Прикладная эконометрика, 2018, т. 50, с. 114–143.
Applied Econometrics, 2018, v. 50, pp. 114–143.

Е. В. Коссова, Б. С. Потанин[1]

Обобщение метода Хекмана и модели с переключением на случай произвольного числа уравнений отбора

В данной статье предлагается модель, обобщающая регрессионную модель с переключением и модель Хекмана на случай произвольного числа бинарных уравнений отбора наблюдений. Рассматриваются два способа оценивания модели при допущении о совместном нормальном распределении случайных ошибок: метод максимального правдоподобия и двухшаговая процедура, обобщающая классический подход Хекмана. Качество оценок модели проверяется при помощи анализа симулированных данных в случае двух уравнений отбора. Результаты данного анализа свидетельствуют о значительном превосходстве точности оценок предложенного метода над методом наименьших квадратов и методом Хекмана.

Ключевые слова: систематическая ошибка отбора; регрессионные модели с переключением.
JEL classification: C34.

Введение

При изучении различных экономических явлений исследователям довольно часто приходится сталкиваться с проблемой селективности выборок. В этом случае существует некоторое правило, которое определяет, будет ли доступно то или иное наблюдение. Классическим примером является наблюдение заработной платы индивидов: на вопрос о заработной плате отвечают лишь занятые респонденты (Heckman, Killingsworth, 1987). Величина целевого показателя может быть связана с правилом отбора наблюдений, если существуют ненаблюдаемые факторы, которые оказывают влияние как на целевой показатель, так и на вероятность того, что наблюдение попадет в выборку. В примере с заработной платой таким ненаблюдаемым фактором может выступать желание сделать карьеру.

Игнорирование селективности имеющихся в распоряжении исследователя выборок может приводить к серьезному смещению результатов при регрессионном оценивании.

Д. Хекманом был предложен метод, учитывающий возникающее смещение (Heckman, 1979). Он рассматривал случай, когда существует одно условие отбора и одно целевое уравнение. Однако некоторые задачи требуют рассмотрения более сложных селективных

[1] **Коссова Елена Владимировна** — Национальный исследовательский университет «Высшая школа экономики», Москва; ekossova@hse.ru.

Потанин Богдан Станиславович — Национальный исследовательский университет «Высшая школа экономики», Москва; bpotanin@hse.ru.

механизмов. Так, например, уравнение отбора может быть не бинарным, а порядковым (Jimenez, Kugler, 1987; Vella, 1993), непрерывным (Garen, 1984) или категориальным (Jeffrey, McFadden, 1984). Последний случай эквивалентен ситуации, когда отбор наблюдения определяется несколькими бинарными правилами, т. е. значением многомерной бинарной переменной.

Данная работа является обобщением классического подхода: рассматривается несколько условий отбора и несколько форм целевого уравнения. Задача решается в максимально общей постановке. Какое именно целевое уравнение имеет место для изучаемого показателя, определяется комбинацией правил отбора. В частном случае, для некоторых из них наблюдение целевого показателя может отсутствовать. Также могут отсутствовать наблюдения по некоторым комбинациям самих правил, что соответствует схеме последовательного принятия решений. Корреляции правил отбора друг с другом (тетрахорические корреляции) и с основными уравнениями являются дополнительными параметрами модели, что позволяет по результатам оценивания проверять гипотезы о характере зависимости случайных ошибок уравнений.

Из-за сложностей реализации метода максимального правдоподобия, работы, рассматривающие многомерный селективный механизм, встречаются редко и ограничиваются двумерным случаем и предположением о независимости правил отбора (Vella, 1998; Cinzia, 2009).

Полученные результаты являются новыми, поскольку в литературе, посвященной данному классу моделей, до сих пор не рассматривался случай произвольного числа уравнений отбора и форм основного уравнения. Алгоритм двухшаговой процедуры в случае более чем одного уравнения отбора также предлагается впервые.

1. Метод Хекмана

Используемая в данной работе система обозначений описана в Приложении.

Формально, модель Хекмана имеет следующий вид:

$$y_i^* = x_i'\beta + \varepsilon_i,$$

$$z_i^* = w_i'\gamma + u_i,$$

$$z_i = \begin{cases} 1, & \text{если } z_i^* \geq 0, \text{ т.е. } u_i \geq -w_i'\gamma, \\ -1, & \text{если } z_i^* < 0, \text{ т.е. } u_i < -w_i'\gamma. \end{cases}$$

$$(\varepsilon_i, u_i) \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}\right),$$

$$y_i = \begin{cases} y_i^*, & \text{если } z_i = 1, \\ \text{не наблюдаем}, & \text{если } z_i = -1, \end{cases} \qquad i \in \{1, \ldots, n\},$$

где y_i относится к наблюдаемой заработной плате индивида i, а z_i — бинарная переменная на трудоустройство, принимающая единичное значение для тех, кто работает, и минус один иначе; x_i и w_i — векторы значений объясняющих факторов, β и γ — векторы их коэффи-

циентов, ε_i и u_i — случайные ошибки, n — число наблюдений. Число компонент векторов x_i и β, w_i и γ определяется числом независимых переменных, включенных в уравнения.

Данную модель можно оценить методом максимального правдоподобия или с помощью двухшаговой процедуры, предложенной Хекманом в работе (Heckman, 1976).

Двухшаговая процедура. При неслучайном отборе математическое ожидание зависимой переменной имеет вид

$$\begin{aligned} \mathrm{E}(y_i) &= \mathrm{E}(y_i^* \mid z_i = 1) = \mathrm{E}(y_i^* \mid z_i^* \geq 0) = \\ &= x_i'\beta + \mathrm{E}(\varepsilon_i \mid u_i \geq -w_i'\gamma) = \\ &= x_i'\beta + \rho\sigma \frac{f(w_i'\gamma)}{F(w_i'\gamma)} = x_i'\beta + \rho\sigma\lambda_i, \end{aligned} \tag{1}$$

где $\lambda_i = \lambda(w_i'\gamma)$, $\lambda(x) = f(x)/F(x) = \dfrac{dF(x)}{dx}\Big/F(x)$ — обратное отношение Миллса, а f и F — функции плотности и распределения стандартного нормального закона соответственно.

Представим y_i в виде $y_i = x_i'\beta + \rho\sigma\lambda_i + v_i$, где v_i — случайная ошибка с нулевым математическим ожиданием и дисперсией, равной условной дисперсии $\mathrm{D}(y_i^* \mid z_i^* \geq 0)$.

Для получения состоятельных оценок параметров предлагается на первом шаге, используя пробит модель, оценить γ и вычислить оценки λ_i, а на втором шаге методом наименьших квадратов (МНК) оценить β и $\rho\sigma$.

При проверке гипотез о коэффициентах необходимо учитывать, что распределение случайной ошибки v является гетероскедастичным. Более того, поскольку на втором шаге вместо истинного обратного отношения Миллса λ используется его оценка, ковариационная матрица оценок коэффициентов должна быть скорректирована с поправкой на зависимость наблюдений.

2. Обобщение метода Хекмана

2.1. Обобщенная модель Хекмана

Предположим, что существует m правил отбора наблюдений целевого показателя, задаваемых значениями бинарных переменных z_{si}, $s \in \{1,...,m\}$. Если для наблюдения i правило s выполнено, бинарная переменная z_{si} принимает значение 1, если не выполнено, то –1. Число всех возможных комбинаций правил составляет 2^m. Перенумеруем те комбинации правил, для которых доступны значения основной переменной. Будем считать, что всего таких комбинаций r $(1 \leq r \leq 2^m)$. Тогда все наблюдения, за исключением тех, для которых значения основной переменной недоступны, разбиваются на r групп в зависимости от имевшей место комбинации правил. Будем считать, что номер группы совпадает с номером соответствующей ей комбинации. Оставшиеся наблюдения образуют группу, которой присвоим нулевой номер. Данная группа может быть пустой, если значения основной переменной доступны для всех наблюдений. Для каждого наблюдения i определим значение функции индекса $g_i = g(z_{1i},...,z_{mi}) = s$ $(0 \leq s \leq r)$, равное номеру группы, к которой относится наблюдение i.

Рассмотрим ситуацию, когда в регрессионном уравнении, связывающем зависимую переменную y_i^* и набор объясняющих переменных x_i, вектор параметров и случайная ошибка ε_i зависят от группы g_i, в которую попало i-е наблюдение, т. е. уравнение имеет вид $y_i^* = x_i'\beta_{g_i} + \varepsilon_{i,g_i}$.

Уравнения отбора и определяемое ими основное уравнение могут быть зависимыми из-за коррелированности входящих в них случайных ошибок, что и порождает проблему селективности. Предполагаем, что совместное распределение случайных ошибок является нормальным.

Формально модель может быть записана следующим образом:

$$y_i^* = x_i'\beta_{g_i} + \varepsilon_{i,g_i},$$

$$z_{si}^* = w_{si}'\gamma_s + u_{si}, \quad s \in \{1,...,m\},$$

$$z_{si} = \begin{cases} 1, & \text{если } z_{si}^* \geq 0, \text{ т.е. } u_{si} \geq -w_{si}'\gamma_s, \\ -1, & \text{если } z_{si}^* < 0, \text{ т.е. } u_{si} < -w_{si}'\gamma_s. \end{cases},$$

$$g_i = g(z_{1i},...,z_{mi}), \quad g_i \in \{0,...,r\},$$
$$u_i' = (u_{1i},...,u_{mi})',$$

$$(\varepsilon_{i,g_i}, u_i')' \sim N\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \sum_{g_i}\right), \quad \text{где} \quad \sum_{g_i} = \begin{bmatrix} \sigma_{g_i}^2 & \rho_{1,g_i}\sigma & \rho_{2,g_i}\sigma & \cdots & \rho_{m,g_i}\sigma \\ \rho_{1,g_i}\sigma & 1 & \rho_{12} & \cdots & \rho_{1m} \\ \rho_{2,g_i}\sigma & \rho_{12} & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{m,g_i}\sigma & \rho_{1m} & \cdots & \cdots & 1 \end{bmatrix};$$

$$y_i = \begin{cases} y_i^*, & \text{если } g_i > 0, \\ \text{не наблюдаем}, & \text{если } g_i = 0, \end{cases} \quad i \in \{1,...,n\}.$$

Как и в одномерном случае, x_i и w_{is} — векторы значений объясняющих факторов для наблюдения i основного уравнения и уравнения отбора s, β_{g_i} и γ_s — векторы коэффициентов, ε_{i,g_i} и u_{si} — случайные ошибки, n — число наблюдений.

В рамках предложенной модели математическое ожидание наблюдаемого целевого показателя является условным:

$$\begin{aligned} \mathrm{E}(y_i) &= \mathrm{E}(y_i^* \mid z_{1i},...,z_{mi}) = \mathrm{E}(y_i^* \mid -z_{1i}u_{1i} \leq z_{1i}w_{1i}'\gamma_1,...,-z_{mi}u_{mi} \leq z_{mi}w_{mi}'\gamma_m) = \\ &= x_i'\beta_{g_i} + \mathrm{E}(\varepsilon_{i,g_i} \mid -z_{1i}u_{1i} \leq z_{1i}w_{1i}'\gamma_1,...,-z_{mi}u_{mi} \leq z_{mi}w_{mi}'\gamma_m). \end{aligned} \tag{2}$$

Обозначим

$$z_i = \begin{bmatrix} z_{1i} \\ \vdots \\ z_{mi} \end{bmatrix}, \ \tilde{z}_i = \begin{bmatrix} \tilde{z}_{1i} \\ \vdots \\ \tilde{z}_{mi} \end{bmatrix} = \begin{bmatrix} z_{1i}w_{1i}'\gamma_1 \\ \vdots \\ z_{mi}w_{mi}'\gamma_m \end{bmatrix}, \ \tilde{u}_i = \begin{bmatrix} -z_{1i}u_{1i} \\ \vdots \\ -z_{mi}u_{mi} \end{bmatrix} \text{ и } \rho_{g_i} = \begin{bmatrix} \rho_{1,g_i} \\ \vdots \\ \rho_{m,g_i} \end{bmatrix}.$$

В этих обозначениях (2) принимает вид:

$$\mathrm{E}(y_i^* \mid z_{1i},...,z_{mi}) = x_i'\beta_{g_i} + \mathrm{E}(\varepsilon_{i,g_i} \mid \tilde{u}_{1i} \le \tilde{z}_{1i},...,\tilde{u}_{mi} \le \tilde{z}_{mi}). \quad (3)$$

Введем обозначения для отношения компонент градиента ∇, гессиана H и тензора D^3, состоящего из третьих производных функции распределения нормального случайного вектора $\tilde{u}_i$ к ее значению[2]:

$$\lambda^{\tilde{u}_i}(x) = \frac{\nabla F_{\tilde{u}_i}(x)}{F_{\tilde{u}_i}(x)}, \quad (4)$$

$$\Lambda^{\tilde{u}_i}(x) = \frac{H(F_{\tilde{u}_i}(x))}{F_{\tilde{u}_i}(x)}, \quad (5)$$

$$\Theta^{\tilde{u}_i}(x) = \frac{D^3(F_{\tilde{u}_i}(x))}{F_{\tilde{u}_i}(x)}. \quad (6)$$

Элементы данных отношений представляют собой не что иное, как значения маргинальных и совместных функций плотности усеченного сверху в точке $x = (x_1,...,x_m)'$ нормального случайного вектора $\tilde{u}_i$ (см. Приложение, п. 3.3, формулы (26)–(28)). Далее для j-й компоненты вектора $\lambda^{\tilde{u}_i}(\tilde{z}_i)$, $j \in \{1,...,m\}$, являющейся j-й маргинальной плотностью, будем использовать обозначение $\lambda_j^{\tilde{u}_i}(\tilde{z}_i)$, для элемента (k,j) матрицы $\Lambda^{\tilde{u}_i}(\tilde{z}_i)$ — обозначение $\Lambda_{kj}^{\tilde{u}_i}(\tilde{z}_i)$ и положим $\lambda_s^{\tilde{u}}(\tilde{z}) = (\lambda_s^{\tilde{u}_1}(\tilde{z}_1),...,\lambda_s^{\tilde{u}_{n_{\tilde{u}}}}(\tilde{z}_{n_c}))'$, $\Lambda_{kj}^{\tilde{u}}(\tilde{z}) = (\Lambda_{kj}^{\tilde{u}_1}(\tilde{z}_1),...,\Lambda_{kj}^{\tilde{u}_{n_{\tilde{u}}}}(\tilde{z}_{n_c}))'$.

