

## Методы оценки показателя эффективности в моделях стохастической производственной границы

Малахов Д.И., Пильник Н.П.

В работе рассматривается вопрос об оценивании эффективности фирм отрасли. В настоящий момент наиболее известными эмпирическими подходами для анализа эффективности являются DEA (*data envelopment analysis*) и SFA (*stochastic frontier analysis*), причем популярность последнего довольно быстро растет. Многообразие SFA-моделей на практике приводит к тому, что, как правило, авторами работ по данной тематике заранее выбирается конкретная модель, на основе которой и делаются окончательные выводы относительно эффективности отрасли или отдельных предприятий. Обзор и совместный анализ методов и моделей оценки стохастической границы производственных возможностей до настоящего времени проведен не был, что и составляет одну из задач настоящего исследования.

В работе описаны 15 наиболее популярных моделей в области SFA-моделирования, обсуждаются их проблемы, особенности оценки и перспективы развития. Отдельно рассмотрены и модели, использующие метод максимального правдоподобия, и модели, использующие метод моментов, также будут отмечены плюсы и минусы от оценивания SFA-моделей классическим методом моментов на реальных данных, так как этот метод не получил распространения на практике, хотя оценки, полученные этим методом, состоятельны и асимптотически нормальны.

В силу отсутствия универсального способа сравнения этих моделей в работе приведены и сопоставлены результаты оценки девяти SFA-моделей на едином массиве данных по российской цементной промышленности. Проведен корреляционный анализ расчетных показателей эффективности и объяснены причины наиболее существенных различий в результатах.

**Ключевые слова:** стохастическая производственная граница; эконометрическое моделирование; эффективность; обзор; метод моментов; бетонная промышленность.

---

Авторы выражают благодарность за полезные комментарии Ершову Э.Б., Левину М.И., Назрулаевой Е.Ю., Пекарскому С.Э., Пересецкому А.А.

Исследование осуществлено в рамках программы фундаментальных исследований НИУ ВШЭ в 2013 г.

**Малахов Дмитрий Игоревич** – студент 2 курса магистратуры факультета экономики НИУ ВШЭ. E-mail: d.malakhov@gmail.com

**Пильник Николай Петрович** – к.э.н., старший преподаватель кафедры математической экономики и эконометрики НИУ ВШЭ. E-mail: u4d@yandex.ru

Статья поступила в Редакцию в сентябре 2013 г.

## Введение

В настоящий момент наиболее известными эмпирическими подходами для анализа эффективности являются DEA (*data envelopment analysis*) и SFA (*stochastic frontier analysis*), причем популярность последнего довольно быстро растет. Количество разнообразных моделей в рамках SFA довольно велико. При этом, как правило, авторами работ по данной тематике заранее выбирается конкретная модель, на основе которой и делаются окончательные выводы относительно эффективности<sup>1</sup> отрасли или отдельных предприятий. Обзор и совместный анализ методов и моделей оценки стохастической границы производственных возможностей, насколько нам известно, до настоящего времени проведен не был, что и составляет одну из задач настоящего исследования.

В работе будут рассмотрены и модели, использующие метод максимального правдоподобия, и модели, использующие метод моментов, также отдельно будут отмечены плюсы и минусы от оценивания SFA-моделей классическим методом моментов на реальных данных, так как этот метод не получил распространения на практике, хотя обладает достаточно «хорошими» свойствами (состоятельность и асимптотическая нормальность). Теоретические сравнения данных методов приведены, например, в работах [Greene, 2008; Kumbhakar, Lovell, 2000].

Важно отметить, что на сегодняшний день нет однозначного и корректного способа сравнения разных SFA-моделей [Greene, 2008]. Сама оценка нескольких моделей параллельно встречается достаточно редко [Kumbhakar et al., 2011; Greene, 2005]. В этой связи возникает потребность в систематизации последних достижений в области SFA-моделирования. В качестве иллюстрации применения методов в данной работе будут описаны девять SFA-моделей на едином массиве данных по российской цементной промышленности.

Раздел 1 посвящен описанию теоретических понятий производственной границы и эффективности, в разделе 2 рассматриваются теоретические положения SFA-моделей, оцениваемых методом максимального правдоподобия, в разделе 3 – SFA-модели, оцениваемые методом моментов. В разделе 4 описаны используемые в качестве иллюстрации данные. Результаты оценки описанных ранее моделей представлены в разделе 5, и, наконец, раздел 6 содержит выводы.

### 1. Граница производственных возможностей: основные понятия

Пионером в области определения эффективности является Дебре [Debreu, 1951], который ввел понятие эффективности выпуска. Далее эту тему развил Фаррел [Farrel, 1957], введя понятие аллокативной эффективности. Сегодня выделяют следующие типы эффективности: технологическая, экономическая и аллокативная. При этом анализируют обычно техническую эффективность. Аллокативная эффективность отвечает за то, на

---

<sup>1</sup> Зачастую в SFA-моделях идет смешение понятий «эффективность» и «неэффективность». Данные термины являются взаимозаменяемыми, так как по сути отражают значения одной и той же величины.

сколько эффективно фирма размещает (закупает и распределяет) ресурсы, а техническая – насколько эффективно их использует:

*Экономическая эффективность = Техническая эффективность + Аллокативная эффективность.*

Подробнее об экономической и аллокативной эффективности см. работу [Kumbhakar, Lovell, 2000].

Технологическую эффективность формально определяют следующим образом: это возможность использовать минимальный вектор ресурсов для производства заданного вектора выпуска или, наоборот, получить максимальный вектор выпуска при заданных ресурсах. Формальное определение выглядит так: «Вектор выпуска-ресурсов  $(y, x) \in GR$  признается технологически эффективным, только если  $\forall (y', x') : (y', -x') \geq (y, -x)$   $(y', x') \notin GR$ , где  $GR$  задает множество всех возможных комбинаций выпуска-ресурсов» [Kumbhakar, Lovell, 2000]. Именно технологическая эффективность по умолчанию используется в SFA-моделях и будет подразумеваться везде далее.

В качестве описания множества возможных вариантов затрат и выпусков, как правило, используется неравенство  $q_{it} \leq f(x_{it}), i = 1, \dots, I, t = 1, \dots, N$ , где  $x_{it}$  – вектор ресурсов, использованный фирмой  $i$  за период  $t$ , а  $q_{it}$  – соответствующий вектор выпусков. Функция  $f$ , одинаковая для всех фирм во все моменты времени, описывает границу производственных возможностей. Поскольку функционирование фирмы на границе не является обязательным условием, поэтому саму производственную функцию записывают как равенство  $q_{it} = f(x_{it}) - U_{it}$ , где  $U_{it}$  – разница между выпуском при наиболее эффективном использовании заданного набора ресурсов и фактическим выпуском. В рамках предположения о виде производственной функции (обычно используется форма Кобба – Дугласа) последнее соотношение можно переписать как  $\ln q_{it} = \ln x_{it}\beta - u_{it}$ . Тогда показатель технической эффективности фирмы  $i$  за период  $t$  может быть рассчитан как

$$TE_{it} = \frac{q_{it}}{\exp(x_{it}\beta)} = \exp(-u_{it}) \in [0; 1].$$

Тем не менее вполне естественно предположить, что не все отклонения от границы производственных возможностей обусловлены деятельностью фирмы. Влияние внешних факторов описывается за счет случайной составляющей  $v_{it}$ :  $\ln q_{it} = \ln x_{it}\beta - u_{it} + v_{it}$ . В этом случае определение технической эффективности сохраняется, а отклонение от границы производственных возможностей вычисляется как  $\varepsilon_{it} = v_{it} - u_{it}$ .

Подробнее о разных типах производственных границ и их свойствах можно прочитать в работе [Kumbhakar, Lovell, 2000].

## 2. Оценивание моделей методом максимального правдоподобия

### 2.1. Базовая модель

В полном соответствии с микроэкономической теорией будем считать, что выпуск фирмы описывается производственной функцией  $q_i = f(z_i, \beta)$ , где  $q_i$  – выпуск  $i$ -й фирмы;  $z_i$  – факторы, влияющие на выпуск, а  $\beta$  – их коэффициенты. Это некая теоретическая идеальная зависимость. На практике же фирмы обычно не могут максимально эффективно использовать все свои ресурсы и произвести максимально возможный объем выпуска, поэтому производственная функция выглядит следующим образом:  $q_i = f(z_i, \beta)\eta_i$ , где  $\eta_i$  – степень эффективности  $i$ -й фирмы. Причем  $\eta_i$  лежит в интервале  $(0; 1]$ , где 1 соответствует абсолютной эффективности производства. Предположим также, что существуют экзогенные шоки в экономике, разные для разных фирм:  $q_i = f(z_i, \beta)\eta_i \exp(v_i)$ , где  $\exp(v_i)$  и есть выражение, описывающее шок. Прологарифмировав, получим  $\ln(q_i) = \ln\{f(z_i, \beta)\} + \ln(\eta_i) + v_i$ .

Допустим, что в распоряжении фирм  $m$  производственных факторов, поэтому из-за линейности производственной функции в логарифмах  $\ln(q_i) = \beta_0 + \sum_{j=1}^m \beta_j \ln(z_{ij}) - u_i + v_i$ ,

где  $u_i = -\ln(\eta_i)$ . Введя ограничение  $u_i > 0$ , получится, что условие на  $\eta_i$  выполняется. Стоит отметить, что в данных моделях неэффективность – это случайная величина, а значит можно найти факторы, влияющие на ее реализацию, что представляет практический интерес, так как в этом смысле неэффективность поддается управлению (хороший пример анализа управляемой неэффективности дан в работе [Афанасьев, 2006]). Также необходимо принять во внимание, что хотя такие модели строятся для производственных функций, возможно построение SFA-моделей и для функции издержек (принципиальных отличий у данного типа моделей нет, в таком случае происходит соответствующее изменение знака у ошибки неэффективности).

