

Правительство Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
**«Национальный исследовательский университет
«Высшая школа экономики»**

Факультет экономики

Специализация: Математические методы анализа экономики

Кафедра математической экономики и эконометрики

ДИПЛОМНАЯ РАБОТА

Операционная эффективность энергораспределяющих предприятий в России

Выполнил

Студент группы 71ММАЭ2

*Тимошилов Иван
Викторович*

Научный руководитель

*Доцент кафедры
математической
экономики и эконометрики*

*Фурманов Кирилл
Константинович*

Москва, 2013

Оглавление

ВВЕДЕНИЕ	4
1. ОБЗОР ТЕОРЕТИЧЕСКИХ МЕТОДОВ.	7
1.1. Непараметрические методы.	7
1.1.1. Общие замечания	7
1.1.2. Метод DEA	7
1.2. Параметрические методы (регрессионные модели затрат)	8
1.2.1. Общие замечания	8
1.2.2. Методы оценивания функции затрат.	10
1.2.2.1. COLS	10
1.2.2.2. MOLS	13
1.2.2.3. SFA	16
1.2.3. Моделирование границы затрат на панельных данных.	18
1.2.3.1. Метод Наименьших Квадратов	20
Регрессия Pooled.	20
Модель с фиксированными эффектами, FE.	20
Модель со случайными эффектами, RE.	21
1.2.3.2. Метод максимального правдоподобия, ML	22
SFA без временного тренда.	22
SFA с временным трендом.	23
Иные модели SFA на панельных данных.	24
1.2.3.3. «Истинные фиксированные эффекты»	24
1.3. Выявление наблюдений-выбросов	25
2. ОПЫТ ДРУГИХ СТРАН.	29
2.1. Источники информации	29
2.2. Методы государственного регулирования отрасли распределения электроэнергии	30
2.3. Применение методов оценивания границы затрат в отрасли распределения электроэнергии в других странах	32
2.3.1. Швейцария 2003	32
2.3.2. Норвегия 2013.	35
3. ОПИСАНИЕ ПЕРЕМЕННЫХ И ИСТОЧНИКИ ДАННЫХ.	38
3.1. Источники данных	38
3.2. Зависимая переменная	38
3.3. Независимые переменные	39

4. ОЦЕНИВАНИЕ МОДЕЛЕЙ	43
4.1. Обозначения и сокращения	43
4.2. Уравнения регрессионных моделей.	44
4.3. Методы оценивания	44
4.4. Результаты оценивания	45
4.5. Сравнение трех моделей	47
4.5.1. МНК за 2011 год	47
4.5.2. Модель со случайными эффектами (RE)	47
4.5.3. Модель Стохастической Границы (SFA, TI)	48
4.6. Сравнение методов оценивания	48
5. ЗАКЛЮЧЕНИЕ	50
6. СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ	53
7. ПРИЛОЖЕНИЯ	56
Приложение 1. Результаты оценивания регрессии (модель 1) методом МНК по данным 2011 года без учета выбросов	56
Приложение 2. Результаты оценивания регрессии (модель 1) методом МНК по данным 2011 года с учетом выбросов	57
Приложение 3. Результаты оценивания модели 1 на панельных данных (метод RE) без учета выбросов.	58
Приложение 5. SFA без учета выбросов	60
Приложение 6. Сравнение различных методов оценивания для моделей 2 и 3.	61

ВВЕДЕНИЕ

В работе излагается подход к анализу операционной эффективности энергосетевых компаний в России. Предмет исследования – их подконтрольные операционные расходы. Цель исследования – разработка методологии сравнения однородных компаний и расчета на основе этого сравнения индивидуальных X-факторов на основе операционных (некапитальных) затрат. Объект исследования – компании, входящие в «Холдинг МРСК» (наиболее крупные в отрасли, достаточно однородные для сравнения).

Задачи исследования: во-первых, проанализировать и обобщить мировой опыт анализа эффективности распределительных сетей; во-вторых, определить наиболее актуальные и применимые методы, подходящие для российской специфики; в-третьих, разработать теоретический дизайн моделей исходя из экономических и технологических соображений (т.е. осуществить выбор функциональной формы и состава независимых переменных); в-четвертых, оценить эти модели наиболее подходящими методами и сравнить результаты; в-пятых, дать рекомендацию по использованию разработанной методологии.

Результаты исследования могут быть использованы двух основных направлениях: во-первых, в инвестиционной деятельности при оценке инвестиционной привлекательности компании, во-вторых, в регуляторике (например в тарифном регулировании).

Энергетические сети – стратегическая отрасль в экономике каждой страны, которая всюду находится под пристальным вниманием государственных органов и многие аспекты ее деятельности регулируются законодательными нормами. В последние 10 лет многие страны (среди них Австрия, Германия, Англия, Норвегия и другие) с целью снижения затрат в

отрасли использовали в регулировании бенчмаркинг, где применялись в том числе эконометрические модели.

Общая канва бенчмаркинга следующая: компании отрасли (обычно речь идет о крупнейших, если их больше 100, либо о всех компаниях отрасли, если их немного) сопоставляются между собой каким-либо методом. На входе используются данные по затратной части (либо операционные затраты, OPEX, либо общие затраты, TOTEX), а также данные по натуральным показателям компании (учет размера фирмы, ее технических особенностей и особенностей среды ее функционирования). На выходе метод дает оценки эффективности (измеряется в долях единицы, где 1 – значение полностью эффективной компании). Такое представление результата позволяет делать выводы об эффективности каждой компании, ранжировать их, а также легко сравнивать различные методы.

Наиболее широко используемые в практике методы – это, во-первых, Анализ Среды Функционирования (Data Envelopment Analysis) – непараметрический метод, основанный на линейном программировании, а во-вторых, параметрический Скорректированный Метод Наименьших квадратов (COLS), построенный на оценивании регрессии обычным МНК с последующей ее корректировкой с целью моделирования функции затрат. Оба метода просты, не предъявляют жестких требований к данным в плане однородности, однако имеют ряд серьезных недостатков. Наконец, в последние годы все более популярным становится метод SFA (Метод Стохастической Границы), который наряду с COLS является параметрическим эконометрическим методом, где, однако, используется предпосылка о специальном виде распределения случайного члена. Метод более сложен и более требователен к данным, однако имеет ряд преимуществ перед COLS и DEA.

В Части 1 работы будет проведен обзор теоретических методов моделирования границы затрат, из которых некоторые отобраны для

дальнейшего применения на российских данных. Также в Части 1 будет проведен обзор популярного метода выявления наблюдений-выбросов.

В Части 2 будут рассмотрен иностранный опыт (главном образом, в европейских странах) государственного регулирования отрасли распределения электроэнергии с разбором конкретных эконометрических моделей.

В Части 3 дается описание данных по компаниям «Холдинга МРСК» и обосновывается выбор набора объясняющих переменных.

В Части 4 на этих данных будут оценены модели рассмотренными в Части 2 методами, в частности COLS и SFA, а также будет проведен сравнительный анализ результатов.

1. Обзор теоретических методов.

1.1. Непараметрические методы.

1.1.1. Общие замечания

Непараметрические методы моделирования границы затрат стали применяться уже 50 лет назад. Их отличительными особенностями является простота и отсутствие необходимости введения каких-либо предпосылок относительно функциональной формы границы затрат. Изложенный в работе Farrell (1957) метод DEA (Data Envelopment Analysis) находит свое применение по сей день не только в частных исследованиях, но и в работе регулирующих органов.

В данном исследовании непараметрические методы моделирования границы затрат не будут применяться к данным, поэтому ниже изложено лишь краткое описание метода DEA.

1.1.2. Метод DEA

Метод DEA основан на линейном программировании. Как и другие методы моделирования границы затрат, метод DEA изначально применялся при моделировании границы производственных возможностей; обобщение до случая границы затрат заключается в добавлении знака (-). Впервые метод был представлен в статье Farrell (1957). Задача нахождения границы производственных возможностей по сути есть задача нахождения огибающей поверхности, как показано на Рис. 1.1:

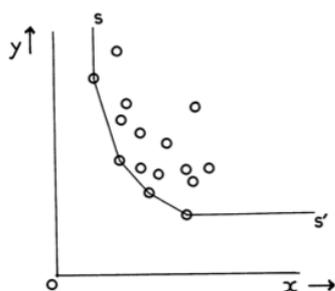


Рис. 1.1. Нахождение границы производственных возможностей методом DEA.

По построению метод присваивает 100% значение эффективности нескольким (граничным) компаниям. Остальные компании (внутри облака)

эффективны менее, чем на 100%. Эффективность рассчитывается из простого соотношения отрезков: отрезок, соединяющий точку внутри границы производственных возможностей и начало координат, делится на две части линией границы. Отношение длины отрезка под границей к длине всего отрезка есть эффективность.

Метод DEA прост, не требует введения предпосылок о функциональной форме зависимости, не требователен к однородности данных. Однако он имеет и ряд недостатков: во-первых, на 100% фирм оказывается всегда несколько, во-вторых, значения эффективности занижены по сравнению с параметрическими методами. В-третьих, применяя DEA, исследователь лишен возможности тестировать статистические гипотезы, хотя это может оказаться весьма востребованным.

В мировой практике DEA очень распространен: почти все регуляторы (например в Германии, Австрии, Великобритании) используют DEA наряду с другими методами оценки эффективности. Однако в последние годы все чаще применяются параметрические методы оценки эффективности. Подробный анализ этих методов представлен ниже в п.1.2.

1.2. Параметрические методы (регрессионные модели затрат)

1.2.1. Общие замечания

Единая логика характерна для всего многообразия регрессионных моделей функции затрат. Затраты компаний (C) рассматриваются как функция от ее натуральных показателей (x), цен на факторы производства (w) и характеристик среды (z): $C = \exp[f(x, w, z) \cdot \varepsilon]$ или в логарифмах¹ $\ln C = f(x, w, z) + \varepsilon$, где ε отражает вклад неучтенных факторов. Для удобства записи обозначим все факторные переменные как x_j ; $\ln C = f(x_1, \dots, x_k) + \varepsilon$. Форма функции $f(x)$ не ограничивается методом, однако в практических работах встречаются в основном только три функциональные формы:

¹ В большинстве исследований зависимая переменная входит в модель в

Линейная: $C = \alpha + \sum_{j=1}^k \beta_j x_j$;

функция Кобба-Дугласа: $\ln[C] = \alpha + \sum_{j=1}^k \beta_j \ln [x_j]$;

функция ТрансЛог: $\ln[C] = \alpha + \sum_{j=1}^k \beta_j \ln[x_j] + \sum_j \sum_l \beta_{jl} \ln [x_j] \ln [x_l]$.

Наиболее гибкая форма у транслогарифмической функции (фактически это разложение в ряд Тэйлора до членов второго порядка функции в логарифмах). Однако обратная сторона гибкости – слишком большое число регрессоров. При количестве факторов $k = 3$ число оцениваемых параметров будет 10, а при $k = 4$ уже 15. Регрессия будет страдать от мультиколлинеарности (МК) факторов, так как линейные и квадратичные члены очень сильно коррелированы, что приведет к неустойчивости и статистической незначимости оценок коэффициентов. Поэтому функцию ТрансЛог имеет смысл использовать, когда в распоряжении исследователя имеется очень мало факторных переменных. В противном случае лучше выбрать менее гибкую функциональную форму, но включить в модель больше факторов.

Функция Кобба-Дугласа менее гибкая по сравнению с транслогарифмической, так как не содержит члены второго порядка, однако она лишена проблемы перегруженности переменными и, следовательно, мультиколлинеарности (МК может присутствовать и по причине коррелированности между собой различных факторов, но в таком случае изменение функциональной формы не имеет значения: проблема все равно останется). Наряду с транслогарифмической, функция Кобба-Дугласа подразумевает логарифмирование всех переменных, что позволяет сделать выборку более однородной. Зачастую одно лишь логарифмирование снимает проблему присутствия в выборке компаний разных масштабов, так что тесты на гетероскедастичность (ГС) дают отрицательный результат.

Линейная функция – самая простая, не предполагает преобразования переменных, однако имеет ряд недостатков. Во-первых, линия регрессии

может уходить в отрицательную область при малых значениях факторов (для малых компаний), что противоречит экономическому смыслу. Во-вторых, без логарифмирования компании разного размера становятся несопоставимы, что делает необходимым учет гетероскедастичности. Этому усложнения легко избежать, применяя функцию Кобба-Дугласа.

Функция Кобба-Дугласа – это компромиссный, промежуточный вариант между простотой и гибкостью функциональной формы, поэтому именно она используется в подавляющем большинстве работ.

Выбор функциональной формы осуществляется исследователем априори и обосновывается только исходя из ее преимуществ перед другими в данном конкретном исследовании. Это является узким местом регрессионного анализа вообще, а в частности параметрических методов оценивания функции затрат. Поэтому широко применяется описанный выше непараметрический метод DEA, одним из основных преимуществ которого является то, что специфицировать для него какую-либо функциональную форму не нужно.

1.2.2. Методы оценивания функции затрат.

1.2.2.1. COLS

При моделировании функции затрат применяется несколько методов оценивания. Самый простой из них – скорректированный метод наименьших квадратов (Corrected Ordinary Least Squares, COLS) и его различные варианты, которые обычно обозначаются в литературе как модифицированный метод наименьших квадратов (Modified Ordinary Least Squares, MOLS).

Впервые COLS как метод оценивания функции затрат был представлен сравнительно давно в работе Winsten (1957). Обоснования теоретических свойств оценок (доказательство их состоятельности), полученных методом COLS, появляется позже в работах Gabrielsen (1975) и Greene (1980).

Расширение метода для панельных данных представлено в работах Schmidt & Sickles (1984) и Simar (1992). Метод COLS заключается в следующем.

