Статьи: эконометрическая теория

Идентификация в классе непараметрических моделей систем одновременных уравнений с выборочной селективностью*

Евгений Ожегов†

Высшая школа экономики, Пермь, Россия

В данной статье рассматривается проблема идентификации непараметрической модели систем одновременных уравнений при наличии выборочной селективности. Для представленной модели сформулированы условия идентификации при наличии исключённых переменных в уравнении участия и уравнениях системы. Исключённые переменные позволяют корректировать ошибки уравнений системы на выборочную селективность и одновременность. Подход представляет собой расширение известной непараметрической процедуры идентификации регрессионных функций на случай нетреугольных систем одновременных уравнений.
Ключевые слова: непараметрическое оценивание, выборочная селективность, системы одновременных уравнений, коррекция ошибок
Классификация JEL: C14, C30, C51

1 Введение

В данной статье рассматривается непараметрическая модель системы одновременных уравнений при условии выборочной селективности. Для идентификации модели сформулированы требования на наличие исключённых переменных для уравнения участия и уравнений системы. Класс идентифицируемых моделей ограничен системой уравнений с аддитивно сепарабельными ошибками, распределёнными совместно с непрерывной функцией плотности, и непрерывно дифференцируемыми регрессионными функциями.

Важным приложением рассматриваемой модели, которому не уделяется должное внимание в литературе, является оценивание функции индивидуального спроса на дифференцированные товары с эндогенными характеристиками. Так при моделировании спроса на товары, характеристики которых выбираются одновременно с объемом потребления, мы должны учитывать структурную взаимосвязь между спросом и характеристиками товара. Примером такого товара является, в частности, кредит, эндогенность характеристик которого подчеркивается в работах Attanasio, Goldberg & Kyriazidou (2008) и LaCour-Little (2007). Для идентификации и состоятельного оценивания функции спроса на кредиты также необходимо учитывать тот факт, что выборка выданных кредитов является сформированной неслучайно как под воздействием самоотбора заёмщиков, так и в силу неслучайного одобрения банком заёмщиков с определёнными характеристиками.

Мы предлагаем модель индивидуального спроса и механизм её идентификации, учитывающие одновременность выбора объема потребления и характеристик товара, выборочную

*Данное научное исследование (№ 14-01-0104) выполнено при поддержке Программы «Научный фонд НИУ ВШЭ» в 2014/2015 гг. Цитировать как: Ожегов, Евгений (2015). «Идентификация в классе непараметрических моделей систем одновременных уравнений с выборочной селективностью», Квантиль, №13, стр. 15–23. Citation: Ozhegov, Evgeniy (2015). “Identification in a class of nonparametric simultaneous equation models with sample selection”, Quantile, No.13, pp. 15–23.

†Адрес: 614000, Пермь, ул. Лебедева, дом 27. Электронная почта: *tos600@gmail.com*

селективность, а также потенциальную разнородность предпочтений, что полностью обусловлено исследовательской необходимостью. Нужно отметить, что не только спрос на основе систем одновременных уравнений не изучен в литературе, но также эконометрика оценивания непараметрических моделей одновременных уравнений при наличии выборочной селективности также является до сих пор не разработанной. Так, наиболее прогрессивными работами по части идентификации и оценивания непараметрических моделей одновременных уравнений являются работы Розы Матцкин, которая сформулировала условия идентификации модели при наличии исключённых инструментов в Matzkin (2015), а также расширила их на случай цензурированных зависимых переменных в Matzkin (2012), оставляя за кадром случай наличия выборочной селективности. Другой близкой работой, изучавшей оценивание систем одновременных уравнений при наличии выборочной селективности, является работа Das, Newey & Vella (2003). В ней авторы рассматривают одно требующее идентификации уравнение с эндогенными регрессорами и предлагают полупараметрические процедуры оценивания такой модели на основе аппроксимации сплайнами и степенными сериями. Мы обобщим предложенный ими подход на случай нетреугольной системы одновременных уравнений с выборочной селективностью, расширив ранговое условие идентификации модели.

