

МАТЕМАТИЧЕСКИЙ АНАЛИЗ ЭКОНОМИЧЕСКИХ МОДЕЛЕЙ

Измерение ранжирующей способности модели стохастической границы с усеченным нормальным распределением неэффективности

© 2025 г. Э. Ахмедов, К. К. Фурманов

Э. Ахмедов,

ФИЦ ИУ РАН, Москва; e-mail: aheldar@mail.ru

К. К. Фурманов,

ЦЭМИ РАН, Москва; e-mail: furmach@menja.net

Поступила в редакцию 03.06.2024

Работа Э. Ахметова над разделами «Модель стохастической границы и коэффициент Харрелла» и «Коэффициент Харрелла при усеченном нормальном распределении неэффективностей» поддержана Российским научным фондом (грант 24-21-00494) «Математическое моделирование планирования инвестиций с целью роста производства в условиях меняющейся конъюнктуры» (<https://rscf.ru/project/24-21-00494/>).

Аннотация. Модель стохастической границы — регрессионная модель, в которой объясняемой величиной выступает выпуск предприятия или его издержки, а необъясненная вариация выпуска (издержек) разделяется на две составляющие: стохастический шок и неэффективность, моделируемые случайными величинами с разными законами распределения. Модель позволяет получить оценки неэффективности отдельных предприятий и отраслей в целом, очищенные от влияния стохастических шоков. В настоящий момент она является одним из основных эконометрических инструментов анализа эффективности и продуктивности. Рассматривается задача измерения точности оценок неэффективности предприятий, полученных с помощью модели стохастической границы с усеченным нормальным распределением неэффективности. В качестве характеристики точности используется коэффициент Харрелла, измеряющий согласованность ранжировок оцененных и истинных показателей неэффективности. Выводится формула для расчета асимптотического коэффициента Харрелла, в которой коэффициент согласованности ставится в зависимость от параметров распределения случайных ошибок модели: стохастических шоков и неэффективностей. На основании этой формулы исследователь может измерить ранжирующую способность модели стохастической границы, подставляя вместо неизвестных параметров модели их оценки. Результат расчетов легко интерпретируется: значение коэффициента Харрелла равно вероятности, с которой модель правильно выбирает более эффективное предприятие из двух случайно отобранных. Применение формулы демонстрируется на исторических данных об эффективности хлопкоперерабатывающих заводов СССР. Полученный результат представляется полезным как для академических исследователей, так и для регулирующих органов.

Ключевые слова: стохастическая граница, ранжирование, неэффективность, коэффициент Харрелла.

Классификация JEL: C2, C52.

УДК: 51–77.

Для цитирования: Ахмедов Э., Фурманов К. К. (2025). Измерение ранжирующей способности модели стохастической границы с усеченным нормальным распределением неэффективности // Экономика и математические методы. Т. 61. № 1. С. 109–117. DOI: 10.31857/S0424738825010107

1. ВВЕДЕНИЕ

Эффективность использования ресурсов — одна из основных целей управления хозяйствующими субъектами — предприятиями, отраслями и регионами. Естественно, что внимание экономистов обращено на задачу объективного измерения эффективности. На конференциях выделяются секции и направления исследования, занятые моделированием продуктивности и эффективности; регулирующие органы различных стран используют эконометрические модели эффективности как инструмент контроля (Jamasp, Pollitt, 2000; Haney, Pollitt, 2013).

Основные направления эконометрического анализа эффективности были намечены еще в середине прошлого века. Большую роль сыграла статья (Farrell, 1957), идеи которой, с одной стороны, были положены в основу оболочечного анализа данных (Charnes, Cooper, Rhodes, 1978), а с другой — подтолкнули исследователей к применению регрессионного анализа для измерения эффективности (Winsten, 1957). В настоящий момент среди регрессионных моделей широко применяются модели стохастической границы (stochastic frontier models, SFA). Академическая литература содержит примеры применения этих моделей для анализа эффективности предприятий различных отраслей (Щетинин, Назруллаева, 2012; Ипатов, Пересецкий, 2013; Цветкова, 2021), банков (Головань, 2006; Головань, Карминский, Пересецкий, 2008), учебных заведений (Kirjavainen, 2012; Aleskerov, Belousova, Petrushchenko, 2017), регионов (Айвазян, Афанасьев, Руденко, 2014).

В настоящей статье решается задача измерения точности ранжировок субъектов по оценкам неэффективности, полученным на основе модели стохастической границы. В качестве показателя точности используется коэффициент согласованности Харрелла между истинными, недоступными исследователю, показателями неэффективности и их оценками.

