Л. Н. Слуцкин

Обобщенный метод моментов

Мы продолжаем обзор наиболее значительных достижений в эконометрической науке, еще не достаточно полно освещенных в отечественной литературе.

Обобщенный метод моментов — ОММ (generalized method of moments — GMM) был введен в эконометрику Л. Хансеном [Hansen (1982)] и является одновременно обобщением метода моментов (ММ) и метода наименьших квадратов (МНК) для оценки параметров модели. В статье мы рассмотрим применение ОММ для нахождения оценок параметров модели линейной регрессии.

1. Метод моментов

Мы напомним, что математическое ожидание $E(x)$ случайной величины x называется ее **моментом (первого порядка) — *m*,** а **выборочным моментом** — ее выборочное среднее:

$$\overline{m}_n = \frac{\Sigma x_i}{n}, \tag{1.1}$$

где n — число наблюдений, а $x_1, \ldots, x_n$ — случайная выборка значений x. Теорема Хинчина утверждает, что $\overline{m}_n$ сходится по вероятности к m, т. е.

$$\operatorname*{plim}_{n\to\infty} \overline{m}_n = m. \tag{1.2}$$

В дальнейшем мы откажемся от требования, чтобы случайные величины $x_1, \ldots, x_n$ были независимыми и даже, чтобы они имели одинаковое вероятностное распределение (но, все они должны иметь одинаковое математическое ожидание $E(x_i) = m, i = 1, 2, \ldots$). Тем не менее, аналог теоремы Хинчина — формула (1.2) остается верной при наложении определенных условий на x_i.

Соответственно для функции $g(x)$ ее момент[1] :

$$m(g(x)) = E(g(x)),$$

а ее выборочный момент:

$$\overline{m}_n(g(x)) = \frac{\Sigma g(x_i)}{n}. \tag{1.3}$$

В отличие от выборочного, момент $m(g(x))$ функции $g(x)$ называют ее **теоретическим моментом**.

Мы заметим, что $m(g(X))$ и $\overline{m}_n(g(X))$ определены также для случая, когда X — вектор, т. е. X представлен конечным набором случайных величин. В этом случае $g(X)$ будет функцией от соответствующего числа переменных. Более того, если $G(X)$ — векторная функция размер-

[1] При $g(x) = x^r$, мы получим начальный момент порядка r.

ности r, то с помощью формул (1.1) и (1.3) мы получим выборочные моменты, которые будут r-мерными векторами.

Как правило, $g(X)$ зависит от вектора параметров $\theta = (\theta_1, \dots, \theta_k)$[2]. Мы будем записывать этот факт в виде $g(X;\theta) = g(X;\theta_1, \dots, \theta_k)$. Соответственно, моменты $m(g(X;\theta))$ и $\overline{m}_n(g(X;\theta))$ также будут зависеть от параметров $\theta_1, \dots, \theta_k$.

Постановка задачи. Предположим, что имеются r функций — $g_1(X;\theta), \dots, g_r(X;\theta)$, и мы наблюдаем значения $X_1, \dots, X_n$, соответствующие вектору $\theta_0 = \theta_1^0, \dots, \theta_k^0$, для которого теоретические моменты $m(g_1(X;\theta_0)), \dots, m(g_r(X;\theta_0))$ равны 0[3]. Мы также предполагаем, что θ_0 — **идентифицируем**. Это означает, что, если $\theta \neq \theta_0$, то хотя бы один из моментов $m(g_1(X;\theta)), \dots, m(g_r(X;\theta))$ не равен 0. Требуется оценить вектор θ_0.

Мы запишем систему из r **уравнений моментов** относительно k неизвестных $\theta_1, \dots, \theta_k$:

$$\begin{aligned} &\overline{m}_n(g_1(X;\theta_1,\dots,\theta_k)) = 0 \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ &\overline{m}_n(g_r(X;\theta_1,\dots,\theta_k)) = 0. \end{aligned} \tag{1.4}$$

Другими словами, мы приравниваем выборочные моменты функций $g_1, \dots, g_r$ к их теоретическим моментам. Естественно, принять решение системы (1.4) $\hat{\theta}_1^0, \dots, \hat{\theta}_k^0$, при условии, что оно существует и единственное, за оценки соответствующих параметров $\theta_1^0, \dots, \theta_k^0$. Как правило, для этого требуется, чтобы число неизвестных и уравнений в (1.4) совпадало, то есть, чтобы $k = r$. Нахождение оценок $\hat{\theta}_1^0, \dots, \hat{\theta}_k^0$ при решении системы (1.4) называется **методом моментов (ММ)**.

1.1. Оценка коэффициентов модели линейной регрессии

Рассмотрим модель линейной регрессии:

$$y_i = \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \dots + \beta_k x_i^{(k)} + \varepsilon_i, \ i = 1,2,\dots,n, \tag{1.5}$$

где случайные члены $\varepsilon_1, \dots, \varepsilon_n$ — произвольные (не обязательно нормальные и, возможно, имеющие различные вероятностные распределения) случайные величины с одним и тем же математическим ожиданием, равным 0. Регрессоры $x^{(1)}, x^{(2)}, \dots, x^{(k)}$ также, в отличие от классической модели линейной регрессии, не предполагаются детерминированными[4]. Мы получим дополнительные ограничения как для случайных членов, так и для регрессоров, при изучении оценок параметров модели $\beta_1, \beta_2, \dots, \beta_k$. Тем самым наш подход принципиально отличается от принятого в курсах эконометрики, когда априори определяются все характеристики модели, а затем, на их основе, выводятся ее свойства.

Положим, что

$$g_j(x^{(j)}, \varepsilon) = x^{(j)}\varepsilon, \ j = 1, 2, \dots, k. \tag{1.6}$$

[2] Вероятностное распределение X также может зависеть от θ.

[3] Если $m(g(X)) = c \neq 0$, то мы рассмотрим $g_1(X) = g(X) - c$.

[4] Для одного и того же регрессора $x^{(j)}$, $x_{i_1}^{(j)}$ и $x_{i_2}^{(j)}$, $i_1 \neq i_2$, могут иметь различные вероятностные распределения, в том числе быть различными константами.