Данная статья имеет следующую структуру. В разделе 2 будет представлена модель системы одновременных уравнений с выборочной селективностью. В разделе 3 будут сформулированы и доказаны необходимые условия идентификации модели. Последний раздел подводит итоги работы.

2 Модель

Традиционные модели оценивания индивидуального спроса на кредитные продукты используют параметрические подходы к оцениванию выбора условий кредитного контракта, таких, например, как сумма займа или соотношение суммы займа к стоимости залога (loan-to-value ratio, LTV). Оценивание таких моделей обычно сопровождается двумя главными вызовами: выборочной селективностью и эндогенностью всех условий договора. Проблема выборочной селективности возникает, потому что некоторые решения принимаются последовательно и часть объясняющих выбор переменных наблюдается частично на разных стадиях процесса кредитования. Так, если процесс одобрения заёмщиков коррелирован с выбором величины ссуды или другими условиями займа, например, через ненаблюдаемый кредитный риск, то сила данной корреляции будет определять величину смещения оценок в уравнении выбора величины ссуды, что доказывается, например в Ross (2000). Теоретическая взаимозависимость выбора на разных этапах процесса кредитования обсуждается также в более ранних работах (см., например, Follain, 1990 и Rachlis & Yezer, 1993).

Ряд работ по оцениванию спроса на кредиты также обсуждает эндогенность всех параметров кредитного договора при использовании их в качестве объясняющих спрос переменных. В работе Ambrose, LaCour-Little & Sanders (2004) подчеркивается эндогенность суммы займа и отношения суммы займа и LTV. Attanazio, Goldberg & Kyriazidou (2008) обсудили и учли эндогенность процентной ставки и срока погашения кредита при моделировании спроса на автокредиты и применили две непараметрические процедуры оценивания спроса при условии неслучайного отбора заёмщиков. Они также нашли эмпирическое подтверждение необходимости использования более гибкой регрессионной функции спроса, а также непостоянной эластичности спроса по процентной ставке и сроке погашения кредита на разных квантилях распределения дохода заёмщиков.

Тем не менее, не только сумма займа может зависеть от других условий кредита. Так, выбор суммы займа или LTV может влиять на предпочтения других условий кредита, что обуславливает необходимость моделирования одновременного выбора всех параметров кредита, а также тот факт, что предпочтения условий кредита могут быть разнородны. При этом интерес представляет не только корректная идентификация выбора объема потребления, но и идентификация всех потенциальных структурных взаимосвязей между выбором объема потребления и характеристик товара.

Подводя итог, процесс кредитования, или, в общем, модель индивидуального спроса на дифференцированный товар с эндогенными характеристиками, может быть представлена следующей эконометрической моделью:

$$d_i = \begin{cases} 1, \; g_0(x_i, z_{0i}) + e_{0i} \geq 0, \\ 0, \; g_0(x_i, z_{0i}) + e_{0i} < 0, \end{cases}$$

$$\begin{cases} y^*_{1i} = g_1(y^*_{-1i}, x_i, z_{1i}) + e_{1i}, \\ \dots \\ y^*_{ki} = g_k(y^*_{-ki}, x_i, z_{ki}) + e_{ki}, \end{cases} \tag{1}$$

$$y_i = \begin{cases} y^*_i, & \text{если } d_i = 1, \\ \text{ненаблюдаемо}, & \text{если } d_i = 0, \end{cases}$$

где d_i — индикатор подписания контракта (наблюдения ненулевого потребления), x_i набор демографических характеристик заёмщиков (потребителей), $y_i = (y_{1i}, \dots, y_{ki}) = (y_{ji}, y_{-ji})$ — набор параметров кредитного договора (характеристик товара), включающий логарифм суммы займа, LTV, процентную ставку и логарифм срока погашения кредита, z_{0i} — набор исключённых переменных для уравнения подписания контракта, а $z_i = (z_{1i}, \dots, z_{ki}) = (z_{ji}, z_{-ji})$ — набор исключённых инструментов для условий кредитного контракта.

