki:

- (a) Plik pdf z raportem,
- (b) Zbiór danych – jeden plik zawierający Państwa zbiór danych w dowolnym formacie,
- (c) Skrypt w pakiecie R, który umożliwi replikację Państwa obliczeń.

Proszę aby powyższe pliki miały nazwę `Imię.Nazwisko1.NumerIndeksu1._Imię.Nazwisko2.NumerIndeksu2. Rozszerzenie`, gdzie `Rozszerzenie` to format odpowiedniego pliku. W nazwach plików proszę nie korzystać z polskich znaków.

Proszę pamiętać, że przesłanie skryptu w pakiecie R z odpowiednimi danymi powinno mi umożliwić replikację opisywanych przez Państwa wyników.

Proszę również przesłać pracę z uczelnianego adresu mailowego. W przeciwnym przypadku praca nie będzie sprawdzana.

- ### 2. Cel i zakres merytoryczny raportu zaliczeniowego.
- Celem raportu jest sprawdzenie nabytych przez Państwa umiejętności prowadzenia analiz ekonometrycznych w badaniach empirycznych. Oprócz adekwatności stosowanych metod ważne są (i) opis problemu ekonomicznego, oraz (ii) motywacja wybranego problemu badawczego.

Nieakceptowalne są modele ekonometryczne, których analiza jest pozbawiona sensu ekonomicznego, np.: tożsamości (np. $Y = I + C + G$), tłumaczenie PKB populacją, itp.

Proszę również pamiętać o odpowiednio (dużej) liczbie próby, która pozwala na wnioskowanie statystyczne.

3. Struktura raportu zaliczeniowego

- (a) **Wprowadzenie zawierające cel analizy ekonometrycznej oraz motywację.** W tej części powinien zostać opisany cel analizy ekonometrycznej. W szczególności powinno zostać opisane, na jakie pytanie badawcze pozwolą odpowiedzieć wyniki Państwa analizy. Jest to również miejsce na opis zjawiska ekonomicznego, które jest przedmiotem Państwa pracy. Wprowadzenie można rozszerzyć o przegląd literatury. Przegląd literatury nie jest obligatoryjny, ale pozwoli Państwu dobrze umotywić temat badawczy poruszany w Państwa raporcie. W szczególności, dodatkowo będzie punktowane odwołanie się do prac opublikowanych w tzw. *Top Five*, a więc *American Economic Review (AER)*, *Journal of Political Economy (JPE)*, *Quarterly Journal of Economics (QJE)* czy *the Review of Economic Studies*. W przeglądzie literatury pomocne mogą się okazać bazy *Google Scholars* czy *IDEAS/RePEc*.
- (b) **Opis zbioru danych oraz wizualizacja dostępnych danych.** W tej części powinny zostać opisane zmienne ekonomiczne wykorzystane w modelowaniu ekonometrycznym oraz źródła danych. Proszę również umieścić informację o transformacjach oryginalnych danych oraz motywacji stojącej za ich przeprowadzeniem. Następnie, można przedstawić wyniki analizy danych, a więc np. analizę korelacyjną, rozkłady częstości czy rozkłady częstości warunkowych. Proszę zwrócić uwagę, że wizualizacja danych silnie zależy od charakterystyki Państwa zbioru danych.
- (c) **Zapis modelu ekonometrycznego dla analizowanego zjawiska/procesu ekonomicznego.** Zapis równania dla modelu ekonometrycznego, który opisuje Państwa zjawisko. W tym miejscu można sformułować swoje oczekiwania odnośnie siły i kierunku zależności czynników wpływających/determinant dla Państwa zmiennej objaśniającej. Jest to również naturalne miejsce na odwołanie się do dotychczasowej literatury, tj. czy to wyników teoretycznych badań czy udokumentowanych prawidłowości empirycznych.
- (d) **Analiza ekonometryczna.** W tej części opisują Państwo wyniki modelowania ekonometrycznego oraz badanie własności statystycznych rozważanego modelu ekonometrycznego.
- Absolutnym minimum są: (i) oszacowania parametrów uzyskane MNK i ich interpretacja, (ii) analiza dopasowania wartości teoretycznych do obserwacji empirycznych, współliniowości zmiennych objaśniających oraz postaci funkcyjnej modelu (np. test RESET dostępny w pakiecie `lmtest`) (iii) badanie własności składnika losowego (normalność rozkładu, autokorelacja lub heteroskedastyczność), (iv) zastosowanie metod uwzględniających niesferyczność składnika losowego, tj. (odpowiedni) odporny estymator macierzy wariancji-kowariancji oraz estymator uogólnionej metody najmniejszych kwadratów (ważona MNK w przypadku danych przekrojowych oraz metod Cochrane'a-Orcutt/ Prais-Winstena w przypadku szeregów czasowych).
 - Proszę pamiętać o **adekwatności wykorzystanych metod/testów**. Przykładowo, zupełnie inne własności składnika losowego powinny być badane w przypadku danych przekrojowych a inne w przypadku analiz opartych na szeregach czasowych.
 - Dodatkowym atutem będzie aplikacja:
 - Metody Zmiennych Instrumentalnych – wraz z dyskusją nt. endogeniczności, wyboru instrumentu/ów oraz odpowiednimi testami statystycznymi pozwalającymi ocenić problem endogeniczności oraz zasadność wykorzystanego instrumentu/ów.
 - Modeli wielorównaniowych – wraz z dyskusją nt. identyfikowalności parametrów strukturalnych.
 - W przypadku, gdy w literaturze wiodącą strategią jest stosowanie metod, które nie były przedmiotem zajęć, można wtedy dodatkowo zastosować te metody. Warto wtedy skonsultować ten wybór.
- (e) **Podsumowanie** raportu powinno zawierać najważniejsze wnioski płynące z Państwa analizy. Naturalnym punktem jest również dyskusja nad otrzymanymi wynikami, ich wiarygodnością oraz potencjalnymi rozszerzeniami Państwa badania.

