

Эконометрический ликбез: инструментальные переменные

Инструментальные переменные и эндогенность: нетехнический обзор*

Питер Эббес[†]

Университет штата Пенсильвания, Университи Парк, США

Настоящее эссе представляет собой нетехнический обзор наиболее свежих результатов, появившихся в эконометрической литературе по инструментальному оцениванию линейной регрессионной модели. Стандартные методы инференции, такие как МНК, дают смещенные и несостоятельные оценки, если регрессоры и ошибки коррелированы. Для преодоления этой проблемы были разработаны методы, использующие инструментальные переменные, однако поиск хороших инструментов всегда затруднителен, и зачастую исследователи-практики имеют дело со слабыми инструментами. В эссе дается обзор последних исследований, связанных со слабыми инструментами, а также рассматриваются некоторые методы, предложенные для работы с такими инструментами, включая «экономные» методы инструментальных переменных, которые не полагаются на наблюдаемые инструменты для идентификации регрессионных параметров при зависимости регрессоров и ошибок.

1 Введение

В (прикладной) статистике стандартная модель линейной регрессии $y = X\beta + \epsilon$ является важным инструментом моделирования влияния набора объясняющих переменных на зависимую переменную. Пусть $y = (y_1, \dots, y_n)'$ – $n \times 1$ вектор наблюдений для зависимой переменной, $X \in \mathbb{R}^{n \times k}$ – $n \times k$ матрица объясняющих переменных (регрессоров), β – неизвестный $k \times 1$ вектор регрессионных параметров, и $\epsilon = (\epsilon_1, \dots, \epsilon_n)'$ – ненаблюдаемый случайный шок. Для идентифицируемости параметров предполагается, что $\text{rank } X = k < n$. Стандартная модель линейной регрессии часто применяется в случае кросс-секционных данных. Во многих ситуациях данные имеют иерархическую структуру, например, данные о сотрудниках фирм или продажах товаров в различных магазинах, или панельные данные о потребителях. Для таких типов данных применяются многоуровневые модели, методы анализа панельных данных, или иерархические линейные модели, которые являются обобщением стандартной модели линейной регрессии (Snijders & Bosker, 1999, Wooldridge, 2002).

Важным предположением в этих моделях является некоррелированность объясняющих переменных (X) и случайной ошибки (ϵ). В этом случае регрессоры называют экзогенными; предполагается, что они определяются вне модели. Невыполнение этого предположения может привести к смещенным и несостоятельным оценкам при применении стандартных методов инференции (например, МНК или ОМНК), а следовательно, к ошибочным выводам и принятию неправильных решений. К сожалению, во многих ситуациях предпосылка о некоррелированности регрессоров и ошибок нарушается. Тогда регрессоры называют

*Перевод Б. Гершмана и С. Анатольева. Цитировать как: Эббес, Питер (2007) «Инструментальные переменные и эндогенность: нетехнический обзор», Квантиль, №2, стр. 3–20. Citation: Ebbes, Peter (2007) “A non-technical guide to instrumental variables and regressor-error dependencies,” *Quantile*, No.2, pp. 3–20.

[†]Адрес: Smeal College of Business, Penn State University, University Park, PA 16802, USA. Электронная почта: pebbes@psu.edu

эндогенными. Зависимость регрессоров и ошибок может возникать по разным причинам: (1) пропущенные существенные переменные, (2) ошибки измерения регрессоров, (3) самоотбор, (4) одновременность, (5) серийная корреляция ошибок при наличии лагированной зависимой переменной среди регрессоров. Ruud (2000) показывает, что проблемы (2)–(5) можно рассматривать как частные случаи проблемы (1). Похожее замечание сделал Wooldridge (2002), заметивший, что различия между возможными источниками эндогенности не всегда ясны, и эти источники могут как усиливать, так и компенсировать друг друга в зависимости от конкретного практического контекста.

В настоящем эссе дается нетехническое изложение проблемы эндогенности в линейных регрессионных моделях. Кратко рассматриваются методы инструментальных переменных (ИП), имеющие долгую историю в эконометрике (Bowden & Turkington, 1984) и разработанные для преодоления этой проблемы. Инструментальные переменные – это переменные, которые коррелируют с «проблемными» регрессорами, но не коррелируют с ошибкой. Когда инструменты доступны, для оценки регрессионных параметров могут применяться двухшаговый метод наименьших квадратов (2ШМНК) или метод максимального правдоподобия с ограниченной информацией. В статье обсуждаются некоторые важные вопросы реализации методов инструментальных переменных, касающиеся, в частности, качества используемых инструментов. Методы инструментального оценивания могут потенциально давать результаты хуже, чем просто МНК при игнорировании зависимости между регрессорами и ошибками, если инструменты плохого качества, то есть являются слабыми или/и эндогенными. Эссе охватывает большинство новейших исследований по этим вопросам и рассматривает ряд методов, предложенных для работы со слабыми инструментами, а также недавно предложенные «экономные» методы инструментального оценивания, которые не основаны на наблюдаемых инструментах для учета зависимости регрессоров и ошибок. Последние используют характеристики распределения эндогенных регрессоров для построения «латентных» инструментов (Ebbes, Wedel & Böckenholt, 2006).

Структура эссе такова. В разделе 2 обсуждается ряд практических приложений линейной регрессионной модели, в которых МНК-оценка является смещенной и несостоятельной при наличии зависимости между регрессорами и ошибкой. Эти приложения мотивируют применение техники инструментальных переменных, которая обсуждается в разделе 3. В разделе 4 кратко представлены альтернативные методы, которые не основаны на использовании наблюдаемых инструментальных переменных, а используют информацию иного рода для оценки регрессионных параметров. В разделе 5 подводятся итоги.

2 Когда МНК-оценку не спасти

Важным предположением регрессионных моделей является некоррелированность регрессоров и ошибок, то есть условие $\mathbb{E}[X\epsilon] = 0$. Рассмотрим далее пять важных приложений, когда это предположение нарушено: пропущенные переменные, ошибки измерения, самоотбор, одновременность, серийная корреляция ошибок вместе с наличием лагированных значений зависимой переменной среди регрессоров.

Пропущенные существенные объясняющие переменные

Проблема пропущенных переменных широко исследуется при оценивании влияния образования на заработную плату (Card, 1999, 2001), когда «способности» являются пропущенной переменной. Индивиды с лучшими способностями, с одной стороны, имеют больший успех на рынке труда и зарабатывают больше, а с другой, более склонны к получению образования. Способности как таковые влияют как на уровень образования, так и на доходы, а регрессор «уровень образования» и регрессионная ошибка не являются независимыми. Аналогично, в

области маркетинга исследователи часто сталкиваются с проблемой пропущенных переменных. Wansbeek & Wedel (1999) утверждают, что предположение об экзогенности регрессоров, включая цену, является недостатком стандартных моделей реакции рынка. Shugan (2004) отмечает растущее внимание рецензентов к проблеме эндогенности в большинстве журналов по маркетингу. Отсутствие экзогенности регрессоров из-за пропущенных ключевых аспектов маркетинговых моделей вызывает все больший интерес при проведении исследований в области маркетинга. Когда менеджеры магазинов устанавливают показатели маркетингового комплекса (например, цену товара или объем рекламы), они основываются на информации о локальном рынке или характеристиках товаров, неизвестных исследователю, например, на знаниях об уровне конкуренции, слухах, изменениях во вкусах, долях различных игроков на локальном рынке или наличии скидок. Эта ненаблюдаемая информация может повлиять на поведение потребителей, что приводит к корреляции ошибок с регрессорами, обычно с ценой, в типичной маркетинговой модели (Villas-Boas & Winer, 1999, Chintagunta, 2001).