Структура статьи следующая. В разд. 2 приводится обзор литературы, посвященной оценкам неэффективности отдельных предприятий и точности этих оценок. В разд. 3 дается формальное описание модели стохастической границы и общее выражение для коэффициента Харрелла между истинными и оцененными показателями неэффективности. В разд. 4 выводится формула, позволяющая на практике рассчитать коэффициент Харрелла по оценкам модели стохастической границы с усеченным нормальным распределением неэффективности. В разд. 5 полученная формула сопоставляется с ранее выведенным выражением для полунормального распределения неэффективности. В разд. 6 дается пример применения формулы для оценки ранжирующей способности модели стохастической границы по историческим данным из работы (Danilin et al., 1985), после чего следует заключение.

2. ОБЗОР ЛИТЕРАТУРЫ

Модель стохастической границы производственных возможностей была одновременно предложена в статьях (Meeusen, Broeck, 1977; Aigner, Lovell, Schmidt, 1977) и по сути представляла собой линейную регрессионную модель с составной случайной ошибкой. Объясняемая величина отражала выпуск предприятия (в поздних версиях вместо выпуска могли фигурировать издержки); объясняющие переменные отражали затраты факторов производства, а случайная ошибка представлялась в виде суммы двух случайных величин: нормально распределенного стохастического шока и неотрицательной неэффективности. По задумке авторов (Meeusen, Broeck, 1977; Aigner, Lovell, Schmidt, 1977) стохастический шок отражал индивидуальные различия предприятий в достижимом объеме производства при полной эффективности, а неэффективность — отклонение наблюдаемого выпуска от достижимого. В статье (Meeusen, Broeck, 1977) предполагалось, что неэффективность имеет экспоненциальное распределение. В статье (Aigner, Lovell, Schmidt, 1977) рассматривались два варианта модели: с экспоненциальным и с полунормальным распределением неэффективности. Позже появились обобщения, в которых неэффективность имела усеченное нормальное распределение (Stevenson, 1980) и гамма-распределение (Greene, 1990).

Первоначально анализ стохастической границы не предполагал определения уровня неэффективности отдельных предприятий — оценивалась неэффективность отрасли в целом, точнее распределение компоненты неэффективности среди предприятий отрасли, что позволяло рассчитать среднюю неэффективность. Начало анализу неэффективности отдельных предприятий было положено в статье (Матеров, 1981), идея которой получила развитие в работе (Jondrow et al., 1982). Оценки неэффективности отдельных предприятий из статьи (Jondrow et al., 1982) широко используются в настоящее время и называются JLMS-оценками — по первым буквам фамилий авторов статьи¹.

Важно заметить, что оценки неэффективности для отдельных предприятий несостоятельны. По мере увеличения объема выборки можно все точнее оценить границу производственных возможностей. Но даже точное знание этой границы не дает точного знания неэффективности, так как отклонения от границы производственных возможностей связаны не только с неэффективностью, но и со стохастическими шоками, индивидуальными для каждого наблюдения, а они

¹ J. Jondrow, C.A.K. Lovell, I.S. Materov, P. Schmidt.

неизвестны исследователю². Это объясняет внимание к точности индивидуальных оценок. Среди исследований в этом направлении отметим прежде всего работу (Horrace, Schmidt, 1996), где выводится интервальная оценка для неэффективности отдельного предприятия, и работу (Horrace, Seth, Wright, 2015), где исследуется неопределенность рангов предприятий по неэффективности.

В то время как в статье (Horrace, Seth, Wright, 2015) в центре внимания — ранг отдельного предприятия. Работа (Никольский, Фурманов, 2023) посвящена точности ранжировок предприятий по неэффективности целиком (т.е. согласованности рейтинга предприятий по оценкам неэффективности и истинных уровней неэффективности, которая в данной работе измерялась одним числом): выводится приближенная оценка коэффициента Харрелла, отражающего согласованность ранжировок предприятий по истинным значениям неэффективности и их оценкам. Значение коэффициента равно вероятности того, что модель стохастической границы правильно выделит более эффективное предприятие из двух случайно отобранных, таким образом коэффициент является оценкой ранжирующей способности модели. Полученный результат применим в случае полунормального распределения неэффективности.

В настоящей статье результаты исследования (Никольский, Фурманов, 2023) уточняются и обобщаются для случая усеченного нормального распределения неэффективности.

