Слабые инструменты[*]

Адриан Паган[†]

Технологический университет Квинсленда, Брисбен, Австралия

Настоящее эссе представляет собой обзор избранной литературы по слабым инструментам. В нем рассматриваются простые модели для иллюстрации проблем, возникающих из-за слабости инструментов, и методы, предложенные для их решения. Поскольку литература по данному вопросу находится в процессе развития, в эссе кратко отражены лишь наиболее свежие результаты.

1 Проблемы, связанные с распределением инструментальной оценки при слабых инструментах

Рассмотрим простую модель

$$y_t = x_t\theta + u_t,$$

где $u_t \sim i.i.d.(0, \sigma_u^2)$, $\mathbb{E}[x_t u_t] \neq 0$. Также имеется набор инструментальных переменных z_t, удовлетворяющих $\mathbb{E}[z_t u_t] = 0$. Предположим, что

$$\begin{pmatrix} z_t \\ x_t \end{pmatrix} \sim i.i.d. \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_z^2 & \sigma_{zx} \\ \sigma_{zx} & \sigma_x^2 \end{pmatrix} \right).$$

Таким образом, имеются один регрессор и один инструмент. Тогда обычная инструментальная оценка имеет следующий вид:

$$\begin{aligned} \hat\theta - \theta_0 &= \left(\sum_t z_t x_t\right)^{-1} \left(\sum_t z_t u_t\right) \\ &= \left(\frac{T^{-1}\sum_t z_t x_t}{\hat\sigma_z \hat\sigma_x}\right)^{-1} \left(\frac{T^{-1}\sum_t z_t u_t}{\hat\sigma_z \sigma_u}\right) \frac{\sigma_u}{\hat\sigma_x} \\ &= \hat\rho_{zx}^{-1} \hat\rho_{zu} \cdot \frac{\sigma_u}{\hat\sigma_x}, \end{aligned} \tag{1}$$

где $\hat\sigma_z$ и $\hat\sigma_x$ – оценки стандартных отклонений z_t и x_t, а $\hat\rho_{zx}$ и $\hat\rho_{zu}$ – оценки коэффициентов корреляции (оценка $\hat\rho_{zu}$ включает истинное стандартное отклонение u_t, σ_u). Из этого выражения видно, что распределение $\hat\theta - \theta_0$ зависит от произведения трех случайных величин, причем считается, что $\sigma_u/\hat\sigma_x$ быстро сходится к константе, так что для изучения остается величина $\hat\rho_{zu}/\hat\rho_{zx}$. Любые проблемы с распределением $\hat\theta$ связаны с $\hat\rho_{zx}$, а именно с тем, насколько случайной является величина $\hat\rho_{zx}$, и насколько вероятны малые по абсолютному значению реализации этой случайной величины. Поскольку $\hat\rho_{zx}$ стоит в знаменателе, для таких реализаций мы получим большие значения $\hat\theta - \theta_0$, что приведет к скошенной плотности распределения $\hat\theta - \theta_0$.

Асимптотическая теория применяется в предположении, что размер выборки достаточно большой, чтобы воспринимать $\hat\rho_{zx}$ как константу и рассматривать

$$T^{1/2}(\hat\theta - \theta_0) = \hat\rho_{zx}^{-1}(T^{1/2}\hat\rho_{zu}) \cdot \frac{\sigma_u}{\hat\sigma_x},$$

[*]Перевод Б. Гершмана и С. Анатольева. Материал подготовлен на основе лекций, прочитанных автором на курсах Economics 4202 в университете Нового Южного Уэльса в 2002–2004 гг. и Economics 607 в университете Джона Гопкинса в 2004 г. Цитировать как: Паган, Адриан (2007) «Слабые инструменты», Квантиль, №2, стр. 71–81. Citation: Pagan, Adrian (2007) “Weak instruments,” Quantile, No.2, pp. 71–81.

[†]Адрес: School of Economics and Finance, Queensland University of Technology, GPO Box 2434, Brisbane, QLD 4001, Australia. Электронная почта: *a.pagan@qut.edu.au*

и, поскольку $T^{1/2}(\hat\rho_{zu}-\rho_{zu}) = T^{1/2}\hat\rho_{zu}$, ожидается, что $T^{1/2}\hat\rho_{zu}$ будет асимптотически иметь стандартное нормальное распределение $N(0,1)$ (если $\rho_{zu}=0$). Кроме того, все другие величины сходятся к своим истинным значениям. Тогда при условии $\rho_{zx}\neq 0$ получаем, что $T^{1/2}(\hat\theta-\theta_0)$ будет иметь асимптотическое распределение $N(0,\rho_{zx}^{-2}\sigma_u^2/\sigma_x^2)$.

Ситуация осложняется, если $\rho_{zx}=0$, поскольку тогда при применении того же подхода в больших выборках происходит деление на нуль. В случае, когда $\rho_{zx}=0$, инструменты называют *нерелевантными*. Они являются *годными*, так как $\rho_{zu}=0$, но представляют малый интерес. Однако можно сказать нечто большее. В частности, предположим, что асимптотически оценка $\hat\rho_{zx}$ также нормально распределена вокруг своего истинного значения (нуля в нашем случае), то есть $T^{1/2}\hat\rho_{zx}$ сходится по распределению к $N(0,v)$ (если $\rho_{zx}=0$, то $v=1$). Тогда имеем

$$\hat\theta-\theta_0 = (T^{1/2}\hat\rho_{zx})^{-1}(T^{1/2}\hat\rho_{zu})\cdot\frac{\sigma_u}{\hat\sigma_x},$$

что в больших выборках реализуется в

$$\frac{N(0,1)}{N(0,1)}\cdot\frac{\sigma_u}{\hat\sigma_x}.$$

Отсюда ясно, что оценка $\hat\theta$ несостоятельна, то есть всегда существует случайный разрыв между $\hat\theta$ и θ_0. Иными словами, в отличие от стандартного случая, когда $\mathbb{V}[\hat\theta-\theta_0]$ стремится к нулю при $T\to\infty$, здесь дисперсия $\hat\theta-\theta_0$ не уменьшается с увеличением размера выборки. Таким образом, в случае нерелевантности инструментов «хорошие» свойства инструментальной оценки не сохраняются. Более того, если две нормальные случайные величины в приведенном выражении независимы, $\hat\theta-\theta_0$ будет иметь распределение Коши, у которого отсутствуют моменты.