### 2.2. Методы оценивания технической эффективности и некоторые проблемы современных SFA-моделей

Кроме задачи оценки ошибки неэффективности важной также представляется задача ранжирования предприятий по эффективности. Наиболее распространенными для этих целей являются следующие подходы.

1. Стандартный способ основан на формуле условной функции плотности распределения  $f(u_{it}|e_{it}) = \frac{f(u_{it}, e_{it})}{f(e_{it})}$ , откуда находится  $E(u_{it}|e_{it})$ , которое и будет обозначать

ранг технической эффективности (далее – JLMS-оценка) [Jondrow et al., 1982].

2. Бэттис и Коелли [Battese, Coelli, 1988] (далее – BC) предложили другой способ оценки технической эффективности  $E(\exp(-u_{it})|e_{it})$ . Хотелось сразу отметить, что оцени-

вание технической эффективности для гамма-распределения и random parameters model производится с использованием численных процедур. Кроме того, применяется  $\exp(u_{it})$  в качестве прокси для технической эффективности. В данной работе было решено остановиться на оценке BC. JLMS и BC дают практически идентичные результаты, однако исследователи чаще используют BC ([Greene, 2004, 2008; Kumbhakar et al., 2011; Kumbhakar, Lovell, 2000]).

3. Иногда для этих оценок используются доверительные интервалы (подробнее см. [Greene, 2008]).

Однако в работе [Greene, 2008] показано, что все эти оценки имеют ряд существенных проблем. Допустим, у нас есть выборка из  $n$  фирм, а некая фирма А имеет в такой выборке значение эффективности  $\hat{u}_{it}^*$ , полученное из JLMS-оценки. Если мы изменим выборку, допустим, добавим другие фирмы, то оценка эффективности станет  $\hat{u}_{it}^*$ , хотя сама фирма А осталась такой же. Иными словами, если выборки будут разные, то оценки также будут разными. Действительно, поскольку на практике повторяющиеся выборки из одной генеральной совокупности могут существенно отличаться друг от друга (а как указано в работе [Greene, 2008], вообще такие выборки несостоятельны по своей сути), то, следовательно, условие будет сильно влиять на само значение эффективности. По большому счету это лишь оценка математического ожидания определенного распределения определенной выборки. Аналогично обстоит дело с доверительными интервалами. Технический интервал строится для  $E(u_{it}|e_{it})$ , а на самом деле мы должны построить его для  $u_{it}$ . Теперь представим, что две фирмы имеют одинаковые значения  $q_{it}, x_{it}$ , но могут иметь разные  $u_{it}$  из-за разных значений экзогенных параметров. А мы будем строить для них один и тот же интервал. Причем могут возникнуть еще и проблемы с несимметричными распределениями, например с экспоненциальным, поскольку доверительный интервал, как правило, подразумевается симметричным.

На сегодняшний день не существует единого способа для определения качества подобных моделей, т.е. авторы всего лишь строят корреляционные таблицы для значений эффективности, полученных из разных моделей. Во многом выбор той или иной модели носит скорее интуитивный характер. Например, если панель длинная, то надо включить тренд. Если в выборке много фирм и они разные, то нужно учесть сильную гетерогенность. Учитывается также, насколько адекватны результаты модели и сошлась ли модель вообще.

SFA-модели, как и любые другие эконометрические модели, должны тестироваться на разнообразные «проблемы», как то гетероскедастичность, автокорреляция и т.д. Отдельного внимания (в силу специфики SFA-моделей) заслуживает тестирование ошибки на симметричность и нормальность. Для тестирования симметричности ошибки (для того чтобы определить, нужно ли использовать SFA-модель или можно ограничиться обычной эконометрической моделью) были предложены процедуры в работах [Lee, 1983; Coelli, 1995]. Однако, как было показано сначала в работе [Carree, 2002], а затем в [Almanidis, Sickles, 2012], «неправильная» асимметричность ошибок может быть не следствием неправильной спецификации модели, а возникает из-за особенностей данных, расширяя тем самым понятие эффективности. Также SFA-модели являются чувствительными к распределению ошибок, поэтому были предложены тесты для определения эмпирического

распределения ошибок моделей (как легко заметить, распределение ошибок в данных моделях достаточно сложное, поэтому обычные тесты не применимы). В работе [Wang et al., 2011] предлагается тест Колмогорова – Смирнова и  $\chi^2$ -тест. Авторы, используя симуляцию Монте – Карло, показывают, что эти два теста (особенно тест Колмогорова – Смирнова) являются достаточно мощными. Также стоит отметить, что пропуск переменных оказывает значимое влияние на показатели эффективности моделей (в работе [Greene et al., 2013] рассматривается связь эндогенности и смещения оценок эффективности).

Одним из главных предположений SFA-моделей является то, что ошибка неэффективности и случайная ошибка некоррелированы. Однако в работе [Smith, 2008] показано, что, благодаря применению метода копул, можно построить SFA-модель с учетом корреляции двух типов ошибок, причем результаты классической модели и модели с учетом корреляции отличаются. Но при этом возникают значительные технические и вычислительные трудности. В качестве критерия сравнения подобных моделей авторы предлагают использовать AIC и BIC. Однако, как показано в работе [Пеникас, 2010], это не совсем корректно. Другое обобщение стандартных SFA-моделей дано в работе [Emvalomatis, 2012]. Авторы предлагают использовать метод копул для учета автокорреляции в ошибках неэффективности. К сожалению, не удалось обнаружить литературы, в которой было бы показано, как можно учесть кросс-секционную корреляцию в панельных моделях. Можно предположить, что в качестве претестов для определения необходимости применения обобщенных моделей можно использовать стандартные тесты для тестирования автокорреляции и кросс-секционной корреляции.

### 2.3. Модель без гетерогенности и тренда (TI-модель)

Данная модель разбирается в работе [Kumbhakar et al., 2011]. Рассмотрим модель панельной регрессии без гетерогенности и тренда в неэффективности. В теории данная модель выглядит следующим образом:

$$\ln(q_{it}) = \beta_0 + \sum_n \beta_n \ln(x_{nit}) + v_{it} - u_i,$$

где  $q_{it}$  – выпуск;  $x_{it}$  – факторы, влияющие на издержки или выпуск фирмы;  $v_{it}$  – стохастическая ошибка;  $u_i$  – ошибка неэффективности (для модели с издержками в качестве регрессанта меняется только знак у ошибки неэффективности).

В TI- и TVD-моделях может быть использовано нормальное или усеченное нормальное распределение, являющееся из-за ненулевого математического ожидания более гибким, чем другие.

### 2.4. Модель без гетерогенности (TVD-модель)

Данная модель рассматривается в работах [Battese, Tessema, 1993; Kumbhakar et al., 2011]. Отличие этой модели от предыдущей – в наличии тренда в ошибке неэффективности. Логично предположить, что чем длиннее панель, тем больше вероятность, что

фирма будет улучшать свою технологию, а значит, неэффективность будет уменьшаться. Теоретическая модель выглядит следующим образом:

$$\ln(q_{it}) = \beta_0 + \sum_n \beta_n \ln(x_{nit}) + v_{it} - u_{it}.$$

Легко заметить, что появился индекс  $t$  у  $u_{it}$ . Как и предыдущая модель, она оценивается методом максимального правдоподобия. Сейчас существует множество способов введения тренда, далее обозначены наиболее популярные.

$$1. \quad \ln(y_{it}) = \beta_{0t} + \sum_n \beta_n \ln(x_{nit}) + v_{it} - u_{it} = \beta_{it} + \sum_n \beta_n \ln(x_{nit}) + v_{it},$$

где  $\beta_{it} = \gamma_{i1} + \gamma_{i2} \cdot t + \gamma_{i3} \cdot t^2$ . Плюсы данного подхода заключаются в гибкости квадратичной функции тренда, однако возникают проблемы с оцениванием коэффициентов (подробнее см.: [Cornwell et al., 1990]).

2. Также можно задавать тренд следующим образом:

$$u_{it} = \beta(t) \cdot u_i,$$

где  $\beta(t)$  – дамми-переменные для тренда. Плюсом данного способа является гибкость, но возникают соответствующие проблемы с плохой определенностью коэффициентов.

3. Можно использовать следующую функциональную форму  $\beta(t)$ :

$$\beta(t) = (1 + \exp(\alpha \cdot t + \lambda \cdot t^2))^{-1}.$$

4. В работе [Battese, Tessema, 1993] было предложено задавать тренд следующим образом:  $\beta(t) = \exp(\alpha \cdot (t - T))$ . Эту модель можно рассматривать как частный случай модели 3.

5. Следующим уровнем обобщения модели 3 является ввод, например, firm-specific-элементов.

Заметим, что в статистическом пакете STATA реализован тренд в соответствии с подходом, предложенным в работе [Battese, Tessema, 1993].