3. МОДЕЛЬ СТОХАСТИЧЕСКОЙ ГРАНИЦЫ И КОЭФФИЦИЕНТ ХАРРЕЛЛА

Рассмотрим регрессионную модель стохастической границы

$$y_i = f(x_i, \beta) - u_i + v_i, \quad i = 1, \dots, n, \quad (1)$$

где y_i — выпуск предприятия i (обычно логарифмированный); x_i — детерминированный вектор затрат факторов производства на предприятии i ; u_i — неэффективность предприятия, v_i — стохастический шок; β — вектор параметров производственной функции; $f(x, \beta)$ — производственная функция, непрерывная по параметрам β ; n — число предприятий в выборке.

Случайные величины $u_1, \dots, u_n, v_1, \dots, v_n$ независимы (это естественно для данных типа «cross-section», но не для панелей). Будем считать, что они непрерывны, все u_i неотрицательны и одинаково распределены, и все v_i одинаково распределены. Эти предпосылки выполнены для первоначальных моделей из работ (Aigner, Lovell, Schmidt, 1977; Meeusen, Broeck, 1977; Stevenson, 1980), однако в современной практике используются варианты, в которых распределения шоков и неэффективностей зависят от объясняющих переменных (см. обзор литературы в статье (Малахов, Пильник, 2013)).

Модель (1) оценивается, как правило, либо методом максимального правдоподобия, либо сочетанием метода наименьших квадратов и метода моментов (Greene, 2002). Оба метода дают состоятельные оценки параметров производственной функции β и параметров распределений неэффективностей и шоков. JLMS-оценка неэффективности u_i имеет вид:

$$u_i = g\left(y_i - f(x_i, \hat{\beta})\right), \quad g(\varepsilon) = E(u | v - u = \varepsilon).$$

Здесь $\hat{\beta}$ — оценка параметров производственной функции. Функция $g(\varepsilon)$ убывающая, поэтому оценки неэффективности тем больше, чем меньше остатки регрессии $y_i - f(x_i, \hat{\beta})$. Это вполне соответствует здравому смыслу: предприятие, которое при фиксированных затратах факторов дает меньший выпуск, считается менее эффективным.

В качестве характеристики согласованности между истинными показателями неэффективности u_i и их оценками $\hat{u}_i$ рассмотрим коэффициент Харрелла, предложенный в работе (Harrell et al., 1982) как измеритель точности медицинских тестов и тесно связанный с коэффициентами ранговой корреляции Кендалла—Сомерса (Newson, 2002). Пусть u_1, u_2 — значения истинных неэффективностей для двух случайно отобранных предприятий, а $\hat{u}_1, \hat{u}_2$ — их оценки. Коэффициент Харрелла C определяется равенством

$$C = P(\hat{u}_1 < \hat{u}_2 | u_1 < u_2). \quad (2)$$

² Это не касается панельных моделей стохастической границы — увеличение числа наблюдений отдельного предприятия позволяет состоятельно оценить неэффективность, если считать, что она постоянна либо меняется от наблюдения к наблюдению неким предсказуемым образом, как в моделях, рассмотренных в (Battese, Coelli, 1988).

Коэффициент показывает, с какой вероятностью оценка неэффективности у относительно неэффективного предприятия будет выше, чем у относительно эффективного, и таким образом измеряет способность модели выделять относительно (не)эффективные предприятия.

Непосредственная оценка вероятности (2) на практике невозможна, так как значения u_i неизвестны, однако эту вероятность можно выразить через состоятельно оцениваемые параметры распределений случайных величин u_i и v_i , как показано в работе (Никольский, Фурманов, 2023).

Выразим событие $\{\hat{u}_1 < \hat{u}_2\}$ через переменные модели (1):

$$C = P(y_1 - f(x_1, \hat{\beta}) > y_2 - f(x_2, \hat{\beta}) | u_1 < u_2).$$

Мы пользуемся тем, что u_i есть убывающая функция от регрессионного остатка $y_i - f(x_i, \hat{\beta})$.

Рассмотрим предел при увеличении объема выборки и получим асимптотический коэффициент Харрелла C_{asy} :

$$\begin{aligned} C_{asy} &= \lim_{n \rightarrow \infty} P(y_1 - f(x_1, \hat{\beta}) > y_2 - f(x_2, \hat{\beta}) | u_1 < u_2) = \\ &= P(y_1 - f(x_1) > y_2 - f(x_2) | u_1 < u_2) = P(v_1 - u_1 > v_2 - u_2 | u_1 < u_2), \end{aligned}$$

при этом мы опираемся на состоятельность оценки $\hat{\beta}$. Так как u_i непрерывны, независимы и одинаково распределены, то $P(u_1 < u_2) = 0,5$ и

$$C_{asy} = P(v_1 - u_1 > v_2 - u_2 | u_1 < u_2) = 2P(\{v_1 - u_1 > v_2 - u_2\} \cap \{u_1 < u_2\}). \quad (3)$$

Введем в рассмотрение величины $\xi = v_1 - v_2$ и $\eta = u_1 - u_2$. Теперь (3) может быть переписана в виде:

$$C_{asy} = 2P(\{\xi > \eta\} \cap \{\eta < 0\}) = 2 \int_{-\infty}^0 f_{\eta}(x) \int_x^{\infty} f_{\xi}(t) dt dx, \quad (4)$$

где $f_{\eta}(x)$ и $f_{\xi}(t)$ — функции плотности случайных величин η и ξ соответственно.

