

# УЧЕТ НЕОДНОРОДНОСТИ В ОБОЛОЧЕЧНОМ АНАЛИЗЕ ДАННЫХ

В.В. Петрущенко

Рассмотрена классическая модель оболочечного анализа данных (ОАД), широко применяющаяся при оценке эффективности в самых разных экономических отраслях. Проанализированы области ее применимости. Изучено влияние неоднородности оцениваемой выборки объектов на значения эффективности, получаемые с помощью ОАД. Дан обзор методов, позволяющих получать более точные оценки эффективности в случае неоднородной выборки.

**Ключевые слова:** эффективность, оболочечный анализ данных, неоднородность выборки.

## ВВЕДЕНИЕ

Обсудим основную проблему, возникающую при попытке оценить эффективность множества компаний. Подразумевается, что каждая фирма в выборке производит какое-то количество результатов  $q_1, \dots, q_M$ , используя ресурсы нескольких типов  $x_1, \dots, x_N$  ( $M$  и  $N$  фиксированы для данной выборки). Естественным подходом здесь было бы агрегировать количество потраченных ресурсов и результатов и принять за эффективность  $i$ -го объекта величину

$$Eff_i = \frac{\text{Aggregated Output}}{\text{Aggregated Input}} = \frac{u_1 q_{1i} + \dots + u_M q_{Mi}}{v_1 x_{1i} + \dots + v_N q_{Ni}}$$

Однако информация о том, как выбирать весовые коэффициенты (их также можно интерпретировать как важность соответствующих результатов или ресурсов)  $u_1, \dots, u_M, v_1, \dots, v_N$ , отсутствует. Оболочечный анализ данных (ОАД, англ. — Data Envelopment Analysis, DEA) как раз и представляет собой аппарат, позволяющий автоматически выбрать данные коэффициенты.

Начнем с базовой мультипликативной формулировки, введенной в работе [1]. Авторы развили и обобщили предложенную в работе [2] модель и показали, что проблема оценки эффективности

может быть сформулирована в терминах математического программирования:

найти

$$\max_{u, v} \left( \theta_i = \frac{u_1 q_{1i} + \dots + u_M q_{Mi}}{v_1 x_{1i} + \dots + v_N q_{Ni}} \right) \quad (1)$$

при ограничениях

$$\begin{aligned} \frac{u_1 q_{1i} + \dots + u_M q_{Mi}}{v_1 x_{1i} + \dots + v_N q_{Ni}} &\leq 1, \quad i = 1, \dots, L; \\ u_j &\geq 0, \quad j = 1, \dots, M; \\ v_k &\geq 0, \quad k = 1, \dots, N, \end{aligned} \quad (2)$$

где  $L$  — число оцениваемых фирм,  $q_{ij}$  —  $j$ -й выходной параметр ( $j = 1, \dots, M$ )  $i$ -й фирмы,  $x_{ki}$  —  $k$ -й входной параметр ( $k = 1, \dots, N$ )  $i$ -й фирмы,  $u$  и  $v$  — векторы неизвестных подходящей размерности. Наконец, значение  $\theta_i \in (0, 1]$  представляет собой эффективность  $i$ -й компании.

Оптимизационная задача (1), (2) может быть представлена в виде задачи линейного программирования в форме, которая в литературе носит название оболочечной формы (от англ. envelopment form):

найти

$$\min_{\lambda, \theta_i} \theta_i \quad (3)$$



при ограничениях

$$\begin{aligned} -q_i + Q\lambda &\geq 0; \\ \theta_i x_i - X\lambda &\geq 0; \\ \lambda &\geq 0, \end{aligned} \quad (4)$$

где  $q_i$  —  $(M \times 1)$ -вектор выходных параметров  $i$ -й фирмы,  $x_i$  —  $(N \times 1)$ -вектор входных параметров  $i$ -й фирмы,  $Q$  —  $(M \times L)$ -матрица выходных параметров всех  $L$  фирм,  $X$  —  $(N \times L)$ -матрица выходных параметров всех фирм,  $\lambda$  —  $(L \times 1)$ -вектор весов, которые можно интерпретировать как теневые цены (см. работу [3]). Как и ранее, значение  $\theta_i$  — эффективность  $i$ -й фирмы.

Формулировка (3), (4) является фундаментальной. Данная версия модели ОАД также носит название ССР (по инициалам авторов-разработчиков). Область применимости данной версии ОАД ограничивается случаем, когда все фирмы из оцениваемой выборки оперируют на оптимальном (постоянном) для себя масштабе. Поэтому в англоязычной литературе модель также часто называется CRS DEA (от англ. Constant Return to Scale — постоянная отдача от масштаба).

Несложно адаптировать модель к случаю переменной отдачи от масштаба, т. е. к ситуации, когда некоторые фирмы ведут производство в неоптимальном для них объеме. Для этого необходимо добавить к ограничениям дополнительное ограничение

$$1^T \lambda = 1, \quad (5)$$

где  $1^T$  — единичный вектор размера  $1 \times L$ .

Модель (3)–(5) носит название VRS DEA (от англ. Variable Return to Scale). Данная модификация была введена в работе [4].

С помощью мультипликативной формы (1), можно установить важную геометрическую интерпретацию обсуждавшихся сформулированных двух моделей в трех простейших случаях: один ресурс — один результат (рис. 1); два ресурса — один результат (рис. 2, а); один ресурс — два результата (рис. 2, б).

Под границей эффективности, или просто границей, подразумевается множество<sup>1</sup> в ресурсно-результативном пространстве, которое «обволакивает» все множество альтернатив<sup>2</sup>.