Модель с пропущенными переменными выглядит следующим образом (Judge, Griffiths, Hill, Lütkepohl & Lee, 1985)

$$\mathbb{E}[y_i|x_i, w_i] = x_i'\beta + w_i'\gamma, \quad (1)$$

где w_i 's – латентные или ненаблюдаемые переменные. Когда ожидание берется условно только на наблюдаемой переменной x_i без учета w_i , имеем

$$\mathbb{E}[y_i|x_i] = x_i'\beta + \mathbb{E}[w_i'|x_i]\gamma, \quad (2)$$

что не равняется $x_i'\beta$, если: (i) $\mathbb{E}[w_i'|x_i] \neq 0$ (то есть пропущенные и включенные регрессоры не ортогональны) и (ii) $\gamma \neq 0$ (то есть пропущенные регрессоры оказывают влияние на y_i). Смещение МНК-оценки для параметра β в этом случае равно $\mathbb{E}[\hat{\beta}_n^{\text{OLS}} - \beta] = \Pi\gamma$, где величина и знак смещения зависят от $\Pi = (X'X)^{-1}X'W$ и γ . Как нетрудно увидеть, пропуск существенных объясняющих переменных затрагивает все оцениваемые коэффициенты вектора β .

Ошибки измерения

Зависимость регрессоров и регрессионных ошибок также возникает, если переменные, участвующие в модели, измеряются с ошибкой. Ошибки измерения могут возникать, например, когда метод или инструмент измерения дают ошибку, величина не имеет физической единицы измерения (например, IQ, способности, или восприятие), или данные из различных источников неправильно агрегируются и компонуются.

Griliches (1977), например, рассматривает проблему ошибок измерения при оценивании влияния образования на доходы. В данном случае нужна адекватная мера образования, которая отражала бы качества работников, за которые работодатели готовы платить. Обычной практикой является использование переменной «количество законченных лет обучения» в качестве меры для «общего уровня образования». Помимо ошибочного сообщения или записи величины «количество лет обучения», сомнительно, что она полностью отражает уровень образования, так как индивиды могут, например, самообразовываться на вечерних курсах или на работе. Кроме того, поскольку большинство исследований по экономике труда основываются на выборочных обследованиях домашних хозяйств, все переменные содержат некоторые ошибки. Даже если эти ошибки малы, их влияние может возрасти при добавлении дополнительных переменных в целях уменьшить смещение из-за пропущенной переменной «способности» (Card, 1999, 2001). Аналогично, Nevo (2000) и Sudhir (2001) утверждают, что мера цены, обычно используемая при оценки (логит) моделей совокупного спроса для измерения степени конкуренции, может измеряться с ошибкой. Переменная, которая используется в большинстве случаев, – это «прейскурантная цена» или агрегированная мера цены, то

есть предполагается, что все потребители сталкиваются с одной и той же ценой (и другими маркетинговыми и товарными характеристиками). Однако это предположение нарушается, если потребители совершают покупки в разных магазинах, регионах или в разные недели, и переменная «цена» может измеряться с ошибкой. В качестве альтернативы было бы лучше использовать цену сделки как меру цены (Sudhir, 2001). Vagozzi, Yi & Nassen (1999) исследуют ошибки измерения в данных, используемых в маркетинговых исследованиях. Например, пункты анкет или шкалы рейтингов, используемые для измерения восприятия, вер, отношения, суждений, или другие теоретические конструкции наверняка ведут к ошибкам измерения, поскольку нет физических единиц для точного замера соответствующих величин. Кроме того, данные для исследований в области маркетинга могут содержать ошибки метода измерения, такие как эффект ореола, эффекты интервьюирующего, или искажения из-за эффекта социальной желательности. Их результаты говорят о том, что ошибки измерения в маркетинговых данных могут быть большими и требуют учета в эмпирических приложениях для улучшения качества принимаемых решений и статистических выводов.

Рассмотрим следующую регрессионную модель с ошибками измерения:

$$y_i = \beta_0 + \beta_1 \chi_i + \epsilon_i.$$

Здесь χ_i – это истинный ненаблюдаемый компонент модели. Вместо него наблюдается x_i , $x_i = \chi_i + \nu_i$, причем $\mathbb{E}[\epsilon_i] = \mathbb{E}[\nu_i] = 0$, $\mathbb{E}[\epsilon_i^2] = \sigma_\epsilon^2 > 0$, $\mathbb{E}[\nu_i^2] = \sigma_\nu^2 > 0$, и $\mathbb{E}[\epsilon_i \nu_i] = \mathbb{E}[\chi_i \epsilon_i] = \mathbb{E}[\chi_i \nu_i] = 0$. Эти два уравнения можно объединить в $y_i = \beta_0 + \beta_1 x_i + u_i$, где $u_i = \epsilon_i - \beta_1 \nu_i$. МНК-оценка параметра β_1 смещена в сторону нуля, поскольку $\mathbb{E}[u_i x_i] = -\beta_1 \sigma_\nu^2 \neq 0$, так что $\mathbb{E}[u_i | x_i] \neq 0$ (см., например, Wansbeek & Meijer, 2000).

Самоотбор

Проблема самоотбора возникает, когда индивиды выбирают себе определенное состояние, например, быть или нет членом профсоюза (Vella & Verbeek, 1998), лечиться или нет (Angrist, Imbens & Rubin, 1996), на основании экономических или других, обычно неизвестных, причин. Например, Angrist (1990) рассматривает влияние статуса ветерана войны во Вьетнаме на доход граждан, чтобы понять, следует ли правительству США давать им компенсацию за возможную потерю личного дохода, вызванную службой в армии. Однако доходы непросто сравнить учитывая лишь статус ветерана, потому что индивиды с меньшими возможностями «на гражданке» скорее поступят на военную службу, и такие индивиды зарабатывали бы меньше независимо от службы в армии.

Hamilton & Nickerson (2003) дают обзор эндогенного принятия решений в стратегическом менеджменте, когда менеджеры осуществляют организационный выбор из нескольких конкурирующих стратегий не случайно, а на основании ожиданий и опыта. Аналогично данные, собранные в интернете, могут страдать от проблемы самоотбора. Определенного рода индивиды чаще бывают в сети и, следовательно, чаще принимают участие в онлайн-опросах, заходят на вебсайты или делают покупки в интернет-магазинах. Если эти ненаблюдаемые индивидуальные характеристики влияют на поведение в сети, предпочтения или восприятие, то часть влияния этих скрытых характеристик неправильно приписывается использованию интернета. Можно предположить, что эти индивиды вели бы себя иначе независимо от частоты посещения интернета. Эти явления важны, например, при исследовании решений о количестве приобретаемого товара в интернет-магазинах по сравнению с обычными («оффлайн-овыми») магазинами или о покупке товаров определенного брэнда в зависимости от категории товара и характеристик магазина.

Проиллюстрируем простую модель с самоотбором:

$$\begin{aligned} y_i &= x_i'(\beta + \delta) + \epsilon_i && \text{если } i \in \text{I}, \\ &= x_i' \beta + \epsilon_i && \text{если } i \in \text{II}, \end{aligned}$$

где I и II обозначают определенные состояния (например, интернет-пользователь или нет). Более компактная запись:

$$y_i = x_i' \beta + d_i x_i' \delta + \epsilon_i,$$

где $d_i = 1$, если $i \in I$, и $d_i = 0$ в противном случае. Из этой записи видно, что d_i является фиктивной переменной, и стандартное оценивание не проходит, если $\mathbb{E}[\epsilon_i | d_i] \neq 0$. Это предположение нарушается в приведенных выше примерах. Более детально о проблеме самоотбора можно узнать, например, из Vella (1998).