Модель (1) содержит систему одновременных уравнений на этапе моделирования выбора условий займа. Более того, обычно выборка индивидуальных данных для оценивания спроса содержит репрезентативную часть домохозяйств (в случае опросных данных), либо все домохозяйства, обратившиеся в банк за кредитом (в случае банковских данных). В этом случае демографические переменные обычно наблюдаются полностью, а параметры кредитного договора только для части домохозяйств, взявших кредит, процесс получения которого, вообще говоря, не случаен. Таким образом, необходимо также учитывать потенциальную выборочную селективность. Для каждого заёмщика i эконометрист наблюдает переменные $(x_i, z_{0i}, z_{1i}, \dots, z_{ki})$, а переменные $(y_{1i}, \dots, y_{ki})$ только в случае $d_i = 1$. Случайные переменные $(e_{0i}, e_{1i} \dots, e_{ki})$ не наблюдаемы эконометристом, а условия, накладываемые на них, описаны далее. Задачей эконометриста является оценивание функций $(g_1, \dots, g_k)$ непараметрическим образом.

Первое уравнение в (1) является уравнением участия, определяющим, будет ли заключён контракт. Вторая часть (1) описывает систему одновременного выбора характеристик контракта. Важным является то, что каждое структурное уравнение системы (1) имеет переменную z_{ji}, исключённую из всех остальных уравнений. На этом факте будет базироваться идентификация.

3 Идентификация модели

Проблема смещения ввиду выборочной селективности впервые обсуждалась в Gronau (1973) и Heckman (1974). Хекман также предложил процедуру оценивания такой модели методом максимального правдоподобия либо используя двухшаговую процедуру в Heckman (1976) и Heckman (1979), при которой ошибка уравнения выбора (уравнения для y) корректируется на потенциальную ковариацию с ошибкой уравнения участия (уравнения для d). Оба подхода, однако, основывались на предположении о том, что совместное распределение ошибок в уравнении участия и уравнении выбора является двумерным нормальным. Следующие работы ослабили эту предпосылку при использовании двухшаговой процедуры с помощью полупараметрических подходов, например, раскладывая функционально неизвестную функцию коррекции ошибки в ряд Фурье (см. Heckman & Robb, 1985), или аппроксимируя её серией степенных функций (см. Newey, 1988). С точки зрения экономической теории эндогенность параметров контракта y_i возникает естественным образом и очень важна, при этом она часто представима в виде структурной системы одновременных уравнений, как в системе (1). Дополнительные предположения о функциональной форме структурных функций в системе (1) зачастую слишком ограничительны и могут вести к неробастным спецификациям. Потому непараметрический подход, используемый в этой статье, предпочтителен. В работе Newey & Powell (2003) впервые была предложена непараметрическая процедуру оценивания треугольной системы одновременных уравнений с неизвестными регрессионными функциями, а в Newey, Powell & Vella (1999) далее была разработана двухшаговая процедура оценивания, использующая коррекцию ошибки на её ковариацию с эндогенными регрессорами и аппроксимирующая функцию коррекции степенной серией, зависящей от ошибок уравнения в приведенной форме. В Newey (2013) также приведен обзор всех известных непараметрических методов инструментальных переменных, а также обсуждается проблема слабых инструментов без обсуждения конструктивных подходов к тестированию силы инструментов, что может быть направлением дальнейших исследований.

В Das, Newey & Vella (2003) предлагается метод оценивания треугольной системы одновременных уравнений при наличии выборочной селективности. Они также предложили аппроксимировать функцию коррекции ошибки серией степенных функций или сплайнов от значения предрасположенности из уравнения участия и остатков уравнений в приведенной форме.

