АСИМПТОТИЧЕСКИЕ ПРИБЛИЖЕНИЯ В СОВРЕМЕННОЙ ЭКОНОМЕТРИКЕ (ОБЗОР)

© 2005 г. С. А. Анатольев

(Москва)

Дается обзор асимптотических методов построения приближений в эконометрике. Рассмотрены как стандартные подходы, имеющие давнюю историю в статистике, так и альтернативные, возникшие за последние годы. Статья ориентирована на экономистов-практиков, использующих эконометрические методы в своих исследованиях и знакомых с основами регрессионного анализа.

1. ВВЕДЕНИЕ

Экономист, определяющий степень статистической значимости коэффициентов регрессии путем сравнения t-статистик по модулю с числом 1.96, пользуется асимптотической теорией. Зачастую он не осознает этого, проделывая по инерции стандартную процедуру; при этом получаемые статистические выводы не оспариваются ни самим автором, ни потенциальной аудиторией. Неявное же использование асимптотической теории – в данном случае приближения истинного распределения t-статистик стандартным нормальным распределением – должно вызывать вопросы типа: а почему используется приближение? а хорошее ли приближение? а не существует ли лучшего приближения?

Ответ на первый вопрос тривиальный: истинное распределение данных наверняка неизвестно, да и рассматриваемая статистика, несомненно, является слишком сложной функцией от данных, поэтому думать о получении ее истинного распределения нереально. Ответы на другие вопросы способны поставить под сомнение используемые стандартные алгоритмы построения выводов. Цель этой статьи – помочь прикладному экономисту разобраться в теории асимптотических приближений и улучшить качество его эконометрической практики.

Современную асимптотическую теорию можно довольно четко разделить на "грубую", прямо апеллирующую к теории вероятностей для вывода распределений нужных статистик, и "утонченную", дополнительно включающую в себя приемы и хитрости, призванные решить определенные проблемы, с которыми "грубая" теория не справляется. "Грубая" теория в свою очередь делится на приводящую к стандартным скоростям сходимости и использованию нормального и χ^2-распределений (*стандартная асимптотика*) и на приводящую к нестандартным скоростям сходимости и появлению распределений, не настолько приятных и распространенных (*нестандартная асимптотика*).

"Утонченная" же теория служит либо для улучшения качества асимптотических приближений, либо для получения возможности использовать распределения, известные с точностью до неизвестных параметров. "Утонченными" приемами являются более точные асимптотические разложения (*асимптотика высоких порядков*) и изменение концепции асимптотической сходимости, а именно, добавление к понятию большой выборки еще и понятия дрейфующих параметров (*альтернативная асимптотика*). Каждому из четырех упомянутых направлений в статье будет посвящен отдельный раздел.

При упоминании конкретных моделей будем придерживаться самых простейших их версий, без излишних деталей моделирования и оценивания. Заинтересованный читатель может обратиться к соответствующей литературе. К сожалению, не для всех затрагиваемых вопросов существуют хорошие обзоры даже в англоязычной литературе, но в тех случаях, когда они существуют, мы даем ссылки преимущественно на такие источники, нежели на оригинальные статьи, где в первый раз встретились обсуждаемые инновации.

Введем обозначения. Как принято в эконометрике, все векторы, например x_i и β, являются векторами-столбцами. Математическое ожидание обозначается через $E[\cdot]$, дисперсия – $V[\cdot]$, ковариация – $C[\cdot,\cdot]$. Символы "$\xrightarrow{p}$" и "$\xrightarrow{d}$" указывают на сходимость по вероятности и по распределению соответственно. Размер выборки – n, будь это случайная выборка или временной ряд.

Пивотальной является статистика, распределение которой не зависит от неизвестных параметров; *пивотизация* (или *стандартизация*) – преобразование статистики с целью сделать ее пивотальной. Термин *асимптотика* обозначает весь спектр результатов асимптотических теорий и практических приемов, соответствующих данной задаче.

2. СТАНДАРТНАЯ АСИМПТОТИКА

При стандартной асимптотике чаще всего приходится иметь дело с нормальным распределением для оценок и χ^2-распределением для тестов, а также со скоростью сходимости $\sqrt{n}$. Поскольку упомянутые асимптотические распределения, затабулированные во многих учебниках, являются одинаковыми по форме и различаются только по нескольким характеристикам (асимптотическая дисперсия в случае асимптотического нормального распределения), имеется возможность сравнивать оценки между собой, в результате чего появляется понятие асимптотической эффективности.

Подобная ситуация типична для гладких параметрических моделей, сопровождаемых случайными выборками или стационарными временными рядами. Скорость сходимости $\sqrt{n}$ – это следствие того, что дисперсия среднего для выборочных средних обратно пропорциональна количеству усредняемых элементов, если эти элементы независимы (или слабо зависимы) и одинаково распределены. Сходимость к теоретическим средним и асимптотическая нормальность обеспечиваются законами больших чисел (ЗБЧ) и центральными предельными теоремами (ЦПТ). Эти предельные теоремы существуют в нескольких форматах: в вариантах сильной и слабой сходимости, для последовательностей независимых и одинаково распределенных величин, для последовательностей независимых, но не однородно распределенных величин, и для стационарных временных рядов (см. (Гнеденко, 1988; Ибрагимов, Линник, 1965)).

