Эконометрический ликбез: ограниченные зависимые переменные

Мультиномиальные модели дискретного выбора*

Золт Шандор†

Университет Гронингена, Гронинген, Нидерланды

В настоящем эссе описаны основные свойства хорошо известных мультиномиальных моделей дискретного выбора: стандартной логит-модели, смешанной логит-модели и пробит-модели.

1 Введение

В эссе описываются некоторые модели дискретного выбора, применяемые для оценивания параметров спроса на товары, приобретаемые в дискретных количествах. В этих моделях различные товары рынка считаются различными выборными альтернативами. В интересующих нас моделях предполагается, что в течение заданного периода времени потребитель, считающийся принимающие решения лицом, приобретает либо одну, либо ноль единиц товара. Набор всех альтернатив, из которых выбирает потребитель, считается конечным и содержащим все товары, доступные на рынке. Обозначим альтернативу как j, а набор всех альтернатив как $\{1, 2, ..., J\}$. В модели дискретного выбора потребитель i, выбирающий альтернативу j, приобретает случайную полезность u_{ij}, которая напрямую исследователем не наблюдается. Во многих ситуациях, однако, вероятность того, что выбрана альтернатива j, наблюдается с небольшой выборочной ошибкой в виде доли рынка. Следовательно, обычно предполагается, что i максимизирует свою полезность, выбирая j таким образом, что u_{ij} наибольшая из всех u_{ir}, $r = 1, ..., J$. Вероятность того, что i выбирает j, можно вычислить, если мы знаем распределения, лежащие в основе функции полезности. В настоящем эссе описаны основные свойства хорошо известных мультиномиальных моделей дискретного выбора: стандартной логит-модели, смешанной логит-модели и пробит-модели.

Изначальным источником логита и пробита было приложение вероятностных моделей к биологическим опытам бинарного выбора в первой половине прошлого века. Первые приложения в экономике использовали биномиальную пробит-модель (например, Farrell, 1954), которая, по-видимому, является победителем в споре между биологами по поводу способности логита и пробита моделировать бинарный выбор. Theil (1969) обобщил биномиальную логит-модель до мультиномиальной логит-модели, открыв дорогу дальнейшим улучшениям и приложениям. В начале 70-х McFadden с соавторами, которые изучали некоторые исследовательские проблемы транспортировки, обобщили логит-модель в нескольких направлениях и сделали ее научно признанной, предоставив теоретическую основу в теории полезности дискретного выбора (например, McFadden, 1973; McFadden & Reid, 1975; McFadden, 1977). Более детальные исторические данные содержатся в Cramer (1991, стр. 39–42) и Anderson, de Palma & Thisse (1992, глава 2).

*Перевод М. Кузина и С. Анатольева. Цитировать как: Шандор, Золт (2009). «Мультиномиальные модели дискретного выбора», Квантиль, №7, стр. 9–19. Citation: Sándor, Zsolt (2009). “Multinomial discrete choice models,” Quantile, No.7, pp. 9–19.

†Адрес: Department of Economics and Econometrics, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands. Электронная почта: *Z.Sandor@rug.nl*

2 Стандартная логит-модель

В этом разделе мы кратко представляем стандартную логит-модель, обсуждаем ее главные теоретические преимущества и практические недостатки. Затем мы выведем некоторые основные формулы, связанные с асимптотическими свойствами оценки максимального правдоподобия модели.

2.1 Спецификация модели

В стандартной логит-модели полезность является линейной функцией свойств альтернативы:

$$u_{ij} = x'_{ij}\beta + \varepsilon_{ij}, \tag{1}$$

где x_{ij} – $K \times 1$ вектор, содержащий характеристики потребителя i и альтернативы j, β – $K \times 1$ вектор параметров, а переменные ε_{ij}, $j = 1, ..., J$, предполагаются случайными и имеющими независимые стандартные распределения экстремальных значений типа I, кумулятивная функция распределения которого равна

$$F\left(\varepsilon_{ij}\right) = \exp\left(-\exp\left(-\varepsilon_{ij}\right)\right), \tag{2}$$

а функция плотности равна

$$f\left(\varepsilon_{ij}\right) = \exp\left(-\varepsilon_{ij}\right) \cdot \exp\left(-\exp\left(-\varepsilon_{ij}\right)\right). \tag{3}$$

В силу принципа максимизации полезности, вероятность того, что i выбирает альтернативу j, есть

$$s_{ij} = \Pr\left\{u_{ij} \geq u_{ir},\ \text{для всех } r = 1, ..., J\right\}.$$

Используя выражение для полезности (1), эта вероятность равна

$$s_{ij} = \Pr\left\{\varepsilon_{ir} \leq \varepsilon_{ij} + x'_{ij}\beta - x'_{ir}\beta,\ \text{для всех } r = 1, ..., J\right\}.$$

Вероятности такого типа обычно вычисляют как J-мерный интеграл, используя совместное распределение случайных элементов:

$$s_{ij} = \int \cdots \int_{\substack{\varepsilon_{ir} \leq \varepsilon_{ij} + x'_{ij}\beta - x'_{ir}\beta \\ \text{для всех } r=1,...,J}} f\left(\varepsilon_{i1}, ..., \varepsilon_{iJ}\right) \mathrm{d}\varepsilon_{i1} ... \mathrm{d}\varepsilon_{iJ}.$$

Так как все ε независимы, имеем

$$s_{ij} = \int \cdots \int_{\substack{\varepsilon_{ir} \leq \varepsilon_{ij} + x'_{ij}\beta - x'_{ir}\beta \\ \text{для всех } r=1,...,J}} f\left(\varepsilon_{i1}\right) \cdots f\left(\varepsilon_{iJ}\right) \mathrm{d}\varepsilon_{i1} ... \mathrm{d}\varepsilon_{iJ},$$

что равно

$$\int_{\mathbb{R}} f\left(\varepsilon_{ij}\right) \int \cdots \int_{\substack{\varepsilon_{ir} \leq \varepsilon_{ij} + x'_{ij}\beta - x'_{ir}\beta \\ r \neq j}} \prod_{r \neq j} f\left(\varepsilon_{ir}\right) \mathrm{d}\varepsilon_{i1} ... \mathrm{d}\varepsilon_{iJ}.$$

Наиболее крупная, $(J-1)$-мерная часть этого интеграла равна произведению $J-1$ одномерных интегралов плотностей, так что мы можем использовать кумулятивные функции распределения:

$$s_{ij} = \int_{\mathbb{R}} f\left(\varepsilon_{ij}\right) \cdot \prod_{r \neq j} F\left(\varepsilon_{ij} + x'_{ij}\beta - x'_{ir}\beta\right) \mathrm{d}\varepsilon_{ij}.$$

Используя выражения для этих функций (2) и (3) , получаем

$$s_{ij} = \int_{\mathbb{R}} \exp\left(-\varepsilon_{ij}\right) \cdot \exp\left(-\sum_{r=1}^{J} \exp\left(-\left(\varepsilon_{ij} + x'_{ij}\beta - x'_{ir}\beta\right)\right)\right) \mathrm{d}\varepsilon_{ij}.$$

Этот интеграл имеет явный вид и может быть вычислен следующим способом:

$$s_{ij} = \frac{1}{\sum_{r=1}^{J} \exp\left(-(x'_{ij}\beta - x'_{ir}\beta)\right)} \exp\left(-\sum_{r=1}^{J} \exp\left(-\left(\varepsilon_{ij} + x'_{ij}\beta - x'_{ir}\beta\right)\right)\right)\Bigg|_{\varepsilon_{ij}=-\infty}^{\infty},$$

что в итоге дает

$$s_{ij} = \frac{\exp(x'_{ij}\beta)}{\sum_{r=1}^{J} \exp\left(x'_{ir}\beta\right)}. \tag{4}$$

Благодаря этой привлекательной явной форме вероятности в (4) стандартная логит-модель и популярна в тех дисциплинах, в которых используются модели выбора.

В следующих разделах мы применим стандартную логит-модель в ситуациях, в которых характеристики варьируются по продуктам и не варьируются по потребителям, то есть когда $x_{ij} = x_j$ для всех $j = 1, ..., J$. Благодаря этому вероятность того, что потребитель i выбирает продукт j, равна

$$s_j = \frac{\exp(x'_j\beta)}{\sum_{r=1}^{J} \exp\left(x'_r\beta\right)} \tag{5}$$

для всех потребителей. Следовательно, это вероятность того, что продукт j приобретается на рынке. Эта величина на рынке с большим числом потребителей равно доле рынка продукта j.

Хотя выражение для вероятности того, что выбрана такая-то альтернатива, выглядит просто и потому доступно для практиков в вычислительном смысле, стандартная логит-модель также имеет серьезный недостаток. Этот недостаток возникает из-за простоты модели для полезности, а именно, из-за того, что полезности всех потребителей зависят от скалярной функции характеристик, $x'_j\beta$, а не от всего вектора характеристик x_j, поэтому отношение вероятностей двух альтернатив j и q не зависит от наличия и свойств любых других альтернатив, так как

$$\frac{s_j}{s_q} = \frac{\exp(x'_j\beta)}{\exp\left(x'_q\beta\right)}. \tag{6}$$

Это свойство известно как независимость от посторонних альтернатив. Это непривлекательное свойство, так как если мы добавляем новую альтернативу к набору всех альтернатив, которые являются близкими заменителями для j, но не для q, мы ожидаем, что s_j снизится значительно сильнее, чем s_q. Поэтому отношение двух вероятностей также должно уменьшиться.

Мы проиллюстрируем эту проблему на простом примере. Предположим, что на рынке есть два товара со следующими свойствами $x_1 = (1, 1, 1)'$ и $x_2 = (.5, 1.5, 1)'$, и это единственные альтернативы для выбора. Пусть $\beta = (1, 1, 1)'$. Тогда $s_1 = s_2 = \frac{1}{2}$. Теперь введем третий продукт со свойствами $x_3 = (1, 1, 1)'$. В этой новой ситуации $s_1 = s_2 = s_3 = \frac{1}{3}$, то есть стандартная логит-модель предсказывает, что потребители будут замещать продукты 1 и 2 в одинаковой степени. Это неверно, так как приблизительно половина потребителей предпочитает продукт 1, который является близким заменителем товару 2, так что часть потребителей, предпочитающих 1, должны переключиться на 3, а те, кто предпочитает 2, должны остаться с 2.

2.2 Логарифмическое правдоподобие и свойства ее первой производной

Так как далее мы используем результаты, связанные с оцениванием стандартной логит-модели, здесь мы приведем некоторые из этих результатов. Оценивание параметров стандартной логит-модели основано на идее о том, что мы находимся в ситуации мультиномиального выбора, когда вероятность s_j того, что выбрана альтернатива j, известна. Тогда если мы наблюдаем большое число выборов, частота альтернативы j, то есть число раз, когда j выбрана, деленное на общее число выборов, должно равняться s_j. Обозначим частоту альтернативы j за f_j. Пусть n – общее число выборов, то есть число потребителей. Тогда $n \cdot f \equiv n \cdot (f_1, ..., f_J)'$ имеет мультиномиальное распределение с параметрами $s_1, ..., s_J$. Поэтому вероятность того, что альтернативы $1, ..., J$ встречаются с частотами $f_1, ..., f_J$, соответственно, на самом деле равна функции правдоподобия, соответствующей стандартной логит-модели:

$$L = C \cdot \left(s_1^{f_1} \cdots s_J^{f_J} \right)^n,$$

где C – коэффициент, соответствующий мультиномиальному распределению. Так как этот коэффициент не зависит от интересующих нас параметров, в дальнейшем мы игнорируем его, когда записываем логарифмическую функцию правдоподобия, которая равна

$$\ln L = n \cdot \sum_{j=1}^{J} f_j \ln s_j. \tag{7}$$

Оценивание параметров можно выполнить максимизацией логарифмической функции правдоподобия. Обозначим за S диагональную матрицу с диагональю $s = (s_1, ..., s_J)'$. Можно записать $\ln L = n \cdot f' \ln s$. Следовательно, производная первого порядка логарифмической функции правдоподобия может быть записана как

$$\frac{\partial \ln L}{\partial \beta} = n \cdot \left(f' \frac{\partial \ln s}{\partial \beta'} \right)' = n \cdot \left(\frac{\partial s}{\partial \beta'} \right)' S^{-1} f,$$

так как

$$\frac{\partial \ln s}{\partial \beta'} = S^{-1} \frac{\partial s}{\partial \beta'}.$$

Для того чтобы вывести $\dfrac{\partial s}{\partial \beta'}$, заметим, что

$$\frac{\partial s_j}{\partial \beta'} = x_j' s_j - s' X s_j,$$

где $X = (x_1 \ ... \ x_J)'$, что приводит к

$$\frac{\partial s}{\partial \beta'} = \left(S - ss' \right) X. \tag{8}$$

Эту формулу можно использовать для того чтобы получить первые производные логарифмической функции правдоподобия:

$$\frac{\partial \ln L}{\partial \beta} = n \cdot X' \left(S - ss' \right) S^{-1} f.$$

Теперь можно вычислить асимптотическую информационную матрицу:

$$\mathcal{I} \left(\beta | X \right) = \mathbb{E} \left[\frac{\partial \ln L}{\partial \beta} \frac{\partial \ln L}{\partial \beta'} \right] = n^2 \cdot X' \left(S - ss' \right) S^{-1} \mathbb{E} \left[f f' \right] S^{-1} \left(S - ss' \right) X.$$

Мультиномиальное распределение имеет свойство $\mathrm{E}[ff'] = \frac{1}{n}S + \left(1 - \frac{1}{n}\right) ss'$. Подставляя это выражение в предыдущую формулу, после некоторых вычислений получаем

$$\mathcal{I}\left(\beta|X\right) = n \cdot X' \left(S - ss'\right) X.$$

Эта формула использовалась напрямую для построения модели с улучшенной эффективностью для экспериментов выбора (см. например, Sandor & Wedel, 2001).