Фирма  $A$  эффективна в смысле обеих моделей, фирмы  $C$  и  $D$  эффективны только в смысле VRS-версии модели ОАД. Эффективность фирмы  $B$

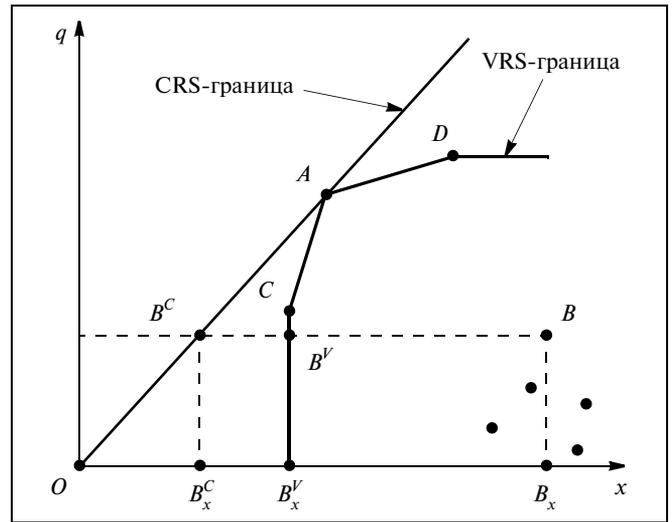


Рис. 1. Интерпретация модели ОАД в случае  $N = M = 1$ :  $x$  — ресурс;  $q$  — выпуск

(см. рис. 1) в VRS- и CRS-модификациях рассчитывается как отношение  $|OB_x^V|/|OB_x|$  и  $|OB_x^C|/|OB_x|$ , соответственно. Такая мера полностью совпадает с интуитивным суждением о том, как должна быть устроена модель оценки относительной эффективности. Заметим, что именно с помощью границы эффективности оценивается все множество альтернатив.

Чуть более сложно выглядят диаграммы для двух остальных простейших случаев (см. рис. 2).

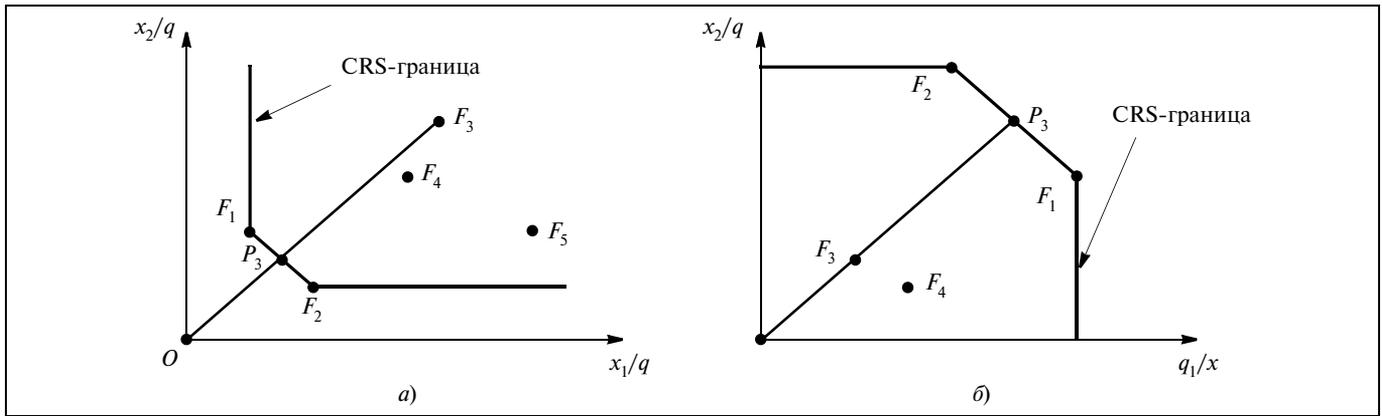
Эффективность фирмы  $F_3$  измеряется в случае  $N = 2, M = 1$  как отношение  $|OP_3|/|OF_3|$ , в случае  $N = 1, M = 2$  как отношение  $|OF_3|/|OP_3|$ . Формальная точка  $P_3$  также называется проекцией фирмы  $F_3$  на границу эффективности.

Отметим, что отдача от масштаба далеко не единственное ограничение области применимости моделей ОАД. Таковыми могут быть, например, наличие управляемых в кратковременном периоде параметров, которые необходимо включать в оценку. Примером может служить оценка эффективности фермерских хозяйств. В данном случае площадь земельных угодий, амбаров и др. — управляемый в краткосрочном периоде параметр. С другой стороны, очевидно, что этот параметр влияет на эффективность ферм, поэтому его необходимо включать в качестве ресурсной переменной. Приведенные постановки ОАД не в силах справиться с такой задачей. Для подобных случаев существует исчерпывающее решение, подробно изложенное в работе [3].

Гораздо более серьезное ограничение на применение моделей оболочечного анализа данных заключается в разнородности оцениваемой вы-

<sup>1</sup> Состоящее в общем случае из пересечения нескольких гиперповерхностей.

<sup>2</sup> Под альтернативами здесь понимаются оцениваемые фирмы.



**Рис. 2. Интерпретация модели ОАД:**  
 $N = 2, M = 1$  (а);  $N = 1, M = 2$  (б)

борки. Данная проблема представляет интерес также и потому, что она не имеет единого общепризнанного решения. Далее будем рассматривать различные адаптации модели ОАД к ситуации, когда оцениваемые альтернативы существенно неоднородны.

Продолжение работы организовано следующим образом. Сначала дается краткий обзор различных применений модели ОАД на практике. Затем обсуждаются различные виды неоднородности, строятся содержательные примеры ситуаций, когда выборка неоднородна и стандартный метод ОАД неприменим. Дается четыре базовых подхода к решению данной проблемы. Далее обсуждается обобщение классического ОАД, основанное на том, что оцениваемые фирмы имеют различные цели и приоритеты. Последняя часть работы посвящена изложению подхода к решению проблемы неоднородных выборок, предложенному в работе [5].

## 1. ПРИМЕНЕНИЯ МЕТОДА ОБОЛОЧЕЧНОГО АНАЛИЗА ДАННЫХ

На практике модель ОАД приобрела широкую известность. Затронем лишь некоторые экономические сферы, в которых она успешно используется. Прежде всего, отметим активное применение описанных моделей к оценке эффективности в области высшего образования различных государств, например, в Великобритании [6–9], Австралии [10, 11], Греции [12], Германии [13, 14], Китая [15], Нидерландов [16].

Список отечественных работ, описывающих применение метода ОАД для оценки эффективности вузов довольно ограничен, укажем лишь на работу [17]. Подробный обзор, посвященный оценке эффективности в высшем образовании, можно найти в работе [18].

В большинстве работ авторы строят оригинальные модификации модели ОАД, чтобы обеспечить ее работу в каждом конкретном случае. Однако их объединяет ряд общих признаков, например, выбор ресурсных и результативных параметров (см. таблицу).