Проиллюстрируем оценивание параметров линейной регрессии

$$y_t = x_t'\beta + \varepsilon_t, \quad E[\varepsilon_t|x_t] = 0$$

методом наименьших квадратов (МНК). МНК-оценка

$$\hat\beta = \left(\frac{1}{n}\sum_{t=1}^{n} x_t x_t'\right)^{-1} \frac{1}{n}\sum_{t=1}^{n} x_t y_t$$

имеет предел по вероятности

$$\hat\beta = \beta + \left(\frac{1}{n}\sum_{t=1}^{n} x_t x_t'\right)^{-1} \frac{1}{n}\sum_{t=1}^{n} x_t \varepsilon_t \xrightarrow{p} \beta + (E[x_t x_t'])^{-1} E[x_t\varepsilon_t] = \beta$$

при $n \to \infty$, где дважды был использован ЗБЧ для последовательностей сумм случайных величин, и асимптотическое распределение

$$\sqrt{n}(\hat\beta - \beta) = \left(\frac{1}{n}\sum_{t=1}^{n} x_t x_t'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t \varepsilon_t \xrightarrow{d} N(0, V_{\hat\beta})$$

при $n \to \infty$, где

$$V_{\hat\beta} \equiv (E[x_t x_t'])^{-1} \sum_{j=-\infty}^{+\infty} E[x_t x_{t-j}' \varepsilon_t \varepsilon_{t-j}] (E[x_t x_t'])^{-1}$$

и где были использованы ЗБЧ и ЦПТ для последовательностей сумм случайных величин (Hamilton, 1994, гл. 10). Остается только пивотизировать нормализованную разность между оценкой и параметром, чтобы вновь созданная статистика не имела неизвестных параметров. Это достигается (легко – для случайной выборки, труднее – для временного ряда) построением состоятельной оценки $\hat V_{\hat\beta}$ для $V_{\hat\beta}$. Тогда t-статистика для скалярного β определяется формулой

$$t_{\hat\beta} = \sqrt{n}(\hat\beta - \beta)/\sqrt{\hat V_{\hat\beta}} \xrightarrow{d} N(0, 1),$$

а Вальдовская *статистика* для векторного β –

$$W_\beta = n(\hat\beta - \beta)'\hat V_\beta^{-1}(\hat\beta - \beta) \xrightarrow{d} \chi^2_{\dim\beta}$$

(например, (Greene, 2000, гл. 9)). Теперь, когда предельные распределения не содержат неизвестных величин, можно приступать к проверке гипотез и построению доверительных интервалов.

Выше было упомянуто, что для случайной выборки состоятельную оценку V_β построить легко. Действительно, оценка

$$\hat V_\beta = n\left(\sum_{t=1}^{n} x_t x_t'\right)^{-1}\sum_{t=1}^{n} x_t x_t'\hat\varepsilon_t^2\left(\sum_{t=1}^{n} x_t x_t'\right)^{-1},$$

где $\hat\varepsilon_t$ – МНК-остатки, является состоятельной ввиду отсутствия бесконечной суммы в формуле для V_β, когда данные серийно независимы (см. (Hamilton, 1994, гл. 8)). В моделях на временных рядах, когда бесконечная сумма присутствует, состоятельно асимптотическую дисперсионную матрицу можно оценить методами, предложенными в (Andrews, 1991), например, по формуле Ньюи–Уэста (Newey, West, 1987; Hamilton, 1994, гл. 10):

$$\hat V_\beta = n\left(\sum_{t=1}^{n} x_t x_t'\right)^{-1}\sum_{j=-m}^{+m}\left(1 - \frac{|j|}{m+1}\right)\sum_{t=\max(1,1+j)}^{\min(n,n+j)} x_t x_{t-j}'\hat\varepsilon_t\hat\varepsilon_{t-j}\left(\sum_{t=1}^{n} x_t x_t'\right)^{-1},$$

где вспомогательный целочисленный параметр m растет, но медленнее, чем n. По построению такая оценка положительно определена благодаря задействованному взвешиванию. Выбор m является дополнительной трудностью. Теоретически оптимальной скоростью роста является $\sqrt[3]{n}$, а на практике хорошо работает правило выбора $m = [4\sqrt[3]{n/100}]$.

В непараметрических моделях, где не делаются предположения о функциональной форме регрессионой зависимости, скорость сходимости ниже по сравнению с параметрическими задачами и регулируется скоростью роста или падения вспомогательного *сглаживающего параметра*. Например, при оценивании значения регрессии $g(x) \equiv E[y|x]$ на скалярном x с помощью ядерных функций, где сглаживающим параметром является *ширина окна* h_n, асимптотическое распределение оценки при благоприятных условиях имеет вид:

$$\sqrt{nh_n}(\hat g(x) - g(x)) \xrightarrow{d} N(B_{\hat g(x)}, V_{\hat g(x)})$$

при $n \to \infty$, $h_n \to 0$, $nh_n \to \infty$. Поскольку h_n асимптотически падает, "непараметрическая" скорость сходимости $\sqrt{nh_n}$ меньше "параметрической" $\sqrt{n}$, что является платой за непараметрический подход (Hardle, Linton, 1994). Кроме того, видно, что асимптотическое нормальное распределение может быть смещено, что не характерно для параметрических задач. Дальнейшее продвижение в теории непараметрической регрессии состоит в поиске "оптимальной" скорости асимптотического падения h_n, которая зависит от множества факторов, и в практической реализации этой оптимальности (подробнее см. в (Hardle, Linton, 1994)).