3 Смешанная логит-модель

Смешанная логит-модель не удовлетворяет условию независимости от посторонних альтернатив. Это происходит из-за того, что в спецификации полезности коэффициенты предполагаются случайными. Таким образом, потребители имеют различные коэффициенты в функциях полезности. Такая спецификация имеет следующее привлекательное свойство: она параметризует отношение потребителей к характеристикам альтернатив через вариацию случайных параметров. Как в стандартной логит-модели, первые экономические приложения смешанной логит-модели, по-видимому, встретились в сфере транспортных исследований (Boyd & Mellman, 1980; Cardell & Dunbar, 1980). Мы отсылаем читателей к более детальным обзорам смешанных логит-моделей McFadden & Train (2000) и Brownstone & Train (1999). В данном разделе мы представим модель и основные асимптотические свойства оценки максимального правдоподобия.

3.1 Спецификация модели

В структуре выбора, аналогичной случаю стандартной логит-модели, полезность потребителя i, выбирающего альтернативу j, определяется как

$$u_{ij} = x_j'\beta_i + \varepsilon_{ij}, \tag{9}$$

где единственное отличие от (1) заключается в том, что здесь β_i – вектор параметров, характерных для потребителя i. Мы предполагаем, что $\beta_i \sim N(\beta, \Sigma)$, $i = 1, ..., n$, независимые и одинаково распределенные случайные величины, где β – $K \times 1$ вектор параметров, а Σ – симметричная положительно определенная матрица параметров. Ошибки ε_{ij} предполагаются независимыми от β_i.

Для удобства вычислений запишем случайный вектор параметров в следующем виде:

$$\beta_i = \beta + \Lambda v_i, \tag{10}$$

где Λ – некая матрица, такая, что $\Sigma = \Lambda\Lambda'$, а v_i – $K \times 1$ случайный вектор, имеющий стандартное нормальное распределение.

Если бы мы знали v_i, мы могли бы использовать формулу (5), поскольку это была бы стандартная логит-модель. В этом случае вероятность того, что альтернатива j выбрана, была бы

$$\Pr\{j|v_i\} = \frac{\exp\left(x_j'\left(\beta + \Lambda v_i\right)\right)}{\sum_{r=1}^{J} \exp\left(x_r'\left(\beta + \Lambda v_i\right)\right)}. \tag{11}$$

Безусловная вероятность того, что альтернатива j выбрана, определяется формулой

$$s_j = \mathbb{E}\left[\Pr\{j|v_i\}\right] = \int_{\mathbb{R}^K} \frac{\exp\left(x_j'\left(\beta + \Lambda v_i\right)\right)}{\sum_{r=1}^{J} \exp\left(x_r'\left(\beta + \Lambda v_i\right)\right)} \phi\left(v_i\right) \mathrm{d}v_i, \tag{12}$$

где $\phi(v_i)$ – функция плотности v_i. Заметим, что поскольку (6) в данном случае не выполняется, то смешанная логит-модель не удовлетворяет свойству независимости от посторонних альтернатив.

Выражение для вероятности в (12) представляет собой интеграл, который в общем случае нельзя вычислить аналитически. Обычно он оценивается с помощью симуляций Монте-Карло или квази-Монте-Карло. Детальное обсуждение технологии такого оценивания выходит за рамки данного эссе, поэтому мы лишь упомянем здесь ее суть. Извлекаем несколько значений случайного вектора v_i, вычисляем функцию под интегралом при каждом его значении, и находим среднее по этим значениям. Полученное среднее и есть (квази-) оценка интеграла методом Монте-Карло.

Чтобы проиллюстрировать, как в смешанной логит-модели моделируется отношение потребителей к различным характеристикам, рассмотрим более простой вид ковариационной матрицы, а именно, когда она диагональна:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_K^2 \end{pmatrix}.$$

В этом случае можно взять

$$\Lambda = \begin{pmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_K \end{pmatrix},$$

так что можно записать полезность в (9) с использованием (10) как

$$u_{ij} = x_j'\beta + \sum_{k=1}^{K} \sigma_k v_{ik} x_{jk} + \varepsilon_{ij}.$$

Записав полезность в таком виде, мы можем проинтерпретировать эффект случайной части параметров. Каждый потребитель i сопоставляется с вектором v_i, элементы которого представляют собой оценку потребителем характеристик. Если предпочтения i таковы, что для него более ценны альтернативы с большими значениями какой-то характеристики, то у такого потребителя будет большое положительное значение v_i, соответствующее этой характеристике. Например, в случае автомобилей, если потребитель i предпочитает большие машины, и «размер» – это характеристика h в этой модели, то этот потребитель будет иметь большое положительное v_{ih}.

Теперь можно объяснить преимущества смешанной логит-модели перед обычной тем, что схемы замен, порождаемые смешанной логит-моделью, подвергаются меньшим ограничениям. Продолжая пример с машинами, если на рынке появляется маленькая машина, то потребитель, о котором шла речь выше, не будет сильно стремиться к переключению на новую маленькую машину. Происходит это из-за того, что для этого потребителя значение $\sigma_h v_{ih}$ несоразмерно больше, чем соответствующие величины для других характеристик, и поэтому член $\sigma_h v_{ih} x_{jh}$ более чувствителен к вариациям в x_{jh}, чем в случае с другими характеристиками. Поэтому полезность потребителя i высока для больших значений x_{jh} и низка для малых.

Описанное выше подразумевает, что чем больше изменения параметров отношения потребителей, σ_k, тем менее вероятно, что потребитель, предпочитающий определенную альтернативу, будет переключаться на несхожую альтернативу. Проиллюстрируем это примером из конца раздела 2.1. В дополнение к случаю стандартной логит-модели здесь мы возьмем $(\sigma_1, \sigma_2, \sigma_3) = \lambda(1,1,1)$. Когда только товары 1 и 2 находятся на рынке, их вероятности $s_1 = s_2 = \frac{1}{2}$, независимо от значения λ. Когда товар 3 появляется на рынке, s_2 становится

0.35 при $\lambda = 1$, 0.42 при $\lambda = 5$, 0.47 при $\lambda = 10$ и 0.49 при $\lambda = 30$. Таким образом, для больших вариаций нет замещения товаром 3 товара 2, в то время как для малых вариаций замещение очень похоже на то, что возникало в стандартной логит-модели.

3.2 Логарифмическое правдоподобие и свойства ее первой производной

В этом разделе мы выведем некоторые полезные формулы для смешанной логит-модели. Обозначим условную вероятность в (11) за $\pi_j(v)$. Тогда

$$s_j = \int_{\mathbb{R}^K} \pi_j(v)\,\phi(v)\,\mathrm{d}v.$$

Модель выбора в данном случае можно сформулировать похожим образом, как для стандартной логит-модели. Тогда функция логарифмического правдоподобия в этой модели равна

$$\ln L = n \cdot \sum_{j=1}^{J} f_j \ln s_j,$$

где f_j вновь обозначает наблюдаемую частоту товара j.

Чтобы сократить обозначения, мы пишем интегралы типа $\int_{\mathbb{R}^K}(\cdot)\phi(v)\mathrm{d}v$ как $\int(\cdot)\mathrm{d}\Phi$ (например, $\int_{\mathbb{R}^K}\pi_j(v)\,\phi(v)\mathrm{d}v \equiv \int \pi_j \mathrm{d}\Phi$). При оценивании методом максимального правдоподобия мы находим первые производные функции логарифмического правдоподобия. Так как нам нужны эти результаты для случая, когда ковариационная матрица случайных коэффициентов Σ диагональна, мы выводим формулы только для этого случая. Для удобства вычислений мы изменили обозначение случайной части коэффициентов с Λv на $V\sigma$, где V – диагональная матрица с диагональю v, а $\sigma = (\sigma_1, ..., \sigma_K)'$. Аналогично случаю стандартной логит-модели,

$$\frac{\partial L}{\partial \beta} = n \cdot \left(\frac{\partial s}{\partial \beta'}\right)' S^{-1} f \quad \text{и} \quad \frac{\partial L}{\partial \sigma} = n \cdot \left(\frac{\partial s}{\partial \sigma'}\right)' S^{-1} f. \tag{13}$$

Далее нам потребуются следующие формулы:

$$\frac{\partial s}{\partial \beta'} = \int \frac{\partial \pi}{\partial \beta'}\mathrm{d}\Phi \quad \text{и} \quad \frac{\partial s}{\partial \sigma'} = \int \frac{\partial \pi}{\partial \sigma'}\mathrm{d}\Phi,$$

где $\pi = (\pi_1, ..., \pi_J)'$. Аналогично, как для стандартной логит-модели, получаем:

$$\frac{\partial \pi}{\partial \beta'} = \left(\Pi - \pi\pi'\right) X \quad \text{и} \quad \frac{\partial \pi}{\partial \sigma'} = \left(\Pi - \pi\pi'\right) XV,$$

где Π – диагональная матрица с диагональю π. Отсюда

$$\frac{\partial s}{\partial \beta'} = \int \left(\Pi - \pi\pi'\right) X \mathrm{d}\Phi \quad \text{и} \quad \frac{\partial s}{\partial \sigma'} = \int \left(\Pi - \pi\pi'\right) XV \mathrm{d}\Phi. \tag{14}$$

Подставляя эти равенства в (13), получаем:

$$\frac{\partial L}{\partial \beta} = n \cdot \int X' \left(\Pi - \pi\pi'\right) \mathrm{d}\Phi \cdot S^{-1} f \tag{15}$$

и

$$\frac{\partial L}{\partial \sigma} = n \cdot \int V X' \left(\Pi - \pi\pi'\right) \mathrm{d}\Phi \cdot S^{-1} f. \tag{16}$$

Асимптотическая информационная матрица определяется как

$$\mathcal{I}(\beta, \sigma | X) = \begin{bmatrix} \mathbb{E}\left[\frac{\partial L}{\partial \beta}\frac{\partial L}{\partial \beta'}\right] & \mathbb{E}\left[\frac{\partial L}{\partial \beta}\frac{\partial L}{\partial \sigma'}\right] \\ \mathbb{E}\left[\frac{\partial L}{\partial \sigma}\frac{\partial L}{\partial \beta'}\right] & \mathbb{E}\left[\frac{\partial L}{\partial \sigma}\frac{\partial L}{\partial \sigma'}\right] \end{bmatrix}.$$

Для упрощения этого выражения используем тот факт, что по предположению f имеет мультиномиальное распределение с $\mathbb{E}[f] = s$. Тогда f обладает свойством $\mathbb{E}[ff'] = \frac{1}{n}S + \left(1 - \frac{1}{n}\right)ss'$. Используя этот факт и формулы (15) и (16), находим:

$$\mathcal{I}(\beta, \sigma|X) = n \cdot \left[\begin{array}{cc} M'S^{-1}M & M'S^{-1}Q \\ Q'S^{-1}M & QS^{-1}Q \end{array} \right], \tag{17}$$

где

$$M = \int \left(\Pi - \pi\pi'\right) X\mathrm{d}\Phi \text{ и } Q = \int \left(\Pi - \pi\pi'\right) XV\mathrm{d}\Phi.$$

Формула (17) использовалась напрямую для построения экспериментов выбора для смешанной логит-модели (см. например, Sandor & Wedel, 2002).