Функция затрат $\ln C = f(x_1, \dots, x_k; \alpha, \beta) + \varepsilon$ (с выбранной функциональной формой и набором факторов) оценивается методом наименьших квадратов (МНК):

$$\sum_{i=1}^n [\ln C_i - f(\vec{x}_i; \alpha, \beta)]^2 \xrightarrow{\alpha, \beta} \min$$

(i – номер наблюдения, n – число наблюдений)

Результат оценивания – оценки коэффициентов регрессии $\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_k$, уравнение регрессии $\ln \hat{C} = f(x_1, \dots, x_k; \hat{\alpha}, \hat{\beta})$ и остатки $e_i = \ln C_i - \ln \hat{C}_i$. На следующем шаге из уравнения линии регрессии параллельным сдвигом вниз получается граница затрат (в противовес методу SFA ее можно назвать нестохастическая или детерминированная граница затрат): $\ln C_i^{fr} = \ln \hat{C}_i + e_{min}$, где $e_{min} = \min(e_i) < 0$. Схематически преобразование изображено на Рис.1.2.:

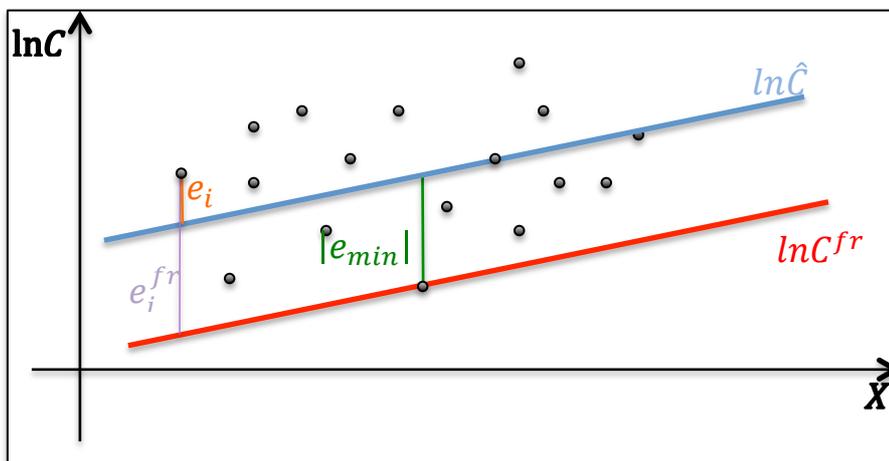


Рис.1.2. Сдвиг линии регрессии в методе COLS

Точка, для которой остаток минимальный (e_{min}), – это лучшая практика, так как она лежит на границе затрат, в то время как другие точки лежат над ней (т.е. все другие компании перерасходуют средства по сравнению с необходимым минимумом, то есть являются неэффективными). Такой результат – свойство метода COLS (всегда будет хотя бы одна 100%

эффективная компания и все другие в той или иной мере неэффективные). Значение $e_i^{fr} = e_i - e_{min}$ есть отклонение от границы затрат, иначе говоря, мера неэффективности (чем больше значение e_i^{fr} , тем больше перерасход средств). Именно она ложится в основу расчета эффективности затрат. Значение e_i^{fr} есть разница между фактическим значением затрат и их эффективным уровнем.

При использовании моделей, где затраты берутся в логарифмах (в частности при использовании функции Кобба-Дугласа), это значение рассчитывается как $e_i^{fr} = \ln[C_i] - \ln[C_i^{fr}] = \ln[C_i/C_i^{fr}]$. Фактические затраты C_i фирмы делятся на эффективные (т.е. объясненные моделью, C_i^{fr}) и неэффективные $C_i - C_i^{fr}$. Показатель эффективности затрат (Efficiency Score, ES) есть отношение эффективных затрат к фактическим $ES_i = C_i^{fr}/C_i$. Это значение всегда будет лежать в пределах от 0 до 1, причем $ES_i = 1$ если $C_i = C_i^{fr}$, то есть для наилучшей практики. Преобразуя формулы, получаем $ES_i = C_i^{fr}/C_i = [C_i/C_i^{fr}]^{-1} = \exp[\ln[C_i/C_i^{fr}]^{-1}] = \exp[-\ln[C_i/C_i^{fr}]] = \exp[-e_i^{fr}]$. Простота конечной формулы, а также свойство полученной по ней эффективности затрат $ES \in (0; 1]$ – еще один аргумент в пользу применения моделей в логарифмах.

Такова схема метода COLS. Ее изложение занимает не более двух страниц печатного текста, что говорит о простоте метода. Метод оценивания регрессии (МНК) всегда дает результат, в отличие от Метода Максимального Правдоподобия, ММП, предполагающий процедуру численной максимизации, которая может не дать решения. К плюсам метода стоит отнести очевидность способа перехода от линии регрессии к границе затрат: в методе COLS этот переход осуществляется по одной лучшей практике.

Однако метод имеет ряд существенных недостатков. Самый большой недостаток метода – интерпретация отклонений от границы затрат

полностью как неэффективность. Иными словами, метод не оставляет места ненаблюдаемым факторам, то есть предполагается, что в модели учтены *все* детерминанты затрат, а необъясненная дисперсия – и есть неэффективность. Эта предпосылка модели очень жесткая и, очевидно, она не выполняется в реальности. Вторым недостатком модели, жестко связанным с первым – занижение показателей эффективности в целом для всех компаний в выборке. По методу COLS иные фирмы могут иметь показатель эффективности 0.3, что крайне маловероятно в действительности. Слишком низкие значения ES, полученные по методу COLS, делают сомнительным прямое их применение в регуляторике (так как вводимые ограничения для отдельных компаний могут быть слишком жесткими), а скорее придают им рекомендательный характер. Также можно использовать ES по методу COLS для ранжирования компаний по эффективности, абстрагируясь от их абсолютных значений.

Указанные проблемы простого метода COLS привели к поиску более сложных, но и более гибких методов. В частности, попытки справиться с занижением значений ES, характерным для COLS, без изменения метода оценивания параметров (МНК как в COLS) привели к появлению различных видов Модифицированного МНК (Modified Ordinary Least Squares, MOLS)².

1.2.2.2. MOLS

Содержательно различные виды Модифицированного МНК (MOLS) повторяют логику COLS, однако либо отличаются способом перехода от линии регрессии к границе затрат, либо тем, что на последнем этапе значения ES «вытягиваются вверх» каким-либо способом. Основным минусом такого подхода является абсолютная произвольность способа, которым значения ES искусственно завышаются, в отличие от единственно логичного выравнивания по первой лучшей практике в методе COLS.

² Терминология впервые встречается в работе Lovell (1993)

Рассмотрим вариант MOLS, изложенный Greene в работе 'The Econometric Approach to Efficiency Analysis'³. Назовем условно этот метод «теоретическим», TMOLS. Остальные разновидности MOLS подробно рассматриваются в Части 2 исследования. Логика TMOLS и сравнение его с COLS представлены на Рис.1.3.

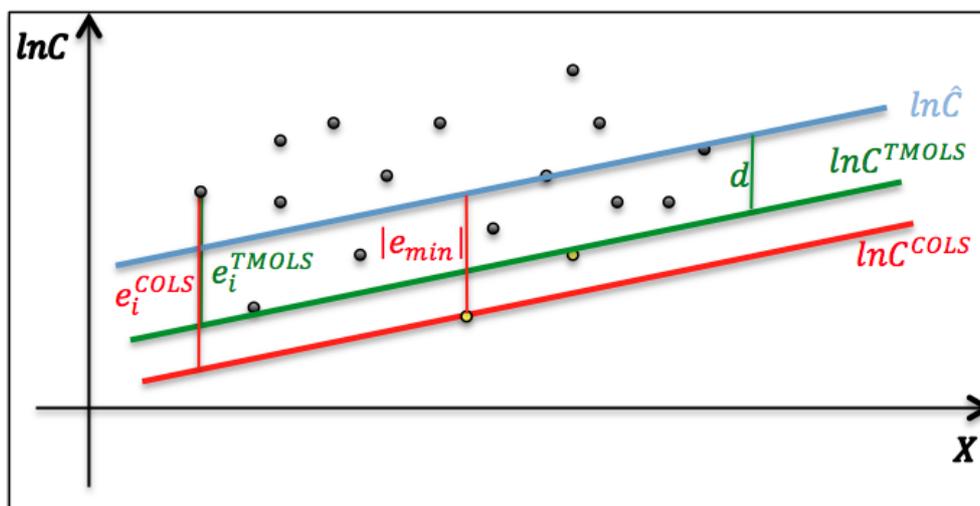


Рис.1.3. Методы TMOLS (Greene) и COLS

Отличие метода TMOLS от COLS заключается лишь в том, что линия регрессии $\ln\hat{C}$ сдвигается вниз не на минимальный остаток, а на некоторое расстояние d , заведомо меньшее, чем $|e_{min}|$. Таким образом разности фактических значений и значений на границе затрат становятся меньше: $e_i^{TMOLS} < e_i^{COLS}$, а значения ES, вычисляемые по прежней формуле $ES_i = \exp[-e_i^{fr}]$, выше: $ES_i^{TMOLS} > ES_i^{COLS}$. Для TMOLS уже не выполняется свойство $ES \in (0; 1]$, так как существуют точки лежащие под границей затрат (выделены светлым), то есть *сверхэффективные* компании. Для них выполняются соотношения $e_i^{TMOLS} < 0$; $ES_i^{TMOLS} > 1$.

Greene предлагает принимать за расстояние d некоторую состоятельную оценку математического ожидания случайной величины ε_i , имеющей ненулевое математическое ожидание. В терминах Greene⁴ в

³ Greene (2008)

⁴ Все выкладки Greene относятся к случаю моделирования производственной функции; при моделировании функции затрат меняется только знак $E[\varepsilon_i]$.

регрессии $\ln[y_i] = \alpha + \beta^T x_i + \varepsilon_i$ математическое ожидание $E[\varepsilon_i] \neq 0$, так как по смыслу эта случайная величина отражает отклонение от границы производственных возможностей (или границы затрат). В первом случае отклонения происходят внутрь границы, то есть $E[\varepsilon_i] < 0$, во втором – за вонне, $E[\varepsilon_i] > 0$. Для того чтобы оценить регрессию $\ln[y_i] = \alpha + \beta^T x_i + \varepsilon_i$ методом наименьших квадратов, необходимо следующее преобразование: $\ln[y_i] = (\alpha + E[\varepsilon_i]) + \beta^T x_i + (\varepsilon_i - E[\varepsilon_i]) = \alpha^* + \beta^T x_i + \varepsilon_i^*$, $E[\varepsilon_i^*] = 0$. МНК даст состоятельные оценки всех параметров регрессии, кроме α , так как неочевидно, каким образом оценить величину $E[\varepsilon_i]$.

Уязвимое место этого метода – выбор способа оценивания величины $E[\varepsilon_i]$ (т.е. расчета величины d), на которую сдвигается линия регрессии. Greene, предлагая использовать *метод моментов* (ММ), пишет следующее⁵: «Остатки МНК e_i – состоятельные точечные оценки соответствующих им случайных величин ε_i^* ⁶. Использовать их среднее бесполезно, так как оно равно нулю по построению. Однако, поскольку искомый сдвиг – величина постоянная, дисперсия или другой центральный момент более высокого порядка остатков МНК (взятых с отрицательным знаком) может служить состоятельной оценкой соответствующего момента величины ε_i . Поэтому, если параметры $E[\varepsilon_i]$ могут быть определены через дисперсию или моменты более высоких порядков величины ε_i^* или иные статистики, состоятельное оценивание параметров модели может быть завершено с помощью ММ. В качестве простого примера, предположим, что ε_i распределены экспоненциально с математическим ожиданием λ . Тогда дисперсия ε_i равна λ^2 , а *стандартное отклонение остатков МНК* $s.d.[e_i]$ является состоятельной оценкой для $E[\varepsilon_i] = \lambda$.⁷ Поскольку это однопараметрическое

⁵ Greene (2008), pp. 17-18.

⁶ В терминах Greene $\varepsilon_i^* = -u_i$; поскольку u_i – избыточное обозначение, далее в выкладках используется ε_i^* .

⁷ Поскольку по построению величины ε_i^* и ε_i отличаются только средним, а другие моменты (в том числе и дисперсия) у них совпадают, $s.d.[e_i] = \hat{\sigma}_{\varepsilon^*} = \hat{\sigma}_{\varepsilon} = \overline{E[\varepsilon_i]}$.

распределение, случайная величина ε_i может быть полностью описана этим одним параметром и функциями от него. Теперь оцененная линия регрессии может быть сдвинута вверх⁸ на величину $\widehat{E}[\varepsilon_i]$. Этот подход MOLS немного менее строгий чем упомянутый выше COLS, поскольку не все получаемые этим способом значения неэффективности будут иметь нужный знак»⁹. Таким образом, в методе TMOLS искомый сдвиг может быть определен как $d = s. d. [e] = \sqrt{\frac{\sum e_i^2}{n-1}}$, что всегда меньше чем $|e_{min}|$.

Метод MOLS справляется со вторым недостатком COLS, а именно с заниженными значениями ES, однако оставляет без внимания первый и важнейший недостаток: при изначальном оценивании регрессии методом наименьших квадратов случайная составляющая модели отвечает только за неэффективность, не оставляя места возможному шуму (неучтенным факторам). С этим недостатком COLS справляется *метод стохастической границы*, Stochastic Frontier Analysis (SFA).

1.2.2.3. SFA

Метод Стохастической Границы (SFA) был изложен почти одновременно в работах Aigner, Lowell, Schmidt (1977) и Meeusen and van den Broeck (1977) как метод оценивания производственной функции (максимальной границы производственных возможностей), который был впоследствии расширен на случай моделирования функции издержек (минимальной границы затрат). Обе могут быть записаны общей формулой $\ln[y] = f(x_1 \dots x_k) + \varepsilon$. Ключевая идея метода заключается в разделении случайного члена ε_i на две независимые составляющие v_i – шум и u_i – неэффективность (где $\forall i u_i \geq 0$): $\varepsilon_i = v_i - u_i$ для производственной функции (неэффективность с отрицательным знаком как движение вниз «внутрь»

⁸ Наверх в случае моделирования границы производственных возможностей; в нашем случае – при моделировании границы затрат – вниз.

