

# **квантиль**

*международный эконометрический журнал  
на русском языке*

**№5  
сентябрь 2008 г.**

## **СОДЕРЖАНИЕ ВЫПУСКА**

### **Эконометрический ликбез: некоторые вопросы микроэконометрики**

Родригес Герман. Модели выживаемости	1
Макфадден Даниэль. Полупараметрический анализ	29

### **В помощь изучающим эконометрику**

Цыплаков Александр. Мини-словарь англоязычных эконометрических терминов, часть 2	41
Анатольев Станислав. Обзор англоязычных учебников по анализу временных рядов	49

### **Задачи и решения**

Задачи 5.1, 5.2, 5.3	57
----------------------	----

### **Статьи: прикладная эконометрика**

Назруллаева Евгения. Оценивание уровня технологического прогресса в российской экономике	59
Демидов Олег. Различные индексы прогнозирования экономической активности в России	83

# **Квантиль**

**№5, сентябрь 2008 г.**

Сайт в Интернете: <http://quantile.ru>

Адрес электронной почты: [quantile@quantile.ru](mailto:quantile@quantile.ru)

Доступ к журналу бесплатный и неограниченный

## **РЕДАКТОР**

Станислав Анатольев

Российская Экономическая Школа (Москва, Россия)

## **РЕДАКЦИОННЫЙ СОВЕТ**

Виктория Зинде-Уолш

Университет МакГилл (Монреаль, Канада)

Рустам Ибрагимов

Гарвардский Университет (Кэмбридж, США)

Анна Микушева

Массачусетский Технологический Институт (Кэмбридж, США)

Алексей Онацкий

Колумбийский Университет (Нью-Йорк, США)

Владимир Павлов

Технологический университет Квинсленда (Брисбен, Австралия)

Константин Тюрин

Университет штата Индиана (Блумингтон, США)

Александр Цыплаков

Новосибирский Государственный Университет (Новосибирск, Россия)

Виктор Черножуков

Массачусетский Технологический Институт (Кэмбридж, США)

## **К СВЕДЕНИЮ АВТОРОВ**

Рукописи для публикации в разделе «Статьи» принимаются в электронном виде по адресу [submit@quantile.ru](mailto:submit@quantile.ru). Работы могут принадлежать любой прикладной сфере экономической науки. Главным требованием является интенсивное использование адекватных эконометрических методов. Рукопись должна быть написана на русском (для русскоязычных авторов) или на английском (для остальных авторов) языке в формате *Microsoft Word* или (предпочтительнее) *LaTeX*, и по объему не превышать 30 страниц формата А4 с двойным междустрочным интервалом. Работы подвергаются контролю качества членами редакционного совета и независимыми референтами. Перспективная работа может быть при необходимости возвращена автору на доработку. Редакция также приглашает к сотрудничеству экспертов по эконометрике, готовых внести вклад в методологические рубрики журнала.

При публикации статьи или методологического эссе в журнале «Квантиль» передача авторских прав не происходит ни полностью, ни частично.

# Эконометрический ликбез: некоторые вопросы микроэконометрики

## Модели выживаемости\*

Герман Родригес<sup>†</sup>

*Принстонский Университет, Принстон, США*

Настоящее эссе представляет собой введение в модели выживаемости в контексте обобщенных линейных моделей. Вводятся понятия функций риска и выживания, а затем рассматриваются наиболее распространенные механизмы цензурирования и получаемые в результате функции правдоподобия. Обсуждаются основные подходы к моделированию времени ожидания, включая модели ускоренной жизни и пропорциональных рисков, и их расширения для случаев меняющихся во времени регрессоров и зависящих от времени коэффициентов. Затем подробно изучается кусочно-экспоненциальная модель выживаемости и отмечается ее эквивалентность модели пуассоновской регрессии. Далее применение этого подхода рассматривается на примере анализа младенческой и детской смертности в Колумбии по данным опроса. В заключении кратко обсуждаются модели в дискретном времени и их эквивалентность модели логистической регрессии.

## 1 Введение

В настоящем эссе рассматриваются модели анализа данных, обладающих следующими тремя основными характеристиками: (а) зависимая переменная, или отклик, – это *время ожидания* до наступления определенного события, (б) наблюдения являются *цензурированными* в том смысле, что для некоторых объектов наблюдения исследуемое событие не наступило на момент анализа данных, и (в) имеются предикторы, или *объясняющие переменные*, чье воздействие на время ожидания мы желаем оценить или учесть. Начнем с некоторых базовых определений.

## 2 Функции риска и выживания

Пусть  $T$  – неотрицательная случайная величина, представляющая собой время ожидания до наступления некоторого события. Для простоты будем использовать терминологию анализа выживаемости, называя исследуемое событие «смертью», а время ожидания – временем «выживания», хотя изучаемые далее методы находят гораздо более широкое применение. Их можно использовать, например, для анализа возраста при вступлении в брак, продолжительности брака, интервалов между последовательными родами у женщин, времени пребывания в городе (или на определенном месте работы) и продолжительности жизни. Наблюдательный демограф заметит, что эти примеры включают проблемы рождаемости, смертности и миграции.

\*Перевод Б. Гершмана и С. Анатольева. Цитировать как: Родригес, Герман (2008) «Модели выживаемости», Квантиль, №5, стр. 1–27. Citation: Rodríguez, Germán (2008) “Survival models,” *Quantile*, No.5, pp. 1–27.

<sup>†</sup>Адрес: Office of Population Research, Princeton University, 241 Wallace Hall, Princeton, NJ 08544, USA. Электронная почта: [grodri@Princeton.edu](mailto:grodri@Princeton.edu)

## 2.1 Функция выживания

Предположим, что  $T$  – непрерывная случайная величина с функцией плотности распределения (ФПР)  $f(t)$  и кумулятивной функцией распределения (КФР)  $F(t) = \mathbb{P}\{T \leq t\}$ , дающей вероятность того, что событие наступило к моменту времени  $t$ .

Часто удобно работать с дополнением КФР, называемым функцией *выживания*:

$$S(t) = \mathbb{P}\{T > t\} = 1 - F(t) = \int_t^{\infty} f(x)dx, \quad (1)$$

и дающим вероятность быть живым в момент времени  $t$ , или в более широком смысле, вероятность того, что исследуемое событие не наступило к моменту времени  $t$ .

## 2.2 Функция риска

Альтернативным способом охарактеризовать распределение величины  $T$  является *функция риска*, или мгновенная интенсивность осуществления события, определяемая как

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{P}\{t < T \leq t + dt | T > t\}}{dt}. \quad (2)$$

Числитель этого выражения – условная вероятность того, что событие произойдет в интервале  $(t, t + dt)$ , если оно не произошло ранее, а знаменатель – ширина интервала. Разделив одно на другое, получаем интенсивность осуществления события в единицу времени. Устремляя ширину интервала к нулю и переходя к пределу, получаем мгновенную интенсивность осуществления события.

Условную вероятность в числителе можно записать в виде отношения совместной вероятности того, что  $T$  принадлежит интервалу  $(t, t + dt)$  и  $T > t$  (что, конечно, совпадает с вероятностью того, что  $T$  принадлежит указанному интервалу), к вероятности условия  $T > t$ . Первая из них равна  $f(t)dt$  для малого  $dt$ , а последняя – это  $S(t)$ , по определению. Деление на  $dt$  и предельный переход дают следующий полезный результат:

$$\lambda(t) = \frac{f(t)}{S(t)}, \quad (3)$$

который некоторые авторы приводят в качестве определения функции риска. Содержательно, интенсивность осуществления события в момент времени  $t$  равна плотности событий в момент  $t$ , деленной на вероятность дожить до этого момента, не испытав событие ранее.

Заметим из уравнения (1), что  $-f(t)$  – это производная  $S(t)$ . Тогда уравнение (3) можно переписать в виде

$$\lambda(t) = -\frac{d}{dt} \log S(t).$$

Если теперь проинтегрировать обе части от 0 до  $t$  и ввести граничное условие  $S(0) = 1$  (поскольку событие не может произойти к моменту времени 0), можно преобразовать приведенное выражение и получить формулу для вероятности дожить до момента времени  $t$  как функции от рисков во все моменты времени до  $t$ :

$$S(t) = \exp \left\{ - \int_0^t \lambda(x)dx \right\}. \quad (4)$$

Это выражение должно быть знакомо демографам. Интеграл в фигурных скобках в этом уравнении называют *кумулятивным риском* и обозначают как

$$\Lambda(t) = \int_0^t \lambda(x)dx. \quad (5)$$

Можно рассматривать  $\Lambda(t)$  как сумму всех рисков при переходе от момента времени 0 к  $t$ .

Приведенные результаты показывают, что функции выживания и риска дают альтернативные, но эквивалентные описания распределения величины  $T$ . Имея функцию выживания, всегда можно ее продифференцировать и получить функцию плотности, а затем найти функцию риска, используя уравнение (3). Имея функцию риска, всегда можно ее проинтегрировать и получить кумулятивный риск, а затем взять от нее экспоненту и найти функцию выживания, используя уравнение (4). Для закрепления введенных понятий рассмотрим пример.

*Пример:* Простейшее распределение времени жизни получается, если предположить постоянный риск, то есть

$$\lambda(t) = \lambda$$

для всех  $t$ . Соответствующая функция выживания имеет вид

$$S(t) = \exp\{-\lambda t\}.$$

Это экспоненциальное распределение с параметром  $\lambda$ . Функцию плотности можно получить, умножив функцию выживания на риск:

$$f(t) = \lambda \exp\{-\lambda t\}.$$

Математическое ожидание равно  $1/\lambda$ . Это распределение играют центральную роль в анализе выживаемости, хотя, возможно, оно является слишком простым, чтобы быть полезным в приложениях само по себе.

### 2.3 Ожидаемая продолжительность жизни

Пусть  $\mu$  обозначает математическое ожидание  $T$ . По определению, значение  $\mu$  можно подсчитать, умножив  $t$  на функцию плотности  $f(t)$  и взяв интеграл, то есть

$$\mu = \int_0^{\infty} t f(t) dt.$$

Интегрируя по частям и используя тот факт, что  $-f(t)$  – это производная  $S(t)$ , удовлетворяющая граничным условиям  $S(0) = 1$  и  $S(\infty) = 0$ , можно показать, что

$$\mu = \int_0^{\infty} S(t) dt. \tag{6}$$

Иными словами, ожидаемая продолжительность жизни – это просто интеграл от функции выживания.

### 2.4 Замечание о несобственных случайных величинах

До сих пор неявно предполагалось, что исследуемое событие обязательно происходит, то есть  $S(\infty) = 0$ . Иначе говоря, по прошествии достаточного времени доля выживших снижается к нулю. Из этого условия следует, что кумулятивный риск должен расходиться, то есть  $\Lambda(\infty) = \infty$ . Интуитивно, событие точно произойдет только в том случае, если кумулятивный риск за долгий период времени достаточно высок.

Есть, тем не менее, множество возможных событий, которые необязательно происходят. Некоторые мужчины и женщины остаются одинокими всю жизнь, вторые и последующие роды могут не произойти, а некоторые люди достаточно счастливы на своем месте работе, чтобы никогда его не покидать. Что делать в таких случаях? Существуют два подхода.

Один подход – заметить, что все равно можно найти функции риска и выживания, которые корректно определены, даже если исследуемое событие может и не произойти. Например,

можно изучать брачность всего населения, включая людей, которые никогда не вступят в брак, и подсчитать доли состоящих в браке и одиноких. В этом примере  $S(t)$  будет отражать долю людей, не состоящих в браке в возрасте  $t$ , а  $S(\infty)$  – долю тех, кто никогда не вступит в брак.

Одно из ограничений данного подхода в том, что, если событие не обязательно должно произойти, время ожидания  $T$  может быть неопределенным (или бесконечным) и в таком случае не является собственной случайной величиной. Ее функция плотности, которую можно вычислить по функциям риска и выживания, будет несобственной, то есть не будет интегрироваться к единице. Очевидно, среднее время ожидания не будет определено. В терминах нашего примера, нельзя подсчитать средний возраст вступления в брак для всего населения, просто поскольку не все люди вступают в брак. Но это ограничение не имеет серьезных последствий, если внимание сосредоточено на функциях риска и выживания, а не на времени ожидания. В примере с браком возможно даже вычислить медианный возраст вступления в брак, если определить его как возраст, к которому половина всего населения вступает в брак.

Альтернативный подход – проводить анализ условно на осуществлении события. В терминах нашего примера, можно было бы изучать брачность (возможно, в ретроспективе) для людей, которые в конечном счете вступают в брак, поскольку для этой группы людей время ожидания  $T$  всегда корректно определено. В таком случае можно подсчитать не только условные функции риска и выживания, но и среднее время ожидания. В нашем примере можно вычислить средний возраст при вступлении в брак для тех, кто на самом деле вступает в брак. Можно даже подсчитать обычную медиану, определенную как возраст, к которому вступила в брак половина той части населения, которая в конечном счете вступает в брак.

Оказывается, что условные функции плотности, риска и выживания для тех, кто испытывает событие, связаны с соответствующими безусловными функциями для всего населения. Условная функция плотности имеет вид

$$f^*(t) = \frac{f(t)}{1 - S(\infty)}$$

и интегрируется к единице. Условная функция выживания имеет вид

$$S^*(t) = \frac{S(t) - S(\infty)}{1 - S(\infty)},$$

и стремится к нулю при  $t \rightarrow \infty$ . Поделив условную функцию плотности на условную функцию выживания, получаем условную функцию риска:

$$\lambda^*(t) = \frac{f^*(t)}{S^*(t)} = \frac{f(t)}{S(t) - S(\infty)}.$$

В качестве упражнения читатель может подсчитать среднее время ожидания для тех, с кем событие случается.

Какой бы подход ни применялся, следует аккуратно указать, какие именно функции риска и выживания используются. Например, условный риск для тех, кто в конечном счете испытывает событие, всегда выше, чем безусловный риск для всего населения. Заметим также, что в большинстве случаев все, что наблюдается, – это произошло событие или нет. Если событие не произошло, может быть, невозможно определить, произойдет ли оно в будущем. В таком случае по данным можно оценить лишь безусловный риск, но этот результат при желании всегда можно выразить в условных величинах, используя приведенные выше выражения.

### 3 Цензурирование и функция правдоподобия

Вторая отличительная черта анализа выживаемости – цензурирование, то есть тот факт, что для некоторых объектов наблюдения исследуемое событие произошло, а значит, известно точное время ожидания, тогда как для других это событие не произошло, и все, что известно, – это то, что время ожидания превышает время наблюдения.

#### 3.1 Механизмы цензурирования

Существует несколько механизмов, способных генерировать цензурированные данные. При цензурировании *типа I* выборка из  $n$  объектов наблюдается в течении фиксированного времени  $\tau$ . Число объектов, испытывающих событие, или число «смертей», случайно, но общая продолжительность исследования фиксирована. Тот факт, что продолжительность фиксирована, может быть важным практическим преимуществом при разработке последующего дополнительного исследования.

При простом обобщении этой схемы, называемом *фиксированным цензурированием*, каждый объект имеет максимально возможный период наблюдения  $\tau_i$ ,  $i = 1, \dots, n$ , который может варьироваться от одного объекта к другому, однако фиксирован заранее. Вероятность того, что объект  $i$  будет жив в конце своего периода наблюдения, равна  $S(\tau_i)$ , а общее число смертей вновь является случайным.

При цензурировании *типа II* выборка из  $n$  объектов наблюдается так долго, сколько необходимо, чтобы  $d$  объектов испытали событие. В этой схеме число смертей  $d$ , которое определяет точность исследования, фиксировано заранее и его можно использовать в качестве параметра. К сожалению, в этом случае общая продолжительность исследования случайна и не может быть точно известна заранее.

При более общей схеме, называемой *случайным цензурированием*, каждый объект имеет потенциальный момент цензурирования  $C_i$  и потенциальную продолжительность жизни  $T_i$ , которые предполагаются независимыми случайными величинами. Наблюдается  $Y_i = \min\{C_i, T_i\}$ , то есть минимум из времени цензурирования и времени жизни, и переменная-индикатор, часто обозначаемая  $d_i$  или  $\delta_i$ , которая указывает, закончено наблюдение в результате смерти или цензурирования.

Все эти схемы объединяет тот факт, что механизм цензурирования *неинформативен*, и все они, в сущности, ведут к той же самой функции правдоподобия. Наиболее слабое предположение, требуемое для получения этой функции правдоподобия, состоит в том, что цензурирование наблюдения не должно давать какой-либо информации относительно перспектив выживания этого конкретного объекта за пределами момента цензурирования. На самом деле базовое предположение, которому мы будем следовать, таково: все, что известно о наблюдении, цензурированном в момент времени  $t$  – это то, что время жизни для него превышает  $t$ .

#### 3.2 Функция правдоподобия для цензурированных данных

Предположим, что имеются  $n$  объектов наблюдения со временем жизни, характеризуемым функцией выживания  $S(t)$  с соответствующей плотностью  $f(t)$  и риском  $\lambda(t)$ . Предположим также, что объект  $i$  наблюдается в течение времени  $t_i$ . Если объект умер в момент  $t_i$ , его вклад в функцию правдоподобия – значение плотности в этот момент времени, которую можно записать как произведение функций выживания и риска:

$$L_i = f(t_i) = S(t_i)\lambda(t_i).$$

Если объект все еще жив в момент времени  $t_i$ , все, что известно при неинформативном цензурировании – это то, что время его жизни превышает  $t_i$ . Вероятность этого события

равна

$$L_i = S(t_i),$$

и отражает вклад цензурированного наблюдения в функцию правдоподобия.

Заметим, что оба варианта вкладов содержат функцию выживания  $S(t_i)$ , поскольку в обоих случаях объект дожил до момента времени  $t_i$ . Смерть «домножает» этот вклад на риск  $\lambda(t_i)$ , а цензурирование – нет. Можно записать оба типа вкладов в виде единого выражения. Для этого пусть  $d_i$  является индикатором смерти и равняется единице, если объект  $i$  умер, и нулю в противном случае. Тогда функцию правдоподобия можно записать в следующем виде:

$$L = \prod_{i=1}^n L_i = \prod_i \lambda(t_i)^{d_i} S(t_i).$$

Логарифмируя и используя выражение, связывающее функцию выживания  $S(t)$  и функцию кумулятивного риска  $\Lambda(t)$ , получаем логарифмическую функцию правдоподобия для цензурированных данных о выживаемости:

$$\log L = \sum_{i=1}^n \{d_i \log \lambda(t_i) - \Lambda(t_i)\}. \quad (7)$$

Для закрепления материала рассмотрим пример.

*Пример:* Предположим, имеется выборка размера  $n$  цензурированных наблюдений из экспоненциального распределения. Пусть  $t_i$  – период наблюдения, а  $d_i$  – индикатор смерти для объекта  $i$ .

В случае экспоненциального распределения  $\lambda(t) = \lambda$  для всех  $t$ . Кумулятивный риск, таким образом, является интегралом от константы, а значит,  $\Lambda(t) = \lambda t$ . Подставляя два этих результата в уравнение (7), получаем логарифмическую функцию правдоподобия

$$\log L = \sum_{i=1}^n \{d_i \log \lambda - \lambda t_i\}.$$

Пусть  $D = \sum_i d_i$  обозначает общее число смертей, а  $T = \sum_i t_i$  – общее время наблюдения (или подверженность риску). Тогда можно переписать логарифмическую функцию правдоподобия как функцию от этих совокупных величин:

$$\log L = D \log \lambda - \lambda T. \quad (8)$$

Дифференцируя это выражение относительно  $\lambda$ , получаем скор-функцию

$$u(\lambda) = \frac{D}{\lambda} - T,$$

и, приравнявая ее к нулю, находим оценку максимального правдоподобия для риска

$$\hat{\lambda} = \frac{D}{T}, \quad (9)$$

которая равна общему числу смертей, деленному на общую подверженность риску. Демографы узнают в этом выражении общее определение коэффициента смертности. Заметим, что эта оценка является оптимальной (в смысле максимального правдоподобия), только если риск является постоянным и не зависит от возраста.

Можно также подсчитать количество информации в выборке, взяв со знаком минус вторую производную скор-функции:

$$I(\lambda) = \frac{D}{\lambda^2}.$$

Для получения ожидаемой информации необходимо найти ожидаемое число смертей, но оно зависит от схемы цензурирования. Например, при цензурировании типа I с фиксированной продолжительностью  $\tau$ , ожидаемое число смертей равно  $n(1 - S(\tau))$ . При цензурировании типа II число смертей фиксируется заранее. При некоторых схемах подсчет данного матожидания может быть довольно затруднителен или даже невозможен.

Простой альтернативный вариант – использовать наблюдаемую информацию, оцененную с помощью ММП-оценки  $\lambda$  из уравнения (9). Используя этот подход, асимптотическую дисперсию ММП-оценки риска можно оценить как

$$\hat{V}(\hat{\lambda}) = \frac{D}{T^2},$$

и затем применять эту оценку для асимптотического тестирования гипотез и построения доверительных интервалов для  $\lambda$ .

В отсутствие цензурированных наблюдений, то есть при  $d_i = 1$  для всех  $i$  и  $D = n$ , полученные выше результаты становятся стандартным ММП-оцениванием для экспоненциального распределения, а ММП-оценка параметра  $\lambda$  равна величине, обратной выборочному среднему.

Интересно заметить, между прочим, что логарифмическая функция правдоподобия для цензурированных данных из экспоненциального распределения, приведенная в уравнении (8), в точности совпадает (не считая констант) с логарифмической функцией правдоподобия, возникающей, если считать  $D$  пуассоновской случайной величиной со средним значением  $\lambda T$ . Чтобы убедиться в этом, следует записать пуассоновскую логарифмическую функцию правдоподобия, когда  $D \sim P(\lambda T)$ , и заметить, что она отличается от уравнения (8) только наличием члена  $D \log(T)$ , являющегося константой, зависящей только от данных, но не от параметра  $\lambda$ .

Таким образом, если считать смерть пуассоновской величиной, условно на время подверженности риску, получаются точно такие же оценки (и стандартные ошибки), как в случае, когда периоды подверженности риску рассматриваются как цензурированные наблюдения из экспоненциального распределения. Этот результат разрабатывается ниже, чтобы связать модели выживаемости и обобщенные линейные модели с пуассоновской структурой ошибок.

## 4 Подходы к моделированию выживаемости

До сих пор речь шла об однородной популяции, в которой продолжительность жизни каждого объекта характеризовалась одной и той же функцией выживания  $S(t)$ . Рассмотрим теперь третью отличительную черту моделей выживаемости – наличие вектора регрессоров, или объясняющих переменных, которые могут воздействовать на время жизни, – и обратимся к общей задаче моделирования этих эффектов.

### 4.1 Модели ускоренной жизни

Пусть  $T_i$  – случайная величина, представляющая собой (возможно, ненаблюдаемое) время жизни  $i$ -го объекта. Поскольку величина  $T_i$  должна быть неотрицательной, можно рассмотреть модель для ее логарифма, скажем, обычную линейную модель

$$\log T_i = \mathbf{x}'_i \boldsymbol{\beta} + \epsilon_i,$$

где  $\epsilon_i$  – надлежащий остаточный член, распределение которого будет специфицировано далее. Эта модель задает распределение логарифма времени жизни для  $i$ -го объекта как простой *сдвиг* стандартного, или базового, распределения, представленного остаточным членом.

Взяв экспоненту от этого уравнения, получаем модель собственно для времени жизни:

$$T_i = \exp\{\mathbf{x}'_i \boldsymbol{\beta}\} T_{0i},$$

где  $T_{0i}$  – экспонента от остаточного члена. Удобно также использовать обозначение  $\gamma_i$  для мультипликативного эффекта регрессоров,  $\exp\{\mathbf{x}'_i\beta\}$ .

Интерпретация параметров стандартна. Рассмотрим, например, модель с константой и фиктивной переменной  $x$ , отражающей бинарный фактор, скажем, принадлежность к группам 1 или 0. Предположим, соответствующий мультипликативный эффект  $\gamma = 2$ , так что коэффициент при  $x$  – это  $\beta = \log(2) = 0,6931$ . Тогда вывод состоит в том, что люди из первой группы живут вдвое дольше, чем из нулевой.

Существует интересная альтернативная интерпретация, которая объясняет название «модель ускоренной жизни». Пусть  $S_0(t)$  обозначает функцию выживания для группы 0, которая будет контрольной группой, а  $S_1(t)$  – для группы 1. Для этой модели

$$S_1(t) = S_0(t/\gamma).$$

Иными словами, вероятность того, что индивид из первой группы доживет до возраста  $t$ , в точности равна вероятности того, что индивид из нулевой группы доживет до возраста  $t/\gamma$ . Для  $\gamma = 2$  получим половину возраста, так что вероятность того, что индивид из первой группы доживет до 40 (или 60) лет будет равна вероятности того, что индивид из нулевой группы доживет до 20 (или 30) лет. Таким образом, можно рассматривать  $\gamma$  как параметр, воздействующий на протекание времени. В нашем примере люди в нулевой группе стареют «в два раза быстрее».

Заметим, что соответствующие функции риска связаны соотношением

$$\lambda_1(t) = \lambda_0(t/\gamma)/\gamma,$$

так что при  $\gamma = 2$  в каждом данном возрасте люди из первой группы будут подвержены вдвое меньшему риску, чем вдвое младшие люди из нулевой группы.

Название «модель ускоренной жизни» происходит из промышленных приложений, когда предметы тестируются при гораздо худших условиях, чем встречающиеся в реальной жизни, чтобы тесты можно было выполнить за более короткое время.

Различные предположения о распределении остаточного члена приводят к различным видам параметрических моделей. Если ошибка  $\epsilon_i$  нормально распределена, получается логнормальная модель для  $T_i$ . Оценивание этой модели для цензурированных данных по методу максимального правдоподобия известно в эконометрической литературе как тобит-модель.

Если же  $\epsilon_i$  имеет распределение экстремального значения с функцией плотности

$$f(\epsilon) = \exp\{\epsilon - \exp(\epsilon)\},$$

то  $T_{0i}$  имеет экспоненциальное распределение, и получается модель экспоненциальной регрессии, где  $T_i$  экспоненциально распределено с риском  $\lambda_i$ , удовлетворяющим логлинейной модели

$$\log \lambda_i = \mathbf{x}'_i\beta.$$

Примером демографической модели, принадлежащей классу моделей ускоренной жизни, является модель Кола–Макнейла для частоты первого брака, в которой доля индивидов, когда-либо состоявших в браке к возрасту  $a$  в данной популяции, записывается в виде

$$F(a) = cF_0\left(\frac{a - a_0}{k}\right),$$

где  $F_0$  – модельное распределение долей женщин, состоявших в браке к определенному возрасту среди когда-либо состоявших в браке, на основе исторических данных по Швеции;  $c$  – доля тех, кто в конечном счете вступают в брак,  $a_0$  – возраст вступления в брак, а  $k$  – скорость протекания брака относительно шведского стандарта.

Модели ускоренной жизни по сути являются обычными моделями регрессии, примененными к логарифму времени жизни, и, не считая факта цензурирования данных, не представляют новых трудностей при оценивании. Как только выбрано распределение остаточного члена, оценивание осуществляется путем максимизации логарифмической функции правдоподобия для цензурированных данных, рассмотренной в предыдущем разделе. Детали можно найти в работе Kalbfleish & Prentice (1980).

## 4.2 Модели пропорциональных рисков

Большой класс моделей, впервые предложенный в Cox (1972), концентрируется непосредственно на функции риска. Простейший представитель этого класса – модель *пропорциональных рисков*, в которой риск в момент  $t$  для индивида с характеристиками  $\mathbf{x}_i$  (не включая константу) имеет вид

$$\lambda_i(t|\mathbf{x}_i) = \lambda_0(t) \exp\{\mathbf{x}'_i\boldsymbol{\beta}\}. \quad (10)$$

В этой модели  $\lambda_0(t)$  – это базовая функция риска, которая измеряет риск для индивидов с  $\mathbf{x}_i = \mathbf{0}$ , служащих точкой отсчета, а  $\exp\{\mathbf{x}'_i\boldsymbol{\beta}\}$  – относительный риск, то есть пропорциональное увеличение или уменьшение риска, связанное с набором характеристик  $\mathbf{x}_i$ . Заметим, что увеличение или снижение риска одинаково для всех моментов времени  $t$ .

В качестве иллюстрации рассмотрим пример с двумя выборками, когда имеется фиктивная переменная  $x$ , означающая принадлежность к первой или нулевой группе. В этом случае модель принимает вид

$$\lambda_i(t|x) = \begin{cases} \lambda_0(t) & \text{если } x = 0, \\ \lambda_0(t)e^\beta & \text{если } x = 1. \end{cases}$$

Таким образом,  $\lambda_0(t)$  представляет собой риск в момент  $t$  в нулевой группе, а  $\gamma = \exp\{\beta\}$  – это отношение риска в первой группе к риску в нулевой группе в любой момент времени  $t$ . Если  $\gamma = 1$  (или  $\beta = 0$ ), риски одинаковы в обеих группах. Если  $\gamma = 2$  (или  $\beta = 0,6931$ ), риск для индивида из первой группы в каждый момент времени вдвое больше риска для индивида того же возраста из нулевой группы.

Заметим, что модель явным образом отделяет эффект времени от эффекта регрессоров. Логарифмируя, легко увидеть, что модель пропорциональных рисков – это простая аддитивная модель для логарифма риска:

$$\log \lambda_i(t|\mathbf{x}_i) = \alpha_0(t) + \mathbf{x}'_i\boldsymbol{\beta},$$

где  $\alpha_0(t) = \log \lambda_0(t)$  – логарифм базового риска. Как во всех аддитивных моделях, предполагается что влияние регрессоров  $\mathbf{x}$  одинаково для всех моментов времени, или возрастов,  $t$ . Нельзя не отметить схожесть между этим выражением и стандартной моделью ковариационного анализа с параллельными прямыми.

Возвращаясь к уравнению (10), можно проинтегрировать обе его части от 0 до  $t$  и получить кумулятивные риски

$$\Lambda_i(t|\mathbf{x}_i) = \Lambda_0(t) \exp\{\mathbf{x}'_i\boldsymbol{\beta}\},$$

которые также пропорциональны. Взяв экспоненту от этого уравнения с противоположным знаком, получаем функции выживания

$$S_i(t|\mathbf{x}_i) = S_0(t)^{\exp\{\mathbf{x}'_i\boldsymbol{\beta}\}}, \quad (11)$$

где  $S_0(t) = \exp\{-\Lambda_0(t)\}$  – базовая функция выживания. Таким образом, эффект регрессоров  $\mathbf{x}_i$  на функцию выживания заключается в возведении ее в степень, равную относительному риску  $\exp\{\mathbf{x}'_i\boldsymbol{\beta}\}$ .

В нашем примере с двумя группами и относительным риском  $\gamma = 2$  вероятность того, что индивид из первой группы доживет до возраста  $t$ , равен квадрату вероятности того, что индивид из нулевой группы доживет до этого же возраста.

### 4.3 Экспоненциальная и вейбулловская модели

Различные виды моделей пропорциональных рисков можно получить, делая различные предположения о базовой функции выживания, или, что эквивалентно, о базовой функции риска. Например, если базовый риск постоянен во времени, то есть  $\lambda_0(t) = \lambda_0$ , получаем модель экспоненциальной регрессии, где

$$\lambda_i(t, \mathbf{x}_i) = \lambda_0 \exp\{\mathbf{x}_i' \boldsymbol{\beta}\}.$$

Интересно, что модель экспоненциальной регрессии принадлежит одновременно к классам моделей пропорциональных рисков и моделей ускоренной жизни. Если базовый риск постоянен, и происходит удвоение или утроение риска, новый риск также будет постоянным (просто более высоким). Возможно, менее очевидным является тот факт, что в случае постоянного базового риска при ускорении течения времени в два или три раза новый риск удваивается или утраивается, но тем не менее остается постоянным во времени, то есть мы остаемся в рамках экспоненциального класса.

Можно задаться вопросом, есть ли другие случаи, при которых две модели совпадают. Есть, но немного. На самом деле, есть всего лишь одно распределение, при котором модели совпадают, и оно включает экспоненциальное как частный случай. Единственный случай, когда два класса совпадают, имеет место при распределении *Вейбулла*, для которого функция выживания принимает вид

$$S(t) = \exp\{-(\lambda t)^p\},$$

а функция риска – вид

$$\lambda(t) = p\lambda(\lambda t)^{p-1},$$

для некоторых параметров  $\lambda > 0$  и  $p > 0$ . При  $p = 1$  данная модель сводится к экспоненциальной и имеет постоянный во времени риск. При  $p > 1$  риск увеличивается со временем. При  $p < 1$  риск уменьшается со временем. Действительно, логарифмируя выражение для функции риска, получаем, что логарифм вейбулловского риска – линейная функция логарифма времени с коэффициентом наклона  $p - 1$ .

Если взять вейбулловский риск в качестве базового, а затем умножить его на константу  $\gamma$  в рамках модели пропорциональных рисков, снова получим распределение Вейбулла, так что данный класс замкнут относительно пропорциональности рисков. Если взять вейбулловскую функцию выживания в качестве базовой и затем ускорить протекание времени в рамках модели ускоренной жизни, разделив время на константу  $\gamma$ , снова получим распределение Вейбулла, так что этот класс замкнут относительно ускорения времени.

Другие подробности об этом распределении можно найти в Cox & Oakes (1984) или Kalbfleish & Prentice (1980), где доказана эквивалентность двух вейбулловских моделей.

### 4.4 Меняющиеся во времени регрессоры

До сих пор в явном виде рассматривались только регрессоры, неизменные во времени. Однако локальная природа модели пропорциональных рисков позволяет легко обобщить ее на случай меняющихся во времени регрессоров. Рассмотрим несколько примеров.

Предположим, производится анализ временных интервалов между родами, и изучаются промежутки между последовательными родами. Один из возможных регрессоров – уровень образования матери, который в большинстве случаев можно считать неизменным во времени.

Предположим теперь, что мы хотим добавить в качестве регрессора индикатор кормления грудью того ребенка, с которого начинается отсчет интервала. Предполагая, что ребенка в принципе кормят грудью, эта переменная будет принимать значение единица («да»), начиная

с родов и до окончания грудного кормления, когда переменная сменит значение на ноль («нет»). Это простой пример ситуации, когда регрессор может менять значение лишь один раз.

Более сложный анализ возникает при включении показателя частоты грудного кормления за сутки. Эта переменная может менять значение ежедневно. К примеру, последовательность значений для одной женщины могла бы иметь вид 4, 6, 5, 6, 5, 4, ...

Пусть  $\mathbf{x}_i(t)$  обозначает значение вектора регрессоров для индивида  $i$  в момент времени  $t$ . Тогда модель пропорциональных рисков можно обобщить следующим образом:

$$\lambda_i(t, \mathbf{x}_i(t)) = \lambda_0(t) \exp\{\mathbf{x}_i(t)' \boldsymbol{\beta}\}. \quad (12)$$

Теперь нет четкого разделения эффектов времени и регрессоров, и иногда может быть сложно идентифицировать эффекты регрессоров, которые сильно коррелируют со временем. Если, например, все дети были отняты от груди в возрасте около 6 месяцев, будет сложно отделить эффект кормления грудью от общих временных эффектов, не имея дополнительной информации. И все же в таких случаях может быть предпочтительным использование меняющегося во времени регрессора как более разумного предиктора риска, чем просто количество прошедшего времени.

Подсчет функций выживания при наличии меняющихся во времени регрессоров немного более затруднителен, поскольку требуется специфицировать траекторию для каждой переменной. В примере с промежутками между родами можно подсчитать функцию выживания для женщин, которые кормят грудью в течение шести месяцев, а затем прекращают. Это реализуется путем использования функции риска, соответствующей  $x(t) = 0$  для месяцев от 0 до 6, а затем риска, соответствующего  $x(t) = 1$ , для месяцев, начиная с шестого. К сожалению, теряется простота уравнения (11): больше нельзя просто возвести в степень базовую функцию выживания.

Меняющиеся во времени регрессоры можно также ввести в контексте моделей ускоренной жизни, но это не так просто и редко осуществляется в практических приложениях. См. подробности в Cox & Oakes (1984, стр. 66).

#### 4.5 Зависящие от времени коэффициенты

Модель также можно обобщить на случай *коэффициентов*, которые меняются во времени и, таким образом, не являются пропорциональными. Вполне возможно, например, что определенные социальные характеристики могут иметь сильное воздействие на риск смертности для детей сразу после рождения, но иметь относительно низкое воздействие позднее. Чтобы учесть модели подобного вида, можно записать

$$\lambda_i(t, \mathbf{x}_i) = \lambda_0(t) \exp\{\mathbf{x}_i' \boldsymbol{\beta}(t)\},$$

где параметр  $\boldsymbol{\beta}(t)$  теперь является функцией от времени.