Может показаться неправдоподобным, что ρ_{zx} в точности равняется нулю. Более вероятно, что эта величина просто мала. Рассмотрим подробнее, как изменится анализ в этом случае. Для этого предположим, что

$$x_t = z_t\pi + \xi_t,$$

где $\mathbb{E}[z_t\xi_t]=0$. Это не что иное, как уравнение в приведенной форме, в котором z_t предполагается экзогенным. Полагая для удобства, что $\mathbb{E}[z_t]=\mathbb{E}[x_t]=0$, получаем

$$\mathbb{E}[z_tx_t]=\mathbb{E}[z_t^2]\pi,$$

так что

$$\rho_{zx}=\frac{\mathbb{E}[z_tx_t]}{\sigma_x\sigma_z}=\frac{\mathbb{E}[z_t^2]\pi}{\sigma_x\sigma_z}=\frac{\pi\sigma_z}{\sigma_x}.$$

Будем называть *слабым инструментом* такой, для которого величина ρ_{zx} мала. Это означает, что π мало.

На первый взгляд кажется, что асимптотическая теория все еще применима в ситуации слабых, но релевантных инструментов, поскольку $\hat\rho_{zx}$ сходится к ненулевому значению. На самом деле это верно, но не следует спешить с выводами. Все-таки при $\rho_{zx}=0$ теория не работает, и интуиция подсказывает, что если ρ_{zx} отличается от нуля на крайне малую величину, то свойства оценки будут ближе к ситуации $\rho_{zx}=0$, чем к случаю $\rho_{zx}\neq 0$. Конечно, это, в сущности, размышления о свойствах оценки в конечных выборках, то есть о том, насколько большим должно быть T, чтобы действовала асимптотическая теория.

Возможно, полезно поразмышлять над этим эвристически, записав $\hat\rho_{zx}$ как $\rho_{zx}+\eta$, где $\eta\sim N(0,v/T)$, так что

$$T^{1/2}(\hat\theta-\theta_0)=\left\{\frac{T^{1/2}\hat\rho_{zu}}{\rho_{xz}+N(0,v/T)}\right\}\frac{\sigma_u}{\hat\sigma_x}.$$

Ясно, что когда T становится большим, член $N(0, v/T)$ исчезает, делая реализации выражения в знаменателе, $\rho_{zx} + N(0, v/T)$, более близкими к ненулевому значению ρ_{zx}. Но в малой выборке и при достаточно большом значении v возможно, что реализация $\hat{\rho}_{zx}$ близка к нулю, что приводит к большому значению для $T^{1/2}(\hat{\theta} - \theta_0)$. В этом случае асимптотическая теория не дает хорошей аппроксимации поведения оценки θ в малых выборках. Поэтому понятно, что очень маленькое значение ρ_{zx} означает, что свойства оценки выглядят скорей как в случае $\rho_{zx} = 0$, чем как в случае $\rho_{zx} \neq 0$.

Хотелось бы иметь представление о том, что происходит по мере приближения ρ_{zx} к нулю. Данная ситуация аналогична той, которая возникает с тестовыми статистиками, которые всегда отвергают неверную нулевую гипотезу в больших выборках, то есть являются состоятельными. Чтобы сравнивать такие тесты, используется идея локальной альтернативы, то есть находятся распределения тестовых статистик при приближении альтернативной гипотезы к нулевой по мере увеличения размера выборки. В данной ситуации поступают аналогично, а именно, полагают $\pi = \phi/\sqrt{T}$. Тогда

$$\begin{aligned} T^{1/2}\hat{\rho}_{zx} &= T^{1/2}\rho_{zx} + T^{1/2}\eta \\ &= \frac{\phi\sigma_z}{\sigma_x} + T^{1/2}\eta \\ &= \frac{\phi\sigma_z}{\sigma_x} + \varepsilon, \end{aligned}$$

где $\varepsilon \sim N(0, v)$. Проанализируем, что происходит в случае слабых инструментов. Получаем

$$\begin{aligned} \hat{\theta} - \theta_0 &= (T^{1/2}\hat{\rho}_{zx})^{-1}(T^{1/2}\hat{\rho}_{zu}) \cdot \frac{\sigma_u}{\hat{\sigma}_x} \\ &\to^d \frac{N(0,1)}{N(\phi\sigma_z/\sigma_x, v)}\frac{\sigma_u}{\sigma_x}. \end{aligned}$$

Плохая новость в том, что теперь $\hat{\theta} - \theta_0$ – это отношение двух случайных величин, которое не сходится к нулю, то есть $\hat{\theta}$ не является состоятельной оценкой θ_0. Более того, $\hat{\theta} - \theta_0$ не может иметь нормальное распределение. Следовательно, когда ρ_{zx} мало, подобный анализ дает лучшее описание поведения $\hat{\theta} - \theta_0$, чем стандартная асимптотика, что было подтверждено с помощью симуляций. Вероятно, распределение $\hat{\theta} - \theta_0$ в конечных выборках будет сильно отличаться от нормального, когда присутствуют слабые инструменты, а оценки коэффициентов будут существенно смещены.