Выражения (3) и (4) не привязаны к конкретному виду распределения участвующих в модели случайных величин. В статье (Никольский, Фурманов, 2023) предлагается приближенный результат, основанный на предпосылке о полуномальном распределении неэффективностей и на нормальном приближении для разности полуномальных величин. В настоящей статье мы опираемся на более общую предпосылку об усеченном нормальном распределении и вместо приближенной формулы предлагаем численный расчет вероятности из выражений (3) и (4). Эта вероятность, как будет показано в следующем разделе, сводится к одномерному интегралу и может быть рассчитана численными методами.

4. КОЭФФИЦИЕНТ ХАРРЕЛЛА ПРИ УСЕЧЕННОМ НОРМАЛЬНОМ РАСПРЕДЕЛЕНИИ НЕЭФФЕКТИВНОСТЕЙ

Дополним список предпосылок следующими утверждениями: случайные величины v_i нормально распределены, $v_i \sim N(0, \sigma_v^2)$; случайные величины u_i имеют усеченное слева в нуле нормальное распределение с параметрами μ и σ_u^2 , $u_i \sim N^+(\mu, \sigma_u^2)$. Заметим, что μ и σ_u^2 — это не математическое ожидание и не дисперсия неэффективности, они равны математическому ожиданию и дисперсии, которые были бы при отсутствии усечения.

По свойству нормального распределения $\xi = v_1 - v_2 \sim N(0, 2\sigma_v^2)$, так что

$$f_{\xi}(x) = \left(\sqrt{4\pi\sigma_v^2}\right)^{-1} \exp(-x^2 / 4\sigma_v^2).$$

Величины u_1 и u_2 имеют усеченное в нуле нормальное распределение, их плотность задается выражением

$$f_u(x) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma_u^2}\right) / \int_0^{+\infty} \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{(t-\mu)^2}{2\sigma_u^2}\right) dt, \quad x \geq 0.$$

Функция плотности при отрицательных значениях аргумента равна нулю.

Для удобства введем обозначение

$$\lambda(\mu, \sigma_u^2) = \frac{1}{\sqrt{2\pi\sigma_u^2}} / \int_0^{+\infty} \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{(t-\mu)^2}{2\sigma_u^2}\right) dt = \left(\left(\frac{\mu}{\sigma_u}\right) \sqrt{2\pi\sigma_u^2} \right)^{-1}, \quad (5)$$

где $\Phi(\mu / \sigma_u)$ — функция распределения стандартного нормального закона. Теперь функцию плотности величин u_1 и u_2 можно описать формулой:

$$f_u(x) = \lambda(\mu, \sigma_u^2) \exp\left(-\frac{(x-\mu)^2}{2\sigma_u^2}\right), \quad x \geq 0.$$

Найдем плотность величины $\eta = u_1 - u_2$ по формуле свертки. При $x \geq 0$:

$$f_\eta(x) = \int_{-\infty}^{+\infty} f_u(t) f_u(t-x) dt = \lambda^2(\mu, \sigma_u^2) \int_x^{+\infty} \exp\left(-\frac{(t-\mu)^2}{2\sigma_u^2} - \frac{(t-x-\mu)^2}{2\sigma_u^2}\right) dt.$$

Выделим полный квадрат в показателе экспоненты:

$$\exp\left(-\frac{(t-\mu)^2}{2\sigma_u^2} - \frac{(t-x-\mu)^2}{2\sigma_u^2}\right) = \exp\left(-\frac{1}{\sigma_u^2} \left[t - \left(\frac{x}{2} - \mu\right) \right]^2 + \frac{x^2}{4}\right).$$

Вынесем постоянный множитель за знак интеграла в функции плотности:

$$\begin{aligned} f_\eta(x) &= \lambda^2(\mu, \sigma_u^2) \int_x^{+\infty} \exp\left(-\frac{1}{\sigma_u^2} \left[t - \left(\frac{x}{2} - \mu\right) \right]^2 + 0,25x^2\right) dt = \\ &= \lambda^2(\mu, \sigma_u^2) \exp(-x^2 / 4\sigma_u^2) \int_x^{+\infty} \exp\left(-\left(t - \left(\frac{x}{2} - \mu\right)\right)^2 / \sigma_u^2\right) dt. \end{aligned}$$