Эта модель обладает очень высокой степенью общности. Например, в случае с двумя выборками модель можно записать в виде

$$\lambda_i(t|x) = \begin{cases} \lambda_0(t) & \text{если } x = 0, \\ \lambda_0(t)e^{\beta(t)} & \text{если } x = 1, \end{cases}$$

что, в сущности, позволяет иметь две произвольные функции риска, по одной для каждой группы. Таким образом, это весьма насыщенная модель.

Обычно форму зависимости коэффициентов от времени необходимо специфицировать параметрически, чтобы было возможно идентифицировать модель и оценить параметры. Очевидными кандидатами являются полиномы от времени, когда  $\beta(t)$  – линейная или квадратичная функция от времени. Cox & Oakes (1984, стр. 76) показывают, как можно использовать быстро затухающие экспоненты для моделирования переменных коэффициентов.

Заметим, что вновь потеряно простое разделение эффектов регрессоров и времени. Подсчет функции выживания в этой модели снова несколько осложняется тем фактом, что коэффициенты теперь зависят от времени, так что они не выносятся за знак интеграла. Простое уравнение (11) не имеет место.

#### 4.6 Общая модель риска

Изложенные расширения модели на случай меняющихся во времени регрессоров и зависящих от времени коэффициентов можно объединить в наиболее общую версию модели риска:

$$\lambda_i(t, \mathbf{x}_i(t)) = \lambda_0(t) \exp\{\mathbf{x}_i(t)' \boldsymbol{\beta}(t)\},$$

где  $\mathbf{x}_i(t)$  – вектор меняющихся во времени регрессоров, представляющих собой характеристики индивида  $i$  в момент времени  $t$ , а  $\boldsymbol{\beta}(t)$  – вектор зависящих от времени коэффициентов, представляющих собой эффект, который эти характеристики оказывают в момент времени  $t$ .

Пример с грудным кормлением и его влиянием на длину промежутков между последовательными родами является хорошей иллюстрацией, отражающей оба эффекта. Статус грудного кормления сам по себе является меняющимся во времени регрессором  $x(t)$ , принимающим значение единица, если женщина кормит грудью ребенка через  $t$  месяцев после родов. Известно, что воздействие, которое грудное кормление может иметь на задержку овуляции, а следовательно, на снижение риска беременности, быстро падает со временем, так что, возможно, его следует моделировать как зависящий от времени эффект  $\beta(t)$ . Опять же, дальнейшие действия требуют спецификации функциональной формы зависимости от времени.

#### 4.7 Подгонка модели

Существует, в сущности, три подхода к подгонке моделей выживаемости:

- Первый и, наверное, наиболее простой – это *параметрический* подход, когда предполагается определенная функциональная форма для базового риска  $\lambda_0(t)$ . Примерами являются модели, основанные на экспоненциальном, вейбулловском, гамма и обобщенном F распределениях.
- Второй подход можно назвать гибким или *полупараметрическим*, когда делаются довольно слабые предположения о базовом риске  $\lambda_0(t)$ . В частности, можно разбить время на достаточно малые интервалы и предположить, что базовый риск постоянен внутри каждого интервала, что приведет к кусочно-экспоненциальной модели.
- Третий подход – *непараметрический*, при котором регрессионные коэффициенты  $\boldsymbol{\beta}$  оцениваются без какой-либо спецификации функции базового риска  $\lambda_0(t)$ . Этот подход основан на функции частного правдоподобия, предложенной в статье Cox (1972).

Подробное обсуждение этих подходов выходит далеко за рамки настоящего эссе. Остановимся подробнее на промежуточном, полупараметрическом подходе, поскольку (а) он достаточно гибкий, чтобы быть полезным инструментом с широким спектром применения, и (б) он тесно связан с моделью пуассоновской регрессии.

### 5 Кусочно-экспоненциальная модель

Рассмотрим подгонку модели пропорциональных рисков в обычном виде

$$\lambda_i(t|\mathbf{x}_i) = \lambda_0(t) \exp\{\mathbf{x}_i' \boldsymbol{\beta}\} \quad (13)$$

при достаточно слабых предположениях о базовом риске  $\lambda_0(t)$ .

### 5.1 Кусочно-постоянный риск

Рассмотрим разбиение времени на  $J$  интервалов с точками разбиения  $0 = \tau_0 < \tau_1 < \dots < \tau_J = \infty$ . Определим  $j$ -й интервал  $[\tau_{j-1}, \tau_j)$ , продолжающийся от  $(j - 1)$ -й границы до  $j$ -й, включая начало и не включая конец.

Предположим далее, что базовый риск *постоянен* внутри каждого интервала, так что

$$\lambda_0(t) = \lambda_j \quad \text{для } t \text{ из } [\tau_{j-1}, \tau_j). \tag{14}$$

Таким образом, базовый риск  $\lambda_0(t)$  моделируется, используя  $J$  параметров  $\lambda_1, \dots, \lambda_J$ , каждый из которых представляет риск для эталонной группы (или индивида) в одном конкретном интервале. Поскольку риск предполагается кусочно-постоянным, соответствующую функцию выживания часто называют кусочно-экспоненциальной.

Конечно, разумный выбор точек разбиения должен позволить достаточно хорошо аппроксимировать почти любой базовый риск, если использовать близко расположенные границы интервалов, когда риск меняется быстро, и более широкие интервалы, когда риск меняется медленнее.

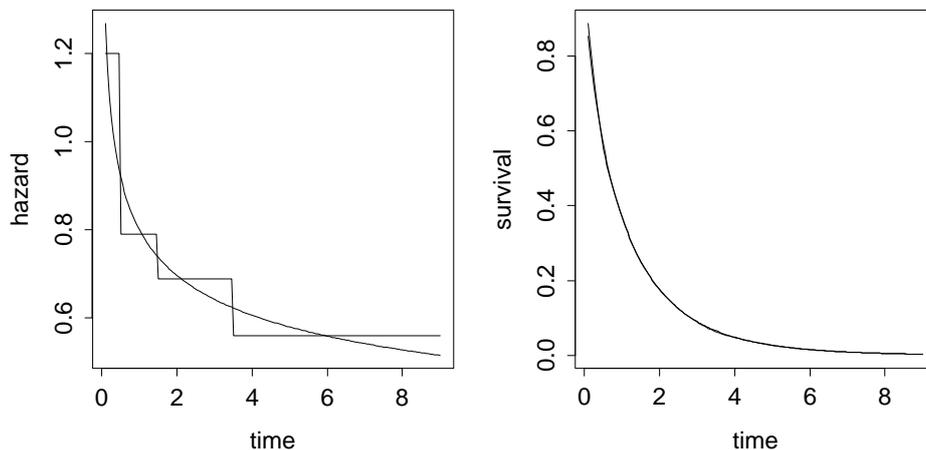


Рис. 1: Аппроксимация кривой выживания с помощью кусочно-постоянной функции риска

На Рис. 1 показано, как распределение Вейбулла с параметрами  $\lambda = 1$  и  $p = 0,8$  можно аппроксимировать с помощью кусочно-экспоненциального распределения с точками разбиения 0,5, 1,5 и 3,5. Диаграмма слева показывает, что кусочно-постоянный риск способен лишь в самых общих чертах повторить гладко убывающую вейбулловскую функцию риска, но при этом, как показано на диаграмме справа, соответствующие кривые выживания неразличимы.

### 5.2 Модель пропорциональных рисков

Теперь добавим регрессоры в контексте модели пропорциональных рисков из уравнения (13), предполагая, что базовый риск является кусочно-постоянным, как в уравнении (14). Запишем модель в виде

$$\lambda_{ij} = \lambda_j \exp\{\mathbf{x}'_i \boldsymbol{\beta}\}, \tag{15}$$

где  $\lambda_{ij}$  – риск, соответствующий индивиду  $i$  в интервале  $j$ ,  $\lambda_j$  – базовый риск в интервале  $j$ , а  $\exp\{\mathbf{x}'_i \boldsymbol{\beta}\}$  – относительный риск для индивида со значениями регрессоров  $\mathbf{x}_i$  по сравнению с базовым в каждый заданный момент времени.

Логарифмируя, получаем аддитивную логлинейную модель

$$\log \lambda_{ij} = \alpha_j + \mathbf{x}'_i \boldsymbol{\beta}, \tag{16}$$

где  $\alpha_j = \log \lambda_j$  – логарифм базового риска. Это стандартная логлинейная модель, в которой временные интервалы влияют на риск. Поскольку константа в явном виде не включена в модель, не требуется накладывать ограничения на  $\alpha_j$ . Если бы мы захотели включить константу, отражающую риск в первом интервале, мы положили бы  $\alpha_1 = 0$ , как обычно.

Модель можно расширить, включив меняющиеся во времени регрессоры и зависящие от времени коэффициенты, но обсуждение деталей мы отложим до изучения оценивания более простой модели пропорциональных рисков.

### 5.3 Эквивалентная пуассоновская модель

Holford (1980) и Laird & Oliver (1981) в независимо написанных статьях, опубликованных почти одновременно, заметили, что кусочная модель пропорциональных рисков из предыдущего раздела эквивалентна определенной модели пуассоновской регрессии. Сначала сформируем результат, а затем набросаем его доказательство.

Вспомним, что наблюдаются  $t_i$  – общее время, прожитое  $i$ -м индивидом, и  $d_i$  – индикатор смерти, который принимает значение единица, если индивид умер, и ноль в противном случае. Введем аналогичные показатели для каждого интервала времени, проживаемого индивидом  $i$ . Можно рассматривать этот процесс как создание ряда псевдонаблюдений, по одному для каждой комбинации индивида и интервала.

Сначала создадим показатели подверженности риску. Пусть  $t_{ij}$  обозначает время, прожитое  $i$ -м индивидом в  $j$ -м интервале, то есть между  $\tau_{j-1}$  и  $\tau_j$ . Если индивид прожил до конца интервала, то есть  $t_i > \tau_j$ , то время, прожитое в интервале, равно ширине интервала и  $t_{ij} = \tau_j - \tau_{j-1}$ . Если индивид умер или был цензурирован в этом интервале, то есть  $\tau_{j-1} < t_i < \tau_j$ , то время, прожитое в интервале, равно  $t_{ij} = t_i - \tau_{j-1}$ , разнице между общим прожитым временем и нижней границей интервала. Рассматриваются только те интервалы, в течение которых индивид жил, но очевидно, что время жизни в интервале равно нулю, если индивид умер до его начала и  $t_i < \tau_{j-1}$ .

Теперь создадим индикаторы смерти. Пусть  $d_{ij}$  принимает значение единица, если индивид  $i$  умер в интервале  $j$ , и ноль в противном случае. Пусть  $j(i)$  обозначает интервал, куда попадает  $t_i$ , то есть интервал, когда индивид  $i$  умер или был цензурирован. Мы используем функциональное обозначение, чтобы подчеркнуть, что этот интервал меняется от одного индивида к другому. Если  $t_i$  попадает, скажем, в интервал  $j(i)$ , то  $d_{ij}$  должно равняться нулю для всех  $j < j(i)$  (то есть для всех более ранних интервалов) и будет равно  $d_i$  для  $j = j(i)$  (то есть для интервала, когда индивид  $i$  наблюдался последний раз).

Теперь кусочно-экспоненциальную модель можно оценить, воспринимая индикаторы смерти  $d_{ij}$  как если бы это были независимые пуассоновские величины со средними значениями

$$\mu_{ij} = t_{ij} \lambda_{ij},$$

где  $t_{ij}$  – время подверженности риску, определенное выше, а  $\lambda_{ij}$  – риск для индивида  $i$  в интервале  $j$ . Логарифмируя это выражение и учитывая, что риски удовлетворяют модели пропорциональных рисков в уравнении (15), получаем

$$\log \mu_{ij} = \log t_{ij} + \alpha_j + \mathbf{x}'_i \boldsymbol{\beta},$$

где  $\alpha_j = \log \lambda_j$ , как и прежде.

Таким образом, кусочно-экспоненциальная модель пропорциональных рисков эквивалентна пуассоновской логлинейной модели для псевдонаблюдений, по одному для каждой комбинации индивида и интервала, где индикатор смерти – это зависимая переменная, а логарифм времени подверженности риску входит в уравнение как сдвиг.

Важно отметить, что не предполагается, что  $d_{ij}$  имеют независимые пуассоновские распределения, поскольку это, очевидно, не так. Если индивид  $i$  умер в интервале  $j(i)$ , он был

жив во все предыдущие интервалы  $j < j(i)$ , так что индикаторы, очевидно, не могут быть независимыми. Более того, каждый индикатор может принимать только значения единица или ноль, так что он не может иметь пуассоновское распределение, при котором значениям, большим единицы, соответствует ненулевая вероятность. Результат более тонкий. Совпадают не распределения, а функции правдоподобия. При данной реализации кусочно-экспоненциального процесса выживаемости можно найти такую реализацию набора независимых пуассоновских наблюдений, которая имеет такую же функцию правдоподобия, а значит, ведет к таким же оценкам и результатам тестирования гипотез.

Доказательство несложно. Вспомним из раздела 3.2, что вклад  $i$ -го индивида в логарифмическую функцию правдоподобия имеет общий вид

$$\log L_i = d_i \log \lambda_i(t_i) - \Lambda_i(t_i),$$

где  $\lambda_i(t)$  – риск, а  $\Lambda_i(t)$  – кумулятивный риск для  $i$ -го индивида в момент  $t$ . Пусть  $j(i)$  обозначает интервал, куда попадает  $t_i$ , как и раньше.

В кусочно-экспоненциальной модели первый член логарифмической функции правдоподобия можно записать как

$$d_i \log \lambda_i(t_i) = d_{ij(i)} \log \lambda_{ij(i)},$$

используя тот факт, что риск равен  $\lambda_{ij(i)}$ , когда  $t_i$  находится в интервале  $j(i)$ , а индикатор смерти  $d_i$  относится непосредственно к последнему интервалу жизни индивида  $i$  и поэтому равен  $d_{j(i)}$ .

Кумулятивный риск во втором члене выражения – это интеграл, который можно записать в виде суммы следующим образом:

$$\Lambda_i(t_i) = \int_0^{t_i} \lambda_i(t) dt = \sum_{j=1}^{j(i)} t_{ij} \lambda_{ij},$$

где  $t_{ij}$  – количество времени, проведенное индивидом  $i$  в интервале  $j$ . Чтобы увидеть это, заметим, что необходимо проинтегрировать риск от 0 до  $t_i$ . Разобьем этот интеграл на сумму интегралов, по одному для каждого интервала, где риск постоянный. Если индивид полностью проживает интервал, вкладом в интеграл будет риск  $\lambda_{ij}$ , умноженный на ширину интервала. Если индивид умирает или цензурируется внутри интервала, вкладом в интеграл будет риск  $\lambda_{ij}$ , умноженный на время, потраченное с начала интервала вплоть до смерти или момента цензурирования, которое равно  $t_i - \tau_{j-1}$ . Но это в точности совпадает с определением времени подверженности риску  $t_{ij}$ .

Одно небольшое отклонение от симметрии в результатах состоит в том, что риск ведет к *одному* члену при  $d_{ij(i)} \log \lambda_{ij(i)}$ , а кумулятивный риск – к  $j(i)$  членам, по одному на каждый интервал от  $j = 1$  до  $j(i)$ . Мы знаем, однако, что  $d_{ij} = 0$  для всех  $j < j(i)$ , так что можно добавить члены при  $d_{ij} \log \lambda_{ij}$  для всех предыдущих  $j$ ; пока  $d_{ij} = 0$ , они не дают никакого вклада в логарифмическую функцию правдоподобия. Этот прием позволяет записать вклад  $i$ -го индивида в логарифмическую функцию правдоподобия как сумму  $j(i)$  вкладов, по одному для каждого интервала, прожитого индивидом:

$$\log L_i = \sum_{j=1}^{j(i)} \{d_{ij} \log \lambda_{ij} - t_{ij} \lambda_{ij}\}.$$

Тот факт, что вклад индивида в логарифмическую функцию правдоподобия – это *сумма* нескольких членов (то есть вклад в функцию правдоподобия – это произведение нескольких членов), означает, что можно воспринимать каждый член как соответствующий независимому наблюдению.

Последний шаг – определить вклад каждого псевдонаблюдения, и здесь заметим, что он совпадает, если не считать константы, с функцией правдоподобия, когда  $d_{ij}$  имеет пуассоновское распределение со средним  $\mu_{ij} = t_{ij}\lambda_{ij}$ . Чтобы убедиться в этом, запишем пуассоновскую логарифмическую функцию правдоподобия в виде

$$\log L_{ij} = d_{ij} \log \mu_{ij} - \mu_{ij} = d_{ij} \log(t_{ij}\lambda_{ij}) - t_{ij}\lambda_{ij}.$$

Это выражение совпадает с логарифмической функцией правдоподобия, полученной выше, за исключением члена  $d_{ij} \log(t_{ij})$ , но это константа, зависящая от данных, но не от параметров, так что ее можно игнорировать с точки зрения оценивания. Доказательство окончено.

Данный результат обобщает наблюдение, сделанное в конце раздела 3.2, о взаимосвязи между функцией правдоподобия для цензурированных экспоненциальных данных и пуассоновской функцией правдоподобия. Расширение состоит в том, что вместо одного «пуассоновского» индикатора смерти для каждого индивида мы имеем по одному подобному индикатору для каждого интервала, прожитого каждым индивидом.

Создание псевдонаблюдений может значительно увеличить размер набора данных, вероятно настолько, что анализ станет невозможным. Заметим, однако, что количество различных комбинаций регрессоров может быть мало, даже если общее число псевдонаблюдений велико. В этом случае можно группировать наблюдения, складывая показатели времени подверженности риску и индикаторы смерти. В этой более общей модели можно определить  $d_{ij}$  как число смертей, а  $t_{ij}$  – как общее время подверженности риску для индивидов с характеристиками  $\mathbf{x}_i$  в интервале  $j$ . Как всегда в агрегированных пуассоновских моделях, оценки, стандартные ошибки и тесты отношения правдоподобия будут в точности такими же, как для индивидуальных данных. Конечно, остатки моделей будут разными, представляя собой точность подгонки агрегированных, а не индивидуальных данных, но это можно рассматривать как низкую цену по сравнению с удобством работы с небольшим количеством объектов.

#### 5.4 Меняющиеся во времени регрессоры

Из предыдущего анализа должно быть очевидно, что можно легко включить в модель меняющиеся во времени регрессоры, если они меняют значение только на границах интервалов. При создании псевдонаблюдений, требуемых для формирования пуассоновской логарифмической функции правдоподобия, обычно повторяют вектор регрессоров  $\mathbf{x}_i$ , создавая копии  $\mathbf{x}_{ij}$ , по одной для каждого интервала. Тем не менее, ничто в нашем анализе не требует, чтобы эти векторы были одинаковыми. Следовательно, можно переопределить вектор  $\mathbf{x}_{ij}$  как представляющий значения регрессоров для индивида  $i$  в интервале  $j$ , и продолжать как обычно, переписав модель в виде

$$\log \lambda_{ij} = \alpha_j + \mathbf{x}'_{ij}\boldsymbol{\beta}.$$

Требование того, чтобы регрессоры сменяли значение только на границах интервала, может показаться ограничительным, но в действительности модель более гибкая, чем кажется на первый взгляд, поскольку всегда можно снова разбить псевдонаблюдения. Например, если бы мы хотели ввести изменение регрессоров для индивида  $i$  в середине интервала  $j$ , можно было бы разбить одно псевдонаблюдение на два, первое со старыми и второе с новыми значениями регрессоров. Каждая половина получает свою собственную меру времени подверженности риску и индикатор смерти, но обе будут помечены как принадлежащие одному интервалу, так что им будет соответствовать один и тот же базовый риск. Все шаги приведенного выше доказательства по-прежнему верны.

Конечно, дальнейшее разбиение наблюдений увеличивает размер набора данных, и всегда будут существовать практические ограничения на то, как долго можно следовать этому

подходу, даже при использовании сгруппированных данных. Альтернативой является использование более простых индикаторов, таких как среднее значение регрессора в интервале, возможно лагированное, чтобы избежать прогнозирования текущего риска с помощью будущих значений регрессоров.

### 5.5 Зависящие от времени коэффициенты

Оказывается, что кусочно-экспоненциальная модель позволяет легко включить в анализ непропорциональные риски или зависящие от времени коэффициенты, снова предполагая, что эти коэффициенты меняются лишь на границах интервалов.

В качестве иллюстрации предположим, что есть один регрессор, принимающий значение  $x_{ij}$  для индивида  $i$  в интервале  $j$ . Предположим далее, что этот регрессор является фиктивной переменной, так что его возможные значения – единица и ноль. На данный момент неважно, является ли значение постоянным для индивида или меняется от интервала к интервалу.

В модели пропорциональных рисков мы бы записали

$$\log \lambda_{ij} = \alpha_j + \beta x_{ij},$$

где  $\beta$  отражает эффект регрессора на логарифм риска в каждый данный момент времени. Беря экспоненту, получаем, что риск при  $x = 1$  равен  $\exp\{\beta\}$ , помноженному на риск при  $x = 0$ , и этот эффект один и тот же во все моменты времени. Это простая аддитивная модель относительно времени и имеющегося регрессора.

Чтобы позволить коэффициенту зависеть от времени, запишем

$$\log \lambda_{ij} = \alpha_j + \beta_j x_{ij},$$

где  $\beta_j$  отражает эффект регрессора на риск в интервале  $j$ . Беря экспоненту, получаем, что риск в интервале  $j$  при  $x = 1$  равен  $\exp\{\beta_j\}$ , помноженному на риск в интервале  $j$  при  $x = 0$ , так что эффект может меняться от интервала к интервалу. Поскольку воздействие регрессора зависит от интервала, получаем своего рода взаимодействие регрессора и времени, что становится более очевидным, если записать модель в виде

$$\log \lambda_{ij} = \alpha_j + \beta x_{ij} + (\alpha\beta)_j x_{ij}.$$

Эти модели напоминают модели ковариационного анализа. Здесь  $\alpha$  играет роль константы, а  $\beta$  – роль коэффициента наклона. Модель пропорциональных рисков имеет различные константы и один и тот же коэффициент наклона, так что она аналогична модели параллельных прямых. Модель с зависящими от времени коэффициентами имеет различные константы и различные коэффициенты наклона и является аналогичной модели с взаимодействием.

Итак, можно учесть непропорциональность рисков просто путем введения взаимодействий со временем. Конечно, можно также тестировать предположение о пропорциональности рисков, проверяя значимость взаимодействия со временем. Теперь мы готовы рассмотреть пример.

## 6 Младенческая и детская смертность в Колумбии

Для иллюстрации применения кусочно-экспоненциальных моделей выживаемости возьмем данные о младенческой и детской смертности в Колумбии из работы Сомосы (Somosa, 1980). Данные собраны в 1976 году в ходе опроса, проведенного в рамках Всемирного обследования рождаемости. Выборка состояла из женщин в возрасте от 15 до 49 лет. Анкета содержала историю материнства, включая пол, дату рождения и (если применимо на момент проведения опроса) возраст смерти для каждого ребенка опрашиваемой женщины.

### 6.1 Подсчет событий и подверженности риску

Как часто случается с данными о выживаемости, большая часть работы уходит на их подготовку для анализа. В рассматриваемом случае мы начали с таблиц, приведенных в статье Сомосы, в которых содержатся данные о выживших детях, сгруппированные по текущему возрасту, и умерших детях, сгруппированные по возрасту смерти. В обеих таблицах возраст указан в соответствии с разбиением возрастов из таблицы 1, использующим короткие интервалы времени в начале жизни, когда риск смерти высок, но быстро падает, и широкие интервалы в более позднем возрасте. Используя эти два бита информации, мы подсчитали количество смертей и время подверженности риску по возрастным группам, предполагая, что дети, которые умерли или были цензурированы внутри интервала, проживали в среднем половину длины интервала.

Таблица 1: Младенческие и детские смерти и подверженность риску по возрасту ребенка и когорте рождения, Колумбия, 1976.

Точный возраст	Когорта рождения					
	1941–59		1960–67		1968–76	
	число смертей	подвер-сть риску	число смертей	подвер-сть риску	число смертей	подвер-сть риску
0–1 мес	168	278,4	197	403,2	195	495,3
1–3 мес	48	538,8	48	786,0	55	956,7
3–6 мес	63	794,4	62	1165,3	58	1381,4
6–12 мес	89	1550,8	81	2294,8	85	2604,5
1–2 года	102	3006,0	97	4500,5	87	4618,5
2–5 лет	81	8743,5	103	13201,5	70	9814,5
5–10 лет	40	14270,0	39	19525,0	10	5802,5

В таблице 1 показаны результаты этих подсчетов в терминах числа смертей и общего числа человеко-лет подверженности риску от рождения и до десятилетнего возраста по категориям возраста ребенка для трех групп детей (или когорт), рожденных в 1941–59, 1960–67 и 1968–76 гг. Цель анализа – оценить величину ожидаемого снижения младенческой и детской смертности в этих когортах и изучить, снижалась ли смертность одинаково быстро во всех возрастах или быстрее в определенных возрастных группах.

### 6.2 Подгонка пуассоновских моделей

Пусть  $y_{ij}$  обозначает число смертей для когорты  $i$  ( $i = 1, 2, 3$ ) в возрастной группе  $j$  ( $j = 1, 2, \dots, 7$ ). В свете результатов предыдущего раздела можно рассматривать  $y_{ij}$  как реализации пуассоновских случайных величин со средними  $\mu_{ij}$ , равными

$$\mu_{ij} = \lambda_{ij} t_{ij},$$

где  $\lambda_{ij}$  – риск, а  $t_{ij}$  – общее время подверженности риску для группы  $i$  в возрасте  $j$ . Иными словами, ожидаемое число смертей равно произведению коэффициента смертности на время подверженности риску.

Предостережение по поводу единиц измерения: риск должен измеряться в тех же самых единицах времени, которые использовались для измерения подверженности риску. В нашем примере время измеряется в годах, и следовательно,  $\lambda_{ij}$  представляет собой риск на человеко-год подверженности. Если бы время измерялось в месяцах,  $\lambda_{ij}$  представляло бы собой риск на человеко-месяц подверженности, и было бы в точности равно одной двенадцатой от размера риска на человеко-год.

Для моделирования рисков будем использовать связующее лог-преобразование, так что линейный предиктор принимает вид

$$\eta_{ij} = \log \mu_{ij} = \log \lambda_{ij} + \log t_{ij},$$

то есть является суммой двух частей,  $\log t_{ij}$ , *сдвига*, или известной части линейного предиктора, и  $\log \lambda_{ij}$ , логарифма риска.

Наконец, рассмотрим логлинейную модель для риска в обычном виде

$$\log \lambda_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta},$$

где  $\mathbf{x}_{ij}$  – вектор регрессоров. В случае, если читатель задумался о том, что произошло с базовым риском, поясним, что он содержится внутри вектора параметров  $\boldsymbol{\beta}$ . Вектор регрессоров  $\mathbf{x}_{ij}$  может включать константу, набор фиктивных переменных, представляющих возрастные группы (например, форму риска в зависимости от возраста), набор фиктивных переменных, представляющих когорты рождения (то есть отвечающих за изменение риска во времени) и даже набор мультипликативных фиктивных переменных, представляющих произведения возрастов и когорт рождения (эффекты взаимодействия).

Таблица 2: Сумма квадратов остатков (СКО) для различных моделей, оцененных по данным о младенческой и детской смертности в Колумбии

Модель	Название	$\log \lambda_{ij}$	СКО	Степени свободы
$\phi$	Нулевая	$\eta$	4239,8	20
$A$	Возраст	$\eta + \alpha_i$	72,7	14
$C$	Когорта	$\eta + \beta_j$	4190,7	18
$A + C$	Аддитивная	$\eta + \alpha_i + \beta_j$	6,2	12
$AC$	Насыщенная	$\eta + \alpha_i + \beta_j + (\alpha\beta)_{ij}$	0	0

В таблице 2 показаны суммы квадратов отклонений для пяти возможных моделей, включая нулевую модель, две однофакторные модели, двухфакторную аддитивную модель и двухфакторную модель с эффектом взаимодействия, которая является насыщенной для этих данных.

### 6.3 Эквивалентные модели выживаемости

Нулевая модель предполагает, что риск постоянен с рождения до десятилетнего возраста и одинаков для всех когорт. Таким образом, она соответствует *экспоненциальной модели выживаемости без регрессоров*. Очевидно, такая модель не подходит под данные, сумма квадратов остатков, равная 4239,8 при 20 степенях свободы – это астрономическая величина. ММП-оценка  $\eta$  равна  $-3,996$  со стандартной ошибкой  $0,0237$ . Беря экспоненту, получаем оценку риска, равную  $0,0184$ . То есть ожидается около 18 смертей на тысячу человеко-лет подверженности риску. Можно проверить, что  $0,0184$  – это просто отношение общего числа смертей к общему времени подверженности риску. Умножая  $0,0184$  на подверженность риску в каждой ячейке таблицы, получаем ожидаемое число смертей. Сумма квадратов остатков, приведенная выше, – это просто удвоенная сумма наблюдаемых смертей, помноженных на логарифм отношения наблюдаемых смертей к ожидаемым.

Модель с возрастом позволяет риску меняться от одной возрастной группы к другой, но предполагает, что риск в каждом возрасте одинаков для всех когорт. Она эквивалентна, таким образом, *кусочно-экспоненциальной модели выживаемости без регрессоров*. Снижение СКО по сравнению с нулевой моделью составляет  $4167,1$  при 6 степенях свободы, и очень значимо. Риск смерти значительно варьируется с возрастом во время нескольких первых

месяцев жизни. Иными словами, риск, очевидно, не является постоянным. Заметим, что при СКО 72,7 и 14 степенях свободы эта модель не подходит под данные. Значит, предпосылка о том, что все когорты подвержены одинаковому риску, не кажется надежной.

В таблице 3 показаны оценки параметров для однофакторных моделей  $A$  и  $C$  и для аддитивной модели  $A+C$  в формате, напоминающем множественный классификационный анализ. Хотя модель  $A$  не подходит под данные, будет поучительным кратко обсудить оценки параметров. Константа, указанная в скобках, соответствует риску  $\exp\{-0,7427\} = 0,4758$ , что составляет приблизительно половину смерти на человеко-год подверженности риску в первый месяц жизни. Оценка для возраста от 1 до 3 месяцев соответствует мультипликативному эффекту  $\exp\{-1,973\} = 0,1391$ , что составляет 86%-ное снижение риска после выживания в первый месяц жизни. Этот понижающий тренд продолжается до возраста 5–10 лет, когда мультипликативный эффект равен  $\exp\{-5,355\} = 0,0047$ , указывая на то, что риск в этих возрастах составляет лишь полпроцента того, каким он был в первый месяц жизни. Можно убедиться в том, что ММП-оценки возрастных эффектов можно подсчитать напрямую по общему числу смертей и общему времени подверженности риску в каждой возрастной группе. Можете ли Вы вычислить СКО вручную?

Рассмотрим теперь модель, включающую только когорту рождения, в которой предполагается, что риск постоянен с момента рождения до десяти лет, но меняется от одной когорты к другой. Эта модель эквивалентна *трем экспоненциальным моделям выживаемости*, по одной для каждой когорты рождения. Как и следовало ожидать, она безнадежно неадекватна, с СКО в несколько тысяч, поскольку не принимает во внимание значительные возрастные эффекты, которые только что обсуждались. Интересно, тем не менее, взглянуть на оценки параметров в таблице 3. В первом приближении, общая норма смертности для старшей когорты равнялась  $\exp\{-3,899\} = 0,0203$ , или около 20 смертей на тысячу человеко-лет подверженности риску. Мультипликативный эффект для когорты, рожденной в 1960–1967 гг., равен  $\exp\{-0,3020\} = 0,7393$ , что соответствует 26%-ному снижению общей смертности. Однако мультипликативный эффект для самой молодой когорты равен  $\exp\{0,0742\} = 1,077$ , что соответствует 8%-ному *увеличению* общей смертности. Можете ли Вы найти объяснение этой явной аномалии? Мы дадим ответ на этот вопрос после рассмотрения следующей модели.

Таблица 3: Оценки параметров для моделей  $A$ ,  $C$  и  $A + C$  по данным о младенческой и детской смертности в Колумбии

Фактор	Категория	Общий эффект	Чистый эффект
База			-0,4485
Возраст	0–1 мес	(-0,7427)	-
	1–3 мес	-1,973	-1,973
	3–6 мес	-2,162	-2,163
	6–12 мес	-2,488	-2,492
	1–2 года	-3,004	-3,014
	2–5 лет	-4,086	-4,115
	5–10 лет	-5,355	-5,436
Когорта	1941–59	(-3,899)	-
	1960–67	-0,3020	-0,3243
	1968–76	0,0742	-0,4784

Рассмотрим теперь аддитивную модель с эффектами как возраста, так и когорты, в которой риск может меняться с возрастом и может быть различным для разных когорт, но эффект возраста (когорты) предполагается одинаковым для каждой когорты (возраста).

Эта модель эквивалентна *модели пропорциональных рисков*, в которой предполагается общая форма риска для каждого возраста, а когорта пропорционально влияет на риск во всех возрастах. Сравнивая модель пропорциональных рисков с моделью *A*, отметим снижение СКО на 66,5 при потере двух степеней свободы, что составляет очень значимый эффект. Это свидетельствует о сильных когортных эффектах, очищенных от влияния возраста. С другой стороны, полученное СКО 6,2 при 12 степенях свободы, очевидно, не является значимым, что означает, что модель пропорциональных рисков адекватно описывает смертность в Колумбии в зависимости от возраста и когорты рождения. Иными словами, предположение о пропорциональности рисков достаточно разумно, откуда следует, что снижение смертности в Колумбии было одним и тем же во всех возрастах.

Рассмотрим оценки параметров в самом правом столбце таблицы 3. Константа – это базовый риск в возрасте 0–1 месяцев для самой первой когорты, то есть рожденных в 1941–59 годы. Параметры возраста, представляющие базовый риск, практически не меняются по сравнению с моделью, учитывающей только возрастной эффект, и соответствуют сильному снижению смертности с момента рождения до десятилетнего возраста, причем половина снижения приходится на первый год жизни. Когортные эффекты, подправленные на возраст, дают более разумную картину снижения смертности во времени. Мультипликативные эффекты для когорт, рожденных в 1960–1967 гг. и 1968–1976 гг., равны  $\exp\{-0,3243\} = 0,7233$  и  $\exp\{-0,4784\} = 0,6120$ , что соответствует снижению смертности на 28 и 38 %% в каждом возрасте, по сравнению с когортой, рожденной в 1941–59 годы. Это удивительное снижение младенческой и детской смертности, которое было одним и тем же для всех возрастов. Иными словами, смертность новорожденных, младенцев и тех, кто только начинает ходить, снизилась примерно на 38 процентов среди этих когорт.

Тот факт, что общий эффект для младшей когорты был положителен, а чистый эффект отрицателен и значителен, можно объяснить следующим образом. Поскольку обследование проводилось в 1976 году, дети, родившиеся между 1968 и 1976 годами, были подвержены по большей части смертности в молодых возрастах, когда нормы смертности значительно выше, чем в старших возрастах. Например, ребенок, родившийся в 1975 году, был подвержен только риску смерти в первый год жизни. Общий эффект игнорирует этот факт и, таким образом, переоценивает смертность в этой группе в возрасте от 0 до 10. Чистый эффект делает необходимую поправку на повышенный риск в молодых возрастах, в сущности сравнивая смертность в данной когорте со смертностью в более ранних когортах, когда они были в том же возрасте, а следовательно, способен раскрыть истинное снижение.

Заключительное предостережение по поводу интерпретации: данные получены ретроспективно со слов матерей, которым было от 15 до 49 лет в момент проведения опроса. Эти женщины представляют собой репрезентативную выборку как матерей, так и рождений для недавних периодов, но несколько смещенную для более ранних периодов. Выборка исключает матерей, умерших до проведения опроса, а также тех, кто был старше в момент рождения ребенка. Например, рождения в 1976, 1966 и 1956 гг. относятся к матерям, которым на момент рождения ребенка было меньше 50, 40 и 30 лет, соответственно. При более аккуратном анализе данных следовало бы включить возраст матери при рождении ребенка в качестве дополнительной контрольной переменной.

#### 6.4 Оценка вероятностей выживания

До сих пор наше внимание было сконцентрировано на риске или смертности, но, конечно, как только подсчитан риск, легко найти кумулятивный риск, а значит, вероятности выживания. В таблице 4 представлены результаты такого упражнения, с использованием оценок параметров из модели пропорциональных рисков в таблице 3.