2 Как обнаружить слабые инструменты

Как обнаружить слабые инструменты? Поскольку проблема возникает, когда ρ_{zx} близко к нулю, логично тестировать гипотезу $\rho_{zx} = 0$, используя оценку $\hat{\rho}_{zx}$. Но поскольку $\hat{\rho}_{zx}$ кратно $\hat{\pi}$, МНК-оценке π в регрессии x_t на z_t, можно вместо этого тестировать гипотезу $\pi = 0$. Для этого подходит либо t-тест, либо F-тест на значимость регрессоров z_t в регрессии для x_t.

Последняя интерпретация становится еще полезней, если отойти от простой модели. Важной модификацией является модель, в которой в оцениваемом уравнении появляются дополнительные регрессоры w_t, наличествующие и в приведенной форме:

$$\begin{aligned} y_t &= x_t\theta + w_t\alpha + u_t, \\ x_t &= z_t\pi + w_t\gamma + \xi_t. \end{aligned}$$

Для анализа этой ситуации запишем уравнения в матричной форме:

$$\begin{aligned} y &= X\theta + W\alpha + u, \\ X &= Z\pi + W\gamma + \xi. \end{aligned}$$

Домножая слева оба уравнения на $M_W = I - W(W'W)^{-1}W'$, получим, что

$$\begin{aligned} M_W y &= M_W X\theta + u^*, \\ M_W X &= M_W Z\pi + \xi^*, \end{aligned}$$

или

$$\begin{aligned} y^* &= X^*\theta + u^*, \\ X^* &= Z^*\pi + \xi^*, \end{aligned}$$

так что вместо y, X и Z остаются y^*, X^*, Z^*. Величины вроде y^* являются остатками регрессии на W, а $\hat{\pi}$ теперь является оценкой π в регрессии X^* на Z^*, что совпадает с оценкой в регрессии X на Z и W. Тогда F-тест на равенство π нулю должен предусматривать присутствие W в регрессии, то есть правильной мерой наличия проблем со слабыми инструментами является корреляция между x_t и z_t *после устранения влияния* w_t. Поскольку F-статистика для $\pi = 0$ в модели без регрессоров равна $(1 - R^2)/R^2$, а при наличии w_t в уравнении в этой формуле R^2 просто заменяется на частный R^2, ясно, что даже когда R^2 большой – x_t хорошо объясняется с помощью z_t и w_t, – частный R^2 может быть очень мал, то есть большая часть x_t объясняется w_t, а не z_t. В этом случае z_t фактически является слабым инструментом. Проблема, конечно, в том, что переменная w_t недоступна в качестве инструмента для x_t, поскольку ее уже «использовали» при оценке α, то есть w_t нужна в качестве инструмента для себя самой. Действительно, большую часть объяснения x_t дает w_t, а не z_t, означая, что R^2 является плохим показателем того, насколько z_t полезны в качестве инструментов.

Staiger & Stock (1997) рекомендуют классифицировать инструменты как слабые, если F-статистика при тестировании $\pi = 0$ в регрессии x_t на z_t и w_t меньше 10.[1] В терминах R^2 это означает, что частный R^2 должен быть больше, чем 0,1. Это разумное практическое правило, которое широко используется. Что именно следует тестировать, когда $\dim(X) > 1$, менее понятно, поскольку в этом случае вектор переменных x_t регрессируется на набор регрессоров z_t. Одно из возможных предложений – использовать канонические корреляции – разработано в Hall, Rudebusch & Wilcox (1996). Stock & Yogo (2005) предложили использовать многомерный аналог коэффициента концентрации и выбирать его наименьшее собственное значение. Авторы приводят таблицу консервативных критических значений для такого теста, поскольку точное распределение найти сложно.

3 Инференция на основе инструментальной оценки

Обратимся теперь к инференции на основе инструментальной оценки. Интерес вызывают два вопроса: во-первых, вывод тестовой статистики для гипотезы $H_0 : \theta = \theta_0$, и во-вторых, способ построения доверительных интервалов. Последний вопрос слишком сложен для рассмотрения в настоящем эссе. Рассмотрим проблему нахождения тестовой статистики, которая давала бы вероятностное значение для тестирования нулевой гипотезы. Обычно для этого используется t-статистика.

Поскольку для оценивания используются моменты

$$\mathbb{E}[m_t] = \mathbb{E}[z_t u_t] = \mathbb{E}[z_t(y_t - x_t\theta)] = 0,$$

из теории метода моментов следует, что дисперсия $T^{1/2}(\hat{\theta} - \theta_0)$ имеет вид

$$V_{\hat{\theta}} = \mathbb{E}\left[\frac{\partial m_t}{\partial \theta}\right]^{-2} \mathbb{V}[m_t] = \sigma_{zx}^{-2}(\sigma_u^2 \sigma_z^2),$$

[1] Как говорится в их статье, это оценка параметра концентрации, основного показателя, который влияет на распределение 2ШМНК-оценки в конечных выборках. Shea (1995) предлагает расширение теста на случай, когда X – многомерная величина, что оказалось полезным, например, в Pagan & Robertson (1996).

поскольку

$$\frac{\partial m_t}{\partial \theta} = -x_t z_t,\ \mathbb{V}[m_t] = \sigma_u^2 \mathbb{E}[z_t^2].$$

При стандартном выводе t-статистики $V_{\hat{\theta}}$ оценивается с помощью $\hat{\sigma}_{zx}^{-2}(\hat{\sigma}_u^2 \hat{\sigma}_z^2)$, то есть

$$\begin{aligned} t_{\hat{\theta}} &= \frac{T^{1/2}(\hat{\theta}-\theta_0)}{\sqrt{\hat{V}_{\hat{\theta}}}} = \frac{T^{1/2}\hat{\sigma}_{zx}^{-1}\hat{\sigma}_{zu}}{\hat{\sigma}_{zx}^{-1}(\hat{\sigma}_u \hat{\sigma}_z)} \\ &= \frac{T^{1/2}\hat{\sigma}_{zu}}{\hat{\sigma}_u \hat{\sigma}_z} \\ &= (T^{1/2}\hat{\rho}_{zu})\frac{\sigma_u}{\hat{\sigma}_u}. \end{aligned}$$