Системы одновременных уравнений

Обычный (или иерархичный) регрессионный анализ не подходит в случае, когда переменные в правой части модели определяются одновременно с зависимыми переменными. Однако часто бывает трудно избавиться от подобного взаимного влияния. Примером может быть экономический агент, принимающий решения относительно образования или участия на рынке труда (Card, 1999, 2001) или установление цен фирмами в условиях конкуренции. Некоторые исследования рассматривают одновременность цены и величины спроса на рынках с дифференцированным продуктом при данной структуре конкуренции. Ценовая политика фирм, обусловленная, например, ненаблюдаемыми характеристиками товара, такими как наличие скидок и общенациональной рекламы, расположение точек продаж и другими параметрами розничной торговли, или же реакция со стороны конкурентов, ведет к эндогенности. Работа Berry (1994) по борьбе с эндогенностью цен в агрегированных моделях с использованием инструментальных переменных широко используется и адаптируется. Например, Nevo (2001) оценивает структурную модель спроса и предложения для отрасли готовых к употреблению зерновых завтраков; Berry, Levinsohn & Pakes (1995) и Sudhir (2001) разрабатывают модель рыночного равновесия с конкурентным ценообразованием на рынке автомобилей для исследования ценообразования на автомобилях и уровня конкуренции. Свежий обзор структурного моделирования в маркетинге содержится в Chintagunta, Erdem, Rossi & Wedel (2006).

Простая модель спроса и предложения для продукта или товара выглядит следующим образом:

$$\begin{aligned} y_t^d &= (x_t^d)' \beta^d + \gamma^d p_t + \epsilon_t^d, \\ y_t^s &= (x_t^s)' \beta^s + \gamma^s p_t + \epsilon_t^s, \end{aligned}$$

где компоненты вектора x_t^d – это факторы, влияющие на спрос или поведение потребителей, а компоненты x_t^s влияют только на поведения производителей. Цена p_t определяется из равенства $y_t^d = y_t^s$. Когда оценивается уравнение спроса $y_t^d = (x_t^d)' \beta + \gamma^d p_t + \epsilon_t^d$, нельзя предполагать, что $\mathbb{E}[\epsilon_t^d | p_t] = 0$, так как цена и величина спроса определяются одновременно, то есть ненаблюдаемые положительные шоки спроса или действия конкурентов сдвигают кривую спроса вверх, что (при прочих равных) означает более высокую равновесную цену. В этом случае МНК нельзя использовать для получения оценок параметров уравнения спроса.

Лагированные значения зависимой переменной среди регрессоров

Присутствие лагированных значений зависимой переменной среди регрессоров нарушает предположение об экзогенности, если имеет место серийная корреляция ошибок. Хорошо известно, что в этом случае МНК нельзя применять (см., например, White, 2001). Рассмотрим следующую модель:

$$\begin{aligned} y_t &= x_t' \beta_1 + y_{t-1} \beta_2 + \epsilon_t, \\ \epsilon_t &= \phi \epsilon_{t-1} + v_t, \end{aligned} \tag{3}$$

где y_t – это, например, объем продаж в момент t , x_t – рекламная деятельность в момент t (детерминированная переменная для простоты), и переменная y_{t-1} включена для отслеживания отложенного эффекта прошлой рекламной компании. Предположим, что $|\phi| < 1$, $|\beta_2| < 1$, v_t – независимые одинаково распределенные случайные величины, $\mathbb{E}[v_t] = 0$, и v_t не зависят от y_t и x_t , а также пусть существуют все моменты второго порядка. Преобразуем $\epsilon_t y_{t-1} = \phi \epsilon_{t-1} y_{t-1} + v_t y_{t-1}$, так что $\mathbb{E}[\epsilon_t y_{t-1}] = \phi \mathbb{E}[\epsilon_{t-1} y_{t-1}]$. Далее, $\mathbb{E}[\epsilon_t y_t] = x_t' \beta_1 \mathbb{E}[\epsilon_t] + \beta_2 \mathbb{E}[y_{t-1} \epsilon_t] + \mathbb{V}[\epsilon_t]$. Пользуясь стационарностью ϵ_t , получаем

$$\mathbb{E}[y_{t-1} \epsilon_t] = \frac{\phi}{1 - \phi \beta_2} \mathbb{V}[\epsilon_t],$$

и $\mathbb{E}[\epsilon_t | y_{t-1}] \neq 0$, если только не выполнено $\phi = 0$. Davidson & MacKinnon (1993) делают более сильное утверждение, указывая на то, что МНК-оценка смещена во всех моделях с лагированными зависимыми переменными среди регрессоров (однако состоятельна при $\phi = 0$). В определенных случаях объясняющие переменные могут играть роль лагированных зависимых переменных, что легко упустить. Подобная ситуация описана в Gönül, Kim & Shi (2000), которые исследуют влияние продажи каталогов на вероятность покупки товаров из этих каталогов. Переменная почтовой рассылки и другие рекламные воздействия на клиентов зачастую являются функциями от прошлых продаж, что неявно создает проблемы наподобие описанных выше.

Из приведенных примеров видно, что зависимость между регрессорами и ошибками возникает в ряде стандартных приложений. Напрямую следует, что МНК-оценка

$$\hat{\beta}_n^{\text{OLS}} = \beta + (X'X)^{-1} X' \epsilon$$

смещена, если $\mathbb{E}[\epsilon|X] \neq 0$, и теряет свою привлекательность. Более того, обычная оценка дисперсии ошибок в этом случае смещена, и истинная дисперсия недооценивается (Greene, 2000). Смещение МНК-оценки не снижается при увеличении выборки, она несостоятельна, так как $\text{plim} \hat{\beta}_n^{\text{OLS}} \neq \beta$ и $\text{plim} \hat{\sigma}_{n,\text{OLS}}^2 < \sigma^2$. Эти проблемы можно уменьшить, по крайней мере в больших выборках, при использовании инструментальных переменных (Bowden & Turkington, 1984, White, 2001). Обсудим далее метод инструментальных переменных.

3 Метод инструментальных переменных

Метод инструментальных переменных предполагает наличие набора переменных Z , называемых инструментами. Инструменты должны быть некоррелированными с ошибкой ϵ , т.е. $\mathbb{E}[\epsilon|Z] = 0$, и объяснять часть вариации эндогенных регрессоров. Следовательно, инструменты Z не должны иметь прямого влияния на y , т.е. быть экзогенными. Стандартная модель регрессии с инструментальными переменными получается добавлением к стандартной линейной регрессионной модели уравнения, связывающего эндогенные регрессоры и инструменты, а именно:

$$\begin{aligned} y &= X\beta + \epsilon, \\ X &= Z\Pi + V, \end{aligned} \tag{4}$$

где y , X , и β определяются так же, как и раньше, Z – матрица инструментальных переменных размера $n \times q$, а V – матрица ошибок размера $n \times k$. Матрица Π отражает эффект влияния инструментов на эндогенные регрессоры. Экзогенные переменные среди X также должны содержаться в наборе Z (Wooldridge, 2002). Для идентифицируемости предполагается, что $q \geq k$ и $\text{rank } Z = q < n$. Корреляция между X и ϵ (эндогенность) возникает из-за ненулевой корреляции между ϵ и V . Эта модель с инструментальными переменными является

частным случаем модели с одновременными уравнениями, широко известной в эконометрике. Наиболее популярными методами оценки β являются двухшаговый МНК (2ШМНК) и метод максимального правдоподобия с ограниченной информацией (МНКОИ), который является ММП-оценкой (4) при нормально распределенных ошибках. Двухшаговый МНК используется чаще, поскольку он реализован во многих стандартных программных пакетах.

Как только набор инструментов доступен, инструментальная оценка для β в (4) вычисляется следующим образом:

$$\hat{\beta}_n^{IV} = (X'P_ZX)^{-1}X'P_Zy, \quad (5)$$

где $P_Z = Z(Z'Z)^{-1}Z'$. Она состоятельна и асимптотически нормальна, если $\text{plim } Z'\epsilon/n = 0$, и $\text{plim } Z'Z/n$ и $\text{plim } Z'X/n$ существуют и имеют полный ранг по столбцам. Несмещенность инструментальной оценки обсуждается ниже. При рассмотрении этой оценки часто используется асимптотический подход, поскольку ее математическое ожидание не существует в случае, когда число инструментов равно числу объясняющих переменных (Wooldridge, 2002). Стандартные методы инференции могут быть использованы для инференции относительно неизвестных параметров или тестирования гипотез. ММПОИ-оценка считается немного сложнее, чем 2ШМНК-оценка. Однако, если только инструменты не являются слишком слабыми, обе оценки имеют одинаковые асимптотические свойства (Davidson & MacKinnon, 1993).