Мы расширим известное семейство двухшаговых процедур оценивания на случай нетреугольной непараметрической системы одновременных уравнений при наличии выборочной селективности и произвольного совместного распределения ошибок всех уравнений.

Введем некоторые предположения, необходимые для идентификации.

Предположение 1. Набор переменных $W = (x, z, z_0)$ статистически независим от набора переменных $(e_0, e_1, \ldots, e_k)$.

Предположение 2. Функция плотности $f_{e_0,e}(e_0, e_1, \ldots, e_k)$ непрерывна на $\mathbb{R}^{k+1}$.

При идентификации каждого отдельного структурного уравнения $y_{ji} = g_j(y_{-ji}, x_i, z_{ji}) + e_{ji}$ следует учитывать, что изменение эндогенных переменных левой части y_{-ji} будет приводить к изменению y_{ji} не только через функцию g_j, но также и через ковариацию между e_{-ji} и e_{ji}, а изменение экзогенных переменных x_i будет приводить к изменению y_{ji} также через ковариацию между e_{0i} и e_{ji}. В таком случае, идентификацию необходимо построить так, чтобы получить вариацию y_{ji} за счет вариации переменных (y_{-ji}, x_i, z_{ji}) при известном ожидаемом изменении e_{ji}.

Если $p = \mathbb{E}[d|x_0, w_0]$ — это значение предрасположенности участия, то ожидаемое значение e_{ji}, условное на (y_{-ji}, x_i, z_{ji}) и $d_i = 1$, можно получить как $\mathbb{E}[e_j|y_{-j}, x, z, z_0, d = 1] = \mathbb{E}[e_j|e_{-j}, g_0(x_0, z_0) + e_0 \geq 0] = \int_{-\infty}^{\infty}\int_{-g_0(x_0,z_0)}^{\infty} e_j f_{e_0,e}(s, r|e_{-j}) ds dr = \phi_j(p, e_{-j})$, где ϕ_j будет известной функцией при известном виде совместного распределения ошибок $f_{e_0,e}$. При произвольном распределении $f_{e_0,e}$ функция ϕ_j будет иметь неизвестный вид, но будет функцией известного набора аргументов, (p, e_{-j}). Таким образом, идентификация будет построена вокруг предварительной идентификации предрасположенности участия p и ошибок e_{-j}, а также функции ϕ_j, что далее позволит идентифицировать регрессионные функции g_j.

Процедура идентификации модели (1) будет состоять из следующих шагов.

На первом шаге необходимо идентифицировать значение предрасположенности $p = \mathbb{E}[d|x_0, z_0]$ из уравнения участия:

$$d_i = \begin{cases} 1, \ g_0(x_i, z_{0i}) + e_{0i} \geq 0, \\ 0, \ g_0(x_i, z_{0i}) + e_{0i} < 0. \end{cases} \quad (2)$$

Для любого маржинального распределения f_{e_0}, $\mathbb{E}[d|x, z_0] = \int_{-g_0(x,z_0)}^{\infty} f_{e_0}(s)ds = \gamma_0(x, z_0)$, если g_0 является непрерывной, что и будем предполагать далее.

Предположение 3. Функция $g_0(x, z_0)$ является непрерывно дифференцируемой с непрерывной функцией распределения почти всюду.

Функция γ_0 при произвольном распределении e_0 и функциональной форме g_0 будет функцией с произвольной функциональной формой, но будет зависеть от известного набора переменных, (x, z_0). Условное ожидание $p = \mathbb{E}[d|x_0, z_0]$ идентифицируемо, т.к. (d, x, z_0) наблюдаемы, а (x, z_0) независимы от e_0 по предположению 1. Таким образом, изменение (x, z_0) будет влиять на d только через функцию γ_0, что позволяет восстановить $p = \mathbb{E}[d|x, z_0]$.