⁹ «Сверхэффективные» компании будут давать противоположный знак; их значения будут лежать «по ту сторону границы»

границы) или $\varepsilon_i = v_i + u_i$ для функции издержек (неэффективность с положительным знаком как движение вверх, «вовне»). Распределение члена v_i , отражающего неучтенные факторы (шум) предполагается симметричным, например нормальным с нулевым средним и с дисперсией σ^2 , а распределение члена u_i , отвечающего за неэффективность и принимающего только неотрицательные значения, предполагается несимметричным, например экспоненциальным с параметром λ : $u_i \sim \exp(\lambda)$, или положительно-нормальным (полуноормальным, half-normal) с параметром разброса σ_u^2 : $u_i \sim N^+(0; \sigma_u^2)$, или усеченным нормальным (truncated normal) с параметром разброса σ_u^2 и модой μ : $u_i \sim N^+(\mu; \sigma_u^2)$. Следует отметить, что величина σ_u^2 не является дисперсией величины u_i ; для величины с нулевой модой $u_i \sim N^+(0; \sigma_u^2)$ дисперсия равна $Var[u_i] = \frac{\pi-2}{\pi} \sigma_u^2$.¹⁰

Любое из перечисленных выше распределений u_i имеет ненулевое математическое ожидание $E[u_i] \neq 0$, что в свою очередь делает всю случайную составляющую ε_i в среднем отличной от нуля $E[\varepsilon_i] \neq 0$. Это означает, что при использовании составного случайного члена в методе SFA нарушается одна из базовых предпосылок теоремы Гаусса-Маркова, что делает неприменимым метод наименьших квадратов как метод оценивания такой регрессии. Метод, подходящий для оценивания регрессии с составным случайным членом – метод максимального правдоподобия (ММП).

Впервые распределение двухчленной случайной величины ε_i (normal – half-normal distribution) было получено в работе Weinstein (1964) и применено к модели стохастической границы в первой работе по SFA Aigner, Lovell, and Schmidt (1977). Если случайные члены распределены как $v_i \sim N(0; \sigma_v^2)$, $u_i \sim N^+(0; \sigma_u^2)$, то функция плотности случайной величины $\varepsilon_i = v_i + u_i$ представляется формулой $f_\varepsilon(x) = \frac{2}{\sqrt{2\pi(\sigma_v^2 + \sigma_u^2)}} \cdot \left[\Phi\left(\frac{\pm x(\sigma_u/\sigma_v)}{\sqrt{\sigma_v^2 + \sigma_u^2}}\right) \right] \cdot \exp\left[\frac{-x^2}{2(\sigma_v^2 + \sigma_u^2)}\right]$,

¹⁰ Greene (2008), p. 32

где $\Phi(\cdot)$ – функция распределения стандартной нормальной случайной величины, а знак $\pm x$ зависит от модели: «+» для модели функции затрат, а «-» для модели производственной функции. Модель удобно перепараметризовать следующим образом: $\sigma_\varepsilon^2 = \sigma_v^2 + \sigma_u^2$; $\lambda = \sigma_u/\sigma_v$. Тогда функция плотности величины ε_i $f_\varepsilon(x) = \frac{2}{\sqrt{2\pi\sigma_\varepsilon^2}} \cdot \left[\Phi\left(\frac{\pm x\lambda}{\sigma_\varepsilon}\right) \right] \cdot \phi\left(\frac{x}{\sigma_\varepsilon}\right)$, где $\phi(\cdot)$ – функция плотности стандартной нормальной случайной величины. Величина λ отражает соотношение разбросов двух случайных компонент: при $\lambda \rightarrow 0$ модели отсутствует неэффективность и она может быть состоятельно оценена обычным МНК, а при $\lambda \rightarrow \infty$ в модели отсутствует шум и она вырождается в модель с детерминированной границей. При оценивании регрессии оценка $\hat{\lambda}$ может служить одним из критериев качества подгонки и признаком применимости метода SFA к используемым данным. Поскольку предпосылка о нормальном-полуномальном распределении случайной величины ε_i довольно ограничительная, несколько иных простых распределений было предложено к использованию в методе SFA, в частности нормальное-экспоненциальное в работах Aigner, Lovell, and Schmidt (1977) и Meeusen and van den Broeck (1977) и нормальное-гамма распределение в работах Stevenson (1980) и Greene (1990). В работе Stevenson (1980) также было рассмотрено расширение нормальной-полуномальной модели. Стивенсон отмечает, что предпосылка о нулевом среднем для положительно-нормального (полуномального) распределения не является необходимым, поэтому распределение случайного члена u_i может иметь второй параметр – среднее μ : $u_i \sim N^+(\mu; \sigma_u^2)$. За таким более общим видом положительно-нормального распределения закрепилось название усеченное нормальное (truncated normal).

1.2.3. Моделирование границы затрат на панельных данных.

Выше были рассмотрены методы оценивания границы затрат на данных типа cross-section (в один момент времени). Спецификой задачи является оценивание регрессии со сложной (состоящей из двух членов)

случайной компонентой: в методе SFA обе компоненты включаются в явном виде в модель и распределение обеих задается априори, а в методах COLS и MOLS разбиение на две компоненты появляются после преобразования уравнения регрессии: $\ln[y_i] = (\alpha + E[\varepsilon_i]) + \beta^T x_i + (\varepsilon_i - E[\varepsilon_i]) = \alpha^* + \beta^T x_i + \varepsilon_i^*$, $E[\varepsilon_i^*] = 0$, где $\alpha^* = \alpha + E[\varepsilon_i]$. Одна из основных проблем моделей на данных одного года (cross-section) – невозможность *без дополнительных предпосылок* состоятельно оценить неэффективность (оценку члена ε_i для COLS или u_i для SFA) в чистом виде для каждой фирмы. МНК дает состоятельные оценки для α^* , но не для α и $E[\varepsilon_i]$ (см. выше п.1.2.2.2 о способах разделения этих величин). В моделях SFA состоятельное оценивание u_i возможно, однако метод их оценивания (ММП) требует априорного определения распределений обоих случайных членов. Выбор распределений на практике происходит произвольным образом, тогда как он напрямую влияет на конечный результат. Использование же панельных данных позволяет состоятельно оценить неэффективность каждой фирмы *без дополнительных предпосылок*.

Краткий обзор различных методов оценивания на панельных данных границы производственных возможностей (легко обобщаемых на случай оценивания границы затрат) изложен в работе Schmidt and Sickles (1984). Ниже приводятся выдержки из этой работы, приспособленные к оцениванию границы затрат (предполагается, что затраты берутся в логарифмах). Рассмотрим уравнение границы затрат на панельных данных, где C – затраты, X – вектор объясняющих переменных, $t = \overline{1, T}$ – индекс времени, а $i = \overline{1, N}$ – индекс компании (N – количество компаний, T – количество периодов времени):

$$\ln[C_{it}] = \alpha + X'_{it}\beta + v_{it} + u_i \quad (1.1)$$

Случайная составляющая v_{it} – шум, u_i – неэффективность i -й компании. Распределения случайных величин априори не заданы, однако

известно, что величина $u_i \geq 0$, а также $E[v_{it}] = 0$, $E[u_i] = \mu > 0$. Удобно преобразовать уравнение:

$$\begin{aligned}\ln [C_{it}] &= (\alpha + \mu) + X'_{it}\beta + v_{it} + (u_i - \mu) \\ \ln [C_{it}] &= \alpha^* + X'_{it}\beta + v_{it} + u_i^* \quad (1.2)\end{aligned}$$

где $E[u_i^*] = 0$. Модель состоятельно можно оценить следующими способами.

1.2.3.1. Метод Наименьших Квадратов

Регрессия Pooled.

Предполагая только $E[v_{it} + u_i^*] = 0$, получаем состоятельные оценки для α^* и β . Однако в такой спецификации не учитывается панельная структура данных (случайная величина не разделяется на шум и индивидуальный эффект), поэтому на практике используются иные методы. Также, для получения состоятельных оценок в данном случае требуется предпосылка о некоррелированности регрессоров и случайных ошибок.

Уравнение (1.2) также представляется в виде

$$\ln [C_{it}] = \alpha_i + X'_{it}\beta + v_{it} \quad (1.3), \text{ где } \alpha_i = \alpha^* + u_i^* = \alpha + u_i.$$

В таком виде модель можно оценить стандартными приемами, описанными в литературе по регрессиям на панельных данных¹¹.

Модель с фиксированными эффектами, FE.

Оценивание модели с фиксированными эффектами *не требует* использования дополнительных предпосылок, таких как некоррелированность регрессоров и индивидуальных эффектов $\text{corr}(X; u) = 0$. Не требуется также знание распределений случайных членов. Оценивание может осуществляться двумя равнозначными способами: или включением в модель дамми-переменных для каждой компании (с исключением свободного члена), или с помощью преобразования 'within'. Последнее

¹¹ Mundlak (1978), Hausman (1978)

заключается в том, что регрессия оценивается на центрированных переменных: $\ln [C_{it}] - \frac{1}{T} \sum_{t=1}^T \ln [C_{it}] = \alpha_i + (X_{it} - \frac{1}{T} \sum_{t=1}^T X_{it})' \beta + v_{it}$. По построению в оценивании не участвуют переменные, не меняющиеся во времени, что является основным минусом FE-метода. Оценки β^{FE} состоятельны при условии $N \rightarrow \infty$ либо $T \rightarrow \infty$; состоятельное оценивание α^{FE} достижимо только при условии $T \rightarrow \infty$. Возвращаясь к специальному случаю моделирования границы затрат, определим лучшую практику и отклонения от нее в терминах метода COLS: $\hat{\alpha}^{FR} = \min \{\hat{\alpha}_i\}$, тогда уравнение границы затрат $\ln [C_{it}] = \hat{\alpha}^{FR} + X_{it}' \hat{\beta}$. Неэффективность каждой фирмы рассчитывается по формуле $\hat{u}_i = \hat{\alpha}_i - \hat{\alpha}^{FR}$. Отметим, что такой метод дает *неизменные во времени* значения неэффективности для каждой фирмы. Границы применимости метода: при $T \rightarrow \infty$ возможно состоятельное оценивание индивидуальных эффектов и, таким образом, относительное ранжирование фирм, а при $N \rightarrow \infty$ возможно отделение общего свободного члена от индивидуальных эффектов и, следовательно, сравнение эффективности со 100% стандартом.

Модель со случайными эффектами, RE.

Модель со случайными эффектами – полный аналог модели с фиксированными эффектами, только с использованием дополнительной предпосылки $corr(X; u) = 0$. Проверка гипотезы о выполнении этой предпосылки осуществляется с помощью теста Хаусмана¹². Этот же тест, построенный на сравнении оценок двух различных моделей, одна из которых использует дополнительную предпосылку, а другая – нет, применяется при решении вопроса о выборе между моделями FE и RE (тогда, когда нет иных доводов в пользу одной или другой модели). В отличие от модели FE, в спецификацию модели со случайными эффектами можно в качестве объясняющих включать переменные, не варьирующие во времени, что и

¹² Hausman (1978), Hausman and Taylor (1981)

является основным плюсом модели. Индивидуальные эффекты u_i в этой модели являются случайными величинами, однако по-прежнему для состоятельного оценивания не требуется знание их функции распределения. Модель (2) со случайными эффектами оценивается *обобщенным методом наименьших квадратов* (GLS). Если ковариационная матрица случайного члена $v_{it} + u_i^*$ неизвестна, а это как раз реалистичный случай, то она оценивается вместе с α^*, β . Обычно предполагается, что ковариационная матрица построена блоками: $v_{it} + u_i^*$ и $v_{jt} + u_j^*$ некоррелированы $\forall i \neq j$ (эффекты разных фирм независимы), а v_{it} некоррелирован с v_{js} $\forall i, j; \forall t, s$, поэтому она определяется всего двумя параметрами σ_u^2 и σ_v^2 . Для состоятельного оценивания σ_v^2 требуется только $N \rightarrow \infty$, поэтому в случае, если N большое, а T – маленькое (обычный случай для панельных данных), оценки RE эффективнее оценок FE. При больших T разница нивелируется. Резюме: модель со случайными эффектами позволяет включать в спецификацию неизменные во времени регрессоры; оценки RE эффективнее чем оценки FE при малых T . Однако модель требует введение предпосылки $corr(X; u) = 0$, которая не всегда выполняется в реальности и требует тестирования.

1.2.3.2. Метод максимального правдоподобия, ML

SFA без временного тренда.

ММП требует введения наиболее жестких предпосылок – о форме распределения случайных членов. Использование панельных данных, как было показано выше, позволяет отказаться от столь ограничительных моделей, однако их применение возможно и имеет ряд преимуществ. Оценки ML всегда состоятельны, и эффективны при $N \rightarrow \infty$ ($\forall T$), что делает метод SFA очень популярным при применении к панелям с короткими временными рядами. Впервые модель стохастической границы (производственных возможностей) на панельных данных была представлена в работе Pitt and Lee (1981). Применительно к задаче оценивания границы затрат, модель Pitt &

Lee записывается следующим образом: $\ln[C_{it}] = \alpha + X'_{it}\beta + v_{it} + u_i$, где v_{it} и u_j независимы $\forall i, j; \forall t$ и распределены нормально и положительно-нормально соответственно: $v_{it} \sim N(0; \sigma_v^2)$, $u_j \sim N^+(0; \sigma_u^2)$. Эта модель является обобщением для случая $T > 1$ самой первой модели SFA, представленной в работе Aigner, Lovell, and Schmidt (1977). Заметим, что как и два предыдущих метода, модель Pitt & Lee дает неизменные во времени оценки эффективности для каждой фирмы. Всё разнообразие распределений u_j , рассмотренное выше для моделей SFA на данных одного года (cross-section), применимо и для панельных регрессий, однако следует помнить, что включение каждого дополнительного параметра утяжеляет модель и снижает эффективность оценивания. Метод оценивания ММП чувствителен к числу параметров. Если выборка не очень большая, а параметров много, то процедура численной максимизации может не сойтись. Поэтому исследователю, ограниченному размером выборки, надлежит искать золотую середину между гибкостью функциональной формы и простотой модели.

SFA с временным трендом.

Предпосылка о неизменности эффективностей фирм во времени может оказаться слишком сильной. В работе Battese и Coelli (1992) предложена SFA-модель производственной функции на панельных данных, где эффективность всех компаний изменяется во времени с постоянным темпом роста. Модель с изменяющейся во времени эффективностью записывается уравнением $\ln[C_{it}] = \alpha + X'_{it}\beta + v_{it} + u_{it}$ (отличие только в члене u_{it} с индексом времени). Неэффективность u_{it} специфицируется следующим образом: $u_{it} = \exp\{-\eta(t - T)\}u_i$, где $u_i \sim N^+(\mu; \sigma_u^2)$, T – максимальный период времени, η – неизвестный параметр темпа роста эффективности. Отметим, что модель имеет как минимум один дополнительный параметр η (если предположить $\mu = 0$) по сравнению с моделью Pitt & Lee. Это усложнение модели может быть излишним, если выборка небольшая, так как эффективность оценивания снизится заметно (выше уже отмечалось, что

процедура максимизации может вовсе не сойтись, что на практике часто случается с моделью Battese & Coelli).