Для i-го наблюдения

$$g_j(x^{(j)},\varepsilon) = x_i^{(j)}\varepsilon_i. \tag{1.7}$$

Заметим, что в правой части (1.7) при различных i мы будем иметь в общем различные случайные величины.

Так как мы собираемся применить метод моментов, потребуем, чтобы для каждого $j = 1, 2,\dots,k$ выполнялось следующее **условие ортогональности**:

$$\boldsymbol{E(x_i^{(j)}\varepsilon_i) = 0,\ i = 1,2,\dots,n.} \tag{1.8}$$

Из (1.5) следует, что формулу (1.6) можно записать в виде функции от $k+1$ переменных $y, x^{(1)},\dots,x^{(k)}$ и k параметров $\beta_1,\dots,\beta_k$

$$g_j(y, x^{(1)},\dots,x^{(k)}; \beta_1,\dots,\beta_k) = x^{(j)}(y - \beta_1 x^{(1)} - \dots - \beta_k x^{(k)}),\ j = 1, 2, \dots, k. \tag{1.9}$$

Для определения параметров $\beta_1,\dots,\beta_k$, мы воспользуемся методом моментов (1.4), и получим следующую систему из k уравнений моментов с k неизвестными:

$$\sum_{i=1}^{n} x_i^{(j)}(y_i - \beta_1 x_i^{(1)} - \beta_2 x_i^{(2)} - \dots - \beta_k x_i^{(k)}) = 0,\ j = 1, 2, \dots, k. \tag{1.10}$$

Систему (1.10) можно записать как одно **уравнение моментов в векторном виде**:

$$X^T(Y - X\beta) = 0, \tag{1.11}$$

где $X = X_n$ — матрица, составленная из значений регрессоров модели (1.5). Мы предполагаем, как обычно, что ранг X равен k.

$$X = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(k)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(k)} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(k)} \end{pmatrix}, \tag{1.12}$$

вектора $Y = (y_1,\dots,y_n)$, $\beta = (\beta_1,\dots,\beta_k)$, и T — символ транспонирования.

Поскольку $\mathrm{rank}(X) = k$, матрица X^TX — невырожденная. Таким образом, решение уравнения (1.11) существует, единственное, и его можно записать в виде:

$$\hat{\beta} = (X^TX)^{-1}X^TY. \tag{1.13}$$

Из (1.11) следует, что вектор остатков $e = (e_1,\dots,e_n)$, соответствующий $\hat{\beta}$, ортогонален каждому из столбцов X (вектору $X^{(j)}$, образованному значениями регрессора $x^{(j)}$, $j = 1,\dots,k$):

$$\sum_{i=1}^{n} x_i^j e_i = 0,\ j = 1,\dots,k, \tag{1.14}$$

где $e_i = y_i - \hat{\beta}_1 x_i^{(1)} - \hat{\beta}_2 x_i^{(2)} - \dots - \hat{\beta}_i^k x_i^{(k)}$.

Полученная методом моментов оценка (1.13), совпадает с МНК-оценкой для β — при нахождении минимума квадратичной формы (по $\beta_1, \dots, \beta_k$):

$$Q = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \beta_1 x_i^{(1)} - \dots - \beta_k x_i^{(k)})^2, \tag{1.15}$$

или в векторной форме:

$$Q = (Y - X\beta)^T (Y - X\beta). \tag{1.15'}$$

Рассмотрим следующие свойства оценки (1.13).

1. Несмещенность, $E(\hat{\beta}) = \beta$.

Из (1.13) имеем:

$$\hat{\beta} = (X^T X)^{-1} X^T Y = (X^T X)^{-1} X^T (X\beta + \varepsilon) = \beta + (X^T X)^{-1} X^T \varepsilon, \tag{1.16}$$

где $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)$. Отсюда получаем:

$$E(\hat{\beta}) = \beta + E((X^T X)^{-1} X^T \varepsilon).$$

Таким образом, оценка $\hat{\beta}$ будет несмещенной тогда и только тогда, когда

$$E((X^T X)^{-1} X^T \varepsilon) = 0. \tag{1.17}$$

Для выполнения (1.17) достаточно потребовать[5] выполнения следующего условия **экзогенности**[6]:

$$\boldsymbol{E(\varepsilon_i | X) = 0, \ \ i = 1, 2, \dots, n}. \tag{1.18}$$

Очевидно, что условие ортогональности (1.8) следует из (1.18).

2. Состоятельность, $\operatorname*{plim}_{n\to\infty} \hat{\beta} = \beta$.

При больших значениях числа наблюдений n, значения $\hat{\beta}$ с большой вероятностью будут близки β. Согласно (1.16), для состоятельности $\hat{\beta}$ необходимо и достаточно, чтобы

$$\operatorname*{plim}_{n\to\infty} (X^T X)^{-1} X^T \varepsilon = 0. \tag{1.19}$$

Если записать

$$(X^T X)^{-1} X^T \varepsilon) = (n^{-1} X^T X)^{-1} (n^{-1} X^T \varepsilon), \tag{1.20}$$

то достаточными условиями для выполнения (1.19) будут:

а)

$$\operatorname*{plim}_{n\to\infty} n^{-1} X^T X = S, \tag{1.21}$$

где S — невырожденная матрица порядка k.

[5] Для доказательства достаточности можно воспользоваться формулой из теории вероятностей $E(\varepsilon x) = E(E(\varepsilon x | x))$.

[6] Обычно под *экзогенностью* регрессоров понимается более общее условие независимости ε_i от X, $i = 1, 2, \dots, n$. Что это не одно и то же, видно на примере гетероскедастичности случайных членов в том случае, когда дисперсия ε_i является функцией от X.

Примечание.

1. В случае, когда регрессоры $x^{(1)}, x^{(2)}, \ldots, x^{(k)}$ — независимые случайные величины, а столбцы X представляют собой случайные выборки значений регрессоров, $s_{i,j} = E(x^{(i)}x^{(j)})$.

2. Условие (1.21) часто не выполняется для нестационарных рядов[7]. Тем не менее, существуют более слабые условия Гренандера [Greene (2003)] на X, имеющие достаточно универсальный характер, при которых выполняется 2.

б)

$$\operatorname*{plim}_{n\to\infty} n^{-1}X^T\varepsilon = 0. \tag{1.22}$$

Примечание. Условие б) представляет собой не что иное, как аналог теоремы Хинчина (1.2). Если оно выполняется, то говорят, что случайные члены **асимптотически не коррелированы с регрессорами**.