## 2.5. Модели FE и RE

Данные модели рассматриваются в работе [Kumbhakar, Lovell, 2000]. Модели FE и RE являются крайне популярными для анализа панельных данных, поэтому нет ничего удивительного в том, что они используются для анализа эффективности. Модель FE:

$$\ln(q_{it}) = \beta_{0i} + \sum_n \beta_n \ln(x_{nit}) + v_{it},$$

откуда  $u_i$  задается за счет нормализации как  $\max_i (\beta_{0i}) - \beta_{0i}$  (т.е. одна фирма признается абсолютно эффективной), а техническая неэффективность –  $\exp(-u_i)$ . Таким образом, достигается ранжирование фирм по эффективности. Однако в такой модели присутствует систематическое смещение оценок эффективности [Greene, 2008], из-за чего многие фирмы получают более низкие ее значения (подробнее об этом см. в разделе 3).

Случайный эффект (RE) для SFA-моделей задается следующим образом:

$$\ln(q_{it}) = (\beta_0 - E(u_i)) + \sum_n \beta_n \ln(x_{nit}) + v_{it} - (u_i - E(u_i)) = \beta_0^* + \sum_n \beta_n \ln(x_{nit}) + v_{it} - u_i^*.$$

Данную модель легко оценить стандартными методами. Оценка эффективности может быть получена из выражения  $\exp(\max_i (u_i^*) - u_i^*)$ .

## 2.6. Модели «True» fixed effect и «True» random effect

Данные модели рассматриваются в работах [Greene, 2004, 2005; Kumbhakar et al., 1995]. Поскольку предыдущие FE- и RE-модели обладали серьезными недостатками (например, оценки эффективности в FE были явно смещенными), то Грин [Greene, 2004] предложил совершенно новый подход к оцениванию подобных моделей в условиях неэффективности. Он назвал их «True», так как при учете неэффективности они действительно выглядят как полноценные панельные модели.

«True» FE-модель имеет вид:  $\ln(y_{it}) = \beta_i + \sum_n \beta_n \ln(x_{nit}) + v_{it} - u_{it}$ . Как легко можно

заметить, в этой модели есть как неэффективность, так и firm-specific коэффициент  $\beta_i$ , поэтому можно теперь отделять эффект особенности каждой фирмы от ее неэффективности. К тому же из-за того, что для FE-модели приходится обсчитывать много дамми-переменных, возникают проблемы с качеством оценок параметров, а в «true» FE-модели дамми-переменные не вводятся, и из-за этого можно использовать метод максимального правдоподобия.

Но в работе [Wang, Ho, 2010] было показано, что все равно есть смещение в ошибке в модели «true» FE, поэтому авторы ввели понятие corrected «true» fixed effect модели. Авторы показали, что если перейти к первым разностям или сделать within-transformation, то можно убрать смещение. Из-за небольшой корректировки формулы расчета неэффективности ранги неэффективности могут сопоставляться с другими моделями. Особенности их модели в следующем: оценки модели состоятельны при  $N \rightarrow \infty$ , можно использовать truncated normal распределение, никакого особенного максимизационного метода не нужно. Другую модернизацию «True» FE предложили авторы работы [Belotti, Llard, 2012].

«True» RE-модель как идея была предложена в статье [Kumbhakar et al., 1995]. Из-за своей сложности «True» RE оценивается там не методом максимального правдоподобия, как, например, в работе [Greene, 2005], а двухшаговым методом наименьших квадратов, как и было изначально предложено авторами оригинальной работы. На первом шаге оце-



нивается обычная RE-модель:  $\ln(q_{it}) = \beta_0 + \sum_n \beta_n \ln(x_{nit}) + v_{it} + \eta_i$ , из нее выражается перманентная ошибка эффективности через  $\eta'_i = \max(\eta_i^*) - \eta_i^*$ . PTE (*persistent technical efficiency*) соответственно может быть оценена как  $\exp(-\eta'_i)$ . На втором шаге оценивается модель SFA только со свободным членом:  $v_{it} = \alpha_0 - u_{it} + e_{it}$ , где  $v_{it}$  – стохастическая ошибка из RE-модели, и получаются значения эффективности, зависящей от времени (*residual technical efficiency*):  $RTE = \exp(-u_{it})$ . Затем считается общая эффективность (*overall technical efficiency*), как  $OTE = PTE \cdot RTE$ . Таким образом, учитываются два типа неэффективности, и это важно, поскольку они зачастую имеют разные причины. Неэффективный менеджмент – это пример постоянной технической неэффективности, сезонные перебои с поставками – соответственно, эффективности, зависящей от времени.

### 2.7. Модель с четырьмя ошибками

Данная модель рассматривается в работе [Kumbhakar et al., 2011]. Поскольку, как отмечал Грин [Greene, 2004, 2005], очень важно отделять гетерогенность от неэффективности, то в работе [Kumbhakar et al., 2011] была предложена модель с четырьмя разными ошибками:  $\ln(q_{it}) = \alpha_0 + \sum_n \beta_n \ln(x_{nit}) + \mu_i + v_{it} - \eta_i - u_{it}$ , где  $\mu_i$  – ненаблюдаемые шоки, постоянные во времени;  $v_{it}$  – стохастическая ошибка;  $\eta_i$  – постоянная техническая неэффективность;  $u_{it}$  – неэффективность, зависящая от времени.

Преобразуем это выражение к виду  $\ln(q_{it}) = \alpha_0^* + \sum_n \beta_n \ln(x_{nit}) + e_{it} + \alpha_i$ , где  $\alpha_0^* = \alpha_0 - E(\eta_i) - E(u_{it})$ ,  $\alpha_i = \mu_i - \eta_i + E(\eta_i)$ ,  $e_{it} = v_{it} - u_{it} + E(u_{it})$ . Теперь все ошибки имеют нулевое математическое ожидание и можно оценить модель как обычную панельную регрессию. Далее берем оценки ошибок и оцениваем следующую регрессию:  $\hat{e}_{it} = E(u_{it}) + (v_{it} + (\hat{e}_{it} - e_{it})) - u_{it}$ , все ошибки имеют нулевое математическое ожидание, а член  $(\hat{e}_{it} - e_{it})$  можно опустить, так как при больших выборках оценка ошибки очень близка к значению самой ошибки. Поэтому можно применить стандартную процедуру оценки SFA-моделей.

Подобным же образом оценивается регрессия для второй ошибки:

$$\hat{\alpha}_i = E(\eta_i) + (\mu_i + (\hat{\alpha}_i - \alpha_i)) - \eta_i,$$

член  $(\hat{\alpha}_i - \alpha_i)$  опять можно опустить и соответственно можно будет воспользоваться оцениванием SFA-моделей. Постоянную техническую эффективность можно оценить как  $PTE = \exp(-\eta_i)$ , а техническую эффективность, зависящую от времени, – как  $RTE = \exp(-u_{it})$ . Общая техническая эффективность  $OTE = PTE \cdot RTE$ .

## 2.8. Модель с изменяющимися коэффициентами (random parameters model)

Данная модель рассматривается в работах [Greene, 2004, 2005]. Поскольку очень важно отделить гетерогенность от неэффективности (если не учесть гетерогенность, то на значение эффективности будут влиять также *особенности* функционирования конкретной фирмы, что приведет к некорректной оценке эффективности), в работе [Greene, 2004] был предложен новый подход к оцениванию параметров модели. Теперь коэффициенты могут меняться от фирмы к фирме, конечно, это создает определенные трудности при оценке, так как приходится использовать метод симулированного максимального правдоподобия. Общий вид такой модели может быть записан (в векторной форме, логарифмы опущены, чтобы не усложнять запись):  $y_{it} = \beta'_i x_{it} + v_{it} - u_{it}$ . Данная модель построена с помощью условного распределения:

$$f(y_{it}|x_{it}, \beta_i) = f(\beta'_i x_{it}), i = \overline{1, N}, t = \overline{1, T},$$

причем  $\beta_i = \beta + \Delta z_i + \Gamma w_i$ ,  $E(\beta_i|z_i) = \beta + \Delta z_i$ ,  $Var(\beta_i|z_i) = \Sigma$ . Теперь коэффициенты модели зависят от параметров  $z_i$ ;  $\Delta$  – матрица соответствующих параметров;  $w_i$  – стохастическая ошибка в этой модели с нулевым математическим ожиданием и ковариационной матрицей  $\Sigma$ ;  $\Gamma$  – матрица без ограничений.

## 2.9. Оценивание неэффективности по кластерам (latent class model)

Данная модель рассматривается в работе [Greene, 2005]. Пусть предполагается, что все наблюдения распределены по  $Q$  классам, причем исследователь не знает истинное распределение по классам (аналогично, записано все в векторной форме).

$f(y_{it}|x_{it}, class\_q) = f(\Theta_q, y_{it}, x_{it}), i = \overline{1, N}, t = \overline{1, T}$ , предполагается, что распределение  $f(\cdot)$  одинаково для всех классов, а каждый класс отличается от другого вектором  $\Theta_q$ . Далее

предполагается, что условная вероятность фирмы  $i$  есть  $P(i|q) = \prod_{t=1}^T P(i, t|q)$ . Причем,

$P(i, t|q) = f(y_{it}|x_{it}, \beta_q, \sigma_q, \lambda_q)$ . Так же задается вероятность принадлежать классу  $q$  как

$P(i, q) = \frac{\exp(z'_i \pi_q)}{\sum_{m=1}^Q \exp(z'_i \pi_m)}$ , это будет не объективная вероятность, а субъективная, основан-

ная на распределении параметров  $z_i$ . Тогда безусловная вероятность фирмы  $i$ , усред-

ненная по всем классам, есть  $P(i) = \sum_{q=1}^Q \Pi(i, q) \cdot P(i|q)$ . Поэтому функция максимального

правдоподобия будет выглядеть как  $\log L = \sum_{i=1}^N \log P(i)$ .