Иные модели SFA на панельных данных.

Вышеописанная модель Battese & Coelli фиксирует зависимость эффективности от времени для всех компаний одинаковым образом. Однако предпосылка, что все компании отрасли с одинаковым темпом наращивают эффективность, нереалистична. Модель, не использующая столь сильное ограничение, предложена в работе Cornwell, Schmidt and Sickles (1990). В модели Cornwell et al. временная зависимость описывается весьма гибкой функцией, так что темп роста эффективности может быть свой для каждой фирмы.

Хороший обзор статей по моделям SFA на панельных данных с кратким описанием методологии каждой из них представлен в работе Battese and Coelli (1993). Среди представленных там моделей следует отметить следующую.

1.2.3.3. «Истинные фиксированные эффекты»

Во всех вышеперечисленных моделях границы затрат на панельных данных ненаблюдаемая неоднородность компаний (индивидуальные эффекты – разной природы для разных моделей) трактуются исключительно как неэффективность. Этот общий недостаток того же рода, что и недостаток модели COLS на данных одного года (cross-section), где МНК-остатки целиком рассматривались как неэффективность. Очевидно, что компании отрасли функционируют в разных условиях, то есть однородность выборки – вопрос весьма спорный, поэтому ненаблюдаемая неоднородность (индивидуальный эффект) в действительности есть сумма неэффективности и неоднородности. Одной из работ, где была ослаблена предпосылка о том, что индивидуальный эффект – это только неэффективность, является статья Greene (2005), которая называется “Reconsidering heterogeneity in panel data

estimators of the stochastic frontier model”¹³. Грин предложил использовать модель, которая получила название «Истинные фиксированные эффекты» (True Fixed Effect model, TFE).

Модель записывается уравнением $\ln[C_{it}] = \alpha_i + X'_{it}\beta + v_{it} + u_{it}$; отличие от стандартной модели SFA в том, что у свободного члена появился индекс i . Такая регрессия может быть оценена с помощью ММП или с помощью двухшаговой процедуры: на первом шаге оценивается стандартная модель с фиксированными эффектами, затем ее остатки становятся зависимой переменной в регрессии SFA на константу.

$$(1.3) \ln[C_{it}] = \alpha_i + X'_{it}\beta + \varepsilon_{it}$$

$$(1.4) \hat{\varepsilon}_{it} = const + v_{it} + u_{it}$$

Следует отметить, что модель TFE довольно сложна и нигде не используется регулирующими органами. Другой ее важный недостаток в том, что как правило результаты оценивания TFE слабо коррелируют с результатами более простых моделей. Для преодоления этого недостатка были разработаны еще более сложные модели (например, True Random Effects), однако следует иметь в виду что для целей регулирования такие модели не подходят, так как регулятору необходимо обосновывать выбор моделей, что сделать проще при использовании простых, очевидных и интуитивно понятных методов.

1.3.Выявление наблюдений-выбросов

При моделировании границы эффективности предельно важно доверие к результатам оценивания, в особенности если они используются в государственном регулировании отрасли. Устойчивость оценок зависит главным образом от нескольких факторов: во-первых, от того, насколько хорошо подобраны объясняющие переменные, во-вторых, от точности исходных данных, в-третьих, от размера выборки, и, наконец, от

¹³ См. также более раннюю работу Greene (2002).

однородности выборки. Наличие серьезных выбросов в выборке приводит к неустойчивости результатов оценивания, делает затруднительным проверку статистических гипотез, так как многие тесты чувствительны к выбросам, а модели, оцениваемые с помощью ММП (в частности SFA) из-за неоднородности выборки зачастую не могут быть оценены (не сходится процедура максимизации). Таким образом, для повышения точности анализа необходимо проводить анализ выборки на наличие выбросов.

Первый и самый очевидный шаг – визуальный анализ переменных на наличие нетипичных наблюдений с помощью диаграмм рассеивания или гистограмм. При выявлении сильных выбросов исследователь может принять решение об исключении их из выборки. Такой способ имеет существенный недостаток – исключаемые наблюдения (с экстремальными значениями) приводят к произвольному смещению выборки и, соответственно, смещению оценок. Второй недостаток – сужение выборки: исключая наблюдения, мы теряем как эффективность оценок, так и возможность оценить эффективность для нетипичных компаний.

Визуальный анализ достаточно субъективен, поэтому эконометристы разработали ряд более-менее объективных мер «нетипичности» наблюдения. Одна из них – расстояние Кука (Cook's distance), предложенная в работе Cook (1977). Эта мера хороша тем, что «нетипичность» наблюдения измеряется в контексте оцениваемой эконометрической модели, то есть ею улавливаются отклонения отдельных точек от общей тенденции, а не от некоторого среднего уровня, как в случае с визуальным анализом.

Расчет расстояния Кука для i -го наблюдения основан на сравнении оценок регрессий по полной выборке и по выборке с исключенным i -м наблюдением. Общая идея следующая: если наблюдение не выбивается из общей тенденции, то оно не вносит существенных изменений в регрессию и оценки коэффициентов будут отличаться слабо; в противном же случае коэффициенты и предсказанные значения для зависимой переменной могут

сильно поменяться. Это изменение в предсказанных значениях и выявляется с помощью расстояния Кука.

Расстояние Кука для i -го наблюдения рассчитывается по следующей формуле $D_i = \frac{\sum_{j \neq i} (\hat{Y}_j - \hat{Y}_j^{-i})^2}{k/n \cdot \sum_j (\hat{Y}_j - Y_j)^2}$, где \hat{Y}_j – предсказанное значение зависимой переменной для j -го наблюдения в регрессии по полной выборке, \hat{Y}_j^{-i} – аналогичная величина в регрессии на выборке с исключенным i -м наблюдением, k – число оцененных параметров, n – число наблюдений. Из формулы видно, что чем меньше меняется уравнение регрессии (и предсказанные значения) при исключении наблюдения, тем меньше расстояние Кука, тем меньше оснований считать наблюдение выбросом.

Пороговое значение расстояния Кука для i -го наблюдения, превышение которого означало бы отнесение i -го наблюдения к числу выбросов, не было предложено в работе Cook (1977). В последующих работах предлагается использование простых правил: в статье Cook, Weisberg (1982) предлагается использование порогового значения 1, в статье Bollen & Jackman (1990) выводится величина $4/n$, где n – число наблюдений. В практической части данного исследования будет использовано второе значение.

Практические рекомендации исследователю при выявлении нетипичных наблюдений в выборке зависят от цели исследования. Если для целей исследования важно получение точных оценок коэффициентов (эластичностей), уравнения регрессии, то исследование ограничивается оцениванием регрессии на короткой выборке (очищенной от выбросов). Если же важно получение некоторого показателя для каждого наблюдения (например, эффективности затрат), то необходимо дополнить полученные по короткой выборке значения значениями по каждому выбросу. Для этого оцениваются отдельно регрессии по выборке, состоящей из типичных наблюдений и одного выброса, по результатам оценивания которых и рассчитываются эффективности для компаний-выбросов.

При определении выбросов с помощью расстояния Кука на панельных данных возникают содержательные трудности. Если абстрагироваться от панельной структуры данных и рассчитать расстояние Кука для всех наблюдений, то может оказаться, что некоторые компании будут отнесены к выбросам в отдельные годы, а в другие годы – к типичным наблюдениям. В этом случае есть как минимум два подхода к тому, что считать «короткой» выборкой (очищенной от выбросов): либо исключить все выявленные выбросы, невзирая на панельную структуру данных, тогда оставшаяся панель будет несбалансированная; либо исключать компанию во все годы, при условии что она была выбросом хотя бы в один год. Последнее условие можно делать менее строгим, например исключать компанию из выборки, если она идентифицировалась как нетипичная в двух периодах или более. В практической части исследования будет использован строгий вариант исключения наблюдений из выборки («хотя бы в один год»).

2. Опыт других стран.

В Части 2 исследования будут подробно рассмотрены две работы по исследованию эффективности энергораспределяющих предприятий в других странах, а также упомянуты особенности моделей в других работах, в том числе по другим отраслям.

2.1. Источники информации

Прежде всего следует отметить обилие литературы по бенчмаркингу операционных или общих затрат в сфере распределения электроэнергии. Это связано прежде всего с тем, что регулирующие органы многих стран заинтересованы в применении новейших теоретических моделей в регулировании своего рынка, поэтому практически все описанные в Части 1 методы применялись к распределительным сетям. Существует ряд рабочих статей, где собрана информация о наиболее популярных теоретических методах, а также их применении в различных странах. Среди таких обзоров следует отметить, во-первых, рабочую статью Австралийских государственных органов Australian Competition and Consumer Commission и Australian Energy Regulator (ACCC/AER) под названием *Benchmarking Opex and Capex in Energy Networks*, во-вторых, отчет по исследованию, выполненному WIK-Consult (Schweinsberg, Stronzik, Wissner, 2011) по заказу тех же австралийских регуляторов под названием *Cost Benchmarking in Energy Regulation in European Countries*, а также обзор-презентацию *Benchmarking and the regulation of electricity distribution companies* под авторством Massimo Filippini (2011).

Следует отметить ряд авторов, принимавших участие в исследованиях по рассматриваемой тематике в разных странах: Тоорaj Jamasb, Michael Pollitt (UK, University of Cambridge), Mehdi Farsi, Massimo Filippini (Switzerland, Swiss Federal Institute of Technology) и William Greene (US, Stern School of Business, New York University). В многочисленных работах

вышеперечисленных авторов оценивается эффективность затрат в исследуемой отрасли (а также в смежных отраслях) методами COLS, MOLS (на данных одного года и на панельных данных), SFA (модель на данных одного года, модель time-invariant и time-varying на панельных данных), а также модели TFE (предложенные Greene) и ряд других модификаций панельного SFA, – то есть весь спектр эконометрических моделей границы затрат. Подробный обзор применявшихся методов представлен в работах Jamasb and Pollitt (2001) и Jamasb and Pollitt (2003).

2.2. Методы государственного регулирования отрасли распределения электроэнергии

Вся масса разработанных методов оценивания границы затрат в конце концов имеет своим приложением регулирование конкретных рынков конкретными регулирующими органами¹⁴. Поэтому здесь большую роль играют предпочтения лиц, принимающих решения. Рассмотрим методы, применяемые разными странами. Многие страны¹⁵ применяют метод «верхней границы цены» (“price cap” или “revenue cap”)¹⁶, в частности австрийский регулятор (E-Control). Метод заключается в ограничении сверху тарифа на электроэнергию для конечного потребителя исходя из некоторой формулы. Этот метод устанавливает ограничение на тариф исходя из так называемого X-фактора. X-фактор включает в себя общее для отрасли требование к увеличению производительности/эффективности затрат (напр. 2% в год) и может дополнительно включать индивидуальный X-фактор (обозначаемый также Xi-фактор). Общий X-фактор позволяет перераспределить выгоды от роста производительности между производителями и потребителями, а индивидуальные вводятся для выравнивания компаний между собой (в терминах эффективности).

¹⁴ Сейчас выносим за скобки те исследования, целью которых являлась инвестиционная деятельность.

¹⁵ Великобритания, Швеция, Норвегия, Австрия, Германия, Мексика и другие.

¹⁶ Часто обозначается в литературе как CPI-X

Австрийский регулятор рассчитывает индивидуальные требования к росту эффективности (индивидуальные X-факторы) по формуле $X_i = (1 - X) \sqrt[8]{ES_i}$, где X_i – индивидуальный X-фактор, X – общий отраслевой X-фактор, ES_i – рассчитанное по модели значение эффективности для i -й компании. Цифра 8 означает что регулятор дает компаниям 8 лет на подстройку к целевым уровням затрат.

Так как каждая эконометрическая модель имеет свои недостатки, опасно полагаться на результаты оценивания одной регрессии. Поэтому регуляторы используют ряд приемов, помогающих снизить риски. Например, в Германии и Норвегии в качестве индивидуального X-фактора используется величина, полученная взвешиванием оценок эффективности по разным моделям; веса при этом задаются экспертно. Норвежская практика имеет еще одну особенность: целевой уровень издержек устанавливается регулятором по формуле $0.6 \times \text{Модельное значение (у регулятора)} + 0.4 \times \text{Плановое значение (у фирмы)}$. Сумма планируемых затрат фирмы (затрат следующего года) при предоставлении регулятору должна быть подкреплена экономическим обоснованием. Такой подход позволяет заметно снизить риск перерегулирования, так как при ошибочно низком модельном значении эффективности, X-фактор вытягивается вверх за счет высокого планового значения.

Другой пример – регулирование отрасли распределения воды в Италии (согласно статье Antoniolì and Filippini, 2003). В статье описана пошаговая процедура регулирования. На первом шаге компании предлагают регулятору тариф. На втором шаге регулятор использует эконометрическую модель для определения верхней границы цены (“price cap”). Далее, предложенный тариф одобряется, если он не превышает верхней границы цены, и пересматривается, если превышает. Компания может получить одобрение «завышенного» тарифа, если она представит достаточные доказательства того, что она функционирует в особых условиях.

Следует отметить, что большая часть регулирующих органов вынуждена «смягчать» модельные результаты расчета эффективностей из-за риска выставить компании заведомо невыполнимые требования и получить в ответ со стороны компании судебный иск. Помимо норвежского способа взвешивания (см. выше) применяется также простое ограничение снизу: так, австрийский регулятор после расчета ES , всем компаниям с $ES < 0.75$ присваивает значение $ES = 0.75$. В Великобритании регулятор (OFGEM) нормирует рассчитанные по модели значения ES по 30% лучших практик, то есть компаниям со значением ES , равным 7-му децилю и выше (30% лучших), присваивается значение $ES = 1$; остальные компании «подтягиваются» линейным сдвигом вверх.