Рассмотрим сначала базовую группу, а именно когорту детей, рожденных до 1960 года. Для получения лог-риска для каждой возрастной группы надо сложить константу и возрастной

Таблица 4: Подсчет вероятностей выживания для трех когорт на основе модели пропорциональных рисков

Возрастная группа (1)	Ширина (2)	База			Выживаемость для когорты		
		лог-риск (3)	риск (4)	кум.риск (5)	<1960 (6)	1960–67 (7)	1968–76 (8)
0–1 мес	1/12	–0,4485	0,6386	0,0532	0,9482	0,9623	0,9676
1–3 мес	2/12	–2,4215	0,0888	0,0680	0,9342	0,9520	0,9587
3–6 мес	3/12	–2,6115	0,0734	0,0864	0,9173	0,9395	0,9479
6–12 мес	1/2	–2,9405	0,0528	0,1128	0,8933	0,9217	0,9325
1–2 года	1	–3,4625	0,0314	0,1441	0,8658	0,9010	0,9145
2–5 лет	3	–4,5635	0,0104	0,1754	0,8391	0,8809	0,8970
5–10 лет	5	–5,8845	0,0028	0,1893	0,8275	0,8721	0,8893

эффект, например, лог-риск для возраста 1–3 мес. равен  $-0,4485 - 1,973 = -2,4215$ . Это дает числа в столбце (3) таблицы 3. Далее берем экспоненту для получения риска в столбце (4), например, риск для возраста 1–3 мес. равен  $\exp\{-2,4215\} = 0,0888$ . Далее подсчитываем кумулятивный риск, умножаем его на ширину интервала и суммируем по всем интервалам. На этом шаге необходимо выразить ширину интервала в тех же единицах, которые используются для подсчета подверженности риску, в данном случае в годах. Таким образом, кумулятивный риск в конце периода 1–3 мес. равен  $0,6386 \times 1/12 + 0,0888 \times 2/12 = 0,0680$ . Наконец, заменим знак и возьмем экспоненту для подсчета функции выживания. Например, базовая функция выживания в возрасте 3 месяцев равна  $\exp\{-0,0680\} = 0,9342$ .

Для подсчета значений функций выживания, показанных в столбцах (7) и (8), для других двух когорт, можно было бы умножить базовые риски на  $\exp\{-0,3242\}$  и  $\exp\{-0,4874\}$ , чтобы получить риски для когорт 1960–67 и 1968–76, соответственно, а затем повторить шаги, описанные выше, для получения функций выживания. Этот подход был бы необходим при наличии зависящих от времени коэффициентов, но в настоящем случае можно воспользоваться упрощением, которое дает модель пропорциональных рисков. А именно, функции выживания для двух младших когорт можно подсчитать как базовую функцию выживания, *возведенную в степень*, равную относительным рискам  $\exp\{-0,3242\}$  и  $\exp\{-0,4874\}$ , соответственно. Например, вероятность дожить до трех месяцев равна 0,9342 для базовой группы, и, соответственно, оказывается равной  $0,9342 \exp\{-0,3242\} = 0,9520$  для когорты, рожденной в 1960–1967 гг., и  $0,9342 \exp\{-0,4874\} = 0,9587$  для когорты 1968–1976 гг.

Заметим, что вероятность умереть во время первого года жизни уменьшилась со 106,7 из тысячи для детей, рожденных до 1960 г., до 78,3 из тысячи для детей, рожденных в 1960–1967 гг., и, наконец, до 67,5 из тысячи для наиболее молодой когорты. Результаты, представленные в терминах вероятностей, часто доступнее для широкой аудитории, чем результаты, представленные в терминах норм риска. (К сожалению, демографы обычно называют вероятность умереть в течение первого года жизни «нормой младенческой смертности». Это неверно, поскольку эта величина представляет собой вероятность, а не норму. В нашем примере норма значительно меняется на протяжении первого года жизни. Если вероятность умереть в течение первого года жизни равна, скажем,  $q$ , то средняя норма равна приблизительно  $-\log(1 - q)$ , что несильно отличается от  $q$  для малых значений  $q$ .)

Концентрируясь на событиях и подверженности риску, мы смогли объединить анализ младенческой и детской смертности в рамках одной модели и использовать всю доступную информацию. Альтернативный подход – сосредоточиться на младенческой смертности (когда смерть наступает на первом году жизни), и решать задачу цензурирования, рассматривая только детей, рожденных по крайней мере за год до обследования, для которых известно, дожили ли они до одного года. Затем можно было бы исследовать вероятность дожития до

возраста 1 год, используя обычные логит-модели. В качестве дополняющего анализа можно было бы взглянуть на дожитие от возраста 1 год до, скажем, 5 лет, работая с данными о детях, рожденных по крайней мере за 5 лет до обследования и доживших до одного года, а затем анализируя, доживают они или нет до пятилетнего возраста, снова используя логит-модель. Будучи простым, данный подход не полностью использует информацию, полагаясь только на полные (нецензурированные) данные. Cox & Oakes (1984) показывают, что этот так называемый подход с уменьшенной выборкой может давать несостоятельные результаты. Другой недостаток этого подхода в том, что он концентрируется на дожитии до ключевых возрастов, но не позволяет исследовать форму риска в промежуточный период.

## 7 Модели в дискретном времени

Обсудим кратко два расширения модели пропорциональных рисков для дискретного времени, начиная с определения функций риска и выживания в дискретном времени, а затем перейдем к моделям на основе логит- и дополнительного лог-лог-преобразований.

### 7.1 Дискретные функции риска и выживания

Пусть  $T$  – дискретная случайная величина, принимающая значения  $t_1 < t_2 < \dots$  с вероятностями

$$f(t_j) = f_j = \mathbb{P}\{T = t_j\}.$$

Определим функцию выживания в момент  $t_j$  как вероятность того, что время жизни  $T$  не меньше  $t_j$ :

$$S(t_j) = S_j = \mathbb{P}\{T \geq t_j\} = \sum_{k=j}^{\infty} f_k.$$

Далее определим риск в момент  $t_j$  как вероятность умереть в этот момент времени при условии, что индивид дожил до этого момента, то есть

$$\lambda(t_j) = \lambda_j = \mathbb{P}\{T = t_j | T \geq t_j\} = \frac{f_j}{S_j}. \tag{17}$$

Заметим, что в дискретном времени риск – это условная вероятность, а не норма. Тем не менее, общий результат, выражающий риск как отношение плотности к функции выживания, остается верным.

Следующий интересный результат для дискретного времени состоит в том, что функцию выживания в момент времени  $t_j$  можно записать в терминах риска за все предыдущие моменты времени,  $t_1, \dots, t_{j-1}$ :

$$S_j = (1 - \lambda_1)(1 - \lambda_2) \dots (1 - \lambda_{j-1}). \tag{18}$$

Иными словами, данный результат утверждает, что, чтобы дожить до момента  $t_j$ , необходимо сначала дожить до  $t_1$ , затем дожить до  $t_2$  при условии дожития до  $t_1$  и так далее, наконец, пережить  $t_{j-1}$  условно на дожитии до этого момента. Этот результат аналогичен результату, связывающему функцию выживания в непрерывном времени с интегральным или кумулятивным риском во все предыдущие моменты времени.

Примером процесса выживаемости, происходящего в дискретном времени, может быть время до зачатия, измеряемое в менструальных циклах. В этом случае возможные значения  $T$  – это все положительные целые числа,  $f_j$  – это вероятность зачатия во время  $j$ -го цикла,  $S_j$  – вероятность зачатия во время  $j$ -го цикла или позже, а  $\lambda_j$  – вероятность зачатия во время

$j$ -го цикла при условии, что оно не произошло ранее. Результат, связывающий функцию выживания и риск, утверждает, что, чтобы добраться до  $j$ -го цикла без зачатия, необходимо его отсутствие во время первого цикла, затем во время второго при условии неудавшегося зачатия во время первого цикла, и т.д., и, наконец, во время  $(j - 1)$ -го цикла при условии его отсутствия на более ранней стадии.

## 7.2 Дискретная функция выживания и логистическая регрессия

Сох (1972) предложил в качестве расширения модели пропорциональных рисков для дискретного времени работать с условными вероятностями смерти в каждый момент времени  $t_j$  при условии дожития до этого момента времени. В частности, он предложил такую модель:

$$\frac{\lambda(t_j|\mathbf{x}_i)}{1 - \lambda(t_j|\mathbf{x}_i)} = \frac{\lambda_0(t_j)}{1 - \lambda_0(t_j)} \exp\{\mathbf{x}'_i\boldsymbol{\beta}\},$$

где  $\lambda(t_j|\mathbf{x}_i)$  – риск в момент  $t_j$  для индивида со значениями регрессоров  $\mathbf{x}_i$ ,  $\lambda_0(t_j)$  – базовый риск в момент  $t_j$ , а  $\exp\{\mathbf{x}'_i\boldsymbol{\beta}\}$  – относительный риск, связанный со значениями регрессоров  $\mathbf{x}_i$ .

Логарифмируя, получаем модель относительно *логита* риска, или условной вероятности умереть в момент  $t_j$  при условии дожития до этого момента:

$$\text{logit } \lambda(t_j|\mathbf{x}_i) = \alpha_j + \mathbf{x}'_i\boldsymbol{\beta}, \quad (19)$$

где  $\alpha_j = \text{logit } \lambda_0(t_j)$  – логит базового риска, а  $\mathbf{x}'_i\boldsymbol{\beta}$  – эффект регрессоров на логит риска. Заметим, что, в сущности, время в модели является дискретным фактором, поскольку для каждого возможного времени смерти  $t_j$  присутствует один параметр  $\alpha_j$ . Интерпретация параметров  $\boldsymbol{\beta}$ , связанных с другими регрессорами, следует той же логике, что и при логистической регрессии.

В действительности есть дальнейшая аналогия с логистической регрессией: можно оценить модель пропорциональных рисков в дискретном времени с помощью оценки логистической регрессии для набора псевдонаблюдений, сгенерированных следующим образом. Предположим, что индивид  $i$  умирает или цензурируется в момент времени  $t_{j(i)}$ . Создадим индикаторы смерти  $d_{ij}$ , которые принимают значение единица, если индивид  $i$  умер в момент времени  $j$ , и ноль в противном случае, по одному для каждого момента времени от  $t_1$  до  $t_{j(i)}$ . С каждым из этих индикаторов свяжем копию вектора регрессоров  $\mathbf{x}_i$  и индекс  $j$ , обозначающий момент времени. Тогда модель пропорциональных рисков (19) можно оценить, воспринимая  $d_{ij}$  как независимые наблюдения Бернулли с вероятностью успеха, задаваемой риском  $\lambda_{ij}$  для индивида  $i$  в момент времени  $t_j$ .

В более общем плане можно сгруппировать псевдонаблюдения с одинаковыми значениями регрессоров. Пусть  $d_{ij}$  обозначает число смертей, а  $n_{ij}$  – общее число индивидов со значениями регрессоров  $\mathbf{x}_i$ , наблюдаемых в момент времени  $t_j$ . Тогда можно воспринимать  $d_{ij}$  как биномиальную величину с параметрами  $n_{ij}$  и  $\lambda_{ij}$ , где последняя удовлетворяет модели пропорциональных рисков.

Доказательство этого результаты следует тем же шагам, что и доказательство эквивалентности пуассоновской функции правдоподобия и функции правдоподобия для кусочно-экспоненциальных данных о выживаемости при неинформативном цензурировании в разделе 5.3, и основано на уравнении (18), в котором вероятность дожить до времени  $t_j$  записана как произведение условных рисков за все предыдущие моменты времени. Важно отметить, что не предполагается, что псевдонаблюдения независимы и имеют бернуллиевское или биномиальное распределение. Мы лишь отмечаем, что функция правдоподобия для модели выживаемости в дискретном времени при неинформативном цензурировании совпадает с биномиальной функцией правдоподобия, которую бы мы получили, воспринимая индикаторы смерти как независимые бернуллиевские или биномиальные случайные величины.

Меняющиеся во времени регрессоры и зависящие от времени коэффициенты можно легко включить в эту модель, следуя тем же шагам, что и ранее. В случае меняющихся во времени регрессоров заметим, что важны только значения регрессоров в дискретные моменты времени  $t_1 < t_2 < \dots$ . Зависящие от времени коэффициенты моделируются как взаимодействия между регрессорами и дискретным фактором (или набором фиктивных переменных), представляющим время.

### 7.3 Дискретная функция выживания и связующее дополнительное лог-лог-преобразование

Альтернативное расширение модели пропорциональных рисков для дискретного времени начинается с функции выживания, которую в рамках модели пропорциональных рисков можно записать в виде

$$S(t_j|\mathbf{x}_i) = S_0(t_j)^{\exp\{\mathbf{x}'_i\boldsymbol{\beta}\}},$$

где  $S(t_j|\mathbf{x}_i)$  – вероятность того, что индивид со значениями регрессоров  $\mathbf{x}_i$  доживет до момента  $t_j$ , а  $S_0(t_j)$  – базовая функция выживания. Вспомнив уравнение (18) для дискретной функции выживания, получаем похожее выражение для дополнения функции риска, а именно:

$$1 - \lambda(t_j|\mathbf{x}_i) = [1 - \lambda_0(t_j)]^{\exp\{\mathbf{x}'_i\boldsymbol{\beta}\}},$$

и, разрешая относительно риска для индивида  $i$  в момент времени  $t_j$ , получаем модель

$$\lambda(t_j|\mathbf{x}_i) = 1 - [1 - \lambda_0(t_j)]^{\exp\{\mathbf{x}'_i\boldsymbol{\beta}\}}.$$

Дополнительное лог-лог-преобразование делает правую часть линейной функцией параметров. Применяя это преобразование, получаем модель

$$\log(-\log(1 - \lambda(t_j|\mathbf{x}_i))) = \alpha_j + \mathbf{x}'_i\boldsymbol{\beta}, \tag{20}$$

где  $\alpha_j = \log(-\log(1 - \lambda_0(t_j)))$  – дополнительное лог-лог-преобразование базового риска.

Модель можно оценить по дискретным данным о выживаемости, генерируя псевдонаблюдения как ранее и подгоняя обобщенную линейную модель с биномиальной структурой ошибок и связующим дополнительным лог-лог-преобразованием. Иными словами, эквивалентность между биномиальной функцией правдоподобия и функцией правдоподобия для функции выживания в дискретном времени при неинформативном цензурировании сохраняется как для логит, так и для дополнительного лог-лог преобразований.

Интересно отметить, что эту модель можно получить, группируя время в модели пропорциональных рисков в непрерывном времени. Чтобы это увидеть, предположим, что время непрерывно, и нам интересна стандартная модель пропорциональных рисков

$$\lambda(t|\mathbf{x}) = \lambda_0(t) \exp\{\mathbf{x}'\boldsymbol{\beta}\}.$$

Предположим, однако, что время сгруппировано в интервалы с границами  $0 = \tau_0 < \tau_1 < \dots < \tau_J = \infty$ , и что все, что наблюдается, – это переживает ли индивид интервал или умирает в нем. Заметим, что эта конструкция накладывает некоторые ограничения на цензурирование. Если индивид цензурируется в некоторой точке внутри интервала, неизвестно, пережил бы он этот интервал или нет. Следовательно, необходимо цензурировать его в конце предыдущего интервала, являющегося последней точкой, для которой имеется полная информация. В отличие от кусочно-экспоненциальной модели в данном случае нельзя использовать информацию о подверженности риску в части интервала. С другой стороны, как оказывается, необязательно предполагать, что риск постоянный в каждом интервале.

Пусть  $\lambda_{ij}$  обозначает дискретный риск или условную вероятность того, что индивид  $i$  умрет в интервале  $j$  при условии того, что он был жив в начале интервала. Эта вероятность равна дополнению условной вероятности выживания в интервале при условии, что индивид был жив в начале интервала, и ее можно записать в виде

$$\begin{aligned}\lambda_{ij} &= 1 - \mathbb{P}\{T_i > \tau_j | T_i > \tau_{j-1}\} \\ &= 1 - \exp \left\{ - \int_{\tau_{j-1}}^{\tau_j} \lambda(t | \mathbf{x}_i) dt \right\} \\ &= 1 - \exp \left\{ - \int_{\tau_{j-1}}^{\tau_j} \lambda_0(t) dt \right\}^{\exp\{\mathbf{x}'_i \boldsymbol{\beta}\}} \\ &= 1 - (1 - \lambda_j)^{\exp\{\mathbf{x}'_i \boldsymbol{\beta}\}},\end{aligned}$$

где  $\lambda_j$  – базовая вероятность смерти в интервале  $j$  при условии дожития до начала этого интервала. Вторая строчка следует из уравнения (4), связывающего функцию выживания и интегральный риск, третья строчка следует из предпосылки о пропорциональности рисков, а последняя строчка определяет  $\lambda_j$ .

Как отмечено в Kalbfleish & Prentice (1980, стр. 37), «эта дискретная модель, таким образом, является единственной подходящей для сгруппированных данных из модели пропорциональных рисков в непрерывном времени». На практике, тем не менее, модель со связующим логит-преобразованием применяется гораздо чаще, чем модель с дополнительным лог-лог-преобразованием, вероятно потому, что логистическая регрессия лучше известна, чем обобщенные линейные модели со связующим дополнительным лог-лог-преобразованием, и поскольку программное обеспечение для первой из них более широко доступно, чем для второй. В действительности логит-модель часто используется в случаях, когда более адекватной была бы кусочно-экспоненциальная модель, возможно потому, что логистическая регрессия лучше известна, чем пуассоновская.

В заключение полезно было бы сделать некоторые предложения относительно выбора подхода к анализу выживаемости при использовании обобщенных линейных моделей:

- Если время на самом деле дискретно, разумно использовать модель в дискретном времени со связующим логит-преобразованием, которая имеет прямую интерпретацию в терминах условных вероятностей, и легко реализуется с помощью стандартного программного обеспечения для логистической регрессии.
- Если время непрерывно, но наблюдается в сгруппированной форме, то связующее дополнительное лог-лог-преобразование кажется более подходящим. В частности, результаты на основе дополнительного лог-лог-преобразования должны быть более устойчивы к выбору категорий, чем результаты на основе логит-преобразования. Тем не менее, в этом случае нельзя учесть частичную подверженность риску при дискретном времени, независимо от применяемого преобразования.
- Если время непрерывно и хочется предполагать, что риск постоянен внутри каждого интервала, то кусочно-экспоненциальный подход на основе пуассоновской функции правдоподобия предпочтителен. Этот подход достаточно устойчив к выбору категорий и уникален в плане возможности использования информации в случаях частичной подверженности риску.

Наконец, если время на самом деле непрерывно, и хочется оценить эффекты регрессоров, не делая каких-либо предположений о базовом риске, то метод частного правдоподобия из Cox (1972) является очень привлекательным.

## Литература

- Cox, D.R. (1972). Regression models and life tables (с обсуждением). *Journal of Royal Statistical Society, Series B* 34, 187–220.
- Cox, D.R. & D. Oakes (1984). *Analysis of Survival Data*. London: Chapman and Hall.
- Holford, T.R. (1980). The analysis of rates and survivorship using log-linear models. *Biometrics* 36, 299–306.
- Kalbfleisch, J.D. & R.L. Prentice (1980). *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- Laird, N. & D. Oliver (1981). Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of American Statistical Association* 74, 231–240.
- Somoza, J.L. (1980). Illustrative analysis: Infant and child mortality in Colombia. *World Fertility Survey Scientific Reports* 10. London: World Fertility Survey.

# Survival models

**Germán Rodríguez**

*Princeton University, Princeton, USA*

This essay is an introduction to survival models in the context of generalized linear models. We introduce the hazard and survival functions and describe the most common censoring mechanisms and the resulting likelihood function. We discuss the main approaches to modeling waiting times, including accelerated life and proportional hazard models, with extensions to time-varying covariates and time-dependent effects. We then focus on the piece-wise exponential survival model and note its equivalence with Poisson regression models. We illustrate this approach with an application to the analysis of infant and child mortality in Colombia using survey data. We conclude with a brief discussion of discrete time models and their equivalence with logistic regression.



# Полупараметрический анализ\*

Даниэль Макфадден†

Калифорнийский Университет, Беркли, США

Настоящее эссе – обзор двух сфер применения полупараметрической эконометрики: анализа цензурированных данных о продолжительности занятости и анализа данных о заявленной готовности платить за природные ресурсы.

## 1 Введение

Многие эконометрические задачи можно рассматривать как один из вариантов следующей модели. Имеется случайный вектор  $(Y, X) \in \mathbb{R}^k \times \mathbb{R}^m$ , такой, что  $X$  имеет (неизвестную) плотность распределения  $g(x)$ , а  $Y$  почти наверное характеризуется (неизвестной) функцией условной плотности  $f(y|x)$ . Также известно преобразование  $t(y, x)$  из  $\mathbb{R}^k \times \mathbb{R}^m$  в множество действительных чисел  $\mathbb{R}$ , и условное математическое ожидание этого преобразования,  $\theta(x) = \mathbb{E}[t(Y, x)|X = x]$ , является объектом эконометрического исследования. Примерами подобных преобразований могут быть: (1)  $t(y, x) \equiv y$ , когда  $\theta(x) = \mathbb{E}[Y|X = x]$  – математическое ожидание  $Y$  при условии  $X = x$ , или *функция регрессии*  $Y$  на  $x$ ; (2)  $t(y, x) = yy'$ , когда  $\theta(x) = \mathbb{E}[YY'|X = x]$  – матрица вторых условных моментов, а в комбинации с первым примером – условная дисперсия  $\mathbb{E}[YY'|X = x] - (\mathbb{E}[Y|X = x])(\mathbb{E}[Y|X = x])'$ ; и (3)  $t(y, x) = \mathbb{I}_A(y)$ , то есть индикатор-функция множества  $A$ , когда  $\theta(x)$  – вероятность события  $A$  при условии  $X = x$ . Примерами из экономических приложений могут быть вектор потребительского спроса  $Y$  и вектор дохода и цен  $x$ , или вектор чистого выпуска фирмы  $Y$  и вектор уровней постоянных затрат и цен на переменные факторы  $x$ .

Определим возмущение  $\varepsilon = \varepsilon(y, x) \equiv t(y, x) - \theta(x)$ . Тогда описанную выше постановку можно сформулировать в виде *обобщенной регрессионной модели*

$$t(y, x) = \theta(x) + \varepsilon,$$

где  $\mathbb{E}[\varepsilon|x] = 0$ . Эконометрические задачи, подходящие под эту модель, можно классифицировать как *полностью параметрические*, *полупараметрические* или *непараметрические*. Модель является полностью параметрической, если *априори* известно, что функция  $\theta$  и распределение ошибки  $\varepsilon$  принадлежат семействам с конечным числом параметров. Модель является непараметрической, если о функциональных формах  $\theta$  и  $\varepsilon$  ничего неизвестно, за исключением, возможно, некоторых свойств регулярности и формы, таких как непрерывная дифференцируемость или вогнутость. Модель является полупараметрической, если она содержит конечный вектор параметров, обычно представляющий первостепенный интерес, но части  $\theta$  и/или распределение  $\varepsilon$  не ограничены семействами с конечным числом параметров. Это определение полупараметрической модели в довольно широком смысле, и оно включает, например, модель линейной регрессии при условиях Гаусса–Маркова, когда распределение ошибок не ограничено параметрическим семейством, и только первые два момента параметризованы. Некоторые эконометристы предпочитают применять термин «полупараметрическая модель» в тех ситуациях, когда задачу можно охарактеризовать с помощью

\*Перевод Б. Гершмана и С. Анатольева. Цитировать как: Макфадден, Даниэль (2008) «Полупараметрический анализ», Квантиль, №5, стр. 29–40. Citation: McFadden, Daniel (2008) “Semiparametric analysis,” Quantile, No.5, pp. 29–40.

†Адрес: University of California, Berkeley, Department of Economics, 549 Evans Hall #3880, Berkeley, CA 94720-3880, USA. Электронная почта: [mcfadden@econ.berkeley.edu](mailto:mcfadden@econ.berkeley.edu)

конечномерного вектора параметров, являющегося объектом анализа, и бесконечномерного вектора шумовых параметров (который может, например, задавать неизвестную функцию), поскольку именно в таких случаях необходимы неклассические статистические методы.

Наиболее распространенный полупараметрический метод в эконометрике – это обыкновенный МНК, который оценивает параметры модели линейной регрессии, не требуя, чтобы распределение ошибок принадлежало семейству с конечным числом параметров. Современная литература по эконометрической теории расширила полупараметрические методы на различные нелинейные модели. Четыре крупнейшие пересекающиеся области их применения – это модели для цензурированных данных о продолжительности (например, продолжительности занятости), модели с ограниченной зависимой переменной (модели с частичной наблюдаемостью) для дискретных или цензурированных данных (например, о статусе занятости, количестве отработанных часов), модели для данных с (естественным или намеренным) эндогенным самоотбором выборки (например, модель определения заработной платы среди самоотобранных работников или модели для выборок типа «случай-контроль») и модели с аддитивными непараметрическими эффектами. В следующей таблице приведены некоторые приложения соответствующих моделей.

Модель	Приложения
Регрессионные и одноиндексные модели для цензурированных данных о продолжительности: $Y x \cong Y x'\beta$ .	Продолжительность занятости, инновационные лаги, мобильность.
Модели с ограниченной зависимой переменной (например, дискретной или цензурированной): $Y^* = x'\beta - \varepsilon$ , $\varepsilon x \sim F(\cdot)$ . Преобразование наблюдаемости $Y = \Psi(Y^*)$ : дискретное: $Y = \text{sgn}(Y^*)$ , цензурированное: $Y = \min(Y^c, Y^*)$ .	Дискретная: статус занятости, выбор брэнда. Цензурированная: количество отработанных часов, уровни расходов.
Эндогенный самоотбор выборки: $Y = x'\beta - \varepsilon$ , $\varepsilon x \sim f(\cdot)$ , $x \sim g(\cdot)$ . Естественный: $(Y, x)$ наблюдаются $\Leftrightarrow Y > 0$ . Намеренный: $(Y, x)$ участвуют в выборке $\Leftrightarrow Y > 0$ .	Естественный: самоотобранные работники, домовладельцы. Намеренный: выборка типа «случай-контроль».
Аддитивные непараметрические эффекты: $Y = x'\beta + H(z) + \varepsilon$ .	Устойчивый анализ политики.

В большинстве случаев основная задача полупараметрического анализа состоит в оценивании регрессионных коэффициентов, которые определяют положение распределения зависимой переменной; тогда неизвестное распределение является (бесконечномерным) шумовым параметром. Также в некоторых приложениях непосредственный интерес представляет некоторый функционал неизвестного распределения, например, условное математическое ожидание зависимой переменной. Конечной целью анализа могут быть точечные оценки или доверительные интервалы для исследуемых объектов или тестирование гипотез относительно параметров. Обычно важно получить меру точности получаемых оценок, включая скорости сходимости, асимптотические распределения и бутстраповские или другие показатели точности оценок в конечных выборках и качества асимптотических приближений.

Настоящее эссе не является обзором всего спектра полупараметрических моделей в эконометрике и не рассматривает свойства полупараметрических оценок, кроме как в иллюстративных примерах. Хороший обзор основ полупараметрического анализа можно найти в Powell (1994). В данном эссе рассматриваются лишь две сферы применения. Первая – это анализ цензурированных данных о продолжительности занятости – возможно, ведущая сфера

прикладного полупараметрического оценивания. Вторая – это анализ данных о заявленной готовности платить за природные ресурсы.

## 2 Модели для цензурированных данных о продолжительности занятости

В центре внимания литературы о продолжительности занятости находится воздействие объясняющих переменных, таких как пол, раса, возраст и уровень образования, на риск прекращения работы. Данные о продолжительности занятости обычно являются цензурированными, поскольку периоды занятости начинаются до начала панельного обследования (и дату начала периода не всегда возможно точно определить, используя ретроспективные вопросы) и/или продолжаются после его окончания, или же из-за выбывания объектов наблюдения из панели. В данном разделе рассматривается только цензурирование справа, то есть до окончания периода занятости. При параметрическом анализе моделей продолжительности обычно используются экспоненциальная или вейбулловская кривые выживания или модель пропорциональных рисков Кокса, которая является полупараметрической.

Horowitz & Newmann (1987), возможно, впервые применили на практике методы полупараметрической цензурированной регрессии для анализа данных о продолжительности занятости. Чтобы придать некоторое содержательное наполнение данному экономическому приложению, рассмотрим риски, которые могут привести к окончанию периода занятости. Во-первых, прекращение работы может быть инициировано работником (увольнение по собственному желанию) или работодателем (сокращение, увольнение). На решение работника об увольнении по собственному желанию воздействуют, по-видимому, неденежные характеристики работы (например, безопасность, разнообразие, установленные правила), альтернативные издержки занятости и характеристики работника, такие как уровень образования, раса, преданность работодателю. На решение фирмы об увольнении сотрудника воздействует ожидаемая производительность работника за вычетом заработной платы. Специфический человеческий капитал работника влияет как на альтернативные издержки занятости, так и на ожидаемую производительность. Альтернативные издержки занятости определяются также ожидаемыми страховыми выплатами по безработице и продолжительностью безработицы. Макроэкономические и продуктовые циклы воздействуют на ожидаемую производительность. Следующие аспекты этого словесного описания важны для моделирования продолжительности занятости:

1. Увольнение по собственному желанию и сокращение являются конкурирующими рисками с пересекающимися, но несовпадающими, наборами объясняющих переменных. При структурном оценивании продолжительности необходимо различать эти два вида рисков. Данные о том, заканчивается ли период занятости в результате увольнения по собственному желанию или нет, значительно способствуют идентификации и оцениванию отдельных рисков.
2. Важные объясняющие переменные, такие как уровень макроэкономической активности и запас специфического человеческого капитала работника, меняются во времени, так что структурная модель должна допускать меняющиеся во времени регрессоры. Это довольно легко учесть в случае дискретного времени, используя разнородные марковские модели, но весьма затруднительно в случае непрерывного времени.
3. Ненаблюдаемые переменные, такие как преданность сотрудника работодателю, различаются в популяции и самоотбираются в процессе выживания. Значит, при структурном моделировании продолжительности необходимо определить распределение этих ненаблюдаемых величин. Наличие ненаблюдаемой разнородности также приводит к самоотбору субпопуляции, которая начинает период занятости в интервале наблюдения. Субпопуляция, начинающая период занятости вблизи начала периода наблюдения, будет в

среднем менее преданной работодателю, чем все работники. Те работники, чей первый наблюдаемый период занятости начинается ближе к концу периода наблюдения, будут в среднем более преданными работодателю, если панель достаточно длинная.

4. В структурной модели продолжительности занятости риск должен зависеть исключительно от экономических переменных, но не напрямую от количества прошедшего времени. Следовательно, модели, предполагающие наличие необъясненного «базового» риска, удаляют вариацию, которая должна иметь структурные источники. С точки зрения структурного оценивания экономических факторов продолжительности занятости акцент на эффекте объясняющих переменных смещается при восприятии базового риска как шумового параметра.
5. Экономическая теория не дает конкретных функциональных форм или распределений ненаблюдаемых величин; предположение о том, что наблюдаемые величины входят в модель как параметрическая аддитивная комбинация следует обосновывать как аппроксимацию. Следовательно, анализ, который предполагает, что наблюдаемые величины входят в модель в виде конкретной аддитивной комбинации при неизвестных преобразованиях или распределениях, на самом деле предполагает слишком много о структуре аддитивной комбинации, и, возможно, слишком мало о неизвестных преобразованиях, которые можно достаточно точно аппроксимировать при помощи гибких семейств с конечным числом параметров.

Процесс, порождающий данные о продолжительности занятости, можно охарактеризовать при помощи *кривой выживания*  $q(t|x)$ , дающей долю популяции с периодами занятости, начинающимися в момент времени 0, которая доживает до момента времени  $t$ , при условии наблюдаемой динамики регрессоров  $x(\cdot)$ . Если присутствуют ненаблюдаемые регрессоры  $\xi$ , распределенные в исходной популяции в соответствии с функцией плотности  $\nu(\cdot|x, 0)$ , а  $q(t|x, \xi)$  – «структурная» кривая выживания, то процесс, порождающий данные, удовлетворяет следующему соотношению:

$$q(t|x) = \int_{-\infty}^{+\infty} q(t|x, \xi) \cdot \nu(\xi|x, 0) d\xi. \quad (1)$$

Функция плотности ненаблюдаемых регрессоров при условии дожития меняется во времени из-за отбора и удовлетворяет уравнению

$$\nu(\xi|x, t) = \nu(\xi|x, 0) \cdot \frac{q(t|x, \xi)}{q(t|x)}. \quad (2)$$

Кривую выживания также можно описать с помощью *функции риска*:

$$h(t|x, \xi) = -\nabla_t \ln(q(t|x, \xi)). \quad (3)$$

*Средняя норма риска* в выжившей популяции равна

$$\begin{aligned} h^*(t|x) &= -\nabla_t \ln(q(t|x)) = \\ &= \frac{\int_{-\infty}^{+\infty} h(t|x, \xi) q(t|x, \xi) \nu(\xi|x, 0) d\xi}{q(t|x)} = \int_{-\infty}^{+\infty} h(t|x, \xi) \nu(\xi|x, t) d\xi. \end{aligned} \quad (4)$$

Обращая уравнение (3), получаем

$$q(t|x, \xi) = \exp\left(-\int_0^t h(s|x, \xi) ds\right) \equiv \exp(-\Lambda(t|x, \xi)), \quad (5)$$

где  $\Lambda(t|x, \xi)$  – так называемый *интегральный риск*. Средняя продолжительность завершённых периодов занятости равна

$$\mathbb{E}[t|x, \xi] = - \int_0^\infty t \cdot \nabla_t q(t|x, \xi) dt = \int_0^\infty q(t|x, \xi) dt, \quad (6)$$

где второе равенство получено путем интегрирования по частям.

Когда интервал наблюдения конечен, некоторые периоды занятости *прерываются* или *цензурируются справа*; функция выживания, определенная вплоть до момента цензурирования, продолжает характеризовать процесс, порождающий данные. Средняя продолжительность периода занятости, завершённого естественным образом (в момент времени  $t$ ) или в результате цензурирования (в момент времени  $t^c$ ) равна

$$\mathbb{E}[\min(t, t^c)] = - \int_0^{t^c} t \cdot \nabla_t q(t|x, \xi) dt + t^c q(t^c|x, \xi) = \int_0^{t^c} q(t|x, \xi) dt. \quad (7)$$

Аналогичные формулы справедливы для средней нормы риска.

При наличии выбывания из выборки момент цензурирования становится случайной величиной с соответствующей функцией выживания  $r(t^c|x, \xi)$ . В этом случае вероятность того, что наблюдение периода занятости продолжается до момента  $t$ , равна  $q(t|x, \xi)r(t|x, \xi)$ ; общий риск завершения наблюдаемого периода занятости естественным путем или в результате цензурирования равен  $h(t|x, \xi) - r'(t|x, \xi)/r(t|x, \xi)$ ; для периода, заканчивающегося в момент времени  $t$ , вероятность цензурирования равна  $h(t|x, \xi)/(h(t|x, \xi) - r'(t|x, \xi)/r(t|x, \xi))$ , а средняя продолжительность наблюдаемых периодов занятости равна

$$\int_0^\infty q(t|x, \xi)r(t|x, \xi) dt.$$

Примером параметрической модели продолжительности, когда вектор  $x$  неизменен во времени, является модель *Вейбулла*:

$$q(t|x) = \exp(-t^\alpha e^{-x'\beta}), \quad (8)$$

где  $\alpha$  – положительный параметр,  $\beta$  – вектор параметров, а  $x$  – вектор регрессоров. Соответствующая функция риска имеет вид

$$h(t|x) = \alpha t^{\alpha-1} e^{-x'\beta}, \quad (9)$$

а средняя продолжительность завершённых периодов равна

$$\mathbb{E}[t|x] = e^{x'\beta/\alpha} \Gamma(1 + 1/\alpha), \quad (10)$$

где  $\Gamma(\cdot)$  – гамма-функция. При  $\alpha = 1$  получаем *экспоненциальную* модель продолжительности.

Имеются три стратегии статистического оценивания цензурированных данных о продолжительности:

1. Полностью параметрический подход, когда предполагается, что  $q(t|x)$  или, в случае ненаблюдаемой разнородности,  $q(t|x, \xi)$  и  $\nu(\xi|x, 0)$  принадлежат семействам с конечным числом параметров.<sup>1</sup>

<sup>1</sup>Типичными примерами являются предположение о вейбулловском или логнормальном распределении для  $q(t|x)$  или экспоненциальном распределении для  $q(t|x, \xi)$  в комбинации с гамма-распределением для  $\xi$ . Параметры распределения можно оценить методом максимального правдоподобия.