- (f) **Literatura.** Spis pozycji bibliograficznych, z których Państwo korzystaliście. Proszę pamiętać, że w spisie literatury powinny się znaleźć pozycje, do których Państwo odwoływaliście się w pracy.

4. Wskazówki:

- (a) **Estetyka prezentacji wyników.** Raport nie może być ciągiem zrzutów z ekranu konsoli programu R. Dlatego, warto się zastawować nad prezentacją wyników. Przykładowo:
- i. Oszacowania parametrów uzyskane kilkoma metodami/ dla kilku specyfikacji można zestawić w jednej tabeli. Można tutaj wykorzystać np. pakiet `stargazer`.
 - ii. Wyniki szeregu testów statystycznych można zestawić również w jednej tabeli. Z drugiej strony, szczegółowe wyniki testów statystycznych, które pozwalają zrozumieć istotę analizowanego problemu, można umieścić w załączniku. Przykładowo, pomocne mogą się tutaj okazać oszacowania regresji pomocniczej w teście na heteroskedastyczność składnika losowego White'a.
 - iii. W przypadku analizy niesferyczności składnika losowego warto posłużyć się odpowiednimi wykresami, które pozwolą Państwu wytłumaczyć istotę tego problemu w Państwa modelu.
- (b) **Przejrzystość i spójność wyводу.** Proponowana powyżej struktura nie ma charakteru obligatoryjnego, a jedynie jest wskazówką dla Państwa. W większości przypadków jej wykorzystanie pozwoli zbudować ramy dla logicznego wyводу. Niemniej proszę pamiętać również o samej treści raportu.
- (c) Szeregi czasowe. Zgodnie z wiedzą z podstawowego kursu ekonometrii proszę pamiętać o ryzyku regresji pozornej (*spurious regression*). Zalecaną strategią jest tutaj przeprowadzenie odpowiedniej transformacji zmiennej (najczęściej różnicowanie), która pozwoli uzyskać stacjonarną postać zmiennej ekonomicznej.
- (d) **Źródła danych**
- i. Proces kolekcjonowania danych jest jednym z etapów badań. Dlatego, **nie będą akceptowalne prace, które bazują na wykorzystaniu wbudowanych w pakiety R gotowych zbiorów danych.** Proszę również **nie korzystać ze zbiorów danych dostępnych danych na ogólnych repozytoriach**, jak np. Kaggle czy UCL.
 - ii. **Przykładowe źródła danych:**
 - Dane makroekonomiczne: Eurostat [\[link\]](#), OECD [\[link\]](#), Penn World Table [\[link\]](#), FRED [Federal Reserve Economic Data; [link](#)], IMF [\[link\]](#), AMECO [\[link\]](#), WDI [\[link\]](#), GUS [\[link\]](#).
 - Dane mikro: European Social Survey [\[link\]](#), Current Population Survey [\[link\]](#), Diagnoza Społeczna [\[link\]](#)
 - iii. Dane z różnych źródeł danych można łączyć. Można również prowadzić odpowiednie zmienne zero-jedynkowe (np. przynależność kraju do OECD, zmienna binarna indukująca kryzys wg NBER, itp.)
 - iv. W niektórych przypadkach dostęp do powyższych zbiorów danych jest również możliwy dzięki odpowiednim pakietom (np. `eurostat` czy `OECD`) w programie R przez odpowiedni interfejs programistyczny (API). Jednak efektywne korzystanie z tego sposobu korzystania ze źródeł danych i tak wymaga wiedzy na temat struktury dostępnych danych/ kodowania zmiennych. Dlatego, korzystanie z tych pakietów może okazać się bardziej czasochłonne na początku niż *ręczne* ściąganie danych ze stron internetowych.
 - v. W szczególnych przypadkach można korzystać ze zbiorów danych udostępnianych przez autorów badań. Prosiłbym w takim przypadku o konsultacje.
- (e) **Odpowiedni język i styl.** Proszę pamiętać, że brak zrozumienia istoty wykorzystywanych metod może odzwierciedlać język raportu. Proszę również unikać skrótów myślowych. Przykłady niefrasobliwych uproszczeń:

- i. *Hipoteza zerowa jest prawdziwa* \implies niezrozumienie istoty wnioskowania statystycznego. Można mieć podstawy do odrzucenia hipotezy zerowej (lub takich podstaw nie mieć) przy pewnym poziomie istotności.
- ii. *Oszacowania modelu* \implies oszacowania parametrów strukturalnych modelu?
- iii. *Model jest heteroskedastyczny* \implies składnik losowy może być heteroskedastyczny, a nie model.
- iv. *Model wyjaśnia 90% zmienności PKB* \implies (i) jakiej zmienności? (ii) a co w przypadku endogeniczności? Czy jest to może charakterystyka danych?