Количество применяемых на практике методов регулирования велико, потенциально оно ограничен лишь фантазией исследователей. Необходимо понимать, что разнообразие методов говорит о недостаточности каждого из них в отдельности. Практически все исследователи в качестве одного из выводов оставляют предостережение: любая модель дает результат, имеющий скорее рекомендательный характер; регулятору не стоит прямолинейно накладывать на компании ограничения, исходя из результатов оценивания единственной регрессии, так как выбор переменных, функциональной формы, метода оценивания и даже метода «смягчения» результата довольно произволен. Таким образом, оптимально использовать результаты нескольких моделей с последующим их взвешиванием¹⁷, а также

2.3. Применение методов оценивания границы затрат в отрасли распределения электроэнергии в других странах

2.3.1. Швейцария 2003

В работе Farsi and Filippini (2003) применялись описанные в Части 1 теоретические методы на примере 59 энергораспределяющих компаний в

¹⁷ Шведский регулятор использует не взвешенное значение эффективности, а максимальное из нескольких моделей.

Швейцарии (поставляют треть объема э/э) за период 1988-1996 гг. (несбалансированная панель, всего 380 наблюдений). Краткий результат исследования следующий: показатели эффективности (ES) в среднем не чувствительны к выбору метода оценивания, однако ранжировки компаний по эффективности различаются значительно, поэтому авторы рекомендуют регулирующему органу применять комбинацию из методов для получения более устойчивого результата. Другая предлагаемая авторами поправка на неточность – практика регулирования, описанная в статье Antonioli and Filippini (2003) для отрасли распределения воды (описано выше в пункте 2.2).

Методы оценивания, применяемые в работе Farsi and Filippini (2003), следующие: COLS, GLS (RE), Within регрессия (FE), SFA (normal – half-normal model). Анализируемый феномен – эффективность совокупных затрат (не только операционных), то есть зависимая переменная всех моделей – *общие затраты* (C). Независимые переменные – *полезный отпуск* (Y, кВт·ч), *цены ресурсов* – труда, капитала и входящей э/э (соответственно P_L, P_K, P_P), *пиковая загрузка сети* (LF, %), *территория обслуживания* (AS, км²), *количество клиентов* (CU, чел). Также в модель включаются дамми-переменные «наличие высоковольтных сетей» (HGRID), «лесистость свыше 40% территории обслуживания» (DW) и «доля вспомогательных доходов более 25% всех доходов» (DOT) и временной тренд (T).

Из содержательных соображений вводятся следующие ограничения на функциональную форму регрессии: функция затрат – однородная первой степени по ценам (однородность 1-й степени – это увеличение затрат в n раз при росте всех цен в n раз, $\beta_K + \beta_L + \beta_P = 1$), неубывающая по ценам и выпуклая. Затраты в работе описываются функцией Кобба-Дугласа (а не функцией ТрансЛог) по двум причинам: во-первых, количество независимых переменных велико, во-вторых, «размерные» переменные в модели (Y, AS, CU) и так сильно коррелированы. Таким образом, уравнение регрессии следующее:

$$\ln\left(\frac{C}{P_P}\right) = \alpha + \beta_K \ln\left(\frac{P_K}{P_P}\right) + \beta_L \ln\left(\frac{P_L}{P_P}\right) + \gamma_Y \ln Y + \gamma_{CU} \ln CU + \gamma_{AS} \ln AS + \gamma_{LF} \ln L \\ + \delta_1 HGRID + \delta_2 DOT + \delta_3 DW + \tau_T T + \varepsilon$$

Результаты оценивания приведены в Табл.2.1 ниже:

	OLS		Random-Effects (GLS)		Random-Effects MLE (Half-Normal)		Fixed-Effects	
	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
lnY	0.851	0.017	0.780	0.032	0.767	0.040	0.677	0.055
lnCU	0.084	0.017	0.153	0.033	0.163	0.048	0.251	0.096
lnAS	0.044	0.004	0.051	0.009	0.047	0.013	-	-
lnLF	-0.243	0.037	-0.239	0.039	-0.230	0.023	-0.213	0.044
lnPL	0.067	0.011	0.041	0.014	0.039	0.014	0.038	0.016
lnPK	0.200	0.009	0.174	0.010	0.171	0.005	0.169	0.010
HGRID	0.063	0.012	0.075	0.027	0.098	0.039	-	-
DOT	0.033	0.010	0.050	0.022	0.040	0.028	-	-
DW	0.014	0.010	0.012	0.023	0.062	0.028	-	-
T	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.002
Constant	-2.236	0.233	-0.793	0.369	-0.653	0.430	-	-
R Square	0.995		0.995				0.991	

Табл.2.1 Результаты оценивания по статье Farsi & Filippini (2003)

Отметим большое сходство оценок разными методами (за исключением FE, где отсутствуют оценки для регрессоров, неизменных во времени). Результаты стабильны исходя и из описательных статистик неэффективностей (оценок \hat{u}_i):

	DOLS	RE (GLS)	FE	ML
Minimum	1.07	1	1	1.07
Maximum	1.46	1.38	2.14	1.36
Average	1.23	1.16	1.35	1.15
Median	1.22	1.16	1.31	1.13
95 percentile	1.41	1.32	1.66	1.30
Number of firms	59	59	59	59

Табл.2.2 Описательные статистики неэффективностей по статье Farsi & Filippini (2003)

Сильные расхождения отмечаются при сопоставлении ранжировок, полученных разными методами.

	DOLS	RE (GLS)	FE	ML
DOLS	1			
RE (GLS)	0.936	1		
FE	0.447	0.514	1	
ML	0.838	0.895	0.417	1

Табл.2.3 Корреляции неэффективностей по статье Farsi & Filippini (2003)

Выводы по статье: модель FE дает заметно отличные от других (COLS, RE, SFA) моделей результаты. Скорее всего это связано с тем, что в ее оценивании не участвуют неизменные во времени независимые переменные, что было отмечено выше. Также, исходя из теоретических свойств метода, для состоятельного оценивания индивидуальных эффектов необходимо выполнение условия $T \rightarrow \infty$, а в выборке всего 9 временных периодов. Таким образом, на коротких временных рядах не следует оценивать модель FE.

2.3.2. Норвегия 2013.

Одно из новейших исследований по оцениванию границы затрат на данных по энергораспределяющим компаниям изложено в рабочей статье Agrell, Farsi, Filippini and Koller (2013). Используемые данные – выборка из 111 норвежских энергораспределяющих компаний за 1998-2002 годы (сбалансированная панель за 5 лет, всего 555 наблюдений). Модель, оцениваемая в статье, описывается следующим уравнением затрат: $\ln\left(\frac{TC}{P_C}\right) = \alpha + \beta_L \ln\left(\frac{P_L}{P_P}\right) + \beta_Q \ln Q + \beta_D \ln D + \beta_{LS} \ln LS + \varepsilon$, где TC – общие издержки, P_L, P_C – цены соответственно труда и капитала, Q – полезный отпуск, D – плотность сети, LS – доля линий высокого напряжения (в этой же статье встречаются обозначения $P = \frac{P_L}{P_P}$, $S = LS$). Используемая функциональная форма – функция Кобба-Дугласа с ограничением на цены (однородность первой степени по ценам факторов). Обоснование выбора такой формы изложено выше в 2.3.1 (по статье Farsi and Filippini (2003)). Заметим, что набор объясняющих переменных значительно уже, чем в упомянутой выше работе, однако авторы предпочитают не использовать функцию ТрансЛог.

Авторы оценивают модель стандартными методами MOLS и SFA для панельных данных, разбивая выборку на кластеры. Также используется непараметрический метод DEA. На первом этапе проводится кластерный или любой иной качественный анализ выборки результатом которого является ее

разбиение на более-менее однородные классы (кластеры); авторы статьи используют метод латентных классов (latent classes approach, LC)¹⁸. На втором этапе оценивается регрессия по каждому классу отдельно (j – номер класса):

$$\ln\left(\frac{TC}{PC}\right) = \alpha^j + \beta_L^j \ln\left(\frac{PL}{PP}\right) + \beta_Q^j \ln Q + \beta_D^j \ln D + \beta_{LS}^j \ln LS + \varepsilon^j$$

Оптимальное число классов – 4, так как в регрессиях с более мелкой кластеризацией существенные переменные оказывались незначимыми. Ниже приведены результаты оценивания и информация о ES (эффективностях) по двум моделям.

	Second step	Class 1	Class 2	Class 3	Class 4
	Variable	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)	Coefficient (SE)
MOLS	Input price ratio (P)	0.8588 *** (0.048)	0.4623 *** (0.018)	0.8644 *** (0.023)	0.6891 *** (0.018)
	Distributed electricity (Q)	0.9080 *** (0.020)	0.7719 *** (0.010)	1.0581 *** (0.010)	0.9856 *** (0.008)
	Density (D)	-0.4236 *** (0.056)	-0.3252 *** (0.017)	-0.9255 *** (0.055)	-0.0464 ** (0.019)
	Share HV network (S)	0.6076 *** (0.202)	0.0802 (0.050)	2.7761 *** (0.164)	0.7582 *** (0.072)
	Constant	5.0162 *** (0.019)	4.6692 *** (0.007)	4.6642 *** (0.010)	4.7317 *** (0.007)
SFA	Input price ratio (P)	0.8961 *** (0.050)	0.4624 *** (0.017)	0.8669 *** (0.022)	0.6880 *** (0.018)
	Distributed electricity (Q)	0.9027 *** (0.020)	0.7719 *** (0.010)	1.0587 *** (0.010)	0.9867 *** (0.008)
	Density (D)	-0.4140 *** (0.055)	-0.3253 *** (0.017)	-0.9329 *** (0.053)	-0.0469 ** (0.019)
	Share HV network (S)	0.8017 *** (0.188)	0.0802 (0.049)	2.7953 *** (0.161)	0.7538 *** (0.071)
	Constant	4.8158 *** (0.020)	4.6472 *** (0.015)	4.6079 *** (0.016)	4.6947 *** (0.011)
	Sigma: $\sigma^2 = \sigma_u^2 + \sigma_v^2$	0.2786 *** (0.002)	0.0764 *** (0.000)	0.1082 *** (0.001)	0.0870 *** (0.000)
Lambda: $\lambda = \sigma_u / \sigma_v$	2.1775 *** (0.395)	0.3866 (0.265)	0.8413 *** (0.275)	0.6256 *** (0.214)	

Табл.2.4 Результаты оценивания в статье Agrell et al. (2013)

Два метода дают весьма схожие результаты, что свидетельствует о высокой однородности выборки (достигнута путем разбиения выборки на классы). Знаки и значимость коэффициентов согласуются с экономическим смыслом.

Значим следующий результат: разбиение на классы позволило авторам получить значения ES более реалистичные, нежели в случае оценивания модели по полной выборке. Можно таким образом сформулировать общее правило: чем большую долю дисперсии затрат занимает необъясненная часть, тем выше оценки неэффективности, так как они – производные остатков регрессии). В Табл.2.5 приводятся описательные статистики обеих

¹⁸ Метод описан в статье Lazarsfeld and Henry (1968)

подходов, где очевидно, что кластеризация позволяет снизить ненаблюдаемую неоднородность выборки и, следовательно, повысить ES.

	Class 1			Class 2			Class 3			Class 4		
	DEA	MOLS	SFA									
Mean	0.568	0.776	0.829	0.695	0.906	0.978	0.807	0.882	0.946	0.771	0.899	0.964
SDev	0.207	0.131	0.093	0.168	0.062	0.006	0.149	0.076	0.018	0.125	0.065	0.010
Min	0.267	0.362	0.531	0.374	0.732	0.916	0.481	0.695	0.883	0.559	0.740	0.933
Max	1	1	0.962	1	1	0.986	1	1	0.976	1	1	0.983

	DEA	MOLS	SFA
Mean	0.554	0.762	0.802
SDev	0.151	0.148	0.104
Min	0.246	0.319	0.498
Max	1	1	0.960

Табл.2.5 Сравнение оценок эффективности (ES), полученных по кластерам (сверху) и по полной выборке (снизу) по статье Agrell et al. (2013)

Вывод по статье: при наличии значительного числа наблюдений (555 – достаточное количество) имеет смысл проводить анализ эффективности в два этапа. На первом этапе проводится кластеризация, на втором – расчет эффективностей на отдельных классах. Это позволяет снизить долю необъясненной дисперсии и повысить оценки эффективности. Однако не стоит проводить кластеризацию при малом числе компаний выборке (и коротких временных рядах), так как регрессионные модели дадут слабую значимость существенных переменных (что показала разбивка на 5 и более классов). В случае с малой выборкой долю необъясненной дисперсии имеет смысл снизить за счет включения большего числа факторов, объясняющих неоднородность компаний (для России – это, например, климатические переменные). Неоднородность выборки можно также снизить, применяя *анализ выбросов* с помощью расстояния Кука (Cook's distance), что рассматривалось выше в п.1.3.

3. Описание переменных и источники данных.

3.1. Источники данных

В данной работе все эконометрические модели будут оценены на данных по компаниям «Холдинга МРСК». Единицей наблюдения в исследовании является региональный филиал МРСК. Все данные собраны из открытых источников. Во-первых, данные по операционным затратам и натуральным показателям компаний – из годовых отчетностей региональных филиалов «Холдинга МРСК», публикуемых в сети¹⁹. Во-вторых, данные по климатическим условиям для различных регионов – из строительных норм и правил (СНиПов²⁰). В-третьих, данные по уровню цен и средней заработной плате взяты из базы РосСтата²¹. Период наблюдения – 2009-2011 годы. Итоговая база данных содержит информацию по 57 филиалам из 63, так как данные по шести филиалам либо частично отсутствуют, либо явно недостоверные. Данные представляют собой сбалансированную (без пропусков) панель за три года, что позволяет оценить по ним все описанные в Части 1 модели.

3.2. Зависимая переменная

Исследуемый феномен – операционная эффективность предприятий – определяет переменную, стоящую в левой части регрессионного уравнения – операционные расходы. *Операционные расходы* (англ. operating expenditures, OPEX; синонимы: текущие расходы, эксплуатационные расходы) – это затраты и платежи, связанные с проведением за определенный период времени финансовых, производственных, хозяйственных операций. Операционные расходы включают затраты на производство и реализацию

¹⁹ Ссылки на сайты филиалов на странице <http://www.holding-mrsk.ru/sites/>

²⁰ СНиП 2.01.07-85* «Нагрузки и воздействия» и СНиП 23-01-99 «Строительная климатология»

²¹ www.gks.ru/cbsd

продукции, административные и финансовые расходы²². Операционные расходы разграничиваются от капитальных затрат протяженностью периода, в который израсходованные средства будут участвовать в хозяйственных процессах. Так, покупка оборудования, например принтера (который будет служить более года / более одного производственного цикла) будет относиться к капитальным затратам, а покупка тонера и бумаги – к операционным²³.

