

М.С. ДУНАЄВСЬКИЙІнститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна,
e-mail: *MaxDunaievskiy@gmail.com*.

М-ОЦІНКИ: ОГЛЯД ТА ЗАСТОСУВАННЯ ДЛЯ ПРИЙНЯТТЯ РІШЕНЬ ЗА УМОВ НЕВИЗНАЧЕНОСТІ¹

Анотація. Розглянуто еволюцію методу робастних оцінок, зокрема основні концепції методів М-оцінок та їхні переваги під час роботи з даними, що містять статистичні викиди. Проведено порівняння методів М-оцінок Хубера, Хампела та методу біквадратних ваг Тьюкі на реальних економічних даних регіонів України. Отримано емпіричне підтвердження ефективності їхнього застосування. Результати порівняння всіх побудованих моделей свідчать про суттєве зміщення оцінок параметрів у разі застосування підходу «наївного» видалення статистичних викидів. На основі побудованих робастних моделей можна з більшою точністю порівняти регіони країни за ефективністю факторів праці та капіталу та відповідно визначити пріоритетність інвестицій, які більшою мірою стимулюватимуть економічний розвиток регіону.

Ключові слова: децентралізація, прийняття рішень, М-оцінки, регіональний розвиток, робастність, невизначеність.

ВСТУП

М-оцінки та їхнє застосування є досить ефективним аналітичним методом, який науковці використовують у своїх емпіричних дослідженнях. Цей метод дає змогу отримати більш точні та стійкі оцінки ключових параметрів моделі навіть за наявності статистичних викидів (outliers) та спостережень з аномальними значеннями, які значно зміщують оцінки, отримані традиційним методом найменших квадратів (МНК) (ordinary least squares). Звичним підходом аналітиків-практиків є видалення спостережень, значення яких є статистичними викидами. Проте цей підхід не завжди є виправданим. По-перше, внаслідок видалення частини спостережень втрачається частина реального сигналу від даних, а відповідно погіршується якість побудованих моделей. По-друге, сама кількість спостережень, доступна досліднику для аналізу, може бути досить обмеженою. Наприклад, можуть бути наявними лише чотири щоквартальних спостереження протягом року для кожного з близько десяти адміністративних районів. Для статистичної значущості побудованих моделей потрібно щонайменше 30 точок даних. Відповідно, у зазначеній ситуації потрібно застосувати кращий метод, ніж «наївне» видалення викидів. На сьогодні розроблено методи оцінювання, які мають майже таку саму ефективність, як МНК (на очищених від статистичних викидів даних) і водночас проявляють високу стійкість до викидів у разі їхнього застосування на неочищених наборах даних. У цій статті наведено короткий огляд еволюції методу робастних оцінок та проведено аналіз на реальних економічних даних регіонів України з порівнянням результатів застосування методів М-оцінок Хубера [1], Хампела [2], та методу біквадратних ваг Тьюкі. Для аналізу та моделювання використано мову програмування R та середовище розробки RStudio [3].

ОГЛЯД МЕТОДІВ М-ОЦІНОК

Загалом, є декілька підходів до знаходження статистично стійких значень параметрів моделей, однак найчастіше використовують саме метод зважених

¹ Роботу виконано за часткової підтримки Національного фонду досліджень України. Грант № 2020.02/0121.