

Відбір екзогенних змінних моделей засобами інтелектуального аналізу даних

Олександра Белз

Кафедра інформаційних систем у менеджменті, Львівський національний
університет імені Івана Франка, УКРАЇНА, м. Львів, проспект Свободи, 18,
E-mail: homepc@list.ru

Abstract – The problem of influence on the predictive characteristics of models of methods of selection of economic indicators that are included in the econometric model as explanatory variables was examined.

Ключові слова – ідентифікація, моделювання, статистичні моделі, екзогенні змінні, інтелектуальний аналіз даних.

I. Вступ

На якість побудованих статистичних моделей вагомий вплив має вибір набору пояснювальних (екзогенних) змінних. Ю. Толбатов зазначає, що суттєвість впливу факторів на показник можна визначити з використанням F -статистики [3]. В. Т. Доля вказує, що “рішення щодо включення незалежних змінних (факторів) у рівняння регресії приймається за наслідками тестування на значущість (не випадковість) їхнього впливу на залежну змінну за t -статистикою Стьюдента, автономність впливу за γ -тестом і достатність вкладу у множинний коефіцієнт детермінації за ρ -тестом” [1].

Метою роботи є дослідження впливу способів відбору економічних показників, які входять до економетричної моделі як пояснювальні змінні, на прогнозні характеристики таких моделей.

II. Основні результати дослідження

У загальному випадку підхід щодо визначення суттєвості впливу на результатний показник окремих факторів з використанням F -статистики вимагає побудови ряду економетричних моделей на підставі повного перебору всеможливих кандидатів у екзогенні змінні [3]. Прийняття рішень щодо включення незалежних змінних у рівняння регресії на підставі ρ -тесту також вимагає побудови ряду економетричних моделей на підставі повного перебору всеможливих кандидатів у екзогенні змінні [1]. Тому такі підходи є неприйнятні для застосування, коли потрібно оперувати із великими масивами незалежних змінних.

В основі відбору суттєвих факторів за t -статистикою Стьюдента і γ -тестом лежить розрахунок парних коефіцієнтів кореляції [1]. Такий підхід не вимагає значних обчислювальних ресурсів і може бути застосований у практичних задачах. Тому у роботі для ідентифікації

моделей соціально-економічних процесів методами статистичного аналізу даних введення пояснювальних змінних будемо здійснювати поступово по мірі спадання значень парних коефіцієнтів кореляції.

Для ідентифікації моделей соціально-економічних процесів методами інтелектуального аналізу даних обрано алгоритми афінитивного аналізу – комбінацію алгоритму Кохонена та алгоритму G-means, що реалізована в аналітичній платформі Deductor Academic [2]. Введення пояснювальних змінних будемо здійснювати поступово по мірі спадання рівня значимості показників. Для визначення рівня значимості показників виконано кластеризацію станів макроекономічного середовища.

Вхідними даними для дослідження є помісячні значення чотирнадцяти макроекономічних показників з січня 2006 р. по грудень 2013 р. З метою дослідження прогнозних характеристик побудованих моделей із статистичної вибірки виокремлено дві частини (навчальну та тестову). На підставі даних навчальної вибірки розраховують коефіцієнти моделей, а на підставі даних тестової вибірки оцінюють якість прогнозування.

ВИСНОВОК

Після проведення необхідних розрахунків виявлено, що якість побудованих моделей на підставі даних навчальної вибірки по мірі введення у модель кожної наступної пояснювальної змінної покращується незалежно від того, чи були застосовані методи статистичного аналізу даних чи методи інтелектуального аналізу даних. Для значення критерію оцінки якості побудованих моделей на підставі даних тестової вибірки така тенденція відсутня. Окрім того, висока якість моделі на підставі даних навчальної вибірки не гарантує хороших прогнозних характеристик такої моделі.

Також зазначимо, що застосування методів статистичного та інтелектуального аналізу даних у всіх випадках покращує прогнозні характеристики моделей. У більшості випадків краща модель досягається на підставі методів інтелектуального аналізу даних.

Література

1. В. Т. Доля. Економетрія: навч. посібник / В. Т. Доля; Харк. нац. акад. міськ. госп-ва. – Х.: ХНАМГ, 2010. – 171 с.
2. Паклин Н. Б. Бизнес-аналитика: от данных к знаниям (+CD): Учеб. пособие / Н. Б. Паклин, В. И. Орешков. – [2-е изд., перераб. и доп.]. – СПб.: Питер, 2010. – 704 с.
3. Толбатов Ю. А. Економетрика: Підручник для студентів екон. спеціальн. вищ. навч. закл / Ю. А. Толбатов. – К.: Четверта хвиля, 1997. – 320 с.