Еще одно поле применения метода ОАД — анализ эффективности различных энергетических предприятий, примером служат работы [19, 20]. Авторы используют стандартную модель ОАД, а также разрабатывают оригинальные модификации для применения в конкретной ситуации.

Различные виды издержек предприятия (совокупные, операционные) часто выбираются в качестве единственного входного параметра. В качестве результативных показателей обычно принимают:

- количество сгенерированной энергии;
- количество переданной энергии;
- количество распределенной энергии;

### Критерии оценки в сфере высшего образования

Входные переменные	Входные переменные
Совокупные издержки	Качество исследовательской работы
Расходы на научные проекты	Число диссертаций
Численность студентов бакалавриата	Число статей в рецензируемых журналах
Численность студентов магистратуры	Численность выпускников
Численность студентов аспирантуры	Процент отчисленных студентов
Рейтинг научной деятельности	Успеваемость студентов
Численность сотрудников университета	Доходы от исследований



- общий объем производства;
- потребление топлива;
- длина сети.

В работах [21, 22] демонстрируется другой подход к анализу эффективности энергетического сектора — за основу берутся параметры, связанные с экологией:

- количество выброшенного в атмосферу углекислого газа (ресурс);
- потери электроэнергии (ресурс);
- эффективность использования природных ископаемых (результат).

В данном случае можно говорить о расчете экологической эффективности энергетических предприятий.

Мы перечислили лишь некоторые области применения модели ОАД, хотя данная методика также применяется для оценки эффективности банковского сектора, систем здравоохранения и городского транспорта, государственного управления и других отраслей экономики. Интересующегося читателя мы отсылаем к обзору [23].

## 2. НЕОДНОРОДНОСТЬ ПРИ ОЦЕНКЕ ЭФФЕКТИВНОСТИ

Рассмотрим причины, по которым может быть вызвана неоднородность. Главные из них, согласно авторам работы [24], это:

- различия в типе управления (государственный или частный);
- наличие государственных привилегий у некоторых фирм;
- особенности расположения оцениваемых объектов (город или сельская местность, плотность населения в месте расположения фирмы);
- различные цели организаций (для университетов — исследовательская и педагогическая направленности);
- различия в законодательстве.

Даже такой краткий список показывает, что существует довольно много причин, ведущих к тому, что выборка может быть неоднородной. Стандартная модель ОАД не способна учесть разнородность объектов.

В обозначениях рис. 1 приведем пример работы модели ОАД на неоднородном наборе альтернатив. В случае с единственным параметром входа и выхода существует возможность, что некоторое подмножество всех фирм будет сосредоточено «далеко» от обеих границ (множество из четырех безымянных точек на рис. 1). Пусть эти точки характеризуют, например, университеты. Экономически это означает, что данные четыре университета находятся в крайне неблагоприятной обстановке (например, из-за малой плотности населения в ок-

руге или из-за недостаточного государственного финансирования).

Применив ОАД, мы увидим, что значения эффективности для некоторых фирм ничтожно малы, что не отражает существующую ситуацию в отрасли. Таким образом, оценки оказываются несправедливо заниженными.

Рассмотрим решения, которые наиболее часто встречаются в работах, посвященных применению ОАД для оценки неоднородной выборки [3].

### 2.1. Разбиение выборки по категориям

Если значения переменной, отвечающей за неоднородность, могут быть упорядочены от меньшего к большему в зависимости от влияния на эффективность, то можно воспользоваться подходом, апробированным в работе [25]. Эффективность  $i$ -й фирмы оценивается только относительно тех фирм, для которых значение параметра разнородности не выше, чем у  $i$ -й фирмы.

Рассмотрим случай оценки эффективности ресторанов быстрого питания. Какие-то из них располагаются в центре города, какие-то — на окраинах, наконец, остальные располагаются в сельской местности. Ясно, что преимущество у тех, кто ближе к городскому центру. В этом случае оценка проводится в три этапа: 1) сравниваются только рестораны из сельской местности; 2) сравниваются рестораны из сельской местности и из пригородов; 3) наконец проводится оценка всей выборки, включая рестораны, расположенные в центре города. Эффективностью сельских ресторанов считаются значения, полученные на первом этапе, пригородных — на втором, расположенных в городских центрах — на третьем.

Таким образом, не сравниваются объекты, один из которых имеет конкурентное преимущество перед другим.

Отметим, что цель работы [25] состояла не в учете неоднородности выборки, а во внесении в стандартную модель ОАД коррективов с учетом невозможности контроля некоторых ресурсов или результатов. Однако нельзя не учитывать при оценке, что рестораны, находящиеся ближе к центру города, получают конкурентное преимущество, поэтому и был разработан описанный прием. В качестве иллюстрации в работе [25] проводится оценка шестидесяти ресторанов быстрого питания.

### 2.2. Проективный метод оболочечного анализа данных

Может случиться так, что не существует естественного упорядочения значений параметра среды в зависимости от влияния на эффективность. Примером такого параметра может служить тип уп-

равления — государственный или частный. Тогда применяется метод, предложенный в работе [26]. Он состоит из трех основных шагов:

— выборка делится на подвыборки относительно выбранного параметра неоднородности; внутри каждой подвыборки используется стандартная модель ОАД;

— находятся проекции всех фирм на соответствующие им границы эффективности;

— применяется стандартный метод ОАД для полученных на предыдущем шаге проекций.

Оба описанных метода [25, 26] обладают рядом ограничений и недостатков. Прежде всего, поскольку множество, с которым мы сравниваем объекты, уменьшается, это ведет к тому, что многие объекты станут эффективными. Далее, только одна переменная среды может быть включена в анализ. Оба метода требуют, чтобы переменную, отвечающую за неоднородность, можно было разбить на различные категории (это не всегда возможно, особенно при учете социальной, культурной или политической неоднородности). Наконец, первый из этих методов подразумевает, что «знак» влияния переменной на эффективность известен априори, что не всегда верно.

Поэтому область применения данных методов довольно сильно ограничена.