4 Пробит

Пробит – это модель, похожая на смешанную логит-модель, в которой все случайные элементы, коэффициенты и ошибки по предположению имеют нормальное распределение. Одно из первых применений мультиномиальной пробит-модели в эконометрическом контексте встретилось в работе McFadden (1976). В этом разделе мы кратко представим модель и покажем способ оценивания вероятностей выбора, известный как GHK-симулятор (по первым буквам Geweke, Hajivassiliou и Keane, разработчиков симулятора для эконометрического контекста в начале 90-х). В отличие от стандартной и смешанной логит-моделей, для пробит-модели мы не приводим функцию логарифмического правдоподобия и свойства ее первых производных. Их краткое описание содержится в Wansbeek & Webel (2001).

4.1 Спецификация модели

Предположим, что полезности, соответствующие J продуктам для потребителя i, равны

$$u_{ij} = x_j'\beta_i + \varepsilon_{ij},$$

где, как и для смешанной логит-модели, $\beta_i \sim N(\beta, \Sigma)$, $i = 1, ..., n$, независимо и одинаково распределены, β – $K \times 1$ вектор параметров, а Σ – симметричная положительно определенная матрица параметров. Ошибка ε_{ij} здесь предполагается нормально распределенной без требования независимости от β_i. Благодаря привлекательному свойству нормального распределения, а именно, тому, что сумма двух нормально распределенных случайных величин нормально распределена, можно выбрать подходящие переменные μ и Ω таким образом, что вектор полезностей можно представить как

$$u_i = \mu + \Gamma e_i \quad \sim N(\mu, \Omega), \tag{18}$$

где $\mu : J \times 1$, $\Gamma : J \times J$, такая что $\Gamma\Gamma' = \Omega$, и $e_i \sim N(0, I_J)$. Если случайные члены e_i независимы, то можно опустить индекс i, и каждому потребителю сопоставить реализацию случайного вектора e. Отметим, что если в полезности в смешанной логит-модели (9) мы заменим предположение о распределении экстремальных значений остатков нормальным распределением, то получим частный случай (18). Причина, по которой последняя спецификация более общая, заключается в том, что в данном случае не предполагается независимости между остатками и случайными коэффициентами.

Из J продуктов выбран один с наибольшей полезностью. Вероятность того, что был выбран продукт j, равна

$$s_j = \Pr\{u_j \geq u_r, \forall r\} = \Pr\{V_j u \leq 0\} = \Pr\{V_j\mu + V_j\Gamma e \leq 0\}$$

с

$$V_j = \begin{bmatrix} 1 & 0 & \dots & -1 & \dots & 0 \\ 0 & 1 & \dots & -1 & \dots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \dots & -1 & \dots & 1 \end{bmatrix},$$

где столбец, состоящий из -1, стоит на j-ом месте. Обозначим $T \equiv V_j\Gamma$ и выберем Γ так, что матрица T нижняя треугольная. Это возможно сделать, так как $TT' = V_j\Omega V_j'$, а эта матрица положительно определена. Тогда

$$s_j = \Pr\left\{Te \leq v\right\}, \text{ где } v \equiv -V_j\mu. \tag{19}$$

Если мы обозначим

$$T = \begin{bmatrix} t_{11} & 0 & \dots & 0 \\ t_{12} & t_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ t_{J1} & t_{J2} & \dots & t_{JJ} \end{bmatrix},$$

то

$$s_j = \Pr\left\{e_1 \leq \frac{v_1}{t_{11}}, e_2 \leq \frac{v_2 - t_{21}e_1}{t_{22}}, \dots,\ e_J \leq \frac{v_J - t_{J1}e_1 - ... - t_{JJ-1}e_{J-1}}{t_{JJ}}\right\}.$$

Пусть

$$D = \left\{(e_1, ..., e_J) \in \mathbb{R}^J : e_1 \leq \frac{v_1}{t_{11}}, ..., e_J \leq \frac{v_J - t_{J1}e_1 - ... - t_{JJ-1}e_{J-1}}{t_{JJ}}\right\}.$$

Тогда формула для вероятности того, что выбрана альтернатива j, выглядит как

$$s_j = \int_D \phi\left(e_1\right)...\phi\left(e_J\right) \mathrm{d}e_1...\mathrm{d}e_J, \tag{20}$$

где ϕ – плотность стандартного нормального распределения.

4.2 Оценивание вероятностей

Чтобы получить более удобные в вычислительном смысле формулы вероятности, заметим, что e имеют усеченное нормальное распределение. Если $e \leq b$ и $e \sim N(0,1)$, тогда можно получить e быстрее, извлекая u равномерно распределенным на $[0,1]$ и беря $e = \Phi^{-1}\left(u \cdot \Phi\left(b\right)\right)$, где Φ – стандартная нормальная функция распределения. Оказывается, что основываясь на этой идее, можно трансформировать интеграл (20) в интеграл, аргументы которого принадлежат единичному гиперкубу.

Для того чтобы произвести трансформацию, можно воспользоваться идеей из предыдущего раздела в J-мерном случае. Это подразумевает трансформацию

$$\begin{aligned} e_1 &= \Phi^{-1}\left(u_1\Phi\left(\frac{v_1}{t_{11}}\right)\right) \equiv \psi_1\left(u_1, ..., u_J\right), \\ e_2 &= \Phi^{-1}\left(u_2\Phi\left(\frac{v_2 - t_{21}e_1\left(u_1\right)}{t_{22}}\right)\right) \equiv \psi_2\left(u_1, ..., u_J\right), \\ &\vdots \\ e_J &= \Phi^{-1}\left(u_J\Phi\left(\frac{v_J - t_{J1}e_1 - ... - t_{JJ-1}e_{J-1}}{t_{JJ}}\right)\right) \equiv \psi_3\left(u_1, ..., u_J\right), \end{aligned} \tag{21}$$

где $u_1, ..., u_J \in [0,1]$, и в последней формуле e_j рассматривается как функция от $u_1, ..., u_j$ для $j = 1, ..., J-1$. Тогда (20) становится

$$\int_{[0,1]^J} \phi\left(\psi_1\left(u_1, ..., u_J\right)\right) ... \phi\left(\psi_J\left(u_1, ..., u_J\right)\right) |J| \, \mathrm{d}u_1 ... \mathrm{d}u_J,$$

где $|J|$ – якобиан, то есть

$$|J| = \begin{vmatrix} \frac{\partial\psi_1}{\partial u_1} & \cdots & \frac{\partial\psi_J}{\partial u_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial\psi_1}{\partial u_J} & \cdots & \frac{\partial\psi_J}{\partial u_J} \end{vmatrix} = \begin{vmatrix} \frac{1}{\phi(\psi_1(...))}\Phi\left(\frac{v1}{t_{11}}\right) & \cdots & \frac{\partial\psi_J}{\partial u_1} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\phi(\psi_J(...))}\Phi\left(\frac{v_J - t_{J1}e_1 - ... - t_{JJ-1}e_{J-1}}{t_{JJ}}\right) \end{vmatrix},$$

что является определителем верхней треугольной матрицы, поэтому якобиан равен произведению диагональных элементов:

$$|J| = \frac{\Phi\left(\frac{v_1}{t_{11}}\right) \cdots \Phi\left(\frac{v_J - t_{J1}e_1 - ... - t_{JJ-1}e_{J-1}}{t_{JJ}}\right)}{\phi\left(\psi_1\left(...\right)\right) \cdots \phi\left(\psi_J\left(...\right)\right)}.$$

Используя полученные формулы, можно преобразовать (20) в

$$\int_{[0,1]^J} \Phi\left(\frac{v_1}{t_{11}}\right) \cdots \Phi\left(\frac{v_J - t_{J1}e_1 - ... - t_{JJ-1}e_{J-1}}{t_{JJ}}\right) \mathrm{d}u_1 ... \mathrm{d}u_J.$$

Тогда получаем формулу для вероятности

$$s_j = \Phi\left(\frac{v_1}{t_{11}}\right) \int_{[0,1]^{J-1}} \Phi\left(\frac{v_2 - t_{21}e_1}{t_{22}}\right) \cdots \Phi\left(\frac{v_J - t_{J1}e_1 - ... - t_{JJ-1}e_{J-1}}{t_{JJ}}\right) \mathrm{d}u_1 ... \mathrm{d}u_{J-1}.$$

Интеграл в этой формуле можно интерпретировать как ожидание от подынтегральной функции при равномерно распределенном случайном $(J-1)$-векторе. Этот интеграл можно оценить с помощью симуляций Монте-Карло, извлекая набор случайных векторов и вычисляя их среднее. Оценка, найденная таким способом, называется GHK-симулятором (Börsch-Supan & Hajivassiliou, 1993) или RIS-симулятором (от *recursive importance sampling*), основанным на усеченной нормальной плотности (Vijverberg, 1997). Vijverberg (1997) также обсуждает другие RIS-симуляторы. Ряд других типов симуляторов представлен в работе Hajivassiliou, McFadden & Ruud (1996).

Благодарности

Автор благодарит Тома Вансбеека за многочисленные полезные комментарии.

Список литературы

Anderson, S.P., A. de Palma & J-F. Thisse (1992). *Discrete Choice Theory of Product Differentiation.* Cambridge: MIT Press.

Börsch-Supan, A. & V.A. Hajivassiliou (1993). Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models. *Journal of Econometrics* 58, 347–368.

Boyd, H.J. & R.E. Mellman (1980). The effect of fuel economy standards on the U.S. automotive market: An hedonic demand analysis. *Transportation Research* 14, 367–378.

Brownstone, D. & K. Train (1999). Forecasting new product penetration with flexible substitution patterns. *Journal of Econometrics* 89, 109–129.

Cardell, N.S. & F. Dunbar (1980). Measuring the societal impacts of automobile downsizing. *Transportation Research* 14, 423–434.

Cramer, J.S. (1991). *The Logit Model: An Introduction for Economists.* London: Arnold.

Farrell, M.J. (1954). The demand for motorcars in the United States. *Journal of the Royal Statistical Society, A* 117, 171–200.

Hajivassiliou, V., D. McFadden & P. Ruud (1996). Simulation of multivariate normal rectangle probabilities and their derivatives: Theoretical and computational results. *Journal of Econometrics* 72, 85–134.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. Глава в *Frontiers in Econometrics* под редакцией P. Zarembka. New York: Academic Press.

McFadden, D. (1976). Quantal choice analysis: A survey. *Annals of Economic and Social Measurement* 5, 363–90.

McFadden, D. (1977). Econometric models of probabilistic choice. Стр. 171–260 в *Structural Analysis of Discrete Data with Econometric Applications* под редакцией C.F. Manski & D. McFadden. Cambridge: MIT Press.

McFadden, D. & F. Reid (1975). Aggregate travel demand forecasting from disaggregated behavioral models. *Transportation Research Record* 534, 24–37.

McFadden, D. & K. Train (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics* 15, 447–470.

Sándor, Z. & M. Wedel (2001). Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research* 38, 430–444.

Sándor, Z. & M. Wedel (2002). Profile construction in experimental choice designs for mixed logit models. *Marketing Science* 21, 55–75.

Theil, H. (1969). A multinomial extension of the linear logit model. *International Economic Review* 10, 251–259.

Vijverberg, W.P.M. (1997). Monte Carlo evaluation of multivariate normal probabilities. *Journal of Econometrics* 76, 281–307.

Wansbeek, T. & M. Wedel (2001). The structure of the multinomial probit Model. Working Paper, University of Groningen.

Multinomial discrete choice models

Zsolt Sándor

University of Groningen, Groningen, Netherlands

This essay briefly describes the main features of some well-known multinomial discrete choice models: the standard logit, the mixed logit, and the probit.