2. Полностью непараметрический подход, когда  $q(t|x)$  оценивается без каких-либо параметрических ограничений, например, при помощи оценки Каплана–Мейера.<sup>2</sup>
3. Одноиндексный полупараметрический подход, когда  $q(t|x)$  зависит от  $x$  через скалярную функцию  $V(x, \beta)$ , которая известна, за исключением конечного вектора параметров  $\beta$ , но  $q(t|v)$  не ограничивается параметрическим семейством. В случае ненаблюдаемой разнородности либо  $q(t|v, \xi)$ , либо  $\nu(\xi|v, t)$  могут быть непараметрическими (но не оба одновременно, если нет дополнительных ограничений, ввиду требований идентификации).<sup>3</sup>

Рассмотрим некоторые альтернативные варианты полупараметрических моделей, которые предлагаются в литературе. Пусть  $x$  – вектор регрессоров, предполагаемый *неизменным во времени*. Пусть далее  $\beta$  – вектор неизвестных параметров,  $V(x, \beta) \equiv x'\beta$  – одноиндексная функция с неизвестными параметрами  $\beta$ , а  $q(t|x'\beta)$  – функция выживания. Пусть  $T^*$  – случайная величина, обозначающая количество прошедшего времени, а  $T^c$  – момент цензурирования, так что наблюдаемая продолжительность соответствует  $T = \min(T^*, T^c)$ . Имеются четыре альтернативные модели для  $T$ :

1. *Модель регрессии*:  $\ln T^* = x'\beta + \varepsilon$ , где  $\varepsilon|x$  имеет неизвестную плотность распределения  $f(\varepsilon)$  с нулевым средним. Относительно функции плотности  $f(\cdot)$  часто предполагают симметричность и гомоскедастичность. Модели соответствует следующая функция выживания:

$$q(t|x'\beta) = 1 - F(\ln t - x'\beta), \quad (11)$$

где  $F(\cdot)$  – кумулятивная функция распределения для  $f(\cdot)$ . Соответствующая функция риска имеет вид

$$h(t|x'\beta) = \frac{f(\ln t - x'\beta)}{t[1 - F(\ln t - x'\beta)]}. \quad (12)$$

Обобщение этой модели допускает гетероскедастичность  $\varepsilon$ , когда дисперсия зависит от индекса  $x'\beta$ , или, в более общем случае, от некоторой другой функции от  $x$ . *Модель цензурированной регрессии* – это просто модель вида

$$\ln T = \min(\ln T^c, x'\beta + \varepsilon). \quad (13)$$

В случае неслучайного цензурирования она обладает тем свойством, что

$$\mathbb{E}[\ln T|x] = \int [1 - F(y - x'\beta)] dy \quad (14)$$

<sup>2</sup>Классическая оценка Каплана–Мейера формулируется для данных о продолжительности в случае отсутствия регрессоров. Предположим, что в данных периоды занятости, начинающиеся в один и тот же момент времени 0, прерываются (естественным образом или в результате цензурирования) в моменты времени  $t_1 < \dots < t_J$ . Пусть  $n_j$  обозначает число периодов, которые завершаются естественным образом в момент времени  $t_j$ , а  $m_j$  – число периодов, цензурируемых в этот момент времени. Общее число периодов, находящихся «в группе риска» в момент времени  $t_j$ , равно  $N_j = \sum_{i=j}^J (n_i + m_i)$ . Оценка Каплана–Мейера для функции риска в момент  $t_j$  имеет вид  $h^*(t_j) = n_j/N_j$ . Соответствующая оценка функции выживания имеет вид  $q^*(t_j) = (1 - h^*(t_j))q^*(t_{j-1})$ , или  $q^*(t_j) = \prod_{i=1}^j (1 - n_i/N_i)$ . При наличии категориальных регрессоров оценка Каплана–Мейера, очевидно, применяется отдельно для каждой клетки для всех возможных комбинаций регрессоров. Используя идею оценки ближайших соседей из непараметрического регрессионного анализа, оценку Каплана–Мейера можно адаптировать для общего случая некатегориальных регрессоров. В случае ненаблюдаемой разнородности, вообще говоря, невозможно идентифицировать функции выживания и плотность распределения ненаблюдаемых регрессоров, когда оба этих объекта являются непараметрическими. Heckman & Singer (1984) установили этот результат, а также предложили полупараметрические методы для оценивания параметрической структурной функции выживания  $q(t|x, \xi, \beta)$  при наличии непараметрической плотности распределения разнородности  $\nu(\xi|x, 0)$ .

<sup>3</sup>Другие полупараметрические подходы включают многоиндексные модели и методы параметризации квантилей без полной параметризации распределения.

является возрастающей функцией от  $x'\beta$ .

2. *Модель с преобразованием (обобщенная модель Бокса–Кокса)*. Предположим,  $G$  является неизвестным монотонно возрастающим преобразованием из  $(0, +\infty)$  на множество действительных чисел, и предположим, что

$$G(T^*) = x'\beta + \varepsilon, \quad (15)$$

где  $\varepsilon|x$  имеет известную или неизвестную плотность распределения  $f(\varepsilon)$ . Соответствующая функция выживания имеет вид

$$q(t|x'\beta) = 1 - F(G(t) - x'\beta), \quad (16)$$

а соответствующая функция риска –

$$h(t|x'\beta) = \frac{G'(t)f(G(t) - x'\beta)}{1 - F(G(t) - x'\beta)}. \quad (17)$$

Опять же, модель можно обобщить на случай гетероскедастичности относительно  $x'\beta$ .

3. *Целенаправленное проецирование (одноиндексная регрессия)*. Предположим,  $H$  – неизвестное преобразование из множества действительных чисел в себя. Предположим, что

$$\ln T^* = H(x'\beta) + \varepsilon, \quad (18)$$

где  $\varepsilon|x$  имеет известную или неизвестную плотность распределения  $f(\varepsilon)$ . Соответствующая функция выживания имеет вид

$$q(t|x'\beta) = 1 - F(\ln t - H(x'\beta)), \quad (19)$$

а функция риска –

$$h(t|x'\beta) = \frac{f(\ln t - H(x'\beta))}{t[1 - F(\ln t - H(x'\beta))]} \quad (20)$$

Распределение ошибок обычно предполагается гомоскедастичным, но некоторые оценки этой модели допускают гетероскедастичность относительно  $x'\beta$ .

4. *Модель пропорциональных рисков*. Предположим, что  $h_0(t)$  – неизвестная неотрицательная функция «базового риска», а регрессоры оказывают пропорциональный эффект на риск, то есть

$$h(t|x) = h_0(t) \exp(-x'\beta). \quad (21)$$

Определим базовый интегральный риск:

$$\Lambda_0(t) = \int_0^t h_0(s) ds. \quad (22)$$

Тогда функция выживания принимает вид

$$q(t|x'\beta) = \exp(-\Lambda_0(t)e^{-x'\beta}), \quad (23)$$

и

$$\ln \Lambda_0(T^*) = x'\beta + \varepsilon, \quad (24)$$

где  $\varepsilon$  имеет распределение экстремальных значений:

$$F(\varepsilon) = 1 - \exp(-e^{-\varepsilon}). \quad (25)$$

Другие распределения ошибки можно получить из модели пропорциональных рисков с ненаблюдаемой разнородностью. Например, следуя работе Lancaster (1979), предположим, что

$$h(t|x, \xi) = h_0(t) \exp(-x'\beta)\xi, \quad (26)$$

где  $\xi$  имеет гамма-распределение,  $\nu(\xi|x, 0) = \xi^{\theta-1}e^{-\xi}/\Gamma(\theta)$ . Тогда, применяя соотношение (1), получаем

$$q(t|x) = \left(1 + e^{\Lambda_0(t)-x'\beta}\right)^{-\theta}, \quad (27)$$

откуда следует, что выполняется уравнение (15), когда  $\varepsilon$  имеет обобщенное логистическое распределение (или  $e^\varepsilon$  имеет распределение Парето):

$$F(\varepsilon) = 1 - (1 + e^\varepsilon)^{-\theta}. \quad (28)$$

Средний риск для (26) равен

$$h^*(t|x) = \frac{\theta h_0(t) e^{\Lambda_0(t)}}{e^{\Lambda_0(t)} + e^{x'\beta}} \quad (29)$$

и больше не принимает форму пропорциональных рисков. Условное распределение ненаблюдаемых регрессоров при данной функции выживания  $\nu(\xi|x, t)$  остается гамма-распределением с параметром  $\theta$ , но относительно преобразованной величины  $(1 + e^{\Lambda_0(t)-x'\beta})\xi$ .

Модель пропорциональных рисков (21) является частным случаем модели с преобразованием, когда ошибка имеет распределение (25). Модель пропорциональных рисков с разнородностью (26) – это также частный случай модели с преобразованием. Когда базовый риск является степенной функцией от  $t$ ,  $h_0(t) = \alpha t^{\alpha-1}$ , модель (21) упрощается до параметрической вейбулловской модели продолжительности, а также может быть интерпретирована как модель цензурированной регрессии с ошибками, имеющими распределение экстремальных значений.

Общая «аддитивная одноиндексная модель», включающая как частные случаи четыре описанные модели, имеет вид

$$G(T^*) = H(x'\beta) + \varepsilon, \quad (30)$$

где  $\varepsilon$  имеет кумулятивную функцию распределения  $F(\cdot)$ . Соответствующая функция выживания имеет вид

$$q(t|x'\beta) = 1 - F(G(T) - H(x'\beta)). \quad (31)$$

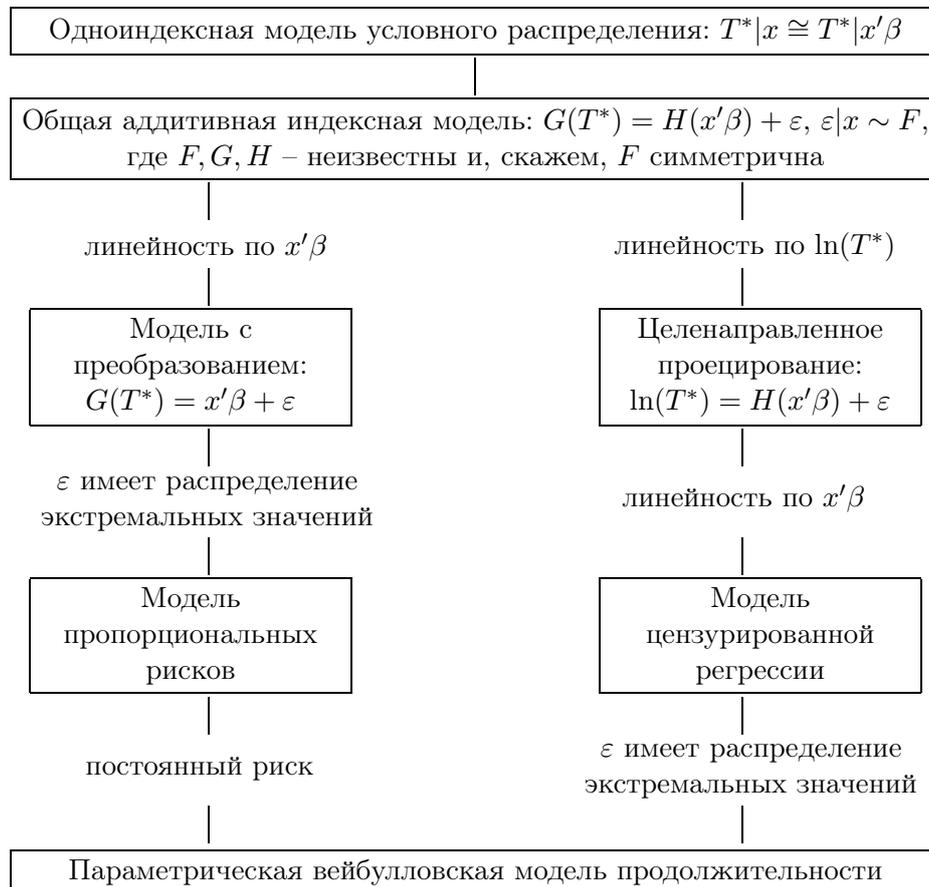
На рисунке 1 показана логическая связь между этими моделями. Все они являются частными случаями *одноиндексной модели*, в которой условное распределение зависимой переменной зависит от регрессоров  $x$  исключительно через индекс  $x'\beta$ . Модель пропорциональных рисков и модель цензурированной регрессии логически различаются, за исключением того факта, что обе они упрощаются до параметрической вейбулловской модели. Обе модели являются частными случаями модели с преобразованием. Модель цензурированной регрессии является частным случаем регрессионной модели целенаправленного проецирования. Модель с преобразованием можно записать как гетероскедастичную модель целенаправленного проецирования: если  $G(T^*) = x'\beta + \varepsilon$ , где  $G(\cdot)$  – монотонно возрастающее преобразование, то  $\ln T^* = H(x'\beta) + \zeta$ , где  $H(x'\beta) = \mathbb{E}_\varepsilon[\ln G^{-1}(x'\beta + \varepsilon)]$ , а  $\zeta$  имеет функцию распределения  $F(G(\exp(\zeta + H(x'\beta))) - x'\beta)$ , которая в общем случае гетероскедастична.

Статистические вопросы, которые возникают при применении этих моделей, включают свойства распределений оценок (асимптотические и, возможно, в конечных выборках), которые получаются при различных предположениях, и эффективность альтернативных оценок.

Рис. 1: Одноиндексные модели

Правила наблюдения:  $T = \min(T^c, T^*)$  для данных, цензурированных справа,  
 $T = \text{sgn}(\ln(T^*))$  для биномиальных моделей дискретного выбора.

(Специфика модели растет по мере продвижения вниз по таблице)



До настоящего времени большая часть исследований сконцентрирована на поиске вычислительно доступных оценок, установлении их состоятельности, асимптотической нормальности и границ эффективности.

Хоровиц и Ньюманн используют две оценки для модели цензурированной регрессии – квантильную оценку (Powell, 1986) и одношаговую полупараметрическую ОМНК-оценку (ПОМНК) (Horowitz, 1986). Другие оценки, предложенные для данной модели, включают гибкие параметрические приближения кумулятивной функции распределения (см., например, Duncan (1986), который рассматривает приближения сплайнами – «метод решета»). Chamberlain (1986) и Cosslett (1987) установили для модели цензурированной регрессии существование положительной границы эффективности для параметрической части. Это означает, что можно использовать достаточно грубые оценки непараметрической части, чтобы достичь  $\sqrt{N}$  асимптотически нормальной оценки для параметрической части. Доказано, что оценки из Powell (1986) и Horowitz (1986) являются асимптотически нормальными. Ни одна из них не достигает границы эффективности в случае IID-ошибок, и в общем случае одна не является эффективнее другой.

Оценивание модели пропорциональных рисков с неизвестной функцией базового риска подробно изучено, см. Kaplan & Meier (1968), Cox (1972), Kalbfleisch & Prentice (1982) и Meyer (1990). Особенно полезный «полупараметрический» метод оценивания этой модели, приме-

нимый, когда продолжительность измеряется в «неделях», – гибко параметризовать базовый риск; Меуер (1990) показал, что этот метод является  $\sqrt{N}$  асимптотически нормальным.

Оценки (одноиндексной) модели целенаправленного проецирования были предложены в Ichimura (1987), Ruud (1986), Stoker (1986) и Powell, Stock & Stoker (1989). Оценка Ичимуры выбирает  $\beta$ , минимизирующую дисперсию  $\ln T$  условно на  $x'\beta$ , используя ядерную оценку условного среднего для получения оценки условной дисперсии. Эта оценка состоятельна, даже если ошибки разнородны относительно индексной функции, так что ее также можно применять для модели с преобразованием. Оценка Ичимуры является  $\sqrt{N}$  асимптотически нормальной, и, как недавно было показано, достигает полупараметрической границы эффективности для гомоскедастичной модели целенаправленного проецирования с нормальными ошибками. Она почти наверняка не является эффективной для модели с преобразованием. Оценки Рууда и Стокера основаны на том факте, что при подходящих условиях регрессия  $\ln T$  на  $x$  пропорциональна  $\beta$ . Эти оценки также  $\sqrt{N}$  асимптотически нормальны.

Оценивание модели с преобразованием, применимое также к модели пропорциональных рисков, реализуется с помощью метода максимальной ранговой корреляции, предложенного в Han (1987) и Doksum (1985).

Newey (1990) установил асимптотическую эффективность некоторых ядерных и квантильных оценок модели цензурированной регрессии, когда ошибки имеют симметричное распределение. Эффективность этих оценок при других условиях не установлена. Проблемой, требующей дальнейших исследований, является построение надежных и практичных оценок дисперсии полупараметрических оценок. Интересный эмпирический вопрос заключается в том, можно ли воспринимать модель цензурированной регрессии или модель пропорциональных рисков как ограничения модели с преобразованием (и каковы подходящие и удобные тестовые статистики).

### 3 Заявленная готовность платить за природные ресурсы

Методом выявления готовности платить (ГП) за природные ресурсы является экспериментальный опрос населения об их условных оценках: участникам обследования задается вопрос, готовы ли они платить величину  $b$ , где  $b$  – ставка, установленная правилами эксперимента. Пусть  $d$  обозначает фиктивную переменную, равную единице при ответе «да» и нулю в противном случае. Выборка из  $n$  наблюдений формируется из пар  $(b, d)$ , а также регрессоров  $x$ , характеризующих респондента. Предположим, что ГП распределена в популяции как  $w = x'\beta - \varepsilon$ , где  $\varepsilon$  имеет кумулятивную функцию распределения  $G(\varepsilon)$ , не зависящую от  $x$ . Тогда  $\mathbb{P}\{d = 1|x'\beta\} = G(x'\beta - b)$ , или

$$d = G(x'\beta - b) + \varepsilon. \quad (32)$$

Предположим, что  $\beta$  и функция  $G$  неизвестны. Эконометрическая задача состоит в том, чтобы оценить  $\beta$  и, если необходимо,  $G$  и при помощи этих оценок измерить положение распределения ГП, условное на  $x$  или безусловное. Это пример регрессионной модели целенаправленного проецирования.

Экспериментальные опросы об условных оценках вызывают споры, поскольку они очень чувствительны к психометрическим контекстным эффектам, таких как якорение, при котором респонденты, не уверенные в своих предпочтениях, воспринимают предлагаемую ставку как сигнал о «политкорректном» диапазоне значений оценки. Также некоторые субъекты, по-видимому, действуют стратегически, намеренно принимая ложно высокую ставку, которую в действительности они не заплатили бы, но которая выражает «протестную» позицию. Эти эффекты делают оценки ГП неточными, а их связь с экономикой благосостояния непрочной.

Почему же в экспериментальных опросах об условных оценках для их выявления применяется формат референдума, а не формат, при котором респондентов просили бы дать свободный ответ о ГП? Одной из причин является то, что открытый формат ведет к гораздо более высокой доле отсутствия ответа, так что метод референдума снижает смещение вследствие самоотбора, вызываемого отсутствием ответов. Другая причина состоит в том, что психологически референдум и открытый формат выявляют весьма различное поведение. Некоторые считают, что формат референдума ближе к механизму выборов, обычно применяемому для принятия общественных решений, и имеется преимущество в подражании этому механизму при принятии общественных решений о природных ресурсах.

Один из вопросов, возникающих при разработке экспериментальных опросов об условных оценках, – выбор уровней ставок  $b$ . Альтернативами являются случайный выбор  $b$  или выбор  $b$  на сетке с определенным размером ячеек. На практике используются грубые сетки, что ограничивает точность полупараметрических оценок. Пусть  $h(b|x)$  – плотность распределения, из которого вытягиваются уровни ставок  $b$ , условно на  $x$ . Оно известно исследователю, поскольку выбирается разработчиками эксперимента.

При эконометрическом анализе данных по референдуму о ГП можно использовать тот факт, что (32) является моделью бинарного выбора и одноиндексной моделью (которая гетероскедастична, но только относительно индекса). Тогда доступными методами для оценивания  $\beta$  являются оценка, основанная на максимуме очков из Manski (1978), полупараметрическая ММП-оценка из Cosslett (1987), оценка из Ichimura (1986), минимизирующая ожидаемую условную дисперсию, оценка из Horowitz (1992), являющаяся гладкой версией оценки, основанной на максимуме очков, и оценка из Klein & Spady (1993). Ключевой результат для модели бинарного выбора состоит в том, что при некоторых условиях гладкости, существуют  $\sqrt{N}$ -состоятельные оценки  $\beta_n$  для  $\beta$ , т.е. величина  $\sqrt{N}(\beta_n - \beta)$  асимптотически нормальна. Непараметрическую оценку  $G$  можно получить совместно с оцениванием  $\beta$ , как в процедуре Косслетта, или при помощи обычных ядерных методов на втором шаге, после того как оценка  $\beta$  подставляется для формирования индекса; ее непараметрическая оценка обязательно будет иметь скорость сходимости меньшую, чем  $\sqrt{N}$ .

Особенно простая оценка параметров индекса  $\beta$  была предложена для этой задачи в Lewbel & McFadden (1997): надо просто оценить с помощью МНК модель

$$\frac{d_i - \mathbb{I}\{b_i < 0\}}{h(b_i|x_i)} = x_i\beta + \zeta_i. \quad (33)$$

Авторы показывают, что оценки коэффициентов в данной регрессии являются состоятельными оценками  $\beta$  и асимптотически нормальны со скоростью сходимости  $\sqrt{N}$ . Эти оценки не являются особо эффективными, но их простота делает их отличной отправной точкой для анализа спецификации модели и построения более эффективных оценок. Авторы также устанавливают, что  $r$ -й момент ГП, условно на  $x = x_0$ , можно  $\sqrt{N}$ -состоятельно оценить следующим образом:

$$M_r = (x_0\beta)^r + r \sum_{i=1}^n (b_i + (x_0 - x_i)\beta)^{r-1} \cdot \frac{d_i - \mathbb{I}\{x_i\beta > b_i\}}{\sum_{j=1}^n h(b_i + (x_j - x_i)\beta|x_j)}. \quad (34)$$

Оценки (33) и (34) – хорошие примеры статистических процедур полупараметрического оценивания, которые устойчивы в том смысле, что они не зависят от параметрических предположений о распределении ГП и представляют собой вычислительно удобную альтернативу непараметрическим оценкам ядерного типа.

## Литература

- Cosslett, S. (1987). Efficiency bounds for distribution-free estimators of the binary choice and the censored regression models. *Econometrica* 55, 559–585.
- Cox, D. (1972). Regression models and life tables. *Journal of Royal Statistical Society B* 34, 187–220.
- Doksum, K. (1985). An extension of partial likelihood methods for proportional hazard models to general transformation models. Working paper, University of California, Berkeley.
- Duncan, G. (1986). A semiparametric censored regression estimator. *Journal of Econometrics* 29, 5–34.
- Han, A. (1987). Nonparametric analysis of generalized regression models: The maximum rank correlation estimator. *Journal of Econometrics* 35, 303–316.
- Heckman, J. & B. Singer (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica* 52, 271–320.
- Horowitz, J. (1986). A distribution-free least squares method for censored linear regression models. *Journal of Econometrics* 29, 59–84.
- Horowitz, J. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica* 60, 505–531.
- Horowitz, J. & G. Newmann (1987). Semiparametric estimation of employment duration models. *Econometric Reviews* 6, 5–40.
- Horowitz, J. & G. Newmann (1989). Computational and statistical efficiency of semiparametric GLS estimators. *Econometric Reviews* 8, 223–225.
- Ichimura, H. (1986). *Estimation of Single Index Models*. Ph.D. Dissertation, MIT.
- Kalbfleisch, J. & R. Prentice (1980). *The Stochastic Analysis of Failure Time Data*. New York: Wiley.
- Kaplan, E. & P. Meier (1958). Nonparametric estimation from incomplete observations. *Journal of American Statistical Association* 53, 487–491.
- Klein, R. & R. Spady (1993). An efficient semiparametric estimator for binary response models. *Econometrica* 61, 387–422.
- Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica* 47, 141–165.
- Lewbel, A. and D. McFadden (1997). Estimating features of a distribution from binomial data. Working paper, University of California, Berkeley.
- Manski, C. (1978). Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics* 3, 205–228.
- Meyer, B. (1987). Unemployment insurance and unemployment spells. *Econometrica* 58, 757–782.
- Newey, W. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics* 5, 99–135.
- Powell, J. (1986). Censored regression quantiles. *Journal of Econometrics* 29, 143–155.
- Powell, J., J. Stock & T. Stoker (1989). Semiparametric estimation of weighted average derivatives. *Econometrica* 57, 1403–1430.
- Powell, J. (1994). Estimation of Semiparametric Models. Глава в *Handbook of Econometrics IV* под редакцией R. Engle & D. McFadden. Amsterdam: North-Holland.
- Ruud, P. (1986). Consistent estimation of limited dependent variable models despite misspecification of distribution. *Journal of Econometrics* 29, 157–187.
- Stoker, T. (1986). Consistent estimation of scaled coefficients. *Econometrica* 54, 1461–1481.

## Semiparametric analysis

Daniel McFadden

*University of California, Berkeley, USA*

This essay surveys two areas of application of semiparametric econometrics: the analysis of censored employment duration data, and the analysis of data on stated willingness-to-pay for natural resources.

# В помощь изучающим эконометрику

## Мини-словарь англоязычных эконометрических терминов, часть 2\*

Александр Цыплаков<sup>†</sup>

*Новосибирский государственный университет, Новосибирск, Россия*

В этой части словаря комментируются англоязычные эконометрические термины *kurtosis*, *skewness*, *critical region*, *significance level*, *confidence level* и др. Акцент вновь делается на уточнении значения терминов с целью избежать возможной путаницы и некорректностей при интерпретации.

### **Kurtosis, skewness**

Куртозис (*kurtosis*) является показателем, отражающим остроту вершины и толщину хвостов одномерного распределения. Термин происходит от греческого слова *κυρτός*, что означает «выгнутый», «раздутый». Этот показатель строится на основе четвертого центрального момента распределения ( $\mu_4$ ). Обычно в эконометрике используют следующее определение:

$$\kappa = \frac{\mu_4}{\sigma^4},$$

где  $\sigma^4$  — квадрат дисперсии. У нормального распределения  $\kappa = 3$ , в связи с чем распространение получил модифицированный показатель

$$\kappa - 3 = \frac{\mu_4}{\sigma^4} - 3,$$

называемый *excess kurtosis*. Прилагательное *excess* здесь буквально означает «превышающий норму».

С употреблением и переводом этого слова связан ряд проблем. Во-первых, термин *kurtosis* может использоваться как родовое название для ряда показателей формы распределения, связанных с остротой вершины и толщиной хвостов. В частности, показатель  $\kappa - 3$  часто тоже называют просто *kurtosis*. Во-вторых, в качестве перевода термина *kurtosis* в русскоязычной литературе закрепилось слово «эксцесс». Возможно, это связано с переводом на русский язык в 1948 г. известной книги Гаральда Крамера по статистике (Крамер, 1975). Крамер использовал термин *excess* без добавления *kurtosis* для показателя  $\kappa - 3$ . В англо-русских словарях для *kurtosis* предлагается перевод «эксцесс». Однако ясно, что показатель  $\kappa$  некорректно называть эксцессом, поскольку он не является «излишком» по отношению к 3, как  $\kappa - 3$ . В связи с этим желательнее переводить *kurtosis* на русский язык как «куртозис» или «куртосис», а *excess kurtosis* как «эксцесс». Слово *excess* в значении  $\kappa - 3$  в английском языке практически не употребляется, и во избежание недоразумений лучше использовать *excess kurtosis*.

**Kurtosis:** куртозис, куртосис.

**Excess kurtosis:** эксцесс.

\*Цитировать как: Цыплаков, Александр (2008) «Мини-словарь англоязычных эконометрических терминов, часть 2», Квантиль, №5, стр. 41–48. Citation: Tsyplakov, Alexander (2008) “A mini-dictionary of English econometric terminology II,” *Quantile*, No.5, pp. 41–48.

<sup>†</sup>Адрес: 630090, г. Новосибирск, Весенний проезд, 6–44. Электронная почта: [tsy@academ.org](mailto:tsy@academ.org)

Имеется ряд терминов, которые показывают, насколько велик куртозис распределения, и сравнивают его с куртозисом нормального распределения. Так, *leptokurtic distribution* — это распределение с положительным эксцессом, имеющее острую вершину и толстые хвосты, а *platykurtic distribution* — это распределение с отрицательным эксцессом, имеющее плоскую вершину и тонкие хвосты. Соответственно, *mesokurtic distribution* — это распределение, похожее по куртозису на нормальное распределение.

**Leptokurtic:** лептокуртический, островершинный, имеющий высокий куртозис, имеющий положительный (выше нормального) эксцесс.

**Platykurtic:** платикуртический, плосковершинный, имеющий низкий куртозис, имеющий отрицательный (ниже нормального) эксцесс.

**Mesokurtic:** мезокуртический, имеющий нормальный куртозис, имеющий нулевой эксцесс.

**Kurtotic:** то же, что *leptokurtic*.

**Leptokurtosis:** высокий куртозис, превышающий куртозис нормального распределения, положительный эксцесс.

**Platykurtosis:** низкий куртозис, меньший куртозиса нормального распределения, отрицательный эксцесс.

**Tail:** хвост (распределения).

**Heavy tail, fat tail, thick tail:** тяжелый хвост, утяжеленный хвост, толстый хвост.

**Long tail:** вытянутый хвост, удлиненный хвост.

**Thin tail:** тонкий хвост.

**Long-tailed:** имеющий удлиненный хвост.

**Heavy-tailed:** имеющий утяжеленный хвост.

**Heavy-tailedness:** утяжеленность хвоста.

Несимметричность распределения может характеризоваться терминами *skewness* или *asymmetry*. В качестве формального показателя асимметрии распределения обычно используется

$$\frac{\mu_3}{\sigma^3},$$

где  $\mu_3$  — третий центральный момент, а  $\sigma^3$  — дисперсия в степени  $3/2$ . Этот показатель также называют *skewness*.

**Asymmetry:** асимметрия.

**Skewness:** скошенность, асимметрия.

**Skew (сущ.):** то же, что *skewness*.

**Asymmetric/asymmetrical distribution:** асимметричное распределение.

**Skewed distribution, skew distribution:** скошенное распределение.

При положительной скошенности распределения правый хвост, как правило, толще левого, а вершина смещена влево. Такое распределение часто называют скошенным вправо (исходя из соотношения толщины хвостов, а не из положения вершины). При отрицательной скошенности, соответственно, левый хвост, как правило, толще правого, а вершина смещена вправо. Такое распределение называют скошенным влево.

Right-skewed distribution: скошенное вправо распределение.

Left-skewed distribution: скошенное влево распределение.

### Critical region, significance level

В этой словарной статье мы обсудим некоторые термины, связанные с проверкой гипотез. Стандартная процедура проверки гипотезы о параметрах статистической модели состоит в том, что выбираются нулевая гипотеза и конкурирующая с ней альтернативная гипотеза, а также некоторая функция данных — тестовая статистика (*test statistic*). Для тестовой статистики определяются две области: область принятия (область значений статистики, при попадании в которую принимается нулевая гипотеза) и критическая область (область, при попадании в которую отклоняется нулевая гипотеза и принимается альтернативная).

В английском языке слово «гипотеза» часто опускается, и говорится просто *the null, the alternative*. Следует также помнить, что «нулевая гипотеза» — это именно *null hypothesis*, а не *zero hypothesis* (хотя такое словоупотребление изредка встречается)<sup>1</sup>.

Чаще всего для одномерных статистик критическая область задается при помощи критического значения (*critical value*). Это граничная точка между критической областью и областью принятия нулевой гипотезы. Обычно если значение тестовой статистики превышает критическое значение, то нулевая гипотеза отклоняется и принимается альтернативная гипотеза. Если значение статистики ниже критического значения, то наоборот.

To test a hypothesis: проверить гипотезу.

Statistical test: статистический тест, статистический критерий.

Test statistic: тестовая статистика, статистика критерия.

Null hypothesis, null,  $H_0$ : нулевая гипотеза.

Alternative hypothesis, alternative,  $H_1$ : альтернативная гипотеза.

To accept a hypothesis: принять гипотезу.

To reject a hypothesis: отклонить, отвергнуть гипотезу.

Critical region: критическая область.

Rejection region: область отклонения (неприятя) нулевой гипотезы, критическая область.

Acceptance region: область принятия нулевой гипотезы.

Critical value: критическое значение.

<sup>1</sup>Zero hypothesis можно понять и в смысле ограничения, состоящего в том, что некий коэффициент равен нулю. Лучше, конечно, такие неоднозначные термины не использовать.

С процедурой проверки гипотезы связаны различные события и их вероятности. Результат проверки может быть верным (принята верная нулевая гипотеза или отклонена неверная нулевая гипотеза) или ошибочным. Ошибка первого рода состоит в отклонении верной нулевой гипотезы, а ошибка второго рода — в принятии неверной нулевой гипотезы.

Вероятность ошибки первого рода обычно обозначается буквой  $\alpha$ . (Отсюда происходит еще один термин для ошибки первого рода —  $\alpha$  error.) Эту вероятность принято называть уровнем значимости или размером теста. Термины *significance level* и *size* обычно употребляются как синонимы, хотя есть некоторые отличия. Например, обычно говорят, что «гипотеза проверяется на таком-то уровне значимости»; термин *size* здесь не годится.

Если проверяется сложная нулевая гипотеза (т. е. гипотеза, которая неоднозначно задает распределение), то вероятность ошибки первого рода меняется в зависимости от распределения, и в качестве уровня значимости (размера) берется верхняя граница по рассматриваемому семейству распределений, удовлетворяющих нулевой гипотезе. Поэтому *type I error probability* и *significance level/size* — это не полные синонимы.

Близкий к уровню значимости термин — это «р-значение» *p-value*. *P* в данном случае — это сокращение слова *probability* («вероятность»). *P*-значение — это уровень значимости, который наблюдался бы в том случае, если бы критическое значение было выбрано равным текущему значению статистики, или другими словами, это граничный уровень значимости — если бы использовался такой уровень значимости, то при текущем значении статистики мы были бы на грани отклонения нулевой гипотезы. Отсюда синоним *p*-значения — *marginal significance level*.

Вероятность не совершить ошибку второго рода называется мощностью теста (*power*). Это единица минус вероятность ошибки второго рода. Мощность обычно обозначается буквой  $\beta$ . (Отсюда происходит еще один термин для ошибки второго рода —  $\beta$  error.)

*Type I error, type 1 error, error of the first kind*: ошибка первого рода.

$\alpha$  error:  $\alpha$ -ошибка, ошибка первого рода.

*Type II error, type 2 error, error of the second kind*: ошибка второго рода.

$\beta$  error:  $\beta$ -ошибка, ошибка второго рода.

*Rejection probability/rate*: вероятность отклонения нулевой гипотезы.

*Test significance level, test significance*: уровень значимости теста (критерия), значимость теста.

*Conventional significance level*: обычный (традиционный) уровень значимости.

*To test a hypothesis at the 5% significance level*: проверять гипотезу на уровне значимости 5%.

*Test size, size of a test*: размер теста (критерия).

*P-value, probability value*: *p*-значение.

*Marginal significance level*: граничный уровень значимости, *p*-значение.

*Power of a test*: мощность теста (критерия).

При использовании асимптотических аппроксимаций бывает трудно контролировать размер теста, так что реальная вероятность отклонения нулевой гипотезы (*actual rejection rate*) в конечных выборках может отличаться от номинальной вероятности (*nominal rejection rate*). При этом исследователь будет отклонять нулевую гипотезу слишком часто или слишком редко (*over-reject/under-reject*).

Nominal rejection rate: номинальная вероятность отклонения.

Nominal test size/significance level: номинальный размер/уровень значимости теста.

Actual (true) rejection rate: реальная (фактическая) вероятность отклонения.

Over-reject, overreject: отвергать нулевую гипотезу слишком часто (по сравнению с номинальным размером).

Under-reject, underreject: отвергать нулевую гипотезу слишком редко (по сравнению с номинальным размером).

### Confidence level

Близкая к проверке гипотез тема — интервальное оценивание (в общем случае — построение доверительных областей). Вероятность накрытия для доверительного интервала — это вероятность того, что интервал содержит (накрывает) оцениваемую величину при данном значении неизвестных параметров. Доверительный интервал строится таким образом, чтобы вероятность накрытия была равна некоторому уровню, называемому доверительным уровнем (коэффициентом доверия). Если вероятность накрытия зависит от неизвестных параметров, то доверительный уровень интервала приравнивается к наименьшей вероятности накрытия по всем возможным значениям параметров, так что вероятность накрытия при любом значении параметров заведомо не меньше доверительного уровня. Таким образом, coverage probability и confidence level близки по значению, но это не полные синонимы. Доверительная граница (доверительный предел) — это граница доверительного интервала.

Confidence interval: доверительный интервал.

Confidence region/set: доверительная область/доверительное множество.

Coverage probability: вероятность накрытия.

Confidence level: доверительный уровень, уровень доверительной вероятности.

Confidence coefficient: коэффициент доверия.

Confidence bound/limit: доверительная граница/доверительный предел.