### 2.3. Включение неоднородностей в задачу линейного программирования

Другой возможный метод — включить параметры среды непосредственно в задачу линейного программирования, составляющую суть метода ОАД [27, 28]. Это возможно сделать либо зная, какие переменные имеют отрицательное влияние на эффективность, а какие — положительное, либо не зная «знаков» влияния переменных среды на эффективность.

Рассмотрим первый вариант. Пусть имеется  $K$  параметров среды, имеющих положительное влияние на эффективность оцениваемой выборки,  $z_i = (z_{1i}, \dots, z_{ki})$  — вектор значений этих переменных для  $i$ -й фирмы,  $Z$  —  $(K \times L)$ -матрица значений параметров среды для всех фирм выборки. Модель ОАД в данном случае примет форму:

найти  $\min_{\lambda, \theta_i} \theta_i$  при ограничениях

$$-q_i + Q\lambda \geq 0;$$

$$\theta_i x_i - X\lambda \geq 0;$$

$$z_i - Z\lambda \geq 0;$$

$$\lambda \geq 0,$$

где все обозначения взяты из выражений (3), (4).

Если  $K$  переменных среды имеют отрицательное влияние на значения эффективности, то похожим образом можно включить их в ОАД:

найти  $\max_{\lambda, \theta_i} \theta_i$  при ограничениях

$$-q_i + Q\lambda \geq 0;$$

$$\theta_i x_i - X\lambda \geq 0;$$

$$-z_i + Z\lambda \geq 0;$$

$$\lambda \geq 0.$$

Если имеются переменные среды, влияющие на эффективность как положительно, так и отрицательно, то итоговая формулировка задачи ОАД представляет собой объединение этих двух задач линейного программирования.

Если тип влияния параметров среды на эффективность исследуемой выборки неизвестен, то предлагается сформулировать задачу линейного программирования в виде:

найти  $\min_{\lambda, \theta_i} \theta_i$  при ограничениях

$$-q_i + Q\lambda \geq 0;$$

$$\theta_i x_i - X\lambda \geq 0;$$

$$z_i - Z\lambda = 0;$$

$$\lambda \geq 0.$$

К недостаткам данного подхода можно отнести то, что параметры среды должны быть непрерывны. Кроме того, дополнительные ограничения существенно увеличивают значения эффективности для всех фирм.

Наконец, для определения знаков влияния параметров среды на эффективность можно воспользоваться эконометрическими методами. Включение категориальных переменных непосредственно в задачу линейного программирования также возможно, по этому поводу см. работу [29].

### 2.4. Кластерный метод

Существует и другой метод, главная идея которого — соединить кластерный анализ, нейронные сети и ОАД [30—37], применяются две его версии. Согласно одной из них необходимо разбить исследуемую выборку на кластеры в зависимости от полученных значений эффективности, затем применить стандартный ОАД к кластерам. В соответствии с другой версией объекты сначала разбиваются на кластеры в зависимости от значений входных и выходных переменных. После этого стандартный метод ОАД применяется для каждого кластера отдельно.

Рассматривая более подробно описанные методы, становится ясно, что все они связаны либо



с разбиением исходной выборки на кластеры, либо с четким определением параметров, которые непосредственно воздействуют на эффективность. Без сомнения, представленные техники хорошо зарекомендовали себя на практике, однако они не учитывают двух обстоятельств. Прежде всего, возможна ситуация, когда сложно выделить конкретные параметры среды. Такое происходит, когда необходимо учесть влияние на эффективность сложных социальных или политических процессов. Описанные методы не годятся для этого случая.

Кроме того, рассмотренные способы учета неоднородности не подразумевают, что исследуемые объекты могут иметь различные цели в рамках одной отрасли. Прекрасный пример ситуации такого рода — университеты, одни из которых могут быть нацелены на проведение исследовательской деятельности, другие — на обучение специалистов в определенной области.

Далее рассмотрим метод, позволяющий учесть возникающую из-за разности в приоритетах разнородность.

### 3. МНОГОЗАДАЧНЫЙ БОЛОЧЕЧНЫЙ АНАЛИЗ ДАННЫХ

Данный метод впервые был представлен в работе [7] и применялся для оценки эффективности английских университетов. Рассмотрим простейший случай двух возможных приоритетов.

Пусть по-прежнему имеется  $N$  ресурсов,  $M$  результатов и  $L$  фирм. Пусть  $x^1 \subset \{x_1, \dots, x_N\}$  — ресурсы, используемые только для достижения первой цели,  $x^{12}$  — ресурсы, используемые для достижения двух целей,  $x^2$  — ресурсы, годные только для второй цели. Очевидно, что  $\{x_1, \dots, x_N\} = x^1 \sqcup x^{12} \sqcup x^2$ , где  $\sqcup$  — квантор несвязного объединения. Аналогичным образом множество результативных параметров разбивается на три группы  $\{q_1, \dots, q_M\} = q^1 \sqcup q^{12} \sqcup q^2$ .

Для  $j$ -го ресурса из множества  $x^{12}$  ведем коэффициент  $\alpha_j$  — вес, с которым данный ресурс используется для достижения первой цели. Соответственно,  $1 - \alpha_j$  — вес, с которым ресурс  $x_j$  используется для достижения второй цели. Аналогично, для  $j$ -го результата из множества  $q^{12}$  вводится коэффициент  $\beta_j$ .

Таким образом, формулируется следующая задача математического программирования:  
найти

$$\max_{u, v, \alpha, \beta} \gamma_i^1 \theta_i^1 + \gamma_i^2 \theta_i^2 \quad (6)$$