Соображения по использованию инструментальных переменных

Проблемой в практических приложениях является поиск качественных инструментов. В целом, нет четких направлений поиска, и вообще инструменты нелегко найти. Кроме того, получение дополнительных данных может обойтись очень дорого. Как таковые, инструменты часто выбираются исходя из предположений или даже просто доступности, что потенциально приводит к их негодности. Условие $\mathbb{E}[\epsilon|Z] = 0$ требует отсутствия прямой связи между инструментами и зависимой переменной, что во многих практических ситуациях весьма спорно.

Wooldridge (2002), например, обсуждает годность инструмента «порядковый номер призыва в армию», используемого в Angrist (1990) для оценки влияния статуса ветерана войны во Вьетнаме на личный доход. Хотя порядковый номер призыва в армию, определяемый из лотереи, является случайным, индивиды, с наибольшей вероятностью подпадающие под призыв, могут продолжить образование для повышения шансов отсрочки от призыва, или же работодатели могут захотеть вкладывать в образование и тренинги работников, с наименьшей вероятностью подпадающих под призыв. Bound, Jaeger & Baker (1995) ставят под вопрос экзогенность инструментов, связанных с «кварталом рождения», используемых в Angrist & Krueger (1991), которые оценивают влияние образования на доход. Они предоставляют свидетельства наличия слабой корреляции между инструментом «квартал рождения» и зависимой переменной «заработная плата», не зависящей от эффекта квартала рождения на образование и достаточно сильной для смещения инструментальной оценки. Card (1999, 2001) дает обширный обзор дискуссии о годности переменных, связанных с биографией семьи (например, образования родителей), и институциональных характеристик системы образования (например, близкое расположение колледжа) как инструментов для эндогенного регрессора «образование». При оценивании спроса лагированные значения показателей рекламной деятельности часто используются в качестве инструментов в моделях рыночной реакции, но они не являются годными, когда существуют эталонные цены, формируемые исторически (см. Bronnenberg & Mahajan, 2001). Yang, Chen & Allenby (2003) замечают, что лагированные значения цены могут не быть подходящими инструментами из-за заблаговременных закупок и накопления запасов. Кроме того, использование лагированных переменных как экзогенных само по себе является потенциальным источником эндогенности (Arellano, 2002). Nevo (2001) использует данные по ценам других рынков в качестве инструментов для цены, но замечает,

что эти инструменты не являются годными, когда имеют место общие (страновые) шоки спроса, или когда рекламная или промоутерская деятельность скоординирована по рынкам. Это вероятно, если один и тот же производитель или продавец действует на нескольких рынках. Хотя носитель издержек мог бы быть потенциальным инструментом для цены, Nevo (2000) заключает, что он редко наблюдаем, а прокси-переменные для издержек обычно недостаточно вариабельны.

Экзогенность инструментов – это лишь один из двух критериев качества инструментов. Вдобавок, доступные инструменты могут быть «слабыми», в том смысле, что они слабо коррелируют с эндогенными регрессорами. Stock, Wright & Yogo (2002) утверждают: «Исследователи-практики часто сталкиваются со слабыми инструментами. Поиск экзогенных инструментов – тяжелый труд, и те свойства, которые делают инструменты экзогенными, [...] также могут делать их и слабыми». К сожалению, статистические свойства инструментальных оценок и основанная на этих оценках инференция являются чувствительными к выбору и корректности инструментов, даже в больших выборках. Следовательно, исследователи, изучающие один и тот же вопрос, но использующие разные наборы инструментов, могут прийти к различным выводам.

Agellano (2002) замечает, что «многие темы [по инструментальному оцениванию], появившиеся [...] в эконометрической литературе в 80-х и 90-х, были на удивление зрелым образом разработаны в статьях Саргана 1958 и 1959 гг.» (Sargan, 1958, 1959). Далее представлен обзор некоторых свежих результатов о слабых инструментах, появившихся в эконометрической литературе (технические детали более подробно изложены в Stock, Wright & Yogo, 2002 и Hahn & Hausman, 2003).

Слабые инструменты

Последние результаты в эконометрической литературе показали, что присутствие слабых инструментов не только снижает точность ИП-оценок, но также может привести к несостоятельности и смещению инструментальной оценки, превосходящему смещение МНК-оценки. Более того, стандартные асимптотические приближения не срабатывают (Staiger & Stock, 1997, Bound, Jaeger & Baker, 1995, Hahn & Hausman, 2002, 2003). Как следствие, на традиционное тестирование гипотез и построение доверительных интервалов нельзя полагаться. Слабые инструменты могут возникать в случаях, когда инструменты не обладают высокой объясняющей силой по отношению к эндогенным регрессорам, или когда число инструментов велико. Обсудим далее три потенциальные ловушки использования инструментальной оценки при наличии слабых инструментов: (1) смещение 2ШМНК-оценки в конечных выборках, (2) ситуации, в которых инструменты потенциально коррелируют с ϵ , и (3) плохое асимптотическое приближение фактического распределения инструментальной оценки.

В конечных выборках ИП- или 2ШМНК-оценка смещена в том же направлении, что и МНК-оценка. На этот факт часто не обращают внимание при проведении эмпирических исследований. Даже когда $\mathbb{E}[\epsilon|Z] = 0$, $\hat{\beta}_n^{IV} = \beta + (X'P_Z X)^{-1} X'P_Z \epsilon$ является в общем случае смещенной, так как $\mathbb{E}[(X'P_Z X)^{-1} X'P_Z \epsilon] \neq 0$. Смещение возникает из-за того, что коэффициенты Π в (4) ненаблюдаемы. Если бы $Z\Pi$ были наблюдаемы, МНК-регрессия y на $Z\Pi$ давала бы несмещенные результаты, но в действительности оценка Π получается из регрессии X на Z . Buse (1992) среди прочих показывает, что это смещение в конечных выборках является функцией от количества инструментов, так что увеличение их числа может увеличить смещение инструментальной оценки. Однако это смещение будет расти лишь пропорционально при росте количества инструментов более быстро, чем доля объясненной вариации в эндогенных регрессорах. Следовательно, добавление важного или сильного инструмента необязательно увеличивает смещение, но добавление менее важных инструментов или наличие слабых инструментов, несомненно, приведет к увеличению смещения. Bound,

Jaeger & Baker (1995) и Hahn & Hausman (2003) показывают, что смещение обратно связано с F -статистикой в регрессии эндогенных объясняющих переменных на инструменты. Их результаты говорят о том, что (частный) R^2 и F -статистика для регрессии первого шага (то есть регрессии X на Z) полезны как грубые показатели качества инструментальных оценок и обязательно должны сообщаться. Распределение ММПОИ-оценки в конечных выборках не имеет конечных моментов, характеризуется толстыми хвостами, а также обычно является менее чувствительной к добавлению избыточных инструментов, чем 2ШМНК-оценка. Тем не менее, когда инструменты слабые, даже эта оценка не всегда решает проблему (Hahn & Hausman, 2003, Kleibergen & Zivot, 2003).