Существуют также статистические термины, которые похожи на термин confidence interval, но имеют другое значение, в связи с чем может возникнуть путаница. В частности, это fiducial interval — один из давно известных видов интервального оценивания, но не очень распространенный из-за того, что применим только в очень специфических случаях.

Другой термин, который может вызвать путаницу, — это credible interval. Он относится к байесовскому интервальному оцениванию. Неискушенные в статистике люди часто неверно интерпретируют обычный классический доверительный интервал как такой фиксированный интервал, в который оцениваемая величина попадает с заданной вероятностью. Такая интерпретация скорее подходит для байесовского апостериорного интервала.

Fiducial interval: фидуциальный интервал.

Fiducial probability: фидуциальная вероятность.

Credible interval: правдоподобный интервал, байесовский апостериорный интервал.

Credible set: правдоподобное множество.

## Ad hoc

*Ad hoc* с латинского буквально переводится «для этого». В научных исследованиях иногда приходится делать предположения или использовать приемы, которые берутся не из какой-то общей теории или общего метода, а придумываются специально для данного случая. Для определения подобных приемов и используется термин *ad hoc*. Он может иметь как нейтральный, так и неодобрительный характер. Образованное от него существительное *ad hockery* не случайно похоже на *trickery* — «жульничество».

*Ad hoc*: специальный, устроенный/придуманный для данной цели, для конкретного случая, специфический, произвольный, необоснованный.

*Ad hoc assumption*: специфическое предположение, произвольное, необоснованное допущение.

*Ad hoc approach*: специальный подход, подход, использующий специфические или произвольные приемы.

*Ad hockery, adhockery, ad hocery*: приемы, правила, решения и т. д. предназначенные специально для данного случая.

## Intercept

*Intercept* буквально означает «пересечение». В эконометрике *intercept* означает точку, в которой линия регрессии пересекает ось  $Y$ . В более широком смысле — это параметр регрессии, равный координате, в которой гиперплоскость регрессии пересекает ось, соответствующую зависимой переменной. Синонимом является *constant* или *constant term*.

*Intercept*: параметр пересечения, константа.

*Constant term*: константа, постоянный член.

## Bayesian, frequentist

Статистические методы делятся на две группы: классические и байесовские. Идея байесовского метода состоит в том, что параметры модели являются случайными величинами, и у исследователя есть некоторые исходные (априорные) представления об их распределении. Эти априорные представления затем корректируются с учетом наблюдаемых данных, результатом чего являются апостериорные представления о распределении параметров. Эти методы называют байесовскими (*Bayesian*) из-за того, что в них часто используется теорема Байеса для условных вероятностей. Классические статистические методы в отличие от байесовских направлены на то, чтобы получать результаты, которые бы не зависели от априорного распределения. Классический подход называют также частотным (*frequentist*). Это связано с тем, что принято различать два взгляда на вероятность — частотный и субъективный (байесовский). (Следует однако понимать, что различие байесовской и классической статистики основано на несколько других принципах, чем различие байесовского и частотного взгляда на вероятность. Байесовская статистика не обязательно основывается на субъективных вероятностях, а классическая статистика не обязательно основывается на частотном определении вероятности.)

*Bayes' theorem*: теорема Байеса.

*Bayes' formula/law/rule/equation*: формула/закон/правило/уравнение Байеса.

Bayesian (прил.): байесовский, бейесовский.

Bayesian (сущ.): байесовец, сторонник байесовского подхода.

Bayesian approach: байесовский подход.

Classical approach: классический подход.

Frequentist: частотник, сторонник частотного подхода.

Frequentist approach: частотный подход.

## Hazard

Одно из ключевых понятий в анализе длительностей (дюраций) — это *hazard rate*. Это показатель вероятности наступления некоторого события при том, что это событие до текущего момента еще не наступило. Для непрерывного распределения длительностей

$$\lambda(d) = \frac{f(d)}{1 - F(d)} = \frac{f(d)}{S(d)},$$

где  $F(d)$  — функция распределения,  $f(d) = F'(d)$  — плотность распределения,  $S(d) = 1 - F(d)$  — так называемая *survival function*. Показатель  $\lambda(d)$  как функцию длительности  $d$  называют *hazard function*. Вместо *hazard function* часто используют просто *hazard*.

Данную терминологию трудно перевести на русский язык. Само слово *hazard* буквально означает «риск». Использование слова *hazard* подразумевает, что изучаемые события имеют неблагоприятный характер (например, при изучении продолжительности жизни это смерть). В теории надежности в качестве синонима используют слово *failure*, которое на русский переводят как «отказ». Имеющиеся в русскоязычной литературе переводы связаны с в основном с теорией надежности или статистикой продолжительности жизни. Например, *hazard rate* переводится как «интенсивность отказов».

Подробнее об этих и связанных с ними понятиях можно прочитать в методологической секции настоящего выпуска журнала «Квантиль». Также с данной темой можно познакомиться в Кокс & Оукс (1978), Wooldridge (2001) и Greene (2007).

Duration: длительность, продолжительность, дюрация.

Survival function, survivor function: функция выживаемости, функция надежности, функция долговечности.

Hazard rate, hazard ratio: интенсивность отказов.

Hazard function: функция риска, функция интенсивности отказов.

Proportional hazard model: модель пропорционального риска, модель пропорциональной интенсивности отказов.

Baseline hazard function: базовая функция риска.

## Proxy

*Proxy variable* или просто *proxy* — это такая переменная, которая сама по себе не представляет интереса, но которая заменяет другую переменную. В экономике часто вместо тех переменных, которые действительно интересны, используют их не совсем совершенные заменители.

Например, природные способности человека невозможно измерить непосредственно, но существуют различные показатели, такие как известный коэффициент IQ. ВВП на душу населения используют как показатель степени экономического развития страны или как показатель качества жизни. Об использовании переменных-заменителей в регрессии см. Wooldridge (2001) или Wooldridge (2008).

Proxy variable, проку: переменная-заменитель, суррогатная переменная, эрзац-переменная, прокси.

## Список литературы

Кокс, Д.Р. & Д. Оукс (1978). *Анализ данных типа времени жизни*. Москва: Финансы и статистика.

Крамер, Г. (1975). *Математические методы статистики*. Москва: Мир.

Greene, W.H. (2007). *Econometric Analysis* (6<sup>th</sup> edition). Prentice Hall.

Wooldridge, J. (2001). *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Wooldridge, J. (2008). *Introductory Econometrics: A Modern Approach* (4<sup>th</sup> edition). South-Western College Publishing.

# A mini-dictionary of English econometric terminology II

Alexander Tsyplakov

*Novosibirsk State University, Novosibirsk, Russia*

This part of the dictionary comments on English econometric terms kurtosis, skewness, critical region, significance level, confidence level, and some others. Emphasis is again placed on accurate definitions of their meaning to avoid possible confusion and incorrect interpretation.

# Обзор англоязычных учебников по анализу временных рядов\*

Станислав Анатольев†

*Российская экономическая школа, Москва, Россия*

Представлен обзор наиболее заметных учебников по эконометрике временных рядов. Эссе выражает как мнение автора, так и мнения эконометристов, выраженные в опубликованных рецензиях.

## Введение

С учебниками по анализу временных рядов дело обстоит сложнее, чем с общеэконометрическими. Если в хорошем учебнике по эконометрике материала хватает, чтобы всесторонне познакомиться с предметом, то чтобы адекватно освоить модели и методы временных рядов, одной книги маловато. Удачно подобранные источники дополняют друг друга, но и этого часто недостаточно, и необходимо обращаться к обзорным (в лучшем случае) статьям в экономических журналах и в томах *Handbook of Econometrics*. Не стоит также сбрасывать со счетов некоторые удачные разделы в общеэконометрических учебниках (см. обзор в Анатольев, 2007), как, например, главы, посвященные единичным корням и коинтеграции у Хайаши (Hayashi, 2000).

Еще одна проблема состоит в том, что даже самые замечательные учебники потихоньку (если не сказать с бешеной скоростью) устаревают. Современные исследования, касающиеся временных рядов, идут до того интенсивными темпами, что ни один учебник «не поспевает» за текущей ситуацией.

В данном эссе мы приводим краткий обзор наиболее заметных книг и учебников, касающихся методологии временных рядов. Мы включили только те источники, которые покрывают всю область или ее большую часть и не специализируются на узких аспектах (например, есть монографии, всецело посвященные сезонности и периодичности, или спектральному анализу, или рядам с единичными корнями). Кроме того, мы упоминаем источники, содержащие именно *эконометрику* временных рядов. Так, например, мы включили в список книгу Миллза, но не книгу Тэйлора (Taylor, 1986) с практически идентичным названием, а также проигнорировали массу учебников по *статистическому* анализу временных рядов.

**James D. Hamilton. Time Series Analysis. Princeton University Press, 1994, 820 стр.**

Этот «главный» учебник покрывает практически весь материал, имеющий отношение к временным рядам, особенно линейный анализ. Главы, посвященные, например, оцениванию линейных ARMA-моделей, асимптотике моделей с единичными корнями, векторным авторегрессиям, являются, пожалуй, одними из наилучших примеров в эконометрической литературе.

\* Автор благодарит Александра Цыплакова за полезные замечания. Цитировать как: Анатольев, Станислав (2008) «Обзор англоязычных учебников по анализу временных рядов», Квантиль, №5, стр. 49–55. Citation: Anatolyev, Stanislav (2008) “Review of English textbooks in time series analysis,” *Quantile*, No.5, pp. 49–55.

† Адрес: 117418, г. Москва, Нахимовский проспект, 47, офис 1721(3). Электронная почта: [sanatoly@nes.ru](mailto:sanatoly@nes.ru)

Будучи несколько устаревшей, правда, книга не отражает тенденции последних лет в эконометрике. Так, например, главы по ARCH-моделированию и обобщенному методу моментов откровенно куцые, практически отсутствует материал по анализу структурных сдвигов и нелинейных регрессионных моделей (например, пороговых авторегрессий). Удивительной неполнотой обладают главы по спектральному анализу и Байесовским методам. В то же время в книге можно найти материал, редко встречающийся в неспециализированных учебниках, например фильтр Кальмана и модели с переключениями режимов (последняя тема – конек автора).

По мнению видного эконометриста Брюса Хансена (Hansen, 1995), книга «продельывает отличную работу по смыканию теории и техники... Это, однако, не учебник теории... Целью является дать прикладным экономистам знания, необходимые для овладения современными технологиями манипулирования временными рядами. Я думаю, книга хороша по этим критериям и будет широко использоваться как учебник». Далее: «Ее сила – в основательном разборе традиционных временных рядов и регрессий и аккуратном анализе текущей литературы по единичным корням и коинтеграции». В то же время Хансен указывает и на недостатки, которые, впрочем, довольно очевидны и частично упоминались выше: «Теоретику книга может показаться недостаточно строгой»; «Хотя книга написана для прикладной аудитории, в ней очень мало эмпирических примеров»; «Недостаточно обсуждение выбора модели, компоненты, являющейся существенной для повседневной эмпирической работы»; «Нелинейные модели временных рядов почти не обсуждаются... Обсуждение ARCH-моделей очень краткое». Хансен советует дополнять чтение книги просмотром обзорных статей из различных источников. В целом Хансен называет учебник отличным и с энтузиазмом рекомендует его изучение.

Однако есть и другое мнение. Эконометрист Бенедикт Петчер в своем обзоре (Pötscher, 1996) перечисляет неверные утверждения, ошибки и прочие несуразности, которые он нашел в книге. Кроме того, он жалуется на нестрогость повествования, отсутствие обсуждения некоторых тем (как действительно актуальных, так и не слишком актуальных), не совсем корректное цитирование, а также методологические недостатки вроде неровности подачи материала по глубине и продвинутости. В конце концов Петчер заключает: «Ценность этой книги сильно уменьшена множеством серьезных перечисленных мной недостатков, так что трудно рекомендовать ее в качестве учебника». Во многом Петчер, конечно, прав, но, по нашему мнению, он сильно сгущает краски. Многие из упомянутых недостатков совершенно не мешают постигать важный материал, а некоторые (например, отсутствие строгости, вольное цитирование и т.д.) даже помогают.

Мы же главным недостатком учебника Гамильтона считаем ее устарелость. Конечно, этот недостаток не мог быть актуален для Хансена и Петчера, опубликовавших свои рецензии в 1995–1996 гг.

У издателя имеется сайт учебника [pup.princeton.edu/titles/5386.html](http://pup.princeton.edu/titles/5386.html), хотя он не слишком информативен. Программы для пакета *RATS* и данные для имеющихся в книге эмпирических примеров находятся на сайте [www.estima.com/Hamilton's Time Series Analysis.shtml](http://www.estima.com/Hamilton's%20Time%20Series%20Analysis.shtml).

### **Philip H. Franses & Dick van Dijk. Nonlinear Time Series Models in Empirical Finance. Cambridge University Press, 2000, 296 стр.**

Книга Франсеса и Ван Дайка является замечательным дополнением к учебнику Гамильтона. Здесь освещены именно те важные вопросы, которые отсутствуют или недостаточно представлены (возможно, в силу устарелости) у Гамильтона: нелинейные регрессии, модели для условной гетероскедастичности, непараметрические методы (точнее, один из них). Причем, вопреки названию, многие модели вполне годятся для анализа и нефинансовых (например, макроэкономических) данных.

Первые две главы книги представляют собой краткий обзор основных концепций анализа временных рядов и особенно финансовых данных, таких как единичные корни, сезонность, прогнозирование, выбросы и т.д. Далее идут две «основные» главы, одна из которых обобщает нелинейные модели для (условно говоря) доходностей, в частности, пороговые регрессии, модели с Марковскими переключениями и модели с гладкими переходами, а другая – модели для волатильности, в основном ARCH-типа. Еще одна массивная глава посвящена нейронным сетям. С одной стороны, радует, что данный материал изложен системно с эконометрических позиций, но, с другой стороны, вызывает недоумение, почему авторы так выделяют из непараметрических и полупараметрических моделей именно нейронные сети.

Одно из главных достоинств книги – обилие эмпирических примеров и возможность их реплицировать. Все программы написаны на эконометрическом языке *GAUSS*, хотя, исходя из нашего опыта, при попытке запуска они часто ругаются и с трудом подвергаются модификации.

У книги есть сайты: у издателя [www.cambridge.org/catalogue/catalogue.asp?isbn=0511034083](http://www.cambridge.org/catalogue/catalogue.asp?isbn=0511034083) и у одного из авторов [people.few.eur.nl/djvandijk/nltsmef/nltsmef.htm](http://people.few.eur.nl/djvandijk/nltsmef/nltsmef.htm). Со второго сайта можно скачать использованные в книге данные и программы.

**Walter Enders. Applied Econometric Time Series. Wiley, 2-е издание, 2004, 460 стр.**

Книга Эндерса представляет собой «легковесный» аналог Гамильтона, и может быть полезна технически менее искушенным читателям. Спектр охваченных тем также более узок, чем у Гамильтона. Достаточно сказать, что в первое издание вошли только линейные модели (если не считать модели из ARCH-класса), и только во втором издании появилась глава, посвященная нелинейному моделированию. Она, впрочем, представляет собой скорее краткий обзор нелинейного анализа временных рядов и далека до соответствующих глав монографии Франсеса и Ван Дайка.

Конечно, «легковесность» книги означает и неглубокое изложение материала. На множество современных важных методов, если они сколь-нибудь нестандартны, автор лишь ссылается в заключительных комментариях к каждой главе. Наличие таких комментариев, правда, само по себе представляет ценность, как и наличествующие эмпирические примеры, а также теоретические и практические упражнения.

Подытоживая, учебник Эндерса будет весьма полезен тем, кто хочет освоить эконометрику временных рядов практически с нуля. В то же время изучившему книгу не стоит обольщаться по поводу вооруженности современными методами анализа временных рядов.

У книги имеется сайт [cba.ua.edu/~wenders/applied-econometric-time-series](http://cba.ua.edu/~wenders/applied-econometric-time-series), где можно найти данные, использованные в книге, и список замеченных ошибок. Есть также сайт у издателя [eu.wiley.com/WileyCDA/WileyTitle/productCd-0471451738.html](http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471451738.html), но он скуден. Существует, кроме того, пособие для преподавателя; см. Enders & Chung (1995) и Chung, Enders, Shao & Yuan (2004).

**Peter J. Brockwell & Richard A. Davis. Time Series: Theory and Methods. Springer-Verlag, 2-е издание, 1991, 577 стр.**

Данная книга – энциклопедия линейного анализа временных рядов, со статистическим уклоном. По ней не стоит изучать предмет, особенно эконометристу, но иногда полезно использовать как справочник. Современные изыскания в книге не отражены; нелинейному анализу уделено крайне мало страниц в конце книги. В то же время странно, что, например, вовсе не затронута тема тестирования на единичные корни, не говоря уже о коинтеграции. Сильный акцент делается на спектральном анализе, что естественно для анализа линейных

процессов. Присутствует также глава, посвященная фильтру Кальмана, и, что интересно, глава, посвященная Гильбертовым пространствам. Последняя служит цели геометрической иллюстрации многих идей на протяжении всего повествования.

В приложении книги содержатся таблицы с данными, использованными в эмпирических примерах. Каждая глава сопровождается набором задач, правда, без решений.

**Christian Gourieroux & Alain Monfort. Time Series and Dynamic Models. Cambridge University Press, 1997, 668 стр.**

Данное произведение – одна из многочисленных книг-учебников Гурьеру, написанных «на коленке». Последний факт не означает, конечно, бесполезность книги – продвинутый читатель может почерпнуть немало полезного: например, имеется материал по причинности, экзогенности и формированию ожиданий. Но рекомендовать ее в качестве учебника вряд ли целесообразно.

Вот что пишет Колин Маккензи в своей рецензии (McKenzie, 1998): «Слишком силен акцент на теории и формулировании результатов в виде теорем, и существенной мотивации результатов и идей местами не хватает. В то время как некоторые методы иллюстрируются с помощью эмпирических примеров, даже продвинутому обучающимся часто будет непонятно, как реализовать некоторые из процедур на практике». Маккензи также отмечает специфичность используемых данных, отсутствие логической связи между некоторыми главами и тенденциозность в подборе материала, а в заключение делает вывод, что вряд ли эта монография будет конкурентом ведущим учебникам по временным рядам.

Каждая глава книги сопровождается набором задач без решений. Книги имеют сайт у издателя [prp.contentdirections.com/mr/cupress.jsp/doi=10.2277/0521411467](http://prp.contentdirections.com/mr/cupress.jsp/doi=10.2277/0521411467).

**Philip H. Franses. Time Series Models for Business and Economic Forecasting. Cambridge University Press, 1998, 280 стр.**

Данная монография – краткий, и возможно, слишком краткий путеводитель по современным эконометрическим методам анализа временных рядов, с акцентом на прогнозировании. В какой-то степени здесь повторяются многие главы книги Франсеса и Ван Дайка, некоторые с сокращением материала, некоторые с расширением. Много места уделено таким темам, как тренды, сезонность и, что является редкостью, анализу выбросов.

Книга написана четким и понятным языком и может служить хорошим источником для исследователя, желающего быстро, пусть и неглубоко, освоить современные методы эконометрики временных рядов. Ганс Раймерс (Reimers, 2000) отмечает: «Для каждого метода детально рассматриваются интуитивная мотивация и практические соображения, что делает книгу легко читаемой. Жизненные примеры, взятые из научных приложений, использованы при обсуждении важных вопросов в прикладном анализе деловых, финансовых и макроэкономических данных». И далее: «Книга будет полезна для студентов и преподавателей прикладных курсов, а также для практиков, желающих получить первое, не очень техническое, впечатление о прогнозировании временных рядов, используя современные, недавно разработанные методы».

С сайта автора [people.few.eur.nl/franses/#research](http://people.few.eur.nl/franses/#research) можно скачать использованные в эмпирических примерах данные.

**Daniel Peña, George C. Tiao & Ruey S. Tsay (редакторы). A Course in Time Series Analysis. Wiley, 2001, 496 стр.**

Данная книга – не учебник, а собрание статей по анализу временных рядов, написанных разными авторами, как эконометристами, так и статистиками, как именитыми, так и не очень. Конечно, по сравнению с учебником материал здесь не очень сбалансирован, зато покрывает максимальное количество тем и, что самое главное, относительно современен. Книга будет наиболее полезна желающему глубоко разобраться в какой-то теме и наименее полезна (а может, и вредна) желающему составить для себя общую картину эконометрики временных рядов.

У одного из редакторов книги имеется сайт [faculty.chicagogsb.edu/ruey.tsay/teaching/ecas](http://faculty.chicagogsb.edu/ruey.tsay/teaching/ecas). На нем, в частности, находятся данные, использованные в трех главах книги.

**Ruey S. Tsay. Analysis of Financial Time Series. Wiley, 1-е издание, 2002, 448 стр.; 2-е издание, 2005, 640 стр.**

Данный учебник посвящен анализу финансовых временных рядов, и эта специфика получила сильное отражение в его содержании. Наряду с «обязательными» темами детально рассматриваются и такие, как модели для высокочастотных финансовых данных, модели в непрерывном времени и стоимостная мера риска. Во втором издании присутствуют еще и метод главных компонент, факторные модели и фильтр Кальмана.

Экономист Йюсси Толви (Tolvi, 2003) отмечает: «... Половина глав строятся на финансовой теории, а другая половина – на статистике. Несколько не хватает интегрированности этих двух точек зрения...» И далее: «К сожалению, это не наилучшим образом написанная книга из тех, что я читал. Хотелось бы увидеть более широкое обсуждение интерпретации некоторых моделей». Тем не менее, нам кажется, что данная монография – один из самых удачных учебников для изучения.

У книги имеются два сайта на странице автора: [faculty.chicagogsb.edu/ruey.tsay/teaching/fts](http://faculty.chicagogsb.edu/ruey.tsay/teaching/fts) посвящен первому изданию, [faculty.chicagogsb.edu/ruey.tsay/teaching/fts2](http://faculty.chicagogsb.edu/ruey.tsay/teaching/fts2) – второму. На сайтах находятся данные, использованные в книге, а также списки выявленных ошибок и опечаток.

**Terence C. Mills. The Econometric Modelling of Financial Time Series. Cambridge University Press, 2-е издание, 1999, 280 стр.**

Данное пособие также посвящено анализу финансовых временных рядов. По сравнению с монографией Тцзя, правда, материал выглядит несколько староватым. Хотя и имеются разделы, посвященные нелинейным моделям, описание методов представляется довольно куцым. А в главе, где делается похвальная попытка обсудить моделирование всего распределения доходностей (гл. 5), автор не заходит слишком далеко.

Эконометрист Питер Педрони в целом книгу хвалит (Pedroni, 2001): «Обсуждение явно направлено на приложения, а не на теорию, и большинство читателей найдут уровень технической детализации меньшим, чем у других учебников типа Гамильтона... Однако, это не книга рецептов. Наоборот, акцент делается на построения концепций и интуиции на протяжении всего повествования. Где необходимо, без стеснения вводятся и более продвинутые концепции... В то же время, чрезвычайно технические детали, не помогающие пониманию методов, избегаются». Педрони заявляет, что ему не хватает в книге обсуждения обобщенного метода моментов и методов оценивания долгосрочной дисперсии. По нашему же мнению, книге не хватает очень много чего. Лучше всего это можно возместить чтением монографий Тцзя и Франсеса и Ван Дайка.

У издателя имеется сайт книги [www.cambridge.org/catalogue/catalogue.asp?isbn=0521624134](http://www.cambridge.org/catalogue/catalogue.asp?isbn=0521624134).  
Данные можно скачать с сайта [lboro.ac.uk/departments/ec/cup](http://lboro.ac.uk/departments/ec/cup).

## Заключение

По традиции, в заключение мы приводим сводную таблицу, дающую некоторое представление о содержании в данных учебниках некоторых тем: «+» означает «имеется», «-» – «отсутствует», «+-» – тема раскрыта, но далеко не полностью, а «-+» – тема затронута, но очень кратко или неадекватно. Вряд ли стоит воспринимать информацию в таблице как справочную, ибо она не отражает ни качества предоставляемого учебниками материала, ни современность подходов, но все же по ней можно судить об их тематическом наполнении. Кроме того, некоторые книги ценны именно благодаря тем редким узким темам, которые не отражены в таблице.

Тема	H	FD	E	BD	GM	F	PTT	T	M
Эконометрические методы в целом	+	-	-	-+	-	-	-	-	-
Асимптотическая теория	+	-	-	+	-	-	-	-	-
Выбор модели	-	+	-+	-+	-+	+	+	-	-
ARMA моделирование	+	-+	+	+	+	+	+	+	+
Пороговые и другие нелинейные авторегрессии	-	+	+	-+	-	+	+	+	+-
Переключения режимов	+	+	+	-	-	+	-	+	+-
Векторные авторегрессии	+	-	+	+	-+	+	+	+	+
ARCH моделирование	-+	+	+	-	-	+	+	+	+
Вопросы сезонности	-	-	+	+	+	+	+	+	-
Анализ выбросов	-	+	-	-	-+	+	+	+	+-
Структурные сдвиги	-+	-	-+	-	-	-	-	-	-
Единичные корни	+	-	+	-	+	+	-	+	+
Коинтеграция	+	-	+	-	+	+	+	+	+
Непараметрические и полупараметрические методы	-	+-	-	-	-	-+	+	+	-+
Фильтр Кальмана	+	-	-	+	+	-	+	+	-+
Спектральный анализ	+-	-	-	+	+	-	+	-	-
Эмпирические иллюстрации	-+	+	+	+	-+	+	-+	+	+

## Список литературы

- Анатольев, Станислав (2007). Обзор англоязычных учебников по эконометрике. *Квантиль* 3, 73–82.
- Enders, W. & P. Chung (1995). Instructor's Manual to Accompany *Applied Econometric Time-Series*. John Wiley & Sons: New York. Второе издание: Chung, P., W. Enders, L. Shao & J. Yuan (2004).
- Hansen, B. (1995). *Time Series Analysis* by James D. Hamilton. *Econometric Theory* 11, 625–630.
- Hayashi, F. (2000). *Econometrics*. Princeton University Press: Princeton.
- McKenzie, C. (1998). *Time Series and Dynamic Models* by Christian Gourieroux & Alain Monfort. *Journal of Applied Econometrics* 13, 681–684.
- Pedroni, P. (2001). *The Econometric Modelling of Financial Time Series* by Terence Mills. *Journal of American Statistical Association* 96, 345–346.
- Pötscher, B.M. (1996). *Time Series Analysis* by James D. Hamilton. *Journal of American Statistical Association* 91, 439–440.
- Reimers, H.-E. (2000). *Time Series Models for Business and Economic Forecasting* by Philip Hans Franses. *Journal of American Statistical Association* 95, 686–687.

Taylor S.J. (1986). *Modelling Financial Time Series*. John Wiley & Sons: Chichester.

Tolvi, J. (2003). *Analysis of Financial Time Series* by R. S. Tsay. *The Statistician* 52, 128–129.

## Review of English textbooks in time series analysis

Stanislav Anatolyev

*New Economic School, Moscow, Russia*

This is a survey of most notable time series econometrics texts written in English. The essay reflects the author's opinion, as well as opinions of econometricians expressed in published book reviews.



# Задачи и решения

Начиная с настоящего выпуска, мы вводим новую рубрику «Задачи и решения». Здесь будут публиковаться эконометрические задачи, представляющие образовательный или исследовательский интерес, а также решения к ним. Задачи рассчитаны не на профессионалов-эконометристов, а скорее на изучающих эконометрику, а также прикладных исследователей, желающих понять некоторые тонкости теории. Читатель имеет возможность в течение трех месяцев с момента выхода выпуска присылать решения задач на адрес [ps@quantile.ru](mailto:ps@quantile.ru). Наиболее интересные ответы будут публиковаться. Принимаются также интересные постановки задач.

## Задача 5.1

Могут ли две случайные величины быть некоррелированными безусловно, но коррелированными условно на третьей? Могут ли две случайные величины быть коррелированными безусловно, но некоррелированными условно на третьей?

## Задача 5.2

Известно, что для простейшей авторегрессии с независимыми и одинаково распределенными инновациями

$$y_t = \rho y_{t-1} + \varepsilon_t$$

в случае единичного корня  $\rho = 1$  МНК-оценка  $\hat{\rho}$  состоятельна для единицы и имеет распределение Дики–Фуллера

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

где  $B(r)$  – стандартный Винеровский процесс на  $[0, 1]$ . Пусть по недоразумению вместо того, чтобы регрессировать  $y_t$  на  $y_{t-1}$ , мы регрессируем  $y_{t-1}$  на  $y_t$ . Выведите асимптотическое распределение полученной МНК-оценки при наличии единичного корня. Состоятельна ли эта оценка для единицы?

## Задача 5.3

Пусть скалярные случайные величины  $x$  и  $y$  имеют одно и то же математическое ожидание  $\mu$ . Покажите, что тест Хаусмана на верность условия на моменты  $\mathbb{E}[y] = \mu$  при верности условия на моменты  $\mathbb{E}[x] = \mu$  асимптотически эквивалентен J-тесту на верность модели, состоящей из обоих условий на моменты. Каково интуитивное объяснение этого результата?



# Статьи: прикладная эконометрика

## Оценивание уровня технологического прогресса в российской экономике\*

Евгения Назруллаева<sup>†</sup>

*Высшая школа экономики, Москва, Россия*

Оценивание индикаторов технологического прогресса является одной из ключевых задач в теории роста. Общепринятым индикатором технического прогресса в литературе является совокупная факторная производительность (СФП), которая представляет собой экзогенно заданный, не объясняемый за счет динамики факторов производства остаток от роста выпуска. В существующих исследованиях при анализе СФП для российской экономики, как правило, основной акцент делается на период трансформационного спада, характеризующийся сильным падением выпуска, которое объясняется снижением СФП, т.е., по сути, техническим «регрессом» в экономике. В данной работе СФП оценивается на основе данных Федеральной службы государственной статистики и Российского экономического барометра по 12 отраслям промышленности в 1995–2004 гг. Оценки СФП, скорректированные на основе наблюдаемых индикаторов загрузки производственных мощностей и с учетом специфики российских статистических данных, свидетельствуют о постепенном улучшении технологий, сопровождающем рост выпуска после 1998 г., и о наличии отрицательного технологического шока до 1998 г.

*Ключевые слова:* совокупная факторная производительность, технологический прогресс, уровень загрузки производственных мощностей, отрасли российской промышленности

*Классификация JEL:* C22, E22, O33

### 1 Введение

Оценивание уровня технологического прогресса тесно связано с изучением природы экономического роста. В прикладной теории роста, основанной на подходе Солоу к измерению технологического прогресса (Solow, 1957), предполагается, что технологический прогресс определяется экзогенно как остаток от роста выпуска, не объясняемого динамикой факторов производства. Этот необъяснимый остаток или совокупная факторная производительность (СФП) в литературе используется в качестве прокси, позволяющей судить об уровне технологического прогресса в экономике. Подходы к оцениванию СФП в большинстве случаев разработаны и используются для изучения технологического прогресса в США и странах Европы. Первые оценки СФП для российской экономики появляются в литературе с 2000-х годов.

Оценивание СФП во многом связано с уровнем агрегирования, который рассматривается при построении соответствующих индикаторов. При анализе СФП на уровне всей экономики полученные индикаторы не учитывают неоднородность фирм в отраслях и неоднородность самих отраслей. В сравнительно новых работах (например, Basu, Kimball & Fernald, 2006),

\*Цитировать как: Назруллаева, Евгения (2008). «Оценивание уровня технологического прогресса в российской экономике», Квантиль, №5, стр. 59–82. Citation: Nazrullaeva, Eugenia (2008). “Measurement of technological progress in Russia,” *Quantile*, No.5, pp. 59–82.

<sup>†</sup>Адрес: 109028, Россия, Москва, Покровский бульвар, 11. Электронная почта: [e.nazrullaeva@gmail.com](mailto:e.nazrullaeva@gmail.com)

посвященных изучению СФП, резюмируется основной подход к оцениванию, и предлагаются способы улучшения свойств стандартных индикаторов СФП.

Мотивацию, лежащую у истоков написания данной работы, можно резюмировать в следующих вопросах:

1. Сопровождается ли рост выпуска в российской экономике, наблюдаемый в последние несколько лет, улучшением технологий? Какова динамика технологического прогресса до и после 1998 г.?
2. Характеризуются ли отрасли российской промышленности возрастающей отдачей от масштаба и наличием несовершенной конкуренции, подразумевающей монопольную власть и возможность извлечения положительной прибыли?

Первый вопрос тесным образом связан с результатами существующих исследований, посвященных анализу динамики СФП в России: отраслевые темпы роста технологического прогресса и темпы роста СФП для всей российской экономики в 1990-х гг. отрицательны. Отрицательный темп роста (или отрицательный шок технологий) означает, по сути, ухудшение используемых в производстве технологий. В данной работе предложен альтернативный подход к оцениванию СФП, призванный проверить, действительно ли период трансформационного спада в российской экономике и последующий кризис 1998 г. сопровождались негативным технологическим шоком. Что если снижение СФП, которое наблюдается согласно эмпирическим оценкам, связано вовсе не с ухудшением технологий? С этой целью в работе оценивается индикатор СФП на уровне отраслей российской промышленности, скорректированный с учетом неполного использования производственных мощностей, в то время как стандартный индикатор СФП, предложенный Солоу, предполагает наличие постоянной отдачи от масштаба и не учитывает, что загрузка производственных мощностей может меняться с течением времени. Второй вопрос продиктован интересом к ситуации, складывающейся в отраслях российской промышленности: в 1990-х годах во время трансформационного спада и впоследствии многие предприятия несли убытки, и их прибыль была отрицательна, что идет вразрез с предположением о возрастающей отдаче. Начиная с 2000-х гг. развитие отраслей было довольно неоднородным: интенсивное развитие одних отраслей (в частности, топливной промышленности) сопровождалось ухудшением положения в других отраслях (например, в легкой промышленности).

В основе подходов к оцениванию СФП в большинстве случаев так или иначе лежит концепция производственной функции, т. е. предполагается, что существующие в экономике технологии позволяют производить максимально возможный уровень выпуска при фиксированной комбинации факторов производства. Другими словами, фирмы в экономике считаются полностью эффективными и функционирующими на границе своих производственных возможностей. Однако на практике данная предпосылка вряд ли выполнена. Скорее можно ожидать, что фирмы оперируют внутри своих производственных множеств. В подходе, который используется в данной работе, проблема неэффективности частично решается за счет дополнительного предположения об изменяющейся во времени загрузке производственных мощностей: фирмы могут не полностью использовать имеющиеся у них в наличии производственные мощности, что было особенно характерно для российской экономики в первой половине 1990-х гг. При оценивании индикаторов технологического прогресса вслед за Basu, Kimball & Fernald (2006) в данной работе контролируются эффекты, связанные с изменяющимся во времени уровнем загрузки производственных мощностей, несовершенной конкуренцией (предполагается непостоянная отдача от масштаба) и наличием в производственной функции дополнительного фактора производства (затрат сырья и материалов). Контроль данных эффектов позволяет элиминировать влияние нетехнологических факторов, связанных с динамикой факторов производства. Динамика СФП анализируется на дезагрегированном уровне по 12 отраслям российской промышленности в период с 1 квартала 1995 г. по 4

квартал 2004 г. в соответствии с отраслевым классификатором ОКОНХ: электроэнергетика (11100); нефтедобывающая промышленность (11210); нефтеперерабатывающая промышленность (11220); газовая промышленность (11230); угольная промышленность (11300); черная металлургия (12100); цветная металлургия (12200); химическая и нефтехимическая промышленность (13000); машиностроение и металлообработка (14000); лесная, деревообрабатывающая и целлюлозно-бумажная промышленность (ЛДЦБ) (15000); легкая промышленность (17000); пищевая промышленность (18000).

Работа структурирована следующим образом. В разделе 2 представлены обзор литературы и основные результаты существующих исследований, которые необходимо учитывать при моделировании СФП с учетом выбранного подхода. Раздел 3 характеризует непосредственно модель, на основе которой строятся индикаторы технологического прогресса для отраслей. В разделе 4 приведено описание данных, а также анализируются основные проблемы, связанные с измерением капитала и уровня загрузки производственных мощностей, и проблемы технического характера, имеющие отношение непосредственно к процедуре оценивания СФП. Полученные эмпирические результаты и их возможная экономическая интерпретация рассматриваются в разделе 5. В разделе 6 представлены основные выводы.