### 3. Оценивание моделей методом моментов

Классический метод моментов дает состоятельные и асимптотически нормальные оценки. Однако в отличие от метода максимального правдоподобия метод моментов не использует всю доступную информацию, что в итоге приводит к большему смещению в оценках. Но в то же время оценки в методе моментов считаются, а не оцениваются.

При гетероскедастичности в панельных данных методом максимального правдоподобия приходится оценивать большое число параметров. Например,  $(I + N + 2)$  параметров для time-invariant-модели с гетероскедастичной  $v$ , где  $I$  – количество фирм,  $N$  – количество регрессоров, что ведет к проблемам со сходимостью этих моделей на реальных данных, особенно если  $T$  гораздо меньше  $I$  по абсолютному значению. Кумбхакар и Ловелл [Kumbhakar, Lovell, 2000] предложили оценивать подобные модели с помощью метода моментов. Как уже было указано в начале, теоретическое сравнение данных методов можно найти в работах [Greene, 2008; Kumbhakar, Lovell, 2000], однако авторы данных работ не делают однозначного вывода, какой из двух методов лучше. Поэтому представляется интересным проверить работоспособность этого метода на практике и сравнить оценки неэффективности с результатами других моделей, особенно с учетом того, что данный метод крайне редко применялся (можно лишь отметить работы [Kumbhakar, Lovell, 2000; Wang, Chen, 2004]).

#### 3.1. МоМ. T-inv без гетероскедастичности

Это обычная time-invariant-модель без гетероскедастичности. Сначала оценивается обычная OLS-регрессия

$$\ln(q_{it}) = (\beta_0 - E(u_i)) + \sum_n \beta_n \ln(x_{nit}) + v_{it} - (u_i - E(u_i)) = \beta_0^* + \sum_n \beta_n \ln(x_{nit}) + v_{it} - u_i^*$$

Предполагается, что в модели нормальное или полунормальное распределение, поэтому  $E(u_i) = \sqrt{\frac{2}{\pi}} \sigma_u$ ,  $V(u_i) = (1 - 2/\pi) \sigma_u^2$ ,  $E(u_i^3) = -\sqrt{\frac{2}{\pi}} \left(1 - \frac{4}{\pi}\right) \sigma_u^3$ . Центральный момент третьего порядка  $\mu_3 = \sqrt{\frac{2}{\pi}} \left(1 - \frac{4}{\pi}\right) \sigma_u^3$ . Затем считается выборочный момент третьего порядка и подставляется в формулу

$$\hat{\sigma}_u = \left( \frac{\hat{m}_3}{\sqrt{2/\pi}} \left( 1 - \frac{4}{\pi} \right)^{-1} \right)^{1/3}.$$

Из RE-модели выражается

$$\hat{\sigma}_v^2 = \left( \frac{1}{I \cdot T} \right) \cdot \sum_i \sum_t \hat{e}_{it}^2,$$

где  $\hat{e}_{it}^2$  – общая ошибка. Затем значения подставляются в нужные формулы для выражения технической эффективности (подробнее об этой и последующих моделях см. [Kumbhakar, Lovell, 2000]).

### 3.2. МоМ. Т-*v* без гетероскедастичности

Если добавить тренд в ошибку, то шаги оценивания немного изменятся. Регрессия будет записываться уже следующим образом:

$$\ln(q_{it}) = (\beta_0 - \beta_t \cdot E(u_i)) + \sum_n \beta_n \ln(x_{nit}) + v_{it} - (u_i - \beta_t \cdot E(u_i)),$$

или после преобразования  $\ln(q_{it}) = \beta_t^* + \sum_n \beta_n \ln(x_{nit}) + v_{it} - u_{it}^*$ , где  $\beta_t^*$  – дамми-переменная на год.

Тогда  $\beta_t \hat{\sigma}_u = \left( \frac{\hat{m}_3}{\sqrt{2/\pi}} \left( 1 - \frac{4}{\pi} \right)^{-1} \right)^{1/3}$ ,  $t = 1, \dots, T$ . Нормализация  $\beta_t = 1$ ,  $t = 2, \dots, T$

дает возможность получить оценки  $\sigma_u$ ,  $\beta_t$ . Отсюда получается

$$\hat{\beta}_0 = (1/T) \sum_t (\hat{\beta}_t^* + \hat{\beta}_t \hat{\sigma}_u \sqrt{2/\pi}).$$

Соответственно, можно легко посчитать оценки  $\sigma_v^2$ ,

$$\hat{\beta}_v^2 = \left( \frac{1}{I \cdot T} \right) \cdot \sum_i \sum_t \hat{e}_{it}^2 - \frac{1}{T} \hat{\sigma}_u^2 \left( 1 - \frac{2}{\pi} \right) \sum_t \hat{\beta}_t^2.$$

Затем оценки, полученные выше, подставляются в  $E(\exp(-u_{it}) | e_{it})$ .

### 3.3. МоМ. T-inv с гетероскедастичностью в $v$

Теперь предположим, что у нас есть гетероскедастичность в  $v$ , однако нет тренда в ошибке. Во многом ход оценивания совпадает с ходом оценивания для гомоскедастичной  $v$ . Сначала мы оцениваем OLS-регрессию, получаем оценки  $\hat{\sigma}_u^2$ . Затем оцениваем RE-регрессию с корректировкой на гетероскедастичность и получаем  $\hat{\sigma}_{vi}^2 = (1/T) \sum_t \tilde{e}_{it}^2$ . Далее необходимо подставить полученные оценки в формулу технической эффективности.

### 3.4. МоМ. T-inv с гетероскедастичностью в $u$

Пусть у нас гетероскедастичность только в  $u$ . На первом шаге мы получаем оценки  $\beta_n$ ,  $\sigma_v^2$  из OLS-регрессии. Преобразуем изначальное уравнение

$$\ln(q_{it}) - \sum_n \beta_n \ln(x_{nit}) = \beta_0 + v_{it} - u_i.$$

Тогда состоятельные оценки будут выглядеть так:

$$\hat{\sigma}_{ui}^2 = \left( \frac{\pi}{\pi - 2} \right) \left( (1/T) \cdot \sum_t (\ln(y_{it}) - \sum_n \hat{\beta}_n \ln(x_{xit}))^2 - \sigma_v^2 \right)$$

и

$$\hat{\sigma}_v^2 = \left( \frac{1}{I \cdot T} \right) \cdot \sum_t \sum_i \hat{e}_{it}^2.$$

Далее все аналогично предыдущим моделям.

### 3.5. МоМ. T-inv с гетероскедастичностью в $u$ и $v$

Так же как и при гетероскедастичности в  $v$ , мы получаем оценки  $\hat{\sigma}_{vi}^2 = (1/T) \sum_t \tilde{e}_{it}^2$ .

Затем оценивается  $\hat{\sigma}_{ui}^2 = \left( \frac{m_{3i}}{\sqrt{2/\pi}} \left( 1 - \frac{4}{\pi} \right) \right)^{2/3}$ ,  $\hat{\beta}_0 = \left( \frac{1}{I \cdot T} \right) \sum_i \sum_t \hat{e}_{it} + (1/I) \sqrt{2/\pi} \sum_i \hat{\sigma}_{ui}$ . Далее считаются значения технической эффективности.

### 3.6. МоМ. T-v с гетероскедастичностью в $v$

Модель предполагает наличие тренда в ошибке и гетероскедастичности в  $v$ . Оценивается модель, как и в гомоскедастичном случае:

$$\ln(q_{it}) = (\beta_0 - \beta_t \cdot E(u_i)) + \sum_n \beta_n \ln(x_{nit}) + v_{it} - (u_i - \beta_t \cdot E(u_i)),$$

или после преобразований  $\ln(q_{it}) = \beta_t^* + \sum_n \beta_n \ln(x_{nit}) + v_{it} - u_{it}^*$ , где  $\beta_t^*$  – дамми-переменная на год. Так же оценивается  $\sigma_u^2$ , а затем

$$\hat{\sigma}_{v_i}^2 = (1/T) \sum_t e_{it}^{*2} - (1/T) \cdot ((\pi - 2)/\pi) \cdot \sum_t \hat{\beta}_t \hat{\sigma}_{it}^2,$$

где  $e_{it}^*$  – общие ошибки из уравнения, приведенного выше. Далее значения подставляются в нужные формулы для расчета технической эффективности.

### 3.7. МоМ. T-v с гетероскедастичностью в $u$ . T-v с гетероскедастичностью в $u$ , $v$

Модель T-v с гетероскедастичностью в  $u$  может быть оценена лишь на основании результатов, полученных из оценивания SFA-модели с гетерогенностью в среднем значении ошибки неэффективности. Поскольку в таком случае надо использовать нормальное/усеченное нормальное распределение, то среднее значение будет зависеть от фирмы и времени. А поскольку дисперсия зависит от среднего, то модель становится гетероскедастичной. Абсолютно аналогичная ситуация характерна и для модели T-v с гетероскедастичностью в  $u$  и  $v$ .

### 3.8. Перспективы SFA-моделей

В настоящее время SFA-модели бурно развиваются. Сейчас всю большую популярность набирают полупараметрические и непараметрические модели (например, [Heshmati et al., 2013]), также исследователи пытаются разработать средства диагностики данных моделей и определить критерии их качества.

Развитие параметрических и полупараметрических подходов к оцениванию позволит значительно увеличить гибкость данных моделей. Также быстро развивается байесовский подход к построению SFA-моделей (например, [Tsonas, 2005]).