В задачах "промежуточного" типа (между параметрическими и непараметрическими моделями), в так называемых *полупараметрических* моделях, скорость сходимости обычно меньше, чем $\sqrt{n}$. Например, известно, что оценка "максимального Скор" (maximum score) в модели бинарного выбора (Manski, 1985) состоятельна, но имеет скорость сходимости $\sqrt[3]{n}$ и сложное асимптотическое распределение. Таким образом, в непараметрических и полупараметрических задачах царят свои "стандарты".

Что касается анализа панельных данных, здесь простор для асимптотической теории еще шире, ибо данные обладают не только "шириной" – числом объектов (фирм, потребителей, и т.д.) n, но и "длиной" – их временной протяженностью n. Следовательно, возможно использовать асимптотические приближения либо при $n \to \infty$ и фиксированном T, либо при $T \to \infty$ и фиксированном n, либо при одновременном $n \to \infty$ и $T \to \infty$. Наиболее популярен, впрочем, асимптотический анализ при $n \to \infty$ и фиксированном T, мотивированный тем, что обычно панельные данные "короткие", т.е. их временная протяженность маленькая, но содержат большое число объектов (Baltagi, 2001). В последнее время, впрочем, усилился интерес к "длинным" панелям и асимптотическим выводам при одновременном $n \longrightarrow \infty$ и $T \longrightarrow \infty$ (например (Hahn, Newey, 2004)).

3. НЕСТАНДАРТНАЯ АСИМПТОТИКА

Когда данные имеют детерминистические или стохастические тренды или когда модели не являются гладкими, асимптотические распределения могут отличаться от нормального или χ^2-распределения. В таком случае говорят о *нестандартной асимптотике*. Неприятности, связанные с нестандартной асимптотикой, следующие. Во-первых, такие распределения могут не иметь возможности быть затабулированными, что приводит к необходимости самому исследователю находить нужные критические значения с помощью имитаций. Во-вторых, в асимптотическом распределении могут "сидеть" неизвестные параметры, да так "глубоко", что не существует пивотизации, которая бы смогла стандартизовать предельное распределение. В таких случаях приходится прибегать к бутстрапу. В-третьих, различия в форме распределений для разных оценок приводят к трудностям сравнения альтернативных оценок одного и того же параметра, и становится проблематичным формулирование концепции асимптотической эффективности.

Самый известный случай появления нестандартной асимптотики – тестирование на единичные корни. Для простейшей авторегрессионной модели без свободного члена $y_t = \rho y_{t-1} + \varepsilon_t$, где ε_t – белый шум, при нулевой гипотезе $H_0 : \rho = 1$ МНК-оценка $\hat{\rho}$ является сверхсостоятельной (т.е. со скоростью сходимости, превышающей $\sqrt{n}$) и имеющей асимптотическое *распределение Дикки–Фуллера*

$$n(\hat{\rho} - 1) \xrightarrow{d} \frac{\int_0^1 B(r)dB(r)}{\int_0^1 B(r)^2 dr} \equiv DF,$$

где $B(r)$ – стандартное броуновское движение на отрезке [0, 1] (Hamilton, 1994, гл. 17). Причина такого нестандартного поведения оценки кроется в наличии стохастического тренда (т.е. единичных корней) у случайного блуждания, которым является y_t, когда нулевая гипотеза верна. Как видно, асимптотическое распределение является пивотальным, так что можно сразу приступать к статистическим выводам, отыскав затабулированное распределение Дикки–Фуллера. Интересно, что пивотальность асимптотического распределения сохраняется после введения в модель свободного члена и детерминистических трендов, а при увеличении порядка авторегресси пивотизация достигается просто. Инструментарием для нахождения вида распределений в моделях со стохастическими трендами является функциональная центральная предельная теорема (ФЦПТ) (см. (Hamilton, 1994, гл. 17)).

Вникнув в принцип построения МНК-оценки, несложно понять причину ее сверхсостоятельности. Разница $\hat{\rho} - 1$ – это дробь, числитель и знаменатель которой есть средние случайных величин с растущей дисперсией, причем рост дисперсии в знаменателе превышает рост в числителе. Напомним, что в стандартном случае, когда данные стационарные, скорость роста дисперсии в числителе и знаменателе одинакова.

Скорость сходимости изменяется, даже если данные содержат детерминистические тренды. Например, в модели $y_t = \mu + \lambda t + \varepsilon_t$, где ε_t – белый шум с дисперсией σ^2, МНК-оценка $\hat{\lambda}$ также является сверхсостоятельной, хотя асимптотическое распределение и является нормальным $\sqrt{n^3}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0, 4\sigma^2)$.

К счастью, при стандартной пивотизации t-статистика имеет стандартное нормальное распределение, т.е. нестандартность скорости сходимости можно игнорировать, если требуется построение обычных статистических выводов (см. (Hamilton, 1994, гл. 16)).