Заметки о моделях с самоотбором выборки[⋆]

Виктор Агиррегабирия[†]

Университет Торонто, Торонто, Канада

Проблемы, связанные с самоотбором выборки, часто встречаются при работе с микроэкономическими моделями и данными по индивидам, домашним хозяйствам или фирмам. За последние тридцать лет в этой области эконометрики были сделаны весьма значительные достижения. Предложены и применены на практике различные типы моделей. Разработаны новые методы оценивания и инференции, как параметрические, так и полупараметрические. Настоящее эссе является кратким введением в эту обширную литературу.

1 Введение

Рассмотрим модель регрессии:

$$Y^* = X^*\beta + \varepsilon, \tag{1}$$

где Y^* и ε – скалярные случайные величины, X^* – $1\times K$ вектор случайных величин, а β – $K\times 1$ вектор параметров. Случайная ошибка ε независима в среднем от X^*, и матрица $\mathbb{E}[X^{*\prime}X^*]$ имеет полный ранг. Тогда при наличии случайной выборки для (Y^*, X^*) МНК-оценка параметров состоятельна и асимптотически нормальна. Ключевая особенность моделей с самоотбором выборки состоит в том, что исследователь не наблюдает случайную выборку для (Y^*, X^*). Вместо нее имеется случайная выборка для пары переменных (Y, X), которые связаны с (Y^*, X^*), но отличаются от них. Переменные (Y^*, X^*) называют латентными. Задача заключается в состоятельном оценивании β по выборке для (Y, X).[1] В зависимости от соотношения между латентными и наблюдаемыми переменными выделяют различные классы моделей с самоотбором выборки. Далее запись $Y = \{X|Z > c\}$ означает, что Y – это случайная величина X при условии, что случайная величина Z больше константы c. Аналогично определяются выражения $Y = \{X|Z < c\}$ и $Y = \{X|b < Z < c\}$.

(a) *Модель усеченной регрессии.* Пусть c – известная константа. Если переменная Y усечена слева в точке c, то

$$(Y, X) = \{(Y^*, X^*)|Y^* > c\}. \tag{2}$$

Если переменная Y усечена справа в точке c, то

$$(Y, X) = \{(Y^*, X^*)|Y^* < c\}. \tag{3}$$

В обоих случаях случайная выборка для пары (Y, X) не является случайной выборкой ни для Y^*, ни для X^*.[2]

[⋆]Перевод Б. Гершмана. Цитировать как: Агиррегабирия, Виктор (2009). «Заметки о моделях с самоотбором выборки», Квантиль, №7, стр. 21–36. Citation: Aguirregabiria, Victor (2009). “Some notes on sample selection models,” Quantile, No.7, pp. 21–36.

[†]Адрес: 150 St. George Street, Toronto, ON, M5S 3G7. Электронная почта: *victor.aguirregabiria@utoronto.ca*

[1]В некоторых приложениях интерес также представляет оценивание функции распределения случайной ошибки ε.

[2]Случайная выборка для пары (Y, X) приводит к случайной выборке для X^* только в случае, когда Y^* и X^* независимо распределены.

Пример 1. Рассмотрим уравнение для логарифма заработной платы, $W^* = X^*\beta + \varepsilon$, где W^* – логарифм заработной платы индивида, а X^* – вектор наблюдаемых характеристик его человеческого капитала. Предположим, что из-за конфиденциальности имеющаяся в распоряжении база данных не содержит информацию (ни о заработной плате, ни о прочих характеристиках) по индивидам с почасовой заработной платой более 800 долл. Тогда наблюдаются переменные (W, X), такие, что $(W, X) = \{(W^*, X^*) | W^* < \ln 800\}$. В этом случае говорят, что зависимая переменная усечена справа, и рассматривают модель усеченной регрессии, поскольку ни W^*, ни X^* не наблюдаются, когда заработная плата превышает 800 долл. в час.

Пусть f_{Y^*} и F_{Y^*} – функция плотности распределения (ФПР) и кумулятивная функция распределения (КФР) случайной величины Y^*, соответственно. Если переменная Y усечена слева в точке c, то ФПР Y имеет вид

$$f_Y(y) = \begin{cases} 0 & \text{при} \quad y \le c, \\ \dfrac{f_{Y^*}(y)}{1 - F_{Y^*}(c)} & \text{при} \quad y > c. \end{cases} \tag{4}$$

Если переменная Y усечена справа в точке c, то

$$f_Y(y) = \begin{cases} \dfrac{f_{Y^*}(y)}{F_{Y^*}(c)} & \text{при} \quad y < c, \\ 0 & \text{при} \quad y \ge c. \end{cases} \tag{5}$$

На рисунке 1 представлены ФПР нормальных случайных величин, усеченных слева и справа.

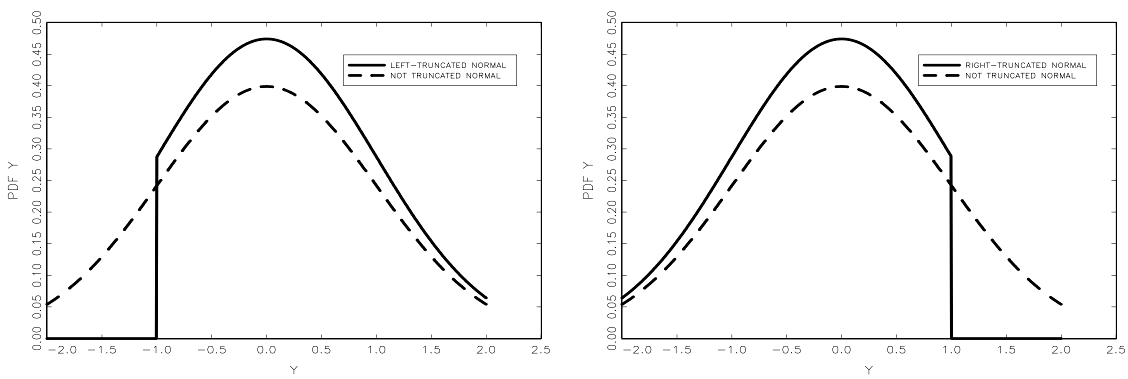

Рис. 1: Нормальные случайные величины, усеченные слева и справа, соответственно.

(b) *Модель цензурированной регрессии (или тобит-модель).* Основное отличие этой модели от модели усеченной регрессии состоит в том, что имеется случайная выборка для экзогенных регрессоров X^*. То есть случайные переменные X и X^* совпадают. Что касается зависимой переменной, то при цензурировании слева

$$Y = \max\{Y^*; c\} = \begin{cases} c & \text{при} \quad Y^* \le c, \\ Y^* & \text{при} \quad Y^* > c. \end{cases} \tag{6}$$

При цензурировании справа

$$Y = \min\{Y^*; c\} = \begin{cases} Y^* & \text{при} \quad Y^* < c, \\ c & \text{при} \quad Y^* \ge c. \end{cases} \tag{7}$$

Пример 2. Рассмотрим уравнение для логарифма заработной платы из примера 1. Теперь в распоряжении исследователя другой набор данных. Он включает информацию по всем индивидам независимо от уровня дохода. Доступна случайная выборка индивидов, содержащая информацию о переменных X. Но из-за конфиденциальности данные по наиболее высоким заработным платам недоступны. Если индивид имеет почасовую заработную плату менее 800 долл., наблюдается ее реальное значение. Но для индивидов, зарабатывающих более 800 долл. в час, данные о заработной плате отсутствуют. Следовательно, для каждого индивида в выборке наблюдается цензурированное справа значение логарифма заработной платы $W = \min\{W^*; \ln 800\}$. Зависимая переменная цензурирована справа, поэтому рассматривается модель цензурированной регрессии.

Пример 3. Рассмотрим следующую модель инвестиционных вложений фирмы в определенный вид оборудования, например, компьютеры. Пусть Q^* – «желаемый» объем инвестиций фирмы в соответствии с некоторой экономической моделью оптимальных инвестиций, например, объем инвестиций, максимизирующий прибыль без ограничения на неотрицательность Q^*, то есть $Q^* = \arg\max_q \Pi(q)$, где $\Pi(q)$ – (межвременная) функция прибыли. Предположим, что из этой модели следует следующее уравнение регрессионного типа: $Q^* = X\beta + \varepsilon$. Вектор X включает характеристики фирмы и рынка капитала, на котором действует фирма, такие как запас оборудования и цена нового капитала. β – вектор параметров, имеющих четкую экономическую интерпретацию в рамках модели. Имеется случайная выборка фирм, для которых наблюдается X и объем инвестиций Q. Глядя на эмпирическое распределение объема инвестиций Q, становится очевидным, что эта переменная всегда положительна с некоторой вероятностной массой в нуле. Эти свойства распределения нельзя объяснить предыдущей моделью регрессии, если не делать необоснованных предположений о распределении ε. Более того, рассмотренная теоретическая модель предполагает, что объем инвестиций Q^* может быть как положительным, так и отрицательным, а это противоречит наблюдаемым значениям Q. Рассмотрим тогда следующую модель для Q: $Q = \arg\max_q \Pi(q)$ при ограничении $q \geq 0$. Если функция прибыли $\Pi(q)$ строго вогнута, легко показать, что $Q = Q^*$ при $Q^* > 0$ и $Q = 0$ при $Q^* \leq 0$. То есть $Q = \max\{Q^*; 0\}$, где $Q^* = X\beta + \varepsilon$. С экономической точки зрения эту модель можно интерпретировать как модель необратимых инвестиций. С эконометрической точки зрения, это модель цензурированной регрессии.

Примеры 2 и 3 представляют две различные модели цензурированной регрессии. Интересно отметить некоторые существенные различия между этими двумя примерами. Они основаны на весьма разных экономических и статистических предпосылках. В примере 2 цензурирование является свойством выборки. Заработная плата индивидов, превышающая 800 долл. в час, не является теоретическим объектом, а реально существует, хоть и не наблюдается в выборке. В примере 3 цензурирование – это предположение модели. Принимая во внимание определенные свойства распределения объема инвестиций, предполагается, что для этой переменной разумно рассматривать модель цензурированной регрессии. Переменная Q^* представляет собой теоретический объект, и для нее невозможно получить случайную выборку. Тем не менее параметры β могут иметь четкую экономическую интерпретацию в рамках этой модели и представляют интерес.

Пусть f_{Y^*} и F_{Y^*} – ФПР и КФР случайной величины Y^*, соответственно. Если переменная Y цензурирована слева в точке c, то ее ФПР имеет вид

$$f_Y(y) = \begin{cases} 0 & \text{при} \quad y < c, \\ F_{Y^*}(c) & \text{при} \quad y = c, \\ f_{Y^*}(y) & \text{при} \quad y > c. \end{cases} \tag{8}$$

Если переменная Y цензурирована справа в точке c, то

$$f_Y(y) = \begin{cases} f_{Y^*}(y) & \text{при} \quad y < c, \\ 1 - F_{Y^*}(c) & \text{при} \quad y = c, \\ 0, & \text{при} \quad y > c. \end{cases} \tag{9}$$

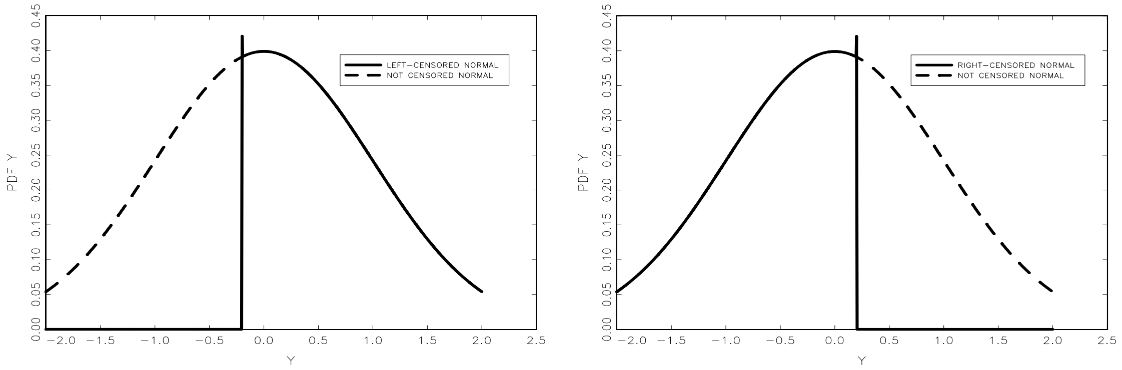

Рис. 2: Нормальные случайные величины, цензурированные слева и справа, соответственно.