## 4. Применение методов оценки: бетонная промышленность России

Данные для исследования были взяты из базы RUSLANA (источник: Bureau Van Dijk). Эта база содержит информацию о более чем 10 млн российских предприятий по 30 переменным бухгалтерской отчетности (баланс, отчет о прибылях и убытках и т.д.). Для анализа были выбраны компании, производящие лишь бетон и цемент в соответствии с классификатором видов экономической деятельности ОКВЭД<sup>2</sup>, в то время как компании,

<sup>2</sup> Производство изделий из бетона для использования в строительстве (26.61); производство товарного бетона (26.63); производство сухих бетонных смесей (26.64); производство изделий из асбестоцемента и волокнистого цемента (26.65).

занимающиеся производством гипса, были удалены из выборки. Это объясняется предпосылками подхода SFA, который предполагает однородность компаний, а производство гипса сильно отличается по технологии от производства цемента и бетона.

Из полученной выборки были удалены предприятия с пропущенными значениями переменных, а также – на основе первичного анализа – удалены наблюдения с подозрительными на выбросы значениями переменных. Таких компаний оказалось около 10% от всей выборки. В итоге сбалансированная панель содержит 299 фирм, отражая их деятельность за период 2006–2010 гг., описываемую следующими показателями:

**lnq** – логарифм выпуска (готовая продукция, изначально в тыс. руб.);

**lnl** – логарифм числа рабочих;

**lnk** – логарифм размера производственных активов фирмы (исначально в тыс. руб.).

Последняя переменная равна сумме основных и оборотных активов фирм. Переменная составляет большую часть общих активов компаний, поэтому косвенно учитывает и размер конкретной фирмы;

**dummy1, dummy2, dummy3, dummy4** – фиктивные переменные для учета временных эффектов (принимает единичные значения для 2010, 2009, 2008 и 2007 гг. соответственно);

**ln\_dolg\_po\_zp** – логарифм задолженности перед работниками (исначально в тыс. руб.);

**ln\_dolg\_po\_dohodam** – логарифм задолженности по выплате доходов (исначально в тыс. руб.);

**ln\_sebestoim** – логарифм себестоимости продукции (исначально в тыс. руб.). Долг по доходам и зарплате, а также издержки использовались в качестве факторов, влияющих на эффективность;

**kap** – логарифм суммы оборотного капитала и собственного капитала (исначально в тыс. руб.). Данная переменная использовалась для учета гетероскедастичности в многоступенчатых моделях (табл. 1).

Таблица 1.

Описательные статистики переменных

Переменная	Количество наблюдений	Среднее	Стандартное отклонение	Минимум	Максимум
Lnq	1495	11,777	1,4328	3,7135	15,5066
lnl	1495	4,8847	1,1015	1,3862	7,517
lnk	1495	11,2416	1,4191	4,3178	15,4986
ln_dolg_po_zp	1495	6,6861	2,2306	0	10,787
ln_dolg_po_dohodam	1495	1,003	2,318	0	11,1336
ln_sebest	1495	11,592	2,0794	2,0795	15,069
Kap	1495	10,073	3,7185	3,7185	14,407

В работе использовалась производственная функция Кобба – Дугласа, поэтому оценки коэффициентов можно рассматривать как эластичности. Оценивалась также

транслогарифмическая функция, однако данная спецификация по объясняющей силе не выигрывает у стандартной функции Кобба – Дугласа: оценки параметров, отвечающие за дополнительную гибкость транслогарифмической спецификации, являются сильно незначимыми, коэффициент детерминации не выше, чем в модели Кобба – Дугласа, при этом оценки эффективности предприятий имеют схожие характеристики распределения.

Итоговый вид модели следующий:

$$\ln(q_{it}) = \beta_0 + \beta_1 \cdot \ln(l_{it}) + \beta_2 \cdot \ln(k_{it}) + \beta_3 \cdot dummy1_t + \\ + \beta_4 \cdot dummy2_t + \beta_5 \cdot dummy3_t + \beta_6 \cdot dummy4_t + e_{it},$$

где в  $e_{it}$  включены ошибка неэффективности  $u_{it}$  и случайное возмущение  $v_{it}$ . Далее все регрессоры для упрощения обозначены как  $\ln(x_{nit})$ .

В качестве аргумента в пользу использования моделей, учитывающих неэффективность, может быть использована гистограмма общих остатков RE-модели (ситуация с остатками FE-модели аналогичная) (рис. 1). Легко заметить смещение остатков влево (коэффициент асимметрии равен  $-0,17$ ), что свидетельствует в пользу гипотезы о наличии в модели ошибки неэффективности.

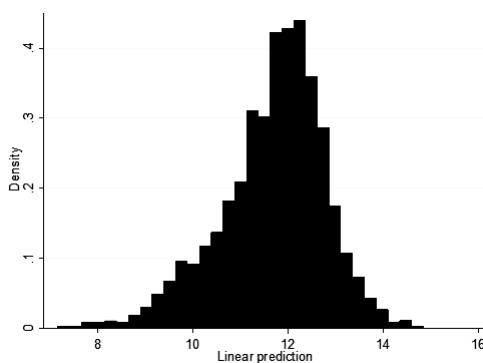


Рис. 1. Распределение общих остатков RE-модели

В дальнейшем с целью сопоставимости результатов во всех случаях оценивалась спецификация модели, приведенная выше, также было решено не выбрасывать незначимые коэффициенты в моделях, чтобы не терять сопоставимость (остатки при выбрасывании незначимых коэффициентов менялись незначимо). Также следует отметить, что полученные модели стабильны и, таким образом, результаты робастны.

## 5. Результаты и сравнение моделей

Непосредственно перед представлением полученных результатов необходимо отметить ряд ключевых моментов.



К сожалению, из-за технических трудностей реализовать «True» FE не удалось. Поэтому была оценена только «True» RE-модель.

Поскольку STATA позволяет учитывать гетероскедастичность в кросс-секционных SFA-моделях, то было решено использовать это свойство, и в качестве фактора гетероскедастичности была выбрана переменная *kap*, которая учитывает структуру капитала фирмы, а следовательно, и гетерогенность.

Модель с изменяющимися коэффициентами (*random parameters model*) оценить не удалось, так как процедура максимизации функции правдоподобия не сошлась на данных.

К сожалению, при оценке неэффективности по кластерам (*latent class model*) пришлось столкнуться с техническими трудностями. STATA11 не может работать с большим объемом данных при оценке этой модели.

При оценке методом моментов модели T-inv с гетероскедастичностью в *u* возникли проблемы с оценкой дисперсий *u*. Они получились отрицательными, поэтому данная модель выпадает из рассмотрения.

При оценке данной модели получилось, что дисперсия *v* отрицательна. Это вполне ожидаемый результат, так как не гарантировано, что  $\left(\frac{1}{I \cdot T}\right) \cdot \sum_i \sum_t \hat{e}_{it}^2$  будет больше, чем

$\frac{1}{T} \hat{\sigma}_u^2 \left(1 - \frac{2}{\pi}\right) \sum_t \hat{\beta}_t^2$ . Основное условие применимости метода моментов – отрицательное

значение третьего момента распределения ошибок, что выполняется во всех моделях.

При оценке методом моментов модели T-v с гетероскедастичностью в *v* дисперсия стохастической ошибки стала отрицательной, как и в гомоскедастичном случае, поэтому для данной модели получить результаты путем оценки методом моментов нельзя.

Модель T-v с гетероскедастичностью в *u* с помощью метода моментов из-за технических сложностей оценить не удалось. Аналогичная ситуация имела место и в случае T-v с гетероскедастичностью в *u* и *v* (метод моментов).

Таблица 2.

Основные результаты одношаговых регрессий

Регрессоры	TI	TVD	RE	FE
LnI	0,4571*** (0,0359)	0,4748*** (0,0284)	0,5564*** (0,0336)	1,0441*** (0,0671)
Lnk	0,5958*** (0,02)	0,569*** (0,0195)	0,5733*** (0,0247)	0,5984*** (0,036)
Dummy1	-0,0036 (0,0282)	-0,1026*** (0,0309)	-0,0028 (0,0278)	0,0081 (0,0273)
Dummy2	0,5524*** (0,0283)	0,3688*** (0,0373)	0,5573*** (0,0279)	0,5865*** (0,0275)
Dummy3	0,5014*** (0,0285)	0,2384*** (0,0443)	0,5085*** (0,0282)	0,5783*** (0,0296)

Окончание табл. 2.

Регрессоры	TI	TVD	RE	FE
Dummy4	0,3189*** (0,0299)	-0,018 (0,0515)	0,3237*** (0,0302)	0,4310*** (0,0361)
Cons	3,1867*** (0,1691)	3,5683*** (0,1707)	2,3372*** (0,2134)	-0,3711 (0,4876)
Техническая эффективность				
среднее	0,5916	0,6007	0,4144	0,2179
стандартное отклонение	0,2018	0,2069	0,1676	0,1792
минимум	0,0407	0,0167	0,0383	0,0173
максимум	0,9527	0,9648	0,99	1

\*\*\* – значимость на 1-процентном уровне, \*\* – на 5-процентном уровне, \* – на 10-процентном уровне.

Поскольку единого критерия качества моделей неэффективности не существует, то обычно используется анализ корреляционных зависимостей, парные диаграммы рассеивания для эффективности и гистограмм значений эффективности (ГЗЭ) (рис. 2–10).