Второе слагаемое в выражении (3) представляет собой условное математическое ожидание компоненты многомерного нормального вектора. Используя лемму 1 (см. Приложение, п. 3.3, формула (29)), представим условное математическое ожидание (3) в виде

$$\mathrm{E}(y_i \mid z_{1i},...,z_{mi}) = x_i'\beta_{g_i} + \sum \sigma_{g_i}\rho_{j,g_i}\lambda_j\,(z_i\ \ z_{ji} \quad (7)$$

Таким образом, на математическое ожидание наблюдаемых значений y_i оказывают влияние как переменные, входящие в основное уравнение, так и переменные, входящие в уравнения отбора наблюдений w_{si}, $s \in \{1,...,m\}$. Уравнение (7) является обобщением уравнения (1).

Аналогично, в силу леммы 1 справедливо следующее выражение для условной дисперсии:

$$\mathrm{D}(y_i \mid z_{1i},...,z_{mi}) = \sigma_{g_i}^2 \left(1 - \sum_{k=1}^{m} \rho_{k,g_i}^2 \tilde{z}_{ki}\lambda_k^{\tilde{u}_i}(\tilde{z}_i) + \right.$$
$$\left. + \sum_{k=1}^{m} z_{ki}\rho_{k,g_i} \sum_{j \ne k} z_{ji}(\rho_{j,g_i} - \rho_{kj}\rho_{k,g_i})\Lambda_{kj}^{\tilde{u}_i}(\tilde{z}_i) - \left(\sum_{k=1}^{m} z_{ki}\rho_{k,g_i}\lambda_k^{\tilde{u}_i}(\tilde{z}_i)\right)^2\right). \quad (8)$$

[2] Отношение (4) можно считать обобщенным обратным отношением Миллса.

2.2. Оценивание обобщенной модели Хекмана методом максимального правдоподобия

Вклад *i*-го наблюдения в функцию правдоподобия может быть записан в виде

$$\mathrm{P}(y_i, z_{1i}, ..., z_{mi}) = \begin{cases} \mathrm{P}(\tilde{u}_{1i} \le \tilde{z}_{1i}, ..., \tilde{u}_{mi} \le \tilde{z}_{mi} \mid \varepsilon_{g_i} = y_i - x_i'\beta_{g_i}) f_{\epsilon_{g_i}}(y_i - x_i'\beta_{g_i}), & \text{если } g_i > 0, \\ \mathrm{P}(\tilde{u}_{1i} \le \tilde{z}_{1i}, ..., \tilde{u}_{mi} \le \tilde{z}_{mi}), & \text{если } g_i = 0. \end{cases}$$

Соответственно, функция правдоподобия имеет вид

$$L(\sum\nolimits_1, ..., \sum\nolimits_r, \beta_1, ..., \beta_r, \gamma_1, ..., \gamma_m) = \prod_{i=1}^{n} r_i,$$

$$r_i = \begin{cases} F_{\tilde{u}_i \mid \varepsilon_{g_i} = y_i - x_i\beta_{g_i}}(z_{1i}w_{1i}'\gamma_1, ..., z_{mi}w_{mi}'\gamma_m) f_{\varepsilon_{g_i}}(y_i - x_i'\beta_{g_i}), & \text{если } g_i > 0, \\ F_{\tilde{u}_i}(z_{1i}w_{1i}'\gamma_1, ..., z_{mi}w_{mi}'\gamma_m), & \text{если } g_i = 0. \end{cases}$$

Здесь $f_{\varepsilon_{g_i}}$ — плотность распределения случайной ошибки ε_{g_i}, которая в силу предположения модели является нормальной с нулевым математическим ожиданием и дисперсией $\sigma^2_{g_i}$. $F_{\tilde{u}_i}$ — совместная функция распределения *m* компонент случайного вектора $\tilde{u}_i$, а $F_{\tilde{u}_i \mid \varepsilon_{g_i} = y_i - x_i\beta_{g_i}}$ — его условная функция распределения, которые, по предположению модели, являются нормальными (см. Приложение, п. 2.1). Отметим, что для вычисления $F_{\tilde{u}_i}$ и $F_{\tilde{u}_i \mid \varepsilon_{g_i} = y_i - x_i\beta_{g_i}}$ достаточно знать соответствующую ковариационную матрицу и математическое ожидание. Ковариационная матрица и математическое ожидание $\tilde{u}_i$ идентичны u_i, за исключением изменения знака некоторых компонент, в зависимости от соблюдения правил отбора, т. е. значений z (см. Приложение, п. 2.2). Для вычисления ковариационной матрицы и математического ожидания $F_{\tilde{u}_i \mid \varepsilon_{g_i} = y_i - x_i\beta_{g_i}}$ используются свойства условного многомерного нормального распределения (см. Приложение, п. 2.1).

Таким образом, изменение знаков компонент u_i и условие на ε_{g_i} влияют лишь на параметры распределения (математическое ожидание и ковариационную матрицу) $\tilde{u}_i$, которое по-прежнему остается нормальным.

Отметим, что случай с последовательным применением правил соответствует ситуации, когда некоторые комбинации правил отбора не наблюдаемы, т. е. селективность присутствует и в правилах отбора. При этом функция правдоподобия будет выглядеть аналогично, с той поправкой, что вклад наблюдений, для которых определены не все значения правил, будет записываться как совместная вероятность значений бинарных переменных, соответствующих оставшимся (наблюдаемым) правилам.

Оценивание данной функции правдоподобия является непростой вычислительной задачей, сложность которой значительно возрастает с увеличением числа правил *m*. По этой причине может быть полезна процедура двухшагового оценивания, обобщающая подход, предложенный в (Heckman, 1979). Данная процедура позволяет получить состоятельные асимптотически нормальные оценки, проигрывающие по эффективности оценкам максимального правдоподобия. Помимо самостоятельной ценности, полученные с помощью двухшаговой процедуры оценки могут быть использованы в качестве начальных точек при нахождении максимума функции правдоподобия.

2.3. Двухшаговая процедура оценивания обобщенной модели Хекмана

Уравнение (7) можно было бы оценить с помощью МНК с поправкой на гетероскедастичность (8), если бы были известны значения обобщенного обратного отношения Миллса. Согласно теореме о двухшаговой процедуре оценивания (Murphy, Topel, 2002), можно получить состоятельные оценки параметров, если оценивать (7) с помощью МНК, используя вместо обобщенных отношений Миллса их состоятельные оценки. В свою очередь, состоятельные оценки обобщенных отношений Миллса можно получить, оценив систему бинарных уравнений, задающих правила отбора наблюдений. Таким образом, оценивание разбивается на два этапа.

1. На первом шаге оценивается система бинарных уравнений отбора, находятся оценки γ_s и ρ_{sk}, $s,k \in \{1,...,m\}$. Рассчитываются оценки обобщенных отношений Миллса[3].

2. На втором шаге, как и в процедуре Хекмана, с помощью МНК оцениваются уравнения (7) для каждого значения $g_i = c$, $c \in \{1,...,r\}$, по n_c наблюдениям, где n_c — число наблюдений, относящихся к группе c, т. е. $\{i : g_i = c\}$:

$$y_i = x_i'\beta_c + \sum_{j=1}^{m} \beta_{\lambda_j} (\lambda_j^{\tilde{u}_i}(\tilde{z}_i) z_{ji}) + v_i, \text{ где } \beta_{\lambda_j} = \sigma_c \rho_{j,c},$$

с заменой обратных отношений Миллса $\lambda_j^{\tilde{u}_i}(\tilde{z}_i)$ их состоятельными оценками $\lambda_j^{\tilde{u}_i}(\hat{\tilde{z}}_i)$, полученными на предыдущем шаге.

Обозначим через e вектор остатков регрессии второго шага, тогда

$$\operatorname*{plim}_{n_c \to \infty} \left(\frac{1}{n_c} \sum_{i=1}^{n_c} \mathrm{D}(y_i \mid z_{1i},...,z_{mi}) - \frac{1}{n_c} e'e \right) = 0. \tag{9}$$

В соответствии с формулами (8) и (9), состоятельная оценка дисперсии может быть записана в виде

$$\hat{\sigma}_c^2 =$$

$$= \frac{1}{n_c} \left(e'e + \sum_{i=1}^{n_c} \left(\sum_{k=1}^{m} \widehat{\beta_{\lambda_k}}^2 \hat{\tilde{z}}_{ki} \hat{\lambda}_k^{\tilde{u}_i}(\hat{\tilde{z}}_i) - \sum_{k=1}^{m} \widehat{\beta_{\lambda_k}} z_{ki} \sum_{j \neq k} z_{ji} (\widehat{\beta_{\lambda_j}} - \hat{\rho}_{kj} \widehat{\beta_{\lambda_k}}) \hat{\Lambda}_{kj}^{\tilde{u}_i}(\hat{\tilde{z}}_i) + \left(\sum_{k=1}^{m} z_{ki} \widehat{\beta_{\lambda_k}} \hat{\lambda}_k^{\tilde{u}_i}(\hat{\tilde{z}}_i) \right)^2 \right) \right). \tag{10}$$

Зная оценки $\hat{\beta}_{\lambda_k}$ и $\hat{\sigma}_c^2$, можно получить состоятельные оценки коэффициентов корреляции:

$$\hat{\rho}_{k,c} = \hat{\beta}_{\lambda_k} \big/ \hat{\sigma}_c, \quad 1 \le k \le m.$$

2.4. Состоятельная оценка асимптотической ковариационной матрицы

Используя предложенную Хекманом методику для классической двухшаговой процедуры, получим состоятельную оценку асимптотической ковариационной матрицы МНК-оценок параметров уравнения (7). Введем диагональную матрицу Δ размерности $n_c \times n_c$,

[3] В случае если некоторые комбинации правил не наблюдаются, оценивается система бинарных уравнений, учитывающая селективное смещение.

такую, что $\Delta_{ii} = 1 - \sigma_c^{-2}\mathrm{D}(y_i \mid z_{1i},...,z_{mi})$. Через $\tilde{X} = \begin{bmatrix} X & \lambda_1^{\tilde{u}}(\tilde{z}) \circ z_1 & ... & \lambda_m^{\tilde{u}}(\tilde{z}) \circ z_m \end{bmatrix}$ обозначим расширенную матрицу переменных размера $n_c \times (\mathrm{rang}(X) + m)$, где $\mathrm{rang}(X)$ определяется числом независимых переменных основного уравнения, а символ $\circ$ означает покомпонентное умножение векторов. Тогда, в силу гетероскедастичности ошибок (8), ковариационная матрица вектора оценок коэффициентов $\tilde{\beta}' = \begin{bmatrix} \beta_c' & \beta_{\lambda_1} & ... & \beta_{\lambda_m} \end{bmatrix}$ имеет вид

$$\mathrm{Cov}(\hat{\tilde{\beta}}) = \sigma^2 (\tilde{X}'\tilde{X})^{-1} \tilde{X}'(I - \Delta)\tilde{X}(\tilde{X}'\tilde{X})^{-1}.$$

Однако, поскольку на втором шаге процедуры оценивается регрессия (7), в которой истинные значения γ и ρ заменены их состоятельными оценками $\hat{\gamma}$ и $\hat{\rho}$, ковариационная матрица должна быть скорректирована:

$$\mathrm{Cov}\left(X\beta_c + \sum_{k=1}^{m} \sigma_c \rho_{k,c} (\lambda_k^{\tilde{u}}(\hat{\tilde{z}}) \circ z_k) + v \right) = \sigma_c^2 (I - \Delta) + \sigma_c^2 \mathrm{Cov}\left(\sum_{k=1}^{m} \rho_{k,c} z_k \circ \lambda_k^{\tilde{u}}(\hat{\tilde{z}}) \right). \qquad (11)$$

Введем матрицу Якоби рассматриваемого вектора функций от коэффициентов при переменных (γ) и коэффициентов корреляций (ρ) m уравнений отбора. Для определенности будем считать, что все m уравнений имеют одну и ту же матрицу переменных W. Тогда матрица Якоби $\Gamma = J\left(\sum_{k=1}^{m} \rho_{k,c} z_k \circ \lambda_k^{\tilde{u}}(\tilde{z}) \right)$ имеет размер $n_c \times \left(\mathrm{rang}(W) \cdot m + (m^2 - m)/2 \right)$.