при ограничениях

$$\frac{\sum_{\{j|q_j \in q^1\}} u_j q_{ji} + \sum_{\{j|q_j \in q^{12}\}} u_j \beta_j q_{ji}}{\sum_{\{j|x_j \in x^1\}} v_j x_{ji} + \sum_{\{j|x_j \in x^{12}\}} v_j \alpha_j x_{ji}} \leq 1, \quad i = 1, \dots, L;$$

$$\frac{\sum_{\{j|q_j \in q^1\}} u_j q_{ji} + \sum_{\{j|q_j \in q^{12}\}} u_j (1 - \beta_j) q_{ji}}{\sum_{\{j|x_j \in x^2\}} v_j x_{ji} + \sum_{\{j|x_j \in x^{12}\}} v_j (1 - \alpha_j) x_{ji}} \leq 1, \quad i = 1, \dots, L;$$

$$u \geq 0; \quad v \geq 0, \quad (7)$$

где  $u$  —  $(M \times 1)$ -вектор весовых коэффициентов для выходных результатов,  $v$  —  $(N \times 1)$ -вектор весовых констант для ресурсных параметров,  $\alpha$  —  $(|x^{12}| \times 1)$ -вектор коэффициентов, определяющих распределение ресурсных параметров по двум приоритетам,  $\beta$  —  $(|q^{12}| \times 1)$ -вектор коэффициентов, определяющих влияние каждого вида деятельности на результаты,  $\theta_i^1 \in (0, 1]$  — значение эффективности  $i$ -й фирмы по первому направлению,  $\theta_i^2 \in (0, 1]$  — аналогичная величина для второго направления деятельности,  $\theta_i$  — средняя эффективность. Величины  $\gamma_i^1$  и  $\gamma_i^2$  определяют важность каждого из двух направлений деятельности для  $i$ -й фирмы.

Отметим, что задача (6), (7) представляет собой уже существенно нелинейную оптимизационную задачу из-за коэффициентов, определяющих важность направлений развития фирм.

В такой постановке величины  $\gamma_i^1$  и  $\gamma_i^2$  должны быть выбраны заранее и удовлетворять условию  $\gamma_i^1 + \gamma_i^2 = 1$ . Данный подход работает, если априори известна стратегия фирм, принадлежащих выборке. Однако исследователь не всегда располагает такой информацией, в этом случае необходимо добавить к ограничениям (7) тождества

$$\gamma_i^1 = \frac{\sum_{\{j|x_j \in x^{12}\}} v_j \alpha_j x_{ji}}{\sum_{\{j|x_j \in x^{12}\}} v_j x_{ji}},$$

$$\gamma_i^1 + \gamma_i^2 = 1, \quad (8)$$

здесь индекс  $i$  фиксирован в соответствии с формулой (6).

Таким образом, согласно выражению (8), приоритеты  $i$ -й фирмы распределяются в соответствии с тем, как эта фирма распределяет ресурсы между ее целями.

Данная модель предлагает решение проблемы оценки эффективности при неоднородности выборки, связанной с наличием разных целей у разных объектов. Она нашла широкое применение в литературе, (см., например, [38—41]).

#### 4. МЕТОД ПОСЛЕДОВАТЕЛЬНОГО ИСКЛЮЧЕНИЯ АЛЬТЕРНАТИВ

Рассмотрим модель, предложенную в работе [5]. Основное ее отличие от рассмотренных заключается в абстрагировании от конкретных показателей, отвечающих за влияние среды. Предлагается каждой выборке поставить в соответствие число  $\mu \in (0, 1]$ , интерпретирующее неоднородность, т. е. чем ближе данное число к единице, тем исследуемая выборка более неоднородная. Данный параметр может быть найден с помощью простого алгоритма, предложенного в работе [5], однако можно воспользоваться экспертной оценкой.

##### 4.1. Простейшие случаи

Рассмотрим работу алгоритма в простейшем случае, когда отрасль характеризуется единственным ресурсным и единственным результативным параметром. В данном случае  $i$ -я фирма может быть представлена как точка  $(x_i, q_i) \in \mathbb{R}^2$ . Геометрическая интерпретация алгоритма оценки фирм  $F_1, \dots, F_6$  представлена на рис. 3.

Согласно модели ОАД с постоянной отдачей от масштаба,  $F_1$  — единственная эффективная фирма. На первом этапе алгоритма в обычном геометрическом смысле рассчитывается центр масс всей выборки — точка  $B_1 = (\bar{x}, \bar{q})$ , где черта означает среднее значение. Новая граница эффективности строится с помощью введенного коэффи-

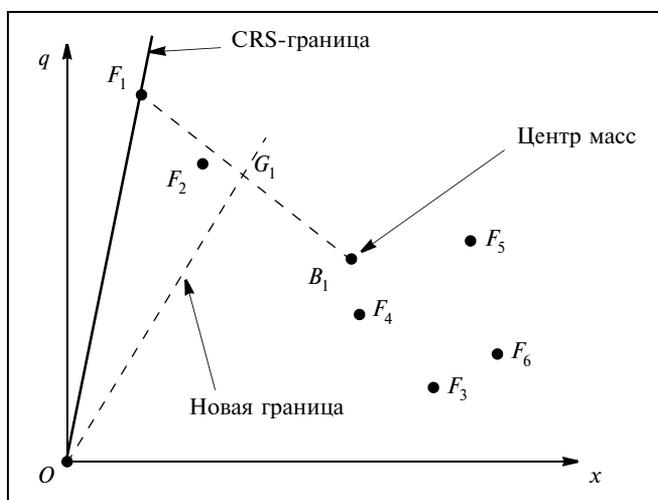


Рис. 3. Графическая интерпретация алгоритма в случае  $N = M = 1$

циента неоднородности  $\mu$ . Для этого вычисляются координаты фирмы, генерирующей новую границу  $G_1 = \mu B_1 + (1 - \mu) F_1$ .

После того, как новая граница сконструирована, относительно нее оцениваются фирмы, которые показали меньшую эффективность, чем генерирующая фирма  $G_1$ . На рис. 3 роль таких фирм играют фирмы  $F_3, \dots, F_6$ . В данном случае фирму  $F_2$  невозможно оценить относительно фирмы  $G_1$ , так как компания  $F_2$  более эффективна. Для того, чтобы произвести оценку  $F_2$ , уже оцененные с помощью новой границы фирмы  $F_3, \dots, F_6$  удаляются из рассмотрения. Строится новый центр масс  $B_2$  и формируется генерирующая условная фирма  $G_2$ , оценивается фирма  $F_2$ .

Алгоритм останавливается тогда, когда оценены все неэффективные в смысле стандартного ОАД фирмы. В случае, представленном на рис. 3, потребуется сделать всего два шага. По построению сходимость алгоритма гарантирована, кроме того, эффективными окажутся лишь те фирмы, которые были эффективными в смысле стандартной ОАД-модели.