Стандартного поведения t-статистики иногда удается достичь при казалось бы совсем нестандартном асимптотическом поведении оценок. Например, при тестировании на коинтеграцию между скалярными интегрированными рядами y_t и x_t при строгой экзогенности последнего

МНК-оценка от регрессии y_t на x_t сверхсостоятельна и асимптотически распределена как бесконечная смесь нормальных распределений:

$$n(\hat{\beta}-\beta) \xrightarrow{d} \left(\int_0^1 B_x(r)dB_e(r)\right) / \int_0^1 B_x(r)^2 dr,$$

где $(B_x(r), B_e(r))$ – двумерное броуновское движение на квадрате $[0, 1] \times [0, 1]$, первая координата которого ассоциирована с рядом x_t, а вторая – с ошибкой e_t регрессии y_t на x_t. Несмотря на сложность асимптотического распределения, и здесь при стандартной пивотизации t-статистика имеет стандартное нормальное распределение! Значит, проверять гипотезы и строить доверительные интервалы можно, не прибегая к нестандартным распределениям.

Совсем иная ситуация складывается, когда регрессионная функция недифференцируема по интересующему исследователя параметру. Рассмотрим пороговую авторегрессию, призванную моделировать бизнес-циклы (см. подробности в (Franses, van Dijk, 2000, гл. 3)):

$$y_t = \begin{cases} \mu_1 + \rho_1 y_{t-1} + \varepsilon_t, & \text{если} \quad y_{t-1} < \gamma, \\ \mu_2 + \rho_2 y_{t-1} + \varepsilon_t, & \text{если} \quad y_{t-1} \geq \gamma, \end{cases}$$

где ε_t – белый шум с дисперсией σ^2, а параметры авторегрессии меняются в зависимости от режима, т.е. того, превышает ли пороговая переменная y_{t-1} порог γ или нет. Оценки нелинейного метода наименьших квадратов (НМНК) для параметров μ_1, ρ_1, μ_2, ρ_2, а также стандартным образом построенные t-статистики имеют стандартную асимптотику, в то время как НМНК-оценка $\hat{\gamma}$ порога γ сверхсостоятельна и имеет нестандартное асимптотическое распределение (Hansen, 1997):

$$n\delta_n(\hat{\gamma}-\gamma) \xrightarrow{d} \sigma^2 \arg \max_{-\infty<r<\infty} \left[B(r) - \frac{1}{2}|r|\right]$$

при $n \longrightarrow \infty$, $\delta_n \longrightarrow 0$, где $B(r)$ – двустороннее броуновское движение на $(-\infty, \infty)$. Здесь δ_n есть функция от определенных моментов y_t, которые можно состоятельно оценить, и от разницы между параметрами авторегрессии в разных режимах. Почему асимптотически δ_n меняется, хотя зависит от постоянных характеристик, будет рассказано в разд. 5. Интересно, что функцию распределения для асимптотического распределения в этом примере можно выписать аналитически (см. (Hansen, 1997); там же это распределение затабулировано). Аналогичная ситуация имеет место в моделях со структурными сдвигами, произошедшими в неизвестные моменты времени при оценивании этих моментов (Andrews, 1993).

Наконец, в последнее время, в связи с повышенным интересом к нелинейным моделям, нестандартная асимптотика часто встречается в контексте тестов, когда при нулевой гипотезе некоторые нетестируемые параметры не являются идентифицируемыми (Hansen, 1996). Рассмотрим пример все той же пороговой авторегрессии. Положим, нам интересна гипотеза о линейности авторегрессии, т.е. что в обоих режимах система ведет себя одинаково. Такая гипотеза формализуется как H_0: $\mu_1 = \mu_2$, $\rho_1 = \rho_2$. Оказывается, что стандартная статистика отношения правдоподобия $LR = n(\bar{\sigma}^2 - \hat{\sigma}^2)/\hat{\sigma}^2$, где $\bar{\sigma}^2$ – средний квадрат остатков от оценивания линейной авторегрессии, а $\hat{\sigma}^2$ – средний квадрат остатков от оценивания пороговой авторегрессии, асимптотически распределена *не как* χ^2. Причина этого в том, что при нулевой гипотезе (т.е. линейности модели) параметр порога не идентифицируется (т.е. любые значения γ одинаково согласуются с данными). Выходом из положения является модификация стандартных статистик путем взятия определенного функционала от их значений при варьировании значений неидентифицируемого параметра; при этом чаще всего используется супремум-функционал. В данном контексте тестовая статистика имеет вид

$$\sup LR = \sup_{\gamma}(n(\tilde{\sigma}^2 - \hat{\sigma}^2(\gamma))/\hat{\sigma}^2(\gamma)) = n(\bar{\sigma}^2 - \hat{\sigma}^2)/\hat{\sigma}^2,$$

где $\bar{\sigma}^2(\gamma)$ – средний квадрат остатков от оценивания пороговой авторегрессии с величиной порога, равной фиксированному γ. В данном случае из-за монотонности статистики отношения правдоподобия для подсчета $\sup LR$-статистики достаточно двух прогонов алгоритма оценивания, но в других случаях приходится честно прогнать процедуру оценивания много раз на сетке для неидентифицируемого параметра. Асимптотическое распределение нестандартно, несмотря на его стандартность при любом фиксированном значении неидентифицируемого параметра, и может быть непивотальным. В работе (Hansen, 1996) предлагается бутстраповский алгоритм для преодоления этой трудности. Аналогичная ситуация имеет место при тестировании на отсутствие структурных сдвигов (Andrews, 1993).