## 2 Обзор литературы

Целью данной работы является анализ технологического прогресса в российской экономике. На сегодняшний день существует ряд исследований, в которых изучается взаимосвязь экономического роста, технологического прогресса и методов оценивания СФП. Подход к оцениванию технологического прогресса, который используется в данной работе, основан на оценке совокупной факторной производительности, предложенный Солоу (Solow, 1957). В соответствии с походом Солоу, СФП измеряет эффективный уровень технологий, позволяющий трансформировать факторы производства в конечную продукцию, т.е. подход основан на концепции производственной функции. Стилизованным фактом является процикличность роста СФП в краткосрочном периоде, т.е. динамика выпуска и динамика СФП положительно коррелированы. Процикличный характер динамики СФП объясняется положительным влиянием шоков спроса, которые увеличивают одновременно объем производства и объем потребляемых в производстве факторов, а также влиянием уровня загрузки производственных мощностей, т.е. без учета загрузки мощностей труда и капитала динамика СФП положительно коррелирует с динамикой выпуска. Исходя из этих соображений, предлагается модифицировать стандартный индикатор технологического прогресса, предложенный Солоу. Предпосылками СФП Солоу являются совершенная конкуренция и постоянная отдача от масштаба, также не учитывается динамика уровня загрузки производственных мощностей. Элиминируя влияние перечисленных нетехнологических факторов, можно получить так называемый «чистый» индикатор СФП, который и будет характеризовать улучшение технологий как таковое.

Данная работа основана на подходе, предложенном в работе Basu, Kimball & Fernald (2006), где анализируются модифицированные отраслевые индикаторы СФП для США в рамках многофакторного KLEMS-подхода (аббревиатура расшифровывается как капитал K, труд L, электроэнергия E, материальные затраты M и услуги S), т.е. предполагается наличие дополнительных факторов, помимо капитала и труда. Basu, Kimball & Fernald (2006) учитывают влияние несовершенной конкуренции и возрастающей отдачи от масштаба, а также динамику уровня загрузки производственных мощностей капитала и труда. Предпосылка о наличии возрастающей отдачи от масштаба позволяет моделировать, во-первых, непропорциональное влияние динамики факторов производства на динамику выпуска, во-вторых, внешние эффекты, т.е. взаимосвязь, существующую между отраслями. Предположение о меняющемся во времени уровне загрузки мощностей необходимо для того, чтобы учесть воз-

можную недозагрузку производственных мощностей, которая ведет к смещенным оценкам капитала и труда. Согласно Basu, Kimball & Fernald (2006), индикатор СФП, очищенный от влияния нетехнологических факторов, контрцикличен, т.к. улучшение технологий должно проявляться в снижении загрузки мощностей (технологический прогресс предполагает использование менее капиталоемких технологий). Авторы оценивают индикаторы технологического прогресса для 29 отраслей (исключая сельское хозяйство и добывающие отрасли промышленности) за 1949–1996 гг., а затем агрегируют полученные отраслевые оценки и строят агрегированный индикатор СФП для экономики США. В отличие от результатов, полученных в Hall (1990), Basu, Kimball & Fernald (2006) не находят подтверждения гипотезе о наличии возрастающей отдачи от масштаба (отдача скорее постоянная). Сравнивая динамику полученного индикатора технологического прогресса с динамикой стандартного СФП-индикатора Солоу, авторы приходят к выводу, что остаток Солоу более волатилен из-за неучтенного влияния нетехнологических факторов.

В работе Inklaar (2005) анализируются потенциальные недочеты подхода Basu, Kimball & Fernald (2006). Вслед за Basu, Kimball & Fernald (2006), Inklaar (2005) использует логлинеаризованную производственную функцию для оценивания параметров отдачи от масштаба на отраслевом уровне для Франции, Германии, Нидерландов и США за 1979–2003 гг. (отрасли агрегируются до секторов: товары длительного пользования, товары кратковременного пользования и не обрабатываемые отрасли). Inklaar (2005) оценивает СФП без учета загрузки мощностей и СФП с учетом ненаблюдаемого уровня загрузки. Согласно Basu, Kimball & Fernald (2006), учет загрузки производственных мощностей ведет к более низким оценкам параметров отдачи от масштаба. В работе Inklaar (2005) полученные результаты действительно свидетельствуют о более низких оценках отдачи от масштаба для Нидерландов и США. Для данных стран гипотеза о наличии в отраслях постоянной отдачи от масштаба не отвергается. Однако для отраслей Франции и Германии характерна значимая возрастающая отдача от масштаба. Более того, предложенная в Basu, Kimball & Fernald (2006) прокси для ненаблюдаемого уровня загрузки производственных мощностей – часы, отработанные в среднем одним работником, – имеет значимое положительное влияние только для отраслей США. В других странах оценка параметра при уровне загрузки мощностей, как правило, отрицательно (причем часто незначимо) влияет на динамику выпуска. Гипотеза Inklaar (2005) заключается в том, что европейские фирмы реагируют на колебания спроса путем изменения числа рабочих, а не отработанных ими часов. Таким образом, подход Basu, Kimball & Fernald (2006) довольно специфичен и, скорее всего, в своем исходном виде применим только для экономических реалий США. Применяя данный подход для российских данных, необходимо дополнительно учитывать специфику отечественной статистики, в особенности при выборе прокси для уровня загрузки производственных мощностей.

Существующие исследования СФП для России можно классифицировать по крайней мере по двум направлениям, в зависимости от того, на каком уровне анализируется динамика технологического прогресса – агрегированном или отраслевом. На агрегированном уровне СФП оценивается в работах Dolinskaya (2002), Воскобойников (2003) и Бессонов (2004).

В работе Dolinskaya (2002) анализируются основные факторы, определяющие динамику агрегированного выпуска в 1991–2007 гг. Падение выпуска в период трансформационного спада объясняется изменениями в динамике факторов производства через динамику уровня загрузки производственных мощностей. При этом в качестве прокси для уровня загрузки мощностей рассматривается оценка, пропорциональная динамике агрегированного выпуска.

Из-за наличия существенных недочетов в статистике по основным фондам, связанных с динамикой относительных цен инвестиционных товаров и отсутствием хороших прокси для основных фондов, которые выбывают из процесса производства при отрицательных шоках спроса, в работе Воскобойников (2003) предлагается модифицировать исходные данные по основным фондам с тем чтобы учесть возрастную структуру капитала. Подход, который

взят в Воскобойников (2003) за основу, был предложен в работе Meinen, Verbiest & Wolf (1998). Он позволяет получать временные ряды по основным фондам, используя данные по валовому накоплению основного капитала (ВНОК) за все предшествующие текущему периоду и концепцию дожития активов. Из динамики основных фондов, оценки для которых получены в работе Воскобойников (2003) за 1961–2001 гг., видно, что снижение выпуска в период трансформационного спада сопровождалось параллельным снижением накопления основного капитала, вызванного снижением спроса на отдельные виды производимых в экономике товаров. Основные фонды в соответствующих отраслях перестали использоваться в производстве и постепенно устаревали. Таким образом, трансформационный спад, согласно работе Воскобойников (2003), объясняется в большей степени снижением эффективности производства как таковым, нежели падением уровня загрузки производственных мощностей, как это предполагается в работе Dolinskaya (2002).

В работе Бессонов (2004) выделены основные особенности российской экономики, связанные с технологическим прогрессом: интенсивная динамика агрегированного выпуска после 1990-х гг. сопровождается довольно инертной динамикой капитала и труда. В результате динамика выпуска слабо связана с динамикой соответствующих факторов производства. Учет уровня загрузки производственных мощностей позволяет элиминировать эффект недогрузки мощностей, характерный для трансформационного периода. Динамика агрегированного индикатора СФП и отраслевых показателей технологического прогресса по 10 отраслям российской промышленности свидетельствует о резком снижении СФП в период 1997–1998 гг., после чего уровень технологий начинает постепенно повышаться. Иначе говоря, снижение выпуска в экономике, согласно результатам в Dolinskaya (2002), Воскобойников (2003) и Бессонов (2004), сопровождалось так называемым технологическим «регрессом» в экономике. Учет уровня загрузки мощностей в работе Бессонов (2004) позволяет получить менее пессимистичную динамику СФП, однако факт технологического «регресса» остается без изменений.

На основе панельных данных по предприятиям для 83 отраслей российской экономики за период 1993–2000 гг. Bessonova, Kozlov & Yudaeva (2002) получают снижение СФП в 1993–1998 гг., что соответствует результатам в Dolinskaya (2002), полученным на агрегированном уровне для всей экономики. Оценивание СФП в работе Bessonova, Kozlov & Yudaeva (2002) основано на концепции транслогарифмической производственной функции, подходе, предложенном в работе Jorgenson (1995). Рассматривая кумулятивную динамику СФП для 9 агрегированных отраслей в 1993–2000 гг., авторы работы отмечают, что отрицательные темпы роста СФП в период 1994–1998 гг. во многом связаны с высокой степенью неопределенности экономической конъюнктуры, вызванной сильным колебанием уровня цен, и кризисом 1998 г., отразившимся на динамике реального обменного курса рубля.

В работе Бессонова (2007) анализируется эффективность российских промышленных предприятий в период 1995–2004 гг. на основе микроданных. Бессонова рассматривает три возможных подхода к оцениванию СФП на основе панельной структуры данных: непараметрический метод, транслогарифмическая производственная функция и модель стохастической границы производственных возможностей. Одним из основных аргументов против стандартного подхода к оцениванию СФП является предпосылка об абсолютной эффективности фирм, оперирующих на границе производственных возможностей: только в этом случае СФП эквивалентен технологическому прогрессу. Подход с точки зрения стохастической производственной границы позволяет ослабить данную предпосылку и учитывать тот факт, что фирмы могут оперировать внутри своих производственных множеств. В таком случае СФП можно разложить на отдельные составляющие: технологический прогресс (который представляет собой сдвиг границы производственных возможностей во времени), уровень неэффективности (как далеко фирма находится от своей производственной границы) и возрастающую отдачу от масштаба. Основные результаты, полученные в работе Бессонова (2007),

следующие:

- Эластичность по капиталу  $\epsilon_K$  для всех отраслей меньше 0,25, эластичность по труду  $\epsilon_L$  высокая (минимальное значение эластичности  $\epsilon_L$  характерно для электроэнергетики и отраслей топливной промышленности), что оправдывает широко распространенную в зарубежной литературе по СФП практику предполагать факторные эластичности заданными как  $\epsilon_L \in [0,7; 0,75]$  и  $\epsilon_K \in [0,25; 0,3]$  и делает данные упрощения состоятельными и для российской экономики;
- Непараметрический метод и метод транслогарифмической производственной функции подтверждают наличие отрицательных отраслевых темпов роста СФП до 1998 г.;
- Уровень неэффективности предприятий (процент от максимально возможного уровня производительности предприятий в отрасли) довольно низкий и составляет порядка 12–15% в 1995–2004 гг., т.е. стандартный подход к оцениванию СФП с использованием производственной функции может быть во многом оправдан;
- Вклад компоненты, отвечающей за возрастающую отдачу от масштаба, в динамику СФП незначим.

В приведенном выше обзоре литературы были рассмотрены основные исследования, в которых оценивается технологический прогресс, и на результаты которых во многом опирается данная работа. Общей чертой всех работ, посвященных оцениванию СФП для российской экономики, является наличие отрицательного шока технологий или технологического «регресса», причем отрицательный шок технологий проявляется и на отраслевом уровне. Таким образом, существующие исследования свидетельствуют в пользу гипотезы о том, что трансформационный спад, если и не был вызван, то по крайней мере сопровождался снижением СФП. Моделирование технологического прогресса на основе подхода Basu, Kimball & Fernald (2006) призвано подтвердить или опровергнуть гипотезу о снижении уровня технологий в российской экономике в 1990-х годах.

### 3 Моделирование технологического прогресса

Basu, Kimball & Fernald (2006) предполагают, что технология производства для каждой отрасли может быть описана производственной функцией однородной степени  $\gamma$ , где  $Y_t$  представляет собой объем выпуска,  $K_t$  – запас капитала,  $N_t$  – число занятых в отрасли,  $A_t$  – уровень технологий:

$$Y_t = F(Z_t K_t, E_t H_t N_t, M_t, A_t), \quad (1)$$

где  $E_t$  – усилия в среднем на одного работника,  $H_t$  – количество часов, отработанных в среднем одним работником,  $Z_t$  – уровень использования капитала в производстве.

Логлинеаризуя сформулированную выше производственную функцию, можно получить следующее представление:

$$dy_t \approx \frac{F_1 Z K}{Y} (dz_t + dk_t) + \frac{F_2 E H N}{Y} (de_t + dh_t + dn_t) + \frac{F_3 M}{Y} dm_t + \frac{F_4 A}{Y} da_t. \quad (2)$$

Параметризуем данное выражение следующим образом:

$$dy_t \approx \mu^* \frac{\gamma^* K^*}{Y^*} (dz_t + dk_t) + \mu^* \frac{w^* N^*}{Y^*} (de_t + dh_t + dn_t) + \mu^* \frac{P_m^* M^*}{Y^*} dm_t + \frac{F_4 A^*}{Y^*} da_t, \quad (3)$$

где  $\mu^*$  – равновесная надбавка к издержкам в предположении о несовершенной конкуренции, представляющая собой наценку над конкурентным уровнем цен ( $P = MC$ ) и характеризующая монопольную власть в отрасли:

$$\mu = \frac{P}{MC} = \frac{AC}{MC} \frac{P}{AC} = \gamma \frac{1}{1 - s_\pi}, \quad (4)$$

где  $AC$  – средние общие издержки,  $MC$  – предельные издержки,  $P$  – средняя цена, которая устанавливается в отрасли.

Если доля прибыли в выручке  $s_\pi = (PY - AC \cdot Y)/PY$  близка к нулю, т.е. предполагается, что доля прибыли по отраслям мала, тогда параметр отдачи от масштаба  $\gamma$  и маржа  $\mu$  равны. В этом случае эластичности по капиталу  $\epsilon_K = F_1 ZK/Y$ , труду  $\epsilon_L = F_2 EHN/Y$  и материальным затратам  $\epsilon_M = F_3 M/Y$  совпадают с долями капитала  $s_K$ , труда  $s_L$  и матзатрат  $s_M$ .

Basu & Fernald (1997) исходят из предположения, что при оценивании параметра отдачи от масштаба должны быть получены либо постоянная отдача (CRS), либо возрастающая отдача (IRS), т.к. убывающая отдача от масштаба означает, что фирмы в отрасли устанавливают цену ниже своих предельных издержек  $MC$ , т.е. их прибыли отрицательные.

Итоговое выражение для логлинеаризованной производственной функции следующее:

$$dy_t = \gamma(dx_t + du_t) + d\tilde{a}_t, \quad (5)$$

$$dx_t = s_K dk_t + s_L (dn_t + dh_t) + s_M dm_t, \quad (6)$$

$$du_t = s_K dz_t + s_L de_t, \quad (7)$$

где  $dx_t$  – темп роста взвешенной комбинации факторов производства,  $du_t$  – темп роста загрузки производственных мощностей.

Если гипотеза о постоянной отдаче от масштаба не отвергается,  $\gamma = 1$  и  $du_t = 0$ , т.е. загрузка производственных мощностей постоянна во времени, в таком случае  $d\tilde{a}_t$  представляет собой стандартный остаток Солоу. Из всего этого следует, что, не учитывая загрузку мощностей, можно получить смещенные оценки темпов роста СФП.

Из работы Бессонов (2004) следует, что соотношение между долями капитала и труда для отечественной экономики в период 1990–2002 гг.  $s_K = 0,5s_L$ . Как правило, в литературе по СФП доли капитала и труда полагаются следующими:  $s_K = 0,7$  and  $s_L = 0,3$ . В исходной работе Basu, Kimball & Fernald (2006) доли явным образом оцениваются из системы уравнений (5)–(7).

Индикатор СФП  $d\tilde{a}_t$  для каждой отрасли может быть получен как остаток из уравнения

$$dy_t = c + \gamma dx_t + \beta du_t + d\tilde{a}_t. \quad (8)$$

В исходном подходе Basu, Kimball & Fernald (2006) уровень загрузки производственных мощностей предполагается ненаблюдаемым технологическим параметром. Микрообоснования для выбора прокси для загрузки мощностей связаны с задачей минимизации издержек фирмой. В предположении, что фирмы оперируют на границе производственных возможностей, наблюдаемые параметры факторов производства могут служить в качестве прокси для загрузки производственных мощностей. Авторы формально выводят, что часы, отработанные в среднем одним работником,  $dh_t$ , пропорциональны изменению ненаблюдаемых уровней загрузки капитала и труда:  $de = \psi dh$ ,  $da = \phi dh$  (Basu, Kimball & Fernald, 2006). Исходя из этих соображений, выражение (5) с учетом прокси для загрузки производственных мощностей модифицируется как

$$dy_t = c + \gamma dx_t + \gamma(\psi s_L + \phi s_K) dh_t + da_t = c + \gamma dx_t + \beta dh_t + d\tilde{a}_t. \quad (9)$$

Исходные гипотезы Basu, Kimball & Fernald (2006) о знаках соответствующих параметров в уравнении (8) следующие: предполагается, что  $\gamma > 1$ , т.е. для отраслей характерна возрастающая отдача от масштаба, и  $\beta > 0$ , т.е. темпы роста загрузки мощностей положительно коррелируют с ростом выпуска. Следует, однако, отметить, что согласно подходу Basu, Kimball & Fernald (2006)  $dh_t$  дважды входит в уравнение (9): во-первых, как часть взвешенной комбинации факторов производства  $dx_t$ , и во-вторых, отдельно как прокси для загрузки производственных мощностей. Такого рода нелинейность может сказываться на оценках коэффициентов: в терминах параметров это означает, что  $\beta = \gamma\xi$ , где  $\xi$  представляет собой отдачу на отработанные часы. Inklaar (2005) отмечает, что в случае, если гипотеза о постоянной отдаче выполнена ( $\gamma = 1$ ), как и было получено для преобладающего числа отраслей в работе Basu, Kimball & Fernald (2006), нелинейность по параметру  $\gamma$  не будет сказываться на результатах оценивания.

#### 4 Оценивание СФП

Несмотря на кажущуюся простоту модели, применяемой для оценивания СФП, существует ряд проблем, связанный с нахождением хороших прокси для измерения факторов производства (капитала, труда и затрат сырья) и уровня загрузки производственных мощностей для соответствующих факторов. Существуют также различные подходы к выбору переменной  $Y_t$ , которая может моделироваться как валовая добавленная стоимость или объем производства. Второй подход обладает существенными недостатками, которые отмечены в работе Bessonova, Kozlov & Yudaeva (2002), а именно: на совокупный объем выпуска оказывает непосредственное влияние динамика цен факторов производства, которая передается на динамику цен конечной продукции (что решается с учетом дефлирования). В исходной работе Basu, Kimball & Fernald (2006) предполагается использование валовой добавленной стоимости, т.е. из объема выпуска исключаются затраты на промежуточное потребление сырья и материалов, т.к.  $M_t$  учитывается в производственной функции отдельно как фактор производства. Однако сами авторы отмечают, что валовая добавленная стоимость может рассматриваться в качестве меры для  $Y_t$  только в предположении о совершенной конкуренции. В случае непостоянной отдачи от масштаба использование валовой добавленной стоимости может приводить к смещению из-за опущенных переменных (Basu & Fernald, 1997).

В данной работе в качестве прокси для объема выпуска  $Y_t$  используется индекс промышленного производства (в классификации ОКОНХ), рассчитываемый В.А. Бессоновым и Э.Ф. Барановым (Центр экономической конъюнктуры при правительстве РФ, ЦЭК). Чтобы избежать проблемы эндогенности между объемом выпуска и промежуточным потреблением, в качестве  $M_t$  рассматриваются затраты на сырье и материалы, т.е. из состава материальных затрат исключаются расходы на потребление топлива и энергии.

##### 4.1 Данные

В данной работе для анализа используется официальная статистическая информация Федеральной службы государственной статистики (ФСГС), покрывающая период с 1 квартала 1995 г. по 4 квартал 2004 г., т.е.  $T = 40$  (в ценах 2000 г., сезонно сглаженные показатели). Выбор периода обусловлен особенностями статистического учета на отраслевом уровне: классификация отраслей до 2005 г. строилась на основе классификатора ОКОНХ (общесоюзный классификатор отраслей народного хозяйства), с 2005 г. был введен ОКВЭД (общероссийский классификатор видов экономической деятельности). Из-за перехода на новый классификатор нарушается преемственность временных рядов по затратам и инвестициям. В ОКОНХ отрасль представляет собой совокупность предприятий, производящих однородную продукцию, т.е. в качестве объекта классификации выделяется предприятие (по профилирующему виду деятельности). Напротив, в ОКВЭД объектом классификации является

вид экономической деятельности, при этом меняются границы существующих видов деятельности, поэтому пересчет показателей из одного классификатора в другой практически не представляется возможным. Именно по этой причине анализируемый в работе временной интервал ограничен 2004-м годом и не включает в себя доступную информацию по 2005–2007 гг. Ограничение периода 1995-м годом обусловлено отсутствием статистической информации по загрузке производственных мощностей до 1994 г., а также попыткой избежать проблем, связанных с активными трансформационными процессами, протекавшими в отраслях как минимум до 1995 г. и негативно отражавшимися на качестве статистической информации.

В работе рассматриваются данные по следующим статистическим показателям:

- Объем промышленного производства: квартальные данные 1995–2004 гг., в ценах производителей 2000 г.;
- Данные по материальным затратам: затраты на приобретение сырья и материалов (по форме ФСГС 5-3, причем в состав затрат исходно не включаются затраты на потребление основного капитала), квартальные данные 1997–2004 гг., ежегодные данные 1995–1996 гг., в ценах производителей 2000 г., исключая расходы на потребление топлива и энергии;
- Данные по основным фондам: ежегодные данные 1995–2004 гг., в ценах 2000 г.;
- Инвестиции в основной капитал (затраты на создание и воспроизводство основных фондов, реконструкцию и техническое перевооружение объектов, приобретение машин, оборудования, etc.): квартальные данные 1999–2004 гг., ежегодные данные 1995–1998 гг., в ценах 2000 г.;
- Средняя списочная численность работников по отраслям и количество часов, отработанных в среднем одним работником: квартальные данные 1995–2004 гг.;
- Уровень загрузки производственных мощностей в промышленности (уровень загрузки капитала и уровень загрузки труда): ежемесячные данные 1995–2004 гг., источник: Российский Экономический Барометр (РЭБ);
- Индекс промышленного производства: ежемесячные данные 1995–2004 гг., источник: Бессонов (2004).

Из-за сложностей, связанных с отсутствием полной квартальной статистической информации (ежеквартальный учет либо не ведется ФСГС в принципе, либо такая информация не полностью публикуется, в особенности на отраслевом уровне), необходима методика заполнения пробелов в квартальных данных по материальным затратам, капиталу и инвестициям исходя из доступной ежегодной информации. Квартальные данные требуются из-за особенностей подхода к оцениванию СФП, который предполагает анализ временных рядов, соответственно, ежегодные российские данные за 10 лет для этой цели не подходят. Гипотеза, которая используется для разбиения квартальных данных до годовых, для простоты предполагает, что коэффициенты сезонности фиксированы на уровне следующего (ближайшего) года, по которому доступны квартальные данные:

$$Z_t = \sum_{k=1}^4 Z_{kt}, \quad (10)$$

$$Z_{1t} = \frac{Z_t}{1 + g_2 + g_2g_3 + g_2g_3g_4}, \quad (11)$$

$$Z_{kt} = g_{k-1}Z_{(k-1)t}, \quad k = 2, 3, 4. \quad (12)$$

Здесь  $Z_t$  – показатель за год,  $t = 1995, 1996$ , для затрат на сырье,  $t = 1995, \dots, 1998$  для инвестиций в основной капитал, млн. руб., в ценах 2000 г.;  $Z_{kt}$  – квартальный показатель,  $k = 2, 3, 4$ , вычисленный на основе годового показателя  $Z_t$ , млн. руб.;  $g_k = Z_k/Z_{k-1}$ ,  $k = 2, 3, 4$ , – коэффициенты сезонности (за 1997 г. для затрат, за 1999 г. для инвестиций).

Представленная выше методика разбиения годовых данных до квартальных в случае статистики по инвестициям и затратам необходима для частичного заполнения пробелов в данных. Ситуация облегчается тем, что ФСГС ведет квартальный учет по соответствующим показателям. С данными по основным фондам ситуация сложнее, т.к. учет по ним ведется ежегодный. Однако в случае основных фондов интересен не сам по себе ряд или его динамика (поскольку, как уже было отмечено ранее, то, что в статистике понимается под основными фондами, не соответствует понятию капитала в производственной функции), а начальное значение капитала  $K_0$  (за 1-й квартал 1995 г.) для последующего рекурсивного моделирования динамики капитала с учетом инвестиций в основной капитал. Квартальные показатели для ряда по основным фондам за 1995 г. были получены путем разбиения соответствующего годового показателя ФСГС в предположении о постоянном ежеквартальном росте:  $1 + g_{year} = (1 + g_{quarter})^4$ , тогда  $g_k = g_{quarter}$  в формулах (10)–(12).

## 4.2 Способы измерения капитала

В эмпирической литературе существует проблема, связанная с незначимостью капитала как фактора при оценивании производственной функции. Burnside, Eichenbaum & Rebelo (1995) отмечают, что большинство рассматриваемых в работах Solow (1964), Lucas (1970), Bernarke & Parkinson (1991) и др. подходов дают похожие результаты: капитал  $K_t$  либо незначим, либо входит в оцениваемое уравнение для производственной функции с отрицательным знаком. Таким образом, одной из первичных задач при оценивании СФП является построение индикатора, адекватно измеряющего капитал, или нахождение соответствующей прокси-переменной для капитала. В работе Griliches & Jorgenson (1967) в качестве прокси для капитала на отраслевом уровне используются расходы отраслей на потребление энергии. Однако для российской экономики такого рода прокси вряд ли будет адекватной оценкой для капитала. Basu, Kimball & Fernald (2006) утверждают, что потребление энергии является хорошим индикатором капитальных активов только для обрабатывающих отраслей, поскольку при использовании такой прокси игнорируются кросс-эффекты взаимосвязи капитала  $K_t$  с численностью работников  $L_t$ .

В работе Бессонов (2004) в качестве прокси для капитала рассматривается объем инвестиций в основной капитал за текущий период  $I_t$ . Преимуществом подобного подхода является то, что учитывается неоднородность возрастной структуры активов. Однако это означает, что во внимание принимаются только активы, существующие в текущем периоде, т.е. влияние капитала предыдущих периодов на текущий запас капитала игнорируется. Таким образом, возрастная структура капитала смещена в сторону более новых капитальных активов, это смещение призвано элиминировать влияние устаревших неэффективных активов. Прокси, предложенная Бессоновым (2004), тесно связана с методом непрерывной амортизации активов, который предполагает, что капитальные активы в каждом периоде формируются по рекурсивной формуле  $K_{t+1} = I_t + (1 - \delta)K_t$ . Для данного подхода необходимо знать оценку для нормы амортизации  $\hat{\delta}$ , которая, помимо всего прочего, предполагается неизменной на протяжении всего рассматриваемого периода ( $\delta = \text{const}$ ). Хотя в официальной статистике ФСГС есть данные по норме выбытия основных средств, по своей природе данный показатель отличается от того, что понимается под  $\delta$  в рекурсивной формуле.

Еще один подход к измерению капитала, который можно отметить, предложен в работе Meinen, Verbiest & Wolf (1998) для голландской статистики по основным фондам, т.е. по нефинансовым капитальным активам, и применяется в работе Воскобойников (2003) для российской статистики за 1961–2001 гг. Этот подход позволяет оценить временной ряд для

капитала, используя статистику по валовому накоплению основного капитала (ВНОК) и понятие дожития капитальных активов. Функция дожития  $S(\tau)$  характеризует вероятность события, при котором актив продолжает функционировать до периода времени  $\tau$ . В основе данного подхода лежит предпосылка о том, что оцениваемые параметры функции дожития постоянны во времени. Проблема заключается в том, что согласно оценкам в Воскобойников (2003), в среднем продолжительность жизни актива составляет порядка 20 лет, т.е. для того, чтобы определить текущую стоимость капитальных активов в отрасли, например, для 1 квартала 1995 г., необходимо знать статистику по ВНОК как минимум с 1975 г.

В данной работе выбран подход к оцениванию капитала, основанный на методе непрерывной амортизации активов. В основе выбора нормы амортизации  $\delta$  на отраслевом уровне лежит следующая логика:  $\min_{\delta} \sum_{t=1}^T (K_t - K_t(\delta))$ , где  $K_t$  представляет собой временной ряд согласно официальной статистике по основным фондам, а  $K_t(\delta)$  – построенный на основе рекурсивной формулы индикатор капитала:  $K_t(\delta) = \sum_{i=0}^t (1 - \delta)^{t-i} I_i + (1 - \delta)^t K_0$ . Однако результаты данной процедуры свидетельствуют в пользу оптимальности предпосылки о постоянном уровне амортизации  $\delta = 0$ . С одной стороны, данная предпосылка является серьезным упрощением, с другой стороны, за предпосылкой  $\delta = 0$  стоит следующая интуиция: во-первых, рассматриваемый в работе период времени охватывает 10 лет, т.е. он сравнительно короток, чтобы исключить из рассмотрения амортизацию, во-вторых, варьируя значения параметра  $\delta \in [0; 0,1]$ , можно влиять на скорость накопления капитала в отрасли, а не на тенденцию в целом.

### 4.3 Измерение уровня загрузки производственных мощностей

При оценивании СФП с учетом изменяющихся во времени уровней использования капитала и труда необходимо располагать адекватными индикаторами уровня загрузки производственных мощностей. В литературе данная проблема, как и в случае с измерением капитала, в большинстве случаев решается нахождением прокси, хотя, например, в работе Shapiro (1996) используется непосредственно статистика по загрузке мощностей. Basu, Kimball & Fernald (2006) в своей работе формализуют подход, предложенный ранее в работе Abbott, Griliches & Hausman (1998), которые рассматривают количество часов, отработанных в среднем одним работником, в качестве прокси как для уровня использования труда в производстве, так и для уровня использования капитала. Данный подход предполагает, что в краткосрочном периоде усилия  $E_t$  и ненаблюдаемый параметр загрузки капитала  $Z_t$  возрастают с увеличением количества рабочих часов. В то время как взаимосвязь усилий работника с количеством отработанных им часов интуитивно очевидна, взаимосвязь между часами и уровнем загрузки капитала проявляется из-за премии рабочих смен (Basu, Kimball & Fernald, 2006). Уровень использования капитала возрастает при увеличении времени в рабочих сменах, т.е. с ростом отработываемых работниками часов увеличивается интенсивность использования капитала.

В нашем случае доступна официальная статистика по уровню использования капитала и труда, хотя существуют и определенные недостатки, с ней связанные. Можно выделить по крайней мере два источника данных по уровню загрузки производственных мощностей, которые опираются на результаты опросов предприятий (Oomes & Dynnikova, 2006):

а) статистика ФСГС по уровню загрузки капитала для 43 категорий промышленных товаров, ежегодные данные, по опросам 7000 предприятий;

б) статистика Российского Экономического Барометра (РЭБ) по уровням загрузки капитала и труда в промышленности, ежемесячные данные, по опросам выборки из 500 предприятий.

ФСГС определяют уровень загрузки производственных мощностей как отношение объема произведенной продукции к максимально возможному уровню выпуска, в то время как в определении РЭБ используется отношение к среднему уровню производства в месяц (Oomes & Dynnikova, 2006). Использование среднего уровня вместо максимально возможного ведет

к получению завышенных оценок загрузки мощностей. Кроме того, в случае оценок РЭБ можно говорить о проблеме репрезентативности выборки, поскольку выборка включает 500 предприятий вместо 7000. Как отмечается в Oomes & Dynnikova (2006), в выборке РЭБ, во-первых, недостаточно часто обновляются предприятия, что ведет к смещению в сторону более старых предприятий, во-вторых, недооценена доля крупных предприятий, т.е. идет смещение из-за отбора в сторону малых предприятий. С этой точки зрения выборка, по которой рассчитывает свои показатели ФСГС, лучше. Однако согласно методологии ФСГС, не существует единого индикатора уровня загрузки капитала в промышленности, вместо этого ФСГС считает индикаторы по отдельным видам промышленной продукции. На основе данной статистической информации В.А. Бессонов путем агрегирования строит сводный индикатор уровня загрузки капитала в промышленности. На Рис. 1 отражены различия между ежегодными показателями уровня загрузки капитала по оценкам ФСГС и РЭБ в период 1995–2004 гг.

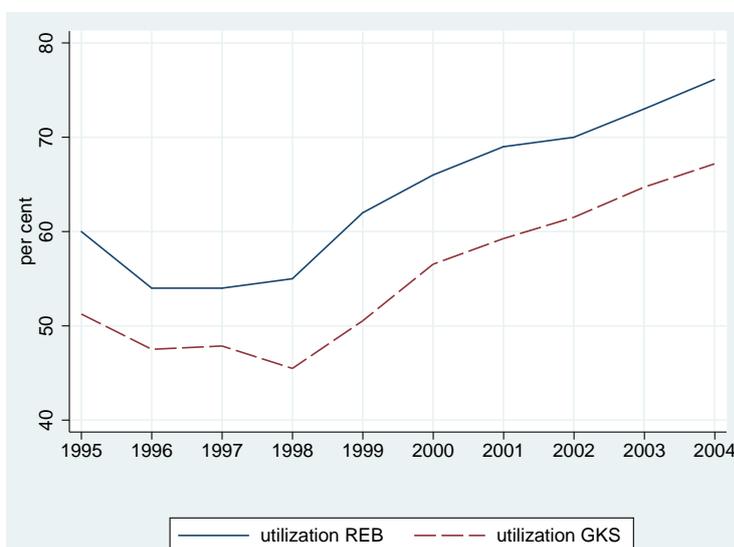


Рис. 1: Загрузка капитала, промышленность: РЭБ и ФСГС.

Как видно из графика, динамика двух данных индикаторов загрузки производственных мощностей практически идентична, однако минимальный уровень загрузки мощностей для индикатора ФСГС характерен для 1998 г. (что соответствует падению выпуска в 3 квартале 1998 г.), в то время как для индикатора РЭБ минимум наблюдается в 1996–1998 гг. Тем не менее, преимущество показателей РЭБ над сводным индикатором ФСГС заключается в том, что в методологии РЭБ доступны также данные по уровню использования труда, что может рассматриваться как прокси для усилий работников в модели. Полагая прокси для уровней загрузки капитала и труда единой, можно получить искаженные результаты оценивания. Из Рис. 2 можно видеть, что уровень загрузки труда существенно выше, чем уровень загрузки капитала с 1 квартала 1995 г. по 4 квартал 2004 г., хотя динамика показателей похожа. Немаловажным также является тот факт, что учет по загрузке мощностей РЭБ ведется ежемесячно, в то время как учет ФСГС – ежегодно. В случае работы с квартальными данными индикаторы РЭБ представляются более удобными.

#### 4.4 Оценивание индикаторов технологического прогресса

Для оценивания уравнения СФП из уравнения (8) используется метод максимального правдоподобия с ограниченной информацией (ММПОИ). Выбор метода оценивания обусловлен стремлением смягчить проблему эндогенности. Эндогенность в данном случае означает нару-

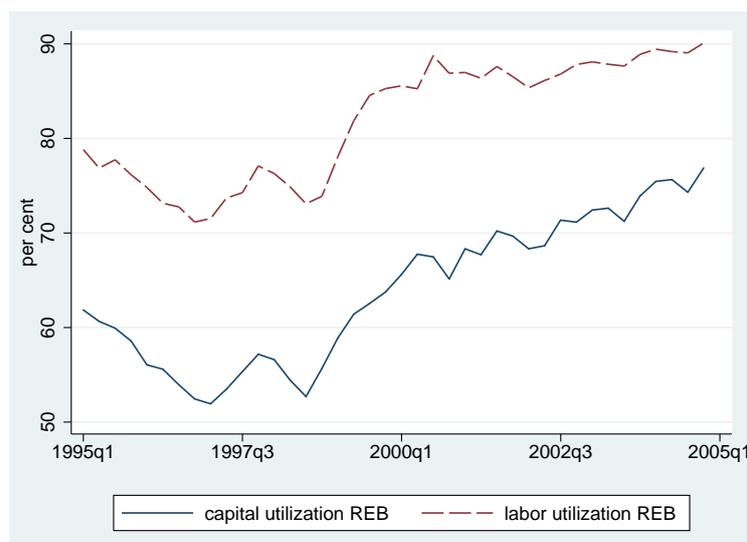


Рис. 2: Загрузка мощностей, промышленность, РЭБ: капитал и труд.