Существует несложная связь между оценками эффективности, полученными с помощью модели последовательного исключения и с помощью стандартного метода ОАД. Предположим, некоторое множество фирм на данном шаге алгоритма оценивается при помощи вспомогательной фирмы  $G_i$ , где  $i$  — номер шага алгоритма. Пусть компания  $F$  принадлежит данному множеству фирм, тогда

$$E_F^{CRS} = E_{G_i}^{CRS} E_F^{New}, \quad (9)$$

где нижний индекс отвечает за фирму, а верхний — за метод оценки эффективности. Формула (9) немедленно следует из геометрической интерпретации модели ОАД, представленной на рис. 1.

В соответствие с формулой (9), построенный алгоритм оценивает неэффективные фирмы менее строго, чем стандартный ОАД (это ясно из рис. 3). Более того, чем фирма эффективнее в смысле стандартного ОАД, тем строже ее оценивает алгоритм последовательного исключения. Это следует из того, что  $E_{G_i}^{CRS}$  увеличивается с увеличением числа проделанных шагов  $i$ , так как чем ближе алгоритм к завершению, тем больше неэффективных фирм было удалено из рассмотрения.

Рассмотрим, как работает алгоритм в случае с двумя входными параметрами и одним показателем ресурса (рис. 4).

Здесь  $F_1, F_2$  — эффективные в смысле стандартного ОАД фирмы. Алгоритм работает похожим

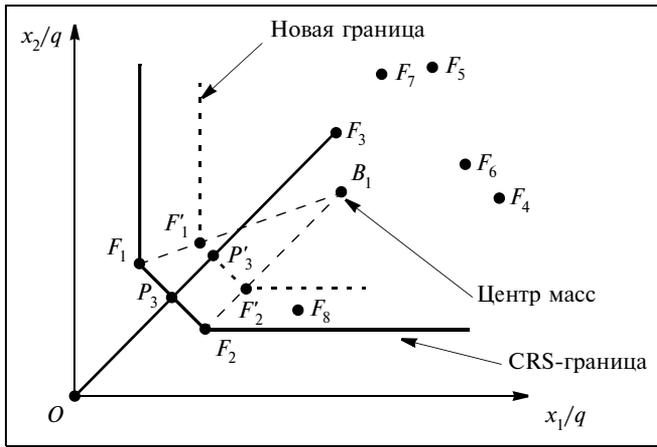


Рис. 4. Графическая интерпретация алгоритма в случае  $N = 2, M = 1$

образом, но теперь сдвигать к центру масс необходимо не одну фирму как на рис. 3, а две:

$$F'_1 = \mu B_1 + (1 - \mu)F_1,$$

$$F'_2 = \mu B_1 + (1 - \mu)F_2.$$

Эффективность фирмы  $F_3$  в новой модели рассчитывается как отношение  $|OP'_3|/|OF_3|$ . Все фирмы последовательно оцениваются относительно новых границ эффективности, полученных с помощью центра масс. Алгоритм также сходится для любой выборки и сохраняет множество эффективных фирм.

#### 4.2. Произвольное число переменных

В данном случае  $i$ -я фирма характеризуется с помощью вектора входных переменных модели  $x_i = (x_{1i}, \dots, x_{Ni})$  и вектора результатов  $q_i = (q_{1i}, \dots, q_{Mi})$ . Как и ранее, мы определяем ресурсную и результитивную части центра масс анализируемой выборки  $b_x = (\bar{x}_1, \dots, \bar{x}_N)$  и  $b_q = (\bar{q}_1, \dots, \bar{q}_M)$ , где, как обычно, черта означает среднее значение параметра.

Пусть оцениваемая выборка определена как множество индексов  $l = \{1, \dots, L\}$ . Обозначим группу 100 %-эффективных в смысле стандартной ОАД модели фирм как  $I_e = \{i_1, \dots, i_S\} \subset I$ . Напомним, что  $N$  и  $M$  — числа входных и выходных параметров соответственно. Пусть  $X_e$  означает матрицу входных параметров всех эффективных фирм из множества  $I_e$  (ее размер  $N \times S$ ), и  $Q_e$  — аналогичная матрица размера  $M \times S$ , содержащая параметры результата всех фирм из  $I_e$ . Также вводится матрица

$$B_x^i = \|b_x^T, \dots, b_x^T, x_i^T\|,$$

где  $b_x^T$  представляет собой транспонированную ресурсную часть центра масс, она повторяется  $S$  раз,  $x_i$  — вектор ресурсов какой-нибудь неэффективной компании, т. е.  $i \in I \setminus I_e$ . Схожим образом определяется матрица размера  $M \times (S + 1)$

$$B_q^i = \|b_q^T, \dots, b_q^T, q_i^T\|,$$

где  $b_q^T$  представляет собой транспонированную результитивную часть центра масс, а  $q_i$  — выходной вектор  $i$ -й неэффективной фирмы, т. е.  $i \in I \setminus I_e$ .

Пусть  $X_e^i$  и  $Q_e^i$  будут матрицы  $X_e$  и  $Q_e$  с одной добавленной к ним справа колонкой  $x_i^T$  и  $q_i^T$  соответственно. Поскольку главная идея данного метода заключается в том, чтобы сдвинуть границу эффективности в сторону центра масс, сформируем две матрицы

$$X_i = \mu B_x^i + (1 - \mu)X_e^i$$

$$\text{и } Q_i = \mu B_q^i + (1 - \mu)Q_e^i. \quad (10)$$

Заметим, что матрицы(10) определены только для неэффективных фирм, т. е.  $i \in I \setminus I_e$ . Последними колонками матриц  $X_i$  и  $Q_i$  являются колонки  $x_i^T$  и  $q_i^T$  соответственно.

Можно сформулировать алгоритм последовательного исключения альтернатив в общем случае. Его первый шаг заключается в решении задачи линейного программирования для всех неэффективных фирм  $i \in I \setminus I_e$ :

$$\min_{\lambda, \theta_i} \theta_i \quad (11)$$

при ограничениях

$$-q_i + Q_i \lambda \geq 0;$$

$$\theta_i^* x_i - X_i \lambda \geq 0;$$

$$\lambda \geq 0,$$

где матрицы  $X_i$  и  $Q_i$  определены выражением (10),  $\lambda$  —  $((S + 1) \times 1)$ -вектор неизвестных констант,  $x_i$  и  $q_i$  — векторы входных и выходных параметров для  $i$ -й неэффективной фирмы. Наконец,  $\theta_i^*$  представляет собой оценку эффективности  $i$ -й фирмы с поправкой на неоднородность оцениваемой выборки.