Обычно $T^{1/2}\hat{\rho}_{zu}$ асимптотически $N(0,1)$, так что возможные проблемы с распределением t-статистики возникают из-за множителя $\sigma_u/\hat{\sigma}_u$. В стандартных условиях этот множитель сходится к 1, что приводит к известному результату об асимптотической нормальности $t_{\hat{\theta}}$. Как меняется этот вывод при наличии слабых инструментов? Ответ кроется в равенстве

$$\begin{aligned} \hat{\sigma}_u^2 &= \frac{1}{T}\sum_t (y_t - x_t\hat{\theta})^2 = \frac{1}{T}\sum_t (y_t - x_t\theta_0 - x_t(\hat{\theta}-\theta_0))^2 \\ &= \frac{1}{T}\sum_t (y_t - x_t\theta_0)^2 + (\hat{\theta}-\theta_0)^2\left(\frac{1}{T}\sum_t x_t^2\right) - 2(\hat{\theta}-\theta_0)\left(\frac{1}{T}\sum_t u_t x_t\right). \end{aligned}$$

Асимптотически, первый член – это σ_u^2, а два других члена обычно исчезают, поскольку $\hat{\theta}$ является состоятельной оценкой для θ_0. Но при слабых инструментах («локально нулевой асимптотике») $\hat{\theta}$ несостоятельна, так что $\hat{\sigma}_u$ асимптотически является случайной величиной, что делает распределение t-статистики нестандартным, так как это отношение нормальной случайной величины к распределению $\hat{\sigma}_u$. Конечно, когда вид этого распределения неизвестен, построение доверительного интервала затруднительно.

4 Построение полезных тестовых статистик

Для случая $\dim(X) = 1$ и $\dim(Z) \geq \dim(X)$ предложено много решений. Наиболее старый метод предлагает обходной путь для тестирования гипотезы $H_0 : \theta = \theta^*$ – с помощью решения другой задачи.[2] А именно, при оценивании θ предполагалось, что $\mathbb{E}[z_t u_t] = 0$. Теперь это условие выглядит как $\mathbb{E}[z_t(y_t - x_t\theta_0)] = 0$, где θ_0 – истинное значение θ, так что

$$\mathbb{E}[z_t(y_t - x_t'\theta^*)] = \mathbb{E}[z_t x_t'(\theta_0 - \theta^*)]. \tag{2}$$

Предполагая, что инструменты не являются нерелевантными, можно проверять нулевую гипотезу $\theta_0 = \theta^*$, тестируя $\mathbb{E}[z_t(y_t - x_t\theta^*)] = 0$.

Равенство (2) можно переформулировать как тест на условные моменты в виде

$$\mathbb{E}[z_t(y_t - x_t'\theta^*) - \gamma] = 0,$$

а $\gamma = 0$ можно тестировать, используя $\hat{\gamma} = T^{-1}\sum_{t=1}^T z_t(y_t - x_t'\theta_0)$. Записывая в матричной форме (и отбрасывая T), получаем $Z'(y - X\theta_0)$. Если нулевая гипотеза $H_0 : \theta = \theta_0$ верна и $u_t \sim i.i.d.(0, \sigma_0^2)$, то $\mathbb{V}[Z'(y - X\theta_0)] = \sigma_0^2 Z'Z$, что дает статистику

$$AR = \frac{(y - X\theta_0)'Z(Z'Z)^{-1}Z(y - X\theta_0)}{\tilde{\sigma}_0^2},$$

[2] Важно отметить, что под θ подразумеваются коэффициенты при эндогенных переменных. Если в исходном соотношении имеются экзогенные переменные, они устраняются, как показано в разделе 2, так что y_t, z_t и x_t будут остатками регрессий для исходных переменных.

где $\tilde{\sigma}_0^2 = T^{-1}\sum_t (y_t - x_t'\theta_0)^2$. Это – *тестовая статистика Андерсона–Рубина* (AR). Она имеет асимптотическое распределение $\chi^2(\dim(Z))$ и вполне хорошо ведет себя в конечных выборках. Тот факт, что она имеет распределение $\chi^2(\dim(Z))$, а не $\chi^2(\dim(\theta))$, печален, поскольку $\dim(Z)$ может значительно превосходить $\dim(\theta)$. Если тестируются *все* параметры θ, тест естественным образом расширяется на случай $\dim(\theta) > 1$, но это скорее редкость.

Есть и другие способы тестирования, более непосредственные, чем AR-тест. Выше было описано не что иное, как тест Вальда. Как насчет LM-теста? Тест Вальда и LM-тест идентичны, когда $\dim(Z) = 1$, просто в LM-тесте $\hat{\sigma}^2$ заменяется на $\tilde{\sigma}_0^2$. Отсюда следует, что распределение t-статистики из LM-теста асимптотически стандартное нормальное. Этот изящный результат впервые получили Wang & Zivot (1996). К сожалению, он не выполнен для более интересных моделей, в частности, для случая, когда инструментов больше, чем регрессоров, требующих инструментирования, то есть $\dim(z_t) > \dim(x_t)$. Чтобы понять, почему так происходит, заметим, что 2ШМНК-оценка для θ имеет вид

$$\hat{\theta} - \theta_0 = (\hat{\pi}'Z'Z\hat{\pi})^{-1}(\hat{\pi}'Z'u), \quad \mathrm{sd}(\hat{\theta}) = \hat{\sigma}_u(\hat{\pi}'Z'Z\hat{\pi})^{-1/2},$$

так что

$$t_{\hat{\theta}} = \frac{(\hat{\pi}'Z'Z\hat{\pi})^{-1}(\hat{\pi}'Z'u)}{\hat{\sigma}_u(\hat{\pi}'Z'Z\hat{\pi})^{-1/2}}.$$