Мы видим, что оценка может быть состоятельной даже при неэкзогенных регрессорах. Например, при выполнении условий а) и б), в качестве регрессоров могут выступать лагированные значения объясняемой переменной y ($y'_i = y_{i-v}, v > 0$).

3. Эффективность.

Эффективность означает, что $\hat{\beta}_i$, $i = 1, \ldots, k$, являются наилучшими несмещенными оценками (то есть, имеют наименьшее значение дисперсии) соответствующих параметров b_i в данном классе оценок. Теорема Гаусса-Маркова утверждает, что в случае, если ковариационная матрица для случайных членов имеет вид:

$$\Sigma_\varepsilon = \|\operatorname{cov}(\varepsilon_i\varepsilon_j)\| = \sigma^2 I, \tag{1.23}$$

где I — единичная матрица порядка n, оценка (1.13) является эффективной в классе линейных оценок по Y, т. е. вида cY, где c — n-мерный вектор.

Примечание. Вектор c в теореме Гаусса-Маркова может зависеть только от исходной информации (значений регрессоров), но не от ненаблюдаемых параметров модели.

Рассмотрим случай, когда условие (1.23) не выполняется, более подробно. Это означает, что или имеются ненулевые корреляционные отношения между случайными членами, соответствующими различным наблюдениям, или присутствует гетероскедастичность. Так как условия несмещенности и состоятельности не зависят от вида Σ_ε, то они выполняются (при тех же условиях) и здесь. Однако оценка $\hat{\beta}$ уже не будет больше эффективной. Чтобы показать это, найдем эффективную оценку β, которая в общем будет отлична от $\hat{\beta}$.

Для этого мы представим матрицу Σ_ε^{-1} в виде произведения PP^T, где P — невырожденная матрица порядка n. В таком случае после преобразования переменных P^TY, P^TX, мы получим модель линейной регрессии:

$$TY = P^TX\beta + \varepsilon'. \tag{1.24}$$

где $\varepsilon' = P^T\varepsilon$. Легко проверяется, что $\Sigma_{\varepsilon'} = \sigma^2 I$. В таком случае, если мы возьмем оценку, определяемую формулой (1.13) для уравнения (1.24), то получим эффективную оценку β:

[7] Можно легко проверить, что (1.21) не выполняется в случае, когда один из регрессоров — время (t =1,2, ...).

$$\hat{\beta}^* = (X^T PP^T X)^{-1} X^T PP^T Y.$$

Отсюда следует, что

$$\hat{\beta}^* = (X^T \Sigma_\varepsilon^{-1} X)^{-1} X^T \Sigma_\varepsilon^{-1} Y. \quad (1.25)$$

Заметим, что оценка (1.25) в отличие от (1.13) будет недоступной (unfeasible), так как ковариационная матрица Σ_ε — ненаблюдаемая.

Оценку $\hat{\beta}^*$ можно также получить из модели (1.24) либо методом моментов, либо, минимизируя соответствующую квадратичную форму. В первом случае мы получим:

$$X^T P(P^T Y - P^T X\beta) = 0.$$

Отсюда:

$$X^T \Sigma_\varepsilon^{-1}(Y - X\beta) = 0. \quad (1.11')$$

Можно заметить, что (1.11′) отличается от соответствующей формулы (1.11) добавлением матрицы Σ_ε^{-1}, что соответствует ортогональности каждого из векторов $X^{(j)}$, $j = 1,\dots,k$, вектору остатков e в n-мерном пространстве, в котором скалярное произведение задано матрицей Σ_ε^{-1}. Формула (1.25) следует из (1.11′) .

Квадратичная форма для модели (1.24) будет:

$$Q = (P^T Y - P^T X\beta)^T (P^T Y - P^T X\beta),$$

откуда получаем, что

$$Q = (Y - X\beta)^T \Sigma_\varepsilon^{-1}(Y - X\beta). \quad (1.15'')$$

Нахождение $\hat{\beta}^*$ при минимизации правой части формулы (1.15″) называется обобщенным методом наименьших квадратов (ОМНК). Если мы сравним формулы (1.15′) и (1.15″), то заметим, что различные произведения компонент вектора $Y - X\beta$ взяты в (1.15″) с множителями (весами), являющимися элементами матрицы Σ_ε^{-1}. В то же время в (1.15′) мы имеем только квадраты этих компонент.

1.2. Инструментальные переменные

При эндогенных (коррелированных со случайными членами) переменных свойства оценок, которые мы получили ранее, могут не выполняться. Такой пример нам дают системы регрессионных уравнений. В этом случае для эндогенных регрессоров находят соответствующие **инструменты или инструментальные переменные**. От инструментов уже требуется, чтобы они были экзогенными.

Пример 1. (Эконометрика [2006]). Модель динамики цен и заработной платы.

$$y_1 = a_{11}x_1 + b_{12}y_2 + e_1,$$
$$y_2 = a_{22}x_2 + a_{23}x_2 + b_{21}y_1 + e_2,$$

где y_1 — темпы изменения месячной заработной платы; y_2 — темпы изменения цен; x_1 — процент безработных; x_2 — темпы изменения постоянного капитала; x_3— темпы изменения цен на импорт сырья.

В данном примере переменные y_1, y_2 — эндогенные, а x_1, x_2, x_3 — экзогенные переменные модели. Для оценки коэффициентов первого уравнения системы можно в качестве инструмента для y_2 взять одну из экзогенных переменных x_2 или x_3 из второго уравнения.