шение предпосылки  $\mathbb{C}[dx_t, d\tilde{a}_t] = 0$ , т.е. шок технологий коррелирован с динамикой факторов производства. Причина возникновения эндогенности, согласно Inklaar (2005), заключается в том, что в ответ на шок технологий фирмы одновременно подстраивают и выпуск, и потребление факторов в производстве. Таким образом, для  $dx_t$  в уравнении (8) необходимы инструменты. В литературе по прикладной теории роста в качестве инструментов при оценивании СФП обычно рассматриваются инструменты Холла–Рамеи (показатели, которые воздействуют на спрос на производимую продукцию). В работе Basu, Kimball & Fernald (2006) инструментами являются темпы роста уровня цен на нефть, темпы роста расходов государства на оборону и «монетарные шоки» из VAR-модели, полученные в работе Burnside (1996). Проблема с выбором инструментов заключается в том, что для получения состоятельных оценок параметров, инструменты должны быть, во-первых, валидны, т.е. экзогенны и некоррелированы с технологией (и другими шоками издержек), и релевантны, т.е. коррелированы с экономической активностью в отрасли, в которой инструментится  $dx_t$ . Однако, как отмечает Inklaar (2005), инструменты Холла–Рамеи, несмотря на свое широкое использование, относительно слабы. Точнее, данные инструменты довольно хорошо работают на уровне отдельной фирмы и зачастую являются слабыми при более высоких уровнях агрегирования (Basu & Fernald, 1997). Согласно Stock & Yogo (2005), инструменты являются слабыми, если полученное Р-значение  $\alpha$ -процентного теста Вальда на первом шаге (т.е. в регрессии эндогенных переменных на инструменты) превышает пороговое значение  $r$ . В работе Stock & Yogo (2005) приведены критические значения для F-статистики в регрессии на первом шаге при наличии единственного эндогенного регрессора в зависимости от порогового уровня  $r$  для теста Вальда и числа инструментов.

Можно выделить по меньшей мере два подхода к выбору «правильных» инструментов в литературе по оцениванию СФП. В частности, в работах Basu & Fernald (1997) и Diewert & Fox (2004) вместо оценивания с использованием инструментальных переменных используется обычный МНК, что позволяет не рассматривать проблемы, связанные со слабыми инструментами и соответствующим смещением ИП-оценок для малых выборок (смещение которых в случае слабых инструментов может быть больше, чем смещение МНК-оценок). Другой подход предложен в работе Shea (1993): вслед за Applebaum (1982) и др. в качестве инструмента для любой из отраслей используется агрегированный объем выпуска. Данный подход, согласно Shea (1993), является не самым предпочтительным: полученные на его основе оценки параметров могут быть смещены из-за влияния на совокупный объем выпуска

общих шоков издержек. Shea (1993) утверждает, что при выборе инструментов для каждой конкретной отрасли необходимо учитывать ее специфику. Критерий выбора, который предлагает автор, основан на межотраслевых балансах: на основе таблиц, характеризующих межотраслевые взаимосвязи, определяются отрасли, которые наиболее тесно связаны друг с другом. Изменения в выпуске одной из отраслей представляют собой экзогенный шок для второй отрасли. Inklaar (2005), следуя схожей логике при выборе инструментов, рассматривает в качестве инструментов для первичных отраслей взвешенные темпы роста объемов производства во вторичных отраслях. Шоки издержек во вторичных отраслях сказываются на ценах готовой продукции в первичных отраслях. В Inklaar (2005) приведены результаты использования построенных таким образом инструментов для стран Европы и показано, что данные инструменты дают лучшие результаты на первом шаге по сравнению со стандартными инструментами Холла–Рамеи: F-статистики существенно выше, и число отраслей, для которых величина смещения ИП-оценок не превышает 10% от смещения МНК-оценок, ниже, чем в случае использования инструментов Холла–Рамеи. Однако методика Shea (1993) довольно ограничена в своем применении. Прежде всего, существуют такие вторичные отрасли, продукция которых используется практически во всех остальных отраслях и составляет значительную долю от их затрат, например, отрасли топливной промышленности. Кроме того, есть разновидность готовой продукции, например в пищевой промышленности, которая не используется как промежуточная продукция в первичных отраслях. В результате построить таким способом инструменты можно далеко не для любой отрасли.

Приведенные выше подходы к выбору инструментальных переменных не всегда пригодны для российских данных. Так, все перечисленные подходы свидетельствуют в пользу слабых инструментов. Наилучшие результаты с точки зрения минимизации смещения в ИП-оценках параметров (с учетом ММПОИ-оценивания) дает, казалось бы, самый простой и стандартный подход – подход Холла–Рамеи. В качестве инструментов в данной работе используются следующие индикаторы и их первые разности (источник данных: IMF November 2007 database):

- темп роста индекса цен на нефть (UK Brent) по отношению к дефлятору ВВП;
- темп роста реального эффективного обменного курса рубля (REER).

Выбор индикатора, характеризующего динамику цен на нефть, согласуется с исходным подходом Холла–Рамеи, однако, как отмечают Basu & Fernald (1997), в случае, если улучшение технологий в экономике связано с благоприятной динамикой цен топливных ресурсов, технологический прогресс будет проявляться тогда, когда цены, наоборот, снижаются. Выбор REER в качестве инструмента для факторов производства во многом связан с логикой, предложенной в работе Вдовиченко, Дынникова и Субботин (2003), в которой рассматривается влияние реального обменного курса рубля на реальный уровень выпуска по отраслям. Ожидается, что в ответ на колебания реального обменного курса отрасли будут подстраивать объем факторов, используемых в производстве продукции.

Basu, Kimball & Fernald (2006) получают оценки параметров, используя оптимизационную процедуру для системы уравнений (5)–(7), при этом в явном виде оценивают параметры  $\hat{s}_K$ ,  $\hat{s}_L$ ,  $\hat{s}_M$ . Использовать схожий подход на российских данных с учетом ограничений на параметры (и с учетом знаков) не удастся: оптимизационная процедура либо не сходится, либо не дает хороших оценок параметров при ограничениях; более того, оценки для тех или иных долей факторов производства в преобладающем большинстве отраслей сходятся к нулю. В работе используется другой подход к оцениванию: предполагается, что доли  $s_K$ ,  $s_L$ ,  $s_M$  фиксированы экзогенно в соответствии с правилом:  $s_K \in [0, 1]$ ,  $s_L$  удовлетворяет соотношению  $s_L = 2s_K$ , которое, как было отмечено несколькими разделами ранее, выполняется для российских данных (Бессонов, 2004; Бессонова, 2007). В таком случае  $s_M = 1 - s_K - s_L$ , если  $s_M > 0$ , и  $s_M = 0$  при  $1 - s_K - s_L < 0$ .

## 5 Результаты оценивания

В период 1995–2004 гг. трансформационные процессы в отрасли и последующий кризис 1998 г. привели к падению выпуска в отраслях. В данном случае концепция агрегированной на отраслевом уровне производственной функции вряд ли будет описывать состояние экономики до 1998 г. Соответственно, может быть полезно рассмотреть укороченный период времени с 1999 по 2004 гг. С одной стороны, при отбрасывании переходной динамики остается довольно мало наблюдений, что само по себе может негативно отразиться на оценках параметров, но с другой стороны, это может прояснить, каким образом U-образная динамика объемов промышленного производства влияет на результаты оценивания СФП.

### 5.1 Рост СФП с учетом загрузки мощностей: 1995–2004 гг.

В данном подразделе анализируется отраслевая динамика СФП с учетом загрузки производственных мощностей. Оценивание СФП основано на подходе Basu, Kimball & Fernald (2006) с учетом индексов загрузки мощностей РЭБ по капиталу и труду в качестве прокси для показателей использования капитала  $Z_t$  и труда  $E_t$ . В соответствии с рассматриваемым подходом предполагается, что отраслевые темпы роста СФП можно вычислить по формуле:

$$dw_t = d\tilde{a}_t + \hat{c} = dy_t - \hat{\gamma}dx_t - \hat{\beta}du_t. \quad (13)$$

Следует также отметить, что при анализе полного периода времени, с 1995 по 2004 гг., отраслевые оценки СФП в соответствии со стандартным подходом Солоу получить не удастся, т.к. динамика комбинации факторов производства сама по себе, без учета загрузки производственных мощностей, не в состоянии объяснить U-образную динамику выпуска. В частности, с этой точки зрения учет загрузки мощностей для российской экономики обладает дополнительным преимуществом.

В таблице 1 представлены оценки параметров отдачи от масштаба и отдачи на использование факторов производства по 12 отраслям промышленности. В соответствии с результатами F-теста для первого шага смещение в оценках из-за использования слабых инструментов при ММПОИ оценивании составляет 5%.

Из таблицы 1 можно также видеть, что значимая оценка параметра отдачи от масштаба была получена только для легкой промышленности. Inklaar (2005) в своей работе анализирует, выполняется ли предпосылка о постоянной отдаче, тестируя гипотезу об отличии полученных оценок параметров для отраслей от  $\gamma_i = 1$  для всех  $i$ . Однако в данном случае оценки свидетельствуют о наличии убывающей отдачи и, кроме того, не отличаются значимо от нуля (хотя при этом значимо отличаются от таковых при постоянной отдаче).

Оценки параметра отдачи на использование факторов значимы и положительны для 9 отраслей из 12, что соответствует гипотезам в исходном подходе Basu, Kimball & Fernald (2006). В соответствии с предположениями, ожидаемые знаки параметров следующие:  $\gamma_i > 1$  и  $\beta_i > 0$ ,  $i = 1, \dots, 12$ . Гипотеза  $\beta_i > 0$  не отвергается в случае отраслей российской промышленности, что свидетельствует об обоснованности выбора агрегированных индексов загрузки мощностей в качестве единой прокси для отраслевых показателей использования факторов  $du_{it}$ . Inklaar (2005) вслед за Basu, Kimball & Fernald (2006) использует количество отработанных часов как прокси и получает для стран Европы отрицательные и в большинстве случаев незначимые оценки для  $\hat{\beta}_i$ .

Полученные результаты для отдачи на использование факторов производства по сути означают, что в период с 1995 по 2004 гг. динамика СФП не отличается от динамики уровня загрузки производственных мощностей. Незначимость параметра  $\gamma$ , с одной стороны, означает, что предположения относительно спецификации производственной функции могут быть не до конца верны, однако с другой стороны, в данном случае об уровне загрузке мощностей

Таблица 1: СФП с учетом загрузки мощностей: 1995–2004 гг.

Отрасль	$\hat{\gamma}$	$\hat{\beta}$	$\epsilon_L, \epsilon_K, \epsilon_M$	F-статистика
машиностроение	−0,013 (0,102)	1,678*** (0,378)	0,7, 0,3, 0	7,087
легкая промышленность	0,481*** (0,131)	2,159*** (0,511)	0,6, 0,3, 0,1	7,895
цветная металлургия	0,020 (0,036)	0,773*** (0,136)	0,7, 0,3, 0	6,817
нефтедобывающая	−0,002 (0,026)	0,112 (0,100)	0,7, 0,3, 0	7,409
нефтеперерабатывающая	0,043 (0,056)	0,190 (0,210)	0,7, 0,3, 0	7,694
газовая промышленность	0,023 (0,031)	0,169 (0,122)	0,7, 0,3, 0	7,520
угольная промышленность	−0,038 (0,054)	0,926*** (0,206)	0,7, 0,3, 0	7,440
черная металлургия	−0,092 (0,066)	0,954*** (0,211)	0,6, 0,3, 0,1	7,251
электроэнергетика	−0,050 (0,040)	0,462*** (0,148)	0,7, 0,3, 0	7,145
химическая промышленность	−0,040 (0,070)	1,420*** (0,243)	0,6, 0,3, 0,1	8,171
ЛДЦБ промышленность	0,069 (0,067)	1,089*** (0,250)	0,6, 0,3, 0,1	7,056
пищевая промышленность	−0,018 (0,057)	0,802*** (0,215)	0,6, 0,3, 0,1	7,659

*Замечания:* «F-статистика» ссылается на первый шаг 2ШМНК. Уровни значимости: \* – 10% уровень, \*\* – 5% уровень, \*\*\* – 1% уровень. В скобках указаны стандартные ошибки. Квантили ММПОИ для порогового значения  $r$  5%-ного теста Вальда гипотезы  $\beta = \beta_0$ : 5,44 ( $r = 0,05$ ), 3,87 ( $r = 0,1$ ), 3,30 ( $r = 0,2$ ), 2,98 ( $r = 0,3$ ). В качестве тестов на сверхидентифицирующие ограничения используются хи-квадрат тест Андерсона–Рубина и F-тест Басманна. Гипотеза о валидности инструментов не отвергается.

можно говорить как о дополнительном виде фактора производства. Рост объема произведенной в экономике продукции сопровождается увеличением объема факторов, использующихся в производстве, и в том числе ростом загрузки производственных мощностей.

Процедура оценивания СФП не работает для трех отраслей топливной промышленности: нефтедобывающей, нефтеперерабатывающей и газовой. С одной стороны, можно говорить о том, что для отраслей топливной промышленности выбор инструментов, использующихся для оценивания СФП, недостаточно обоснован, поскольку инструменты Холла–Рамай включают индикатор, отвечающий за динамику цен на нефть. Однако использование инструментальных переменных в рамках описанных ранее подходов не позволяет улучшить результаты. Таким образом, можно говорить о том, что сам по себе подход, предполагающий существование производственной функции в отраслях топливной промышленности, неверен, т.е. динамика объемов выпуска в данном случае определяется никак не факторами производства такими как труд, капитал, материальные затраты или загрузка мощностей.

Вслед за Vasu, Kimball & Fernald (2006) можно построить агрегированный показатель темпов роста СФП на уровне промышленности как взвешенную с весами (14) комбинацию отраслевых темпов роста СФП в соответствии с формулой (15):

$$dw = \sum_i \left( \frac{\theta_i}{1 - s_{Mi}} \right) dw_i, \quad (14)$$

$$\theta_i = \frac{P_i Y_i - P_{Mi} M_i}{\sum_i (P_i Y_i - P_{Mi} M_i)} \equiv \frac{P_i^V V_i}{P^V V}. \quad (15)$$

Веса  $\theta_i$  рассчитываются как доля валовой добавленной стоимости отрасли  $i$  в совокупной валовой добавленной стоимости.

На Рис. 3 изображена кумулятивная динамика агрегированного индикатора СФП в уровнях, т.е. уровень СФП в 1 квартале 1995 г. принимается равным 1, СФП в остальных кварталах рассчитывается на основе полученных темпов роста. Также на данном графике представлена динамика объема производства в промышленности и показателей загрузки производственных мощностей капитала и труда. Из графика видно, что падение выпуска и уровней загрузки мощностей, характерное до 1998 г., сопровождалось схожим снижением СФП, т.е. технологическим «регрессом».

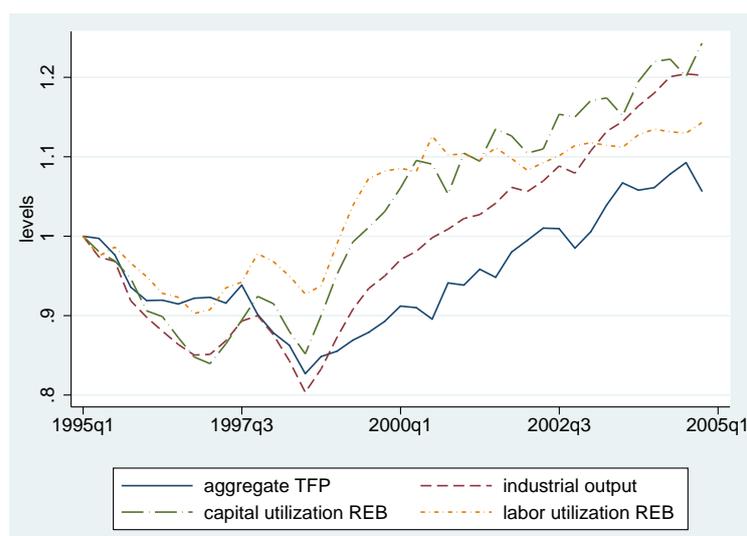


Рис. 3: Кумулятивная динамика: СФП, выпуск и загрузка мощностей.

Однако после 1998 г. рост выпуска сопровождается постепенным ростом СФП, что может свидетельствовать в пользу гипотезы о том, что наблюдающийся в последние несколько лет рост хотя бы отчасти вызван улучшением используемых в экономике технологий. Можно также отметить, что динамика агрегированного индикатора СФП проциклична, т.е. рост загрузки производственных мощностей соответствует росту выпуска и улучшению технологий, в то время как в работе Basu, Kimball & Fernald (2006), наоборот, снижение загрузки мощностей рассматривается как фактор, свидетельствующий о технологическом прогрессе. Динамика отраслевых индикаторов СФП в уровнях представлена на Рис. 4.

Динамика технологического прогресса на отраслевом уровне для большинства отраслей отражает наличие структурного сдвига в СФП после 1998 г. Тем не менее, есть отрасли, в которых динамика СФП довольно инертна на протяжении всего рассматриваемого периода, в частности, в цветной металлургии. Постепенное улучшение в динамике СФП после 1990-х гг. в легкой промышленности сходит к минимуму уже в 2002 г., после чего СФП в данной отрасли начинает медленно снижаться. Интересно также отметить машиностроение, где в 2002 г. наблюдается снижение СФП. По оценкам в Бессонова (2007) на основе данных по предприятиям, результат схожий: в 2002 г. рост СФП в машиностроении отрицателен, а к 2004 г. процесс улучшения технологий в отрасли заканчивается.

## 5.2 Оценивание СФП: проблема отдачи от масштаба

Как уже было отмечено ранее, для того чтобы понять, насколько результаты из предыдущего подраздела меняются в зависимости от выбора периода времени, отраслевые индикаторы СФП рассчитываются за период 1999–2004 гг., т.е. исключается структурный сдвиг.

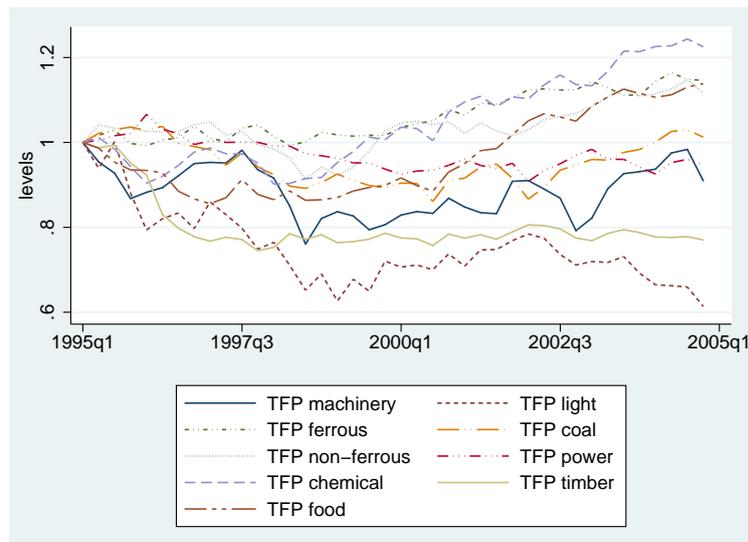


Рис. 4: Кумулятивная динамика СФП по отраслям.

В случае укороченного временного интервала можно сопоставить два подхода к оцениванию СФП: стандартный подход Солоу с учетом непостоянной отдачи от масштаба и подход Basu, Kimball & Fernald (2006) с поправкой на загрузку мощностей.

Стандартный подход Солоу (см. таблицу 2) для укороченного временного интервала дает значимо отличные от 0 и 1 оценки параметров отдачи от масштаба  $\hat{\gamma}_i$  для 6 отраслей из 12. При этом гипотезы как о постоянной, так и о возрастающей отдаче от масштаба отвергаются. Оценки, свидетельствующие об убывающей отдаче от масштаба, могут быть следствием проблемы, связанной с оценкой капитальных активов, которая обсуждалась в работе ранее: довольно интенсивное увеличение отраслевых капитальных активов в соответствии с модельными оценками, используемыми в данной работе, вместе с ростом (в среднем, по всем отраслям) численности работников и отработанных часов перевешивают динамику объемов производства, что и приводит к получению оценок  $\hat{\gamma}_i$  существенно меньших, чем 1.

Подход к оцениванию СФП с учетом загрузки мощностей для более короткого временного интервала не дает существенных улучшений в отраслевых оценках параметров отдачи от масштаба (см. таблицу 3), а именно только для двух отраслей из 12 получаемые оценки  $\hat{\gamma}_i$  значимо отличаются от 0. Оценки отдачи от использования факторов  $\hat{\beta}_i$  для выборки без структурного сдвига ниже, чем для полного временного интервала.

Чем объясняется полученная убывающая отдача от масштаба в отраслях российской экономики? В работе Basu & Fernald (1997) отмечена проблема с оцениванием отдачи от масштаба: для некоторых отраслей США авторы получают, что оценки значимо ниже, чем  $\gamma_i = 1$ . Точнее, для отраслей, производящих товары длительного пользования (в случае, когда отраслевой выпуск измеряется как валовая добавленная стоимость), оценки отдачи от масштаба практически нулевые с высокими стандартными ошибками (Basu & Fernald, 1997), что похоже на ситуацию, наблюдаемую с оценками отдачи от масштаба в данной работе. Более того, интуиция может подсказывать, что для российской экономики характерна ситуация с возрастающей отдачей от масштаба, которая может вести к возникновению множественности равновесий в динамических моделях общего равновесия, т.е. наблюдается нестабильность экономического развития. Еще одно возможное объяснение проблемы с оценками отдачи от масштаба заключается в том, что параметр  $\gamma_i$  явным образом предполагается постоянным на протяжении всего временного интервала, т.е. между динамикой выпуска и факторов производства должно существовать стабильное во времени соотношение (Basu & Fernald, 1997). Однако параметр отдачи от масштаба может меняться с течением

Таблица 2: Остаток Солоу с непостоянной отдачей от масштаба: 1999–2004 гг.

Отрасль	$\hat{\gamma}$	$\epsilon_L, \epsilon_K, \epsilon_M$	F-статистика
машиностроение	0,228** (0,090)	0,7, 0,3, 0	10,258
легкая промышленность	0,470*** (0,125)	0,6, 0,3, 0,1	8,027
цветная металлургия	0,102*** (0,039)	0,7, 0,3, 0	9,593
нефтедобывающая	-0,023 (0,020)	0,7, 0,3, 0	9,146
нефтеперерабатывающая	-0,047 (0,044)	0,7, 0,3, 0	9,697
газовая промышленность	0,018 (0,030)	0,7, 0,3, 0	9,614
угольная промышленность	0,089 (0,066)	0,7, 0,3, 0	9,778
черная металлургия	0,162** (0,069)	0,6, 0,3, 0,1	10,542
электроэнергетика	-0,018 (0,042)	0,7, 0,3, 0	10,503
химическая промышленность	0,152*** (0,056)	0,6, 0,3, 0,1	11,934
ЛДЦБ промышленность	0,112*** (0,033)	0,6, 0,3, 0,1	11,807
пищевая промышленность	0,016 (0,034)	0,6, 0,3, 0,1	12,855

Замечания: См. замечания к таблице 1.

времени, соответственно аппроксимировать коэффициенты  $\hat{\gamma}_{it}$  постоянным коэффициентом  $\hat{\gamma}_i$  не всегда обоснованно. Более того, убывающая отдача от масштаба, наблюдающаяся на уровне отраслей, может быть причиной агрегирования данных: агрегирование индивидуальных данных по предприятиям до уровня отраслей не учитывает неоднородность фирм по размеру и степень концентрации фирм в отрасли.

### 5.3 Темпы роста технологического прогресса

Рассчитанные за 1995–2004 гг. темпы роста индикаторов СФП (как отраслевых, так и агрегированного) до 1998 г. отрицательны, т.е. в 3 квартале 1998 г. наблюдается ярко выраженный структурный сдвиг. Чтобы изучить и сопоставить средние отраслевые темпы роста, в работе используется период времени с 1999 по 2004 гг., т.к. в 1995–2004 гг. отрицательные темпы роста до 1998 г. практически полностью уравниваются положительной динамикой СФП после 1998 г. (в результате отраслевые темпы роста близки к 0). В таблице 4 представлены отраслевые темпы роста СФП, темпы роста агрегированного индикатора СФП по промышленности, а также, для сопоставления, темпы роста объема производства в промышленности.

Довольно высокая степень разброса в отраслевых темпах роста СФП объясняется тем, что индикаторы рассчитываются как остатки из соответствующих регрессий: несмотря на сезонное сглаживание анализируемых показателей, в остатках присутствует определенная степень сезонности.

Наиболее высокий темп роста технологий характерен для химической промышленности: 1,24% в среднем за квартал после 1999 г. Пищевая промышленность также характеризуется существенным улучшением технологий производства (1,17%). Преобладание негативных шоков технологий в легкой промышленности, электроэнергетике и лесной промышленности ведет к тому, что темпы роста СФП в данных отраслях в лучшем случае близки к 0, а в худшем – отрицательны. Что касается агрегированного уровня, промышленности в целом,

Таблица 3: СФП с учетом загрузки мощностей: 1999–2004 гг.

Отрасль	$\hat{\gamma}$	$\hat{\beta}$	$\epsilon_L, \epsilon_K, \epsilon_M$	F-статистика
машиностроение	0,191* (0,107)	0,359 (0,531)	0,7, 0,3, 0	6,581
легкая промышленность	0,208* (0,115)	1,620*** (0,569)	0,7, 0,3, 0	6,553
цветная металлургия	0,033 (0,043)	0,480** (0,192)	0,7, 0,3, 0	5,828
нефтедобывающая	0,010 (0,022)	-0,286 (0,100)	0,7, 0,3, 0	5,880
нефтеперерабатывающая	-0,070 (0,055)	0,219 (0,261)	0,7, 0,3, 0	6,465
газовая промышленность	0,016 (0,038)	0,054 (0,184)	0,7, 0,3, 0	6,129
угольная промышленность	0,014 (0,077)	0,606* (0,206)	0,7, 0,3, 0	6,244
черная металлургия	0,017 (0,079)	0,883*** (0,292)	0,6, 0,3, 0,1	6,261
электроэнергетика	-0,063 (0,050)	0,427* (0,251)	0,7, 0,3, 0	6,700
химическая промышленность	0,051 (0,060)	0,942*** (0,295)	0,6, 0,3, 0,1	7,493
ЛДЦБ промышленность	0,062 (0,038)	0,484*** (0,198)	0,6, 0,3, 0,1	7,175
пищевая промышленность	-0,007 (0,042)	0,242 (0,227)	0,6, 0,3, 0,1	8,254

Замечания: См. замечания к таблице 1.

динамика объема производства (рост в 1,55% за квартал) более интенсивна по сравнению с динамикой технологического прогресса (0,94%). Таким образом, в среднем примерно половина роста выпуска в российской промышленности экзогенна: она не объясняется технологическим прогрессом.

## 6 Заключение

В данной работе был проанализирован модифицированный подход к оцениванию совокупной факторной производительности (СФП) по отраслям российской промышленности для оценки уровня технологического прогресса в экономике. Основные результаты работы следующие.

1. Полученные оценки для темпов роста СФП на агрегированном уровне свидетельствуют о том, что рост объемов производства, который наблюдается в российской экономике после 1998 г., сопровождается постепенным улучшением технологий. На отраслевом уровне наиболее высокие темпы роста технологического прогресса (в среднем за период с 1999 по 2004 гг.) характерны для химической промышленности и пищевой промышленности.
2. Резюмировать ситуацию, происходящую с динамикой технологического прогресса до 1998 г., можно следующим образом: что такое отрицательный шок технологий и насколько оправданно в него верить?
  - Если в возможность технологического «регресса» не верить, то падение СФП после 1998 г. можно объяснять наличием ошибок измерения, что связано с особенностями измерения СФП как остатка. Соответственно, учет дополнительных нетехнологических эффектов мог бы позволить элиминировать отрицательные темпы роста

Таблица 4: Темпы роста технологического прогресса: 1999–2004 гг., за квартал, %%

Отрасль	Среднее	Ст.отклонение	Минимум	Максимум
машиностроение	0,51	4,17	–8,82	9,17
легкая промышленность	–0,38	4,71	–9,22	10,92
цветная металлургия	0,48	1,33	–1,57	2,85
угольная промышленность	0,50	2,64	–5,39	5,70
черная металлургия	0,73	1,98	–2,63	4,83
электроэнергетика	–0,76	1,82	–4,48	2,91
химическая промышленность	1,24	2,32	–2,77	6,70
ЛДЦБ промышленность	–0,05	1,63	–2,71	3,56
пищевая промышленность	1,17	1,76	–1,98	5,10
агрегированный СФП	0,94	1,93	–3,28	5,09
совокупный выпуск	1,55	1,27	–0,82	4,77

СФП. Кроме того, можно спорить и о природе технологического «регресса»: имеет ли место ухудшение технологий, если снижение СФП практически полностью объясняется снижением загрузки мощностей.

- Если верить в то, что технологии могут не только улучшаться, но еще и ухудшаться, то эффект снижения СФП до 1998 г. можно связать с трансформационными процессами, протекавшими в экономике в 1990-х годах, а также с последующим кризисом 1998 г. Последствиями данных процессов стала дезорганизация хозяйственных связей между предприятиями в экономике, а также реорганизация институциональной среды как таковой. «Капитал» и «труд» в том виде, в котором они подразумеваются в производственной функции, вовсе не соответствуют тому, что под капиталом и трудом понимается в статистических данных: снижение качества факторов производства, связанное с процессами накопления человеческого капитала и отсутствием навыков для взаимодействия с новыми институтами привели к технологическому «регрессу» в экономике. Вера в технологический «регресс» позволяет обосновать и убывающую отдачу от масштаба, т.к. предприятия во многих отраслях получали отрицательную прибыль в период трансформационного спада.

3. Полученные оценки отраслевых СФП с учетом загрузки производственных мощностей капитала и труда не согласуются с гипотезами возрастающей или постоянной отдачи от масштаба. Оценки отдачи от масштаба близки к 0 и существенно меньше 1, т.е. в данных мы наблюдаем убывающую отдачу. Объяснение данного феномена может заключаться в рассматриваемой в литературе проблеме оценивания отдачи от масштаба. Vasu & Fernald (1997) отмечают, что разброс в оценках тесным образом связан, во-первых, с выбором индикаторов для соответствующих переменных производственной функции, в особенности с правильным измерением индикатора капитальных активов, во-вторых, с процедурой агрегирования данных по предприятиям до отраслей. Структурный сдвиг, наблюдающийся в данных в 1998 г. не воздействует существенно на получаемые оценки параметров. Анализ укороченного временного интервала, с 1999 по 2004 гг., улучшает оценки, полученные с использованием стандартного подхода Солоу с учетом непостоянной отдачи от масштаба: в случае временного интервала 1995–2004 гг. динамика факторов производства без учета загрузки мощностей не объясняет U-образную динамику выпуска. Оценки, полученные для укороченного интервала времени, значимо отличны от 0 для половины из анализируемых отраслей, однако свидетельств в пользу

гипотез о постоянной или возрастающей отдаче по-прежнему нет. Оценки параметров для СФП с учетом уровня загрузки мощностей для 1999–2004 гг. практически не отличаются от исходных оценок для 1995–2004 гг. Основное отличие заключается в том, что при исключении структурного сдвига из выборки влияние загрузки производственных мощностей на динамику выпуска снижается (параметры отдачи на использование факторов меньше).

В качестве направлений для дальнейшего исследования в области измерения СФП для российской экономики с использованием отраслевой статистики можно выделить, во-первых, поиск других альтернатив рассмотренным индикаторам для капитала и загрузки производственных мощностей, во-вторых, подбор или построение альтернатив используемым в оценивании инструментам, позволяющих более качественно решить проблему эндогенности. Помимо технических вопросов с оцениванием, само по себе измерение технологического прогресса и построение хороших оценок, характеризующих улучшение/ухудшение технологий в экономике, может способствовать дальнейшему анализу в области экономического роста и НТП, т.е. позволит изучить, в частности, каким образом шоки технологий влияют на динамику макроэкономических показателей.

## Благодарности

Данная работа была подготовлена в рамках исследовательского проекта «Волатильность и рост» в РЭШ в 2007–2008 гг. под руководством Олега Замулина и Кирилла Сосунова. Автор благодарен научным руководителям, Владимиру Бессонову, Евгении Бессоновой, Дэвиду Брауну, Ирине Денисовой, Илье Воскобойникову за помощь при написании работы, а также всем участникам XXIII Научной Конференции РЭШ за ценные комментарии и замечания.

## Список литературы

- Бессонов, В.А. (2004). О динамике совокупной факторной производительности в российской переходной экономике. *Экономический журнал ВШЭ* 4, 542–587.
- Бессонов, В.А. & И.Б. Воскобойников (2006). О динамике основных фондов и инвестиций в российской переходной экономике. *Экономический журнал ВШЭ* 2, 193–228.
- Бессонова, Е. (2007). Оценка эффективности производства российских промышленных предприятий. *Прикладная эконометрика* 2, 1–23.
- Вдовиченко, А., О. Дынникова & В. Субботин (2003). О влиянии реального обменного курса на различные сектора российской экономики. Экономическая экспертная группа.
- Воскобойников, И.Б. (2003). Оценка совокупной факторной производительности российской экономики в период 1961–2001 гг. с учетом корректировки динамики основных фондов. *Препринт* WP2/2003/03, ГУ–ВШЭ.
- Энтов, Р., О. Луговой, Е. Астафьева, В. Бессонов, И. Воскобойников, М. Турунцева & Д. Некипелов (2003). Факторы экономического роста российской экономики. *Научные труды ИЭПП* 70Р.
- Abbott, T.A., Z. Griliches & J. Hausman (1998). Short run movements in productivity: Market power versus capacity utilization. Глава в *Practicing Econometrics: Essays in Method and Application* под редакцией Z. Griliches. Cheltenham: Edward Elgar.
- Applebaum, E. (1982). The estimation of the degree of oligopoly power. *Journal of Econometrics* 19, 287–299.
- Baily, M.N., E.J. Bartelsman & J. Haltiwanger (2001). Labor productivity: Structural change and cyclical dynamics. *Review of Economics and Statistics* 83, 420–433.
- Basu, S. (1998). Technology and business cycles: How well do standard models explain the facts? *Conference Series* 2, Federal Reserve Bank of Boston.
- Basu, S. & J. Fernald (1997). Returns to scale in U.S. production: Estimates and implications. *Journal of Political Economy* 105, 249–283.

- Basu, S., J. Fernald & M. Kimball (2006). Are technology improvements contractionary? *American Economic Review* 96, 1418–1448.
- Bernarke, B. & M. Parkinson (1991). Procyclical productivity and competing theories of the business cycle: Some evidence from interwar U.S. manufacturing industries. *Journal of Political Economy* 99, 439–459.
- Bessonova, E., K. Kozlov & K. Yudaeva (2002). Trade liberalization, foreign direct investment, and productivity of Russian firms. Manuscript, Centre for Economic and Financial Research.
- Brown, D.J. & J.S. Earle (2000). Competition and firm performance: Lessons from Russia. *CEPR Discussion Papers* DP2444.
- Burnside, C. (1996). Production function regressions, returns to scale, and externalities. *Journal of Monetary Economics* 37, 177–201.
- Burnside, C., M. Eichenbaum & M. Rebelo (1995). Capital utilization and returns to scale. *NBER Macroeconomics Annual* 1995, 67–110.
- Chang, Y. & J.H. Hong (2006). Do technological improvements in the manufacturing sector raise or lower employment? *American Economic Review* 96, 352–368.
- Diewert, W.E. & K.J. Fox (2004). On the estimation of returns to scale, technical progress and monopolistic markups. Manuscript, University of British Columbia.
- Dolinskaya, I. (2002). Explaining Russia's output collapse: Aggregate sources and regional evidence. *IMF Staff Papers* 49, 1.
- Griliches, Z. & J. Mairesse (1983). Comparing productivity growth: An exploration of French and U.S. industrial and firm data. *European Economic Review* 21, 89–119.
- Hall, R.E. (1990). Invariance properties of Solow's productivity residual. Глава в *Growth, Productivity, Unemployment: Essays to Celebrate Bob Solow's Birthday* под редакцией P. Diamond. Cambridge: MIT Press.
- Inklaar, R. (2006). Cyclical productivity in Europe and the U.S.: Evaluating the evidence on the returns to scale and input utilization. *Economica* 296, 822–841.
- Jorgenson, D. (1995). *Productivity. Том I: Postwar U.S. Economic Growth*. Cambridge: MIT Press.
- Jorgenson, D.W. & K.J. Stiroh (2000). Raising the speed limit: U.S. economic growth in the information age. *Brookings Papers on Economic Activity* 1, 125–211.
- Lucas, R.E.J. (1970). Capacity, overtime, and empirical production functions. *American Economic Review Papers and Proceedings* 60, 23–27.
- Marchetti, D.J. & F. Nucci (2005). Price stickiness and the contractionary effect of technology shocks. *European Economic Review* 49, 1137–1163.
- Meinen, G., P. Verbiest & P.-P. Wolf (1998). Perpetual inventory method. Service lives, discard patterns, and depreciation methods. Heerlen-Voorburg: Statistics Netherlands.
- OECD Manual. Measuring productivity. Measurement of aggregate and industry-level productivity growth. *Statistics Portal*, 1–156.
- Oomes, N. & O. Dynnikova (2006). The utilization-adjusted output gap: Is the Russian economy overheating? *IMF Working Paper* 06/68.
- Shapiro, M.D. (1996). Macroeconomic implications of variation in the workweek of capital. *Brookings Papers on Economic Activity* 2, 79–119.
- Shea, J. (1993). The input-output approach to instrument selection. *Journal of Business & Economic Statistics* 2, 145–155.
- Solow, R.M. (1957). Technical change and the aggregate production function. *Review of Economics and Statistics* 39, 312–320.
- Solow, R.M. (1964). Draft of presidential address: On the short-run relation of employment and output. Manuscript, MIT.
- Stock, J.H. & M. Yogo (2005). Testing for weak instruments in linear IV regression. Глава в *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas Rothenberg* под редакцией J.H. Stock & D.W.K. Andrews. Cambridge: Cambridge University Press.