На следующем шаге необходимо идентифицировать ошибки e_{-j} за счет изменения переменных, не влияющих на y_j. Для этого необходимо идентифицировать уравнения для каждой y_j в приведенной форме, скорректированные на выборочную селективность, получив из них ошибки e_j:

$$e_j = y_j - \mathbb{E}[y_j|x, z, z_0, d = 1]. \quad (3)$$

Если e_j имеет совместное маржинальное распределение с e_0 с функцией плотности f_{e_0,e_j}, то

$$\begin{aligned} \mathbb{E}[e_j|x, z, z_0, d = 1] &= \mathbb{E}[e_j|g_0(x, z_0) + e_0 \geq 0] \\ &= \int_{-\infty}^{\infty} \int_{-g_0(x,z_0)}^{\infty} e_j f_{e_0,e_j}(s, r)dsdr = \lambda_j(p). \end{aligned} \quad (4)$$

Тогда y_j можно разложить на регрессионную функцию в приведенной форме и функцию коррекции ошибки λ_j в силу независимости (x, z, z_0) и e_j:

$$\mathbb{E}[y_j|x, z, z_0, d = 1] = \gamma_j(x, z) + \lambda_j(p). \quad (5)$$

Если p — это значение предрасположенности, полученное на предыдущем шаге оценивания, то зафиксировав значения (x, z), мы можем идентифицировать λ_j за счет изменения z_0, являющейся аргументом p. При этом необходимо, чтобы z_0 действительно имела существенное влияние на p.

Предположение 4. Почти наверное $\partial g_0(x, z_0)/\partial z_0 \neq 0$.

Далее же, зная λ_j, мы можем идентифицировать γ_j, меняя значения (x, z), зафиксировав z_0. При этом идентификация по данному принципу будет возможна, если восстанавливаемые функции λ_j и γ_j будут непрерывно дифференцируемы, что и будем предполагать.

Предположение 5. Функции $\gamma_j(x, z)$ и $\lambda_j(p)$ являются непрерывно дифференцируемыми с непрерывной функцией распределения почти всюду.

Наконец, следует идентифицировать каждое уравнение структурной формы, корректируя ошибки на выборочную селективность и одновременность выбора y, используя значение предрасположенности и ошибки приведенной формы. Если e_j имеет совместное распределение с e_0 и e_{-j} с функцией плотности $f_{e_0,e}$, то

$$\begin{aligned}\mathbb{E}[e_j|y_{-j},x,z,z_0,d=1] &= \mathbb{E}[e_j|e_{-j},g_0(x_0,z_0)+e_0\geq 0]\\ &= \int_{-\infty}^{\infty}\int_{-g_0(x_0,z_0)}^{\infty} e_j f_{e_0,e}(s,r|e_{-j})dsdr = \phi_j(p,e_{-j}). \qquad (6)\end{aligned}$$

Тогда y_j можно разложить на регрессионную функцию в структурной форме и функцию коррекции ошибки как

$$\mathbb{E}[y_j|y_{-j},x,z,z_0,d=1] = g_j(y_{-j},x,z_j)+\phi_j(p,e_{-j}). \qquad (7)$$

Если p и e_{-j} — это значение предрасположенности и ошибки уравнений в приведенной форме, идентифицированные на предыдущих шагах, то зафиксировав значения (y_{-j},x,z_j), мы можем идентифицировать ϕ_j за счет изменения z_0, являющейся аргументом p, а также за счет изменения z_{-j}, влияющих на e_{-j}, если ϕ_j являются непрерывно дифференцируемой. Дополнительным необходимым условием идентификации теперь также должно являться то, что z имеют существенное влияние на y и, соответственно, на e в уравнениях приведенной формы.

Предположение 6. Почти наверное $\mathrm{rank}(\partial\gamma(x,z)/\partial z)=\dim(y)$.

Далее же, зная ϕ_j, мы можем идентифицировать g_j, меняя значения (y_{-j},x,z_j), зафиксировав (z,z_0), при условии, что g_j являются непрерывно дифференцируемыми.