Интеграл может быть выражен через функцию стандартного нормального распределения:

$$\int_x^{+\infty} \exp\left(-\left(t - \left(\frac{x}{2} - \mu\right)\right)^2 / \sigma_u^2\right) dt = \Phi\left(\frac{\sqrt{2}}{\sigma_u}(\mu - 0,5x)\right) \sqrt{\pi\sigma_u^2}.$$

В результате плотность величины η для положительных значений аргумента имеет вид:

$$f_\eta(x) = \lambda^2(\mu, \sigma_u^2) \exp(-x^2 / 4\sigma_u^2) \Phi\left(\left(\sqrt{2} / \sigma_u\right)(\mu - 0,5x)\right) \sqrt{\pi\sigma_u^2}, \quad x \geq 0.$$

Величина η распределена симметрично относительно нуля, так как является разностью двух одинаково распределенных случайных величин u_1 и u_2 , поэтому функция плотности η четная:

$$f_\eta(x) = \lambda^2(\mu, \sigma_u^2) \exp(-x^2 / 4\sigma_u^2) \Phi\left(\left(\sqrt{2} / \sigma_u\right)(\mu - 0,5|x|)\right) \sqrt{\pi\sigma_u^2}. \quad (6)$$

Осталось подставить функции плотности ξ и η в выражение (4) и рассчитать соответствующие интегралы. Заметим, что $\int_x^{+\infty} f_\xi(t) dt$ — интеграл плотности нормального распределения с нулевым математическим ожиданием и дисперсией $2\sigma_v^2$:

$$\int_x^{+\infty} f_\xi(t) dt = 1 - \Phi\left(x / \sqrt{2\sigma_v^2}\right) = \Phi\left(-x / \sqrt{2\sigma_v^2}\right). \quad (7)$$

Подставляем выражения (6) и (7) в (4) и получаем выражение для асимптотического коэффициента Харрелла:

$$C_{asy} = 2\lambda^2(\mu, \sigma_u^2) \sqrt{\pi\sigma_u^2} \int_{-\infty}^0 \exp(-x^2 / 4\sigma_u^2) \Phi\left(\left(\sqrt{2} / \sigma_u\right)(\mu - 0,5|x|)\right) \Phi\left(-x / \sqrt{2\sigma_v^2}\right) dx.$$

Так как интегрирование ведется по отрицательной области, то можем раскрыть модуль и подставить выражение для $\lambda(\mu, \sigma_u^2)$ из формулы (5). Тогда

$$C_{asy} = \left(\Phi^2(\mu / \sigma_u) \sqrt{\pi\sigma_u^2}\right)^{-1} \int_{-\infty}^0 \exp(-x^2 / 4\sigma_u^2) \Phi\left(\left(\sqrt{2} / \sigma_u\right)(\mu + 0,5x)\right) \Phi\left(-x / \sqrt{2\sigma_v^2}\right) dx. \quad (8)$$

Хотя аналитическое выражение для коэффициента Харрелла вывести не удалось, численный расчет интеграла в формуле (8) не составляет представляет сложным. Полученное выражение может быть использовано для измерения ранжирующей способности модели стохастической границы, если вместо неизвестных параметров модели μ , σ_v^2 , σ_u^2 подставлять их оценки. При этом погрешность оценивания скажется и на значении коэффициента Харрелла.

5. СОПОСТАВЛЕНИЕ С ИМЕЮЩИМИСЯ РЕЗУЛЬТАТАМИ ДЛЯ ПОЛУНОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

В статье (Никольский, Фурманов, 2023) приводится приближенная формула для коэффициента Харрелла в модели с полуноормальным распределением неэффективности:

$$C_{asy} \approx 2\Phi_2\left(0,0;\sqrt{\sigma_u^2(1-2/\pi)/[\sigma_u^2(1-2/\pi)+\sigma_v^2]}\right). \quad (9)$$

Здесь $\Phi_2(x,y;\rho)$ — значение совместной функции распределения двух стандартных нормальных случайных величин с коэффициентом корреляции ρ в точке (x,y) .

Таблица 1. Значения асимптотического коэффициента Харрелла, рассчитанные по формуле (8) численным интегрированием (C_{asy}) и по приближенной формуле (9) ($C_{asy}^{прибл}$), в зависимости от дисперсии неэффективности σ_u^2 в случае полунормального распределения ($\mu = 0$)

σ_u^2	C_{asy}	$C_{asy}^{прибл}$
0,01	0,519	0,519
0,1	0,558	0,560
0,25	0,590	0,593
0,5	0,624	0,628
1	0,667	0,673
2	0,716	0,725
4	0,769	0,780
10	0,834	0,846
100	0,940	0,948

по формулам (8) и (9). Впрочем, основная ценность формулы (8) не в уточнении уже имевшего результата для полунормальной модели, а в ее применимости для более общего случая усеченного нормального распределения.