Операционные расходы – статья отчета о Прибылях и Убытках. Этот показатель аудирруется, поэтому достоверность данных по нему высокая. Как и другие показатели деятельности компаний, данные по операционным расходам взяты с сайтов филиалов «Холдинга МРСК».

3.3. Независимые переменные

Чтобы определить состав объясняющих переменных, необходимо рассмотреть все технологические процессы, связанные с операционной деятельностью распределительных сетей. Основные процессы следующие: эксплуатация (обслуживание и текущий ремонт) линий, эксплуатация подстанций, эксплуатация автоматических систем управления и релейной защиты и административно-управленческие процессы. Исходя из этой разбивки, можно выделить параметры компаний, влияющие на величину затрат по каждому процессу²⁴.

²² Определение согласно словарю ФИНАМ, <http://www.finam.ru/dictionary/>

²³ Пример по книге Aswath Damodaran *'Applied Corporate Finance' (2011)*

²⁴ Подробный разбор драйверов затрат см., например, в рабочей статье "A comparative analysis: Aurora Energy's Network cost structure", *Benchmark Economics (May 2011)*, pp. 43-57.

Процессы	Основные типы затрат	Показатели, влияющие на величину затрат	
		Показатели компании	Факторы среды
Эксплуатация линий Обходы и текущее обслуживание Тек. и кап. ремонт	Персонал Материалы: провод, кабель, опоры ЛЭП и пр.	Протяженность линий (высокого напряжения, среднего, низкого) Плотность сети	Температурная зона Гололёдная нагрузка Ветровая нагрузка
Эксплуатация подстанций Тек. обслуживание Тек. и кап. ремонт	Персонал Материалы: трансформаторы, выключатели и т.п. Технологическое потребление э/э	Трансформаторная емкость (выс., ср., низ. напряжения) Возраст оборудования	Температурная зона Переходы через ноль
Эксплуатация релейной защиты и автоматических систем управления	Персонал Материалы: Автоматизированные системы контроля и учёта энергоресурсов и пр.	Трансформаторная емкость Протяженность сети Квалификация персонала	Температурная зона
Административно-управленческие процессы	Персонал Информационные системы	Трансформаторная емкость Протяженность сети	

Табл. 3.1 Драйверы затрат

Для того чтобы модель давала точный результат, по возможности нужно включать в нее как можно больше объясняющих переменных из приведенной выше таблицы, так как теоретически все они оказывают влияние на величину операционных затрат. С точки зрения теоретической эконометрики это также правомерно: невключение в уравнение регрессии значимых переменных приводит к смещению и несостоятельности МНК-оценок²⁵. Однако на практике включить в уравнение все переменные может оказаться невозможным по ряду причин.

Во-первых, по некоторым показателям могут отсутствовать данные. В нашем случае такими показателями являются цены материалов (опоры ЛЭП, провода, трансформаторы, бетон и др. – в базе РосСтата отсутствует разбивка по регионам), квалификация персонала (ненаблюдаемая переменная), разбивка по видам напряжения протяженности сетей и трансформаторной емкости (информация не публикуется в отчетах). Проблема отсутствия данных решается подбором прокси-переменной: для

²⁵ См. например Доугерти К., “Введение в эконометрику”, стр. 202-206.

всех ценовых переменных прокси-переменная – это *стоимость потребительской корзины в регионе*. Для квалификации персонала прокси-переменная – аварийность, так как вторая по частоте причина аварий (после падения деревьев) – человеческий фактор, однако включение аварийности в модель приведет к появлению новой проблемы – эндогенности, поэтому лучшим выходом из ситуации является невключение ее в модель. Разбивку по видам напряжения²⁶ также придется проигнорировать, так как аппроксимировать ее нечем.

Во-вторых, в выборке может не оказаться достаточного числа наблюдений для того, чтобы в модель включать 10-15 переменных. Если рассматривать модели на данных одного года (cross-section) по имеющимся 57 наблюдениям, эта проблема как раз возникает. В этом случае можно отказаться от одновременного включения в модель сильно взаимосвязанных переменных (например, *количества подстанций и трансформаторной емкости*), а включить только одну из них. Другим выходом из этой ситуации будет включение в модель переменной *полезный отпуск* вместо всех «размерных» переменных (характеризующих помимо всего прочего размер компании): *количество точек подключения* и уже упомянутых выше переменных *количество подстанций и трансформаторная емкость*. Этот прием применялся в ряде зарубежных исследований, например в Норвегии (см. п. 2.3.2).

В-третьих, влияние некоторых величин может рассматриваться как априори незначимое, например не имеет смысла включать в модель цену электроэнергии, потребляемой для технологических нужд, так как объем этого потребления несопоставимо мал по сравнению с потерями электроэнергии (для сетевой компании характерен уровень потерь 5-15% на

²⁶ Обслуживание высоковольтных линий и подстанций стоит дороже по причине сложности оборудования и повышенных требований к надежности.

высоковольтных линиях и до 30% на линиях <1кВ, тогда как технологическое потребление составляет не более 1%).

Итоговый список доступных объясняющих переменных следующий:

<i>Переменная</i>	<i>Ед.изм.</i>
Число точек подключения	шт
Длина сети	км
Количество подстанций	шт
Трансформаторная емкость	МВА
Полезный отпуск	млн квт·ч
Плотность сети ²⁷	шт/км
Заработная плата ²⁸	руб
Уровень цен ²⁹	руб
Средняя температура января ³⁰	°С
Толщина стенки гололеда ³¹	мм
Число дней в году с переходом через ноль ³²	дни

Табл. 3.2 Список переменных

Более подробно дизайн модели будет рассмотрен в Части 4 работы.

²⁷ Плотность сети = Число точек подключения / Длина сети

²⁸ База данных РосСтата. Показатель «Средняя заработная плата работников» в отрасли «Производство, передача и распределение электроэнергии» в разрезе регионов России.

²⁹ База данных РосСтата. Показатель «Средняя стоимость фиксированного набора товаров и услуг» в разрезе регионов России.

³⁰ СНиП 23-01-99 «Строительная климатология»

³¹ СНиП 2.01.07-85* «Нагрузки и воздействия»

³² СНиП 23-01-99 «Строительная климатология»

4. Оценивание моделей

В Части 4 работы будут оценены различные модели границы затрат всеми изложенным выше параметрическими методами, а также проведен сравнительный анализ применявшихся методов. Методом DEA эффективности оцениваться не будут.

4.1. Обозначения и сокращения

Переменные, входящие в уравнение регрессии, обозначены следующим образом:

<i>Переменная</i>	<i>Обозначение</i>	<i>В логарифмах</i>
Операционные расходы	OPEX	opex
Число точек подключения	CONNECT	connect
Длина сети	LENGTH	length
Количество подстанций	SUBST	subst
Трансформаторная емкость	TRANSF	transf
Полезный отпуск	OUTPUT	output
Плотность сети	DENSITY	density
Заработная плата	WAGE	wage
Уровень цен	PRICE	price
Средняя температура января	JANUARY	january ³³
Толщина стенки гололеда	ICE	ice
Число дней в году с переходом через ноль	ZERO	zero

Табл. 4.1 Обозначения переменных регрессии

Выше было отмечено, что все исследователи оценивают границу затрат с ограничением на цены факторов $wage + price = 1$, поэтому имеет смысл ввести новые переменные $wage_p = \log \left[\frac{WAGE}{PRICE} \right]$, $opex_p = \log \left[\frac{OPEX}{PRICE} \right]$. Дополнительно в некоторых моделях используется относительная характеристика «средняя величина подстанции»: $trans_s = \log \left[\frac{TRANSF}{SUBST} \right]$.

³³ $january = \log (-JANUARY - \min[JANUARY]) + 100$. Такое преобразование делает возможным интерпретацию коэффициента как процентное изменение в затратах при сдвиге на 1 градус. Ожидаемый знак коэффициента после преобразования (+).

4.2. Уравнения регрессионных моделей.

В данном исследовании использовался следующий подход к формированию состава переменных для моделей. На первом шаге определялся состав переменных, которые обязательно должны присутствовать в регрессии. Эти переменные следующие: зависимая – $opex_p$, независимые – $wage_p, connect, density$. Другие переменные – $transf, output, transf_s$ – добавляются к ним по очереди, так что получается 3 модели. Климатические переменные добавляются одновременно, однако в итоговых моделях остаются только значимые на 10%. Значимость в этом случае проверялась при предварительном оценивании методами МНК (Pooled регрессия) и SFA (на панельных данных, Pitt&Lee).

Таким способом были сформированы три конкурирующие модели:

$$(1) \quad opex_p = \alpha + \beta_w wage_p + \beta_t transf + \beta_c connect + \beta_d density \\ + \beta_j january + \beta_z zero + \varepsilon$$

$$(2) \quad opex_p = \alpha + \beta_w wage_p + \beta_o output + \beta_c connect + \\ + \beta_d density + \beta_z zero + \varepsilon$$

$$(3) \quad opex_p = \alpha + \beta_w wage_p + \beta_o output + \beta_{ts} transf_s + \beta_c connect \\ + \beta_d density + \beta_z zero + \varepsilon$$

4.3. Методы оценивания

Каждая из вышеперечисленных моделей оценивалась следующими способами: COLS и TMOLS на данных каждого года отдельно, COLS и MOLS на панельных данных согласно работе Schmidt and Sickles (1984) с использованием модели RE, SFA на панельных данных (модели с постоянной эффективностью Pitt&Lee и с изменяющейся во времени эффективностью Battese&Coelli), TFE (двухшаговое оценивание) – всего 7 методов. Каждая регрессия оценивалась дважды – с учетом анализа выбросов с помощью

расстояния Кука и без него³⁴. Результат оценивания границы затрат – уравнение самой границы и значения эффективностей (ES) для всех компаний. Заключительная часть работы – сравнительный анализ полученных рядов ES (всего 42 ряда).

4.4. Результаты оценивания

Рассчитанные значения эффективностей (ES) обозначаются следующим образом (# - номер модели):

Метод	Обозначение	
	Без учета выбросов	С учетом выбросов
COLS на данных одного года	CS COLS ES#	CS COLS ES Cook#
MOLS на данных одного года	CS MOLS ES#	CS MOLS ES Cook#
COLS на панельных данных	RE COLS ES#	RE COLS ES Cook#
MOLS на панельных данных	RE MOLS ES#	RE MOLS ES Cook#
SFA (постоянная эффективность)	SFA ES#	SFA ES Cook#
SFA (переменная эффективность)	TVD ES#	TVD ES Cook#
True Fixed Effects	TFE ES#	TFE ES Cook#

Табл. 4.2 Обозначения рядов эффективностей

Следует отметить, что методы COLS и MOLS на данных одного года, а также SFA с переменной эффективностью дают для каждого года свои значения эффективности; остальные методы предполагают постоянную эффективность.

Предварительный анализ результатов оценивания позволяет сделать несколько выводов. Во-первых, не достигается сходимость по модели SFA с переменной эффективностью при введении анализа выбросов. Модель SFA-TVD (Battese&Coelli) сильно чувствительна к изменениям выборки. При анализе выбросов основная выборка сокращается и поэтому снижается эффективность оценивания. Во-вторых, неудовлетворительный результат дает оценивание модели TFE, что видно из описательных статистик:

³⁴ При оценивании методом MOLS на панельных данных с учетом анализа выбросов сначала рассчитываются все остатки (в том числе для выбросов), и только затем они корректируются на общее стандартное отклонение (рассчитываемое по полной выборке). Расстояние Кука рассчитывается отдельно для моделей на данных одного года и на панельных данных (при методе оценивания МНК).

Variable	Obs	Mean	Std. Dev.	Min	Max
TFE_ES1	177	.9998077	3.52e-07	.9998065	.999809
TFE_ES_Cook1	177	.9663736	.0126044	.9112565	.9889017

Табл. 4.3 Описательные статистики эффективностей по методу TFE

Показатели эффективности для данного метода практически не варьируют; метод TFE исключается из дальнейшего анализа. Для остальных методов приведем корреляционные³⁵ матрицы (не включаются одновременно COLS и MOLS, так как они дают одинаковые ранжировки):

	CS_CO~S1	RE_CO~S1	SFA_ES1	TVD_ES1	CS_CO~k1	RE_CO~k1	SFA_E~k1
CS_COLS_ES1	1.0000						
RE_COLS_ES1	0.9262	1.0000					
SFA_ES1	0.8371	0.9290	1.0000				
TVD_ES1	0.8579	0.9504	0.9888	1.0000			
CS_COLS_E~k1	0.8840	0.7666	0.7035	0.7015	1.0000		
RE_COLS_E~k1	0.8453	0.8575	0.8213	0.8103	0.8709	1.0000	
SFA_ES_Cook1	0.9199	0.9753	0.9044	0.9214	0.7978	0.8895	1.0000

Табл. 4.4 Корреляции Спирмена (все методы)

Вывод: представленные в матрице методы дают согласованные результаты, что говорит об устойчивости каждого из них. Следует отметить, что анализ выбросов меняет ранжировки компаний (корреляции выделены полужирным). Это говорит о том, влияние выбросов довольно существенно и их учет при оценивании желателен.

В приложениях 1-5 приведены результаты оценивания модели 1 различными способами, а также некоторые тесты. В частности тест Хаусмана на эндогенность в модели с индивидуальными эффектами дает положительный результат, то есть предпосылка модели RE о некоррелированности эффектов с регрессорами не выполняется, поэтому из моделей на панельных данных предпочтительнее использовать SFA.