Вторая проблема, связанная со слабыми инструментами, – это относительная несостоятельность инструментальной оценки по сравнению с МНК-оценкой в случае, когда инструменты потенциально коррелируют с ϵ , то есть сами являются эндогенными. Bound, Jaeger & Baker (1995) показывают, что относительная несостоятельность инструментальной оценки по сравнению с МНК равна (для простоты предположим, что $k = q = 1$)

$$\frac{\text{plim } \hat{\beta}_n^{\text{IV}} - \beta}{\text{plim } \hat{\beta}_n^{\text{OLS}} - \beta} = \frac{\rho_{z,\epsilon} / \rho_{x,\epsilon}}{\rho_{x,z}},$$

где $\rho_{x,z}$ – коэффициент корреляции между x и z , а другие элементы определяются аналогично. Когда инструменты слабые, $\rho_{x,z} \rightarrow 0$, то есть даже слабая корреляция между z и ϵ может давать большую относительную несостоятельность инструментальной оценки, приводя к *большей* несостоятельности последней, чем для МНК.

В-третьих, если инструменты слабые, то даже в больших выборках классические (первого порядка) асимптотические приближения неадекватны. Это демонстрируют (в том числе) Nelson & Startz (1990), которые замечают, что при слабых инструментах классическая асимптотическая дисперсионная матрица будет больше, а асимптотическое распределение β обладает большим разбросом. Однако также показано, что асимптотическое распределение очень плохо приближает реальную плотность распределения в конечной выборке (распределение является бимодальным, с тяжелыми хвостами и сконцентрировано вокруг предела по вероятности МНК-оценки, а не вокруг истинного значения). Если асимптотическая дисперсия $\hat{\beta}_n^{\text{IV}}$ ниже, то есть когда инструменты более сильные, классическое приближение лучше. Как следствие, в присутствии слабых инструментов процедуры инференции, основанные на классической асимптотике, дают неверные результаты. Хотя методы для анализа конечных выборок могут быть использованы в таких ситуациях, их использование на практике затруднено из-за ограничительных предположений, вычислительных трудностей при работе с распределениями и отсутствия четкой схемы тестирования гипотез и построения доверительных интервалов. Проблема слабых инструментов важна не только в малых выборках, ее нельзя игнорировать и в больших. Это продемонстрировано в Bound, Jaeger & Baker (1995), которые показывают, что для исследования Angrist & Krueger (1991) можно получить аналогичные результаты при использовании искусственных случайных инструментальных переменных, несмотря на 329500 наблюдений в выборке.

Проверка инструментов на пригодность

Поскольку (асимптотические) свойства инструментальной оценки чувствительны к выбору инструментов, желательно иметь какую-то меру слабости. Результаты недавних исследований предлагают обязательно сообщать R^2 - и F -статистики, полученные для регрессии первого шага. Stock, Wright & Yogo (2002), например, заключают, что F -статистика для регрессии первой стадии должна быть больше 10, чтобы инференция, основанная на 2ШМНК-оценке, была достоверной. Bowden & Turkington (1984) утверждают, что следует искать инструменты, максимизирующие все канонические корреляции с X . Staiger & Stock (1997) разработали

основанную на данных меру относительного смещения, большие значения которой должны предупредить исследователя о возможных проблемах, связанных с корреляцией инструментов и ошибок. Bowden & Turkington (1984) и Verbeek (2000) (среди прочих) предлагают тест на приемлемость инструментов, когда $q > k$. Тест отвергает нулевую гипотезу, когда данные свидетельствуют против совместной годности инструментов, хотя невозможно определить, какие именно инструменты негодные. Метод Bowden & Turkington (1984) может быть использован для проверки приемлемости дополнительного набора инструментов, но этот тест ничего не говорит о возможной слабости инструментов. Hahn & Hausman (2003) утверждают, что этот тест отвергает нулевую гипотезу слишком часто в присутствии слабых инструментов, что является существенным недостатком теста, так как он часто применяется для тестирования экономической теории, заключенной в модели.

Hahn & Hausman (2002) разработали тест на пригодность инструментальных переменных, который одновременно проверяет как экзогенность, так и релевантность инструментов. Тест сравнивает 2ШМНК-оценки в прямой и обратной ИП-регрессиях, которые, как показывается, эквивалентны при нулевой гипотезе о правильности стандартной асимптотики. Тест-статистика легко считается и имеет t -распределение при нулевой гипотезе. Отвержение нулевой гипотезы может означать либо нарушение предположения об экзогенности инструментов, либо их слабость. Hahn & Hausman (2002) предлагают двухшаговый метод, основанный на этом тесте, для принятия решения об использовании либо 2ШМНК-оценки, либо ММПОИ-оценки, либо ни одной из них. Ebbes (2004) разработал другой метод для проверки приемлемости наблюдаемых инструментов, который основан на методе латентных инструментальных переменных (Ebbes, Wedel, Böckenholt & Steerneman, 2005). В отличие от теста Hahn & Hausman (2002), он может быть использован для проверки инструментов на слабость и эндогенность как по отдельности, так и одновременно. Более того, даже если инструменты непригодны, оценки регрессионных параметров тем не менее можно использовать, так как метод латентных инструментальных переменных не основан ни на качестве, ни на доступности наблюдаемых инструментов. Симуляции показывают, что этот метод не ведет к проблемам с размером теста в присутствии слабых инструментов в отличие от классического теста на сверхидентифицирующие ограничения.

Выбор количества инструментов

Смещение инструментальной оценки в конечных выборках является функцией от количества инструментов, откуда следует, что не стоит использовать их слишком много, хотя условие идентифицируемости требует их по крайней мере столько же, сколько имеется регрессоров ($q \geq k$). Более того, увеличение количества инструментов ведет к потере степеней свободы, и регрессия первого шага (X на Z) страдает от чрезмерной подгонки. Sargan (1958) заключает, что «если первые несколько инструментов хорошо подобраны, обычно наблюдается не улучшение, а наоборот, ухудшение доверительных интервалов по мере роста числа инструментов свыше трех или четырех». Кроме того, он отмечает, что «оценки могут быть сильно смещены, если инструментов становится слишком много». В отличие от этих конечновыборочных результатов, асимптотическая теория говорит о том, что инструментальная оценка с одним дополнительным инструментом по крайней мере настолько же эффективна, то есть инструменты можно добавлять без ухудшений (Davidson & MacKinnon, 1993).

Bowden & Turkington (1984) предлагают применять метод главных компонент для снижения размерности матрицы $Z'Z$ и выбора первых p главных компонент в качестве инструментов. Однако этот подход не учитывает корреляцию между X и Z , то есть силу инструментов. Donald & Newey (2001) разработали критерий среднеквадратичной ошибки, который следует минимизировать выбором инструментов. Авторы обнаружили, что их метод выбора инструментов обычно дает улучшение качества оценки. В большинстве случаев ММПОИ-оценка лучше, чем 2ШМНК-оценка, хотя последняя дает лучший результат, когда степень эндо-

генности низкая. В случае слабых инструментов явно следует использовать меньшее число инструментов.

Тестирование на зависимость между регрессорами и ошибками

Учитывая возможные опасности при использовании инструментальной оценки и проблему поиска инструментов, желательно для начала провести тест на коррелированность регрессоров с ошибками. К сожалению, невозможно напрямую рассматривать $X'\epsilon$, так как ϵ не наблюдается, а МНК-оценивание дает $X'\hat{\epsilon} = 0$ по построению. Исключением является тестирование гипотезы $X'\alpha = 0$ в многоуровневых моделях со случайными эффектами, где $\alpha = (\alpha_1, \dots, \alpha_n)'$ являются случайными эффектами, так как имеется готовый статистический тест (см., например, Ebbes, Böckenholt & Wedel, 2004). Для тестирования на эндогенность в стандартной регрессионной модели, однако, необходимы инструменты хорошего качества. Тогда можно применить тест на основе общей процедуры Хаусмана (Hausman, 1978), который основан на разности между $\hat{\beta}_n^{\text{OLS}}$ и $\hat{\beta}_n^{\text{IV}}$. Хаусман предложил тестовую статистику, имеющую асимптотическое χ^2 -распределение при нулевой гипотезе о некоррелированности регрессоров и ошибок. Недостаток этой процедуры в том, что для расчета $\hat{\beta}_n^{\text{IV}}$ необходимы инструменты, хотя впоследствии исследователь может заключить, что они не нужны. Более того, этот тест чувствителен к слабым инструментам (Staiger & Stock, 1997). Действительно, тест Хаусмана может неправильно принять гипотезу о возможности использования МНК-оценки, так как она смещена при наличии слабых инструментов (Hahn & Hausman, 2003). Ebbes, Wedel, Böckenholt & Steerneman (2005) предлагают тест без использования инструментов для проверки на коррелированность регрессоров и ошибок в линейной регрессионной модели с одним эндогенным регрессором и нормальными ошибками и показывают, что этот тест обладает достаточной мощностью для широкого класса условий.