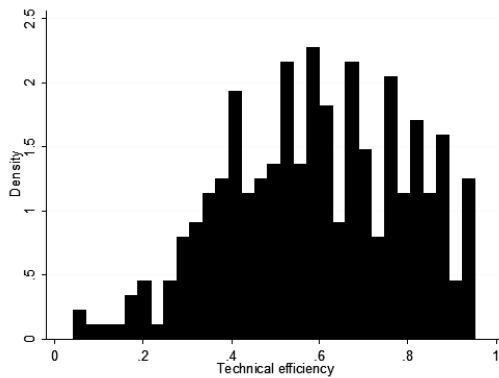


Рис. 2. ГЗЭ из TI

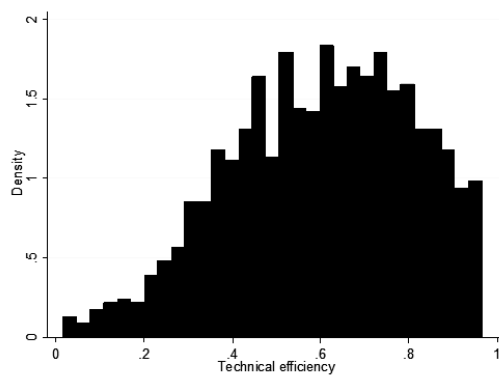


Рис. 3. ГЗЭ из TVD

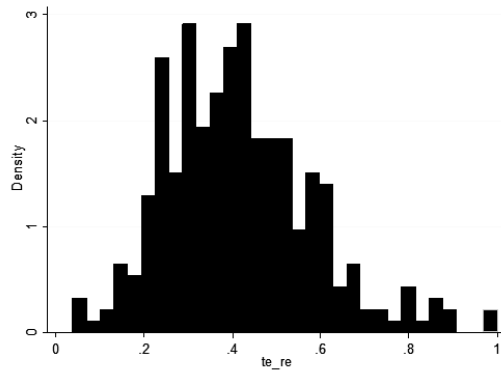


Рис. 4. ГЗЭ из RE

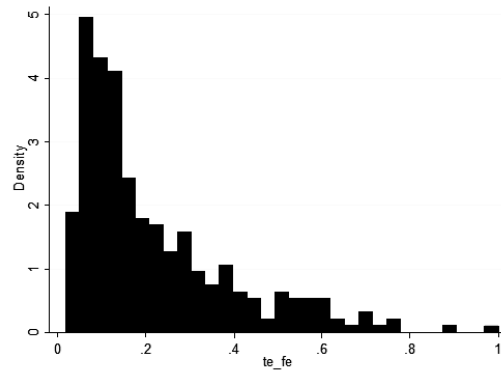


Рис. 5. ГЗЭ из FE

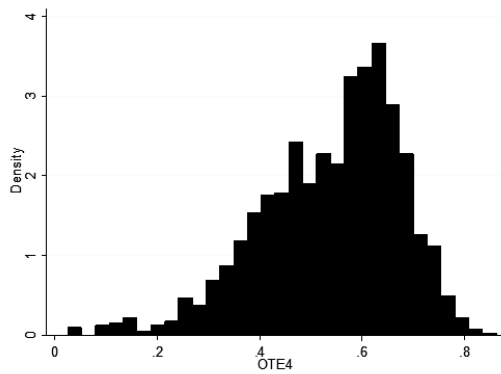


Рис. 6. ГЗЭ из 4 errors model

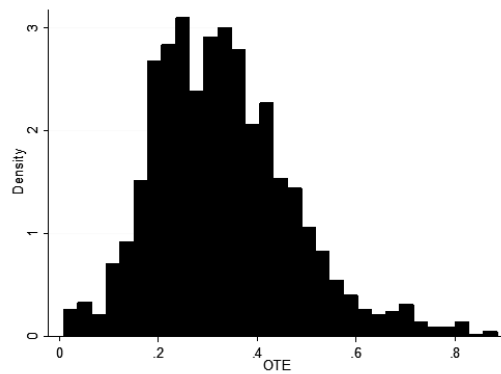


Рис. 7. ГЗЭ из «True» RE

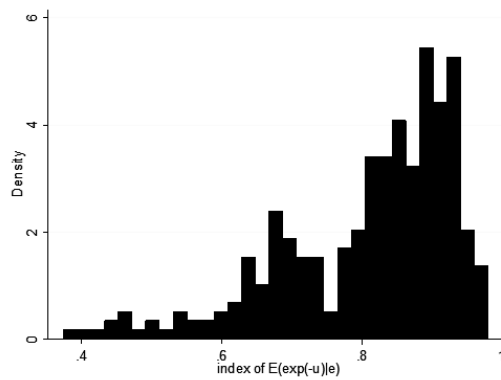
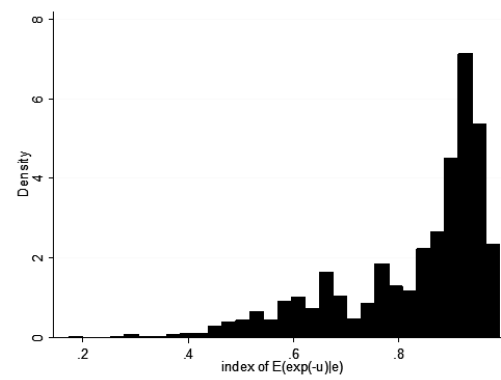


Рис. 8. ГЗЭ из МоМ T-inv

Рис. 9. ГЗЭ из МоМ T-inv с гетероскедастичностью  $\nu$

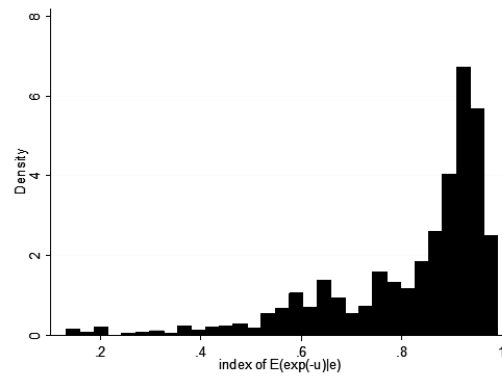


Рис. 10. ГЗЭ из МоМ T-inv  
с гетероскедастичностью  $u$  и  $v$

Далее были построены ранговые корреляции по Спирману (табл. 3).

Таблица 3.

Ранговые корреляции значений эффективности по Спирману

	TI	TVD	RE	FE	4 errors model	«True» RE	МоМ T-inv без гетероскедастичности	МоМ T-inv с гетероскедастичностью $v$	МоМ T-inv с гетероскедастичностью $u$ и $v$
TI	1								
TVD	0,9610	1							
RE	0,9810	0,9394	1						
FE	0,6342	0,5878	0,7567	1					
4 errors model	0,9005	0,8586	0,9051	0,6450	1				
«True» RE	0,9597	0,9156	0,9729	0,7233	0,9694	1			
МоМ T-inv без гетероскедастичности	0,9810	0,9394	1	0,7567	0,9051	0,9729	1		
МоМ T-inv с гетероскедастичностью $v$	0,7871	0,7185	0,7792	0,5195	0,7499	0,7674	0,7792	1	
МоМ T-inv с гетероскедастичностью $u$ и $v$	0,7781	0,7077	0,7700	0,5154	0,7424	0,7581	0,7700	0,9945	1

Все корреляции значимы на 1-процентном уровне. Отметим, что в целом значения эффективности фирм сходны друг с другом. Основываясь на проведенном анализе, можно заключить, что модели на российских данных в большинстве своем неплохо согласуются между собой. Вместе с тем сильно выделяется FE-модель и метод моментов (отличное от других моделей распределение значений неэффективности и, как следствие, относительно невысокая корреляция значений эффективности с другими моделями).

Возникает естественный вопрос, почему часть фирм попала в группу крайне неэффективных и есть ли различия между кластерами неэффективных и эффективных (на основании результатов оценки моделей) фирм. Для ответа на этот вопрос на первом этапе была осуществлена кластеризация. Разбивать на большое количество групп 299 относительно однородных фирм нелогично, поэтому требовалось проверить, нужно ли делить предприятия на две группы (условно, «эффективные» и «неэффективные»). Для этого были рассмотрены статистики Дуда – Харта и Калински – Харабаза. Обе они дали вполне определенный результат: лучше одна большая группа, чем две меньшего размера. Поэтому, согласно результатам кластеризации и логического анализа, нет необходимости разбивать выборку.

Кроме того, следует отметить следующие недостатки FE-модели. Во-первых, она совершенно не учитывает гетерогенность. Во-вторых, ошибка неэффективности не меняется по времени, что для панели с пятью годами является достаточно спорным моментом. В-третьих, как отмечается в работе [Greene, 2008], есть систематическое занижение оценок эффективности, связанное с особенностью конструирования ковариационных матриц. Поэтому такое распределение предприятий по неэффективности легко объясняется неучтенной гетерогенностью. Указанное различие в корреляции может быть обусловлено особенностями конструирования FE-модели. Самыми подходящими, на наш взгляд, для оценки эффективности в этой ситуации являются следующие модели: модель с четырьмя ошибками, «True» RE- и TVD-модели. Они наиболее полно учитывают гетерогенность данных и временной тренд.

Отдельно следует отметить и особенности метода моментов. В принципе, модели с гетероскедастичностью в достаточной мере учитывают гетерогенность, однако, как уже отмечалось, данный метод не лишен своих недостатков. Во-первых, метод моментов дает смещенные оценки, причем сила смещения в среднем выше, чем у метода максимального правдоподобия. Во-вторых, метод моментов использует не всю доступную информацию в силу своих свойств. В-третьих, данный метод оценки моделей очень чувствителен к базовым (начальным) значениям, которые получаются из оценки RE-моделей, что делает его малоприменимым на практике. Вероятно, именно поэтому данный метод не имеет широкого применения в современной литературе по SFA.