В случае $\dim(Z) = \dim(X) = 1$, $\hat{\pi}$ – скалярная величина, которая сокращается в числителе и знаменателе, оставляя нас с

$$t_{\hat{\theta}} = \frac{Z'u}{\hat{\sigma}_u(Z'Z)^{1/2}},$$

и тогда достаточно просто заменить $\hat{\sigma}_u$ на $\tilde{\sigma}_0$, чтобы обеспечить асимптотическую нормальность $t_{\hat{\theta}}$. Но если $\dim(Z) > \dim(X)$, $\hat{\pi}$ не сокращается, и возникает вопрос о распределении $\hat{\pi}$ при локально нулевой асимптотике в случае слабых инструментов

$$\hat{\pi} = \frac{\phi}{\sqrt{T}} + (Z'Z)^{-1}Z'\xi.$$

Чтобы $\hat{\pi}$ не ушло в нуль, необходимо, чтобы $T^{1/2}\hat{\pi}$ было асимптотически

$$N(\phi, \sigma_\xi^2\sigma_z^{-2}).$$

Для анализа последствий таких действий заметим, что

$$t_{\hat{\theta}} = \frac{[(T^{1/2}\hat{\pi}')(T^{-1}Z'Z)(T^{1/2}\hat{\pi})]^{-1}(T^{1/2}\hat{\pi}')(T^{-1/2}Z'u)}{\hat{\sigma}_u[(T^{1/2}\hat{\pi}')(T^{-1}Z'Z)(T^{1/2}\hat{\pi})]^{-1/2}},$$

и, учитывая, что $\hat{\pi}$ и $T^{-1/2}Z'u$ имеют предельные распределения, t-статистика – это произведение (и отношение) множества (асимптотически) нормальных случайных величин. Следовательно, даже при замене $\hat{\theta}$ на θ_0 при оценивании σ_u проблемы, связанные со слабыми инструментами, остаются.

Есть и другие случаи, когда $\hat{\pi}$ исчезает из тестовой статистики. Например, если $\dim(\theta) = \dim(Z)$, $\hat{\pi}$ будет квадратной матрицей и исчезнет из квадратичной формы, приводя к обычной χ^2-статистике. Следовательно, в этом случае, при условии, что $\tilde{\sigma}_0^2$ используется в качестве оценки для σ_u^2, тестовая статистика действительно является $\chi^2(\dim(\theta))$-распределенной случайной величиной. Этот случай интересен, поскольку он возникает в структурных векторных авторегрессиях с долгосрочными ограничениями (см. Pagan & Robertson, 1998), хотя там слабые инструменты возникают из-за близкого к свойствам рядов с единичным корнем поведения x_t, и локально нулевая асимптотика уже неприменима, так как тогда π должно быть порядка δ/T, а не $\delta/\sqrt{T}$.

Возвращаясь к случаю $\dim(\theta) = 1$ и $\dim(Z) > 1$, предположим, что берется другая оценка π. Для ее получения рассмотрим систему

$$\begin{aligned} y_t &= x_t\theta + u_t, \\ x_t &= z_t'\pi + \xi_t. \end{aligned}$$

Если

$$\begin{pmatrix} u_t \\ \xi_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{u\xi} \\ \sigma_{u\xi} & \sigma_\xi^2 \end{pmatrix}\right),$$

то можно записать

$$\xi_t = \frac{\sigma_{u\xi}}{\sigma_u^2} u_t + \eta_t,$$

где η_t не зависит от u_t. Полагая $\theta = \theta_0$, можно оценить u_t, а также $\sigma_{u\xi}/\sigma_u^2$, так что регрессия $x_t - (\hat\sigma_{u\xi}/\hat\sigma_u^2)\hat u_t$ на z_t даст альтернативную оценку для π, скажем, $\tilde\pi$. Оказывается, это оценка ММПОИ при $\theta = \theta_0$. Ясно, что $\tilde\pi$ является функцией от η_t, а значит, не зависит от u_t, откуда следует, что (условно на Z) $Z'u$ не зависит от $\tilde\pi$. Эта независимость означает, что теперь при выводе распределения t-статистики, использующей $\tilde\pi$ вместо $\hat\pi$, его можно рассматривать условно на $\tilde\pi$. Следовательно, t-статистика будет асимптотически нормальна. Замена $\hat\sigma^2$ на σ_0^2 и $\hat\pi$ на $\tilde\pi$ означает применение LM-теста на $\theta = \theta_0$, что и предложил Kleibergen (2002). Эта статистика напрямую обобщается на случай $\dim(\theta) > 1$. Как и в большей части литературы по этой тематике, критическим предположением является экзогенность Z (или возможность преобразований условно на Z), что не выполняется в случае, когда слабость инструментов возникает из-за поведения наподобие рядов с единичным корнем.

Проблемы, вызванные свойствами $\hat\pi$, можно решить другими способами. Одно из предложений – рассматривать распределение условно на какой-нибудь функции от $\hat\pi$, в частности параметре концентрации $\phi = \sigma_\xi^{-2}\pi'Z'Z\pi$. Его оценка $\hat\phi$ – это F-статистика для гипотезы $\pi = 0$. Hillier & Forchini (2004) обсуждают логику таких действий, утверждая, что естественно брать распределение $\hat\theta$ условно на результате теста $\pi = 0$, поскольку он содержит информацию о θ из выборки. Авторы получают моменты этого условного распределения и сравнивают метод с МНК-оцениванием. Если параметр концентрации очень мал, результаты говорят о предпочтительности МНК-оценки. Hillier & Forchini (2004) также выводят распределения для случая $\dim(\theta) > 1$, используя минимальное собственное значение многомерного аналога параметра концентрации. Они отмечают, однако, что эти выражения настолько сложны для расчетов, что лучше выбрать какую-нибудь альтернативную функция от π.