6. ПРИМЕР ПРИМЕНЕНИЯ

Таблица 2. Пример расчета коэффициента согласованности Харрелла (C_{asy}) по параметрам μ , σ_u^2 , σ_v^2 (Danilin et al., 1985)

Выборка заводов	μ	σ_u^2	σ_v^2	C_{asy}
Вся выборка	-0,2734	0,0345	0,1167	0,562
До 1961 г.	-0,2579	0,0219	0,0876	0,553
С 1961 г.	-0,3316	0,0410	0,1005	0,569

Примечание. Расчет выполнен по формуле (8).

Мы приводим оценки только тех параметров, которые имеют отношение к согласованности ранжировок, т.е. дисперсии стохастического шока σ_u^2 и параметров распределения неэффективности μ и σ_v^2 .

Согласно данным, представленным в табл. 2, оцененная модель стохастической границы имеет низкую ранжирующую способность как для всей выборки, так и для ее частей. Из двух случайно отобранных заводов модель верно выделяет более эффективный завод с вероятностью в пределах 0,55–0,57. Это не означает, что модель неприменима для иных целей, помимо ранжирования. Авторы статьи (Danilin et al., 1985) при анализе неэффективности отдельных заводов использовали оценки, основанные не на условном математическом ожидании, как предложено в работе (Jondrow et al., 1982), а на условной моде, как в статье (Матеров, 1981). К этим оценкам формула (8) неприменима, так как при ее выводе мы опирались на предпосылку, что оценка неэффективности есть строго убывающая функция остатка регрессии, что не выполняется для оценок на основе моды.

7. ЗАКЛЮЧЕНИЕ

Выведенная в статье формула (8) позволяет сравнить точность оценок неэффективности отдельных предприятий: она дает исследователю ответ на вопрос, насколько вероятно эти оценки

Так как полунормальное распределение — это частный случай усеченного нормального при $\mu = 0$, то мы можем сравнить расчеты по формулам (8) и (9) для частичной проверки представленного в настоящей статье результата. Заметим, что коэффициент Харрелла не меняется, если участвующие случайные величины умножить на одно и то же число, поэтому мы можем, не теряя общности, постановить $\sigma_v^2 = 1$ — всегда можно нормировать величины u_i и v_i , чтобы обеспечить это равенство. В табл. 1 приведены значения коэффициента Харрелла при разной дисперсии неэффективности σ_u^2 . Расчеты не только показывают связь ранжирующей способности модели с дисперсией неэффективности, но и служат проверкой для выведенной в настоящей статье формулы (8) — в ее пользу свидетельствует близость значений, полученных

В статье (Danilin et al., 1985) приводятся оценки модели стохастической границы с усеченным нормальным распределением неэффективности для 171 хлопкоперерабатывающего завода СССР. По этим данным мы рассчитали коэффициент Харрелла между показателями неэффективности заводов и их оценками. Таблица 2 содержит результаты расчетов как для целой выборки, так и для ее частей: в исходной статье представлены оценки отдельно для заводов, основанных до 1961 г. и начиная с 1961 г.

приводят к правильному выводу при сравнении неэффективности, что очень важно, в частности для регулирования работы предприятий. Регулирующие органы могут опираться на оценки эконометрических моделей (практика их применения рассматривается в работах (Jamasb, Pollitt, 2000; Haneу, Pollitt, 2013)), но ценность этих оценок низка, если модель не в состоянии отличить эффективное предприятие от неэффективного. Такие случаи можно выявить, пользуясь результатами настоящей статьи. Низкие (близкие к 0,5) значения коэффициента Харрелла свидетельствуют о том, что эконометрическая модель примерно в такой же степени способна привести к точному рейтингу неэффективности, как и случайное перемешивание списка предприятий.