Рассмотрим модель:

$$y_i = \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \ldots + \beta_k x_i^{(k)} + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{1.26}$$

в которой первые r, $r < k$, регрессоров $x^{(1)}, x^{(2)}, \ldots, x^{(r)}$ — экзогенные, а остальные — эндогенные. Предположим, что для эндогенных переменных $x^{(r+1)}, x^{(r+2)}, \ldots, x^{(k)}$ мы смогли найти инструментальные переменные $z^{(r+1)}, z^{(r+2)}, \ldots, z^{(k)}$. Обозначим матрицу, составленную из новых и старых регрессоров:

$$W = [X_r; Z_{k-r}], \tag{1.27}$$

где X_r — матрица порядка $n \times r$, соответствующая старым регрессорам, а Z_{k-r} — матрица порядка $n \times (k - r)$, соответствующая инструментам. Тогда уравнение моментов (1.11) можно записать в виде:

$$W^T (Y - X\beta) = 0. \tag{1.28}$$

Если матрица $W^T X$ — невырожденная, мы получим оценку для $\boldsymbol{\beta}$:

$$\hat{\beta}_{IV} = (W^T X)^{-1} W^T Y, \tag{1.29}$$

называемой **простой IV оценкой**. Если $W = X$, т. е. каждая экзогенная переменная является инструментом для самой себя, то мы получим формулу (1.13). Из (1.28) следует, что вектор остатков e, соответствующий $\hat{\beta}$, ортогонален каждому из столбцов W.

Оценка $\hat{\beta}_{IV}$ минимизирует квадратичную форму[8]:

$$Q = (Y - X\beta)^T W W^T (Y - X\beta). \tag{1.30}$$

Подставив $X\beta + \varepsilon$ вместо Y в (1.29), получим:

$$\hat{\beta}_{IV} - \beta = (W^T X)^{-1} W^T \varepsilon. \tag{1.31}$$

Так как элементы матрицы $(W^T X)^{-1} W^T$ могут быть коррелированны со случайными членами, то в отличие от (1.16) $\hat{\beta}_{IV}$, как правило, больше не будет несмещенной оценкой β. Также следует, что $\hat{\beta}_{IV}$ — состоятельная оценка β, тогда и только тогда, когда выражение в правой части (1.31) асимптотически стремится к 0. Например, $\hat{\beta}_{IV}$ будет состоятельной оценкой, если выполняются два условия аналогично (1.21) и (1.22).

а)

$$\operatorname*{plim}_{n \to \infty} n^{-1} W^T X = S, \tag{1.21'}$$

где S — невырожденная матрица порядка k.

б)

$$\operatorname*{plim}_{n \to \infty} n^{-1} W^T \varepsilon = 0. \tag{1.22'}$$

Как мы уже отмечали, при обсуждении формулы (1.22), для выполнения (1.22′) не обязательно требовать, чтобы столбцы W были экзогенны по отношению к случайным членам. Для этого часто достаточно, чтобы выполнялось следующее **условие предопределенности (predeterminedness condition)**:

$w^{(j)}_i$ и ε_i —независимы при любых *i* и *j*, *i* = 1, 2, …, *n*; *j* = 1, 2, …, *k*.

[8] Доказательство этого факта в более общем случае будет приведено в следующем разделе.

Заметьте, что случайный член должен быть независим только от значений инструментов *того же самого наблюдения*. Очевидно, что условие ортогональности (1.8) следует из предопределенности.

Результаты этого раздела резюмируем следующим образом. В каждом из рассмотренных случаев, оценки параметров модели линейной регрессии могут быть получены одним из двух способов:

- **метод моментов.** Нахождение решения системы уравнений, полученных из условия ортогональности (с различными весами, как в случае (1.11′);
- **минимизация квадратичной формы.** $Q = (Y - X\beta)^T \Phi (Y - X\beta)$, где Φ — симметричная положительно определенная матрица ($\Phi = I$ в (1.15′), $\Phi = \Sigma_\varepsilon^{-1}$ в (1.15″), $\Phi = WW^T$ в (1.30).

2. Обобщенный метод моментов

Произвольная симметричная положительно определенная матрица ψ_n порядка n задает в n-мерном пространстве R^n длину вектора $A(a_1, \dots, a_n)'$ формулой:

$$|A|_{\psi_n} = [(a_1, \dots, a_n)\psi_n(a_1, \dots, a_n)']^{1/2}. \tag{2.1}$$

Формула (2.1) является обобщением формулы для длины вектора:

$$|A| = (a_1^2, \dots, a_n^2)^{1/2}, \tag{2.2}$$

в случае, когда ψ_n — единичная матрица I_n. В действительности [Гельфанд (1998)], в некотором базисе в R^n $|A|_{\psi_n}$ будет иметь вид (2.2). Элементы ψ_n можно понимать, как некоторые **веса**, с которыми попарные произведения $a_i a_j$ входят в формулу для $|A|_{\psi_n}$.

Теперь, снова вернемся к примеру 1. В качестве инструмента для y_2 мы можем взять любую из переменных x_2, x_3. Возможно, существуют и другие экзогенные переменные, которые также можно было бы использовать в этом же качестве. Но, если мы включим в W более одной переменной, то система (1.28) станет **сверхидентифицируемой**, т. е. в ней будет больше уравнений, чем переменных, и, как правило, такая система не имеет решений. Следовательно, мы должны выбрать ровно один инструмент для y_2. Но, исключив из рассмотрения другие экзогенные переменные, мы, тем самым, сужаем полезную информацию о модели. В таком случае поступают следующим образом. Обратимся снова к модели:

$$y_i = \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + \dots + \beta_k x_i^{(k)} + \varepsilon_i, \quad i = 1, 2, \dots, n. \tag{2.3}$$

Рассмотрим матрицу W_n, столбцы которой образованы значениями экзогенных (предопределенных) переменных (инструментов) $z^{(1)}, z^{(2)}, \dots, z^{(r)}$. **Мы будем считать, что $r \geq k$, и $\mathrm{rank}(W_n^T X_n) = k$.** Запишем уравнение моментов:

$$W_n^T (Y_n - X_n \beta) = 0. \tag{2.4}$$

Так как система (2.4), в общем случае, не будет иметь решений, то мы в качестве оценки β возьмем вектор, при котором левая часть (2.4) наиболее близка к 0, т. е. имеет наименьшую длину (заданной некоторой положительно определенной матрицей ψ_n) в R^n.

Определение. Если ψ_n — симметричная положительно определенная матрица порядка r (матрица весов), то k-мерный вектор:

$$\hat{\beta}_{OMM}(\Psi_n) = \underset{\beta_1,\dots,\beta_k}{\arg\min}[W_n^T(Y_n - X_n\beta)]^T \Psi_n [W_n^T(Y_n - X_n\beta)],$$

или, в более компактной форме:

$$\hat{\beta}_{OMM}(\Psi_n) = \underset{\beta_1,\dots,\beta_k}{\arg\min}(Y_n - X_n\beta)^T W_n \Psi_n W_n^T (Y_n - X_n\beta), \tag{2.5}$$

называется **оценкой β (соответствующей ψ_n), полученной обобщенным метод моментов**.