Moreira (2003) рассматривает тест отношения правдоподобия (LR) для гипотезы $\theta = \theta_0$. Если F-статистику в тесте на равенство нулю коэффициентов при z_t в регрессии $x_t - (\hat\sigma_{u\xi}/\hat\sigma_u^2)\hat u_t$ на z_t обозначить за $\tilde r$ (что оценивает модифицированный «параметр концентрации» $r = \sigma_\eta^{-2}\pi'Z'Z\pi$), то, как показывает Kleibergen (2005),

$$LR = \frac{1}{2}[AR - \tilde r + \sqrt{(AR + \tilde r)^2 - 4\tilde r(AR - LM)}].$$

Как было отмечено ранее, AR и $\tilde r$ должны быть независимы, а значит, в этом случае логично искать распределение условно на $\tilde r$. Распределение этого условного LR-теста (CLR) можно найти численно по алгоритму, недавно предложенному в Andrews, Moreira & Stock (2005).

Все внимание в описанной выше литературе уделялось тому, чтобы отыскать тестовую статистику, при определенных условиях независимую от π. Хотелось бы также, чтобы тест был мощным, и этим вопросом в последнее время занимались Poskitt & Skeels (2005) (используя асимптотику малого параметра концентрации) и Andrews, Moreira & Stock (2006). Последние рассматривают построение точечно оптимальных инвариантных тестов. Они также отмечают, что функция мощности для LM-теста немонотонна, и CLR-тест является лучшим из трех перечисленных тестов.

Близкая проблема, которую только недавно начали изучать, – тестирование ограничений на подмножества θ, например, $\theta' = (\theta_1', \theta_2')$, и требуется проверить гипотезу $H_0 : \theta_1 = \theta_{10}$. Если к параметрам θ_2 не относятся проблемы со слабыми инструментами, то результаты, перечисленные выше, выполнены, поскольку инструментальная оценка θ_2 имеет «хорошие» свойства. Если же они также подвержены смещению из-за слабости инструментов, следует внести некоторые поправки. В недавних статьях Dufour & Taamouti (2005) и Kleibergen (2007) исследуется, каким образом это можно сделать; в первой статье используются проекционные методы для получения стандартных тестов, тогда как во второй устанавливаются некоторые ограничения на распределение $\hat{\theta}_1$.

Множество и других вопросов рассматривается в литературе. Один из них касается ОММ-оценивания. ОММ-оценка имеет вид

$$\hat{\theta} - \theta_0 = \left(-\sum_t \frac{\partial m_t}{\partial \theta} \right)^{-1} \sum_t m_t(\theta_0),$$

где обычно предполагается, что $T^{-1}\sum_t \partial m_t/\partial\theta$ сходится к константе. Но может случиться, что $\mathbb{E}[\partial m_t/\partial\theta]$ либо равняется нулю, либо очень мало, и тогда возникают те же проблемы, что и со слабыми инструментами (в простейшем случае с инструментальными переменными $-\partial m_t/\partial\theta = z_t x_t$). Таким образом, может оказаться, что асимптотика для ОММ-оценки не работает. Используя обобщенное информационное неравенство, имеем $\mathbb{E}[-\partial m_t/\partial\theta] = \mathbb{E}[m_t L_{\theta t}]$, где $L_{\theta t}$ – скор-функции для θ, и тогда возникают весьма реальные проблемы с ОММ-оценкой, если выбраны моменты m_t, не коррелированные со скор-функциями, так что следует немного поэкспериментировать, чтобы определить, какие моменты являются подходящими. В такой ситуации полезно иметь теоретическую модель, поскольку для нее можно запустить симуляции при выбранных значениях параметров и изучить вопрос о выборе моментов с помощью численных методов. Несовпадение распределения ОММ-оценки с тем, которое предполагает асимптотическая теория, хорошо изучено, например, в Kocherlakota (1990).

5 Примеры слабых инструментов

Оценивание кривой Филлипса в новокейнсианской модели

Новокейнсианская модель, приобретающая все большую популярность в макроэкономических исследованиях, имеет вид

$$\begin{aligned}
\pi_t &= \delta_1 \mathbb{E}_t[\pi_{t+1}] + \delta_2 \pi_{t-1} + \lambda x_t + \varepsilon_{AS,t},\\
x_t &= \mu \mathbb{E}_t[x_{t+1}] + (1-\mu)x_{t-1} - \phi(r_t - \mathbb{E}_t[\pi_{t+1}]) + \varepsilon_{IS,t},\\
r_t &= \rho r_{t-1} + (1-\rho)(\beta \mathbb{E}_t[\pi_{t+1}] + \gamma x_t) + \varepsilon_{MP,t},
\end{aligned}$$

где π_t – темп инфляции, y_t – разрыв между текущим и потенциальным уровнями выпуска, а r_t – ставка процента, определяемая правилом денежной политики. В модели присутствуют три типа шоков: шок совокупного предложения $\varepsilon_{AS,t}$, шок со стороны кривой IS (шок спроса) $\varepsilon_{IS,t}$, и шок денежной политики $\varepsilon_{MP,t}$. Пусть кривая Филлипса оценивается стандартным образом, то есть предполагаются рациональные ожидания, и $\mathbb{E}_t[\pi_{t+1}]$ заменяется на π_{t+1}. Тогда требуется инструмент для π_{t+1}. Но инструмент также нужен и для x_t, так как это эндогенная переменная. Какие инструменты доступны? Известно, что в решении этой системы (при отсутствии серийной корреляции шоков) π_t, x_t и r_t являются функциями только своих первых лагов. Следовательно, среди доступных инструментов только x_{t-1}, r_{t-1} и π_{t-1}.