На рисунке 2 представлены ФПР нормальных случайных величин, цензурированных слева и справа.
(c) *Модель с самоотбором выборки.* В простой модели с самоотбором выборки Y^* наблюдается только для индивидов, для которых определенная бинарная переменная, D, равна единице. Эта бинарная переменная не является независимой от Y^*.

$$Y = \{Y^*|D = 1\} \tag{10}$$

и ФПР$(Y^*|D = 1) \neq$ ФПР$(Y^*|D = 0)$. Заметим, что если D и Y^* независимы, то случайные переменные Y и Y^* совпадают и проблема самоотбора выборки отсутствует. Есть два вида моделей с самоотбором выборки: усеченного типа и цензурированного типа. В модели усеченного типа X^* также не наблюдается при $D = 0$. То есть в модели усеченного типа

$$(Y, X) = \{(Y^*, X^*)|D = 1\}. \tag{11}$$

В модели цензурированного типа имеется случайная выборка для X^* (то есть $X = X^*$). Тогда

$$(Y, X) = (\{Y^*|D = 1\}, X^*). \tag{12}$$

В такой модели цензурированного типа иногда удобно определить Y следующим образом: $Y = Y^*$ при $D = 1$, и $Y = 0$ при $D = 0$. Или в более компактной форме: $Y = DY^*$. Заметим, что модели усеченной и цензурированной регрессии являются частными случаями модели с самоотбором выборки. При $D = \mathbb{I}\{Y^* > c\}$ модель с самоотбором выборки становится моделью регрессии, усеченной/цензурированной слева, а при $D = \mathbb{I}\{Y^* < c\}$ получаем модель регрессии, усеченную/цензурированную справа.

Пример 4. Рассмотрим снова уравнение для логарифма заработной платы из примеров 1 и 2. Однако теперь интерес представляет не исключительно популяция работающих индивидов, а все индивиды, составляющие рабочую силу, занятые и нет. В таком случае W^* интерпретируется как латентная рыночная заработная плата индивида, и она существует независимо от того, работает индивид или нет. Имеется случайная выборка занятых и незанятых индивидов. Таким образом, есть случайная выборка для X^*, и рассматривается модель с самоотбором выборки цензурированного типа. Но рыночная заработная плата W^* наблюдается только для занятых индивидов. Пусть D – индикатор события «индивид работает». Тогда имеется случайная выборка для переменной W, где $W = \{W^*|D = 1\}$. Индикатор занятости D зависит от разных факторов, включая характеристики человеческого капитала, наблюдаемые и не наблюдаемые эконометристом. Следовательно, D и W^* не являются независимыми, и возникает проблема самоотбора выборки.

Спецификация модели с самоотбором выборки должна включать некоторые предположения о совместном распределении Y^* и D. Распространенная спецификация имеет вид

$$D = \mathbb{I}\{Z\gamma - u > 0\}, \tag{13}$$

где $\mathbb{I}\{\cdot\}$ – индикаторная функция, Z – вектор наблюдаемых переменных, γ – вектор параметров, и u не наблюдается. Переменные (X, Z) экзогенны, то есть независимы от случайных величин (u, ε). Условно на (X, Z) ненаблюдаемые величины u и ε не являются независимо распределенными.

(d) *Обобщенная модель с самоотбором выборки.* Рассмотрим следующую систему J линейных уравнений:

$$\begin{array}{rcl} Y_1^* & = & X^*\beta_1 + \varepsilon_1, \\ Y_2^* & = & X^*\beta_2 + \varepsilon_2, \\ \vdots & & \vdots \\ Y_J^* & = & X^*\beta_J + \varepsilon_J. \end{array} \tag{14}$$

Предположим, что наблюдается случайная выборка для X^*, то есть имеет место модель с самоотбором выборки цензурированного типа, где $X = X^*$.[3] Однако для каждого индивида в выборке не наблюдается весь набор J зависимых переменных $(Y_1^*, Y_2^*, ..., Y_J^*)$. Вместо этого, для каждого индивида наблюдается дискретная переменная $D \in \{1, 2, ..., J\}$ и зависимая переменная Y, такая, что

$$Y = \sum_{j=1}^{J} \mathbb{I}\{D = j\} Y_j^*. \tag{15}$$

Каждый индивид наблюдается только в одном *режиме*. Важно, что дискретная переменная D не является независимой от случайных ошибок ε_j в системе линейных уравнений.

Пример 5 (Модель Роя)[4]. Рассмотрим индивида, выбирающего между двумя возможными занятиями, 1 и 2. Предположим, что этот индивид выбирает работу, которая дает наибольший (за всю жизнь) заработок. При данных наблюдаемых и ненаблюдаемых характеристиках индивида доходы от двух занятий равны

$$\begin{array}{l} W_1^* = X\beta_1 + \varepsilon_1, \\ W_2^* = X\beta_2 + \varepsilon_2. \end{array} \tag{16}$$

Вектор X содержит наблюдаемые характеристики человеческого капитала, такие как образование и опыт работы. Векторы параметров β_1 и β_2 измеряют отдачу от характеристик человеческого капитала в занятиях 1 и 2, соответственно. ε_1 и ε_2 представляют собой отдачу от ненаблюдаемых (эконометристом, но не индивидом) характеристик человеческого капитала. Каждый индивид имеет только одно занятие. Пусть D – индикатор события «индивид выбирает занятие 1». Тогда наблюдаемый заработок индивида, W, можно представить в виде

$$W = DW_1^* + (1 - D)W_2^*. \tag{17}$$

При предположении, что индивиды максимизируют доход, получаем, что

$$D = \mathbb{I}\{W_1^* > W_2^*\} = \mathbb{I}\{X(\beta_1 - \beta_2) - (\varepsilon_2 - \varepsilon_1) > 0\}. \tag{18}$$

Ясно, что ненаблюдаемая величина $\varepsilon_2 - \varepsilon_1$ в уравнении для бинарной переменной выбора не является независимой от ненаблюдаемых величин в уравнениях для доходов, ε_1 и ε_2. В этих

[3] Можно также рассматривать версию этой модели, в которой переменная X^* усечена в некотором режиме $j \in \{1, 2, ..., J\}$.

[4] См. Roy (1951) и Heckman & Honoré (1990).

условиях требуется по случайной выборке индивидов с характеристиками X и заработными платами W оценить параметры β_1 и β_2.

Пример 6 (Эффекты воздействия). Задача состоит в оценке воздействия программы субсидирования инвестиций на объем капитальных инвестиций со стороны фирм. Пусть Q_1^* и Q_0^* – объемы инвестиций, осуществляемых фирмой при наличии воздействия (субсидии) и при его отсутствии, соответственно. Q_1^* и Q_0^* – латентные переменные. *Эффект воздействия* (ЭВ) для отдельной фирмы определяется как $TE = Q_1^* - Q_0^*$. Интерес представляет оценка *среднего эффекта воздействия* (СЭВ), который определяется как $ATE = \mathbb{E}[Q_1^* - Q_0^*]$. Также интерес может представлять условный средний эффект воздействия, $ATE(X) = \mathbb{E}[Q_1^* - Q_0^*|X]$, где X – вектор экзогенных характеристик фирмы. Имеется случайная выборка фирм. Каждая фирма наблюдается только один раз, либо при наличии воздействия ($D = 1$), либо при его отсутствии ($D = 0$). То есть при $D = 1$ наблюдается $Q = Q_1^*$, а при $D = 0$ наблюдается $Q = Q_0^*$. Обычно участие в программе субсидирования не является полностью случайным. Исследователь не располагает идеальными экспериментальными данными. Бинарная переменная воздействия D зависит от наблюдаемых характеристик Z и ненаблюдаемой величины u, которая может коррелировать с Q_0^* или/и Q_1^*. Необходимо, используя доступную выборку для (Q, D, X, Z), состоятельно оценить эффект программы предоставления субсидий на объем инвестиций, измеренный безусловным или условным средним эффектом воздействия.

Пример 7 (Модель инвестиций с издержками приспособления). Рассмотрим модель инвестиций в капитал, похожую на модель из примера 3. Теперь инвестиции не являются полностью необратимыми, то есть фирмы могут дезинвестировать или продавать подержанный капитал. Пусть K_t обозначает запас капитала фирмы, который продуктивен в момент времени t. Пусть $\Pi(K_t, K_{t-1})$ – (межвременная) функция прибыли. Прибыль зависит как от K_t, так и от K_{t-1} из-за наличия издержек приспособления. А именно, имеет место асимметрия между ценой нового капитала и ценой подержанного капитала, или, иными словами, между стоимостью капитала при $K_t > K_{t-1}$ и при $K_t < K_{t-1}$.

$$\Pi(K_t, K_{t-1}) = \begin{cases} \Pi^{(+)}(K_t, K_{t-1}) & \text{при} \quad K_t \geq K_{t-1}, \\ \Pi^{(-)}(K_t, K_{t-1}) & \text{при} \quad K_t \leq K_{t-1}. \end{cases} \tag{19}$$

Функции $\Pi^{(+)}$ и $\Pi^{(-)}$ непрерывны, дифференцируемы и строго вогнуты по K_t. Функция прибыли Π всюду непрерывна, но имеет излом (в котором она недифференцируема) в точке $K_t = K_{t-1}$. Определим $K_t^{(+)} \equiv \arg\max_k \Pi^{(+)}(k, K_{t-1})$ и $K_t^{(-)} \equiv \arg\max_k \Pi^{(-)}(k, K_{t-1})$. При указанных условиях легко показать, что $K_t^{(+)} < K_t^{(-)}$, и оптимальный уровень капитала в момент времени t равен

$$K_t = \begin{cases} K_t^{(+)} & \text{при} \quad K_{t-1} < K_t^{(+)}, \\ K_{t-1} & \text{при} \quad K_t^{(+)} \leq K_{t-1} \leq K_t^{(-)}, \\ K_t^{(-)} & \text{при} \quad K_{t-1} > K_t^{(+)}. \end{cases} \tag{20}$$

Модель дополняется спецификацией $K_t^{(+)}$ и $K_t^{(-)}$ в терминах наблюдаемых и ненаблюдаемых величин. Например, $K_t^{(+)} = \alpha^{(+)} + X_t\beta + \varepsilon_t$ и $K_t^{(-)} = \alpha^{(-)} + X_t\beta + \varepsilon_t$, где $\alpha^{(+)}$ и $\alpha^{(-)}$ – параметры, и $\alpha^{(+)} < \alpha^{(-)}$. По случайной выборке для (K_t, K_{t-1}, X_t) требуется оценить параметры $\alpha^{(+)}$, $\alpha^{(-)}$ и β.

2 Оценивание модели усеченной регрессии

2.1 Смещение МНК-оценки

Рассмотрим модель усеченной регрессии, задаваемую выражением $(Y, X) = \{(Y^*, X^*)|Y^* > c\}$, где $Y^* = X^*\beta + \varepsilon$. Поскольку константа c известна, без ограничения общности положим $c = 0$.[5] Предположим, что регрессия Y на X оценивается с помощью МНК. Рисунок 3 графически иллюстрирует смещение МНК-оценки. Истинный наклон линии регрессии равен 1.5, а его МНК-оценка равна 1.15 ($s.e. = 0.05$).[6]

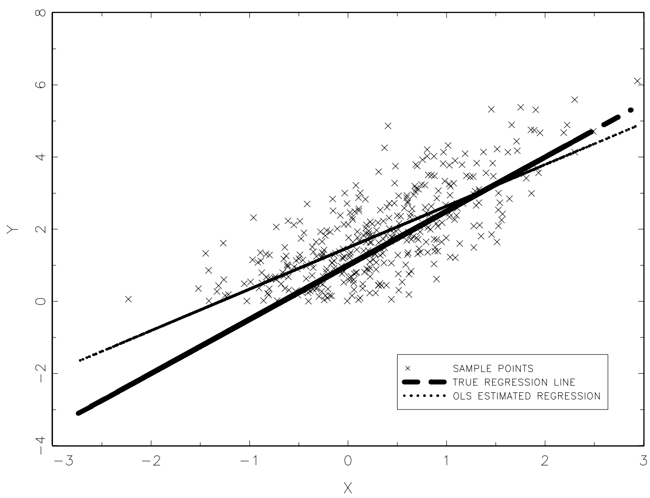

Рис. 3: Смещение МНК-оценки в модели усеченной регрессии.