Решения проблемы слабых инструментов, основанные на инструментальных переменных

Hahn & Hausman (2003) и Stock, Wright & Yogo (2002) дают обзор большей части эконометрических исследований о способах решения проблемы слабых инструментов в практических приложениях. Далее следует краткое изложение этих способов.

Хорошо известно, что асимптотическое приближение первого порядка для распределения инструментальной оценки является плохим при наличии слабых инструментов. Некоторые исследователи предлагают улучшенные асимптотические приближения для распределения оценки в конечных выборках в такой ситуации. Staiger & Stock (1997) разработали альтернативную асимптотику, которая моделирует коэффициенты регрессии первого шага как локально нулевые, то есть слабо коррелированные, без предположения о нормальности. В рамках этой асимптотики они показали, что если инструменты являются слабыми, то 2ШМНК- и ММПОИ-оценки несостоятельны и имеют нестандартные асимптотические распределения, причем смещение ММПОИ-оценки меньше такового у 2ШМНК-оценки, особенно в малых выборках. Более того, авторы выводят свойства процедур инференции (t -тестирование, нормы покрытия доверительных интервалов и тестирование на сверхидентифицирующие ограничения). Bekker (1994) разработал альтернативную асимптотику для моделей с нормальными ошибками, в которых растут и размер выборки, и количество инструментов. Симуляции показывают, что эта асимптотика дает лучшее приближение для среднего и большого числа инструментов, и что ММПОИ-оценке следует отдавать предпочтение перед обычной инструментальной оценкой. Однако результаты Беккера применимы лишь к случаю с нормальными ошибками и не улавливают ненормальность, наблюдаемую в точном распределении в конечной выборке, когда присутствуют слабые инструменты (Staiger & Stock, 1997).

Помимо работы по поиску лучших альтернатив обычной асимптотике первого порядка, разработаны «полностью робастные» тесты для проверки гипотез и методы построения доверительных интервалов для β с приблизительно правильным размером и нормой покрытия при наличии слабых инструментов. Один из таких робастных тестов на проверку гипотезы $\beta = \beta_0$ – это статистика Андерсона–Рубина (Anderson & Rubin, 1949), которая не зависит от степени недоидентификации. Однако, ей может не хватать мощности из-за потери степеней свободы при увеличении количества инструментов. K -статистика (Kleibergen, 2002) имеет схожие асимптотические свойства с минимальным числом степеней свободы. Bekker & Kleibergen (2003) исследуют ее распределение в конечных выборках при нормальности ошибок. Предлагались также и другие тесты, см., например, Staiger & Stock (1997). Stock, Wright & Yogo (2002) сравнивают мощности нескольких тестов при разных условиях. Учитывая двойственность проверки гипотез и построения доверительных интервалов, робастные тесты могут быть использованы и для получения доверительных интервалов. Когда инструменты слабые, эти интервалы могут быть неограничены, указывая на очень ограниченное количество информации для проведения инференции о β .

Указанные методы проверки гипотез и построения доверительных интервалов не дают точечной оценки β . Кроме того, они могут быть вычислительно сложными. Были предложены некоторые альтернативы 2ШМНК-оценке, которые являются более робастными и потенциально более надежными, если инструменты слабые. Оценки с нулевым смещением второго порядка, такие как ММПОИ-оценка, часто предлагаются в качестве робастных альтернатив. Эти оценки, однако, не имеют моментов в конечных выборках, что может стать проблемой при практической реализации (Hahn & Hausman, 2003). Другими альтернативами являются инструментальные переменные «складного ножа» (Angrist, Imbens & Krueger, 1999), k -оценка Фуллера (Fuller, 1977) или подправленная на смещение 2ШМНК-оценка (Donald & Newey, 2001). Stock, Wright & Yogo (2002) находят, что эти частично робастные оценки являются более надежной альтернативой 2ШМНК-оценке в случаях со слабыми инструментами. Тем не менее, Hahn & Hausman (2003) рекомендуют быть очень осторожными при применении оценок, распределение которых не имеет моментов (например, ММПОИ-оценки). Они обнаруживают, что 2ШМНК-оценка, 2ШМНК-оценка «складного ножа» и k -оценка Фуллера ведут себя лучше, и утверждают, что «пессимизм по поводу инструментов преувеличен для 2ШМНК-оценки, поэтому, возможно, 2ШМНК-оценка часто ведет себя лучше, чем ожидается, в смысле среднеквадратической ошибки в ситуации со слабыми инструментами». Тест на спецификацию, разработанный в Hahn & Hausman (2002) можно использовать для выбора из альтернативных инструментальных оценок. Как Stock, Wright & Yogo (2002), так и Hahn & Hausman (2003) подчеркивают, что большая часть литературы по слабым инструментам предполагает экзогенность инструментов. Невыполнение этого условия, особенно вкупе со слабыми инструментами, ведет к дополнительным сложностям, и в этом случае МНК может давать наилучшие результаты по сравнению с предложенными выше способами учета слабых инструментов.

4 Альтернативные подходы к решению проблемы эндогенности

В некоторых приложениях сама природа процесса, порождающего данные, или же сама причина эндогенности подразумевают подходящие инструменты или даже другой подход к оцениванию. Wooldridge (2002) предлагает три альтернативных способа решения проблемы пропущенных переменных, включая МНК с прокси-переменными и использование индикаторов ненаблюдаемых переменных. Последний метод требует инструментального оценивания, но отличается от классического метода ИП. Индикаторный подход предполагает существование возможно неправильно измеренной прокси-переменной для пропущенной переменной w , которая нуждается в инструменте, в то время как классический метод инструментальных

переменных оставляет пропущенную переменную w внутри ошибки, и все элементы x , коррелированные с w , требуют инструментов. Когда одни и те же кросс-секционные объекты наблюдаемы во времени, инструментальные переменные не нужны, так как можно применить оценку с фиксированными эффектами для учета пропущенных переменных, если корреляция между регрессорами и ошибкой возникает из независимых от времени источников (см., например, Ebbes, Böckenholt & Wedel, 2004). Аналогично, Card (1999) предлагает обзор исследований, использующих данные по братьям и близнецам для оценки отдачи от образования, и утверждает, что проблема пропущенной переменной «способности» устраняется при оценивании внутрисемейных данных. Для моделей с ошибками измерения, авторегрессионных моделей и систем одновременных уравнений процесс, порождающий данные, может предложить подходящие инструменты. Например, модели с ошибками измерения можно оценивать, используя инструментальные переменные, когда инструменты измеряются с ошибкой (Wansbeek & Meijer, 2000, White, 2001). Другой метод оценивания моделей с ошибками измерения, предложенный в Wald (1940), использует информацию по наблюдаемой переменной группировки. Эта группировка должна быть независима от ошибок и различать высокие и низкие значения ненаблюдаемой истинной переменной. Аналогично, в авторегрессионных моделях лаги более высокого порядка могут служить инструментами для модели с лагированными зависимыми переменными в качестве регрессоров. В системах одновременных уравнений экзогенные переменные, не включенные в уравнение, часто могут служить инструментами и легко доступны (Greene, 2000).