Примечание. В определении $\hat{\beta}_{OMM}(\Psi_n)$, матрица ψ_n может зависеть от исходной информации, представленной матрицами X_n и W_n также, как и от размера выборки n.

Из условия для оптимизации квадратичной формы[9] [Магнус, Нейдеккер (2002)], получаем:

$$\hat{\beta}_{OMM}(\Psi_n) = (X_n^T W_n \Psi_n W_n^T X_n)^{-1} X_n^T W_n \Psi_n W_n^T Y_n. \tag{2.6}$$

В принципе, мы могли бы взять за определение $\hat{\beta}_{OMM}(\Psi_n)$ формулу (2.6). Но в таком случае, было бы непонятно, откуда она появилась.

Пример 2.

а) Рассмотрим случай $r = k$, то есть, число инструментов равно числу регрессоров. В таком случае, если $W_n^T X_n$ — невырожденная матрица порядка r, то формула (2.6), при любой ψ_n, даст нам простую IV оценку (1.29). Соответственно, квадратичная форма в (2.5) совпадает с Q в (1.30) при $\psi_n = I_r$.

б) Если $r > k$, то мы получим **обобщенную IV оценку** при $\psi_n = (W_n^T W_n)^{-1}$. А соответствующую квадратичную форму легко можно получить из (2.5).

Сейчас мы займемся изучением вопроса, при каких условиях оценка $\hat{\beta}_{OMM}(\Psi_n)$ — состоятельна. Для этого нам надо сделать, во-первых, предположение относительно W_n:

2-а. Матрица $W_n^T X_n$ имеет полный ранг, равный k так же как и $\operatorname{p}\lim_{n\to\infty} n^{-1} W_n^T X_n$.

Матрицы ψ_n мы будем подбирать так, чтобы выполнялось условие:

2-б. $\operatorname{p}\lim_{n\to\infty} \Psi_n$ существует и равен положительно определенной матрице.[10]

Если мы подставим $X_n\beta + \varepsilon_n$, $\varepsilon_n = (\varepsilon_1,\dots,\varepsilon_n)$ вместо Y_n в (2.6), то получим:

$$\hat{\beta}_{OMM}(\Psi_n) = \beta + (n^{-1} X_n^T W_n \Psi_n n^{-1} W_n^T X_n)^{-1} n^{-1} X_n^T W_n \Psi_n n^{-1} W_n^T \varepsilon_n. \tag{2.7}$$

Из 2-а,б следует, что $\operatorname{p}\lim_{n\to\infty} n^{-1} X_n^T W_n \Psi_n n^{-1} W_n^T X_n$ — невырожденная матрица порядка k, и

[9] Для полноты изложения, мы приведем доказательство. Необходимое условие для минимума в (2.5), $\frac{\partial[(Y_n - X_n\beta)^T W_n \Psi_n W_n^T (Y_n - X_n\beta)]}{\partial\beta} = -2X_n^T W_n \Psi_n W_n^T (Y_n - X_n\beta) = 0$. Отсюда следует (2.6). Достаточность следует из того факта, что гессиан квадратичной формы $(Y_n - X_n\beta)^T W_n \Psi_n W_n^T (Y_n - X_n\beta)$ равен $2X_n^T W_n \Psi_n W_n^T X_n$ — положительно определенная квадратичная форма.

[10] Неотрицательная определенность следует из существования $\operatorname{p}\lim_{n\to\infty} \Psi_n$.

$\operatorname{p}\lim\limits_{n\to\infty} n^{-1}X_n^T W_n \Psi_n$ существует. Таким образом, состоятельность $\hat{\beta}_{\text{ОММ}}(\Psi_n)$ сводится к асимптотической некоррелированности случайных членов с инструментами $z^{(1)}, z^{(2)}, \dots, z^{(r)}$:

2-в. $\operatorname{p}\lim\limits_{n\to\infty} n^{-1}W_n^T \varepsilon_n = 0.$

Для изучения дальнейших свойств $\hat{\beta}_{\text{ОММ}}(\Psi_n)$ нам потребуется вспомнить следующие понятия из теории вероятностей.

Определение. Состоятельная оценка $\hat{\theta}_n$ векторного параметра θ размерности k называется **асимптотически нормальной**, если

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{pac} N(0, Q), \tag{2.8}$$

где $N(0, Q)$ — k-мерное нормальное распределение с ковариационной матрицей Q порядка k.

Мы будем записывать (2.8) в виде — $\hat{\theta}_n \overset{a}{\sim} N(\theta, Q/n)$. Матрица Q/n называется **асимптотической ковариационной матрицей для** $\hat{\theta}_n$.

В случае, когда $k = 1$, состоятельная оценка $\hat{\theta}_n$ действительного параметра — асимптотически нормальна, если

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{pac} N(0, \sigma^2), \tag{2.9}$$

где $N(0, \sigma^2)$ — стандартное нормальное распределение. Записывается — $\hat{\theta}_n \overset{a}{\sim} N(\theta, \sigma^2/n)$.

Рассмотрим r-мерный вектор $n^{-1}W_n^T \varepsilon_n$. По 2-в, его вероятностный предел равен 0. Можно показать [Hayashi (2000)], что при определенных, достаточно общих условиях[11] на W_n и ε_n,

$$\sqrt{n}(n^{-1}W_n^T \varepsilon_n) \xrightarrow{pac} N(0, \Omega) \tag{2.10}$$

с некоторой положительно определенной матрицей Ω, в общем зависящей от модели (2.3) и W_n.

Ниже приведены два основных свойства $\hat{\beta}_{\text{ОММ}}(\Psi_n)$. Мы предполагаем справедливость 2-а, б, в. Вероятностные пределы матриц $n^{-1}W_n^T X_n$ и ψ_n будем обозначать через S_{W^TX} и S_ψ соответственно.

1. Оценка $\hat{\beta}_{\text{ОММ}}(\Psi_n)$ — состоятельна, что доказано выше.

