Автокорреляция как следствие неправильной спецификации модели

Автокорреляция в модели регрессии формально вызывается зависимостью между значениями случайной составляющей в выборке. Но этот вопрос можно рассмотреть и более глубоко. Причиной наличия случайной составляющей может быть какая-либо неточность в спецификации модели, например пропуск какой-либо важной объясняющей переменной или использование неподходящей математической функции (см. раздел 2.1). Следовательно, автокорреляция нередко может объясняться неправильной спецификацией модели; в этом случае, по-видимому, лучше вместо использования механической процедуры «исправления» непосредственно попытаться устранить ошибки в спецификации. Конечно, обычно лучше устранять причину, чем симптом.

Автокорреляция, вызванная неправильной спецификацией переменных

Явная автокорреляция может быть вызвана пропуском важной объясняющей переменной, и положение можно исправить, если эта переменная будет определена и включена. (Пример дан в упражнении 10.4.) Другая ее причина может заключаться в том, что не принята во внимание структура модели, включающая запаздывание. Метод Кокрана—Оркатта является эффективным способом отражения структуры запаздывания в модели, которая ранее была статической. Возможно, будет признана предпочтительной более общая спецификация. Мы видели, что при наличии автокорреляции в модели (7.21) ее можно устранить в случае парной регрессии путем преобразования модели к виду (7.27). Это можно переписать таким образом:

$$y_t = \alpha(1-\rho) + \rho y_{t-1} + \beta x_t - \beta\rho x_{t-1} + \varepsilon_t \qquad (7.34)$$

Фактически мы оцениваем регрессионную зависимость y_t от y_{t-1}, x_t, x_{t-1}, налагая ограничение, заключающееся в требовании равенства коэффициента при x_{t-1}, произведению коэффициентов при других двух переменных в правой части уравнения. Так как уравнение является нелинейным по параметрам, мы не можем для его оценивания использовать МНК. Вместо этого мы применяем метод Кокрана-Оркатта или какой-либо другой подобный ему метод оценивания, в сущности, нелинейной регрессии.

В целом мы не имеем права заранее утверждать, что указанное ограничение обосновано. Кроме того, мы должны проверять все ограничения, где это возможно, и в данном случае сделать это несложно. Мы вводим другую, не включающую ограничения модель:

$$y_t = \lambda_0 + \lambda_1 y_{t-1} + \lambda_2 x_t + \lambda_3 \rho x_{t-1} + \varepsilon_t \qquad (7.35)$$

и проверяем, равно ли λ_3 величине $-\lambda_1\lambda_2$. Если это ограничение не отклонено, мы принимаем предположение, что модель адекватно представлена выражениями (7.20) и (7.21), и продолжаем оценивать ее параметры, используя метод Кокрана-Оркатта или другой подобный ему метод. Если ограничение отклонено, то непосредственно оценивается регрессия (7.35) с использованием обычного МНК.

Следует отметить, что если лучшей спецификацией модели окажется (7.35), то из этого следует, что мы отказываемся от гипотезы, что случайный член формируется авторегрессионным процессом (7.21) и тест Дарбина—Уотсона перестает быть применимым при оценивании регрессии (7.20). Тем не менее он может быть полезен в диагностических целях, и часто первым указанием на наличие какой-либо проблемы в исходной регрессии служит d-статистика, недостаточно близкая к двум.

Теоретические положения, обосновывающие рассматриваемую процедуру проверки, здесь не представлены (они кратко излагаются в работе Д.Хендри и Г.Майзона [Hendry, Mizon, 1978]). Для данного случая подходит тестовая статистика

$$T \log (RSS_R/RSS_y), \qquad (7.36)$$

где RSS_R и RSS_y — необъясненные суммы квадратов отклонений соответственно в вариантах с ограничением и без ограничений; логарифмы вычисляются по основанию e и T — количество наблюдений в выборке. В больших выборках статистика, лежащая в основе критерия, имеет распределение χ^2 с числом степеней свободы, равным количеству налагаемых ограничений.