Обозначим через $\Gamma_{\gamma_{ks}}$ столбец Γ, относящийся к s-му коэффициенту, $1 \le s \le \mathrm{rang}(W)$, k-го, $1 \le k \le m$, уравнения отбора, а через $\Gamma_{\rho_{lk}}$ — столбец, относящийся к коэффициенту корреляции между ошибками l-го и k-го селективных уравнений. Всего различных коэффициентов корреляции $(m^2 - m)/2$. Тогда, дифференцируя линейную комбинацию маргинальных плотностей усеченного многомерного нормального распределения по указанным параметрам, можно получить столбцы якобиана.

Для i-го элемента столбца $\Gamma_{\gamma_{ks}}$, используя формулы (31) и (32) (см. Приложение, п. 3.4), получаем:

$$\left(\Gamma_{\gamma_{ks}}\right)_i = W_{ik}\left(\sum_{j \ne k} \rho_{j,c} z_{ji} z_{ki} \left(\Lambda_{kj}^{\tilde{u}_i}(\tilde{z}_i) - \lambda_k^{\tilde{u}_i}(\tilde{z}_i)\lambda_j^{\tilde{u}_i}(\tilde{z}_i) \right) - \rho_{k,c} \left(\tilde{z}_{ki}\lambda_k^{\tilde{u}_i}(\tilde{z}_i) + (\lambda_k^{\tilde{u}_i}(\tilde{z}_i))^2 + \sum_{j \ne k} \rho_{kj} \Lambda_{kj}^{\tilde{u}_i}(\tilde{z}_i) z_{ki} z_{ji} \right) \right).$$

Для столбцов, соответствующих коэффициентам корреляции ρ_{lk}, применяя формулы (33) и (34) (см. Приложение, п. 3.4), имеем:

$$\begin{aligned} \left(\Gamma_{\rho_{lk}}\right)_i = {} & z_{ki}\rho_{l,c}\left(\left(\frac{z_{li}z_{ki}\rho_{lk}\tilde{z}_{ki} - \tilde{z}_{li}}{1-\rho_{lk}^2} - \lambda_l^{\tilde{u}_i}(\tilde{z}_i) \right) \Lambda_{lk}^{\tilde{u}_i}(\tilde{z}_i) + \sum_{j \ne l,k} \frac{z_{li}z_{ji}(\rho_{lk}\rho_{kj} - \rho_{lj})}{1-\rho_{lk}^2} \Theta_{ljk}^{\tilde{u}_i}(\tilde{z}_i) \right) + \\ & + z_{li}\rho_{k,c}\left(\left(\frac{z_{li}z_{ki}\rho_{lk}\tilde{z}_{li} - \tilde{z}_{ki}}{1-\rho_{lk}^2} - \lambda_k^{\tilde{u}_i}(\tilde{z}_i) \right) \Lambda_{lk}^{\tilde{u}_i}(\tilde{z}_i) + \sum_{j \ne l,k} \frac{z_{ki}z_{ji}(\rho_{lk}\rho_{lj} - \rho_{kj})}{1-\rho_{lk}^2} \Theta_{ljk}^{\tilde{u}_i}(\tilde{z}_i) \right) + \\ & + \sum_{j \ne l,k} z_{ji}z_{li}z_{ki}\rho_{j,c}\left(\Theta_{ljk}^{\tilde{u}_i}(\tilde{z}_i) - \lambda_j^{\tilde{u}_i}(\tilde{z}_i)\Lambda_{lk}^{\tilde{u}_i}(\tilde{z}_i) \right). \end{aligned}$$

Наконец, обозначим через $\hat{\Gamma}$ матрицу Γ, рассчитанную при $\hat{\tilde{z}}, \hat{\rho}$ и $\hat{\sigma}$.

Для получения асимптотической оценки второго слагаемого в (11) воспользуемся многомерным дельта-методом:

$$\widehat{\text{As.Cov}}\left(\sigma_c \sum_{k=1}^{m} z_k \rho_k \hat{\lambda}_k^{\tilde{u}}(\hat{\tilde{z}})\right) = \sigma_c^2 \hat{\Gamma} \widehat{\text{As.Cov}}\left(\begin{bmatrix}\hat{\gamma}\\ \hat{\rho}\end{bmatrix}\right)\hat{\Gamma}',$$

где $\widehat{\text{As.Cov}}\left(\begin{bmatrix}\hat{\gamma}\\ \hat{\rho}\end{bmatrix}\right)$ — оценка асимптотической ковариационной матрицы параметров уравнений отбора, т. е. матрица размера $\left(\text{rang}(W)\cdot m + (m^2 - m)/2\right) \times \left(\text{rang}(W)\cdot m + (m^2 - m)/2\right)$, порядок столбцов которой соответствует порядку столбцов матрицы Якоби.

Таким образом, оценка асимптотической ковариационной матрицы оценок коэффициентов второго шага обобщенной процедуры Хекмана имеет вид[4]

$$\widehat{\text{As.Cov}}(\hat{\tilde{\beta}}) = \hat{\sigma}_c^2 (\tilde{X}'\tilde{X})^{-1} \tilde{X}' \left[(I - \hat{\Delta}) + \hat{\Gamma} \widehat{\text{As.Cov}}\left(\begin{bmatrix}\hat{\gamma}\\ \hat{\rho}\end{bmatrix}\right)\hat{\Gamma}' \right] \tilde{X} (\tilde{X}'\tilde{X})^{-1}.$$

2.5. Предельные эффекты

Для тех независимых переменных, которые входят только в основное уравнение и отсутствуют в уравнениях отбора, предельный эффект, очевидно, равен значению соответствующего коэффициента. Однако для тех независимых переменных, которые входят как в основное уравнение, так и в одно или несколько уравнений отбора, предельный эффект отражает совокупное влияние выбранной переменной как на целевой показатель (коэффициент бета), так и на величину смещения. Рассмотрим произвольную переменную ψ, которая входит хотя бы в одно уравнение отбора[5] и имеет коэффициент β_ψ в основном уравнении. Ее предельный эффект равен

$$\frac{\partial E(y_i \mid z_{1i}, ..., z_{mi})}{\partial \psi} = \beta_\psi + \sigma \sum_{j=1}^{m} \rho_j z_{ji} \frac{\partial \lambda_j^{\tilde{u}_i}(\tilde{z}_i)}{\partial \psi}. \tag{12}$$

Обозначим γ_ψ^j — коэффициент при ψ в j-м уравнении, а $\Upsilon_\psi^j(\tilde{z}_i) = z_{ji}\sigma_{g_i}\rho_{j,g_i} \cdot \partial \lambda_j^{\tilde{u}_i}(\tilde{z}_i)/\partial\psi$ — часть предельного эффекта ψ, связанную с j-м уравнением отбора. Υ_ψ^j может быть рассчитана дифференцированием плотности усеченного многомерного нормального распределения по формулам (31) и (32):

$$\Upsilon_\psi^j(\tilde{z}_i) = z_{ji}\sigma_{g_i}\rho_{j,g_i} \sum_{k=1}^{m} z_{ki}\gamma_\psi^k \frac{\partial \lambda_j^{\tilde{u}_i}(\tilde{z}_i)}{\partial \tilde{z}_k} = z_{ji}\sigma_{g_i}\rho_{j,g_i} \left(\sum_{k\neq j} z_{ki}\gamma_\psi^k \left(\Lambda_{jk}^{\tilde{u}_i}(\tilde{z}_i) - \lambda_j^{\tilde{u}_i}(\tilde{z}_i)\lambda_k^{\tilde{u}_i}(\tilde{z}_i) \right) \right) -$$

$$- \gamma_\psi^j \sigma_{g_i}\rho_{j,g_i} \left(\tilde{z}_{ji}\lambda_j^{\tilde{u}_i}(\tilde{z}_i) + (\lambda_j^{\tilde{u}_i}(\tilde{z}_i))^2 + \sum_{k\neq j} z_{ji}z_{ki}\rho_{jk}\Lambda_{jk}^{\tilde{u}_i}(\tilde{z}_i) \right).$$

[4] Данная процедура оценивания ковариационной матрицы для одномерного случая впервые была предложена в (Heckman, 1979).

[5] Будем рассматривать общую совокупность объясняющих переменных. Если в какое-то уравнение рассматриваемая переменная не входит, коэффициент при ней считается равным нулю.

Подставляя полученный результат в (12), имеем:

$$\frac{\partial E(y_i \mid z_{1i},...,z_{mi})}{\partial \psi} = \beta_\psi + \sum_{j=1}^{m} \mathrm{Y}_\psi^j(\tilde{z}_i).$$

3. Оценка качества обобщенного метода Хекмана

3.1. Обзор исследований свойств и качества оценок модели Хекмана

Подробный обзор исследований, посвященных изучению свойств и качества оценок методом Хекмана, можно найти в (Puhani, 2000). Остановимся на основных выводах.

Во-первых, многие исследователи отмечают, что превосходство метода Хекмана над МНК, а также метода максимального правдоподобия (ММП) над двухшаговой процедурой растет пропорционально величине корреляции между ошибками в оцениваемых уравнениях (Nelson, 1984; Stolzenberg, Relles, 1990; Nawata, 1993, 1994). Однако в некоторых работах не удавалось обнаружить соответствующей зависимости, или же ее надежность оказывалась спорной (Hay et al., 1987; Manning et al., 1987; Zuehlke, Zeman, 1990).

Во-вторых, когда ошибки не были распределены нормально, выявить преимущество одного из методов не удавалось (Hay et al., 1987; Paarsch, 1984), а также наблюдалось превосходство в точности оценок непараметрических подходов (Marchenko, Genton, 2012; Chavent et al., 2010).

В-третьих, двухшаговая процедура начинает работать хуже, когда отсутствуют так называемые ограничения исключений (exclusion restrictions) (Manning et al., 1987). Под этим подразумевается высокая степень схожести набора регрессоров в обоих уравнениях, т. е. когда мало таких переменных, которые есть в одном из уравнений, но отсутствуют в другом. Это приводит к тому, что из-за близости обратного отношения Миллса к линейной функции в основном уравнении появляется регрессор, в некоторой степени коллинеарный остальным независимым переменным, что приводит к снижению эффективности оценок (Zuehlke, Zeman, 1990). Также в одном из исследований отмечается, что если уникальная для уравнения отбора переменная коррелирует с независимой переменной только из основного уравнения, то это приводит к относительно более качественным оценкам ММП (Rendtel, 1992).

Наконец, один из исследователей по результатам двух работ пришел к выводу, что при высоких уровнях корреляции между ошибками (более 0.9) оценки двухшагового метода ведут себя очень нестабильно (Nawata, 1993, 1994).

В плане организации симуляций, исследователи использовали несколько фиксированных уровней параметров. Например, что репрезентативно в отношении остальных работ в этой области, в одном из исследований результаты рассматривались при уровнях корреляции ошибок 0.5 и 0.9 и доле нецензурированных наблюдений — 25, 50 и 75% (Manning et al., 1987).

Серьезным недостатком упомянутых выше исследований являлось то, что анализ проводился на очень маленьких выборках. В некоторых из них выборки составляли всего 50–200 наблюдений (Paarsch, 1984; Zuehlke, Zeman, 1990; Nawata, 1993, 1994). Это могло приводить к существенному снижению эффективности ММП оценок ввиду отсутствия надлежащей асимптотики.

В данном исследовании осуществлена попытка преодолеть указанный недостаток за счет использования выборок большего объема, в диапазоне от 300 до 5000 наблюдений на 1000

симуляций. Также используется альтернативный подход к вариации параметров: она происходит случайным образом в каждой из симуляций.

3.2. Анализ обобщенной модели Хекмана на симулированных данных

Рассмотрим частный случай обобщенной модели Хекмана, когда зависимая переменная наблюдается при условии выполнения двух правил отбора.

Обозначим $NR(\mu,\Sigma,n)=\begin{bmatrix} X_1 \sim N(\mu,\Sigma) \\ \vdots \\ X_n \sim N(\mu,\Sigma) \end{bmatrix}$ — случайную выборку из независимых m-мерных нормальных случайных векторов X_i, $i \in \{1,...,n\}$ с математическим ожиданием μ и ковариационной матрицей Σ.

Значения m объясняющих переменных для каждого наблюдения генерируются из многомерного стандартного нормального закона с недиагональной корреляционной матрицей.