4. АСИМПТОТИКА ВЫСОКИХ ПОРЯДКОВ

Для сложных статистик при наличии сильных нелинейностей или при работе с маленькими выборками приближение, полученное с помощью стандартных асимптотических выводов, может быть довольно неточным. В этом случае можно попробовать улучшить приближение с помощью асимптотики высоких порядков. Если стандартная асимптотика апеллирует к стохастическим разложениям первого порядка, асимптотика высоких порядков соответственно апеллирует к последующим членам таких разложений. Проиллюстрируем метод для линейной модели с единственной переменной в правой части $y_t = x_t\beta + \varepsilon_t$, где ε_t – ошибка, нескоррелированная с переменной x_t, т.е. $E[x_t\varepsilon_t] = 0$, и все переменные стационарны. Рассмотрим МНК-оценку и ее стандартную асимптотику:

$$\hat{\beta} = \beta + \frac{1}{\sqrt{n}}\frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t\varepsilon_t}{\frac{1}{n}\sum_{t=1}^{n} x_t^2} = \beta + \frac{1}{\sqrt{n}}\frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t\varepsilon_t}{E[x_t^2]} + o_p\left(\frac{1}{\sqrt{n}}\right),$$

где $o_p(\cdot)$ обозначает слагаемое большего порядка малости, чем аргумент, в стохастическом смысле. Поскольку согласно ЦПТ

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t\varepsilon_t = \xi + o_p(1), \quad \text{где} \quad \xi \sim N(0, E[x_t^2\varepsilon_t^2]),$$

то в результате имеем:

$$\hat{\beta} = \beta + \frac{1}{\sqrt{n}}\frac{\xi + o_p(1)}{E[x_t^2]} + o_p\left(\frac{1}{\sqrt{n}}\right) = \beta + \frac{1}{\sqrt{n}}\xi + o_p\left(\frac{1}{\sqrt{n}}\right), \quad \text{где} \quad \zeta \sim N\left(0, \frac{E[x_t^2\varepsilon_t^2]}{(E[x_t^2])^2}\right),$$

что с точностью до обозначений совпадает с нашими представлениями о стандартной асимптотике. В частности, из этого следует, что

$$E[\hat{\beta}] = \beta + o\left(\frac{1}{\sqrt{n}}\right)$$

или, как говорят, МНК-оценка *асимптотически первого порядка не смещена*.

Попробуем продолжить разложение:

$$\hat{\beta} = \beta + \frac{1}{\sqrt{n}}\frac{1}{E[x_t^2]}\frac{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t\varepsilon_t}{1 + \frac{1}{n}\sum_{t=1}^{n}\frac{x_t^2 - E[x_t^2]}{E[x_t^2]}} =$$

$$= \beta + \frac{1}{\sqrt{n}}\frac{1}{E[x_t^2]}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t\varepsilon_t\left(1 - \frac{1}{\sqrt{n}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{x_t^2 - E[x_t^2]}{E[x_t^2]} + o_p\left(\frac{1}{\sqrt{n}}\right)\right) =$$

$$= \beta + \frac{1}{\sqrt{n}}\frac{1}{E[x_t^2]}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t\varepsilon_t - \frac{1}{n}\frac{1}{E[x_t^2]}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} x_t\varepsilon_t\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{x_t^2 - E[x_t^2]}{E[x_t^2]} + o_p\left(\frac{1}{n}\right) =$$

$$= \beta + \frac{1}{\sqrt{n}}\zeta + \frac{1}{n}\zeta\psi + o_p\left(\frac{1}{n}\right).$$

Теперь мы имеем, кроме слагаемых нулевого (β) и первого ($\zeta/\sqrt{n}$) порядков, слагаемое второго порядка ($\zeta\psi/n$) наряду с остаточным членом, представляющим еще более высокие порядки. Таким образом, асимптотическое нормальное распределение для МНК-оценки может быть уточнено путем анализа слагаемого второго порядка. Например, полезным упражнением является подсчет *смещения второго порядка*:

$$E[\hat{\beta}] = \beta + \frac{1}{n}E[\zeta\psi] + o\left(\frac{1}{n}\right) = \beta - \frac{1}{n}\frac{1}{(E[x_t^2])^2}\sum_{j=-\infty}^{+\infty} E[x_t\varepsilon_t x_{t-j}^2] + o_p\left(\frac{1}{n}\right).$$

Выражение для смещения существенно упрощается, если выборка случайная; тогда

$$E[\hat{\beta}] = \beta - \frac{1}{n}\frac{E[x_t^3\varepsilon_t]}{(E[x_t^2])^2} + o_p\left(\frac{1}{n}\right),$$

ибо $E[x_t\varepsilon_t x_{t-j}^2] = 0$ при $j \neq 0$ из-за серийной независимости данных. Если к тому же мы имеем дело с регрессией, т.е. $E[\varepsilon_t|x_t] = 0$, то смещение второго порядка равно нулю, как, впрочем, и смещения более высоких порядков. Если же мы имеем дело с регрессией на временных рядах, то суммируемые ковариации могут не быть нулевыми. Например, в случае авторегрессии первого порядка без свободного члена $y_t = \rho y_{t-1} + \varepsilon_t$, где ε_t – серийно независимый белый шум и $|\rho| < 1$, подсчет математических ожиданий, фигурирующих в предыдущей формуле, приводит к следующему виду смещения МНК-оценки $\hat{\rho}$:

$$E[\hat{\rho}] = \rho - \frac{1}{n}2\rho + o_p\left(\frac{1}{n}\right),$$

что отражает действительно имеющее место смещение МНК-оценок корней характеристического полинома авторегрессии в сторону нуля, тем большее, чем ближе корень к единице.