На основании полученных результатов можно заключить, что фирмы цементной промышленности, в среднем, имеют значение эффективности около 0,5. Поэтому можно сделать вывод о том, что эффективность всей отрасли находится на достаточно среднем уровне. Это вполне ожидаемый результат, поскольку бетонная промышленность представлена в основном заводами, технология производства на которых десятилетиями почти не меняется и спрос на продукцию которых неэластичный, а значит, предприниматели не особо стремятся к наибольшей эффективности производства.

## 6. Заключение

В работе описаны 15 моделей оценки стохастической производственной границы (SFA), 8 из них используют метод максимального правдоподобия, 7 – метод моментов.

В качестве иллюстрации было оценено 9 моделей для анализа эффективности фирм бетонной и цементной отраслей. Корреляция оценок разных моделей оказалась достаточно высока, что согласуется с результатами других исследователей ([Greene, 2008; Kumbhakar et al., 2011; Kumbhakar, Lovell, 2000]). На общем фоне выделяются МоМ-модели. Это легко объясняется рядом их существенных недостатков: нестабильность результатов, неэффективность. Наиболее адекватные модели для оценки эффективности бетонной промышленности – это модель с четырьмя ошибками (модель 6 из [Kumbhakar et al., 2011]), «True» RE (правда, в интерпретации [Kumbhakar et al., 1995]) и TVD-модель (в виде оригинальной [Battese, Tessema, 1993]), так как они учитывают тренд и гетерогенность.

В статье был проведен анализ последних достижений в области моделирования эффективности на примере SFA-моделей. Можно сделать следующие выводы относительно современного SFA-моделирования.

1. При прочих равных, лучше использовать MML-подход, так как он позволяет извлечь больше информации из данных, чем классический метод моментов.
2. Отсутствуют формальные критерии качества моделей, поэтому сложно выбрать корректную модель для анализа конкретных данных. В целом необходимо полагаться на уровень гетерогенности, длину панели и цель исследования.
3. Слаборазвитый pre- и post-анализ не позволяет формализовать должным образом необходимость и корректность использования SFA-моделей в каждом конкретном случае.

\* \*  
\*

## СПИСОК ЛИТЕРАТУРЫ

- Афанасьев М.Ю. Модель производственного потенциала с управляемыми факторами эффективности // Прикладная эконометрика. 2006. Вып. № 4. С. 43–55.
- Афанасьев М.Ю., Айвазян С.А. Оценка экономической эффективности перехода к производственному потенциалу // Прикладная эконометрика. 2009. Вып. № 3(15). С. 74–89.
- Бессонова Е. Оценка эффективности производства российских промышленных предприятий: препринт РЭШ. 2007.
- Пеникас Г.И. Модели «копула» в приложении к задачам финансов // Журнал новой экономической ассоциации. 2010. № 7. С. 24–44.
- Aigner D., Lovell C.A.K., Schmidt P. Formulation and Estimation of Stochastic Frontier Production Function Models // Journal of Econometrics. 1977. Vol. 6. P. 21–37.
- Almanidis P., Sickles R. Skewness Issue in Stochastic Frontiers Models: Fact or Fiction? // Exploring Research Frontiers in Contemporary Statistics and Econometrics. 2012.
- Battese G.E, Coelli T.J. Prediction of Firm-level Technical Efficiencies: With a Generalized Frontier Production Function and Panel Data // Journal of Econometrics. 1988. Vol. 38. P. 387–399.

- Battese G.E., Coelli T.J.* A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data // *Empirical Econometrics*. 1995. Vol. 20. P. 335–332.
- Battese G.E., Tessema G.A.* Estimation of Stochastic Frontier Production Functions with Time-varying Parameters and Technical Efficiencies Using Panel Data from Indian Villages // *Econometrics and Applied Statistics*. 1993. Vol. 9. P. 313–333.
- Bauer G.S., Berger A.N., Humphrey D.B.* Efficiency and Productivity Growth in US Banking. Oxford University Press, 1993. Vol. 12. P. 386–413.
- Belotti F., Llardi G.* Consistent Estimation of the «True» Fixed Effects Stochastic Frontier Model, Ceis Tor Vergata // *Research Paper Series*. 2012. Vol. 10. Iss. 5. № 231.
- Berg S.* Water Utility Benchmarking: Measurement, Methodology, and Performance Incentives 2012. International Water Association, 2012.
- Carree M.A.* Technological Inefficiency and the Skewness of the Error Component in Stochastic Frontier Analysis // *Economics Letter*. 2002. № 77. P. 101–107.
- Coelli T.J.* Estimators and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis // *Journal of Productivity Analysis*. 1995. Vol. 6. P. 247–268.
- Coelli T.J.* An Introduction to Efficiency and Productivity Analysis. 2<sup>nd</sup> ed. Springer, 2005.
- Cornwell C., Schmidt P., Sickles R.C.* Production Frontiers with Cross-sectional and Time Series Variation in Efficiency Levels // *Journal of Econometrics*. 1990. Vol. 46. P. 185–200.
- Debreu G.* The Coefficient of Resource Utilization // *Journal of the Econometric Society*. 1951. Vol. 19. P. 273–292.
- Devereux M., Schiantarelli F.* Investment, Financial Factors, and Cash Row: Evidence from U.K. Panel Data // *Empirical Econometrics*. 1990. Vol. 12. P. 306–318.
- Emvalomatis G.* Adjustment and Unobserved Heterogeneity in Dynamic Stochastic Frontier Models // *Journal of Product Analysis*. 2012. Vol. 37. P. 7–16.
- Farrel M.G.* The Measurement of Productive Efficiency // *Journal of the Royal Statistical Society*. 1957. Vol. 120. P. 97–110.
- Greene W.* Distinguishing between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organization's Panel Data on National Health Care Systems // *Econometric and Health Economics*. 2004. Vol. 13. P. 958–990.
- Greene W.* Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model // *Journal of Econometrics*. 2005. Vol. 126. P. 269–303.
- Greene W., Mutter R., Spector W., Rosko M., Mukamel D.* Investigating the Impact of Endogeneity on Inefficiency Estimates in the Application of Stochastic Frontier Analysis to Nursing Homes // *Journal of Productivity Analysis*. 2013. Vol. 39. P. 101–110.
- Greene W.* *Econometric Analysis*. 7<sup>th</sup> ed. Prentice Hall, 2008.
- Heshmati A., Kumbhakar S.C., Sun K.* Estimation of Productivity in Korean Electric Power Plants: A Semiparametric Smooth Coefficient Model. IZA Discussion Papers. № 7277. 2013.
- Jondrow J., Lovell C.A.K., Materov I.S., Schmidt P.* On the Estimation of Technical Inefficiency in Stochastic Frontier Production Function Model // *Journal of Econometrics*. 1982. Vol. 19. P. 233–239.
- Kim J.W., Lee J.Y.* Sources of Productive Efficiency: International Comparison of Iron and Steel Firms // *Resources Policy*. 2006. Vol. 31. P. 239–246.
- Kumbhakar S.C., Heshmati A., Hjalmarsson L.* Efficiency of the Swedish Pork Industry: A Farm Level Study Using Rotating Panel Data 1978–1988 // *European Journal of Operational Research*. 1995. Vol. 180. P. 519–533.
- Kumbhakar S.C., Lien G.D., Hardaker J.B.* Technical Efficiency in Competing Panel Data Models: A Study of Norwegian Grain Farming // *Applied Economics*. 2011. Vol. 12. P. 218–232.
- Kumbhakar S.C., Lovell C.A.K.* *Stochastic Frontier Analysis*. Cambridge University Press, 2000.
- Lee L.-F.* A Test for Distributional Assumptions for the Stochastic Frontier Functions // *Journal of Econometrics*. 1983. Vol. 22. P. 245–267.

*Mundlak Y., Butzer R., Larson D.F.* Heterogeneous Technology and Panel Data: The Case of the Agricultural Production Function. World Bank Policy Research Working Paper. № 4536. 2011.

*Revoredo-Giha C., Milne C.E., Lea P.E., Cho W.J.* Efficiency of Scottish Farms: A Stochastic Cost Frontier Analysis // *Empirical Econometrics*. 2009. Vol. 10. P. 40–57.

*Smith M.D.* Stochastic Frontier Models with Dependent Error Components // *The Econometrics Journal*. 2008. Vol. 11. Iss. 1. P. 172–192.

*Styrin K.A.* What Explains Differences in Efficiency across Russian Banks? // *Economics Education and Research Consortium*. 2005. Final Report.

*Talluri S.* Data Envelopment Analysis: Models and Extensions // *Decision Line*. 2000. May.

*Talluri S., Sarkis J.* A Decision Model for Evaluation of Flexible Manufacturing Systems in the Presence of Both Cardinal and Ordinal Factors // *International Journal of Production Research*. 1999. Vol. 37. P. 2927–2938.

*Tsionas E.G.* An Introduction to Efficiency Measurement Using Bayesian Stochastic Frontier Models // *Global Business and Economics Review*. 2005. Vol. 3. P. 287–311.

*Wang H-J., Chen Yi-Yi.* A Method of Moments Estimator for a Stochastic Frontier Model with Errors in Variables // *Economic Letters*. 2004. Vol. 85. P. 221–228.

*Wang H-J., Ho C-W.* Estimating Fixed-effect Panel Stochastic Frontier Models by Model Transformation // *Journal of Econometrics*. 2010. Vol. 157. P. 286–296.