Рассмотрим случай, когда по основной переменной доступны лишь те наблюдения, по которым наблюдаются оба признака. Во всех остальных случаях переменная не наблюдается. Формально это можно записать как

$$g_i(z_{1i},z_{2i})=\begin{cases}1, & \text{если } z_{1i}=z_{2i}=1,\\ 0, & \text{в противном случае,}\end{cases} \qquad i \in \{1,...,n\}.$$

Случайные ошибки симулируются следующим образом:

$$\zeta=(\varepsilon,u_1,u_2)=NR\left(\begin{bmatrix}0\\0\\0\end{bmatrix},\begin{bmatrix}\sigma^2 & \rho_1\sigma & \sigma\rho_2\\ \rho_1\sigma & 1 & \rho_0\\ \rho_2\sigma & \rho_0 & 1\end{bmatrix},n\right).$$

В каждой симуляции коэффициенты при переменных являются случайными величинами, равномерно распределенными в интервале от –1 до 1. Единственное исключение составляют константы в уравнениях отбора γ_{10} и γ_{20}, которые распределены в интервале от 0 до 1 с целью увеличения числа нецензурированных наблюдений в выборке, что соответствует структуре данных, часто встречающейся на практике. Генерация случайных ковариационных матриц ошибок и независимых переменных осуществляется с помощью метода, предложенного (Joe, 2006), где $\sigma \sim U(0,1)$. Следует отметить, что в результате применения данного метода ρ_0, ρ_1 и ρ_2 также оказались распределены равномерно, в интервале от –1 до 1. Симуляции выполняются для выборок объемом 300, 1000 и 5000 наблюдений. На каждую из этих выборок приходилось по 1000 симуляций.

Таким образом, в каждой симуляции на первом этапе случайным образом генерировались параметры модели и экзогенные переменные. На втором этапе, с их использованием, рассчитывались истинные значения y, z_1 и z_2. На третьем этапе оставлялись лишь те значения y, для которых $z_1=z_2=1$. Наконец, на основе полученных данных производилась оценка параметров модели при помощи предложенного и конкурирующих методов в лице МНК и классического метода Хекмана, игнорирующего второе уравнение отбора. Для каждого

объема выборки (300, 1000 и 5000 наблюдений) соответствующая процедура повторялась 1000 раз. Оцениваемая функция правдоподобия имела вид

$$L(\beta,\gamma_1,\gamma_2,\rho_0,\rho_1,\rho_2,\sigma)=\prod_{z_1=1,z_2=1}F_{-u_1,-u_2|\varepsilon=y_i-x_i\beta}(w_{1i}'\gamma_1,w_{2i}'\gamma_2)f_\varepsilon(y_i-x_i'\beta)$$

$$\prod_{z_1=1,z_2=-1}F_{-u_1,u_2}(w_{1i}'\gamma_1,-w_{2i}'\gamma_2)\prod_{z_1=-1,z_2=1}F_{u_1,-u_2}(-w_{1i}'\gamma_1,w_{2i}'\gamma_2)\prod_{z_1=-1,z_2=-1}F_{u_1,u_2}(-w_{1i}'\gamma_1,-w_{2i}'\gamma_2).$$

В качестве показателя, отражающего качество оценки коэффициента в i-й симуляции, будем использовать модуль отклонения

$$D_i=\left|\frac{\beta_{si}-\hat{\beta}_{si}}{\beta_{si}}\right|.$$

Если для оценки методом Q и для оценки методом W верно $D_{iQ}>D_{iW}$, то будем считать, что оценки метода W «победили». В качестве относительного отклонения оценки возьмем

$$\bar{D}=\frac{1}{1000}\sum_{i=1}^{1000}\left|\frac{\beta_{si}-\hat{\beta}_{si}}{\beta_{si}}\right|.$$

Аналогичным способом оценивалось и качество оценки σ.

3.3. Первая спецификация

$$y_i=\beta_0+\beta_1x_{1i}+\beta_2x_{2i}+\beta_3x_{3i}+\varepsilon_i,$$
$$z_{1i}=\gamma_{10}+\gamma_{11}x_{1i}+\gamma_{12}x_{4i}+u_{1i},$$
$$z_{2i}=\gamma_{20}+\gamma_{21}x_{2i}+\gamma_{22}x_{5i}+u_{2i},$$
$$i\in\{1,...,n\},\ \ n\in\{300,1000,5000\}.$$

Данная спецификация соответствует ситуации, когда каждое уравнение содержит хотя бы одну исключительную переменную. Уравнения отбора не имеют общих объясняющих переменных, основное уравнение имеет общую переменную с каждым из уравнений отбора.

Назовем обобщенные модели Хекмана, оцененные при помощи метода максимального правдоподобия и двухшаговой процедуры, «ММП» и «ДП» соответственно. Аналогичные модели, оцененные при помощи классического метода Хекмана с учетом лишь первого уравнения отбора z_1, обозначим «ММП-0» и «ДП-0». Наконец, модель, оцененную при помощи классического метода наименьших квадратов, обозначим «МНК».

Результаты симуляции (табл. 1) показывают, что лучшим способом оценки коэффициентов, независимо от объема выборки, является обобщенный метод Хекмана, реализованный при помощи метода максимального правдоподобия. Превосходство последнего над МНК растет по мере увеличения объемов выборки. Аналогичная картина наблюдается и в отношении ДП оценок, несколько уступающим МНК оценкам на маленькой выборке.

При условии двойного смещения оценки классического метода Хекмана оказались достаточно сильно смещенными, вследствие чего их доля побед над МНК незначительна, хоть и увеличивается по мере роста объемов выборки.

Полученные результаты (табл. 2) говорят об увеличении относительного преимущества ММП оценок над МНК оценками по мере роста модуля корреляции ошибок уравнений отбора с ошибкой основного уравнения[6]. Аналогичная картина наблюдалась и для коэффициентов при других независимых переменных, а также на малой и большой выборках и при различных интервалах ρ_0.

Таблица 1. Сравнение качества оценок коэффициентов при различных способах оценивания

Число наблюдений	β_0	β_1	β_2	β_3	σ	β_0	β_1	β_2	β_3	σ
	Доля побед ММП оценок над МНК					Доля побед ДП оценок над МНК				
300	0.558	0.506	0.534	0.530	0.611	0.428	0.458	0.481	0.508	0.459
1000	0.748	0.648	0.646	0.581	0.760	0.634	0.602	0.617	0.566	0.627
5000	0.843	0.786	0.781	0.733	0.875	0.787	0.749	0.758	0.724	0.814
	Среднее отклонение ММП оценок					Среднее отклонение ДП оценок				
300	0.761	0.323	0.291	0.260	0.129	1.177	0.414	0.356	0.282	0.293
1000	0.389	0.177	0.154	0.132	0.053	0.613	0.232	0.201	0.142	0.120
5000	0.193	0.074	0.074	0.060	0.025	0.305	0.091	0.089	0.061	0.044
	Доля побед ММП-0 оценок над МНК					Доля побед ДП-0 оценок над МНК				
300	0.489	0.469	0.489	0.501	0.545	0.465	0.441	0.491	0.488	0.484
1000	0.549	0.572	0.515	0.516	0.715	0.529	0.531	0.517	0.520	0.637
5000	0.643	0.657	0.575	0.595	0.858	0.633	0.647	0.571	0.592	0.809
	Среднее отклонение ММП-0 оценок					Среднее отклонение ДП-0 оценок				
300	0.869	0.314	0.297	0.267	0.102	1.143	0.363	0.301	0.276	0.150
1000	0.635	0.207	0.267	0.155	0.069	0.771	0.250	0.263	0.153	0.099
5000	0.458	0.150	0.199	0.101	0.051	0.554	0.169	0.203	0.102	0.072
	Среднее отклонение МНК оценок					Доля побед ММП оценок над ДП				
300	0.699	0.307	0.295	0.262	0.114	0.627	0.576	0.570	0.539	0.670
1000	0.796	0.296	0.268	0.194	0.107	0.641	0.577	0.570	0.542	0.660
5000	0.819	0.244	0.230	0.140	0.102	0.645	0.612	0.592	0.540	0.625

Таблица 2. Доля побед ММП оценок над МНК в отношении коэффициента β_1 (при 1000 наблюдениях)

ρ_2	ρ_1							
	(–1,–0.75)	(–0.75,–0.5)	(–0.5,–0.25)	(–0.25, 0)	(0, 0.25)	(0.25, 0.5)	(0.5, 0.75)	(0.75, 1)
(–1,–0.75)	1.000	0.400	0.583	0.667	0.737	0.833	0.750	0.600
(–0.75,–0.5)	0.846	0.733	0.762	0.588	0.593	0.704	0.579	0.875
(–0.5,–0.25)	0.778	0.733	0.565	0.577	0.296	0.533	0.700	0.778
(–0.25, 0)	0.714	0.840	0.727	0.476	0.483	0.588	0.615	0.778
(0, 0.25)	0.800	0.810	0.722	0.438	0.500	0.600	0.741	0.833
(0.25, 0.5)	0.800	0.826	0.400	0.360	0.526	0.583	0.619	0.700
(0.5, 0.75)	1.000	0.714	0.818	0.632	0.500	0.733	0.636	0.692
(0.75, 1)	1.000	0.909	0.636	0.667	0.727	0.667	0.400	1.000

[6] При очень большом числе симуляций данная таблица окажется симметричной.

3.4. Вторая спецификация

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i,$$
$$z_{1i} = \gamma_{10} + \gamma_{11} x_{1i} + \gamma_{12} x_{2i} + \gamma_{13} x_{3i} + u_{1i},$$
$$z_{2i} = \gamma_{20} + \gamma_{21} x_{1i} + \gamma_{22} x_{2i} + \gamma_{23} x_{3i} + u_{2i},$$
$$i \in \{1,...,n\}, \ n \in \{300, 1000, 5000\}.$$

В рамках данной спецификации предполагается, что на основную зависимую переменную и переменные отбора влияют одни и те же факторы. Такая спецификация представляет наибольший интерес, т. к. соответствует наиболее распространенной на практике ситуации.

Таблица 3. Сравнение качества оценок коэффициентов при различных способах оценивания

Число наблюдений	β_0	β_1	β_2	β_3	σ	β_0	β_1	β_2	β_3	σ
	Доля побед ММП оценок над МНК					Доля побед ДП оценок над МНК				
300	0.473	0.436	0.441	0.434	0.544	0.355	0.332	0.372	0.354	0.332
1000	0.673	0.562	0.572	0.593	0.718	0.545	0.487	0.467	0.502	0.507
5000	0.809	0.757	0.748	0.761	0.870	0.738	0.695	0.686	0.705	0.722
	Среднее отклонение ММП оценок					Среднее отклонение ДП оценок				
300	1.245	0.436	0.547	0.524	0.544	1.834	0.332	0.747	0.755	0.332
1000	0.631	0.299	0.289	0.275	0.077	0.998	0.414	0.399	0.390	0.234
5000	0.251	0.121	0.120	0.104	0.028	0.393	0.182	0.179	0.154	0.078
	Доля побед ММП-0 оценок над МНК					Доля побед ДП-0 оценок над МНК				
300	0.444	0.443	0.450	0.443	0.522	0.395	0.416	0.414	0.404	0.408
1000	0.608	0.517	0.522	0.537	0.715	0.570	0.483	0.502	0.503	0.592
5000	0.702	0.566	0.577	0.578	0.856	0.677	0.552	0.550	0.556	0.772
	Среднее отклонение ММП-0 оценок					Среднее отклонение ДП-0 оценок				
300	0.953	0.443	0.474	0.450	0.522	1.28	0.416	0.563	0.526	0.408
1000	0.672	0.305	0.325	0.297	0.077	0.881	0.366	0.394	0.351	0.128
5000	0.510	0.234	0.225	0.209	0.054	0.546	0.276	0.258	0.236	0.078
	Среднее отклонение МНК оценок					Доля побед ММП оценок над ДП				
300	0.779	0.647	0.390	0.403	0.717	0.659	0.647	0.616	0.625	0.717
1000	0.917	0.316	0.328	0.349	0.107	0.667	0.615	0.654	0.630	0.697
5000	0.801	0.283	0.274	0.266	0.102	0.612	0.642	0.607	0.619	0.694

Согласно полученным результатам (см. табл. 3), на малых выборках оценки ММП и ДП процедур хуже, чем МНК. Однако на средней выборке ММП более точен, а на большой оба обобщенных метода Хекмана показывают результаты существенно лучшие, чем МНК. Ухудшение относительно качества оценок в случае с ДП может быть связано с упомянутой выше проблемой отсутствия exclusion restriction, вследствие чего оценки утрачивают эффективность из-за мультиколлинеарности. Тем не менее, оценки ММП оказались достаточно устойчивыми к этой проблеме, что говорит о целесообразности использования обобщенного

метода Хекмана даже в тех случаях, когда наборы факторов, влияющих на зависимые переменные в основном уравнении и уравнениях отбора, совпадают.

4. Применение обобщенного метода Хекмана для оценивания уравнения заработной платы по данным РМЭЗ за 2016 год

В качестве примера применения обобщенного метода Хекмана рассмотрим классическую задачу оценивания минцеровского уравнения заработной платы. Отметим тот факт, что наблюдения по зарплате доступны лишь для тех индивидов, которые и работают, и согласились предоставить информацию о своих трудовых доходах. Таким образом, отбор может быть неслучайным не только по признаку занятости, но и по наличию ответа на вопрос о заработной плате.

Приводит ли к смещению при оценивании заработной платы удаление из выборки тех респондентов, которые отказались отвечать на вопрос о доходах?