2. $$\sqrt{n}(\hat{\beta}_{\text{ОММ}}(\Psi_n) - \beta) \xrightarrow{pac} N(0, Q(S_\Psi)), \tag{2.11}$$

где

$$Q(S_\Psi) = (S_{W^TX}^T S_\Psi S_{W^TX})^{-1} S_{W^TX}^T S_\Psi \Omega S_\Psi S_{W^TX} (S_{W^TX}^T S_\Psi S_{W^TX})^{-1}, \tag{2.12}$$

что следует из (2.7) для $\hat{\beta}_{\text{ОММ}}(\Psi_n) - \beta$[12].

[11] Точная формулировка этих условий предполагает знакомство с теорией эргодических стационарных процессов.

[12] Мы воспользовались при доказательстве следующим результатом из теории вероятности: если X_n — r-мерный вектор, A_n — матрица порядка $k \times r$, так что $\sqrt{n}X_n \xrightarrow{pac} N(0, \Sigma)$ и $\operatorname{p}\lim\limits_{n\to\infty} A_n = A$, то $\sqrt{n}A_nX_n \xrightarrow{pac} N(0, A\Sigma A^T)$.

Определение. Асимптотически нормальная оценка $\hat{\theta}_n$ называется **асимптотически эффективной** в данном классе M асимптотически нормальных оценок, если для любой оценки $\hat{\theta}'_n$ из M, матрица Q' — Q — неотрицательно определена.

Асимптотическая эффективность означает, что любая линейная комбинация компонент $\hat{\theta}_n$ имеет асимптотически не большую дисперсию, чем такая же линейная комбинация компонент $\hat{\theta}'_n$.

Естественным было бы попытаться узнать, когда $\hat{\beta}_{ОММ}(\Psi_n)$ будет асимптотически эффективной в класс оценок, имеющих свойство (2.12). Результат Хансена [Hansen (1982)], который отвечает на этот вопрос, является центральным для обобщенных методов моментов:

Оценка $\hat{\beta}_{ОММ}(\Psi_n)$ будет асимптотически эффективной, если $S_\psi = \Omega^{-1}$.

Обозначим асимптотически эффективную оценку — $\hat{\beta}_{ЭОММ}(\hat{\Omega}_n^{-1})$ ($\operatorname{p}\lim_{n\to\infty} \hat{\Omega}_n = \Omega$). Асимптотическая ковариационная матрица для $\hat{\beta}_{ЭОММ}(\hat{\Omega}_n^{-1})$ легко находится из формулы (2.12):

$$n^{-1}Q(\Omega^{-1}) = n^{-1}(S_{W^TX}^T \Omega^{-1} S_{W^TX})^{-1}. \tag{2.13}$$

Асимптотическая ковариационная матрица (2.13) однозначно определяется матрицами S_{W^TX} и W.

Из примера 2-а следует, что в случае, когда число инструментов равно числу регрессоров ($r = k$), оценка $\hat{\beta}_{IV} = (W^TX)^{-1}W^TY$ является асимптотически эффективной. Асимптотическая ковариационная матрица для $\hat{\beta}_{IV}$ вычисляется по формуле (2.13). В отсутствие новых инструментальных переменных ($W = X$), мы получаем, что $\hat{\beta}_{МНК} = (X^TX)^{-1}X^TY$ — асимптотически эффективна. Этот результат довольно неожиданный, так как мы знаем, что для конечных выборок $\hat{\beta}_{МНК}$ — эффективна в случае: $\Sigma_{\varepsilon_n} = \sigma^2 I_n$.

Итак, подведем итоги по данному разделу.

1. С каждой положительно определенной матрицей ψ_n порядка n однозначно ассоциируется оценка $\hat{\beta}_{ОММ}(\Psi_n)$ вектора параметров модели β (при условии наличия n наблюдений и матрицы W_n).
2. Если $\operatorname{p}\lim_{n\to\infty} \Psi_n$ — положительно определенная матрица, то (при выполнении ряда условий на X, W и ε), $\hat{\beta}_{ОММ}(\Psi_n)$ — состоятельная асимптотически нормальная оценка β.
3. Если $\operatorname{p}\lim_{n\to\infty} \hat{\Omega}_n = \Omega$, то $\hat{\beta}_{ОММ}(\hat{\Omega}_n^{-1})$ — асимптотически эффективна ($\hat{\beta}_{ЭОММ}(\hat{\Omega}_n^{-1})$).

3. ГАС-оценка[13]

Для нахождения эффективной оценки $\hat{\beta}_{ЭОММ}(\hat{\Omega}_n^{-1})$ мы можем воспользоваться формулой (2.6) ($\Psi_n = \hat{\Omega}_n^{-1}$). Таким образом, поиск $\hat{\beta}_{ЭОММ}(\hat{\Omega}_n^{-1})$ сводится к нахождению $\hat{\Omega}_n$.

Из (2.10) следует, что $n\Omega$ является асимптотической ковариационной матрицей для вектора $W_n^T \boldsymbol{\varepsilon}_n$. Ковариационная матрица для $W_n^T \boldsymbol{\varepsilon}_n$ вычисляется по формуле:

$$E(W_n^T \boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}_n^T W_n) = E[E(W_n^T \boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}_n^T W_n | W_n)] = E(W_n^T \Sigma_{\varepsilon_n} W_n), \tag{3.1}$$

[13] В нашем изложении мы следовали [Hayashi (2000)] и [Davidson, MacKinnon (2004)].

где Σ_{ε_n} — ковариационная матрица для случайных членов. По этой причине, $\hat{\Omega}_n$ носит название **гетероскедастичной автокорреляционной состоятельной (ГАС)** оценки Ω[14].

Рассмотрим сначала случай, при котором Σ_ε имеет диагональный вид (гетероскедастичность без автокорреляции). Тогда для нахождения ГАС-оценки $\hat{\Omega}_n$ поступают следующим образом[15]:

1. Берется состоятельная оценка $\hat{\beta}$. Например, простая или обобщенная IV оценка.
2. Вычисляется вектор остатков
$$\boldsymbol{e}_n = Y_n - X_n\hat{\beta}. \tag{3.2}$$
3. Полагаем[16]
$$\hat{\Omega}_n = n^{-1}W_n^T\Sigma_{e_n}W_n, \tag{3.3}$$

где Σ_{e_n} — диагональная матрица с элементами $e_1^2,\dots,e_n^2$ на главной диагонали.