Формально $Y = \{Y^*|Y^* > 0\} = X^*\beta + \varepsilon^{Trun}$, где $\varepsilon^{Trun} \equiv \{\varepsilon|Y^* > 0\}$. Следовательно,

$$\mathbb{E}[Y|X] = \mathbb{E}[X^*\beta + \varepsilon^{Trun}|X] = X\beta + \mathbb{E}[\varepsilon^{Trun}|X]. \tag{21}$$

Величина $\mathbb{E}[\varepsilon^{Trun}|X]$ отражает эффект *самоотбора выборки* в условном среднем Y при данном X. Заметим, что $\mathbb{E}[\varepsilon^{Trun}|X] = \mathbb{E}[\varepsilon|\varepsilon > -X\beta]$, что, вообще говоря, не равно нулю и зависит от X. Если ε не зависит от X, эффект самоотбора выборки зависит от X только через индекс $X\beta$. Тогда его можно представить в виде функции $s(X\beta)$. Легко показать, что смещение $s(X\beta)$ – убывающая функция от индекса $X\beta$. Чтобы убедиться в этом, заметим, что

$$\begin{cases} \text{при } X\beta \to +\infty & s(X\beta) \to \mathbb{E}[\varepsilon|\varepsilon > -\infty] = \mathbb{E}[\varepsilon] = 0, \\ \text{при } X\beta \to -\infty & s(X\beta) \to \mathbb{E}[\varepsilon|\varepsilon > +\infty] = +\infty. \end{cases} \tag{22}$$

Следовательно, $s(X\beta)$ отрицательно зависит от $X\beta$. В модели регрессии, усеченной справа, смещение вследствие самоотбора также убывает по $X\beta$. Принимая во внимание то, что $\mathbb{E}[Y|X] = X\beta + s(X\beta)$, можно записать следующее уравнение регрессии Y на X:

$$Y = X\beta + s(X\beta) + \tilde{\varepsilon}. \tag{23}$$

Случайная ошибка $\tilde{\varepsilon}$ равна $\varepsilon^{Trun} - s(X\beta)$, и, по построению, независима в среднем от X. Приведенное выражение показывает несостоятельность МНК-оценки, не учитывающей самоотбора выборки. Игнорирование самоотбора выборки ведет к тому, что случайная ошибка в регрессии равна $s(X\beta) + \tilde{\varepsilon}$ и отрицательно коррелирует с $X\beta$.

[5]Если c не равна нулю, всегда можно переопределить Y^* как исходную переменную Y^* за вычетом c.

[6]Процесс, порождающий данные, таков, что X^* и ε – независимые стандартные нормальные случайные величины, $Y^* = 1.0 + 1.5X^* + \varepsilon$, и усечение происходит слева в точке $y = 0$. Размер выборки $n = 500$.

2.2 Оценивание методом максимального правдоподобия

Для получения ММП-оценки параметра β необходимо ввести в модель дополнительную предпосылку о виде распределения ε. Типичным в этом классе моделей является предположение о том, что ε – IID с распределением $N(0,\sigma^2)$. Тогда логарифмическая функция правдоподобия для этой модели и выборки имеет вид

$$\log\mathrm{L}(\beta,\sigma)=\sum_{i=1}^{n}\ln\mathbb{P}\{Y=y_i|X=x_i\},$$

и условные вероятности равны

$$\mathbb{P}\{Y=y_i|X=x_i\}=\mathbb{P}\{Y^*=y_i|X=x_i;Y^*>0\}=$$
$$=\frac{\mathbb{P}\{\varepsilon=y_i-x_i\beta\}}{\mathbb{P}\{\varepsilon>-x_i\beta\}}=\frac{\frac{1}{\sigma}\phi\left(\frac{y_i-x_i\beta}{\sigma}\right)}{\Phi\left(\frac{x_i\beta}{\sigma}\right)}, \tag{24}$$

где $\phi(\cdot)$ и $\Phi(\cdot)$ – ФПР и КФР стандартного нормального распределения. Следовательно, логарифмическую функцию правдоподобия можно записать следующим образом:

$$\log\mathrm{L}(\beta,\sigma)=-n\ln(\sigma)-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-x_i\beta)^2-\sum_{i=1}^{n}\ln\Phi\left(\frac{x_i\beta}{\sigma}\right). \tag{25}$$

Первая часть этого выражения – логарифмическая функция правдоподобия для классической модели линейной регрессии. Вторая часть учитывает усечение. Заметим, что, в отличие от модели бинарного выбора, логарифмическая функция правдоподобия в данном случае зависит не только от β/σ, но и от β и σ по отдельности, так что оба параметра можно идентифицировать.[7]

Логарифмическая функция правдоподобия $\log\mathrm{L}(\beta,\sigma)$ не является глобально вогнутой по (β,σ). Это важный момент. Максимизация глобально вогнутых функций – очень простая задача, то есть можно использовать простые алгоритмы, такие как метод Ньютона или BHHH. Однако максимизация функций, не являющихся всюду вогнутыми, вычислительно более сложна, поскольку требует глобального поиска на всем пространстве параметров, чтобы гарантировать получение глобального, а не локального максимума. Тем не менее, для этой модели легко перепараметризовать логарифмическую функцию правдоподобия и получить глобально вогнутую функцию. Определим параметры $\theta=1/\sigma$ и $\gamma=\beta/\sigma$ и рассмотрим логарифмическую функцию правдоподобия в терминах этих параметров:

$$\log\mathrm{L}(\gamma,\theta)=n\ln(\theta)-\frac{1}{2}\sum_{i=1}^{n}(\theta y_i-x_i\gamma)^2-\sum_{i=1}^{n}\ln\Phi\left(x_i\gamma\right).$$

Функция $\log\mathrm{L}(\gamma,\theta)$ глобально вогнута по (γ,θ). Заметим, что существует взаимно однозначное соответствие между (γ,θ) и (β,σ). Следовательно, по свойству *инвариантности к перепараметризации*, ММП-оценки параметров (β,σ) имеют вид $\hat\sigma_{MLE}=1/\hat\theta$ и $\hat\beta_{MLE}=\hat\gamma/\hat\theta$. Дисперсионную матрицу можно получить, используя дельта-метод.

В контексте линейных регрессионных моделей МНК-оценка состоятельна, если регрессоры не коррелируют со случайной ошибкой. Состоятельность МНК-оценки робастна к гетероскедастичности, серийной корреляции и ненормальности ошибки. Гетероскедастичность является типичной характеристикой большинства кросс-секционных данных. Следовательно,

[7] Заметим также, что в этой модели можно получить остатки $\hat\varepsilon$, которые являются состоятельными оценками ε.

важный вопрос состоит в том, является ли полученная ММП-оценка робастной к гетероскедастичности ε. Остается ли ММП-оценка состоятельной, если ε гетероскедастична, а функция правдоподобия соответствует гомоскедастичной модели? Ответ: нет. Симуляции Монте–Карло показывают, что МНК-оценка может становиться серьезно смещенной. Эта проблема мотивирует изучение других оценок, робастных к гетероскедастичности и ненормальности случайной ошибки.

2.3 Симметрично урезанная МНК-оценка

Джеймс Пауэлл заложил основу полупараметрического оценивания моделей усеченной и цензурированной регрессии. В Powell (1984) предложены оценки наименьших абсолютных отклонений (НАО), которые робастны к гетероскедастичности и ненормальности ошибок. В Powell (1986) рассматривается другая робастная оценка, основанная на симметричном усечении (или цензурировании) хвостов распределения зависимой переменной. Остановимся на симметрично урезанной МНК-оценке (СУМНК-, или STLS-оценке).

Рассмотрим модель усеченной слева регрессии и определим следующую зависимую переменную:[8]

$$\tilde{Y} \equiv \{Y^* | 0 < Y^* < 2X\beta\} = \{Y | Y < 2X\beta\}. \tag{26}$$

Переменная $\tilde{Y}$ усечена слева и справа. Заметим, что точки усечения переменной $\tilde{Y}$ (то есть 0 и $2X\beta$) находятся на одинаковом расстоянии от условного среднего $\mathbb{E}[Y^*|X] = X\beta$. При таком «симметричном усечении» получаем

$$\mathbb{E}[\tilde{Y}|X] = \mathbb{E}[X\beta + \varepsilon | X, 0 < X\beta + \varepsilon < 2X\beta] = X\beta + \mathbb{E}[\varepsilon | X, -X\beta < \varepsilon < X\beta]. \tag{27}$$

В линейной регрессии $\tilde{Y}$ на X выражение $\mathbb{E}[\varepsilon | X, -X\beta < \varepsilon < X\beta]$ представляет собой эффект *самоотбора выборки*. Ясно, что он равен нулю, если функция плотности распределения ε симметрична относительно нуля.

Таким образом, можно получить состоятельную оценку β с помощью МНК-оценивания регрессии $\tilde{Y}$ на X. Эта оценка робастна к гетероскедастичности ε. Более того, предположение о симметричности распределения ε является более общим, чем предположение о нормальности. Однако $\tilde{Y}$ не наблюдается. Чтобы получить случайную выборку для $\tilde{Y}$ нужно усечь наблюдаемую зависимую переменную Y справа в точке $2X\beta$. Но значение β неизвестно. Чтобы справиться с этой проблемой, рассмотрим следующий критерий:

$$Q(\beta) = \sum_{i=1}^{n} \mathbb{I}\{y_i < 2x_i\beta\} \left(y_i - x_i\beta\right)^2 . \tag{28}$$

Эта функция представляет собой симметрично усеченную сумму квадратов остатков. Оценка СУМНК определяется как значение β, минимизирующее этот критерий. Эта оценка состоятельна и асимптотически нормальна. Асимптотическая дисперсионная матрица СУМНК-оценки имеет вид

$$\begin{array}{rcl} \mathbb{V}[\hat{\beta}_{STLS}] & = & C^{-1}DC^{-1}, \text{ где} \\ C & = & \mathbb{E}[\mathbb{I}\{Y < 2X\beta\} XX'], \\ D & = & \mathbb{E}[\mathbb{I}\{X\beta > 0\} \min\{\varepsilon^2; (X\beta)^2\} XX']. \end{array} \tag{29}$$

Заметим, что функция $Q(\beta)$ не является непрерывной и дифференцируемой относительно β во многих различных точках (стольких, сколько имеется точек в выборке). Следовательно, ее минимизация может быть затруднительной. Простой метод нахождения (локального) минимума состоит в следующем. Шаг 1: возьмем начальное значение β, скажем $\hat{\beta}^{(1)}$.

[8] Аналогично для модели усеченной справа регрессии можно определить $\tilde{Y} \equiv \{Y^* | 2X\beta < Y^* < 0\} = \{Y | 2X\beta < Y\}$.

Это может быть, например, МНК-оценка по всей выборке $\{y_i, x_i\}$. Шаг 2: получим усеченную переменную $\tilde{y}_i^{(1)} = \{y_i | y_i < 2x_i\hat{\beta}^{(1)}\}$. То есть удалим все наблюдения с $y_i > 2x_i\hat{\beta}^{(1)}$. Шаг 3: оценим с помощью МНК регрессию $\tilde{y}_i^{(1)}$ на x_i для получения нового значения β, $\hat{\beta}^{(2)}$. Итерируем шаги 2 и 3 до сходимости, то есть до тех пор, пока не выполнится условие $||\hat{\beta}^{(k)} - \hat{\beta}^{(k-1)}|| <$ *малое число*. По достижении сходимости эта процедура дает локальный минимум $Q(\beta)$. Для проверки на глобальность минимума необходимо выполнить глобальный поиск, повторяя эту процедуру для различных начальных значений β.

Данный метод прост и особенно полезен, когда имеется большая выборка и масштаб усечения незначительный. Для относительно малых выборок или при значительном усечении потеря эффективности, связанная с усечением, может быть очень серьезной, а оценки – неточными.

2.4 Тест Хаусмана на гетероскедастичность и ненормальность

Для осуществления теста Хаусмана нужна оценка, которая эффективна при H_0 и несостоятельна при H_1, и оценка, состоятельная как при H_0, так и при H_1. Следовательно, можно использовать ММП-оценку и оценку Пауэлла для построения теста на гетероскедастичность и ненормальность. Нулевая гипотеза состоит в том, что $\varepsilon_i \sim$ IID $N(0, \sigma^2)$, а тест-статистика имеет вид

$$\mathcal{H} = (\hat{\beta}_{STLS} - \hat{\beta}_{MLE})' \left(\mathbb{V}[\hat{\beta}_{STLS}] - \mathbb{V}[\hat{\beta}_{MLE}]\right)^{-1} (\hat{\beta}_{STLS} - \hat{\beta}_{MLE}), \tag{30}$$

и при H_0 имеет распределение хи-квадрат с k степенями свободы.