В рамках предложенной в разделе 2.1 модели для ответа на поставленный вопрос необходимо проверить гипотезу о равенстве нулю коэффициента корреляции ошибок уравнения заработной платы и бинарного уравнения ответа на вопрос.

В качестве выборки используются данные опроса РМЭЗ по индивидам за 2016 год[7]. Рассматриваются женщины в возрасте 25–55 лет. Указанный возрастной период выбран в соответствии с большинством исследований, посвященных предложению труда женщинами (Heckman, Killingsworth, 1987).

В соответствии с классическим подходом (Mincer, 1974; Heckman et al., 2006) в качестве зависимой переменной выступает логарифм заработной платы, а к экзогенным переменным относятся стаж (эффект которого предполагается квадратичным) и образование. Эффект образования, как правило, оказывается положительным, а влияние стажа характеризуется вогнутостью: линейная часть положительна, а квадратичная — отрицательна (Heckman et al., 2006). В уравнение также включены характеристики места проживания респондента и его состояния здоровья.

В уравнение занятости были включены факторы, традиционно используемые при анализе предложения труда среди женщин: возраст, образование, зарплата мужа, число несовершеннолетних детей и нетрудовые доходы (Heckman, Killingsworth, 1987; Blundell, MaCurdy, 1999).

Факторы, влияющие на принятие респондентом решения об ответе на вопрос о своих трудовых доходах, изучены довольно слабо (Turrell, 2000, Kim et al., 2007). Поэтому предполагается, что в уравнение ответа на вопрос о зарплате входят те же экзогенные переменные, что и в уравнение занятости, за исключением нетрудовых доходов, поскольку те, кто не отвечают на вопрос о зарплате, часто не предоставляют информацию и о своих нетрудовых доходах.

---

[7] «Российский мониторинг экономического положения и здоровья населения НИУ ВШЭ (RLMS-HSE)», проводимый Национальным исследовательским университетом «Высшая школа экономики» и ЗАО «Демоскоп» при участии Центра народонаселения Университета Северной Каролины в Чапел Хилле и Института социологии РАН. Сайты обследования RLMS-HSE — http://www.cpc.unc.edu/projects/rlms и http://www.hse.ru/rlms.

Ниже приведена табл. 4 с оценками параметров уравнения заработной платы. Во втором и четвертом столбце таблицы представлены оценки, полученные при одновременном оценивании уравнения заработной платы и двух уравнений отбора (занятости и ответа на вопрос о величине зарплаты) ММП и описанным выше двухшаговым методом. В третьем и пятом столбце — оценки, полученные с помощью обычной модели Хекмана с единственным уравнением отбора по занятости, игнорирующей неслучайный характер отказа ответа на вопрос о величине зарплаты.

Таблица 4. Оценки уравнения заработной платы

Переменные	Обобщенный метод (ММП)	Модель Хекмана (ММП)	Обобщенный метод (2 шага)	Модель Хекмана (2 шага)
Константа	9.112***	9.171***	9.180***	9.193***
	(0.067)	(0.062)	(0.070)	(0.064)
Среднее образование	0.112**	0.121***	0.094	0.120***
	(0.047)	(0.046)	(0.066)	(0.046)
Среднее специальное образование	0.207***	0.195***	0.222***	0.190***
	(0.047)	(0.046)	(0.070)	(0.046)
Высшее образование	0.528***	0.499***	0.569***	0.490***
	(0.048)	(0.045)	(0.091)	(0.046)
Стаж	0.0165***	0.0160***	0.0161***	0.0154***
	(0.004)	(0.004)	(0.0057)	(0.004)
Квадрат стажа	–0.00036***	–0.00036***	–0.00034**	–0.00034***
	(0.00012)	(0.00012)	(0.00015)	(0.00012)
Проживает в Москве	0.863***	0.865***	0.859***	0.863***
	(0.042)	(0.041)	(0.055)	(0.042)
Проживает в Санкт-Петербурге	0.654***	0.692***	0.587***	0.692***
	(0.077)	(0.076)	(0.127)	(0.076)
Проживает в областном центре	0.283***	0.288***	0.267***	0.284***
	(0.029)	(0.029)	(0.040)	(0.029)
Проживает в городе	0.188***	0.191***	0.176***	0.187***
	(0.029)	(0.029)	(0.039)	(0.029)
Проживает в поселке городского типа	0.183***	0.207***	0.140*	0.206***
	(0.047)	(0.045)	(0.077)	(0.045)
Инвалидность	–0.160*	–0.148***	–0.160	–0.133
	(0.086)	(0.078)	(0.122)	(0.085)
σ	0.500***	0.487***	0.586***	0.488***
	(0.012)	(0.0073)	(0.014)	(0.007)
ρ_0 (Занятость, Ответ)	0.251		0.265	
	(0.169)		(0.166)	
ρ_1 (ЗП, Занятость)	–0.173**	–0.166**	–0.206**	–0.214*
	(0.076)	(0.067)	(0.106)	(0.110)
ρ_2 (ЗП, Ответ)	0.407***		0.977	
	(0.154)		(0.784)	
Логарифм функции правдоподобия	–4234.00	–4333.04[8]	–4739.55	–5048.41

[8] Данное значение представляет собой сумму логарифмов правдоподобия одномерной модели Хекмана и бинарного уравнения ответа на вопрос о зарплате.

Окончание табл. 4

Переменные	Обобщенный метод (ММП)	Модель Хекмана (ММП)	Обобщенный метод (2 шага)	Модель Хекмана (2 шага)
LR тест (одномерная модель Хекмана — модель с ограничениями)	$LR = 2(-4234.00+4333.04) \approx 200$		$LR = 2(-4739.55+5048.41) \approx 618$	
	$\chi^2_{2;0.005} = 10.6$			
AIC	8361.99	8732.09	9585.11	10162.82
Число наблюдений	Всего — 3733, работающих — 2707, ответивших — 2424			

Примечание. В скобках — стандартные отклонения; *** — $p < 0.01$, ** — $p < 0.05$, * — $p < 0.1$.

Интерпретация эффектов переменных не является целью данного параграфа. Отметим только, что знаки коэффициентов соответствуют ожиданиям. Главный вопрос — есть ли смещение из-за отказа ответить на вопрос о величине зарплаты? Как видно из таблицы, корреляция случайных ошибок уравнения зарплаты и ответа на вопрос о ее величине значима и положительна. Следовательно, существуют ненаблюдаемые факторы, увеличивающие заработную плату и вероятность ответа на вопрос о ее величине. И, как показывает значение *LR* статистики, использование одномерной модели Хекмана в данном случае неправомерно. Однако игнорирование возникающего смещения не приводит к существенным изменениям оценок коэффициентов при объясняющих переменных, что видно при сопоставлении их значений (столбцы 2 и 3, 4 и 5).

Корреляция ошибок уравнений занятости и ответа на вопрос о зарплате незначима, следовательно, может быть установлена направленность частных предельных эффектов. Переменные, положительно влияющие на вероятность занятости (ответа), также, через частный предельный эффект, увеличивают (уменьшают) ожидаемую заработную плату, и наоборот. Однако в уравнении ответа значимыми переменными оказались лишь высшее образование (коэффициент 0.557***) и проживание в Санкт-Петербурге (коэффициент –0.431**) (см. Приложение, п. 4). По этой причине предельные эффекты заработной платы работающих и ответивших на вопрос о зарплате индивидов при использовании обычной модели Хекмана будут отличаться от предельных эффектов обобщенной модели лишь для высшего образования и типа населенного пункта.

Согласно полученным результатам, средние предельные эффекты высшего образования для классической и обобщенной моделей Хекмана, оцененных методом максимального правдоподобия, практически совпадают и составляют 0.529 и 0.528 соответственно, аналогичный результат имеет место и для проживания в Санкт-Петербурге.

По всей видимости, столь малое различие в коэффициентах и предельных эффектах объясняется тем, что работающих и не ответивших на вопрос о зарплате женщин в выборке очень мало — менее 8% (283 человека). Было бы интересно посмотреть, насколько существенно смещение в уравнении заработной платы для мужчин, поскольку мужчины значительно чаще, чем женщины, отказываются отвечать на вопрос о заработной плате.

Отметим также, что оценка дисперсии при использовании обобщенной двухшаговой процедуры оказалась, по всей видимости, завышенной, а коэффициент корреляции уравнений зарплаты и ответа большим и незначимым, что, по всей видимости, вызвано относительно низкой эффективностью двухшаговой процедуры.

Заключение

В данной работе предложена регрессионная модель с переключением, регулируемым системой бинарных уравнений. Предполагается, что форма регрессионного уравнения основной зависимой переменной соответствует определенным комбинациям бинарных правил. При этом как наблюдения основной переменной, так и комбинации самих правил в некоторых случаях могут отсутствовать. Частным случаем описанной модели является модель селективных выборок (модель Хекмана) и модель с переключением (switch model).

В статье разработаны два способа оценивания модели: с помощью метода максимального правдоподобия и с помощью двухшаговой процедуры (метода максимального правдоподобия плюс метод наименьших квадратов), реализованные в среде R[9].

Получены выражения для предельных эффектов зависимой переменной, позволяющие определять направление и степень влияния на нее независимых факторов. Показано, что на зависимую переменную оказывают влияние не только переменные, входящие в основное уравнение, но и переменные, входящие в уравнения отбора наблюдений.

Качество предложенных способов оценивания для случая одного основного уравнения и двух правил отбора было изучено на симулированных данных. Оказалось, что для обобщенного случая сохраняются тенденции, присущие одномерным моделям: рост относительного преимущества метода Хекмана по мере увеличения корреляции между ошибками в основном и в селективных уравнениях, а также порождаемое квазимультиколлинеарностью существенное падение эффективности двухшаговой процедуры при условии схожести набора регрессоров для основной переменной и цензурирующих факторов. Кроме того, оценки обобщенного метода Хекмана в рамках использовавшихся симуляций оказались существенно более точными, чем оценки классического метода и МНК.

Список литературы

Blundell R., MaCurdy T. (1999). Labor supply: A review of alternative approaches. In: *Handbook of Labour Economics. Ashenfelter O., Card D. (eds.)*. Vol. 3, 1559–1659.

Chavent M., Liquet B., Saracco J. (2010). A semiparametric approach for a multivariate sample selection model. *Statistica Sinica*, 20 (2), 513–536.

Cinzia D. (2009). Sample selection correction in panel data models when selectivity is due to two sources. *Institute of Public Policy and Public Choice Working Papers*, 137.

Garen J. (1984). The retums to schooling: A selectivity bias approach with a continuous choice variable. *Econometrica*, 52 (5), 1199–1218.

Hay J., Leu R., Rohrer P. (1987). Ordinary least squares and sample-selection models of health-care demand: Monte Carlo comparison. *Journal of Business, Economic Statistics*, 5, 499–506.

Heckman J. (1976). The common structure of statistical models of truncation, sample selection and limited sependent variables and a simple estimator for such models. *Annals of Economic Social Measurement*, 5 (4), 475–492.

Heckman J. (1979). Sample selection bias as a specification error. *Econometrica*, 47 (1), 153–161.

Heckman J., Killingsworth M. (1987). Female labor supply: A survey. In: *O. Ashenfelter and R. Layard (eds.). Handbook of Labor Economics*. Vol. 1, 103–204.

[9] Код может быть предоставлен авторами по запросу.

Heckman J., Lochner L., Todd P. (2006). Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. *Handbook of the Economics of Education*, vol. 1, 307–458.

Jeffrey D., McFadden D. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, 52 (2), 345–362.

Jimenez E., Kugler B. (1987). The earnings impact of training duration in a developing country. *Journal of Human Resources*, 22 (2), 228–247.

Joe H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97, 2177–2189.

Kim S., Egerter S., Cubbin C., Takahashi E., Braveman P. (2007). Potential implications of missing income data in population-based surveys: An example from a pospartum survey in California. *Public Health*, 122, 753–763.

Manjunath B., Wilhelm S. (2012). Moments calculation for the double truncated multivariate normal density. *ArXiv e-prints*. arXiv:1206.5387.

Manning W., Duan N., Rogers W. (1987). Monte Carlo evidence on the choice between sample selection and two-part models. *Journal of Econometrics*, 35, 59–82.

Marchenko Y., Genton M. (2012). A Heckman selection-t model. *Journal of the American Statistical Association*, 107, 304–317.

Mincer J. (1974). *Schooling, experience and earnings*. New York: National Bureau of Economic Research.

Murphy K., Topel R. (2002). Estimation and inference in two step econometric models. *Journal of Business and Economic Statistics*, 20, 88–97.

Nawata K. (1993). A note on the estimation of models with sample selection biases. *Economics Letters*, 42, 15–24.

Nawata K. (1994). Estimation of sample selection bias models by the maximum likelihood estimator and Heckman's two-step estimator. *Economics Letters*, 45, 33–40.

Nelson F. (1984). Efficiency of the two-step estimator for models with endogenous sample selection. *Journal of Econometrics*, 24, 181–196.

Paarsch H. (1984). A Monte Carlo comparison of estimators for censored regression models. *Journal of Econometrics*, 24, 197–213.