В общем случае присутствия корреляционных связей между случайными ошибками запишем:
$$n^{-1}E(W_n^T\boldsymbol{\varepsilon}_n\boldsymbol{\varepsilon}_n^TW_n) = n^{-1}\sum_{t=1}^{n}\sum_{s=1}^{n}E(\varepsilon_t\varepsilon_sW_t^TW_s), \tag{3.4}$$

где W_t — строка W с номером t. Из (3.4) следует, что
$$n^{-1}E(W_n^T\boldsymbol{\varepsilon}_n\boldsymbol{\varepsilon}_n^TW_n) = \Gamma(0) + \sum_{j=1}^{n-1}(\Gamma(j) + \Gamma^T(j)), \tag{3.5}$$

где
$$\Gamma(j) = n^{-1}\sum_{t=j+1}^{n}E(\varepsilon_t\varepsilon_{t-j}W_t^TW_{t-j}),\ j \ge 0. \tag{3.6}$$

Из (3.5) следует (при тех же условиях, что и для (2.10):
$$\Omega = \Gamma(0) + \lim_{n\to\infty}\sum_{j=1}^{n-1}(\Gamma(j) + \Gamma^T(j)). \tag{3.7}$$

В качестве оценки для $\Gamma(j)$, $j \ge 0$, естественно взять
$$\hat{\Gamma}(j) = n^{-1}\sum_{t=j+1}^{n}e_te_{t-j}W_t^TW_{t-j}, \tag{3.8}$$

где e_t то же, что и в (3.2). Оценки, которые мы сейчас обсудим, отличаются только тем, как различные исследователи обходят трудность с бесконечным числом слагаемых в (3.7).

Оценка Хансена-Уайта (Hansen-White estimator). В правой части (3.7) берется конечное число членов, $j = 1,\dots,p$, где p — некоторый порог (lag truncation parameter), зависящий

[14] В англоязычной литературе — heteroskedasticity autocorrelation consistent (HAC) estimator.

[15] Предполагается, что все инструменты в W — стационарны.

[16] Оценки типа (3.3) были впервые исследованы Уайтом [White (1980)].

от n. Другими словами, слагаемые в (3.7) просто урезаются при $j>p$. Таким образом, мы имеем:

$$\hat{\Omega}_{HW}=\Gamma(0)+\sum_{j=1}^{p}(\Gamma(j)+\Gamma^{T}(j)). \tag{3.9}$$

Главный недостаток $\hat{\Omega}_{HW}$, что она может не удовлетворять одному из основных требований к ГАС-оценкам — быть положительно определенной.

Оценка Ньюэя-Уеста (Newey-West). Эта оценка также как и предыдущая урезает все слагаемые в (3.7), кроме конечного числа. Однако в отличие от нее, приписывает оставшимся членам веса, уменьшающиеся в процессе их удаления от j=0.

$$\hat{\Omega}_{NW}=\Gamma(0)+\sum_{j=1}^{p}(1-\frac{j}{p+1})(\Gamma(j)+\Gamma^{T}(j)). \tag{3.10}$$

Оценка Ньюэя-Уеста, в отличие от Хансена-Уайта, положительно определенная. Весовой коэффициент $1-j/(p+1)$ в (3.10) называется ядром Бартлета (Bartlett kernel).

Для того чтобы быть ГАС-оценками, в обоих случаях требуется, чтобы p монотонно возрастало с ростом n.

Наконец, что касается практического вычисления $\hat{\beta}_{ЭОММ}(\hat{\Omega}_{n}^{-1})$, то в EViews предусмотрен целый набор ГАС-оценок на все случаи жизни. Для пространственных данных предлагается оценка (3.3). Кроме Ньюэя-Уеста, имеется также оценка с квадратичным спектральным ядром (quadratic spectral — QS) [Hayashi (2000)], имеющая лучшие асимптотические свойства, чем у Ньюэя-Уеста.

4. Тестирование модели

После того, как определили ГАС-оценку $\hat{\Omega}_{n}$ и нашли $\hat{\beta}_{ЭОММ}(\hat{\Omega}_{n}^{-1})$ по формуле (2.6), мы можем протестировать модель. Тестирование в ОММ носит асимптотический характер и, следовательно, применимо к выборкам большого размера.

***t*-отношение (*t*-ratio).** Согласно формулам (2.11) и (2.13),

$$\sqrt{n}(\hat{\beta}_{ЭОММ}(\hat{\Omega}_{n}^{-1})-\beta)\xrightarrow{рас}N(0,Q(\Omega^{-1})), \tag{4.1}$$

где

$$Q(\Omega^{-1})=(S_{W^{T}X}^{T}\Omega^{-1}S_{W^{T}X})^{-1}. \tag{4.2}$$

Проверяется нулевая гипотеза H_0, что $\beta_j=0$ (против H_1, что $\beta_j\neq 0$), $1\leq j\leq k$. t-отношение находится по формуле:

$$t_j=\frac{\sqrt{n}\hat{\beta}_{ЭОММ,j}(\hat{\Omega}_{n}^{-1})}{\sqrt{\lambda_{j,j}}}=\frac{\hat{\beta}_{ЭОММ,j}(\hat{\Omega}_{n}^{-1})}{SE_j}, \tag{4.3}$$

где $\hat{\beta}_{ЭОММ,j}(\hat{\Omega}_{n}^{-1})$ — j-я компонента $\hat{\beta}_{ЭОММ}(\hat{\Omega}_{n}^{-1})$, $\lambda_{j,j}$ — j-й диагональный элемент матрицы $(X_n^T W_n \hat{\Omega}_n^{-1} W_n^T X_n)^{-1}$, которая получается из (4.2) заменой предельных матриц на их оценки, стандартная ошибка — $SE_j=\sqrt{\lambda_{j,j}/n}$.

Из (4.1) следует, что $t_j \xrightarrow{pac} N(0;1)$[17].

