

ОПРЕДЕЛЕНИЕ ОПТИМАЛЬНОЙ СТРУКТУРЫ РЕГРЕССИОННОЙ МОДЕЛИ

Решетняк Ю.Б., Стрижаченко А.В.

Национальный фармацевтический университет, г. Харьков

Известно, что включение в эконометрическую модель дополнительных факторов (объясняющих переменных) всегда приводит к росту коэффициента детерминации R^2 и уменьшению суммы квадратов остатков модели RSS . Таким образом, происходит «улучшение» модели. Однако, лишние факторы снижают эффективность оценок параметров модели, и приводят к расширению доверительных интервалов их оценок.

Обычно целесообразность добавления группы новых регрессоров определяется с помощью одного из вариантов теста Вальда. Сравняются длинная и короткая регрессии (длинная содержит дополнительные объясняющие переменные). Нулевая гипотеза предполагает одновременное равенство нулю коэффициентов при дополнительных переменных.

F -статистика выглядит следующим образом:

$$F = \frac{\frac{RSS_k - RSS_m}{m - k}}{\frac{RSS_m}{n - m}},$$

где k - количество регрессоров в короткой регрессии, m - количество регрессоров в длинной регрессии, n - количество наблюдений. Если расчётное значение F -статистики больше, чем $F_{\alpha} = F_{1-\alpha}(m - k, n - m)$, где α - уровень значимости, то нулевая гипотеза отвергается.

Другой способ решения задачи оптимизации количества регрессоров состоит в использовании штрафных критериев Акаике AIC и Шварца BIC . $AIC = n \ln(RSS/n) + 2p$; $BIC = n \ln(RSS/n) + \ln(n)p$, где n - количество наблюдений, p - число параметров модели, RSS - сумма квадратов остатков модели, полученная при оценке коэффициентов методом наименьших квадратов. Среди нескольких альтернативных моделей предпочтение отдается той, у которой значение AIC и BIC меньше.

В рамках алгоритма, связанного с рассмотрением статистических гипотез, возможно только попарное сравнение моделей с разным числом факторов. Таким образом, необходим анализ $\sim m^2$ пар гипотез (m - максимальное количество факторов). При использовании штрафных критериев каждому набору объясняющих переменных отвечает одно значение критерия. Следовательно, при анализе приходится иметь дело с относительно небольшим числом (m) значений применяемого критерия. В тоже время такой подход является слишком упрощенным и недостаточно обоснованным.

Представляется, что штрафные критерии можно применять на начальном этапе построения регрессионной модели при первичном определении ее структуры.