

Głównym celem naukowym projektu jest rozwój i popularyzacja narzędzi używanych do analizy danych panelowych - danych, gdzie obserwowanych jest kilka podmiotów (np. państw) na przestrzeni czasu (np. lat). Obserwowane mogą być jednocześnie różne cechy podmiotów - np. dla państw może to być PKB, poziom bezrobocia lub poziom inwestycji w danym roku. Są dwie główne kwestie, które analiza danych panelowych musi zaadresować. Pierwszą z nich jest niepewność wyboru modelu, tzn. które cechy powinniśmy włączyć do naszej analizy. Drugą z nich jest endogeniczność, tzn. jakie są relacje między cechami, które włączyliśmy do analizy. Wyniki tego projektu znacząco ułatwią adresowanie obu tych kwestii.

Droga do osiągnięcia celu projektu składa się z dwóch równoległych ścieżek:

1. Rozwój metodologii i teorii, które są drogowskazem, jak podejść do analizy danych panelowych w konkretnych sytuacjach.
2. Utworzenie łatwego w obsłudze i szeroko dostępnego oprogramowania, które pozwoli z łatwością przeprowadzać analizę danych panelowych w praktyce.

Stąd głównym produktem projektu będzie darmowy pakiet w środowisku R, zawierający jako swoje podstawy rozwiniętą metodologię.

Pierwsza kwestia, niepewność wyboru modelu, będzie zaadresowana przy pomocy Bayesowskiego Wążenia Modeli. Metodologia ta pozwala na przeprowadzenie statystycznej analizy, w wyniku której dostajemy informację o tym, jak istotna - średnio - jest wybrana cecha podmiotu podczas przewidywania innej cechy, która nas interesuje. Na przykład, jesteśmy w stanie stwierdzić, jak istotny jest poziom bezrobocia lub inwestycji do przewidywania PKB kraju w przyszłości.

Druga kwestia to endogeniczność, czyli odpowiedź na pytanie jakie relacje zachodzą między cechami, które zdecydowaliśmy się umieścić w modelu. Do osiągnięcia tego celu musi zostać utworzona odpowiednia funkcja wiarygodności. Funkcja wiarygodności mówi nam o tym, jak prawdopodobne jest, że określone wartości parametrów wygenerowały dane, które zaobserwowaliśmy. Wybór funkcji wiarygodności decyduje o wynikach całej analizy, a także determinuje, w jakich sytuacjach możemy stosować daną metodologię. Stąd, jednym z rezultatów tego projektu będzie wygenerowanie jak najbardziej ogólnej funkcji wiarygodności, która jednocześnie pozwala na uzyskanie wartościowych wniosków na temat obserwowanych cech.

Obie kwestie są znane w środowisku akademickim i były nieraz poruszane przez wielu znakomitych naukowców takich jak Enrique Moral-Benito czy Adrian E. Raftery. Celem tego projektu jest budowa na tych mocnych fundamentach. Jednak zaaplikowanie istniejących metodologii było dotychczas co najmniej olbrzymim wyzwaniem dla ogółu środowiska akademickiego. Wierzmy, że następujące czynniki są za to odpowiedzialne. Po pierwsze, istniejąca teoria musi zostać ulepszona i rozszerzona tak, aby było możliwe przeprowadzanie analizy w wielu różnych sytuacjach. Po drugie, narzędzia i oprogramowanie służące do zaaplikowania metodologii w praktyce muszą być łatwo dostępne dla wszystkich naukowców.

Wszystkie powyższe kwestie zostaną zaadresowane przez główny produkt tego projektu - przyjazny dla użytkownika pakiet R zawierający w sobie rozwiniętą metodologię. Język programowania R jest jednym z najczęściej używanych narzędzi do analizy podczas badań naukowych. R jest oprogramowaniem zupełnie darmowym, a jednocześnie niesamowicie rozwiniętym. Wybór tego środowiska zapewni dostępność utworzonych w ramach projektu narzędzi. Rozwinięcie metodologii razem z szczegółowymi instrukcjami i filmami instruktażowymi dla pakietu ułatwi użytkowanie, a tym samym wspomże dalszą popularyzację i rozwój narzędzi. Stąd przeprowadzenie tego projektu będzie bardzo istotnym osiągnięciem z perspektywy analizy danych panelowych i badań naukowych.