***J*-статистика Хансена.** Рассмотрим экстремальное (минимальное) значение квадратичной формы (2.5), соответствующее $\hat{\beta}_{\text{ЭОММ}}(\hat{\Omega}_n^{-1})$:

$$J(\hat{\Omega}_n^{-1}) = (Y_n - X_n\hat{\beta}_{\text{ЭОММ}})^T W_n \hat{\Omega}_n^{-1} W_n^T (Y_n - X_n\hat{\beta}_{\text{ЭОММ}}), \tag{4.4}$$

где мы для краткости обозначили $\hat{\beta}_{\text{ЭОММ}}(\hat{\Omega}_n^{-1})$ через $\hat{\beta}_{\text{ЭОММ}}$. Второй фундаментальный результат, полученный Хансеном [Hansen (1982)] утверждает, что

$$n^{-1}J(\hat{\Omega}_n^{-1}) \xrightarrow{pac} \chi^2(r-k) \tag{4.5}$$

при $r > k$[18]. При значениях $n^{-1}J(\hat{\Omega}_n^{-1})$, превышающих критическое значение (при заданном уровне значимости) χ^2-статистики, условие ортогональности (1.8) (экзогенности, предопределенности) инструментов $w^{(1)}, \dots, w^{(r)}$ к случайным членам отвергается. Тут нужно заметить, что причиной больших значений χ^2-статистики могут быть неправильная спецификация модели, а также невыполнение условий, которые нам потребовались при нахождении $\hat{\Omega}_n^{-1}$.

Пример 3 [Dahlberg, Johansson(2000)]. Этот пример также приведен в книге Грина [Greene (2003)]. В работе исследовались статьи расходов нескольких сот шведских муниципалитетов за период 1979–1987 гг. В качестве объясняющих рассматривались лагированные переменные: расходы (*S* — spending), доходы (*R* — revenues) и гранты (*G* — grants), а также фиктивные переменные по годам. Отдельные модели строились в зависимости от значения максимального лага m ($m = 1, 2, 3$). При $m = 3$ общее количество объясняющих переменных равнялось 14. Уравнение модели выглядит следующим образом:

$$\Delta S_{i,t} = \lambda_t + \sum_{j=1}^{m}\beta_j\,\Delta S_{i,t-j} + \sum_{j=1}^{m}\gamma_j\,\Delta R_{i,t-j} + \sum_{j=1}^{m}\delta_j\,\Delta G_{i,t-j} + u_{i,t}, \tag{4.6}$$

где Δ — оператор взятия разности, $u_{i,t}$ — случайная ошибка, соответствующая i-му муниципалитету, $i = 1, 2, \dots, 265$, и t-му году ($t = 1983, \dots, 1987$).

В качестве инструментальных переменных использовалось дополнительно 16 переменных. Инструментальные переменные строились как значения расходов за определенный год. Причем, если данный год участвовал в наблюдении, то эта инструментальная переменная принималась равной 0. Авторы оценили коэффициенты для каждой из переменной, стандартные ошибки и t-отношения. Приведем часть данных в табл. 1.

Значение χ^2-статистики с 16 степенями свободы, соответствующее 5% уровню значимости — $\chi^2_{0,95}(16) = 26{,}3$. Таким образом, мы видим, что $J = 22{,}83 < 26{,}3$, и модель принимается. В том числе делается вывод о не эндогенности инструментов. Однако для модели, где в качестве объясняемой переменной рассматривались расходы — R, $J = 20{,}54$, что свидетельствовало о неверной спецификации.

[17] t_j не является t-статистикой, так как мы не предполагаем, что случайная ошибка ε нормально распределена.

[18] Поэтому этот тест также называется **тестом на сверхидентифицируемость**. Как легко видеть, при $r = k$, уравнение моментов (2.4) имеет решение, и, следовательно, $J \equiv 0$.

Таблица 1

Оценка параметров модели методом ОММ, *lag* = 3

Переменная	ОММ-оценка	SE	t-ratio
Year 1983	–0,00	0,00	–12,32*
………	………	………	………
Year 1987	0,00	0,00	5,87*
$S(t-1)$	1,15	0,34	3,36*
$S(t-2)$	–0,04	0,23	–0,17
$S(t-3)$	–0,56	0,22	–2,59*
$R(t-1)$	–1,24	0,36	–3,42*
$R(t-2)$	0,08	0,27	0,28
$R(t-3)$	0,65	0,27	2,41*
$G(t-1)$	0,02	0,82	0,02
$G(t-2)$	1,55	0,76	2,05*
$G(t-3)$	1,79	0,69	2,58*
$J = 22{,}83$			

Примечание. * — значимость на уровне 5%.

5. Заключение

Со времени опубликования Хансеном своих результатов [Hansen (1982)], обобщенный метод моментов получил большое распространение, в основном в теоретической, но также и в прикладной эконометрике. По своему характеру, ОММ, основанный на асимптотических свойствах оценок, может быть эффективно использован лишь для больших выборок. Преимуществом ОММ является отсутствие привязки к априори заданному (обычно нормальному) распределению ошибок. Также надо отметить большое методологическое значение метода. Практически все известные оценки (мы рассмотрели только часть из них) могут быть получены при его применении. Напоследок заметим, что применение ОММ для нелинейной регрессии основано на принципах, изложенных выше. Детали можно найти у Грина [Greene (2003)].

Список литературы

Айвазян С. А., Иванова С. С. Эконометрика. М.: Маркет ДС, 2007.

Гельфанд И. М. Лекции по линейной алгебре. М.: Добросвет МЦНМО, 1998.

Магнус Я., Нейдеккер Х. Матричное дифференциальное исчисление с приложениями к статистике и эконометрике. М.: Физматлит, 2002.

Рао С. Р. Линейные статистические методы и их применения. М.: Наука, 1968.

Эконометрика (под ред. Елисеевой). М.: Финансы и Статистика, 2006.

Dahlberg M., Johansson E. An examination of the dynamic behavior of local governments using GMM bootstrapping methods // *Journal of Applied Econometrics.* 2000. № 15. P. 401–416.

Davidson R., MacKinnon J. Econometric theory and methods. New York: Oxford University Press, 2004.

Greene W. Econometric analysis. New York: Prentice Hall, 2003.

Hansen L. Large sample properties of generalized method of moment's estimators, Econometrica, 1982. № 50.

Hayashi F. Econometrics. Princeton: Princeton University Press, 2000.

White H., A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity, Econometrica. 1980. № 48. P. 817–838.