³⁵ Используется ранговая корреляция Спирмена

4.5. Сравнение трех моделей

4.5.1. МНК за 2011 год

Variable	Obs	Mean	Std. Dev.	Min	Max
CS_COLS_ES1	59	.6701723	.1352566	.4039328	1
CS_COLS_E~k1	59	.676803	.1186318	.4497029	1
CS_COLS_ES2	59	.6350336	.1310399	.3828953	1
CS_COLS_E~k2	59	.7237326	.1222289	.4633239	1
CS_COLS_ES3	59	.7163308	.1269013	.4441614	1
CS_COLS_E~k3	59	.7331925	.1123543	.4997522	1

Табл. 4.5 Описательные статистики эффективностей по трем моделям

	CS_CO~S1	CS_CO~k1	CS_CO~S2	CS_CO~k2	CS_CO~S3	CS_CO~k3
CS_COLS_ES1	1.0000					
CS_COLS_E~k1	0.8840	1.0000				
CS_COLS_ES2	0.6458	0.5529	1.0000			
CS_COLS_E~k2	0.6113	0.7107	0.8563	1.0000		
CS_COLS_ES3	0.8520	0.7523	0.8595	0.7548	1.0000	
CS_COLS_E~k3	0.7633	0.8526	0.7952	0.9188	0.8802	1.0000

Табл. 4.6 Корреляции Спирмена (все модели, метод МНК за 2011 год)

Выводы: Модели 1 и 3 согласуются между собой, Модель 2 отличается значительно (особенно от модели 1). Использование анализа выбросов позволяет в среднем увеличить значения ES.

4.5.2. Модель со случайными эффектами (RE)

Variable	Obs	Mean	Std. Dev.	Min	Max
RE_COLS_ES1	59	.7108971	.1327004	.4143685	1
RE_COLS_E~k1	59	.7870367	.1233613	.4930606	1
RE_COLS_ES2	59	.71869	.136195	.4089408	1
RE_COLS_E~k2	59	.8010969	.140574	.4204724	1
RE_COLS_ES3	59	.7567183	.1254531	.4488186	1
RE_COLS_E~k3	59	.8059615	.1247499	.4808943	1

Табл. 4.7 Описательные статистики эффективностей по трем моделям

	RE_CO~S1	RE_CO~k1	RE_CO~S2	RE_CO~k2	RE_CO~S3	RE_CO~k3
RE_COLS_ES1	1.0000					
RE_COLS_E~k1	0.8575	1.0000				
RE_COLS_ES2	0.7033	0.6439	1.0000			
RE_COLS_E~k2	0.6768	0.7854	0.8842	1.0000		
RE_COLS_ES3	0.8548	0.7449	0.8895	0.8021	1.0000	
RE_COLS_E~k3	0.7158	0.8811	0.8126	0.9138	0.8237	1.0000

Табл. 4.8 Корреляции Спирмена (все модели, метод RE)

Выводы: Все модели согласуются между собой. Использование анализа выбросов позволяет в среднем увеличить значения ES (вывод тот же, что и для МНК на данных одного года).

4.5.3. Модель Стохастической Границы (SFA, TI)

Variable	Obs	Mean	Std. Dev.	Min	Max
SFA_ES1	59	.7906803	.1501301	.3672765	.9729666
SFA_ES_Cook1	59	.8194366	.1587446	0	.9775792
SFA_ES2	59	.7962515	.1545709	.4029825	.9697367
SFA_ES_Cook2	59	.8207304	.133584	.453447	.973343
SFA_ES3	59	.8018477	.1347251	.4172881	.9742712
SFA_ES_Cook3	59	.8160298	.1860545	0	.9769503

Табл. 4.9 Описательные статистики эффективностей по трем моделям

	SFA_ES1	SFA_E~k1	SFA_ES2	SFA_E~k2	SFA_ES3	SFA_E~k3
SFA_ES1	1.0000					
SFA_ES_Cook1	0.9044	1.0000				
SFA_ES2	0.7487	0.7150	1.0000			
SFA_ES_Cook2	0.7368	0.7527	0.9623	1.0000		
SFA_ES3	0.8625	0.8146	0.9464	0.9119	1.0000	
SFA_ES_Cook3	0.8244	0.8500	0.9149	0.9451	0.9536	1.0000

Табл. 4.10 Корреляции Спирмена (все модели, метод SFA-TI)

Следует отметить, что результаты оценивания моделей SFA сильнее всего коррелируют между собой (меньше чувствительность к смене объясняющих переменных). Недостаток метода SFA – чувствительность к изменениям в выборке: при использовании анализа выбросов не сошлись процедуры максимизации моделей 1 и 3 (о чем свидетельствуют нули в таблице 4.9)

4.6. Сравнение методов оценивания

Ниже представлено сравнение эффективностей, полученных по модели 1 (за 2011 год) без учета выбросов.

Variable	Obs	Mean	Std. Dev.	Min	Max
CS_COLS_ES1	59	.6701723	.1352566	.4039328	1
CS_MOLS_ES1	59	.8298981	.1674931	.5002043	1.238335
RE_COLS_ES1	59	.7108971	.1327004	.4143685	1
RE_MOLS_ES1	59	.8393949	.1566865	.4892675	1.180754
SFA_ES1	59	.7906803	.1501301	.3672765	.9729666
TVD_ES1	59	.7888748	.1390878	.4017265	.9736849

Табл. 4.11 Описательные статистики эффективностей по разным методам

Расположим методы в порядке убывания строгости оценок эффективности (опираясь на средние значения) в терминах предпочтения (предпочтительнее менее жесткие методы):

$$CS_{COLS} < RE_{COLS} < SFA(TVD) < SFA(TI) < CS_{MOLS} < RE_{MOLS}.$$

Этот результат сохраняется и для 2 и 3 моделей (см. Приложение б). Метод MOLS искусственно завышает значения эффективностей. Его недостаток в том, что по сравнению с COLS сильнее завышаются значения эффективных компаний, а значения низкоэффективных компаний меняются слабо. Это в частности приводит к появлению ряда сверхэффективных компаний (со значениями $ES > 1$), что создает дополнительные трудности при регулировании. Оптимальные значения дает метод SFA, однако и он имеет серьезный недостаток: процедура максимизации зачастую не сходится, в особенности при сильно неоднородной выборке. Так как все методы имеют свои преимущества и недостатки, общераспространенной практикой сегодня является взвешивание результатов разных моделей. Рассмотрим два ряда эффективностей, полученных из следующих моделей: RE и SFA(TI) (два метода на панельных данных, дающие $ES \leq 1$).

$$ES1 = [RE_{COLS}(1) + SFA_{TI}(1) + RE_{COLS}(2) + SFA_{TI}(2) + RE_{COLS}(3) + SFA_{TI}(3)]/6$$

$$ES2 = \max [RE_{COLS}(1), SFA_{TI}(1), RE_{COLS}(2), SFA_{TI}(2), RE_{COLS}(3), SFA_{TI}(3)]$$

Variable	Obs	Mean	Std. Dev.	Min	Max
CS_COLS_ES1	59	.6701723	.1352566	.4039328	1
CS_MOLS_ES1	59	.8298981	.1674931	.5002043	1.238335
RE_COLS_ES1	59	.7108971	.1327004	.4143685	1
RE_MOLS_ES1	59	.8393949	.1566865	.4892675	1.180754
SFA_ES1	59	.7906803	.1501301	.3672765	.9729666
TVD_ES1	59	.7888748	.1390878	.4017265	.9736849
ES1	59	.7625141	.129344	.4113426	.9700506
ES2	59	.8402829	.1319138	.4488186	1

Табл. 4.12 Описательные статистики эффективностей по разным методам + взвешивание

	CS_CO-S1	RE_CO-S1	SFA_ES1	ES1	ES2
CS_COLS_ES1	1.0000				
RE_COLS_ES1	0.9262	1.0000			
SFA_ES1	0.8371	0.9290	1.0000		
ES1	0.8203	0.8921	0.8905	1.0000	
ES2	0.7846	0.8550	0.8463	0.9468	1.0000

Табл. 4.13 Ранговые корреляции эффективностей по разным методам + взвешивание

Из Табл. 4.12 видно, что наименее жесткие значения эффективности дает метод максимального из нескольких моделей. Однако эти значения эффективности (ES2) слабее коррелируют с изначальными рядами по сравнению с усредненным рядом (ES1).

5. Заключение

В работе были решены следующие задачи: во-первых, был проведен обзор и анализ теоретических методов моделирования границы затрат. Для дальнейшего применения были отобраны следующие методы: скорректированный МНК (COLS), модифицированный МНК (MOLS), расширения этих методов для случая панельной структуры данных, метод стохастической границы (SFA) – с постоянной и с переменной эффективностью и метод «истинных фиксированных эффектов» (TFE). Не использовались в дальнейшем непараметрические методы, а также некоторые эконометрические модели: модель с фиксированными эффектами (FE) и модель Pooled по причине несоответствия их теоретических свойств данным.

Во-вторых, был проанализирован зарубежный опыт государственного регулирования отрасли распределительных сетей. Было установлено, что наиболее популярным методом регулирования является метод «верхней границы цены» (CPI-X), для применения которого необходим расчет индивидуальных X-факторов (X-фактор – требование к снижению затрат, % в год). Были подробно проанализированы конкретные применяемые на практике эконометрические модели (в Швейцарии и Норвегии), а также кратко охарактеризованы общие для всех исследований моменты. На данном этапе был сделан вывод об оптимальности функции Кобба-Дугласа для моделирования границы затрат, а также отмечено, какие объясняющие переменные используются в мировой практике.

В-третьих, были проанализированы технологические процессы энергопередающей компании и составлен желательный для включения в модель набор объясняющих переменных, среди которых полезный отпуск, трансформаторная емкость, плотность сети, климатические переменные и цены факторов. Согласно этому списку были собраны данные из открытых источников (сайт «Холдинга МРСК», РосСтат, СНИПы).

В-четвертых, на собранных данных были оценены три конкурирующие модели всеми отобранными способами. Неудовлетворительными оказались результаты оценивания методом TFE. Были оценены все те же модели с применением анализа выбросов, который позволяет в среднем увеличить получаемые значения эффективности затрат. Однако не имеет смысла применять анализ выбросов при оценивании сложных SFA-моделей (Battese&Coelli), так как при сокращении выборки перестает сходиться процедура максимизации. Все методы были сопоставлены между собой; их можно расположить следующим образом в порядке убывания «строгости» оценок эффективности: COLS (данные одного года) – COLS (панельные данные) – SFA (переменная во времени эффективность) – SFA (постоянная эффективность) – MOLS (данные одного года) – MOLS (панельные данные).

Метод MOLS самый мягкий, однако он имеет существенный недостаток: ряд фирм оказывается сверхэффективными, что затрудняет применение этого метода на практике. Метод COLS самый простой и интуитивно понятный, однако его недостаток в том, что ошибка в регрессии целиком интерпретируется как неэффективность, а значит всегда при оценке модели COLS имеет место занижение оценок эффективности (наиболее жесткий метод). Метод SFA занимает некое оптимальное срединное положение в плане жесткости оценок, однако и он имеет ряд недостатков. Основной недостаток – риск несходимости процедуры максимизации, в особенности на сильно неоднородных выборках. Другой недостаток – требование к введению слишком сильных предпосылок, так как для оценки методом максимального правдоподобия необходимо специфицировать распределение ошибки. Однако несмотря на эти недостатки, метод SFA набирает популярность и всё больше стран используют SFA в целях государственного регулирования.

В последнее десятилетие появился класс сложных моделей SFA (одна из них – рассмотренная в работе модель TFE), учитывающих одновременно неоднородность и неэффективность. Теоретические работы наращивают

сложность моделей в попытке учесть как можно больше нюансов. В связи с этим следует помнить, что ни одна модель не будет идеально подходить ко всем задачам, что каждый метод имеет как свои достоинства, так и недостатки. Поэтому оптимальное решение для исследователя – не полагаться на результаты оценивания одной модели, а использовать некоторую комбинацию результатов нескольких моделей (в работе было предложено два способа: взвешивание результатов или использование максимума). С помощью этого приема исследователь может достичь более устойчивых к изменениям выборки результатов, не прибегая к использованию слишком сложного инструментария.

Следует помнить (об этом напоминают в конце своих работ ведущие исследователи в области моделирования границы затрат), что всякий результат моделирования не следует слепо применять в регулировании; модельные выводы скорее имеют рекомендательный характер. Поэтому государственному регулятору лучше стремиться к использованию по возможности более мягких методов, чтобы избежать перерегулирования отдельных компаний.

6. Список использованной литературы

- Agrell P. J., Farsi M., Filippini M. and Koller M., 2013. “Unobserved heterogeneous effects in the cost efficiency analysis of electricity distribution systems”. *Economics Working Paper Series*, Working Paper 13/171 January 2013.
- Aigner, D., C.A.K. Lovell, and P. Schmidt, 1977. “Formulation and estimation of stochastic frontier function models”. *Journal of Econometrics*, 6, 21–37.
- Aurora Energy’s Network cost structure”, *Benchmark Economics (May 2011)*, pp. 43-57
- Australian Competition & Consumer Commission, Australian Energy Regulator (ACCC/AER), 2012. “Benchmarking Opex and Capex in Energy Networks”, Working Paper no.6, May 2012, *ACCC/AER working paper series*.
- Battese, G., and T. Coelli, 1992. “Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India”. *Journal of Productivity Analysis*, 3, pp. 153–169.
- Battese, G., and T. Coelli, 1993. “A Stochastic Frontier Production Function. Incorporating a Model for Technical Inefficiency Effects”. *University of New Zealand*, Working Paper No 69 October 1993.
- Bollen, K. A., & Jackman, R. W., 1990. "Regression diagnostics: An expository treatment of outliers and influential cases". *Modern methods of data analysis*, 257-291.
- Cook, R. D., 1977. “Detection of Influential Observation in Linear”. *Technometrics*, 19(1).
- Cook, R. D., and S. Weisberg, 1982. "Residuals and influence in regression." *Monographs on statistics and applied probability*.
- Cornwell, C., Schmidt P. and Sickles R. C., 1990. “Production frontiers with cross-sectional and time-series variation in efficiency levels”. *Journal of Econometrics*, 46 (1990): 185-200.
- Damodaran, A., 2011. “Applied Corporate Finance” 3rd ed. *JOHN WILEY & SONS, INC.*
- Farrell, M., 1957. “The Measurement of Productive Efficiency”. *Journal of the Royal Statistical Society A, General*, 120, pp. 253–281.
- Farsi M. and Filippini M., 2003. “Regulation and measuring cost efficiency with panel data models : application to electricity distribution utilities”, Working Paper

No. 03-05, *Department of Economics, University of Lugano, Switzerland*, January 2003.

Farsi, M., M. Filippini and W. H. Greene, 2006. "Application of Panel Data Models in Benchmarking Analysis of the Electricity Distribution Sector", *Annals of Public and Cooperative Economics*, 77 (3): 271-290.