Вывод формул для смещения второго порядка полезен еще и тем, что предоставляет возможность производить *аналитическую корректировку смещения*, т.е. подправлять оценки для обнуления смещения второго порядка. Для рассмотренной выше авторегрессии скорректированной на смещение оценкой будет следующая:

$$\tilde{\rho} = \hat{\rho} + \frac{1}{n}2\hat{\rho} = \frac{n+2}{n}\hat{\rho}.$$

Заметим, что оценивание смещения второго порядка с использованием смещенных, но состоятельных, оценок вместо неизвестных истинных параметров не влияет на несмещенность второго порядка скорректированной оценки.

Разложения высоких порядков можно продолжить и далее, рассмотрев слагаемые третьего порядка. Известен интересный результат, что, хотя оценка методом максимального правдоподобия (ММП) является смещенной второго порядка, при благоприятных условиях скорректированная на смещение ММП-оценка является асимптотически эффективной третьего порядка, т.е. ее среднеквадратическая ошибка третьего порядка (учитывающая и смещение, и дисперсию третьего порядка) минимальна среди таковых у конкурентов.

Помимо стохастических разложений, описанных в данном разделе, популярными также являются так называемые *разложения Эджворта* для кумулятивных функций распределения различных статистик вокруг их предельных распределений. Эти разложения можно использовать для более точных приближений квантилей (подробнее в (Бандорфф-Нильсен, Кокс, 1999)).

5. АЛЬТЕРНАТИВНАЯ АСИМПТОТИКА

Иногда приближение обычными асимптотическими выводами дает откровенно плохие результаты. Асимптотические распределения могут быть совсем не похожими на реальные либо из-за малости выборки, либо из-за каких-то характеристик распределения данных. Привлечение асимптотики высоких порядков при этом кардинально не меняет ситуацию. Возможно также, что асимптотическое распределение настолько сильно зависит от неизвестных параметров, что пивотизация оценок не представляется возможной, и использовать такую асимптотику проблематично. В таких случаях существуют приемы, позволяющие получать альтернативные асимптотические приближения либо более точные, либо более удобные в использовании.

Здесь уместно вспомнить, что асимптотическое распределение оценки, получаемое при $n \longrightarrow \infty$, есть распределение для некоторого абстрактного объекта, в реальности не существующего. Такое распределение при оценке "на бесконечной выборке" удобно использовать для приближения распределения при оценке, построенной на реальной выборке с конечным n, в частности, из-за стандартности асимптотического распределения. При этом мы надеемся, что приближение хорошее из-за того, что n большое, но в то же время осознаем, что оно конечно в конкретной задаче. Возникает естественный вопрос – почему именно n должно уходить в бесконечность в нашей теоретизации и нет ли более удобной теоретизации, дающей лучшее приближение? Действительно, дело именно так и может обстоять, и ниже мы приводим примеры таких альтернативных подходов. Во всех описываемых моделях присутствует так называемый *дрейфующий параметр*, который, будучи фиксированным при стандартном подходе, теперь меняется при стремлении n к бесконечности, не буквально, конечно, а в рамках новой теоретизации.

Практики, имеющие дело с линейными моделями с инструментальными переменными, особенно специалисты в экономике труда, нередко жалуются на качество используемых инструментов. Хотя им удается находить экзогенные инструменты, эти инструменты слабо скоррелированы с переменными в правой части. И даже маленькой, но ненулевой корреляции достаточно, чтобы оценка двушагового метода наименьших квадратов была состоятельна и асимптотически нормальна, реальное распределение 2ШМНК-оценки далеко от нормального, даже для довольно больших выборок. Вместо куполообразного симметричного распределения в действительности имеет место бимодальное смещенное распределение с бесконечным средним. Такая ситуация – идеальный полигон для применения альтернативной асимптотики.

Рассмотрим простейшую линейную модель с одним регрессором $y_t = x_t\beta + \varepsilon_t$, где ε_t – ошибка, скоррелированная с переменной x_t, т.е. $E[x_t\varepsilon_t] \neq 0$. Пусть даны k инструментов, заключенных в векторе z_t со свойством $E[z_t\varepsilon_t] = 0$, но эти инструменты *слабые* в том смысле, что $E[z_tx_t] \approx 0$, хотя $E[z_tx_t] \neq 0$, или, другими словами, $\pi \approx 0$ во вспомогательной регрессии $x_t = z_t'\pi + v_t$. Понятно, что хотя формально для 2ШМНК-оценки $\hat{\beta}$ справедливо

$$\hat{\beta} \xrightarrow{p} \beta, \quad \sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N\left(0, \frac{E[z_t^2\varepsilon_t^2]}{(E[z_tx_t])^2}\right)$$