Предположение 7. Функции $g_j(y_{ij},x,z_j)$ и $\phi_j(p,e{-}j)$ являются непрерывно дифференцируемыми с непрерывной функцией распределения почти всюду.

Формализуем все введенные условия для идентификации регрессионных функций g_j.

Теорема 1. *Если верны предположения 1–7, то каждая регрессионная функция g_j идентифицируема вплоть до аддитивной константы.*

Доказательство. См. Приложение. □

Условия введенной теоремы являются вполне естественными и неограничительными.

Первая группа условий (предположения 3, 5, 7) требует от регрессионных функций и функций коррекции ошибок быть непрерывно дифференцируемыми. Данное условие при этом может быть ослаблено в части экзогенных регрессоров x, позволяя им быть дискретными и входить в регрессионную функцию аддитивно сепарабельно (более подробно см. в Das, Newey & Vella, 2003).

Вторая группа условий (предположения 4, 6) накладывает требования на наличие и релевантность исключённых переменных. Так, требуется хотя бы одна значимая исключённая переменная z_0 в уравнении участия, а также хотя бы один значимый исключённый инструмент z_j для каждой эндогенной переменной y_j, при этом эффекты исключённых инструментов на эндогенные переменные в уравнениях приведенной формы не должны быть коллинеарны.

Также требуется, чтобы все включённые и исключённые переменные были ортогональны ошибкам (предположение 1), а распределение ошибок было непрерывным (предположение 2).

4 Заключение

В данной статье рассмотрены проблемы идентификации непараметрических моделей систем одновременных уравнений при наличии выборочной селективности. Так при игнорировании одновременности и выборочной селективности идентификация каждой отдельной структурной регрессионной функции будет невозможна, т.к. невозможно будет разделить вариацию зависимой переменной на вариацию регрессионной функции и ковариацию ошибок.

Данное исследование представляет процедуру идентификации, позволяющую корректировать ошибки уравнений системы на выборочную селективность и одновременность выбора при произвольном совместном распределении ошибок и виде регрессионных функций. Процедура представляет собой расширение непараметрической процедуры, предложенной в работе Das, Newey & Vella (2003) на случай нетреугольных систем одновременных уравнений.

Для представленной модели сформулированы условия идентификации. Так, модель будет идентифицируема при наличии хотя бы одного исключённого регрессора для уравнения выбора, хотя бы одного исключённого релевантного инструмента для каждой зависимой переменной системы одновременных уравнений. Регрессионные функции и функции коррекции ошибок должны также быть непрерывно дифференцируемыми.

Данная статья естественно развивает сферу применения двухшаговых процедур коррекции ошибки на модели систем одновременных уравнений нетреугольной формы, предлагая простую последовательную процедуру идентификации. Однако, существенным ограничением при её использовании пока является аддитивность ошибок, ограничивающая сферу применения модели непрерывными зависимыми переменными.

Список литературы

Ambrose, B., M. LaCour-Little & A. Sanders (2004). The effect of conforming loan status on mortgage yield spreads: A loan level analysis. *Real Estate Economics* 32, 541–69.

Attanazio, O., P. Goldberg & E. Kyriazidou (2008). Credit constraints in market for consumer durables: Evidence from micro data on car loans. *International Economic Review* 49, 401–36.

Das, M., W. Newey & F. Vella (2003). Nonparametric estimation of sample selection models. *Review of Economic Studies* 70, 33–58.

Follain, J.R. (1990). Mortgage choice. *Real Estate Economics* 18, 125–44.

Gronau, R. (1973). Wage comparisons: A selectivity bias. NBER Working Paper #13.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica* 42, 679–94.

Heckman, J. (1976). The common structure of statistical models of truncation, sample selection, and limited dependent variables and a sample estimator for such models. *Annals of Economic and Social Measurement* 5, 475–92.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica* 47, 153–61.