Puhani P. (2000). The Heckman correction for sample selection and its critique. *Journal of Economic Surveys*, 14 (1), 53–68.

Rendtel U. (1992). On the choice of a selection-model when estimating eegression models with selectivity. *DIW-Discussion Paper*, 53.

Stolzenberg R., Relles D. (1990). Theory testing in a world of constrained research design, the significance of Heckman's censored sampling bias correction for nonexperimental research. *Sociological Methods and Research*, 18 (4), 395–415.

Turrell G. (2000). Income non-reporting: implications for health inequalities research. *Epidemiol Community Health*, 54, 207–214.

Vella F. (1993). A simple estimator for models with censored endogenous regressors. *International Economic Review*, 34 (2), 441–457.

Vella F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources*, 33, 127–169.

Zuehlke T., Zeman A. (1990). A comparison of two-stage estimators of censored regression models. *The Review of Economics and Statistics*, 72, 185–188.

Поступила в редакцию 18.12.2017;
принята в печать 21.04.2018.

Приложение
Некоторые свойства многомерного нормального распределения

1. Обозначения

Обозначим через F_X и f_X функцию распределения и плотность случайного вектора X. Положим $X^{(i)}$ в качестве вектора X без i-й компоненты X_i. Через $X^{(i,j)}$ обозначим вектор X без i-й и j-й компонент, $X_{i,j}$ — вектор, состоящий из компонент X_i и X_j.

Через Σ_{ij}^{-1} обозначим элемент обратной матрицы Σ^{-1} в i-й строке и j-м столбце. Аналогично, Σ_{ij} является (i,j)-элементом матрицы Σ. Через Σ_{i*} и Σ_{*j} обозначим i-ю строку и j-й столбец матрицы Σ соответственно.

2. Некоторые факты

2.1. Условное распределение

Предположим, что многомерный нормальный вектор $X \sim N(\mu, \Sigma)$ состоит из двух частей $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$, где X_1 и X_2 являются случайными векторами длины n_1 и n_2 соответственно. Представим ковариационную матрицу вектора X в блочном виде: $\Sigma = \begin{bmatrix} \Sigma^{11} & \Sigma^{12} \\ \Sigma^{21} & \Sigma^{22} \end{bmatrix}$ с соответствующими размерами $\begin{bmatrix} n_1 \times n_1 & n_2 \times n_1 \\ n_1 \times n_2 & n_2 \times n_2 \end{bmatrix}$.

В соответствии со свойствами многомерного нормального распределения, при фиксировании значений вектора X_2 вектор X_1 будет подчиняться многомерному нормальному закону $(X_1 \mid X_2 = \alpha) \sim N(\tilde{\mu}, \tilde{\Sigma})$, где условная математическое ожидание и условная ковариационная матрица имеют вид

$$\tilde{\mu} = \mu_1 + \Sigma^{21}(\Sigma^{22})^{-1}(a - \mu_2), \tag{13}$$

$$\tilde{\Sigma} = \Sigma^{11} - \Sigma^{21}(\Sigma^{22})^{-1}\Sigma^{12}. \tag{14}$$

2.2. Изменение знака

Обозначим через $\tilde{X}$ распределение, полученное из $X \sim N(\mu, \Sigma)$ путем изменения знака i-й компоненты на противоположный. Тогда, используя правила умножения случайного нормального вектора на скалярную матрицу, нетрудно показать, что отличия от исходного распределения будут заключаться в следующем: $\tilde{\mu}_i = -\mu_i$, $\tilde{\Sigma}_{ij} = -\Sigma_{ij}$, $\forall i \neq j$. Повторяя аналогичную процедуру, можно получить распределение вектора X, в котором были изменены знаки $k \leq n$ компонент.

2.3. Дифференцирование

Пусть случайный вектор $X \sim N(\mu, \Sigma)$, а $x = (x, ..., x)' \in \mathbb{R}$. Тогда справедливы следующие формулы дифференцирования[10]:

[10] Доказательства формул могут быть предоставлены авторами по запросу.

$$\frac{\partial f_X(x)}{\partial x_i} = -f_X(x)\Sigma_{i*}^{-1}(x-\mu), \tag{15}$$

$$\frac{\partial f_X(x)}{\partial \Sigma_{ij}} = \frac{1}{2}\frac{\partial^2 f_X(x)}{\partial x_i \partial x_j} = \frac{1}{2} f_X(x)\left(\sum_{l=1}^{n}\sum_{k=1}^{n}(x_l-\mu_l)(x_k-\mu_k)\Sigma_{ik}^{-1}\Sigma_{jl}^{-1} - \Sigma_{ij}^{-1}\right),\ i \neq j \tag{16}$$

или $$\frac{\partial f_X(x)}{\partial(\mathrm{Cov}(X_i, X_j))} = \frac{\partial^2 f_X(x)}{\partial x_i \partial x_j} = f_X(x)\left(\sum_{l=1}^{n}\sum_{k=1}^{n}(x_l-\mu_l)(x_k-\mu_k)\Sigma_{ik}^{-1}\Sigma_{jl}^{-1} - \Sigma_{ij}^{-1}\right),\ i \neq j, \tag{17}$$

$$\frac{\partial f_X(x)}{\partial \Sigma_{ii}} = \frac{1}{2}\frac{\partial^2 f_X(x)}{\partial^2 x_i} = \frac{1}{2} f_X(x)\left(\left(\Sigma_{i*}^{-1}(x-\mu)\right)^2 - \Sigma_{ii}^{-1}\right), \tag{18}$$

$$\frac{\partial F_X(x)}{\partial x_i} = f_{X_i}(x_i) F_{X^{(i)}|X_i=x_i}(x_1,...,x_{i-1},x_{i+1},...x_n), \tag{19}$$

$$\begin{aligned}&\frac{\partial F_{X^{(i)}|X_i=x_i}(x_1...x_{i-1}x_{i+1}...x_n)}{\partial x_i} = \\ &= -\sum_{k\neq i}\frac{\Sigma_{ik}}{\Sigma_{ii}} f_{X_k^{(i)}}(x_k) F_{X^{(i,k)}|X_i=x_i, X_k=x_k}(x_1,...,x_{i-1}x_{i+1},...,x_{k-1}x_{k+1},...,x_n),\end{aligned} \tag{20}$$

$$\frac{\partial F_X(x)}{\partial \Sigma_{ij}} = \frac{1}{2} f_{X_{i,j}}(x_i, x_j) F_{X^{(i,j)}|X_i=x_i, X_j=x_j}(x_1,...,x_{j-1}x_{j+1},...,x_{i-1}x_{i+1},...,x_n),\ i \neq j$$

или $$\frac{\partial F_X(x)}{\partial(\mathrm{Cov}(X_i, X_j))} = f_{X_{i,j}}(x_i, x_j) F_{X^{(i,j)}|X_i=x_i, X_j=x_j}(x_1,...,x_{j-1}x_{j+1},...,x_{i-1}x_{i+1},...,x_n),\ i \neq j, \tag{21}$$

$$\frac{\partial F_X(x)}{\partial \Sigma_{ii}} = \frac{1}{2}\left(\frac{\partial f_{X_i}(x_i)}{\partial x_i} F_{X^{(i)}|X^i=x_i}(x_1,...,x_{i-1},x_{i+1},...,x_n) + f_{X_i}(x_i)\frac{\partial F_{X^{(i)}|X^i=x_i}(x_1,...,x_{i-1},x_{i+1},...,x_n)}{\partial x_i}\right), \tag{22}$$

где $\frac{\partial f_{X_i}(x_i)}{\partial x_i}$ и $\frac{\partial F_{X^{(i)}|X^i=x_i}(x_1,...,x_{i-1},x_{i+1},...,x_n)}{\partial x_i}$ вычисляются по формулам (15) и (20) соответственно.

3. Усеченное многомерное нормальное распределение

3.1. Определение

Для произвольных векторов $a, b \in R^n$ введем множество

$$B_{a,b} = \{x \in R^n : a_1 \le x_1 \le b_1, ..., a_n \le x_n \le b_n\}.$$

Если $X \in R^n$ — многомерная нормальная величина с математическим ожиданием μ и ковариационной матрицей Σ, то усеченная (на множество $B_{a,b}$) многомерная случайная величина $X^{a,b}$ имеет плотность распределения, задаваемую формулой

$$f_{X^{a,b}}(x) = \begin{cases}\dfrac{f_X(x)}{\mathrm{P}(a_1 \le x_1 \le b_1, ..., a_n \le x_n \le b_n)}, & \text{если } x \in B_{a,b},\\ 0, & \text{в противном случае.}\end{cases}$$

3.2. Моменты

Первый и второй моменты компонент случайного вектора $X^{a,b}$ могут быть вычислены по следующим формулам (Manjunath, Wilhelm, 2012):

$$\mathrm{E}(X_i^{a,b}) = \mu_i + \sum_{j=1}^{n}\left(f_{X_j^{a,b}}(a_j) - f_{X_j^{a,b}}(b_j)\right)\cdot\Sigma_{ij}, \tag{23}$$

$$\begin{aligned}\mathrm{Cov}(X_i^{a,b}, X_j^{a,b}) &= \Sigma_{ij} + \sum_{k=1}^{n}\frac{\Sigma_{ik}\Sigma_{jk}}{\Sigma_{kk}}\left(a_k f_{X_k^{a,b}}(a_k) - b_k f_{X_k^{a,b}}(b_k)\right) + \\ &+\sum_{k=1}^{n}\Sigma_{ik}\sum_{q\neq k}\left(\Sigma_{jq} - \frac{\Sigma_{kq}\Sigma_{jk}}{\Sigma_{kk}}\right)\left(f_{X_{k,q}^{a,b}}(a_k,a_q) - f_{X_{k,q}^{a,b}}(a_k,b_q) - f_{X_{k,q}^{a,b}}(b_k,a_q) + f_{X_{k,q}^{a,b}}(b_k,b_q)\right) - \\ &-\left(\sum_{k=1}^{n}(f_{X_k^{a,b}}(a_k) - f_{X_k^{a,b}}(b_k))\cdot\Sigma_{ik}\right)\left(\sum_{k=1}^{n}(f_{X_k^{a,b}}(a_k) - f_{X_k^{a,b}}(b_k))\cdot\Sigma_{jk}\right).\end{aligned} \tag{24}$$

3.3. Вывод некоторых свойств усеченного сверху многомерного нормального распределения

В рамках данной работы представляет интерес случай, когда

$$X \sim N\left(\begin{bmatrix}\mu\\0\\\vdots\\0\end{bmatrix}, \begin{bmatrix}\sigma^2 & \sigma\rho_{12} & \cdots & \sigma\rho_{1n}\\ \sigma\rho_{12} & 1 & \cdots & \rho_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma\rho_{1n} & \rho_{2n} & \cdots & 1\end{bmatrix}\right),\ a = (-\infty,\cdots,-\infty),\ b = (\infty, b_2,\cdots,b_n). \tag{25}$$

Обозначим связанную этими ограничениями многомерную случайную величину через X^b. Маргинальная плотность i-й компоненты X^b принимает вид[11] для $i \in \{2,...,n\}$:

$$\begin{aligned}f_{X_i^b}(x_i) &= \int_{-\infty}^{\infty}\int_{-\infty}^{b_2}\cdots\int_{-\infty}^{b_{i-1}}\int_{-\infty}^{b_{i+1}}\cdots\int_{-\infty}^{b_n} f_{X^b}(x_1,...,x_{i-1},x_i,x_{i+1},...,x_n)dx_n...dx_{i+1}dx_{i-1}...dx_1 = \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{b_2}\cdots\int_{-\infty}^{b_{i-1}}\int_{-\infty}^{b_{i+1}}\cdots\int_{-\infty}^{b_n}\frac{f_X(x_1,...,x_{i-1},x_i,x_{i+1},...,x_n)}{F_{X^{(1)}}(b_2,...,b_n)}dx_n...dx_{i+1}dx_{i-1}...dx_2dx_1 = \\ &= \int_{-\infty}^{b_2}\cdots\int_{-\infty}^{b_{i-1}}\int_{-\infty}^{b_{i+1}}\cdots\int_{-\infty}^{b_n}\frac{f_{X^{(1)}}(x_2,...,x_{i-1},x_i,x_{i+1},...,x_n)}{F_{X^{(1)}}(b_2,...,b_n)}dx_n...dx_{i+1}dx_{i-1}...dx_2 = \\ &= \frac{1}{F_{X^{(1)}}(b_2,...,b_n)}\cdot\frac{\partial F_{X^{(1)}}(b_2,...,b_{i-1},x_i,b_{i+1},...,b_n)}{\partial x_i}.\end{aligned}$$

[11] Предложенное преобразование позволяет рассчитывать плотность усеченного многомерного распределения через плотности и функции многомерного нормального распределения. Это, во-первых, существенно ускоряет скорость работы сопряженных с расчетом функций правдоподобия оптимизационных алгоритмов и, во-вторых, значительно упрощает задачу дифференцирования соответствующей плотности.