К сожалению, формула (8) не применима в тех случаях, когда параметры распределений шоков и неэффективностей не одинаковы для всех предприятий. Исследование этих случаев затрудняется, в частности тем, что при различных распределениях случайных компонент модели не существует единого коэффициента ранговой корреляции или конкордации.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- Айвазян С.А., Афанасьев М.Ю., Руденко В.А. (2014). Оценка эффективности регионов РФ на основе модели производственного потенциала с характеристиками готовности к инновациям // *Экономика и математические методы*. Т. 50. № 4. С. 34–70. [Aivazian S.A., Afanasiev M.Yu., Rudenko V.A. (2014). Efficiency estimation of Russian regions based on the productive potential model including the characteristics of readiness to innovate. *Economics and Mathematical Methods*, 50, 4, 34–70 (in Russian).]
- Головань С.В. (2006). Факторы, влияющие на эффективность российских банков // *Прикладная эконометрика*. Т. 2. С. 3–17. [Golovan S. (2006). Factors influencing the efficiency of Russian banks performance. *Applied Econometrics*, 2, 3–17 (in Russian).]
- Головань С.В., Карминский А.М., Пересецкий А.А. (2008). Эффективность российских банков с точки зрения минимизации издержек с учетом факторов риска // *Экономика и математические методы*. Т. 44. № 4. С. 28–38. [Golovan S.V., Karminsky A.M., Peresetsky A.A. (2008). Cost efficiency of Russian banks, taking into account the risk factors. *Economics and Mathematical Methods*, 44, 4, 28–38 (in Russian).]
- Ипатова И.Б., Пересецкий А.А. (2013). Техническая эффективность предприятий отрасли производства резиновых и пластмассовых изделий // *Прикладная эконометрика*. Т. 32. С. 71–92. [Ipatova I., Peresetsky A. (2013). Technical efficiency of Russian plastic and rubber production firms. *Applied Econometrics*, 32, 71–92 (in Russian).]
- Малахов Д.И., Пильник Н.П. (2013). Методы оценки показателя эффективности в моделях стохастической производственной границы // *Экономический журнал ВШЭ*. Т. 17. № 4. С. 660–686. [Malakhov D., Pilnik N. (2013). Methods of estimating of the efficiency in stochastic frontier models. *HSE Economic Journal*, 17 (4), 660–686 (in Russian).]
- Матеров И.С. (1981). К проблеме полной идентификации модели стохастических границ производства // *Экономика и математические методы*. Т. 17. № 4. С. 784–788. [Materov I.S. (1981). On full identification of the stochastic production frontier model. *Economics and Mathematical Methods*, 17, 4, 784–788 (in Russian).]
- Никольский И.М., Фурманов К.К. (2023). Измерение точности ранжировок предприятий по эффективности в модели стохастической границы // *Прикладная эконометрика*. Т. 71. С. 128–142. DOI: 10.22394/1993-7601-2023-71-128-142 [Nikolskiy I.M., Furmanov K.K. (2023). Assessing the accuracy of efficiency rankings obtained from a stochastic frontier model. *Applied Econometrics*, 71, 128–142. DOI: 10.22394/1993-7601-2023-71-128-142 (in Russian).]
- Цветкова А.Н. (2021). Динамика технической эффективности российских предприятий в 2013–2018 годах // *Прикладная эконометрика*. Т. 63. С. 91–116. DOI: 10.22394/1993-7601-2021-63-91-116 [Tsvetkova A. (2021). Technical efficiency trends of Russian firms in 2013–2018. *Applied Econometrics*, 63, 91–116. DOI: 10.22394/1993-7601-2021-63-91-116 (in Russian).]
- Щетинин Е.И., Назруллаева Е.Ю. (2012). Производственный процесс в пищевой промышленности: взаимосвязь инвестиций в основной капитал и технической эффективности // *Прикладная эконометрика*. Т. 28. С. 63–84. [Shchetynin Ye.I., Nazrullaeva E. Yu. (2012). Effects of fixed capital investments on technical efficiency in food industry. *Applied Econometrics*, 28, 63–84 (in Russian).]
- Aigner D., Lovell C.A.K., Schmidt P. (1977). Formulation and estimation of stochastic frontier function models. *Journal of Econometrics*, 6, 21–37.
- Aleskerov F.T., Belousova V.Yu., Petrushchenko V.V. (2017). Models of data envelopment analysis and stochastic frontier analysis in the efficiency assessment of universities. *Automation and Remote Control*, 78 (5), 902–923. DOI: 10.1134/S0005117917050125