3 Модель цензурированной регрессии (тобит)

3.1 Смещение МНК-оценки

Рассмотрим модель цензурированной регрессии, такую, что имеется случайная выборка $X = X^*$ и $Y = \max\{Y^*, c\}$, где $Y^* = X\beta + \varepsilon$. Снова без ограничения общности можно положить $c = 0$. Предположим, что оценивается регрессия Y на X. Рисунок 4 графически иллюстрирует смещение МНК-оценки. Истинный наклон линии регрессии равен 1.5, а его МНК-оценка равна 1.10 ($s.e. = 0.04$).[9]

Формально $Y = \max\{X\beta + \varepsilon, 0\}$, или в форме регрессионного уравнения $Y = X\beta + \varepsilon^{Cens}$, где $\varepsilon^{Cens} \equiv \max\{\varepsilon, -X\beta\}$. Следовательно,

$$\mathbb{E}[Y|X] = X\beta + \mathbb{E}[\varepsilon^{Cens}|X] = X\beta + \mathbb{E}[\max\{\varepsilon, -X\beta\}|X]. \tag{31}$$

Величина $\mathbb{E}[\varepsilon^{Cens}|X]$ представляет собой эффект *самоотбора выборки* в условном среднем Y при данном X. Заметим, что $\mathbb{E}[\varepsilon^{Cens}|X] = \mathbb{E}[\max\{\varepsilon, -X\beta\}|X]$, что, вообще говоря, не равно нулю и зависит от X. Если ε не зависит от X, эффект самоотбора выборки зависит от X только через индекс $X\beta$, то есть $\mathbb{E}[\varepsilon^{Cens}|X] = s(X\beta)$, и $s(\cdot)$ – убывающая функция. Тогда, учитывая, что $\mathbb{E}[Y|X] = X\beta + s(X\beta)$, можно записать следующее регрессионное уравнение: $Y = X\beta + s(X\beta) + \tilde{\varepsilon}$, где ошибка $\tilde{\varepsilon} \equiv \varepsilon^{Cens} - s(X\beta)$ независима в среднем от X. МНК-оценка, игнорирующая эффект самоотбора выборки $s(X\beta)$, несостоятельна.

[9]Процесс, порождающий данные, таков, что X^* и ε – независимые стандартные нормальные случайные величины, $Y^* = 1.0 + 1.5X^* + \varepsilon$, а цензурирование слева происходит в точке $y = 0$. Размер выборки $n = 500$.

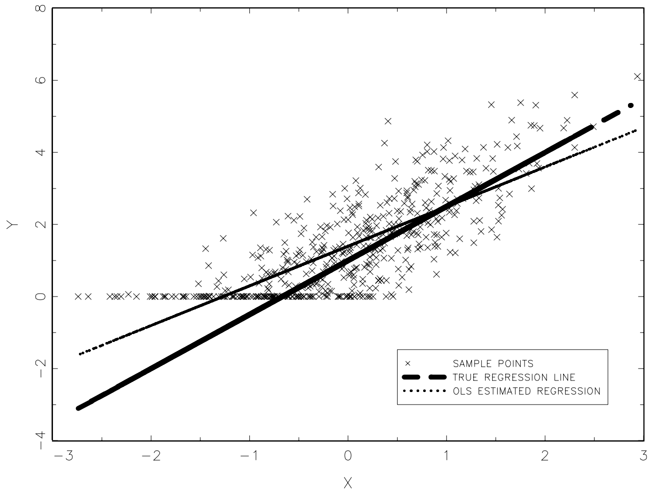

Рис. 4: Смещение МНК-оценки в модели цензурированной регрессии

3.2 Оценивание методом максимального правдоподобия

Логарифмическая функция правдоподобия для этой модели и выборки имеет вид

$$\text{logL}(\beta, \sigma) = \sum_{i=1}^{n} \ln \mathbb{P}\{Y = y_i | X = x_i\},$$

где условные вероятности равны

$$\mathbb{P}\{Y = y_i | X = x_i\} = \begin{cases} \mathbb{P}\{Y^* = y_i | X = x_i\} = f_\varepsilon(y_i - x_i\beta) & \text{при} \quad y_i > 0, \\ \mathbb{P}\{Y^* < 0 | X = x_i\} = F_\varepsilon(-x_i\beta) & \text{при} \quad y_i = 0. \end{cases} \tag{32}$$

При предположении о том, что $\varepsilon_i \sim \text{IID } N(0, \sigma^2)$, логарифмическая функция правдоподобия имеет вид

$$\text{logL}(\beta, \sigma) = -n_1 \ln(\sigma) - \frac{1}{2\sigma^2} \sum_{y_i > 0} (y_i - x_i\beta)^2 + \sum_{y_i = 0} \ln \Phi\left(\frac{-x_i\beta}{\sigma}\right), \tag{33}$$

где n_1 – число наблюдений с $y_i > 0$. Все замечания по поводу модели усеченной регрессии, также справедливы в случае цензурирования.

3.3 Симметрично урезанная МНК-оценка

Рассмотрим модель цензурированной слева регрессии и определим следующую зависимую переменную:

$$\tilde{Y} = \begin{cases} 0 & \text{при} \quad Y^* \le 0, \\ Y^* & \text{при} \quad 0 < Y^* < 2X\beta = \min\{Y; 2X\beta\}, \\ 2X\beta & \text{при} \quad Y^* \ge 2X\beta. \end{cases} \tag{34}$$

Переменная $\tilde{Y}$ цензурирована и слева, и справа. Ясно, что точки цензурирования $\tilde{Y}$ (то есть 0 и $2X\beta$) находятся на одинаковом расстоянии от $X\beta$, условного среднего Y^*. При таком симметричном цензурировании получаем, что $\tilde{Y} = \min\{Y; 2X\beta\} = \min\{\max\{X\beta + \varepsilon; 0\}; 2X\beta\}$.[10] Или, в виде наподобие регрессионного уравнения:

$$\tilde{Y} = X\beta + \mathbb{I}\{\varepsilon < -X\beta\}(-X\beta) + \mathbb{I}\{\varepsilon > X\beta\}(X\beta) + \mathbb{I}\{-X\beta \le \varepsilon \le -X\beta\}\varepsilon. \tag{35}$$

[10] Заметим, что $\max\{X\beta + \varepsilon; 0\} = X\beta + \max\{\varepsilon; -X\beta\}$. Следовательно, $\min\{\max\{X\beta + \varepsilon; 0\}; 2X\beta\} = \min\{X\beta + \max\{\varepsilon; -X\beta\}; 2X\beta\} = X\beta + \min\{\max\{\varepsilon; -X\beta\}; X\beta\}$. Или, что эквивалентно, $X\beta + \mathbb{I}\{\varepsilon < -X\beta\}(-X\beta) + \mathbb{I}\{\varepsilon > X\beta\}(X\beta) + \mathbb{I}\{-X\beta \le \varepsilon \le -X\beta\}\varepsilon$.

В линейной регрессии $\tilde{Y}$ на X, эффект самоотбора – это условное матожидание случайного члена $\mathbb{I}\{\varepsilon < -X\beta\}(-X\beta) + \mathbb{I}\{\varepsilon > X\beta\}(X\beta) + \mathbb{I}\{-X\beta \leq \varepsilon \leq -X\beta\}\varepsilon$. Как и в случае усечения, этот эффект равен нулю, если функция плотности распределения ε симметрична относительно нуля.

СУМНК-оценка модели цензурированной регрессии определяется как значение β, минимизирующее следующий критерий:

$$Q(\beta) = \sum_{i=1}^{n} \left(\min\{y_i; 2x_i\beta\} - x_i\beta\right)^2. \tag{36}$$

Эта функция представляет собой сумму квадратов остатков линейной регрессии $\tilde{Y}$ на X. Получаемая оценка состоятельна, асимптотически нормальна и робастна к ненормальности и гетероскедастичности ε.

4 Модели с самоотбором выборки

Рассмотрим модель с самоотбором выборки, в которой $Y = (1-D)Y_0^* + DY_1^*$, где

$$\begin{aligned} Y_0^* &= X\beta_0 + \varepsilon_0, \\ Y_1^* &= X\beta_1 + \varepsilon_1, \end{aligned} \tag{37}$$

и

$$D = \mathbb{I}\{Z\gamma - u > 0\}. \tag{38}$$

Ненаблюдаемые величины ε_0, ε_1 и u не являются независимо распределенными. Например, предположим, что D – индикатор события «индивид является членом профсоюза», Y_1^* – заработная плата индивида, если он входит в профсоюз, а Y_0^* – если не входит. Задача состоит в оценке параметров β_0 и β_1. Иногда особый интерес представляет средний эффект воздействия $ATE(X) = X(\beta_1 - \beta_0)$, то есть средняя отдача от членства в профсоюзе для индивида с характеристиками X.

4.1 Смещение МНК-оценки

Можно построить следующие два вида МНК-оценок векторов β_0 и β_1: (а) *совместную МНК-оценку*, когда с помощью МНК оценивается регрессия Y на X и DX, то есть $Y = X\beta_0 + DX(\beta_1 - \beta_0) + e$; (б) *отдельные МНК-оценки*, когда по отдельности оцениваются регрессия $Y = X\beta_0 + e_0$ для подвыборки наблюдений с $D = 0$ и регрессия $Y = X\beta_1 + e_1$ для подвыборки наблюдений с $D = 1$. Ясно, что при отсутствии ограничений на параметры β_0 и β_1 между уравнениями эти оценки совпадают, а следовательно, можно рассматривать лишь один способ оценивания, скажем (б).

По построению, случайная ошибка $e_j \equiv \{\varepsilon_j | D = j\}$. Следовательно,

$$\begin{aligned} \mathbb{E}[e_0|X] &= \mathbb{E}[\varepsilon_0|X, D = 0] = \mathbb{E}[\varepsilon_0|X, u \geq Z\gamma], \\ \mathbb{E}[e_1|X] &= \mathbb{E}[\varepsilon_1|X, D = 1] = \mathbb{E}[\varepsilon_1|X, u < Z\gamma]. \end{aligned} \tag{39}$$

Если ε и u не являются независимыми, и если только X и Z не являются независимыми (что крайне нереалистично при использовании неэкспериментальных данных), эти эффекты самоотбора коррелируют с X. Значит, ошибки e_0 и e_1 коррелируют с X, и соответствующие МНК-оценки для β_0 и β_1 несостоятельны.

Дадим интерпретацию этому смещению в контексте примера об отдаче от участия в профсоюзе. МНК-оценка $\beta_1 - \beta_0$ в регрессии $Y = X\beta_0 + DX(\beta_1 - \beta_0) + e$ представляет собой комбинацию двух эффектов: (1) реальной отдачи от участия в профсоюзе, $\beta_1 - \beta_0$ и (2) того факта, что работники, которые решают вступить в профсоюз, обычно также являются теми, для кого велик «эффект воздействия», или разница заработных плат $Y_1^* - Y_0^*$. Первый элемент – это тот причинно-следственный эффект, который требуется оценить. Второй элемент – это «ложный» эффект, не являющийся следствием участия в профсоюзе. Для иллюстрации предположим, что X – это просто константа. Предположим также, что участие в профсоюзе имеет два эффекта: оно увеличивает константу, то есть $\beta_1 > \beta_0$, и сокращает дисперсию заработных плат, то есть $\varepsilon_1 = \lambda\varepsilon_0$, где $\lambda < 1$. Допустим, что единственный фактор, влияющий на решение об участии в профсоюзе, – это разница заработных плат (модель Роя), так что $Z\gamma - u = Y_1^* - Y_0^* = (\beta_1 - \beta_0) + (\varepsilon_1 - \varepsilon_0) = (\beta_1 - \beta_0) - (1-\lambda)\varepsilon_0$. В этом примере ясно, что

$$\begin{aligned} \text{plim}\hat{\beta}_0^{OLS} &= \mathbb{E}[Y|D=0] = \beta_0 + \mathbb{E}\left[\varepsilon_0 \,\middle|\, \varepsilon_0 > \frac{\beta_1 - \beta_0}{1-\lambda}\right] > \beta_0, \\ \text{plim}\hat{\beta}_1^{OLS} &= \mathbb{E}[Y|D=1] = \beta_1 + \lambda\mathbb{E}\left[\varepsilon_0 \,\middle|\, \varepsilon_0 < \frac{\beta_1 - \beta_0}{1-\lambda}\right] < \beta_1. \end{aligned} \tag{40}$$

Следовательно, в данном примере оценка $\hat{\beta}_0^{OLS}$ переоценивает β_0, поскольку не участвующие в профсоюзах работники имеют большие значения ε_0 (то есть большую производительность), $\varepsilon_0 > (\beta_1 - \beta_0)/(1-\lambda)$. Также оценка $\hat{\beta}_1^{OLS}$ недооценивает β_1, поскольку участвующие в профсоюзах работники имеют более низкие значения ε_0, то есть $\varepsilon_0 < (\beta_1 - \beta_0)/(1-\lambda)$. В результате при изложенных предпосылках МНК-оценка $\beta_1 - \beta_0$ недооценивает истинную отдачу от участия в профсоюзе.