Но π_{t-1} уже присутствует в выражении для кривой Филлипса, так что в итоге в качестве инструментов для π_{t+1} остаются только x_{t-1} и r_{t-1}. Как покажет любая регрессия, эти инструменты являются слабыми, то есть фактически оценить δ_1 и δ_2 невозможно. На практике часто предполагают ограничение $\delta_2 = 1 - \delta_1$ («гибридная модель»), что преобразует кривую Филлипса в

$$\Delta\pi_t = \delta_1 \mathbb{E}_t[\pi_{t+1} - \pi_{t-1}] + \lambda x_t + \varepsilon_{AS,t},$$

и тогда π_{t-1} оказывается хорошим инструментом для $\pi_{t+1} - \pi_{t-1}$. Таким образом, иногда проблему слабых инструментов можно обойти, адаптируя модель из некоторых теоретических соображений.

Уравнения Эйлера в задачах управления запасами

Слабые инструменты возникают в различных ситуациях. Иногда выбору инструментов не уделяется должного внимания, о чем свидетельствуют, например, дискуссии об отдаче от образования, когда выводы очень чувствительны к выбору инструментов. В других ситуациях сложности могут возникать либо из-за природы используемой модели, либо из-за взаимодействия свойств модели с данными. Возможным примером последней причины может служить «тест на общие факторы» Vahid & Engle (1993), когда необходимы инструменты для темпов роста выпуска и потребления, а найти корреляцию между этими темпами роста удается редко.[3] В некоторых случаях возможно определить источник слабого инструмента. Рассмотрим такой пример.

Параметры, возникающие в системе условий первого порядка, связанной с уравнениями Эйлера, определяющими оптимальный выбор переменных управления, часто оценивают с помощью ОММ, что может привести к проблеме слабых инструментов. Во многих ситуациях эта возможность становится действительностью. Gregory, Pagan & Smith (1993) показали, что при использовании уравнений Эйлера из линейно-квадратичной оптимизационной модели инструменты нерелевантны при оценки коэффициента дисконтирования, если переменные роста являются интегрированными первого порядка. Похожий пример приводится в статье Fuhrer, Moore & Schuh (1995). Авторы обнаружили крайне плохие результаты применения ОММ-оценки в определенных условиях, даже при числе наблюдений, измеряемом тысячами.

Fuhrer, Moore & Schuh (1995) получают условия первого порядка, описывающие оптимальный выбор запасов в линейно-квадратичной задаче минимизации

$$\sum_{j=0}^{\infty} \beta^j \mathbb{E}_t[C_Y(Y_{t+j}) + C_N(t_{t+j}, S_{t+j})]$$

при ограничении $N_{t+j} = N_{t+j-1} + Y_{t+j} - S_{t+j}$, где Y_t – выпуск, S_t – продажи, N_t – уровень запасов, и

$$\begin{aligned} C_Y(Y_{t+j}) &= (\delta/2)Y_{t+j}^2 + (\alpha/2)(\Delta Y_{t+j})^2, \\ C_N(N_{t+j}, S_{t+j}) &= (\phi/2)(N_{t+j} - \omega S_{t+j})^2. \end{aligned}$$

Показатели ОММ настолько плохи, что авторы рекомендуют использовать метод максимального правдоподобия, даже если используется неверная плотность (эта оговорка возникает из-за необходимости специфицировать форму процесса для продаж при выписывании

[3]Это, по существу, тест на серийную корреляцию после инструментирования, и, следовательно, распределение статистики зависит от других оцениваемых параметров. Поскольку инструментальная оценка параметров линейной модели подвержена влиянию слабых инструментов, это справедливо и для распределения соответствующей статистики. Этот вопрос подробно не рассматривается в литературе, но получено, что тесты на серийную корреляцию могут давать странные результаты, если они основаны на слабых инструментах.

функции правдоподобия). Имеет смысл задуматься об источнике плохих показателей ОММ-оценки в модели Fuhrer, Moore & Schuh (1995), так как понимание этих причин важно для оценки подобных рекомендаций.

Уравнение Эйлера для приведенной оптимизационной задачи выглядит как

$$\mathbb{E}_t \left[\delta \left(Y_t - \beta Y_{t+1}\right) + \alpha(\Delta Y_t - 2\beta \Delta Y_{t+1} + \beta \Delta Y_{t+2}) + \phi \left(N_t - \omega S_t\right)\right] = 0, \tag{3}$$

что дает систему условий на моменты

$$\mathbb{E} \left[z_t(\delta \left(Y_t - \beta Y_{t+1}\right) + \alpha(\Delta Y_t - 2\beta \Delta Y_{t+1} + \beta \Delta Y_{t+2}) + \phi \left(N_t - \omega S_t\right))\right] = 0, \tag{4}$$

где z_t – инструменты из информационного множества, используемые в условном ожидании. Из (4) ясно, что невозможно идентифицировать все входящие параметры, и исследователи прибегают к различным вариантам нормализации. Fuhrer, Moore & Schuh (1995) работают с пятью типами нормализации. Из них два, обозначим их A и B, имеют следующий вид: $A : \delta = 1$, $B : \phi = 1$. Для анализа этих различных условий используем два упрощающих предположения. Во-первых, предположим, что $S_t = S_{t-1} + e_t$, где $e_t \sim i.i.d.(0, \sigma^2)$, то есть продажи являются случайным блужданием. Во-вторых, пусть $\beta = 1$.[4] Когда S_t является $I(1)$-рядом, можно показать, что N_t также является $I(1)$, а ω – параметр коинтеграции. При этих предположениях нормализация типа A предполагает линейную модель, в которой ΔY_{t+1} является зависимой переменной, а $N_t - \omega S_t$ и $\Delta^2 Y_{t+2}$ – регрессорами, для последнего из которых нужны инструменты. При нормализации типа B, $N_t - \omega S_t$ является независимой переменной, а ΔY_{t+1} и $\Delta^2 Y_{t+2}$ – регрессоры, и для обоих из них требуются инструменты. В качестве инструментов обычно применяются приращения Y_t и S_t, поскольку переменные в уровнях являются плохими инструментами для ΔY_{t+1}, так как в этом случае $I(1)$-переменные берутся как инструменты для $I(0)$ переменных. $\Delta Y_{t+1} = \Delta S_{t+1} + \Delta^2 N_{t+1}$ и, следовательно, чтобы лаги ΔY_t были хорошими инструментами, необходимо наличие серийной корреляции в $\Delta^2 N_t$. Если вывести выражение для процесса N_t при предложенной спецификации для S_t, можно установить, что серийная корреляция в $\Delta^2 N_t$ очень мала. Это означает, что такие нормализации, как B, приводят к проблеме слабых инструментов. Следовательно, ясен источник вывода авторов о том, что (стр. 143) «в случае сглаживающей модели, нормализация B требует 30000 наблюдений для сходимости к истинному значению».