Heckman, J. & R. Robb (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics* 30, 239–67.

LaCour-Little, M. (2007). The home purchase mortgage preferences of low- and moderate-income households. *Real Estate Economics* 35, 265–90.

Matzkin, R. (2012). Identification in nonparametric limited dependent variable models with simultaneity and unobserved heterogeneity. *Journal of Econometrics* 166, 106–115.

Matzkin, R. (2015). Estimation of nonparametric models with simultaneity. *Econometrica* 83, 1–66.

Newey, W. (2013). Nonparametric instrumental variables estimation. *American Economic Review* 103, 550–556.

Newey, W. & J. Powell (2003). Instrumental variables estimation of nonparametric models. *American Economic Review* 103, 550–556.

Newey, W., J. Powell & F. Vella (2003). Nonparametric estimation of triangular simultaneous equations models. *Econometrica* 67, 565–603.

Philips, R. & A. Yezer (1996). Self-selection and tests for bias and risk in mortgage lending: Can you price the mortgage if you don't know the process? *Journal of Real Estate Research* 11, 87–102.

Rachlis, M. & A. Yezer (1996). Serious flaws in statistical tests for discrimination in mortgage markets. *Journal of Housing Research* 4, 315–336.

Ross, S.L. (2000). Mortgage lending, sample selection and default. *Real Estate Economics,* 8, 581–621.

Приложение

Лемма 1. *Если выполнены предположения 1–5, то каждая γ_j и λ_j идентифицируема вплоть до аддитивной константы.*

Доказательство аналогично T.2.1 в Das, Newey & Vella (2003): Любая наблюдаемо эквивалентная модель для уравнения (3) должна иметь $\mathbb{E}[y_j|x,z,z_0,d=1]=\hat{\gamma}_j(x,z)+\hat{\lambda}_j(p)$. Рассмотрим $f_1(x,z)+f_2(p)=0$, где $f_1(x,z)=\gamma_j(x,z)-\hat{\gamma}_j(x,z)$, и $f_2(p)=\lambda_j(p)-\hat{\lambda}_j(p)$. Если g_0, γ_j и λ_j непрерывно дифференцируемы, то f_1 и f_2 тоже непрерывно дифференцируемы. Продифференцируем $f_1+f_2=0$ по переменным (z_0,x,z):

$$\begin{aligned}
0 &= \frac{\partial f_2(p)}{\partial p}\frac{\partial g_0(x,z_0)}{\partial z_0},\\
0 &= \frac{\partial f_1(x,z)}{\partial x}+\frac{\partial f_2(p)}{\partial p}\frac{\partial g_0(x,z_0)}{\partial x},\\
0 &= \frac{\partial f_1(x,z)}{\partial z}.
\end{aligned}\tag{8}$$

Первое условие и $\frac{\partial g_0(x,z_0)}{\partial z_0}\neq 0$ дает $\frac{\partial f_2(p)}{\partial p}=0$, что означает, что f_2 — константа, а $\hat{\lambda}_j(p)=\lambda_j(p)+C_j$. Тогда второе условие даст $\frac{\partial f_1(x,z)}{\partial x}=0$. Это значит, что $f_1(x,z)$ — тоже константа, а $\hat{\gamma}_j(x,z)=\gamma_j(x,z)+C_j$. $\square$

Доказательство теоремы 1. По лемме 1 функции $\gamma_j(x,z)$, $\lambda_j(p)$ идентифицируемы, а значит и e_{-j} идентифицируемы. Докажем идентификацию регрессионных функций g_j. Любая наблюдаемо эквивалентная модель для уравнения (7) должна иметь $\mathbb{E}[y_j|y_{-j},x,z,z_0,d=1]=\hat{g}_j(y_{-j},x,z_j)+\hat{\phi}_j(p,e_{-j})$. Рассмотрим $f_3(y_{-j},x,z_j)+f_4(p,e_{-j})=0$, где $f_3(y_{-j},x,z_j)=g_j(y_{-j},x,z_j)-\hat{g}_j(y_{-j},x,z_j)$, а $f_4(p,e_{-j})=\phi_j(p,e_{-j})-\hat{\phi}_j(p,e_{-j})$.