4.2 Оценивание методом максимального правдоподобия

Зависимые переменные модели – это Y и D, а экзогенные объясняющие переменные – X и Z. Логарифмическая функция правдоподобия для этой модели и выборки имеет вид

$$\text{logL}(\beta, \gamma, \Omega) = \sum_{i=1}^{n} \ln \mathbb{P}\{Y = y_i, D = d_i | X = x_i, Z = z_i\} \tag{41}$$

с вероятностями

$$\begin{aligned} \mathbb{P}\{Y = y_i, D = 0 | X = x_i, Z = z_i\} &= \mathbb{P}\{\varepsilon_0 = y_i - x_i\beta_0; u_i > z_i\gamma\} \\ &= \int_{z_i\gamma}^{+\infty} f_{\varepsilon_0,u}(y_i - x_i\beta_0, u)du \end{aligned} \tag{42}$$

и

$$\begin{aligned} \mathbb{P}\{Y = y_i, D = 1 | X = x_i, Z = z_i\} &= \mathbb{P}\{\varepsilon_1 = y_i - x_i\beta_1; u_i < z_i\gamma\} \\ &= \int_{-\infty}^{z_i\gamma} f_{\varepsilon_1,u}(y_i - x_i\beta_1, u)du, \end{aligned} \tag{43}$$

где $f_{\varepsilon_0,u}$ и $f_{\varepsilon_1,u}$ – функции плотности совместных распределений (ε_0, u) и (ε_1, u), соответственно.

При ММП-оценивании данной модели обычно предполагается, что тройка $(\varepsilon_0, \varepsilon_1, u)$ имеет совместное нормальное распределение. Дисперсия u нормируется к 1. Параметры, которые входят в данную функцию правдоподобия, – это β_0, β_1, γ, стандартные отклонения σ_0 и σ_1 и ковариации σ_{0u} и σ_{1u}. В общем, эта функция правдоподобия не является глобально вогнутой и может иметь несколько локальных максимумов. Более того, в отличие от моделей усеченной и цензурированной регрессии, не существует перепараметризации, при которой функция правдоподобия всюду вогнута. Следовательно, при реализации оптимизационного алгоритма необходимо выбирать различные начальные значения параметров и сравнивать получающиеся по достижении сходимости значения функции правдоподобия в надежде получить глобальный максимум.

4.3 Двухшаговый метод Хекмана

Хекман (Heckman, 1976, 1979) предложил альтернативный двухшаговый подход, который дает состоятельные оценки в модели с самоотбором выборки и очень легок в применении. Вычислительная простота этого двухшагового метода делает его очень привлекательным на практике. Однако есть по крайней мере одна другая причина, по которой двухшаговый метод Хекмана так популярен в практических приложениях. Как и в случаях моделей усеченной и цензурированной регрессии, ММП-оценка в общей модели с самоотбором выборки не является робастной к гетероскедастичности и ненормальности. Хотя двухшаговый подход Хекмана был предложен в контексте параметрической модели с нормальными и гомоскедастичными ошибками, одно из наиболее привлекательных свойств получаемой оценки состоит в том, что ее можно расширить на полупараметрический случай с ненормальными и гетероскедастичными ошибками.

Рассмотрим сначала эту оценку в контексте полностью параметрической модели с нормальными и гомоскедастичными ошибками. Заметим, что

$$\begin{aligned}\mathbb{E}[Y|X,Z,D=0] &= X\beta_0 + \mathbb{E}[\varepsilon_0|X,Z,D=0]\\ &= X\beta_0 + \frac{1}{1-F_u(Z\gamma)}\int_{Z\gamma}^{+\infty}\mathbb{E}[\varepsilon_0|u]f_u(u)du\end{aligned} \tag{44}$$

и

$$\begin{aligned}\mathbb{E}[Y|X,Z,D=1] &= X\beta_1 + \mathbb{E}[\varepsilon_1|X,Z,D=1]\\ &= X\beta_1 + \frac{1}{F_u(Z\gamma)}\int_{-\infty}^{Z\gamma}\mathbb{E}[\varepsilon_1|u]f_u(u)du.\end{aligned} \tag{45}$$

При нормальности $(u,\varepsilon_0,\varepsilon_1)$ эти выражения принимают вид

$$\begin{aligned}\mathbb{E}[Y|X,Z,D=0] &= X\beta_0 + \sigma_{0u}\lambda(-Z\gamma),\\ \mathbb{E}[Y|X,Z,D=1] &= X\beta_1 - \sigma_{1u}\lambda(Z\gamma),\end{aligned} \tag{46}$$

где функцию

$$\lambda(c) \equiv \frac{\phi(c)}{\Phi(c)}$$

называют обратным отношением Миллса, или лямбдой Хекмана.

Основываясь на этом результате, Хекман предложил следующую двухшаговую процедуру. Шаг 1: оценим γ методом максимального правдоподобия в пробит-модели $D = \mathbb{I}\{Z\gamma - u > 0\}$. Получим $\{z_i\hat{\gamma}\}$ для каждого наблюдения в выборке и найдем оценки лямбд Хекмана, $\hat{\lambda}_{0i} = \phi(-z_i\hat{\gamma})/\Phi(-z_i\hat{\gamma})$ и $\hat{\lambda}_{1i} = \phi(z_i\hat{\gamma})/\Phi(z_i\hat{\gamma})$. Шаг 2: с помощью МНК оценим регрессию Y на X и $\hat{\lambda}_0$, используя подвыборку наблюдений с $D=0$, и регрессию Y на X и $\hat{\lambda}_1$, используя подвыборку наблюдений с $D=1$. Эта процедура дает состоятельные оценки β_0, β_1, σ_{0u} и σ_{1u}. Amemiya (1985, с. 370–371) приводит выражение для корректировки стандартных ошибок оценок параметров с учетом ошибки оценивания в переменных $\hat{\lambda}_0$ и $\hat{\lambda}_1$.

Каким образом данная процедура учитывает смещение вследствие самоотбора? Это достигается путем включения в регрессию (оценки) эффекта самоотбора $\hat{\lambda}$. Как по отдельности идентифицировать причинно-следственный эффект X на Y (через $X\beta_j$) и смещение вследствие самоотбора $\hat{\lambda}_j$? Или, другими словами, почему $\hat{\lambda}$ и X неколлинеарны? Есть две возможные причины. Во-первых, в Z могут быть переменные, которых нет в X (то есть выполняется условие невключения). В этом случае, если эти переменные имеют достаточную объясняющую силу в пробит-модели, $\hat{\lambda}_j$ имеет источник вариации, который *не зависит* от X. Во-вторых, $\hat{\lambda}$ – нелинейная функция от $Z\hat{\gamma}$. Даже если $Z \subseteq X$, переменная $\hat{\lambda}$ имеет выборочную вариацию, которая *линейно независима* от X. Первый источник идентификации называют *идентификацией через инструментальные переменные*, и он не зависит от предположений о функциональной форме, то есть идентификация возможна, даже если модель специфицирует непараметрическую взаимосвязь между Y_j^* и X. Второй источник идентификации называют *идентификацией через условие невключения* и для него критичны параметрические предположения, то есть линейность взаимосвязи между Y_j^* и X и нормальность ошибок.

Предшествующее обсуждение указывает на дополнительную причину, по которой желательно ослабить предположение о нормальности. Даже если интерес представляют линейные эффекты X на Y_j^*, лучше, если идентификация этих эффектов не полагается исключительно на предположение о линейности и параметрическое предположение о распределении ненаблюдаемых величин. Опишем расширение двухшаговой процедуры Хекмана, которое допускает произвольное распределение ненаблюдаемых величин. Рассмотрим модель с самоотбором выборки, в которой ненаблюдаемые величины $(\varepsilon_0, \varepsilon_1, u)$ независимы от (X, Z) и имеют произвольное распределение с носителем в евклидовом пространстве и заданной мерой Лебега. На самом деле, можно разрешить гетероскедастичность $(\varepsilon_0, \varepsilon_1, u)$, если дисперсии и ковариации этих величин зависят от (X, Z) только через индекс $Z\gamma$. Без дополнительных предположений из модели следует, что

$$\begin{aligned}
&\mathbb{E}[Y|X, Z, D = 0] = X\beta_0 + s_0(Z\gamma), \\
&\mathbb{E}[Y|X, Z, D = 1] = X\beta_1 + s_1(Z\gamma),
\end{aligned} \tag{47}$$

где теперь функциональная форма эффектов самоотбора $s_0(\cdot)$ и $s_1(\cdot)$ неизвестна. Тем не менее, известно, что они являются одноиндексными функциями: они зависят только от $Z\gamma$. При наличии оценки γ из модели бинарного выбора[11] можно с произвольной точностью аппроксимировать элементы $s_j(Z\hat{\gamma})$, используя полином порядка q от $Z\hat{\gamma}$. То есть на втором шаге можно с помощью МНК оценить следующие регрессии:

$$\begin{aligned}
&\{Y|D = 0\} = X\beta_0 + \sum_{j=1}^{q} \rho_{0j}(Z\hat{\gamma})^j + e_0, \\
&\{Y|D = 1\} = X\beta_1 + \sum_{j=1}^{q} \rho_{1j}(Z\hat{\gamma})^j + e_1.
\end{aligned} \tag{48}$$

Некоторые исследователи также предлагают использовать полином от оцененной лямбды Хекмана или от оценки вероятности из модели дискретного выбора (то есть оценки вероятности воздействия). Можно также использовать другие виды полупараметрических оценок для частично линейных моделей (см. Robinson, 1983, и Yatchew, 2003). Ясно, что идентификация β_0 и β_1 основана только на условии невключения. Из этого также следует, что для идентификации параметров при данном подходе индекс $Z\hat{\gamma}$ должен обладать достаточной выборочной вариацией, независимой от X. Кроме того, необходимо обосновать выполнение условия невключения, исходя из экономических соображений и понимания задачи.

Литература

Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press: Cambridge, Massachusetts.

Heckman, J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such model. *Annals of Economic and Social Measurement* 15, 475–492.

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

[11] Параметрическая спецификация модели дискретного выбора в данном случае неважна. В пробит-модели можно также использовать полином от z_i.

Heckman, J., & B. Honoré (1990). The empirical content of the Roy model. *Econometrica* 58, 1121–1149.

Powell, J. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* 25, 303–325.

Powell, J. (1986). Symmetrically trimmed least squares estimation for Tobit models. *Econometrica* 54, 1435–1460.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers (New Series)* 3, 135–146.

Robinson, P. (1988). Root-N-consistent semiparametric regression. *Econometrica* 56, 931–954.

Yatchew, A. (2003). Semiparametric regression for the applied econometrician. Глава в сборнике *Themes in Modern Econometrics* под редакцией P.C.B. Phillips. Cambridge University Press.

Some notes on sample selection models

Victor Aguirregabiria

University of Toronto, Toronto, Canada

Sample selection problems are pervasive when working with micro economic models and datasets of individuals, households or firms. During the last three decades, there have been very significant developments in this area of econometrics. Different type of models have been proposed and used in empirical applications. And new estimation and inference methods, both parametric and semiparametric, have been developed. These notes provide a brief introduction to this large literature.