Этот пример показывает, почему так сложно выдвинуть общий подход к проблеме слабых инструментов. Если бы ряд S_t не был близок к ряду с единичным корнем, инструменты могли бы быть очень эффективными, так что исход сильно зависит от природы соответствующих процессов.[5]

Список литературы

Andrews, D.W.K., M.J. Moreira & J.H. Stock (2006). Optimal two-sided invariant similar tests for instrumental variables regression. *Econometrica* 74, 715–752.

Andrews, D.W.K., M.J. Moreira & J.H. Stock (2005). Performance of conditional Wald tests in IV regression with weak instruments. *Journal of Econometrics*, в печати.

Dufour, J.-M. & M. Taamouti (2005). Projection-based statistical inference in linear statistical models with possibly weak instruments. *Econometrica* 73, 1351–1365.

Forchini, G. & G. Hillier (2003). Conditional inference for possibly unidentified structural equations. *Econometric Theory* 19, 707–743.

[4] В исходном анализе данных и симуляциях $\beta = 0{,}995$, и процесс для продаж по существу содержит единичный корень при слабой автокорреляции. Fuhrer, Moore & Schuh (1995) отмечают, что если процесс для продаж неверно специфицирован как AR(1)-процесс вместо AR(3), «оцененный лаговый коэффициент в AR(1) будет приблизительно равен сумме коэффициентов для AR(3)-процесса» (стр. 143). Это дает корень, равный 0,956.

[5] Тот же вывод имеет место для примера из статьи Gregory, Pagan & Smith (1993).

Fuhrer, J., G. Moore & S. Schuh (1995). Estimating the linear-quadratic inventory model: Maximum likelihood versus generalized method of moments. *Journal of Monetary Economics* 35, 115–158.

Gregory, A.W., A.R. Pagan & G.W. Smith (1993). Estimating linear quadratic models with integrated processes. Глава в *Models, Methods and Applications of Econometrics: Essays in Honor of A.R. Bergstrom* под редакцией P.C.B. Phillips. Cambridge: Basil Blackwell.

Hall, A.R., G.D. Rudebusch & D.W. Wilcox (1996). Judging instrument relevance in instrumental variables estimation. *International Economic Review* 37, 283–298.

Kleibergen, F. (2002). Pivotal statistics for testing parameters in instrumental variables regression. *Econometrica* 70, 1781–1803.

Kleibergen, F. (2006). Testing. Глава в *The New Palgrave Dictionary of Economics*, под редакцией S. Durlauf & L. Blume. Palgrave Macmillan, в печати.

Kleibergen, F. (2007). Subset statistics in the linear IV regression model. Manuscript, Brown University.

Kocherlakota, N.R. (1990). On tests of representative consumer asset pricing models. *Journal of Monetary Economics* 26, 285–304.

Moreira, M.J. (2003). A conditional likelihood ratio test for structural models. *Econometrica* 71, 1027–1048.

Nelson, C.R. & R. Startz (1990). The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one. *Journal of Business* 63, 125–140.

Pagan, A.R. & J.C. Robertson (1998). Structural models of the liquidity effect. *Review of Economics and Statistics* 80, 202–217.

Poskitt, D.S. & C.L. Skeels (2005). Small concentration asymptotics and instrumental variables inference. Research Paper No.948, University of Melbourne.

Shea, J. (1997). Instrument relevance in multivariate linear models: A simple measure. *Review of Economics and Statistics* 79, 348–352.

Staiger, D. & J.H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica* 65, 557–586.

Stock, J.H. & J. Wright (2000). GMM with weak identification. *Econometrica* 68, 1055–1096.

Stock, J.H. & M. Yogo (2005). Testing for weak instruments in linear IV regression. Глава в *Identification and Inference for Econometric Models: A Festshrift in Honor of Thomas Rothenberg* под редакцией J.H. Stock & D.W.K. Andrews. Cambridge: Cambridge University Press.

Stock, J.H., J.H. Wright & M. Yogo (2002). A survey of weak instruments and weak identification in GMM. *Journal of Business & Economic Statistics* 20, 518–529.

Vahid, F. & R. Engle (1993). Common trends and common cycles. *Journal of Applied Econometrics* 8, 341–360.

Wang, J. & E. Zivot (1996). Inference on a structural parameter in instrumental variables regression with weakly correlated instruments. *Econometrica* 66, 1389–1404.

West, K.D. & D.W. Wilcox (1996). A comparison of alternative instrumental variable estimators of a dynamic linear model. *Journal of Business & Economic Statistics* 14, 281–293.

Weak instruments

Adrian Pagan

Queensland University of Technology, Brisbane, Australia

This essay is a selective guide to the literature on weak instruments. We use simple models to illustrate the problems raised by weak instruments and the suggestions that have been made to solve them. Because the literature is one that is still developing we only briefly touch on some recent developments.