При вычислении значения  $\theta_i^*$  может быть два различных исхода. А именно, оно может быть строго меньше единицы и может быть равно единице.

Если  $i$ -я неэффективна в смысле стандартной ОАД модели фирма получает в результате нахождения минимума (11) значение  $\theta_i^*$ , мы исключаем такую фирму из рассмотрения. Алгоритм можно сформулировать следующим образом.

*Шаг 1.* Вычисляется центр масс всей выборки.

*Шаг 2.* Для всех  $i \in I \setminus I_e$  вычисляются матрицы  $X_i$  и  $Q_i$ .

*Шаг 3.* Для всех неэффективных компаний вычисляются новые оценки эффективности согласно модели (11), (12). Все фирмы, получающие значения  $\theta_i^* < 1$  исключаются из рассмотрения.

*Шаг 4.* Если какие-то из фирм получают оценку эффективности  $\theta_i^* = 1$ , то для них процедура повторяется, начиная с шага 1.

На данный момент остался неучтенным лишь один случай. Теоретически возможна ситуация, когда матрицы (10) организованы таким образом, что при запуске модели (11), (12) все неэффективные компании получают значения  $\theta_i^* = 1$ . Это означает, что изначальная граница эффективности, определенная с помощью стандартной модели DEA, отодвинута слишком далеко. В таком случае необходимо уменьшить значение показателя неоднородности  $\mu$  и начать процедуру сначала. В качестве алгоритмического решения новое значение показателя неоднородности может вычисляться, например, как  $\mu^2$ .

Отметим важные свойства построенной модели. Прежде всего, сходимость алгоритма обеспечена по построению для любой выборки. Далее, множество  $I_e$  эффективных фирм по отношению к стандартному ОАД не меняется. Наконец, хотя в многомерном случае и не удается сохранить свойство (9), однако его прямой аналог сохраняется. Поскольку на каждом этапе мы используем модель ОАД, то можно вычислить координаты проекции каждой неэффективной фирмы из множества  $i \in I \setminus I_e$  на соответствующую границу эффективности, определенную с помощью модели (11), (12). Тогда

$$E_F^{CRS} = E_P^{CRS} E_F^{New},$$

где  $E_F^{CRS}$  — эффективность фирмы  $F$  в стандартной модели ОАД,  $E_P^{CRS}$  — эффективность проекции  $P$  фирмы  $F$  на границу, определенную с помощью модели (11), (12). Наконец,  $E_F^{New}$  — оценка эффективности фирмы  $F$  согласно построенному алгоритму.

В качестве иллюстрации модели в работе [5] приведен расчет эффективности пилотной выборки, состоящей из 29-ти российских университетов.

## ЗАКЛЮЧЕНИЕ

Рассмотрена модель оболочечного анализа данных и ее модификации для случая, когда оцениваемая выборка неоднородна. Представлен обзор основных подходов к определению неоднородности. В частности, рассмотрены четыре метода, основанные на знании параметров среды, которые оказывают существенное влияние на эффективность. Такие методы могут лишь ограниченно применяться для оценки эффективности неоднородного множества альтернатив, поскольку не всегда удается выделить круг причин, вызывающих неоднородность.

Подход, предложенный в работе [5], базируется на том, что неоднородность возникает из-за разницы в приоритетах оцениваемых компаний. Как показывает многочисленная практика, с этим аспектом неоднородности модель справляется успешно. Однако разница в целях далеко не единственная причина неоднородности выборки.

Наконец, рассмотрен наиболее общий подход к экзогенному определению неоднородности [7]. Предложенный алгоритм основан на последовательном исключении неэффективных альтернатив и позволяет исключить влияние неоднородности на оценку эффективности в наиболее общем виде.

Автор признателен Международной лаборатории анализа и выбора решений НИУ — ВШЭ за финансовую поддержку при проведении исследований и особо благодарен Ф.Т. Алескерову за ценные советы и полезные комментарии.

## ЛИТЕРАТУРА

1. Charnes A., Cooper W., Rhodes E. Measuring the Efficiency of Decision Making Units // European Journal of Operational Research. — 1978. — N 2. — P. 429–444.
2. Farrell M.J. The Measurement of Productive Efficiency // Journal of the Royal Statistical Society, Series A. — 1957. — N 3. — P. 253–290.
3. Coelli T.J., Rao D.S.P., O'Donnell C.J., Battese G.E. An Introduction to Efficiency and Productivity Analysis. — N.-Y.: Springer, 2005.
4. Banker R.D., Charnes A., Cooper W. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, Management Science. — 1984. — N 30. — P. 1078–1092.
5. Алескеров Ф., Петрущенко В. DEA by sequential exclusion of alternatives // HSE Working Paper WP7. — 2013. — N 02.
6. Athanassopoulos A.D., Shale E. Assessing the Comparative Efficiency of Higher Education Institutions in the UK by the Means of Data Envelopment Analysis, Education Economics. — 1997. — N 5. — P. 117–134.



