Влияние отсутствия переменной, которая не включена в уравнение

К каким результатам приведет включение в уравнение регрессии переменной, которой там не должно быть? Каковы последствия невключения переменной, которая должна там присутствовать? Что произойдет, если при наличии трудностей в поиске исходных данных вы решите использовать вместо них «заменители»? В данной главе, представляющей собой предварительную попытку решения этих вопросов, основное внимание сосредоточено на последствиях неправильной спецификации переменной. Более сложный предмет – процедура выбора модели – будет затронут в последней главе книги.

В главе показано, каким образом могут быть проверены простейшие ограничения по параметрам. Глава завершается рассмотрением проблем введения переменных с запаздыванием (лагом) и описания фактора времени в моделях, основанных на данных временных рядов.

1. Моделирование

Построение эконометрической модели включает спецификацию составляющих ее соотношений, выбор переменных, входящих в каждое соотношение, а также определение математической функции, представляющей каждое соотношение. Последний элемент был рассмотрен в главе 4 и затем еще раз в главе 5. В данной главе мы рассмотрим второй из вышеперечисленных элементов и будем по-прежнему предполагать, что модель состоит только из одного уравнения. Вопрос о применении регрессионного анализа в моделях, состоящих из систем одновременных уравнений, будет рассмотрен в главе 11.

Если точно известно, какие объясняющие переменные должны быть включены в уравнение при проведении регрессионного анализа, то наша задача – ограничиться оцениванием их коэффициентов, определением доверительных интервалов для этих оценок и т. д. Однако на практике мы никогда не можем быть уверены, что уравнение специфицировано правильно. Экономическая теория должна указывать направление, но теория не может быть совершенной. Не будучи уверенными в ней, мы можем включить в уравнение переменные, которых там не должно быть, и в то же время мы можем не включить другие переменные, которые должны там присутствовать.

Свойства оценок коэффициентов регрессии в значительной мере зависят от правильности спецификации модели. Результаты

неправильной спецификации переменных в уравнении могут быть в обобщенном виде выражены следующим образом.

Несмещенной называют статистическую оценку, математическое ожидание которой равно оцениваемому параметру.

Смещенной называют статистическую оценку, математическое ожидание которой не равно оцениваемому параметру.

1. **Если опущена переменная, которая должна быть включена, то оценки коэффициентов регрессии, вообще говоря, хотя и не всегда, оказываются смещенными**. Стандартные ошибки коэффициентов и соответствующие t-тесты в целом становятся некорректными.
2. **Если включена переменная, которая не должна присутствовать в уравнении, то оценки коэффициентов регрессии будут несмещенными, однако, вообще говоря (хотя и не всегда), – неэффективными. Стандартные ошибки будут в целом корректны, но из-за неэффективности регрессионных оценок они будут излишне большими**.

Мы начнем с рассмотрения этих двух случаев, а затем перейдем к более широким аспектам спецификации модели.

2. Влияние отсутствия в уравнении переменной, которая должна быть включена

Проблема смещения

Предположим, что переменная y зависит от двух переменных x_1 и x_2 в соответствии с соотношением:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + u \qquad (6.1)$$

однако вы не уверены в значимости x_2. Считая, что модель должна выглядеть как

$$y = \alpha + \beta_1 x_1 + u \qquad (6.2)$$

вы оцениваете регрессию

$$y = a + b_1 x_1 \qquad (6.3)$$

и вычисляете b_1 по формуле $Cov(x_1,y)/D(x_1)$ вместо правильного выражения, данного в уравнении (5.12). По определению, b_1 является несмещенной оценкой величины β_1, если $E(b_1)$ равняется β_1. Практически, если соотношение (6.1) верно, то

$$E\left\{\frac{Cov(x_1, y)}{D(x_1)}\right\} = \beta_1 + \beta_2 \frac{Cov(x_1, x_2)}{D(x_1)} \qquad (6.4)$$

Сначала мы дадим интуитивное объяснение этого, а затем – формальное доказательство.

В разделе 5.2 показано, что если опустить x_2 в регрессионном соотношении, то переменная x_1 будет играть двойную роль: отражать свое прямое влияние и заменять переменную x_2 в описании ее влияния. Данное кажущееся опосредованное влияние величины x_1 на $\boldsymbol{y}$ будет зависеть от двух факторов: от видимой способности x_1 имитировать поведение x_2 и от влияния величины x_2 на $\boldsymbol{y}$.

Кажущаяся способность переменной x_1 объяснять поведение x_2 определяется коэффициентом наклона h в псевдорегрессии:

$$x_2 = g + hx_1$$

Величина h, естественно, рассчитывается при помощи обычной формулы для парной регрессии, в данном случае $Cov(x_1, x_2)/D(x_1)$. Влияние величины x_2 на $\boldsymbol{y}$ определяется коэффициентом β_2. Таким образом, эффект имитации посредством величины β_2 может быть записан как $\beta_2 Cov(x_1, x_2)/D(x_1)$. Прямое влияние величины x_1 на $\boldsymbol{y}$ описывается с помощью β_1. Таким образом, при оценивании регрессионной зависимости $\boldsymbol{y}$ от переменной x_1 (без включения в нее переменной x_2) коэффициент при x_1 определяется формулой:

$$h = \beta_1 + \beta_2 \frac{Cov(x_1, x_2)}{D(x_1)}$$

При условии, что величина x_1 не является стохастической, ожидаемым значением коэффициента будет сумма первых двух членов этой формулы. Присутствие второго слагаемого предполагает, что математическое ожидание коэффициента будет отличаться от истинной величины β_1, другими словами, оценка будет смещенной.

Формальное доказательство соотношения (6.4) не представляет труда. Выполним ряд теоретических преобразований оценки b_1.

$$b_1 = \beta_1 + \beta_2 \frac{Cov(x_1, x_2)}{D(x_1)} + \frac{Cov(x_1, u)}{D(x_1)}$$

Если величины x_1 и x_2 являются нестохастическими, то при вычислении математического ожидания величины b_1 первые два члена в уравнении (6.7) остаются неизменными, а третий будет равен нулю. Отсюда мы получаем формулу (6.4).

Этим подтверждается наш интуитивный вывод, что b_1 смещена на величину, равную $\beta_2 \frac{Cov(x_1, x_2)}{D(x_1)}$. Направление смещения будет зависеть от знака величин β_2 и $Cov(x_1,x_2)$ **.** Например, если β_2 положительна, а также положительна ковариация, то смещение будет положительным, а b_1 будет в среднем давать завышенные оценки β_2. Самостоятельно вы можете рассмотреть и другие случаи.

Есть, однако, один исключительный случай, когда оценка b_1 остается несмещенной. Это случается, когда выборочная ковариация между x_1 и x_2 в точности равняется нулю. Если $Cov(x_1,x_2) = 0$, то смещение исчезает. Действительно, коэффициент, полученный с использованием парной регрессии, будет точно таким же, как если бы вы оценили правильно специфицированную множественную регрессию. Конечно, величина смещения здесь равнялась бы нулю и при $\beta_2 = 0$, но в этом случае неправильной спецификации не возникает.