Filippini, M., 2011. "Benchmarking and the regulation of electricity distribution companies", presentation in Swiss Federal Institute of Technology, Zurich.

Gabrielsen, A., 1975. "On Estimating Efficient Production Functions," Working Paper No. A-85, Chr. Michelsen Institute, Department of Humanities and Social Sciences, Bergen, Norway.

Greene, W., 1980. "Maximum Likelihood Estimation of Econometric Frontier Functions," *Journal of Econometrics*, 13, pp. 27–56.

Greene, W. H., 1990. "A gamma-distributed stochastic frontier model". *Journal of econometrics*, 46(1), 141-163.

Greene, W., 2002. "Alternative Panel Data Estimators for the Stochastic Frontier Model", Manuscript. Department of Economics, Stern School of Business, New York University.

Greene, W., 2005. "Reconsidering heterogeneity in panel data estimators of the stochastic frontier model". *Journal of Econometrics*, 126, 269–303.

Greene, W. H., 2008. "The econometric approach to efficiency analysis". *The measurement of productive efficiency and productivity growth*, 92-250.

Hausman, J. A., 1978. "Specification tests in econometrics. *Econometrica*", *Journal of the Econometric Society*, 1251-1271.

Hausman, J. A. and W. Taylor, 1981. "Panel data and unobservable individual effects" *Econometrica*, 49, 1377-1399.

Jamasb, T. and M. Pollitt (2001) "Benchmarking and regulation: international electricity experience" *Utilities Policy*, 9 (3), 107-130.

Jamasb, T. & Pollitt, M., 2003. International benchmarking and regulation: an application to European electricity distribution utilities. *Energy Policy*, 31(15), 1609-1622.

Lazarsfeld, Paul F., and Neil W. Henry, 1968. "*Latent Structure Analysis*", Boston, MA, USA: Houghton Mifflin.

Lovell, K., 1993. "Production Frontiers and Productive Efficiency," *The Measurement of Productive Efficiency*, H. Fried, K. Lovell, and S. Schmidt, eds., Oxford University Press, Oxford, UK.

Meeusen, W., and J. Van den Broeck, 1977. "Efficiency estimation from Cobb-Douglas production functions with composed error". *International Economic Review*, 18, 435–444.

Mundlak, Y., 1978. "On the pooling of time series and cross section data". *Econometrica: journal of the Econometric Society*, 69-85.

Pitt M.M., and Lee L.F., 1981. "The Measurement and Sources of Technical Inefficiency in the Indonesian Waving Industry," *Journal of Development Economics*, 9, 43-64.

Schmidt, P., and R. Sickles, 1984. "Production Frontiers and Panel Data," *Journal of Business and Economic Statistics*, 2, pp. 367–374.

Schweinsberg, Stronzik, Wissner, 2011. "Cost Benchmarking in Energy Regulation in European Countries," WIK-Consult (Final Report)

Simar, L., 1992. "Estimating efficiencies from frontier models with panel data: a comparison of parametric, non-parametric and semi-parametric methods with bootstrapping". *Journal of productivity analysis*, 3(1-2), 171-203.

Stevenson, R. E., 1980. "Likelihood functions for generalized stochastic frontier estimation". *Journal of Econometrics*, 13(1), 57-66.

Weinstein, M. A., 1964. "Query 2: The sum of values from a normal and a truncated normal distribution". *Technometrics*, 6(1), 104-105.

Winsten, C., 1957. "Discussion on Mr. Farrell's Paper," *Journal of the Royal Statistical Society, Series A, General*, 120, pp. 282–284.

Доугерти К. Введение в эконометрику. – Б. м. : ИНФРА-М, 2001.

РосСтат, Центральная База Статистических данных. www.gks.ru/cbsd

СНиП 2.01.07-85* «Нагрузки и воздействия»

СНиП 23-01-99 «Строительная климатология»

7. Приложения

Приложение 1. Результаты оценивания регрессии (модель 1) методом МНК по данным 2011 года без учета выбросов

year = 2011

Source	SS	df	MS	Number of obs =	59
Model	23.6721498	6	3.94535829	F(6, 52) =	82.50
Residual	2.48662072	52	.047819629	Prob > F =	0.0000
Total	26.1587705	58	.451013284	R-squared =	0.9049
				Adj R-squared =	0.8940
				Root MSE =	.21868

opex_p	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
wage_p	.4548246	.1508283	3.02	0.004	.1521655	.7574837
transf	.6208125	.0629741	9.86	0.000	.4944457	.7471793
connect	.265842	.0701313	3.79	0.000	.1251132	.4065708
density	-.1782803	.0943894	-1.89	0.065	-.3676865	.011126
january	.7830553	.6194121	1.26	0.212	-.4598859	2.025996
zero	.6813981	.2615574	2.61	0.012	.1565448	1.206251
_cons	-16.93697	3.10908	-5.45	0.000	-23.17579	-10.69814

Variable	VIF	1/VIF
connect	5.62	0.177984
density	3.36	0.297316
transf	2.81	0.355867
wage_p	1.37	0.731774
january	1.19	0.840911
zero	1.12	0.893079
Mean VIF	2.58	

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	59	-59.72346	9.697704	7	-5.395408	9.147354

Note: N=Obs used in calculating BIC; see [R] BIC note

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
 Ho: Constant variance
 Variables: fitted values of opex_p

chi2(1) = 0.20
 Prob > chi2 = 0.6568

Краткие выводы: в регрессии отсутствуют проблемы мультиколлинеарности ($\max VIF < 10$), гетероскедастичности ($\text{Prob} > \text{chi}2 > 0.05$), незначима только одна климатическая переменная. Все коэффициенты имеют верный знак.

Приложение 2. Результаты оценивания регрессии (модель 1) методом МНК по данным 2011 года с учетом выбросов

Source	SS	df	MS	Number of obs =	52
Model	20.4583983	6	3.40973305	F(6, 45) =	135.31
Residual	1.13396781	45	.025199285	Prob > F =	0.0000
Total	21.5923661	51	.423379727	R-squared =	0.9475
				Adj R-squared =	0.9405
				Root MSE =	.15874

opex_p	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
wage_p	.9639943	.1738275	5.55	0.000	.6138877 1.314101
transf	.5478135	.0580861	9.43	0.000	.430822 .6648049
connect	.3622753	.0719717	5.03	0.000	.2173168 .5072338
density	-.3124783	.0984205	-3.17	0.003	-.5107073 -.1142493
january	.7343175	.5369599	1.37	0.178	-.3471752 1.81581
zero	.378196	.201569	1.88	0.067	-.0277848 .7841767
_cons	-16.32184	2.604005	-6.27	0.000	-21.56657 -11.0771

Variable	VIF	1/VIF
connect	10.37	0.096471
density	6.17	0.162137
transf	3.64	0.274730
wage_p	1.60	0.626296
january	1.38	0.722132
zero	1.09	0.915809
Mean VIF	4.04	

Model	Obs	ll(null)	ll(model)	df	AIC	BIC
.	52	-50.9333	25.67874	7	-37.35748	-23.69877

Note: N=Obs used in calculating BIC; see [R] BIC note

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

Ho: Constant variance

Variables: fitted values of opex_p

chi2(1) = 1.55

Prob > chi2 = 0.2124

В очищенной от выбросов выборке появляется мультиколлинеарность (maxVIF>10). Остальные выводы прежние (см. Прил.1).

Приложение 3. Результаты оценивания модели 1 на панельных данных (метод RE) без учета выбросов.

```

Random-effects GLS regression           Number of obs   =       177
Group variable: id                     Number of groups =        59

R-sq:  within = 0.0067                  Obs per group:  min =         3
        between = 0.9113                  avg =           3.0
        overall = 0.9032                  max =           3

corr(u_i, X) = 0 (assumed)              Wald chi2(6)    =       505.35
                                           Prob > chi2     =        0.0000
    
```

opex_p	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
wage_p	.3828273	.1092096	3.51	0.000	.1687805	.5968742
transf	.5778471	.0600642	9.62	0.000	.4601234	.6955709
connect	.3350491	.0675532	4.96	0.000	.2026474	.4674509
density	-.2794645	.0893224	-3.13	0.002	-.4545331	-.1043958
january	.8834601	.5981137	1.48	0.140	-.2888211	2.055741
zero	.54854	.261164	2.10	0.036	.036668	1.060412
_cons	-17.00525	3.056768	-5.56	0.000	-22.9964	-11.01409
sigma_u	.20700512					
sigma_e	.07466725					
rho	.88487256	(fraction of variance due to u_i)				

Тест на адекватность модели RE:

Breusch and Pagan Lagrangian multiplier test for random effects

$$opex_p[id,t] = Xb + u[id] + e[id,t]$$

Estimated results:

	Var	sd = sqrt(Var)
opex_p	.4605092	.6786082
e	.0055752	.0746672
u	.0428511	.2070051

Test: Var(u) = 0

chibar2(01) = 130.06
 Prob > chibar2 = 0.0000

Дисперсия случайного эффекта отлична от нуля (модель адекватна)

Тест Хаусмана на эндогенность:

	---- Coefficients ----		(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
	(b) F1	(B) R1		
wage_p	.1414169	.3828273	-.2414105	.1160337
transf	-.5119949	.5778471	-1.089842	.3195085
connect	.91719	.3350491	.5821409	.3988983
density	-1.112889	-.2794645	-.8334243	.3731283

b = consistent under Ho and Ha; obtained from xtreg
 B = inconsistent under Ha, efficient under Ho; obtained from xtreg

Test: Ho: difference in coefficients not systematic

chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
 = 16.18
 Prob>chi2 = 0.0028

В модели присутствует эндогенность, использование RE теоретически некорректно.

Приложение 4. Определение id выбросов (на панельных данных)

COOK_PD_OU T1	year			Total
	2009	2010	2011	
3	1	1	1	3
4	1	1	1	3
6	1	0	1	2
18	1	1	1	3
27	1	1	1	3
47	1	0	0	1
55	1	0	0	1
57	0	1	1	2
58	1	1	1	3
Total	8	6	7	21

Приложение 5. SFA без учета выбросов

Модель с постоянной эффективностью Number of obs = 177
 Group variable: id Number of groups = 59

Obs per group: min = 3
 avg = 3
 max = 3

Wald chi2(6) = 2516.06
 Prob > chi2 = 0.0000

Log likelihood = 117.39561

opex_p	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
wage_p	.2343961	.0554796	4.22	0.000	.1256581	.3431342
transf	.5784765	.0519037	11.15	0.000	.4767471	.6802059
connect	.420227	.0658532	6.38	0.000	.2911571	.5492968
density	-.3120523	.0523424	-5.96	0.000	-.4146414	-.2094631
january	1.875545	.3323033	5.64	0.000	1.224243	2.526848
zero	.3475221	.137526	2.53	0.012	.0779761	.6170681
_cons	-21.80035	1.684309	-12.94	0.000	-25.10153	-18.49916

/mu	-.4649066	.9880125	-0.47	0.638	-2.401376	1.471562
/lnsigma2	-1.439956	1.057322	-1.36	0.173	-3.51227	.632358
/ilgtgamma	3.663874	1.096972	3.34	0.001	1.513848	5.8139

sigma2	.2369382	.2505201			.0298291	1.882043
gamma	.9750076	.0267308			.8196308	.9970231
sigma_u2	.2310165	.2505471			-.2600467	.7220798
sigma_v2	.0059217	.0007674			.0044175	.0074258

Модель с переменной эффективностью Number of obs = 177
 Group variable: id Number of groups = 59

Time variable: year Obs per group: min = 3
 avg = 3
 max = 3

Wald chi2(6) = 1962.45
 Prob > chi2 = 0.0000

Log likelihood = 121.83381

opex_p	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
wage_p	.2270324	.0604056	3.76	0.000	.1086396	.3454251
transf	.6172886	.0537376	11.49	0.000	.5119647	.7226124
connect	.3640391	.0723571	5.03	0.000	.2222217	.5058565
density	-.2773391	.0560618	-4.95	0.000	-.3872184	-.1674599
january	1.799714	.3586662	5.02	0.000	1.096741	2.502687
zero	.4068917	.1554912	2.62	0.009	.1021345	.7116489
_cons	-21.41208	1.837515	-11.65	0.000	-25.01354	-17.81061

/mu	-.049863	.4019388	-0.12	0.901	-.8376486	.7379225
/eta	.0648614	.020951	3.10	0.002	.0237983	.1059246
/lnsigma2	-2.107451	.8005009	-2.63	0.008	-3.676404	-.5384979
/ilgtgamma	3.050931	.841914	3.62	0.000	1.40081	4.701052

sigma2	.1215474	.0972988			.0253138	.5836243
gamma	.9548227	.0363171			.8023124	.9909961
sigma_u2	.1160562	.0972621			-.074574	.3066865
sigma_v2	.0054912	.0007133			.0040931	.0068893

Параметр темпа роста эффективности /eta значим, поэтому предпочтительно использование модели TVD.

**Приложение 6. Сравнение различных методов оценивания
для моделей 2 и 3.**

Variable	Obs	Mean	Std. Dev.	Min	Max
CS_COLS_ES2	59	.6350336	.1310399	.3828953	1
CS_MOLS_ES2	59	.8270448	.1706616	.498669	1.302364
RE_COLS_ES2	59	.71869	.136195	.4089408	1
RE_MOLS_ES2	59	.8345076	.1581429	.4748421	1.161151
SFA_ES2	59	.7962515	.1545709	.4029825	.9697367
TVD_ES2	59	.2384639	.0445553	.136762	.3366821

Примечание: SFA(TVD) для модели 2 не сошелся.

Variable	Obs	Mean	Std. Dev.	Min	Max
CS_COLS_ES3	59	.7163308	.1269013	.4441614	1
CS_MOLS_ES3	59	.8461725	.1499034	.5246699	1.18126
RE_COLS_ES3	59	.7567183	.1254531	.4488186	1
RE_MOLS_ES3	59	.8532182	.1414514	.5060538	1.127524
SFA_ES3	59	.8018477	.1347251	.4172881	.9742712
TVD_ES3	59	.798145	.123231	.4597777	.9728678

В других моделях сохраняется порядок «жесткости» моделей:

$$CS_{COLS} < RE_{COLS} < SFA(TVD) < SFA(TI) < CS_{MOLS} < RE_{MOLS}.$$