Может возникнуть вопрос о количестве налагаемых ограничений. До сих пор анализировалась исходная модель с одной объясняющей переменной. В этом случае ограничение было только одно: λ_3 равно $-\lambda_1\lambda_2$. При наличии $к$ объясняющих переменных количество ограничений также было бы равно $к$. Если исходная модель имеет вид:

$$y_t = \alpha + \beta_1 x_{1t} + ... + \beta_k x_{kt} + u_t ,$$

где u_t, формируется на основе соотношения (7.20), то преобразованная модель будет представлена выражением:

$$y_t = \alpha(1-\rho) + \rho y_{t-1} + \beta_1(x_{1t} - \rho x_{1t-1}) + ... + \beta_k(x_{kt} - \rho x_{kt-1}) + \varepsilon_t ,$$

и, таким образом, каждой объясняющей переменной соответствует ограничение, состоящее в том, что коэффициент при лаговом значении объясняющей переменной должен равняться произведению со знаком «минус» коэффициента при текущем значении этой переменной и коэффициента при y_{t-1} .

Пример

Преобразованная по методу Кокрана—Оркатта логарифмическая регрессия между расходами на жилье, располагаемым личным доходом и относительной ценой имеет следующий вид (в скобках приведены

(7.41)

$R2 = 0{,}9996$; $RSS = 0{,}0008$; $d = 1{,}94$; $h = 0{,}16$.

стандартные ошибки):

$$\log y, = 4{,}47 + 0{,}40 \log x, - 0{,}26 \log p_t; \qquad (7.39)$$
(1,05) (0,11) (0,14)

$R^2 = 0{,}9994$; $RSS = 0{,}0014$; р = 0,98; </=1,93.

Результаты оценивания регрессии (7.34) по МНК без учета ограничений могут быть представлены в виде:

$$\log y, = 0{,}73 + 0{,}87 \log y_{t\_x} + 0{,}22 \log x, - \text{(с.о.)}\ (0{,}48)\ (0{,}06)\ (0{,}09)$$

$$-0{,}11 \log x,\_, -0{,}19 \log/»,+ 0{,}01 \log p,\_{}_t; \qquad (7.40)$$
(0,11) (0,14) (0,17)

$R^2 = 0{,}9997$; $RSS = 0{,}0008$; $d = 2{,}27$; $h = -0{,}67$.

Рассмотрим это уравнение, прежде чем применить *тест на общий фактор.* Мы получаем оценку р из коэффициента при log y,_,. Верно ли, что коэффициент при log x,_, приблизительно равен умноженному на —0,87 коэффициенту при log x, и что коэффициент при log /»t_1 приблизительно равен умноженному на -0,87 коэффициенту при log *p*,? Очевидно, нет, по меньшей мере на первый взгляд. Статистика, лежащая в основе критерия, рассчитывается как 24 log(0,0014/0,0008), что равняется 13,4. Критическое значение *у}* с двумя степенями свободы при уровне значимости в 1% составляет 9,2 (см. табл. А.4). Следовательно, ограничение подлежит обоснованному отклонению (но при этом не нужно забывать, что данный тест следует использовать только для больших выборок). Еще одно свидетельство в пользу уравнения (7.40) обеспечивается тем, что А-тест показывает отсутствие статистически значимой автокорреляции.

Если мы выполним /-тест применительно к коэффициентам уравнения без ограничений, то увидим, что только одна лаговая переменная (log y,_,) имеет значимый коэффициент. Это означает, что мы можем опустить два других лаго- вых члена. Если мы сделаем это и повторно оценим регрессию (снова используя обычный МНК), то получим:

Здесь нет статистически значимой автокорреляции. *Вывод*: Ярко выраженная автокорреляция в первоначальной регрессии между расходами на жилье, доходом и ценой фактически объясняется пропуском лаговой зависимой переменной.

Резюме

В связи с проведенным анализом следует отметить, что если при оценивании регрессии мы получаем rf-статистику, которая явно указывает на автокорреляцию, то в первую очередь следует выполнить общий факторный тест, используя как преобразование по методу Кокрана—Оркатга, так и вариант без ограничений. Если ограничение не отклоняется, следует придерживаться результата, полученного по методу Кокрана—Оркатта. Если оно отклоняется, следует сосредоточиться на варианте без ограничений и попробовать внести новые усовершенствования. Например, не всегда необходимо сохранять все лаговые переменные.