*Wang W.S., Schmidt P.* On the Distribution of Estimated Technical Efficiency in Stochastic Frontier Models // *Journal of Econometrics*. 2009. Vol. 148. P. 36–45.

*Wang W. S., Amsler C., Schmidt P.* Goodness of Fit Tests in Stochastic Frontier Models // *Journal of Productivity Analysis*. 2011. Vol. 35. P. 95–118.



## Methods of Estimating of the Efficiency in Stochastic Frontier Models

**Malakhov Dmitry<sup>1</sup>, Pilnik Nikolay<sup>2</sup>**

<sup>1</sup> National Research University «Higher School of Economics»,  
20, Myasnitskaya ul., Moscow, 101990, Russian Federation  
E-mail: d.malakhov@gmail.com

<sup>2</sup> National Research University «Higher School of Economics»,  
20, Myasnitskaya ul., Moscow, 101990, Russian Federation  
E-mail: u4d@yandex.ru

This paper discusses the problems of modeling efficiency of firms. There are two the most popular methods to estimate efficiency of firms: DEA (data envelopment analysis) and SFA (stochastic frontier analysis), and popularity of the last one is fast growing. There are a lot of different SFA-models, so most researches often choose in advance one or two models, which they are going to estimate. So survey of different SFA-models is one of goals of this paper.

We discuss 15 popular SFA-models. Also we discuss problems of SFA models and their prospects. In our paper we compare models, estimated by classical method of moments (MoM), and models, estimated by maximum likelihood approach (MML). Today there are no such papers, so we try to discuss pros and cons of using method of moments approach in SFA models. Interesting, that this method is very unpopular today, but its' estimates are asymptotical normal and consistent.

Because there are no formal criteria to compare different SFA-models, we investigate the estimation results from 9 SFA-models on the concrete industry data. We use correlation analysis of estimates of efficiency ranks and also we try to find out the causes of the most serious differences between models.

**Key words:** stochastic frontier analysis; econometric modeling; efficiency; method of moments; survey; concrete industry.

**JEL Classification:** C4.

\* \*  
\*

### References

- Afanas'ev M.Yu. (2006) Model' proizvodstvennogo potentsiala s upravlyaemyimi faktorami effektivnosti [Model of the Production Capacity with Managed Efficiency Factors]. *Prikladnaya ekonometrika*, no 4, pp. 43–55.
- Afanas'ev M.Yu., Aivazyan S.A. (2009) Otsenka ekonomicheskoi effektivnosti perekhoda k proizvodstvennomu potentsialu [Evaluation of Economic Efficiency of Transition to Production Potential]. *Prikladnaya ekonometrika*, no 3(15), pp. 74–89.
- Bessonova E. (2007) *Otsenka effektivnosti proizvodstva rossiiskikh promyshlennykh predpriyatii* [Estimation of Efficiency of Russian Industrial Enterprises]. Working Paper, Moscow: RESh.
- Penikas G.I. (2010) Modeli «kopula» v prilozhenii k zadacham finansov [Model «Copula» in the Annex to the Tasks of Finance]. *Zhurnal novoi ekonomicheskoi assotsiatsii*, no 7, pp. 24–44.
- Aigner D., Lovell C.A.K., Schmidt P. (1977) Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics*, vol. 6, pp. 21–37.
- Almanidis P., Sickles R. (2012) Skewness Issue in Stochastic Frontiers Models: Fact or Fiction? *Exploring Research Frontiers in Contemporary Statistics and Econometrics*.
- 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*, vol. 38, pp. 387–399.
- Battese G.E., Coelli T.J. (1995) A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data. *Empirical Econometrics*, vol. 20, pp. 335–332.
- Battese G.E., Tessema G.A. (1993) Estimation of Stochastic Frontier Production Functions with Time-varying Parameters and Technical Efficiencies Using Panel Data from Indian Villages. *Econometrics and Applied Statistics*, vol. 9, pp. 313–333.
- Bauer G.S., Berger A.N., Humphrey D.B. (1993) *Efficiency and Productivity Growth in US Banking*. Oxford University Press, vol. 12, pp. 386–413.
- Belotti F., Llardi G. (2012) *Consistent Estimation of the «True» Fixed Effects Stochastic Frontier Model*, *Ceis Tor Vergata*. Research Paper Series, vol. 10, iss. 5, no 231.
- Berg S. (2012) *Water Utility Benchmarking: Measurement, Methodology, and Performance Incentives 2012*, International Water Association.
- Carree M.A. (2002) Technological Inefficiency and the Skewness of the Error Component in Stochastic Frontier Analysis. *Economics Letter*, no 77, pp. 101–107.
- Coelli T.J. (1995) Estimators and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis. *Journal of Productivity Analysis*, vol. 6, pp. 247–268.
- Coelli T.J. (2005) *An Introduction to Efficiency and Productivity Analysis*, 2<sup>nd</sup> ed., Springer.
- Cornwell C., Schmidt P., Sickles R.C. (1990) Production Frontiers with Cross-sectional and Time Series Variation in Efficiency Levels. *Journal of Econometrics*, vol. 46, pp. 185–200.
- Debreu G. (1951) The Coefficient of Resource Utilization. *Journal of the Econometric Society*, vol. 19, pp. 273–292.
- Devereux M., Schiantarelli F. (1990) Investment, Financial Factors, and Cash Row: Evidence from U.K. Panel Data. *Empirical Econometrics*, vol. 12, pp. 306–318.
- Emvalomatis G. (2012) Adjustment and Unobserved Heterogeneity in Dynamic Stochastic Frontier Models. *Journal of Product Analysis*, vol. 37, pp. 7–16.
- Farrel M.G. (1957) The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society*, vol. 120, pp. 97–110.
- Greene W. (2004) Distinguishing between Heterogeneity and Inefficiency: Stochastic Frontier Analysis of the World Health Organization's Panel Data on National Health Care Systems. *Econometric and Health Economics*, vol. 13, pp. 958–990.

Greene W. (2005) Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model. *Journal of Econometrics*, vol. 126, pp. 269–303.

Greene W., Muttter R., Spector W., Rosko M., Mukamel D. (2013) Investigating the Impact of Endogeneity on Inefficiency Estimates in the Application of Stochastic Frontier Analysis to Nursing Homes. *Journal of Productivity Analysis*, vol. 39, pp. 101–110.

Greene W. (2008) *Econometric Analysis*, 7<sup>th</sup> ed., Prentice Hall.

Heshmati A., Kumbhakar S.C., Sun K. (2013) *Estimation of Productivity in Korean Electric Power Plants: A Semiparametric Smooth Coefficient Model*. IZA Discussion Papers, no 7277.

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*, vol. 19, pp. 233–239.

Kim J.W., Lee J.Y. (2006) Sources of Productive Efficiency: International Comparison of Iron and Steel Firms. *Resources Policy*, vol. 31, pp. 239–246.

Kumbhakar S.C., Heshmati A., Hjalmarsson L. (1995) Efficiency of the Swedish Pork Industry: A Farm Level Study Using Rotating Panel Data 1978–1988. *European Journal of Operational Research*, vol. 180, pp. 519–533.

Kumbhakar S.C., Lien G.D., Hardaker J.B. (2011) Technical Efficiency in Competing Panel Data Models: A Study of Norwegian Grain Farming. *Applied Economics*, vol. 12, pp. 218–232.

Kumbhakar S.C., Lovell C.A.K. (2000) *Stochastic Frontier Analysis*, Cambridge University Press.

Lee L.-F. (1983) A Test for Distributional Assumptions for the Stochastic Frontier Functions. *Journal of Econometrics*, vol. 22, pp. 245–267.

Mundlak Y., Butzer R., Larson D.F. (2011) *Heterogeneous Technology and Panel Data: The Case of the Agricultural Production Function*. World Bank Policy Research Working Paper, no 4536.

Revoredo-Giha C., Milne C.E., Lea P.E., Cho W.J. (2009) Efficiency of Scottish Farms: A Stochastic Cost Frontier Analysis. *Empirical Economics*, vol. 10, pp. 40–57.

Smith M.D. (2008) Stochastic Frontier Models with Dependent Error Components. *The Economics Journal*, vol. 11, iss. 1, pp. 172–192.

Styrin K.A. (2005) *What Explains Differences in Efficiency across Russian Banks?* Economics Education and Research Consortium, final report.

Talluri S. (2000) Data Envelopment Analysis: Models and Extensions. *Decision Line*. May.

Talluri S., Sarkis J. (1999) A Decision Model for Evaluation of Flexible Manufacturing Systems in the Presence of Both Cardinal and Ordinal Factors. *International Journal of Production Research*, vol. 37, pp. 2927–2938.

Tsionas E.G. (2005) An Introduction to Efficiency Measurement Using Bayesian Stochastic Frontier Models. *Global Business and Economics Review*, vol. 3, pp. 287–311.

Wang H.-J., Chen Yi-Yi (2004) A Method of Moments Estimator for a Stochastic Frontier Model with Errors in Variables. *Economic Letters*, vol. 85, pp. 221–228.

Wang H.-J., Ho C.-W. (2010) Estimating Fixed-effect Panel Stochastic Frontier Models by Model Transformation. *Journal of Econometrics*, vol. 157, pp. 286–296.

Wang W.S., Schmidt P. (2009) On the Distribution of Estimated Technical Efficiency in Stochastic Frontier Models. *Journal of Econometrics*, vol. 148, pp. 36–45.

Wang W. S., Amsler C., Schmidt P. (2011) Goodness of Fit Tests in Stochastic Frontier Models. *Journal of Productivity Analysis*, vol. 35, pp. 95–118.