Отметим, что

$$f_{X_i^b}(b_i) = \frac{1}{F_{X^{(1)}}(x)} \cdot \frac{\partial F_{X^{(1)}}(x)}{\partial x_i}\bigg|_{x=b}, \quad i \in \{2,...,n\}. \tag{26}$$

Аналогичным образом получаем выражения для совместных плотностей двух и трех компонент[12]:

$$f_{X^b_{i,j}}(x,y) = \frac{1}{F_{X^{(1)}}(b_2,...,b_n)} \cdot \frac{\partial^2 F_{X^{(1)}}(b_2,...,b_{i-1},x,b_{i+1},...,b_{j-1},y,b_{j+1},...,b_n)}{\partial x \partial y}, \tag{27}$$

$$i, j \in \{2,...,n\}, i \neq j,$$

$$f_{X^b_{i,j,k}}(x,y,z) = \frac{1}{F_{X^{(1)}}(b_2,...,b_n)} \cdot$$

$$\cdot \frac{\partial^3 F_{X^{(1)}}(b_2,...,b_{i-1},x,b_{i+1},...,b_{j-1},y,b_{j+1},...,b_{k-1},z,b_{k+1},...,b_n)}{\partial x \partial y \partial z}, \tag{28}$$

$$i, j, k \in \{2,...,n\}, \quad i \neq j \neq k.$$

Подставляя полученные выражений для маргинальной и совместных плотностей в формулы (23) и (24), получаем следующую лемму.

Лемма 1. *Математическое ожидание и дисперсия первой компоненты* X^b *могут быть представлены в виде*

$$\mathrm{E}(X_1^b) = \mu - \sigma \sum_{i=2}^{n} \rho_{1i} f_{X_i^b}(b_i), \tag{29}$$

$$\mathrm{D}(X_1^b) = \sigma^2 \left(1 - \sum_{k=2}^{n} \rho_{1k}^2 b_k f_{X_k^b}(b_k) + \sum_{k=2}^{n} \rho_{1k} \sum_{j \neq k} (\rho_{1j} - \rho_{kj}\rho_{1k}) f_{X^b_{k,j}}(b_k, b_j) - \left(\sum_{k=2}^{n} \rho_{1k} f_{X_k^b}(b_k)\right)^2\right). \tag{30}$$

3.4. Дифференцирование в точке верхней границы

Рассмотрим многомерную случайную величину X^b. Для упрощения записи введем следующие обозначения:

$$F(x) = F_X(x), \quad f(x_i) = f_{X_i}(x_i), \quad F(x_i) = F_{X_i}(x_i), \quad F(x \mid x_i) = F_{X^{(i)}|X_i = x_i}(x^{(i)}),$$

$$f(x_i, x_j) = f_{X_{i,j}}(x_i, x_j), \quad F(x \mid x_i, x_j) = F_{X^{(i,j)}|X_i = x_i, X_j = x_j}(x^{(i,j)}),$$

$$f(x_i, x_j, x_k) = f_{X_{i,j,k}}(x_i, x_j, x_k), \quad F(x \mid x_i, x_j, x_k) = F_{X^{(i,j,k)}|X_i = x_i, X_j = x_j, X_k = x_k}(x^{(i,j,k)}).$$

[12] Чтобы развернуть эти формулы, достаточно воспользоваться указанными выше правилами дифференцирования многомерного нормального распределения.

1. Дифференцирование маргинальной функции плотности в точке границы *по верхней границе*:

$$\frac{\partial f_{X_i^b}(b_i)}{\partial b_j} = f_{X_{i,j}^b}(b_i, b_j) - f_{X_i^b}(b_i) f_{X_j^b}(b_j), \quad i \neq j, \tag{31}$$

$$\frac{\partial f_{X_i^b}(b_i)}{\partial b_i} = -b_i f_{X_i^b}(b_i) - f_{X_i^b}(b_i)^2 - \sum_{j \neq i} f_{X_{i,j}^b}(b_i, b_j)\rho_{ij}. \tag{32}$$

Для доказательства (31) воспользуемся формулами (15), (19) и (26):

$$\frac{\partial f_{X_i^b}(b_i)}{\partial b_j} = \frac{\partial\left(\frac{f(b_i)F(b \mid b_i)}{F(b)}\right)}{\partial b_j} = \frac{F(b) f(b_i, b_j) F(b \mid b_i, b_j) - f(b_i)F(b \mid b_i) f(b_j) F(b \mid b_j)}{F(x)^2} =$$

$$= \frac{f(b_i, b_j)F(b \mid b_i, b_j)}{F(b)} - \frac{f(b_i)F(b \mid b_i)}{F(b)} \cdot \frac{f(b_j)F(b \mid b_j)}{F(b)} = f_{X_{i,j}^b}(b_i, b_j) - f_{X_i^b}(b_i) f_{X_j^b}(b_j).$$

Для доказательства (32) воспользуемся формулами (15), (19), (20) и (26):

$$\frac{\partial f_{X_i^b}(b_i)}{\partial b_i} = \frac{\partial\left(\frac{f(b_i)F(b \mid b_i)}{F(b)}\right)}{\partial b_i} =$$

$$= \frac{\left(-f(b_i)\sum_{j \neq i} \rho_{i,j} f(b_j \mid b_i) F(b \mid b_i, b_j) - b_i f(b_i) F(b \mid b_i)\right) F(b) - \left(f(b_i)F(b \mid b_i)\right)^2}{F(b)^2} =$$

$$= \frac{-\sum_{j \neq i} \rho_{i,j} f(b_i, b_j) F(b \mid b_i, b_j)}{F(b)} - \frac{b_i f(b_i) F(b \mid b_i)}{F(b)} - \left(\frac{f(b_i)F(b \mid b_i)}{F(b)}\right)^2 =$$

$$= -b_i f_{X_i^b}(b_i) - f_{X_i^b}(b_i)^2 - \sum_{j \neq i} f_{X_{i,j}^b}(b_i, b_j)\rho_{ij}.$$

2. Дифференцирование маргинальной функции плотности в точке границы *по корреляции*:

$$\frac{\partial f_{X_i^b}(b_i)}{\partial \rho_{jk}} = f_{X_{i,j,k}^b}(b_i, b_j, b_k) - f_{X_i^b}(b_i) f_{X_{j,k}^b}(b_j, b_k), \quad j \neq i, k \neq i, \tag{33}$$

$$\frac{\partial f_{X_i^b}(b_i)}{\partial \rho_{ij}} = \left(\frac{\rho_{ij} b_j - b_i}{1 - \rho_{ij}^2} - f_{X_i^b}(b_i)\right) f_{X_{ij}^b}(b_i, b_j) + \sum_{k \neq i,j} \frac{\rho_{ij}\rho_{kj} - \rho_{ki}}{1 - \rho_{ij}^2} f_{X_{i,j,k}^b}(b_i, b_j, b_k). \tag{34}$$

Чтобы доказать (33), воспользуемся формулами (15), (21) и (26):

$$\frac{\partial f_{X_i^b}(b_i)}{\partial \rho_{jk}} = \frac{\partial\left(\dfrac{f(b_i)F(b\,|\,b_i)}{F(b)}\right)}{\partial \rho_{jk}} =$$

$$= \frac{F(b)f(b_i,b_j,b_k)F(b\,|\,b_i,b_j,b_k) - f(b_i)F(b\,|\,b_i)f(b_j,b_k)F(b\,|\,b_j,b_k)}{F(x)^2} =$$

$$= \frac{f(b_i,b_j,b_k)F(b\,|\,b_i,b_j,b_k)}{F(b)} - \frac{f(b_i)F(b\,|\,b_i)}{F(b)} \cdot \frac{f(b_j,b_k)F(b\,|\,b_j,b_k)}{F(b)} =$$

$$= f_{X_{i,j,k}^b}(b_i,b_j,b_k) - f_{X_i^b}(b_i) f_{X_{j,k}^b}(b_j,b_k).$$

Чтобы доказать (34), без потери общности, для наглядности положим $i = n$ и $j = n-1$. Сначала рассмотрим

$$\frac{\partial F(b\,|\,b_n)}{\partial \rho_{(n-1),n}} = \sum_{i \neq n} \frac{\partial F(b\,|\,b_n)}{\partial \mathrm{E}(X_i^b \,|\, X_n^b = b_n)} \cdot \frac{\partial E(X_i^b \,|\, X_n^b = b_n)}{\partial \rho_{n-1,n}} +$$

$$+ \sum_{i \neq n} \sum_{j \neq n} \frac{\partial F(b\,|\,b_n)}{\partial \mathrm{Cov}(X_i^b, X_j^b \,|\, X_n^b = b_n)} \cdot \frac{\partial \mathrm{Cov}(X_i^b, X_j^b \,|\, X_n^b = b_n)}{\partial \rho_{n-1,n}}.$$

В соответствии с (13) условное математическое ожидание имеет вид

$$\mathrm{E}(X^b \,|\, X_n^b = b_n) = \begin{bmatrix} \rho_{1n} b_n \\ \vdots \\ \rho_{n-1,n} b_n \end{bmatrix}.$$

Поскольку $\rho_{n-1,n}$ располагается лишь в последней строке данного вектора, то, применяя (19), получаем:

$$\sum_{i \neq n} \frac{\partial F(b\,|\,b_n)}{\partial \mathrm{E}(X_i^b \,|\, X_n^b = b_n)} \cdot \frac{\partial \mathrm{E}(X_i^b \,|\, X_n^b = b_n)}{\partial \rho_{n-1,n}} = -b_n f(b_{n-1}\,|\,b_n) F(b\,|\,b_n, b_{n-1}).$$

Согласно формуле (14), условная (по n-й компоненте) ковариационная матрица имеет вид

$$\mathrm{Cov}(X^b \,|\, X_n^b = b_n) = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1,n-1} \\ \rho_{12} & 1 & \cdots & \rho_{2,n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,n-1} & \rho_{2,n-1} & \cdots & 1 \end{bmatrix} - \begin{bmatrix} \rho_{1n}^2 & \rho_{1n}\rho_{2n} & \cdots & \rho_{1n}\rho_{n-1,n} \\ \rho_{1n}\rho_{2n} & \rho_{2n}^2 & \cdots & \rho_{2n}\rho_{n-1,n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n}\rho_{n-1,n} & \rho_{2n}\rho_{n-1,n} & \cdots & \rho_{n-1,n}^2 \end{bmatrix}.$$

Заметим, что $\rho_{n-1,n}$ встречается лишь в правой части разности, причем только в $(n-1)$-м столбце и $(n-1)$-й строке, откуда следует, что

$$\sum_{i\neq n}\sum_{j\neq n}\frac{\partial F(b\,|\,b_n)}{\partial \mathrm{Cov}(X_i^b, X_j^b\,|\,X_n^b=b_n)}\cdot\frac{\partial \mathrm{Cov}(X_i^b, X_j^b\,|\,X_n^b=b_n)}{\partial\rho_{n-1,n}}=$$

$$=\sum_{k\le n-2}\frac{\partial F(b\,|\,b_n)}{\partial \mathrm{Cov}(X_k^b, X_{n-1}^b\,|\,X_n^b=b_n)}\cdot\frac{\partial \mathrm{Cov}(X_k^b, X_{n-1}^b\,|\,X_n^b=b_n)}{\partial\rho_{n-1,n}}+$$
$$+\frac{\partial F(b\,|\,b_n)}{\partial \mathrm{Cov}(X_{n-1}^b, X_{n-1}^b\,|\,X_n^b=b_n)}\cdot\frac{\partial \mathrm{Cov}(X_{n-1}^b, X_{n-1}^b\,|\,X_n^b=b_n)}{\partial\rho_{n-1,n}}=$$

$$=-\sum_{k\le n-2}\rho_{kn}\frac{\partial F(b\,|\,b_n)}{\partial \mathrm{Cov}(X_k^b, X_{n-1}^b\,|\,X_n^b=b_n)}-2\rho_{n-1,n}\frac{\partial F(b\,|\,b_n)}{\partial \mathrm{Cov}(X_{n-1}^b, X_{n-1}^b\,|\,X_n^b=b_n)}.$$

Используя (21) и (26), получаем:

$$-\sum_{k\le n-2}\rho_{kn}\frac{\partial F(b\,|\,b_n)}{\partial \mathrm{Cov}(X_k^b, X_{n-1}^b\,|\,X_n^b=b_n)}=-\sum_{k\le n-2}\rho_{kn}f(b_{n-1},b_k\,|\,b_n)F(b\,|\,b_n,b_{n-1},b_k).$$

Применяя (22) и (26), имеем:

$$-2\rho_{n-1,n}\frac{\partial F(b\,|\,b_n)}{\partial \mathrm{Cov}(X_{n-1}^b, X_{n-1}^b\,|\,X_n^b=b_n)}=$$
$$=-\rho_{n-1,n}\left(f'(b_{n-1}\,|\,b_n)F(b\,|\,b_n,b_{n-1})-\rho_{n-1,n}f(b_{n-1}\,|\,b_n)\frac{\partial F(b\,|\,b_n,b_{n-1})}{\partial b_{n-1}}\right)=$$
$$=\rho_{n-1,n}\left(f(b_{n-1}\,|\,b_n)F(b\,|\,b_n,b_{n-1})\frac{b_{n-1}-\rho_{n-1,n}b_n}{1-\rho_{n-1,n}^2}+\right.$$
$$\left.+\rho_{n-1,n}f(b_{n-1}\,|\,b_n)\sum_{k\le n-2}\frac{\rho_{k,n-1}-\rho_{kn}\rho_{n-1,n}}{1-\rho_{n-1,n}^2}f(b_k\,|\,b_n,b_{n-1})F(b\,|\,b_k,b_{n-1},b_n)\right).$$