7. *Beasley J.E.* Determining teaching and research efficiencies, *Journal of the Operational Research Society*. — 1995. — N 46. — P. 441–452.
8. *Flegg A., Allen D., Field K., Thurlow T.* Measuring the Efficiency and Productivity of British Universities: An Application of DEA and the Malmquist Approach, unpublished, 2004.
9. *Johnes J.* Data envelopment analysis and its application to the measurement of efficiency in higher education // *Economics of Education Review*. — 2006. — N 25. — P. 273–288.
10. *Abbott M., Doucouliagos C.* The efficiency of Australian universities: a data envelopment analysis // *Economics of Education Review*. — 2003. — N 22. — P. 89–97.
11. *Avkiran N.K.* Investigating technical and scale efficiencies of Australian Universities through data envelopment analysis // *Socio-Economic Planning Sciences*. — 2001. — N 35. — P. 57–80.
12. *Katharakis G., Katharaki M.* A comparative assessment of Greek universities' efficiency using quantitative analysis // *Intern. Journal of Educational Research*. — 2010. — N 49. — P. 115–128.
13. *Kempkesa G., Pohl C.* The efficiency of German universities — some evidence from nonparametric and parametric methods // *Applied Economics*. — 2010. — N 42. — P. 2063–2079.
14. *Warning S.* Performance Differences in German Higher Education: Empirical Analysis of Strategic Groups // *Review of Industrial Organization*. — 2004. — N 24. — P. 393–408.
15. *Johnes J., Yu L.* Measuring the research performance of Chinese higher education institutions using data envelopment analysis // *China Economic Review*. — 2008. — N 19. — P. 679–696.
16. *Cherchye L., Vanden Abeele P.* On research efficiency A micro-analysis of Dutch university research in *Economics and Business Management // Research Policy*. — 2005. — N 34. — P. 495–516.
17. *University efficiency evaluation with using its reputational component / I.V. Abankina, F.T. Aleskerov, V.Y. Belousova, et al. // Lecture Notes in Management Science*. — 2012. — N 4. — P. 244–253.
18. *Оценка результативности университетов с помощью Обобщенного Анализа Данных / И. Абанкина, Ф. Алескеров, В. Белоусова и др. // Вопросы образования*. — 2013. — N 2 (в печати).
19. *Farsi M., Filippini M.* A benchmarking analysis of electricity distribution utilities in Switzerland // *CEPE working paper*. — 2005. — N 43.
20. *Goto M., Tsutsui M.* A Multi-division Efficiency Evaluation of U.S. Electric Power Companies Using a Weighted Slacks-based Measure // *Socio-Economic Planning Sciences*. — 2009. — N 43. — P. 201–208.
21. *Meyer R.* Benchmarking Economies of Vertical Integration in U.S. Electricity Supply: An Application of DEA // *Competition and Regulation in Network Industries*. — 2011. — N 12. — P. 299.
22. *Vaninsky A.* Environmental Efficiency of Electric Power Industry of the United States: A Data Envelopment Analysis Approach, *World Academy of Science // Engineering and Technology*. — 2008. — N 40. — P. 584–590.
23. *Murillo-Zamorano L. R.* Economic efficiency and frontier techniques // *Journal of Economic Surveys*. — 2004. — N 18. — P. 33–77.
24. *Fried H.O., Schmidt S.S., Yaisawamg S.* Incorporating the Operating Environment into a Nonparametric Measure of Technical Efficiency // *Journal of Productivity Analysis*. — 1999. — N 12. — P. 249–267.
25. *Banker R.D., Morey R.C.* Efficiency Analysis for Exogenously Fixed Inputs and Outputs // *Operations Research*. — 1986. — N 34. — P. 513–521.
26. *Charnes A., Cooper W., Rhodes E.* Evaluating Program and Managerial Efficiency: An Application of Data Envelopment Analysis to Program Follow Through // *Management Science*. — 1981. — N 27. — P. 668–697.
27. *Bessent A.M., Bessent E.W.* Comparing the Comparative Efficiency of Schools through Data Envelopment Analysis // *Educational Administration Quarterly*. — 1980. — N 16. — P. 57–75.
28. *Ferrier G.D., Lovell C.A.K.* Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence // *Journal of Econometrics*. — 1990. — N 46. — P. 229–245.
29. *Banker R.D., Morey R.C.* The Use of Categorical Variables in Data Envelopment Analysis // *Management Science*. — 1986. — N 32. — P. 1613–1627.
30. *Hirschberg J.G., Lye J.N.* Clustering in a Data Envelopment Analysis Using Bootstrapped Efficiency Scores, Department of Economics // *Working Papers Series 800 The University of Melbourne*, 2001.
31. *Lemos C.A.A., Lima M.P., Ebecken N.F.F.* DEA Implementation and Clustering Analysis using the K-means algorithm, *Data Mining VI — Data Mining // Text Mining and Their Business Applications*. — 2005. — N 1. — P. 321–329.
32. *Use of data envelopment analysis and clustering in multiple criteria optimization / M. Marroquin, M. Pena, C. Castro, et al. // Intelligent Data Analysis*. — 2008. — N 12. — P. 89–101.
33. *Meimand M., Cavana R.Y., Laking R.* Using DEA and survival analysis for measuring performance of branches in New Zealand's Accident Compensation Corporation // *Journal of the Operational Research Society*. — 2002. — N 53. — P. 303–313.
34. *Samoilenko S., Osei-Bryson K.M.* Determining sources of relative inefficiency in heterogeneous samples: Methodology using Cluster Analysis, DEA and Neural Networks // *European Journal of Operational Research*. — 2010. — N 206. — P. 479–487.
35. *Schreyogg J.C., von Reitzenstein C.* Strategic groups and performance differences among academic medical centers // *Health Care Management Review*. — 2008. — N 33. — P. 225–233.
36. *Sharma M.J., Yu S.J.* Performance based stratification and clustering for benchmarking of container terminals // *Expert Systems with Applications*. — 2009. — N 36. — P. 5016–5022.
37. *Shin H.W., Sohn S.Y.* Multi-attribute scoring method for mobile telecommunication subscribers // *Expert Systems with Applications*. — 2004. — N 26. — P. 363–368.
38. *Salerno C.* Using Data Envelopment Analysis to Improve Estimates of Higher Education Institution's Per-student Education Costs // *Education Economics*. — 2006. — N 14. — P. 281–295.
39. *Tsai P.F., Mar Molinero C.* A Variable Returns to Scale Data Envelopment Analysis Model for the Joint Determination of Efficiencies with an Example of the UK Health Service // *European Journal of Operational Research*. — 2002. — N 141. — P. 21–38.
40. *Yu M.* Measuring the efficiency and return to scale status of multi-mode bus transit — evidence from Taiwan's bus system // *Applied Economics Letters*. — 2008. — N 15. — P. 647–653.
41. *Cinar Y.* Research and Teaching Efficiencies of Turkish Universities with Heterogeneity Considerations: Application of «Multi-Activity DEA» and «DEA by Sequential Exclusion of Alternatives' Models». — 2013. — HSE Working Paper WP7. — N 04.

Статья представлена к публикации членом редколлегии Ф.Т. Алескеровым.

**Всеволод Владимирович Петрущенко** — стажер-исследователь, Научно-исследовательский университет — Высшая школа экономики, г. Москва,  
✉ goroddt@yandex.ru.