при $n \longrightarrow \infty$, нас такое приближение не устраивает. В работе (Staiger, Stock, 1997) было предложено приближать реальное распределение $\hat{\beta}$ асимптотическим при одновременном стремлении $n \longrightarrow \infty$ и $\pi \longrightarrow 0$ или, точнее, $\pi = c/\sqrt{n}$, где c – некоторый вектор констант. То есть дрейфующим параметром является в данном случае π, показатель скоррелированности инструментов и регрессора. Тогда асимптотический анализ приводит к следующим качественно новым выводам: 2ШМНК-оценка несостоятельна и смещена в ту же сторону, в которую смещена несостоятельная МНК-оценка, а асимптотическое распределение 2ШМНК-оценки не нормально, а является смещенной бесконечной смесью нормальных распределений:

$$\hat{\beta} \xrightarrow{p} \beta + \sqrt{\frac{V[\varepsilon_t]}{V[v_t]}}\frac{\zeta_1'(\theta + \zeta_2)}{(\theta + \zeta_2)'(\theta + \zeta_2)}, \quad \text{где} \quad \begin{pmatrix}\zeta_1 \\ \zeta_2\end{pmatrix} \sim N\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}1 & \rho \\ \rho & 1\end{pmatrix} \otimes I_k\right),$$

где

$$\theta = \sqrt{\frac{E[z_tz_t']}{V[v_t]}}c, \quad \rho = \frac{C[\varepsilon_t, v_t]}{\sqrt{V[\varepsilon_t]V[v_t]}}$$

(подробнее в (Staiger, Stock, 1997)). Понятно, что, хотя такое асимптотическое распределение адекватнее отражает происходящее в реальности, пользоваться им труднее, чем стандартной асимптотикой. Дополнительной проблемой является невозможность состоятельного оценивания ключевого параметра c и построения состоятельных тестов (т.е. мощность которых асимптотически приближается к единице) касательно β. В настоящий момент тема слабых инструментов бурно развивается и ищутся решения возникающих проблем (см. (Stock, Wright, Yogo, 2002)).

Другим примером альтернативного подхода является *асимптотика большого числа инструментов* (Bekker, 1994). Нередко на практике встречаются случаи, когда набор используемых инструментов велик, а выборка не очень большая, хотя инструменты сильные. Стандартная асимптотика хорошо работает, когда количество инструментов ничтожно по сравнению с размером выборки, и в данной ситуации приближение стандартной асимптотикой может быть плохим. Альтернативный подход заключается в том, что при $n \longrightarrow \infty$ делается предположение $k \longrightarrow \infty$, $k/n \longrightarrow \alpha$, где k – число инструментов, являющееся дрейфующим параметром, а α – некоторая константа. Если эта константа равна нулю, подход равносилен стандартной асимптотике с малым числом инструментов, а если она строго положительна, ситуация соответствует их большому количеству. При этом, опять же, мы не понимаем асимптотический рост множества инструментов буквально; это очередная абстракция, призванная дать альтернативное приближение истинным распределениям. При таком подходе оказывается, что 2ШМНК-оценка структурных параметров (β) становится несостоятельной. Зато ММП-оценка состоятельна, асимптотически нормальна, правда, с большей дисперсией, чем при стандартном подходе. Увеличение дисперсии – это плата за использование чрезмерно большого числа инструментов, и лучше отражает реальную ситуацию, чем стандартная асимптотика, игнорирующая этот факт.

Улучшения возможны не только для приближений стандартной асимптотикой. Вернемся к примеру с тестированием на единичный корень в простейшей авторегрессии $y_t = \rho y_{t-1} + \varepsilon_t$, где ε_t – белый шум. Пусть значение ρ близко к единице, но строго меньше ее, т.е. мы имеем дело с *почти единичным корнем*. Известно, что в такой ситуации реальное распределение МНК-оценки $\hat{\rho}$ обладает многими чертами распределения Дикки–Фуллера, хотя формально здесь применима именно стандартная асимптотика, где $\hat{\rho}$ является $\sqrt{n}$-состоятельной и асимптотически нормальной оценкой. Чтобы улучшить приближение в (Phillips (1987) было предложено моделировать ρ как параметр, дрейфующий к единице, т.е. $\rho \longrightarrow 1$ при $n \longrightarrow \infty$, или, точнее, $\rho = \exp(c/n)$, где c – некоторая константа. Если $c = 0$, альтернативное приближение должно давать распределение Дикки–Фуллера. В (Phillips, 1987) доказывается, что

$$n(\hat{\rho} - \rho) \xrightarrow{d} \frac{\int_0^1 J_c(r)dB(r)}{\int_0^1 J_c(r)^2 dr},$$

где $J_c(r) = \int_0^1 \exp(c(r - s))dB(s)$ – процесс Орнштейна–Уленбека, а $B(r)$ – стандартное броуновское движение на отрезке [0, 1]. К сожалению, воспользоваться таким асимптотическим распределением проблематично, ибо в нем глубоко "зашит" неизвестный и состоятельно неоцениваемый параметр c. В настоящее время делаются попытки сделать этот подход более конструктивным.

В приведенных трех примерах целью разработки альтернативных методов был поиск *более адекватного приближения* для реальных распределений используемых статистик, чем те, к которым приводит стандартная методология. В других случаях целью является *конструктивность использования* полученных распределений. Примером может служить пороговая авторегрессия, рассмотренная в разд. 3. Вспомним, что параметр δ_n асимптотически уходил в ноль, являясь дрейфующим параметром; при этом асимптотическое распределение НМНК-оценки $\hat{\gamma}$ порога γ получалось пивотальным. В этом-то и вся суть: пивотальное распределение удобно применять на практике и, в частности, табулировать; при обычном подходе асимптотическое распределение было бы непивотальным и проблематичным для использования. Дрейфование δ_n к нулю означает, что разница между параметрами авторегрессии в разных режимах должна быть небольшой, чтобы оправдать предположение $\delta_n \longrightarrow 0$. Если это так, то полученное приближение распределения $\hat{\gamma}$ должно хорошо приближать истинное (правда, отмечено, что приближение хорошее даже для умеренно большой разницы между параметрами). Аналогичный анализ применяется при оценивании моментов структурных сдвигов, когда величина структурного сдвига декларируется теоретически стремящейся к нулю, а на практике небольшой (см. (Andrews, 1993)).