Далее мы рассмотрим три других подхода к проблеме зависимости между регрессорами и ошибками: (1) «экономные» методы инструментальных переменных, (2) методы для моделирования спроса, издержек и конкуренции и (3) пространственную эконометрику.

«Экономные» методы инструментальных переменных

Ebbes, Wedel & Böckenholt (2006) рассматривают три метода оценивания, которые не основываются на наблюдаемых инструментальных переменных для идентификации регрессионных параметров в моделях с эндогенностью: подход «моментов высокого порядка» (НМ, *higher moments*) (Erickson & Whited, 2002, Lewbel, 1997), метод «идентификации через гетероскедастичность» (И, *identification through heteroskedasticity*) (Rigobon, 2003 и Hogan & Rigobon, 2003) и метод «латентных инструментальных переменных» (LIV, *latent instrumental variables*) (Ebbes, Wedel, Böckenholt & Steerneman, 2005). Учитывая проблемы классического инструментального оценивания, в частности проблему слабых инструментов и малую доступность наблюдаемых инструментов, «экономные» методы инструментальных переменных могут быть полезной альтернативой.

В НМ-подходе инструменты строятся по имеющимся данным, используя моменты высоких порядков. Состоятельное оценивание требует (среди прочего), чтобы ошибки измерения и ошибки в структурных уравнениях были независимы и имели моменты всех порядков, однако дальнейших предположений о формах распределения не делается. И-оценка также основана на стратегии идентификации с использованием моментов более высокого порядка и предполагает, что доступна некоторая информация о гетероскедастичности ошибок. Для подсчета И-оценки требуется наблюдаемая переменная группировки, которая описывает структуру гетероскедастичности в ошибках. LIV-подход аппроксимирует ненаблюдаемые инструменты латентной дискретной переменной. LIV-модель принадлежит классу моделей со смесью нормальных распределений, когда ошибки в регрессионной модели нормально распределены. В отличие от НМ- и И-оценок, этот подход основан на правдоподобии, а идентифицируемость может быть доказана из свойств моделей со смесями вероятностных распределений. Оценки максимального правдоподобия для регрессионных параметров LIV-модели можно получить, используя только зависимые и независимые переменные. Ebbes, Wedel & Böckenholt (2006) докладывают о симуляциях, которые демонстрируют «за» и «против» этих трех методов

при различных условиях. Вывод состоит в том, что при разумном применении эти «экономные» методы ИП могут быть полезными альтернативами в ситуациях, когда есть эндогенный регрессор, но недоступны наблюдаемые инструменты хорошего качества.

Методы моделирования спроса, издержек и конкуренции

Ряд исследований рассматривают эндогенность цены на рынках дифференцированных продуктов. Цена эндогенно определяется спросом и предложением. Berry (1994) и Berry, Levinsohn & Pakes (1995) разработали модель рыночного равновесия с логистической функцией спроса, которая (модель) адаптируется для возможности применения традиционного метода инструментальных переменных. Этот метод применим как для агрегированных, так и для дезагрегированных данных или их комбинации. Итоговая система получается путем объединения модели дискретного выбора для потребительского поведения индивида и функции издержек. Эти две модели заключены в систему установления цен фирмами на рынках дифференцированных товаров. Совместное оценивание ведет к потенциально более эффективным оценкам, чем при использовании только модели спроса, где берется инструмент для цены. Более того, полная система дает детальную информацию о структуре издержек и природе конкуренции. Такие равновесные модели, однако, накладывают больше требований к данным, и некорректная спецификация поведения фирм может привести к смещенным оценкам. Данный подход широко применялся и модифицировался, см., например, Chintagunta, Erdem, Rossi & Wedel (2006), где дан свежий обзор.

Пространственная эконометрика

Недавно опубликованы два исследования в области маркетинга, которые решают проблему эндогенности переменных маркетингового комплекса, используя пространственные зависимости в наблюдаемых на рынке данных, постоянные или несильно варьирующиеся во времени. Эти зависимости вызваны тем, что экономические агенты организованы в пространстве или имеют схожее расположение магазинов. Bronnenberg & Mahajan (2001) идентифицируют корреляцию между переменными маркетингового комплекса и ошибкой, накладывая измеримую пространственную структуру на случайные ошибки в модели. Пространственная карта возникает из ненаблюдаемых действий розничных продавцов в различных местах торговли с несколькими соседствующими рынками. Bronnenberg & Mahajan строят пространственную карту, исходя из географической близости. Можно оценить и протестировать эффект ненаблюдаемого поведения продавцов, включая карту в модель. Результаты авторов применительно к анализу рынка мексиканской еды говорят о том, что ненаблюдаемая компонента в зависимой переменной связана с переменными маркетингового комплекса. van Dijk, van Heerde, Leeflang & Wittink (2004) рассматривают оценку эластичности площадей витрин, отведенных под разные товары, на основе эндогенных данных о площадях витрин для выкладки товаров. Оценивание этих эластичностей затруднено из-за минимальной вариации (во времени) показателей площадях витрин, отведенных под разные товары. Авторы, основываясь на работе Bronnenberg & Mahajan (2001), предлагают моделировать корреляцию между площадью витрин и случайными ошибками, используя пространственную структуру, основанную на схожести характеристик магазинов, потребителей или конкурентов. Их результаты применительно к часто используемым товарам ежедневной гигиены дают оценки эластичностей, превосходящие по качеству оценки из модели с пространственной структурой на основе географической близости в терминах правильности прогнозов. Поскольку продавцы обычно принимают решение о распределении витрин по разным товарам на основе характеристик магазинов, потребителей или конкурентов, ожидается, что схожесть двух географически близких магазинов ниже, чем схожесть двух магазинов с одинаковым профилем в разных регионах.

5 Заключение

Методы инструментальных переменных разработаны для преодоления проблемы зависимости между регрессорами и ошибками. Из настоящего обзора ясно, однако, что поиск подходящих инструментов в каждой конкретной ситуации может быть проблематичным. Традиционные методы инструментальных переменных полагаются на экономическую теорию или интуицию в отыскании инструментов. К сожалению, зачастую инструменты хорошего качества просто недоступны, и инструменты могут быть слабыми или/и эндогенными. Хорошо известно, что качество процедур инференции на основе инструментальной оценки напрямую зависит от качества используемых инструментов. Это явилось темой некоторых недавних исследований в эконометрике, и был сделан ряд предложений для улучшения инференции при наличии слабых инструментов. В целом, меньшее количество инструментов предпочитается большему. Более того, R^2 - или F -статистика в регрессии первого шага должны быть всегда рассчитаны и сообщаться как меры силы инструментов. Большинство результатов по проблеме слабых инструментов выводятся при условии экзогенности инструментов, и вопрос о том, что делать в ситуациях с эндогенными инструментами, остается открытым.

Для исследователей-практиков поиск подходящих инструментов затруднителен, а выбор обычно небольшой. Альтернативные «экономные» методы инструментальных переменных, описанные выше, пытаются идентифицировать регрессионные параметры не через наблюдаемые инструменты, а путем использования характеристик распределений эндогенных регрессоров, и могут считаться альтернативой или дополнением к классическому инструментальному оцениванию. Без наличия инструментов хорошего качества техника классического инструментального оценивания ненадежна. Теория, с одной стороны, говорит о том, что наилучшими инструментами являются переменные, сильно коррелированные с эндогенными регрессорами. С другой стороны, однако, чем больше они коррелированы, тем меньше надежда на то, что сами эти инструменты не коррелируют с ошибками (см. Greene, 2000).

Благодарности

Данное эссе основано на второй главе моей диссертации (Ebbes, 2004). Я бы хотел поблагодарить моих научных руководителей, профессоров Михеля Веделя, Ульфа Бёкенхольта и Тона Штернемана за ценные комментарии и предложения по ранним версиям эссе.