Если g_0, γ_j, g_j, λ_j и ϕ_j являются непрерывно дифференцируемыми, то и $f_3(y_{-j},x,z_j)$ и $f_4(p,e_{-j})$ тоже непрерывно дифференцируемы. Продифференцируем тождество $f_3+f_4=0$ по (z_j,z_{-j},x,z_0):

$$\begin{aligned}
0 &= \frac{\partial f_3(y_{-j},x,z_j)}{\partial z_j}+\frac{\partial f_3(y_{-j},x,z_j)}{\partial y_{-j}}\frac{\partial\gamma_{-j}(x,z)}{\partial z_j},\\
0 &= \frac{\partial f_3(y_{-j},x,z_j)}{\partial y_{-j}}\frac{\partial\gamma_{-j}(x,z_j)}{\partial z_{-j}},\\
0 &= \frac{\partial f_3(y_{-j},x,z_j)}{\partial y_{-j}}\frac{\partial\gamma_{-j}(x,z_j)}{\partial x}+\frac{\partial f_3(y_{-j},x,z_j)}{\partial x}+\frac{\partial f_4(p,e_{-j})}{\partial p}\frac{\partial g_0(x,z_0)}{\partial x},
\end{aligned}\tag{9}$$

$$0 = \frac{\partial f_4(p, e_{-j})}{\partial p} \frac{\partial g_0(x, z_0)}{\partial z_0}.$$

Последнее условие и $\frac{\partial g_0(x,z_0)}{\partial z_0} \neq 0$ дают $\frac{\partial f_4(p,e_{-j})}{\partial p} = 0$. Выполнение рангового условия $\mathrm{rank}[\frac{\partial \gamma(x,z)}{\partial z}] = \dim(y)$ обеспечивает, что для каждого j найдется z_j с $\frac{\partial \gamma_j(x,z)}{\partial z_j} \neq 0$ и $\frac{\partial \gamma_{-j}(x,z)}{\partial z_{-j}} \neq 0$ соответственно, что делает второе условие эквивалентным $\frac{\partial f_3(y_{-j},x,z_j)}{\partial y_{-j}} = 0$. Заменяя $\frac{\partial f_3(y_{-j},x,z_j)}{\partial y_{-j}} = 0$ в третьем условии и используя $\frac{\partial f_4(p,e_{-j})}{\partial p} = 0$, получим, что $\frac{\partial f_3(y_{-j},x,z_j)}{\partial x} = 0$. Наконец, $\frac{\partial f_3(y_{-j},x,z_j)}{\partial y_{-j}} = 0$ первом условии дает $\frac{\partial f_3(y_{-j},x,z_j)}{\partial z_j} = 0$.

Все полученные результаты говорят о том, что $f_3(y_{-j}, x, z_j) = g_j(y_{-j}, x, z_j) - \hat{g}_j(y_{-j}, x, z_j)$ является константой, а значит $\hat{g}_j(y_{-j}, x, z_j) = g_j(y_{-j}, x, z_j) + C_j$. $\square$

Identification in a class of nonparametric simultaneous equation models with sample selection

Evgeniy Ozhegov

Higher School of Economics, Perm, Russia

We consider identification of the nonparametric simultaneous equation model with the presence of sample selection. For the proposed model we introduce necessary conditions for its identification if excluded variables for selection and outcome equations are available. Our approach extends the well known class of nonparametric two-step identification procedures for the case of non-triangular simultaneous equations.

Keywords: nonparametrics, sample selection, simultaneous equations, control function

JEL Classification: C13, C30, C51