6. ЗАКЛЮЧЕНИЕ

Асимптотическая теория в эконометрике развивается в настоящее время бурными темпами. Новые ЗБЧ и ЦПТ для стандартной асимптотики разрабатываются довольно вяло, хотя можно отметить появление предельных теорем для нестационарных временных рядов с перемешиванием и с "почти эпохальной" зависимостью (см. (Davidson, 1994)). Продвижения же в области асимптотики высоких порядков и альтернативной асимптотики поражают. Асимптотика высоких порядков позволила сравнивать асимптотически (первого порядка) эквивалентные оценки и сделать ряд важных выводов (см. (Rilstone, Srivastava, Ullah, 1996; Newey, Smith, 2004)). Особенно же впечатляют продвижения в задачах со слабыми инструментами и более общего плана задачах со слабой идентификацией (хороший обзор таких задач содержится в (Stock, Wright, Yogo, 2002); см. также недавние достижения в (Kleibergen, 2002; Hahn, Kuersteiner, 2002; Chao, Swanson, 2004; Anatolyev, 2004)). Остается еще масса сложных нерешенных проблем, и в скором времени мы наверняка увидим еще больший расцвет подобных нестандартных подходов. Хочется при этом надеяться, что прикладные экономисты, в том числе и в России, не останутся в стороне от этих достижений.

СПИСОК ЛИТЕРАТУРЫ

Бандорфф-Нильсен О., Кокс Д. (1999): Асимптотические методы в математической статистике. М.: Мир.

Гнеденко Б. В. (1988): Курс теории вероятностей. М.: Наука.

Ибрагимов А.И., Линник Ю.В. (1965): Независимые и стационарно связанные величины. М.: Наука.

Anatolyev S. (2004): Inference When a Nuisance Parameter is Weakly Identified Under the Null Hypothesis // *Economics Letters*. Vol. 84.

Andrews D.W.K. (1991): Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation // *Econometrica*. Vol. 59.

Andrews D.W.K. (1993): Tests for Parameter Instability and Structural Change with Unknown Change Point // *Econometrica*. Vol. 61.

Baltagi B. (2001): Econometric Analysis of Panel Data. N.Y.: John Wiley & Sons.

Bekker P.A. (1994): Alternative Approximations to the Distributions of Instrumental Variable Estimators // *Econometrica*. Vol. 62.

Chao J.C., Swanson N.R. (2004): Consistent Estimation with a Large Number of Weak Instruments. New Brunswick: Rutgers University–New Brunswick.

Davidson J. (1994): Stochastic Limit Theory: An Introduction for Econometricians. Oxford: Oxford University Press.

Franses P., van Dijk D. (2000): Nonlinear Time Series Models in Empirical Finance. Cambridge: Cambridge University Press.

Greene W. (2000): Econometric Analysis: 4th edition. New Jersey: Prentice-Hall.

Hahn J., Kuersteiner G. (2002): Discontinuities of Weak Instrument Limiting Distributions // *Economic Letters*. Vol. 75.

Hahn J., Newey W. (2004): Jackknife and Analytical Bias Reduction for Nonlinear Panel Models // *Econometrica*. Vol. 72.

Hamilton J. (1994): Time Series Analysis. Princeton: Princeton University Press.

Hansen B. (1996): Inference when a Nuisance Parameter is not Identified Under the Null Hypothesis // *Econometrica*. Vol. 64.

Hansen B. (1997): Inference in TAR Models // *Studies in Nonlinear Dynamics and Econometrics*. Vol. 2.

Hardle W., Linton O. (1994): Applied Nonparametric Methods. In: "*Handbook of Econometrics*". Vol. 4. North Holland: Elsevier Science.

Kleibergen F. (2002): Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression // *Econometrica*. Vol. 70.

Manski C.F. (1985): Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator // *J. of Econometrics*. Vol. 27.

Newey W.K., Smith R.J. (2004): Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators // *Econometrica*. Vol. 72.

Newey W.K., West K.D. (1987): A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix // *Econometrica*. Vol. 55.

Phillips P. (1987): Towards a Unified Asymptotic Theory for Autoregression // *Biometrika*. Vol. 74.

Rilstone P., Srivastava V.K., Ullah A. (1996): The Second-Order Bias and Mean-Squared Error of Nonlinear Estimators // *J. of Econometrics*. Vol. 75.

Staiger D., Stock J.H. (1997): Instrumental Variables Regression with Weak Instruments // *Econometrica*. Vol. 65.

Stock J.H., Wright J.H., Yogo M. (2002): A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments // *J. of Business and Economic Statistics*. Vol. 20.

Поступила в редакцию
22.07.2004 г.

Asymptotic Approximations in Modern Econometrics (Survey)

S. A. Anatolyev

This article surveys the state of the art in asymptotic methods of constructing approximations in econometrics. Standard approaches that go deep back to the history of statistics together with more recent alternative approaches are reviewed. The article is directed towards practicing economists who use econometric methods in their research and are familiar with basics of the regression analysis.