Список литературы

- Anderson, T.W. & Rubin, H. (1949). Estimators on the parameters of a single equation in a complete set of stochastic equations. *Annals of Mathematical Statistics* 21, 570–582.
- Angrist, J.D. (1990). Lifetime earnings and the vietnam era draft lottery: Evidence from social security administrative records. *American Economic Review* 80, 313–336.
- Angrist, J.D., G.W. Imbens & A.B. Krueger (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics* 14, 57–67.
- Angrist, J.D., G.W. Imbens & D.B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association* 91, 444–455.
- Angrist, J.D. & A.B. Krueger (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics* 56, 979–1014.
- Arellano, M. (2002). Sargan's instrumental variables estimation and the generalized method of moments. *Journal of Business & Economic Statistics* 20, 450–459.
- Bagozzi, R.P., Y. Yi & K.D. Nassen (1999). Representation of measurement error in marketing variables: Review of approaches and extension to three-faced designs. *Journal of Econometrics* 89, 393–421.

- Bekker, P.A. (1994). Alternative approximations to the distributions of instrumental variable estimators. *Econometrica* 62, 657–681.
- Bekker, P.A. & F. Kleibergen (2003). Finite-sample instrumental variables inference using an asymptotic pivotal statistic. *Econometric Theory* 19, 744–753.
- Berry, S., J. Levinsohn & A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica* 63, 841–890.
- Berry, S.T. (1994). Estimating discrete-choice models of product differentiation. *RAND Journal of Economics* 25, 242–262.
- Bound, J., D.A. Jaeger & R.M. Baker (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association* 90, 443–450.
- Bowden, R.J. & D.A. Turkington (1984). *Instrumental Variables*. New York: Cambridge University Press.
- Bronnenberg, B.J. & V. Mahajan (2001). Unobserved retailer behavior in multimarket data: Joint spatial dependence in market shares and promotion variables. *Marketing Science* 20, 284–299.
- Buse, A. (1992). The bias of instrumental variables estimators. *Econometrica* 60, 173–180.
- Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics* 3A, 1801–1863. Amsterdam: Elsevier.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica* 69, 1127–1160.
- Chintagunta, P., T. Erdem, P. Rossi & M. Wedel (2006). Structural modeling in marketing: Review and assessment. *Marketing Science*, в печати.
- Chintagunta, P.K. (2001). Endogeneity and heterogeneity in a probit demand model: Estimation using aggregate data. *Marketing Science* 20, 442–456.
- Davidson, R. & J.G. MacKinnon (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- van Dijk, A., H. J. van Heerde, P.S.H. Leeflang, & D.R. Wittink (2004). Similarity-based spatial methods for estimating shelf space elasticities from correlational data. *Quantitative Marketing and Economics* 2, 257–277.
- Donald, S.G. & W.K. Newey (2001). Choosing the number of instruments. *Econometrica* 69, 1161–1191.
- Ebbes, P. (2004). *Latent Instrumental Variables: A New Approach to Solve for Endogeneity*. PhD thesis, University of Groningen.
- Ebbes, P., U. Böckenholt & M. Wedel (2004). Regressor and random-effects dependencies in multilevel models. *Statistica Neerlandica* 58, 161–178.
- Ebbes, P., M. Wedel & U. Böckenholt (2006). Frugal IV alternatives to identify the parameter for an endogenous regressor. *Journal of Applied Econometrics*, в печати.
- Ebbes, P., M. Wedel, U. Böckenholt & A.G.M. Steerneman (2005). Solving and testing for regressor-error (in) dependence when no instrumental variables are available: With new evidence for the effect of education on income. *Quantitative Marketing and Economics* 3, 365–392.
- Erickson, T. & T.M. Whited (2002). Two-step GMM estimation of the errors-in-variables model using high-order moments. *Econometric Theory* 18, 776–799.
- Fuller, W. (1977). Some properties of a modification of the limited information estimator. *Econometrica* 45, 939–953.
- Gönül, F.F., B.-D. Kim & M. Shi (2000). Mailing smarter to catalog customers. *Journal of Interactive Marketing* 14, 2–16.
- Greene, W.H. (2000). *Econometric Analysis*. New Jersey: Prentice-Hall.

- Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica* 45, 1–22.
- Hahn, J. & J. Hausman (2002). A new specification test for the validity of instrumental variables. *Econometrica* 70, 163–189.
- Hahn, J. & J. Hausman (2003). Weak instruments: Diagnosis and cures in empirical econometrics. *Recent Advances in Econometric Methodology* 93, 118–125.
- Hamilton, B.H. & J.A. Nickerson (2003). Correcting for endogeneity in strategic management research. *Strategic Organization* 1, 51–78.
- Hausman, J.A. (1978). Specification tests for econometrics. *Econometrica* 46, 1251–1271.
- Hogan, V. & R. Rigobon (2003). Using unobserved supply shocks to estimate the returns to education. Technical report, University College Dublin.
- Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl & T.-C. Lee (1985). *The Theory and Practice of Econometrics*. New York: John Wiley & Sons.
- Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica* 70, 1781–1803.
- Kleibergen, F. & E. Zivot (2003). Bayesian and classical approaches to instrumental variables regression. *Journal of Econometrics* 114, 29–72.
- Lewbel, A. (1997). Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica* 65, 1201–1213.
- Nelson, C.R. & R. Startz (1990). Some further results on the exact small sample properties of the instrumental variable estimator. *Econometrica* 58, 967–976.
- Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics & Management Strategy* 9, 513–548.
- Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69, 307–342.
- Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics* 85, 777–792.
- Ruud, P.A. (2000). *An Introduction to Classical Econometric Theory*. New York: Oxford University Press.
- Sargan, J.D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica* 26, 393–415.
- Sargan, J.D. (1959). The estimation of relationships with autocorrelated residuals by the use of instrumental variables. *Journal of the Royal Statistical Society, Series B* 21, 91–105.
- Shugan, S.M. (2004). Endogeneity in marketing decision models. *Marketing Science* 23, 1–3.
- Snijders, T.A.B. & R.J. Bosker (1999). *Multilevel Analysis*. London: SAGE Publications.
- Staiger, D. & J.H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.
- Stock, J.H., J.H. Wright & M. Yogo (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics* 20, 518–529.
- Sudhir, K. (2001). Competitive pricing behavior in the auto market: A structural analysis. *Marketing Science* 20, 42–60.
- Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources* 33, 127–169.
- Vella, F. & M. Verbeek (1998). Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men. *Journal of Applied Econometrics* 13, 163–183.
- Verbeek, M. (2000). *A Guide to Modern Econometrics*. Chichester: John Wiley & Sons.

- Villas-Boas, J.M. & R.S. Winer (1999). Endogeneity in brand choice models. *Management Science* 45, 1324–1338.
- Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics* 11, 284–300.
- Wansbeek, T. & E. Meijer (2000). *Measurement Error and Latent Variables in Econometrics*. Amsterdam: Elsevier.
- Wansbeek, T. & M. Wedel (1999). Marketing and econometrics: editors' introduction. *Journal of Econometrics* 89, 1–14.
- White, H. (2001). *Asymptotic Theory for Econometricians*. New York: Academic Press.
- Wooldridge, J.M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.
- Yang, S., Y. Chen & G.M. Allenby (2003). Bayesian analysis of simultaneous demand and supply. *Quantitative Marketing and Economics* 1, 251–275.

A non-technical guide to instrumental variables and regressor–error dependencies

Peter Ebbes

Penn State University, University Park, USA

We provide a non-technical summary of most of the recent results that have appeared in the econometric literature on instrumental variables estimation for the linear regression model. Standard inferential methods, such as OLS, are biased and inconsistent when the regressors are correlated with the error term. Instrumental variables methods were developed to overcome this problem, but finding instruments of good quality is cumbersome in any given situation and empirical researchers are often confronted with weak instruments. We review most of the recent studies on weak instruments and point to several methods that have been proposed to deal with such instruments, including “frugal” IV alternatives that do not rely on observed instruments to identify the regression parameters in presence of regressor–error dependencies.