- Battese G.E., Coelli T.J.** (1988). Prediction of firm-level technical efficiencies: With a generalized frontier production function and panel data. *Journal of Econometrics*, 38, 387–399.
- Charnes A., Cooper W.W., Rhodes E.** (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2 (6), 429–444. doi: 10.1016/0377-2217 (78)90138-8
- Danilin V.I., Materov I.S., Rosefield S., Lovell C.A.K.** (1985). Measuring enterprise efficiency in the Soviet Union: A stochastic frontier analysis. *Economica*, 52 (206), 225–233.
- Farrell M.J.** (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society. Series A (General)*, 120, 253–290.
- Greene W.H.** (1990). A gamma-distributed stochastic frontier model. *Journal of Econometrics*, 46, 141–163. DOI: 10.1016/0304-4076 (90)90052-U
- Greene W.H.** (2002). *Econometric analysis*. 5th ed. Upper Saddle River: Prentice Hall.
- Haney A.B., Pollitt M.G.** (2013). International benchmarking of electricity transmission by regulators: A contrast between theory and practice. *Energy Policy*, 62, 267–281.
- Harrell F.E., Califf R.M., Pryor D.B., Lee K.L., Rosati R.A.** (1982). Evaluating the yield of medical tests. *Journal of the American Medical Association*, 247, 2543–2546.
- Horrace W.C., Schmidt P.** (1996). Confidence statements for efficiency estimates from stochastic frontier models. *Journal of Productivity Analysis*, 7, 257–282.
- Horrace W.C., Seth R.-S., Wright I.** (2015). Expected efficiency ranks from parametric stochastic frontier models. *Empirical Economics*, 48, 2, 829–848. DOI: 10.1007/s00181-014-0808-8
- Jamasb T., Pollitt M.** (2000). Benchmarking and regulation: International electricity experience. *Utilities Policy*, 9 (3), 107–130.
- Jondrow J., Lovell C.A.K., Materov I.S., Schmidt P.** (1982). On the estimation of technical inefficiency in stochastic frontier production function model. *Journal of Econometrics*, 19, 233–239.
- Kirjavainen T.** (2012). Efficiency of Finnish general upper secondary schools: An application of stochastic frontier analysis with panel data. *Education Economics*, 20 (4), 343–364.
- Meeusen W., Broeck J. van den** (1977). Efficiency estimation from Cobb–Douglas production functions with composed error. *International Economic Review*, 18 (2), 435–444.
- Newson R.** (2002). Parameters behind “nonparametric” statistics: Kendall’s tau, Somers’ D and median differences. *Stata Journal*, 2, 45–64.
- Stevenson R.E.** (1980). Likelihood functions for generalized stochastic frontier estimation. *Journal of Econometrics*, 13, 57–66.
- Winsten C.B.** (1957). Discussion on Mr. Farrell’s paper. *Journal of the Royal Statistical Society. Series A (General)*, 120, 282–284.

Assessing the accuracy of efficiency rankings obtained from a stochastic frontier model with truncated normal distribution of inefficiency

© 2025 E. Ahmedov, C.C. Furmanov

E. Ahmedov,

Federal Research Center “Computer Science and Control” RAS, Moscow, Russia; e-mail: aheldar@mail.ru

C.C. Furmanov,

Central Economics and Mathematics Institute, Russian Academy of Sciences (CEMI RAS), Moscow, Russia; e-mail: furmach@menja.net

Received 03.06.2024

Eldar Ahmedov's work on “Stochastic frontier model and Harrell's C-index” and “Harrell's C-index in the case of truncated normal inefficiency” is supported by Russian Science Foundation grant no. 24-21-00494 “Mathematical modeling of investment planning for the purpose of production growth in a changing environment” (<https://rscf.ru/en/project/24-21-00494/>).

Abstract. A stochastic frontier model is a regression model where an explained variable is either output of a firm or its costs, and unexplained variation of this variable is divided into two components: inefficiency and stochastic shock. These components are modeled by random variables with different families of distributions. The model allows estimation of inefficiency at firm level and at industry level refined from the effects of stochastic shocks. At present it is the basic instrument for analyzing the efficiency and productivity. We consider a problem of assessing the accuracy of inefficiency estimators, obtained via stochastic frontier model with truncated normal distribution of inefficiency components. We propose using Harrell's C-index as a measure of concordance between true inefficiencies and their estimates. We derive the expression for the asymptotic C-index as a function of distribution parameters of model's random components (stochastic shocks and inefficiencies). The derived expression can be used by practitioners for assessing the ranking ability of a stochastic frontier model. The value of C-index has clear interpretation: it is the probability of choosing a more efficient firm from two randomly selected ones. For demonstration purposes, we provide historical data on cotton refining plants in the Soviet Union. The obtained result may be useful both for academic researchers and for regulatory agencies.

Keywords: stochastic frontier, ranking, inefficiency, Harrell's C.

JEL Classification: C2, C52.

UDC: 51–77.

For reference: **Ahmedov E., Furmanov C.C.** (2025). Assessing the accuracy of efficiency rankings obtained from a stochastic frontier model with truncated normal distribution of inefficiency. *Economics and Mathematical Methods*, 61, 1, 109–117. DOI: 10.31857/S0424738825010107 (in Russian).