В результате объединения выражений и раскрытия скобок получаем:

$$\sum_{i\neq n}\sum_{j\neq n}\frac{\partial F(b\,|\,b_n)}{\partial \mathrm{Cov}(X_i^b, X_j^b\,|\,X_n^b=b_n)}\cdot\frac{\partial \mathrm{Cov}(X_i^b, X_j^b\,|\,X_n^b=b_n)}{\partial\rho_{n-1,n}}=$$

$$=f(b_{n-1}\,|\,b_n)F(b\,|\,b_n,b_{n-1})\frac{\rho_{n-1,n}b_{n-1}-\rho_{n-1,n}^2b_n}{1-\rho_{n-1,n}^2}+$$

$$+\sum_{k\le n-2}\frac{\rho_{n-1,n}\rho_{k,n-1}-\rho_{n-1,n}^2\rho_{kn}}{1-\rho_{n-1,n}^2}f(b_{n-1},b_k\,|\,b_n)F(b\,|\,b_k,b_{n-1},b_n)-$$

$$-\sum_{k\le n-2}\rho_{kn}f(b_{n-1},b_k\,|\,b_n)F(b\,|\,b_n,b_{n-1},b_k)=$$

$$=f(b_{n-1}\,|\,b_n)F(b\,|\,b_n,b_{n-1})\frac{\rho_{n-1,n}b_{n-1}-\rho_{n-1,n}^2b_n}{1-\rho_{n-1,n}^2}+$$

$$+\sum_{k\ne n,n-1}\frac{\rho_{n-1,n}\rho_{k,n-1}-\rho_{kn}}{1-\rho_{n-1,n}^2}f(b_{n-1},b_k\,|\,b_n)F(b\,|\,b_k,b_{n-1},b_n).$$

Таким образом, имеем:

$$\frac{\partial F(b\,|\,b_n)}{\partial\rho_{n-1,n}}=f(b_{n-1}\,|\,b_n)F(b\,|\,b_n,b_{n-1})\frac{\rho_{n-1,n}b_{n-1}-\rho_{n-1,n}^2b_n}{1-\rho_{n-1,n}^2}+$$

$$+\sum_{k\ne n,n-1}\frac{\rho_{n-1,n}\rho_{k,n-1}-\rho_{kn}}{1-\rho_{n-1,n}^2}f(b_{n-1},b_k\,|\,b_n)F(b\,|\,b_k,b_{n-1},b_n)-b_nf(b_{n-1}\,|\,b_n)F(b\,|\,b_n,b_{n-1})=$$

$$=f(b_{n-1}\,|\,b_n)F(b\,|\,b_n,b_{n-1})\frac{\rho_{n-1,n}b_{n-1}-b_n}{1-\rho_{n-1,n}^2}+$$

$$+\sum_{k\ne n,n-1}\frac{\rho_{n-1,n}\rho_{k,n-1}-\rho_{kn}}{1-\rho_{n-1,n}^2}f(b_{n-1},b_k\,|\,b_n)F(b\,|\,b_k,b_{n-1},b_n).$$

Наконец, используя полученный результат и формулы (26), (27) и (28), получаем:

$$\frac{\partial f_{X_n^b}(b_n)}{\partial\rho_{n-1,n}}=\frac{\partial\left(\dfrac{f(b_n)F(b\,|\,b_n)}{F(b)}\right)}{\partial\rho_{n-1,n}}=\frac{\rho_{n-1,n}b_{n-1}-b_n}{1-\rho_{n-1,n}^2}\cdot\frac{f(b_n)F(b)f(b_{n-1}\,|\,b_n)F(b\,|\,b_n,b_{n-1})}{F(b)^2}+$$

$$+\sum_{k\ne n,n-1}\frac{\rho_{n-1,n}\rho_{k,n-1}-\rho_{kn}}{1-\rho_{n-1,n}^2}\cdot\frac{f(b_n)F(b)f(b_{n-1},b_k\,|\,b_n)F(b\,|\,b_k,b_{n-1},b_n)}{F(b)^2}-$$

$$-\frac{f(b_n,b_{n-1})F(b\,|\,b_n,b_{n-1})f(b_n)F(b\,|\,b_n)}{F(b)^2}=\frac{\rho_{n-1,n}b_{n-1}-b_n}{1-\rho_{n-1,n}^2}\frac{f(b_n,b_{n-1})F(b\,|\,b_n,b_{n-1})}{F(b)}+$$

$$+\sum_{k\ne n,n-1}\frac{\rho_{n-1,n}\rho_{k,n-1}-\rho_{kn}}{1-\rho_{n-1,n}^2}\cdot\frac{f(b_k,b_{n-1},b_n)F(b\,|\,b_k,b_{n-1},b_n)}{F(b)}-$$

$$-\frac{f(b_n,b_{n-1})F(b\,|\,b_n,b_{n-1})}{F(b)}\cdot\frac{f(b_n)F(b\,|\,b_n)}{F(b)}=$$

$$=\left(\frac{\rho_{n-1,n}b_{n-1}-b_n}{1-\rho_{n-1,n}^2}-f_{X_n^b}(b_n)\right)f_{X_{n,n-1}^b}(b_n,b_{n-1})+\sum_{k\ne n,n-1}\frac{\rho_{n-1,n}\rho_{k,n-1}-\rho_{kn}}{1-\rho_{n-1,n}^2}f_{X_{n,n-1,k}^b}(b_n,b_{n-1},b_k).$$

4. Оценки обобщенной модели Хекмана: заработная плата, занятость, ответ. ММП

Независимая переменная	Зарплата	Занятость	Ответ
Константа	9.112***	–1.253*	0.481
	(0.067)	(0.596)	(0.903)
Среднее образование	0.112**	0.195***	–0.091
	(0.047)	(0.090)	(0.138)
Среднее специальное образование	0.207***	0.577***	0.212
	(0.047)	(0.093)	(0.147)
Высшее образование	0.528***	0.955***	0.557***
	(0.048)	(0.093)	(0.153)
Стаж	0.0165***		
	(0.0044)		
Квадрат стажа	–0.00036***		
	(0.00012)		
Возраст		0.094***	0.013
		(0.030)	(0.044)
Квадрат возраста		–0.00118***	–0.0000
		(0.00038)	(0.0005)
Проживает в Москве	0.863***	0.064	0.035
	(0.042)	(0.098)	(0.148)
Проживает в Санкт-Петербурге	0.654***	0.016	–0.431**
	(0.077)	(0.163)	(0.206)
Проживает в областном центре	0.283***	0.326***	–0.052
	(0.029)	(0.068)	(0.103)
Проживает в городе	0.188***	0.387***	0.007
	(0.029)	(0.068)	(0.102)
Проживает в поселке городского типа	0.183***	0.035	–0.269
	(0.046)	(0.102)	(0.138)
Число детей в возрасте до двух лет		–1.241***	–0.354
		(0.086)	(0.200)
Число детей в возрасте 3–6 лет		–0.200***	0.082
		(0.057)	(0.093)
Число детей в возрасте 7–18 лет		0.127**	0.069
		(0.058)	(0.076)
Помощь родственников в уходе за детьми		0.252***	0.032
		(0.068)	(0.093)
Состоит в официальном браке		0.144	0.069
		(0.183)	(0.238)
Логарифм дохода мужа		–0.0446**	0.0005
		(0.0178)	(0.0237)
Логарифм нетрудового дохода		–0.0910***	
		(0.0058)	
Инвалидность	–0.160*	–0.802***	–0.351
	(0.086)	(0.127)	(0.283)
σ	0.500***		
	(0.012)		
ρ_0	0.251		
	(0.169)		
ρ_1	–0.173**		
	(0.077)		
ρ_2	0.407***		
	(0.154)		
Число наблюдений	3733	3733	3733

Примечание. В скобках — стандартные отклонения; *** — $p < 0.01$, ** — $p < 0.05$, * — $p < 0.1$.

Kossova E. V., Potanin B. S. Heckman method and switching regression model multivariate generalization. *Applied Econometrics*, 2018, v. 50, pp. 114–143.

Elena Kossova
National Research University Higher School of Economics (NRU HSE), Moscow, Russian Federation; ekossova@hse.ru

Bogdan Potanin
National Research University Higher School of Economics (NRU HSE), Moscow, Russian Federation; bogdanpotanin@gmail.com

Heckman method and switching regression model multivariate generalization

The article is devoted to simultaneous estimation of one continuous and various binary equations under assumption of disturbances joint normality. It generalizes Heckman selection and switch-probit models to multivariate case. Following Heckman's univariate model implementation both two step and maximum likelihood procedures are provided. In order to test model performance and correctness we execute analysis on simulated data. It shows that when there are two selection equations generalized model estimates accuracy noticeably outperforms those that are obtained using least squares or Heckman's methods.

Keywords: sample selection; switching regression model.

JEL classification: C34.

References

Blundell R., MaCurdy T. (1999). Labor supply: A review of alternative approaches. In: *Handbook of Labour Economics. Ashenfelter O., Card D. (eds.)*. Vol. 3, 1559–1659.

Chavent M., Liquet B., Saracco J. (2010). A semiparametric approach for a multivariate sample selection model. *Statistica Sinica*, 20 (2), 513–536.

Cinzia D. (2009). Sample selection correction in panel data models when selectivity is due to two sources. *Institute of Public Policy and Public Choice Working Papers*, 137.

Garen J. (1984). The retums to schooling: A selectivity bias approach with a continuous choice variable. *Econometrica*, 52 (5), 1199–1218.

Hay J., Leu R., Rohrer P. (1987). Ordinary least squares and sample-selection models of health-care demand: Monte Carlo comparison. *Journal of Business, Economic Statistics*, 5, 499–506.

Heckman J. (1976). The common structure of statistical models of truncation, sample selection and limited sependent variables and a simple estimator for such models. *Annals of Economic Social Measurement*, 5 (4), 475–492.

Heckman J. (1979). Sample selection bias as a specification error. *Econometrica*, 47 (1), 153–161.

Heckman J., Killingsworth M. (1987). Female labor supply: A survey. In: *O. Ashenfelter and R. Layard (eds.). Handbook of Labor Economics*. Vol. 1, 103–204.

Heckman J., Lochner L., Todd P. (2006). Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. *Handbook of the Economics of Education*, vol. 1, 307–458.

Jeffrey D., McFadden D. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica*, 52 (2), 345–362.

Jimenez E., Kugler B. (1987). The earnings impact of training duration in a developing country. *Journal of Human Resources*, 22 (2), 228–247.

Joe H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis*, 97, 2177–2189.

Kim S., Egerter S., Cubbin C., Takahashi E., Braveman P. (2007). Potential implications of missing income data in population-based surveys: An example from a pospartum survey in California. *Public Health*, 122, 753–763.

Manjunath B., Wilhelm S. (2012). Moments calculation for the double truncated multivariate normal density. *ArXiv e-prints*. arXiv:1206.5387.

Manning W., Duan N., Rogers W. (1987). Monte Carlo evidence on the choice between sample selection and two-part models. *Journal of Econometrics*, 35, 59–82.

Marchenko Y., Genton M. (2012). A Heckman selection-t model. *Journal of the American Statistical Association*, 107, 304–317.

Mincer J. (1974). *Schooling, experience and earnings*. New York: National Bureau of Economic Research.

Murphy K., Topel R. (2002). Estimation and inference in two step econometric models. *Journal of Business and Economic Statistics*, 20, 88–97.

Nawata K. (1993). A note on the estimation of models with sample selection biases. *Economics Letters*, 42, 15–24.

Nawata K. (1994). Estimation of sample selection bias models by the maximum likelihood estimator and Heckman's two-step estimator. *Economics Letters*, 45, 33–40.

Nelson F. (1984). Efficiency of the two-step estimator for models with endogenous sample selection. *Journal of Econometrics*, 24, 181–196.

Paarsch H. (1984). A Monte Carlo comparison of estimators for censored regression models. *Journal of Econometrics*, 24, 197–213.

Puhani P. (2000). The Heckman correction for sample selection and its critique. *Journal of Economic Surveys*, 14 (1), 53–68.

Rendtel U. (1992). On the choice of a selection-model when estimating eegression models with selectivity. *DIW-Discussion Paper*, 53.

Stolzenberg R., Relles D. (1990). Theory testing in a world of constrained research design, the significance of Heckman's censored sampling bias correction for nonexperimental research. *Sociological Methods and Research*, 18 (4), 395–415.

Turrell G. (2000). Income non-reporting: implications for health inequalities research. *Epidemiol Community Health*, 54, 207–214.

Vella F. (1993). A simple estimator for models with censored endogenous regressors. *International Economic Review*, 34 (2), 441–457.

Vella F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources*, 33, 127–169.

Zuehlke T., Zeman A. (1990). A comparison of two-stage estimators of censored regression models. *The Review of Economics and Statistics*, 72, 185–188.

Received 14.03.2018; accepted 21.04.2018.