## Источники данных

IMF November 2007 database, commodity prices and exchange rates.

«Движение рабочей силы: среднесписочная численность работников и количество отработанных человеко-часов». ФСГС, 1995–2004, Отдел статистической информации ГУ–ВШЭ.

«Заработная плата в России: эволюция и дифференциация». Под редакцией В.Е. Гимпельсона и Р.И. Капелюшниковой. Москва: Издательский дом ГУ–ВШЭ, 2008.

«Затраты на производство и реализацию продукции (работ, услуг)». ФСГС, 1995–2004, Центр анализа данных ГУ–ВШЭ, Отдел статистической информации ГУ–ВШЭ, [stat.hse.ru/main.html](http://stat.hse.ru/main.html).

«Инвестиции в России». ФСГС, 2001, 2003, 2005, [www.gks.ru](http://www.gks.ru).

«Инвестиции в экономику РФ». ФСГС, Отраслевые, 1999–2004. Центр анализа данных ГУ–ВШЭ, Отдел статистической информации ГУ–ВШЭ, [stat.hse.ru/main.html](http://stat.hse.ru/main.html).

«Основные показатели работы промышленности». Ежемесячный сборник ФСГС, 1995–2004. Отдел статистической информации ГУ–ВШЭ.

«Российский статистический ежегодник». ФСГС, 2001–2006, [www.gks.ru](http://www.gks.ru).

# Measurement of technological progress in Russia

Eugenia Nazrullaeva

*Higher School of Economics, Moscow, Russia*

Measurement of technological progress is one of core issues in the growth theory. Applied growth accounting based on the standard Solow residual approach to technology measurement assumes exogenous nature of productivity growth not explained by the dynamics of factor inputs. This exogenous unexplained residual is considered as a proxy for the technological change. The existing research in TFP measurement for Russian economy concentrates on the analysis of the transition process and considers the collapse in output as being caused by the fall in TFP, i.e. by technological “regress”. The declining TFP is sometimes explained by the fall in efficiency of production process or by the fall in capacity utilization. In this paper the TFP is estimated using Russian industry level data from the Federal State Statistics Service and Russian Economic Barometer from 1995 to 2004. TFP measurement based on observable estimates for capital and labor utilization accounting for the specifics of Russian industrial statistics reveals that the output growth after 1998 is accompanied by a slow technological improvement, and that a negative technology shock is observed before 1998.

*Keywords: total factor productivity, technological progress, capital and labor utilization, Russian production industries*

*JEL Classification: C22, E22, O33*

# Различные индексы прогнозирования экономической активности в России\*

Олег Демидов<sup>†</sup>

*Droege & Comp., Москва, Россия*

Рассматриваются различные способы вычисления индексов прогнозирования экономической активности в России. Первый способ – это методика, используемая российским Центром Развития и основанная на концепции «циклов роста». Второй подход объединяет в себе динамический метод главных компонент и динамический факторный анализ. Третий подход – это методика американского Национального Бюро Экономических Исследований, состоящая в построении диффузионных индексов с использованием динамической факторной модели. Данная работа является попыткой раскрыть преимущества и недостатки этих трех методов в применении к российским данным и предложить наилучшую методику для прогнозирования экономической активности в России.

*Ключевые слова: индекс экономической активности, опережающие и текущие индикаторы, метод динамических главных компонент, факторная модель, Россия*

*Классификация JEL: C32, E37, O11*

## 1 Введение

Система текущих и опережающих индикаторов – один из самых широко используемых методов прогнозирования будущей экономической активности. Этот метод впервые был применен в Соединенных Штатах в 1930-х годах. Регулярные публикации по используемым индикаторам были начаты в конце 1960-х и продолжаются по сей день (см. Stock & Watson, 1989, 2002). В 1980-х в OECD Statistics Directorate начали строить опережающие индикаторы для стран Евросоюза (см. Forni, Hallin, Lippi & Reichlin, 2001). Что касается России, то такие индикаторы строятся в Центре Развития с использованием своей собственной методологии (см. Smirnov, 2000, 2006). Однако методология Центра Развития имеет несколько недостатков. Данная статья является попыткой приложить европейскую процедуру и процедуру Стока и Уотсона к российским данным, а также улучшить некоторые аспекты данных методологий.

Обычно текущие и опережающие индикаторы экономической активности – это взвешенные средние определенных временных рядов, и необходим критерий того, как взвешивать различные компоненты индексов, чтобы учесть временные сдвиги между этими компонентами. Это порождает методологические проблемы: какой статистический критерий надо выбрать для построения этих средних, и какой метод необходим, чтобы реализовать оценивание для больших панелей временных рядов?

Метод, предложенный Центром Развития, состоит из следующих шагов. Во-первых, проводится датирование поворотных точек бизнес-циклов. Во-вторых, выбираются начальные опережающие индикаторы согласно тому критерию, что их поворотные точки предшествуют поворотным точкам бизнес-циклов. В-третьих, единый опережающий индикатор вычисляется как взвешенное среднее всех выбранных опережающих индикаторов с весами, обратно пропорциональными волатильностям этих индикаторов.

\*Цитировать как: Демидов, Олег (2008). «Различные индексы прогнозирования экономической активности в России», Квантиль, №5, стр. 83–102. Citation: Demidov, Oleg (2008). “Different indexes for forecasting economic activity in Russia,” *Quantile*, No.5, pp. 83–102.

<sup>†</sup>Адрес: 115054, Россия, г. Москва, ул. Бахрушина, 32/1. Электронная почта: [demidovs@inbox.ru](mailto:demidovs@inbox.ru)

У этой процедуры есть несколько очевидных недостатков. Во-первых, требуется датирование поворотных точек всех переменных, что порой вызывает достаточно большие трудности из-за высокой волатильности макроэкономических переменных. Во-вторых, процедура отбора имеет неопределенность, поскольку некоторые переменные могут быть как опережающими, так и запаздывающими в различных точках. Наконец, простое усреднение по опережающим индикаторам не принимает во внимание временные сдвиги и взаимную корреляцию между переменными.

В то же время метод, предложенный в Forni, Hallin, Lippi & Reichlin (2001), позволяет избежать некоторые из этих недостатков. Европроцедура использует основную идею динамической факторной модели: благодаря сильным совместным движениям макроэкономических временных рядов динамика каждой переменной может быть представлена в виде суммы малоразмерной компоненты, общей для всех переменных, и ортогонального идиосинкратического остатка (см. Sargent & Sims, 1977). Факторные модели дают статистическую основу для построения индексов из большого количества временных рядов. Модель, использованная в работе Forni, Hallin, Lippi & Reichlin (2001), носит название обобщенной динамической факторной модели. Она объединяет в себе динамический факторный анализ и метод динамических главных компонент. Так что оценщик строится таким образом, чтобы принять во внимание временные сдвиги между переменными, правильно взвешивая опережающие и запаздывающие переменные. Это обстоятельство позволяет оценивать модель, используя все переменные – опережающие, текущие и запаздывающие – без необходимости классифицировать их априори.

Однако метод, предложенный в Forni, Hallin, Lippi & Reichlin (2001), имеет ряд недостатков, которые будут обсуждаться позже, и будут предложены возможные усовершенствования этого метода. Также необходимо принять во внимание то, что построение индексов для России имеет ряд особенностей по сравнению с построением индексов для Еврозоны, таких как отсутствие достаточного объема качественных данных и специфика российских бизнес-циклов.

Более того, вероятно, европроцедура не является оптимальной с точки зрения прогнозирования, так как требует использования приближенной динамической факторной модели. Поэтому также необходимо изучить предсказательные свойства методологии Stock & Watson (2002), которая использует точную статическую модель для прогнозирования. Все методы прогнозирования сравниваются с помощью среднеквадратичной ошибки предсказания (СКОП) вневыборочного прогноза для нахождения наилучшего способа прогнозирования будущего ВВП.

Работа организована следующим образом. Раздел 2 обсуждает теоретические основы циклических индикаторов Центра Развития, опережающих индикаторов европроцедуры и диффузионных индексов Стока и Уотсона. Процедура оценивания и результаты для России представлены в разделе 3. Завершает работу Заключение.

## 2 Теоретическая часть

### 2.1 Система циклических индикаторов Центра Развития

Как уже было сказано, методология Центра Развития включает следующие шаги. Сначала поворотные точки бизнес-циклов датируются по поворотным точкам индекса промышленного производства с удаленной сезонностью. Затем рассматриваются поворотные точки 40 индикаторов, чтобы решить, являются ли эти индикаторы опережающими. После тестирования этих переменных 7 из них включены в систему опережающих индикаторов для России (детали см. в Smirnov, 2006):

- Средняя месячная цена нефти марки Urals;

- Доля предприятий с увеличившимся или неизменным спросом;
- Доля предприятий, не имеющих избыточных запасов конечной продукции;
- Денежный агрегат М2, включая рубли и иностранную валюту;
- Обратный индекс реального эффективного обменного курса рубля, вычисленный ЦБ РФ, против корзины валют российских торговых партнеров;
- Индекс РТС;
- Межбанковская процентная ставка.

Наконец, единый опережающий индекс вычислен по следующей процедуре:

1. Для всех подобранных опережающих индикаторов  $X^i$  ( $i$  – номер ряда) вычисляются симметричные ежемесячные приросты ( $t$  – текущий момент времени):

$$x_t^i = 200 \frac{X_t^i - X_{t-1}^i}{X_t^i + X_{t-1}^i}. \quad (1)$$

2. Оцениваются средние значения  $x_{av}^i$  и стандартные отклонения  $s^i$  полученных приростных рядов:

$$x_{av}^i = \frac{\sum_{t=1}^T x_t^i}{T}, \quad s^i = \sqrt{\frac{\sum_{t=1}^T (x_t^i - x_{av}^i)^2}{T-1}}, \quad (2)$$

где  $T$  – число периодов наблюдений.

3. Для каждого  $t$  рассчитываются усредненный прирост  $g_t$ , а также среднее  $g_{av}$  и стандартное отклонение  $s^g$  ряда  $g_t$ :

$$g_t = \frac{\sum_{i=1}^n (x_t^i / s^i)}{n}, \quad g_{av} = \frac{\sum_{t=1}^T g_t}{T}, \quad s^g = \sqrt{\frac{\sum_{t=1}^T (g_t - g_{av})^2}{T-1}}, \quad (3)$$

где  $n$  – число исходных индикаторов.

4. Шаги 1–2 выполняются для индексов промышленного производства  $y$ . Результат – среднее  $y_{av}$  и стандартное отклонение  $s^y$  приростного ряда.
5. Показатель  $g_t$  корректируется так, чтобы его волатильность была равна волатильности приростов индекса промышленного производства:

$$G_t = g_t \frac{s^y}{s^g}. \quad (4)$$

6. По рекурсивной формуле рассчитываются значения сводного опережающего индекса  $Z_t$  путем возврата от приростов к агрегату:

$$Z_1 = \frac{200 + G_1}{200 - G_1}, \quad Z_t = Z_{t-1} \frac{200 + G_t}{200 - G_t}. \quad (5)$$

7. Полученный индекс  $Z$  приводится к той же базе, которую имеет индекс промышленного производства (1996 = 100). Для этого все значения  $Z_t$  делятся на среднемесячный уровень 1996 г. и умножаются на 100. В итоге получается, что сводный опережающий индекс имеет ту же базу, что и индекс промышленного производства, а симметричные приросты двух этих показателей имеют одинаковую волатильность.

## 2.2 Построение опережающего и текущего индикаторов по методике Forni, Hallin, Lippi & Reichlin (2001)

В работе Forni, Hallin, Lippi & Reichlin (2001) было предложено оценивать обобщенную динамическую факторную модель по панели, содержащей макроэкономические переменные всех стран Еврозоны, и извлекать ненаблюдаемую общую компоненту из каждого ряда панели. Для этого требуется «очищать» переменные как от ошибок измерения, так и от локальных, то есть присущих только этой переменной, компонент. Тогда текущий индикатор для Еврозоны определяется как взвешенное среднее общих («очищенных») компонент ВВП всех стран. Поэтому так построенный индекс со стохастической размерностью возможно больше единицы учитывает взаимную корреляцию внутри и между странами и суммирует ту часть динамики ВВП, которая наиболее коррелирована с остальной экономикой и поэтому наиболее полезна для оценивания бизнес цикла Еврозоны.

Информацию о ведуще-запаздывающих связях между переменными в панели и текущим индикатором можно затем восстановить по оценкам ковариаций между общими компонентами переменных и текущим индикатором. Эта информация позволяет определить опережающие переменные во всех экономиках Еврозоны, которые затем могут быть агрегированы в единый опережающий индикатор. Наконец, опережающий индикатор можно использовать для прогнозирования текущего индикатора.

Полностью теоретическую основу обобщенной динамической факторной модели можно найти в оригинальной статье Forni, Hallin, Lippi & Reichlin (2001). В данном разделе будут представлены только общее описание метода, его критика и возможные пути улучшения.

### «Очищение» переменных с помощью метода динамических главных компонент

Главная идея здесь заключается в том, чтобы разделить каждую переменную на две ортогональные компоненты, первая из которых представляет часть индивидуальной динамики, сильно коррелированную с остальной панелью, а вторая – не интересующую нас идиосинкратическую часть.

Предположим, что все макроэкономические ряды, преобразованные должным образом, представляют собой реализацию стационарного в широком смысле  $n$ -мерного векторного процесса  $x_t = (x_{1t}, \dots, x_{nt})'$  с нулевым средним. Необходимо выделить с помощью малого числа  $q$  «агрегированных индексов» часть динамики, общую для всех процессов  $x_{it}$  для  $i = 1, \dots, n$ :

$$z_{ht} = p_h(L)x_t, \quad h = 1, \dots, q,$$

где  $q$  процессов  $z_{ht}$ ,  $h = 1, \dots, q$ , являются линейными комбинациями различных лагов переменных из  $x_t$ ,  $L$  – оператор лагирования и  $p_h(L) - 1 \times n$  вектор двустороннего линейного фильтра.

Наконец, рассмотрим декомпозицию

$$x_t = \gamma_t^q + \zeta_t^q = C^q(L)z_t^q + \zeta_t^q = K^q(L)x_t + \zeta_t^q, \quad (6)$$

где  $\gamma_t^q = (\gamma_{1t}^q, \dots, \gamma_{nt}^q)$  – проекция  $x_t$  на настоящее, прошлое и будущее  $z_t^q = (z_{1t} \dots z_{qt})'$ , и  $\zeta_t^q$  – остаточный вектор. Потребуем, чтобы фильтр  $p_h(L)$  и соответствующие процессы  $z_{ht}$ ,  $h = 1, \dots, q$ , удовлетворяли условию максимизации объясненной дисперсии для данного  $q$

$$\sum_{j=1}^n \mathbb{V}[\gamma_{jt}^q]. \quad (7)$$

Процессы  $z_{1t}, \dots, z_{nt}$  называются «главными компонентами временного ряда» или «динамическими главными компонентами» процесса  $x_t$ . Эти компоненты так же связаны с собственными числами и векторами матрицы спектральной плотности процесса  $x_t$ , как статистические главные компоненты связаны с собственными числами и векторами ковариационной матрицы.

Точнее, пусть  $P_x(k) = \mathbb{E}[x_t x'_{t-k}]$  – матрица корреляционной функции стационарного процесса  $x_t$ , и  $\Sigma(\theta)$ ,  $-\pi < \theta < \pi$ , – матрица спектральной плотности  $x_t$ . Матрицы корреляционной функции и спектральной плотности связаны преобразованием Фурье, то есть

$$P_x(k) = \int_{-\infty}^{\infty} e^{i\theta k} \Sigma(\theta) d\theta, \quad \Sigma(\theta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\theta k} P_x(k) dk.$$

Пусть вектор  $p_h(e^{-i\theta})$  будет собственным вектором, соответствующим  $h$ -му собственному числу матрицы  $\Sigma(\theta)$ , у которой собственные числа упорядочены по убыванию. Кроме того, обозначим это собственное число как  $\lambda_h(\theta)$  и положим  $\lambda_h = \frac{1}{2\pi} \int_{-\pi}^{\pi} \lambda_h(\theta) d\theta$ . Тогда максимальная объясненная дисперсия (7) равна  $\lambda_1 + \dots + \lambda_q$ , и процент объясненной дисперсии дается отношением

$$\frac{\lambda_1 + \dots + \lambda_q}{\lambda_1 + \dots + \lambda_n}. \quad (8)$$

Далее, есть возможность выписать явные выражения для фильтров  $C^q(L)$  и  $K^q(L)$ , заданных в (6). Эти выражения выглядят так:

$$\begin{aligned} C^q(L) &= (p_1(F)' \cdots p_q(F)'), \\ K^q(L) &= C^q(L) C^q(F)' = p_1(F)' p_1(L) + \cdots + p_q(F)' p_q(L), \end{aligned}$$

в тесной аналогии со статическим случаем. Поэтому фильтр  $K^q(L)$  можно найти следующим образом. Во-первых,

$$K^q(e^{-i\theta}) = p_1(e^{-i\theta})' p_1(e^{-i\theta}) + \cdots + p_q(e^{-i\theta})' p_q(e^{-i\theta}).$$

Во-вторых, используя обратное преобразование Фурье, можно вычислить

$$K^q(k) = \int_{-\infty}^{\infty} e^{i\theta k} K^q(e^{-i\theta}) d\theta.$$

Наконец, получаем выражение для  $K^q(L)$ :

$$K^q(L) = \int_{-\infty}^{\infty} K^q(k) L^k dk.$$

Также необходимо найти матрицу спектральной плотности главных компонент  $\Sigma^\gamma(\theta)$ , которая задается выражением

$$\Sigma^\gamma(\theta) = K^q(e^{-i\theta}) \Sigma(\theta) K^q(e^{i\theta})'.$$

Следует отметить, что все фильтры и матрица спектральной плотности должны оцениваться из конечной выборки процесса  $x_t$  длины  $T$ . Более того, все непрерывные модели должны быть заменены на дискретные аналоги. Оценитель, используемый здесь (который обозначен как  $\chi_{nt}^T$  в силу причин, которые будут ясны ниже), детально описан в Приложении.

## Главные компоненты и обобщенная динамическая факторная модель

Процедура «очистки» основана на выборе малого числа  $q$  и поэтому может содержать существенную неопределенность. Однако, если предположить, что  $x_t$  генерируется факторной моделью, то можно построить достаточно четкий критерий выбора  $q$ . В динамическом факторном подходе переменные представляются в виде суммы двух ненаблюдаемых компонент: общей компоненты, формируемой малым числом факторов, присущих всем переменным системы, и идиосинкратической компоненты, которая некоррелирована с общими компонентами и присуща отдельной переменной. Если придерживаться этой модели, то отделение идиосинкратической части от общих компонент представляется достаточно естественной процедурой «очистки». Чтобы лучше понять данную факторную модель, будет удобно полагать, что вектор  $x_t$  формируется первыми  $n$  элементами бесконечной последовательности  $x_{jt}, j = 1, \dots, \infty$ . Чтобы подчеркнуть зависимость от  $n$ , будем писать  $x_{nt}$  вместо  $x_t$ . В данной модели

$$x_{nt} = \chi_{nt} + \xi_{nt} = B_n(L)u_t + \xi_{nt}, \quad (9)$$

где  $\chi_{nt}$  – общая компонента,  $u_t = (u_{1t}, \dots, u_{qt})'$  – вектор общих шоков, ковариационно стационарный  $q$ -векторный процесс с невырожденной матрицей спектральной плотности,  $B_n(L)$  – матрица двустороннего квадратично суммируемого фильтра, и идиосинкратические компоненты  $\xi_{nt}$  ортогональны к  $u_{t-k}$  для каждого  $k$ .

Используя некоторые предположения, Forni, Hallin, Lippi & Reichlin (2001) показывают глубокую связь между представлениями (6) и (9). Основная интуиция здесь в следующем. При вычислении главных компонент мы берем среднее по переменным, входящим в  $x_t$ . Когда  $n$  большое, мы получаем некий результат типа закона больших чисел. Идиосинкратические компоненты исчезают, а остаются только линейные комбинации различных лагов главных компонент. Такие линейные комбинации охватывают почти то же динамическое пространство, что и общие факторы.

Данный результат дает простой критерий для выбора числа используемых главных компонент. Если модель (9) справедлива, то собственные числа  $\lambda_{hn} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \lambda_{hn}(\theta) d\theta$  ограничены для  $h > q$  и расходятся для  $h \leq q$  при  $n \rightarrow \infty$ . Следовательно, при больших  $n$  можно ожидать большой «скачок» при переходе от  $\lambda_{qn}$  к  $\lambda_{q+1,n}$ . Поэтому главные компоненты стоит добавлять в процедуру оценивания до тех пор, пока увеличение в объясненной вариации не станет меньше некоторого выбранного значения. Точнее, если обозначить через  $\lambda_{hn}^T$ , где  $T$  – число наблюдений, оценку  $\lambda_{hn}$ , и выбрать  $\alpha \in (0, 1)$ , то критерий выбора  $q = q^*$  будет следующим:

$$\frac{\lambda_{q^*n}^T}{\sum_{h=1}^n \lambda_{hn}^T} > \alpha, \quad \frac{\lambda_{q^*+1,n}^T}{\sum_{h=1}^n \lambda_{hn}^T} < \alpha. \quad (10)$$

### Процедура оценивания

Теперь, после описания теоретической основы метода, можно представить полную процедуру для построения текущего и опережающего индикаторов и идентификации поворотных точек бизнес-цикла. В данном разделе обсуждается подход Forni, Hallin, Lippi & Reichlin (2001) и предлагаются возможные усовершенствования этого подхода.

Процедура оценивания состоит из следующих шагов.

**Шаг 1: выбор переменных для включения в панель** На первом шаге процедуры необходимо решить, какие переменные стоит включать в массив данных, по которому будут вычисляться общие компоненты. Как было показано, с теоретической точки зрения все

доступные переменные должны быть включены в процедуру оценивания. На практике, однако, не рекомендуется использовать переменные с большой идиосинкратической компонентой, поскольку она может быть неправильно интерпретирована как дополнительная общая компонента. Для выбора массива данных Forni, Hallin, Lippi & Reichlin (2001) предлагают следующую последовательность действий. Они выделяют 2 набора переменных: «ядро», содержащее переменные, которые точно надо использовать, и набор «кандидатов» на включение. Затем они преобразовывают переменные так, чтобы получить стационарность, и нормализуют их, вычитая среднее и деля на стандартное отклонение. Стационарность достигается взятием первой разности (или разности в логарифмах). Очевидно, однако, что в данных, по крайней мере российских, присутствуют большие сезонные компоненты, которые также должны быть исключены. В данной статье сезонность удаляется с помощью регрессии на сезонные фиктивные переменные, хотя есть и другие, возможно более лучшие, способы.

Используя ядро, Forni, Hallin, Lippi & Reichlin (2001) фиксируют  $q^*$  согласно критерию (10) и вычисляют соответствующую степень общности как процент объясненной дисперсии (8). Как только эта степень  $\mu$  зафиксирована, они оценивают по очереди каждого кандидата. Для оценивания формируется расширенная система, состоящая из ядра и кандидата, для нее вычисляется значение (8) с  $q = q^*$ , и кандидат проходит экзамен, если значение (8) больше, чем  $\mu$ . Успешные кандидаты не добавляются к ядру до тех пор, пока не будут оценены все кандидаты. Итоговый массив данных включает в себя ядро и всех успешных кандидатов.

На этом этапе также оцениваются вектор общих компонент  $\chi_{nt}$  и матрица его спектральной плотности  $\Sigma^{\chi}(\theta)$  с  $q = q^*$ , как это описано в Приложении.

Следует отметить, что процедура отбора переменных, предложенная в Forni, Hallin, Lippi & Reichlin (2001), имеет ряд недостатков. Во-первых, имеется неопределенность в выборе ядра и кандидатов. Одна из проблем состоит в том, что данные для некоторых переменных из ядра могут быть недостаточно хороши, что может привести к получению малого  $\mu$  и включению переменных, слабо связанных с ВВП. Другая проблема возникает, если, наоборот, выбрано слишком хорошее ядро с высоким  $\mu$ , тогда можно потерять переменные с немного меньшей связью с ВВП, которые имеют опережающую динамику и могут быть эффективно использованы при построении опережающего индекса. Во-вторых, предложенный критерий включения переменных в итоговое оценивание не принимает во внимание временные сдвиги и корреляции переменных относительно ВВП. Чтобы по возможности исправить эти недостатки, была разработана альтернативная процедура отбора, которая будет представлена в последнем параграфе этого раздела.

**Шаг 2: текущий индикатор** Текущий индикатор  $\Delta C_t$  равен общей компоненте ВВП, умноженной на вычисленное стандартное отклонение ВВП. Пусть ВВП соответствует первой компоненте  $x_{nt}$ . Тогда текущий индикатор есть

$$\Delta C_t = \chi_{1nt} \sigma_1.$$

Агрегированный индикатор в уровнях вычисляется путем добавления дрейфа  $\delta$  к текущему индикатору, как кумулятивная сумма  $\Delta C_t + \delta$ :

$$C_t = C_0 + \delta t + \sum_{\tau=1}^t \Delta C_{\tau}.$$

**Шаг 3: процикличные и противоцикличные переменные** Теперь можно классифицировать общие компоненты  $\chi_{jt}$  как находящиеся в фазе или противофазе по отношению к текущему индикатору. Для этого с использованием оценки  $\Sigma^{\chi}(\theta)$  вычисляется спектральная плотность каждой общей компоненты по отношению к  $\Delta C_t$ . Затем вычисляется аргумент каждой из этих плотностей, который является фазовым сдвигом переменной относительно

российского ВВП. Обозначим фазовый сдвиг общей компоненты  $\chi_{jt}$  как  $\phi_j(\theta)$ ,  $-\pi < \theta \leq \pi$ . На нулевой частоте фазовый сдвиг может быть либо 0, либо  $\pi$  в зависимости от того, положительна или отрицательна долгосрочная корреляция. Пусть долгосрочная корреляция  $j$ -й компоненты с текущим индикатором равна  $\rho_j$ , тогда

$$\rho_j = \Sigma_{1j}^X(0).$$

Если она положительна, то переменная проциклична, иначе – противоциклична.

Долгосрочную корреляцию также можно использовать в процедуре отбора переменных как индикатор связи переменной с ВВП.

**Шаг 4: опережающие, текущие и запаздывающие переменные** Для такой классификации переменных необходимо вычислить фазовый сдвиг переменной  $\chi_{jt}$ ,  $j = 1, \dots, n$ , относительно  $\Delta C_t$  на типичной частоте бизнес цикла  $\theta^* > 0$ . Обозначим этот фазовый сдвиг как  $\psi_j(\theta^*)$ . Тогда классифицируем  $\chi_{jt}$  как текущую, если  $|\psi_j(\theta^*)|$  меньше, чем заданное значение  $\tau$ , опережающую, если  $\psi_j(\theta^*) < -\tau$ , и запаздывающую, если  $\psi_j(\theta^*) > \tau$ . Фазовый сдвиг  $\psi_j(\theta^*)$  равен

$$\psi_j(\theta^*) = \text{Arg}(\Sigma_{1j}^X(\theta^*)),$$

где  $\text{Arg}$  означает аргумент комплексного числа. Он также может быть использован в процедуре отбора переменных. Например, опережающие переменные иногда можно включить в финальное оценивание, даже если они не очень сильно коррелированы с ВВП.

**Шаг 5: опережающий и запаздывающий индикаторы** Построение опережающего индикатора как обычного среднего ведущих переменных, предложенное Forni, Hallin, Lippi & Reichlin (2001), не учитывает фазовые сдвиги и корреляцию переменных относительно ВВП. Похоже, что корреляции переменных с ВВП могут быть частично учтены усреднением с весами, пропорциональными долгосрочным корреляциям  $\rho_j$ :

$$\Delta L_t = \frac{\sum_{j \in \ell} \rho_j \chi_{jt}}{\sum_{j \in \ell} |\rho_j|}.$$

Здесь  $\ell$  обозначает набор  $j$  таких, что  $\chi_{jt}$  являются опережающими.

Аналогичное выражение имеет место для запаздывающего индикатора за исключением того, что вместо опережающих переменных берутся запаздывающие.

**Шаг 6: Добавление значений в крайних точках** Следует отметить, что общие компоненты  $\chi_{jnt}^T$  вычисляются с помощью двустороннего фильтра  $K_n^T(L)$  длины  $2M + 1$  (см. Приложение). Поэтому в начале и в конце периода наблюдений, то есть при  $t = 1, \dots, M$  и  $t = T - M + 1, \dots, T$ , невозможно построить оценки общих компонент стандартным образом. По этой причине Forni, Hallin, Lippi & Reichlin (2001) заполняют оценки в крайних точках проекциями  $\chi_{jnt}^T$  на текущие и опережающие либо запаздывающие переменные, соответственно.

Такие проекции  $\chi_{jnt}^T$  не используют все возможности, которые дает процедура оценивания. Здесь можно использовать весь двусторонний фильтр и все переменные для заполнения значений в крайних точках.

Если быть более точным, то в начале интервала наблюдений при  $t = 1, \dots, M$  необходимо сделать следующие шаги. Сначала находим  $x_0$  как проекцию  $\chi$  на момент 0, используя проекцию фильтра  $K_n^T(L)$

$$x_0 = \chi_0^p = \sum_{k=-M}^{-1} K_k x_{0-k}$$

(здесь отброшены индексы  $T$  и  $n$ , индекс  $p$  обозначает проекцию). Затем вычисляем  $\chi_M$ , используя полный фильтр  $K_n^T(L)$ :

$$\chi_M = \sum_{k=-M}^M K_k x_{M-k}.$$

Далее аналогичная процедура повторяется для  $t = M - 1, \dots, 1$ .

В конце периода наблюдений при  $t = T - M + 1, \dots, T$  используется такой же метод добавления значений общих компонент, как и в начале периода.

**Шаг 7: поворотные точки** Поворотные точки наилучшим образом идентифицируются по текущему индикатору в уровнях без дрейфа. Поворотные точки определяются как даты  $t^*$ , в которые текущий индикатор  $C_t$  достигает локальных минимумов или максимумов.

**Шаг 8: прогноз** Проекция  $\Delta C_{t+1}$  на опережающие и текущие переменные является прогнозом текущего индикатора на один шаг вперед. В работе Forni, Hallin, Lippi & Reichlin (2001) он вычисляется с использованием той же модели, которая применяется для добавления значений в крайних точках.

Критика здесь та же, что и для шагов 5 и 6. Поэтому для прогноза текущего индикатора на период вперед можно использовать метод из шага 6 с добавлением одной дополнительной точки в конце интервала наблюдений. Эту точку также можно рассматривать в качестве опережающего индикатора.

### Альтернативная процедура отбора переменных

Во-первых, все правильно преобразованные переменные включаются в оценивание, и находятся соответствующие  $q^*$ ,  $\mu$ ,  $\rho_j$ ,  $\psi_j(\theta^*)$ .

Во-вторых, выбираются некие  $\mu^*$  и  $\rho^*$ , и находятся переменные с  $\rho_j < \rho^*$ . Тогда эти переменные исключаются по очереди, и вычисляются новые  $\mu_j$ . Каждый раз исключаем только одну переменную из начальной панели. После этого находим наименьший  $\mu_j$ , и если  $\mu - \mu_j > \mu^*$ , то, возможно, есть необходимость исключить эту переменную. Единственное, что может спасти переменную от исключения в данной ситуации – это ее опережающая динамика относительно ВВП. Здесь следует принимать во внимание временной сдвиг переменной и экономическую интуицию относительно связи этой переменной с ВВП.

Наконец, после решения об исключении переменной берется новая панель и повторяется заново та же последовательность действий. Процедура заканчивается, когда больше не остается кандидатов на исключение.

### 2.3 Диффузионные индексы

Здесь приводится общее описание методологии Stock & Watson (2002).

Пусть  $y_t$  – скалярный временной ряд, а  $X_t$  –  $N$ -мерный векторный временной ряд. Переменная  $y_t$  является переменной для прогнозирования, в то время как  $X_t$  содержит полезную информацию для прогнозирования  $y_{t+1}$ . Предполагается, что  $X_t$  может быть представлен факторной структурой

$$X_t = \Lambda_t F_t + e_t, \tag{11}$$

где  $F_t - r \times 1$  вектор общих факторов,  $\Lambda_t - N \times r$  матрица коэффициентов при общих факторах и  $e_t - N \times 1$  вектор идиосинкратических шумов. В общем случае идиосинкратические шумы коррелированы по времени и по рядам.

Главная цель – это оценивание  $\mathbb{E}[y_{t+1}|X_t]$ . Здесь  $y_{t+1}$  моделируется как

$$y_{t+1} = \beta_t' F_t + \epsilon_{t+1}, \quad (12)$$

где  $\mathbb{E}[\epsilon_{t+1}|X_t, y_t, \beta_t, X_{t-1}, y_{t-1}, \beta_{t-1}, \dots] = 0$ . Это выражение включает в себя три предположения: что  $\mathbb{E}[y_{t+1}|X_t, y_t, \beta_t, X_{t-1}, y_{t-1}, \beta_{t-1}, \dots]$  зависит от  $X_t$  только через  $F_t$  и никак иначе, что лаги  $F_t$  не входят в (12), и что лаги  $y_t$  не входят в (12).

Матрица коэффициентов  $\Lambda_t$  ( $N \times r$ ) и коэффициенты  $\beta_t$  ( $r \times 1$ ) меняются со временем согласно AR(1)-процессам

$$\begin{aligned} \Lambda_t &= \Lambda_{t-1} + h\zeta_t, \\ \beta_t &= \beta_{t-1} + \eta_t, \end{aligned} \quad (13)$$

где  $h$  – диагональная  $N \times N$  матрица, а  $\eta_t$  и  $\zeta_t$  – соответственно  $r \times 1$  и  $N \times r$  стохастические шумы.

Следуя Стоку и Уотсону, динамическая факторная модель оценивается в ее статической форме. Здесь используется подход квазиправдоподобия, в том смысле, что оценивание мотивировано при сильных параметрических предположениях, но состоятельность оцениваемых факторов показана при более слабых непараметрических предположениях (детали см. в Stock & Watson, 2002). Для определения стратегии оценивания предположим, что  $h = 0$ , то есть  $\Lambda_t = \Lambda_0 = \Lambda$ , и  $e_{it}$  – IID  $N(0, \sigma_e^2)$  и независимы между рядами. Также отклонимся от рассмотрения  $F_t$  в рамках динамической факторной модели, в которой  $F_t$  подчиняется стохастическому процессу. Вместо этого будет рассматривать  $F_t$  как оцениваемую  $r \times T$  матрицу неизвестных неслучайных параметров. В таких предположениях метод максимального правдоподобия находит  $(F, \Lambda)$  как решение задачи нелинейных наименьших квадратов с целевой функцией

$$V_{NT}(F, \Lambda) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (X_{it} - \Lambda_i' F_t)^2. \quad (14)$$

Пусть  $(F^*, \Lambda^*)$  минимизирует значение  $V_{NT}(F, \Lambda)$ . Тогда удовлетворяются условия первого порядка

$$\Lambda_i^* = \left( \sum_{t=1}^T F_t^* F_t^{*'} \right)^{-1} \sum_{t=1}^T F_t^* X_{it}, \quad (15)$$

$$F_t^* = \left( \sum_{i=1}^N \Lambda_i^* \Lambda_i^{*'} \right)^{-1} \sum_{i=1}^N \Lambda_i^* X_{it}. \quad (16)$$

Неизвестные параметры можно оценить с помощью решения двух задач на собственные числа. Первая задача ставится заменой (15) в (14) и решается выбором в качестве  $F$  собственных векторов, соответствующих  $k$  наибольшим собственным числам  $T \times T$  матрицы  $\frac{1}{N} \sum_{i=1}^N X_i X_i'$ , где  $X_i = (X_{i1}, X_{i2}, \dots, X_{iT})'$ .

Вторая задача на собственные числа ставится заменой (16) в (14) и решается выбором в качестве  $\Lambda$  собственных векторов, соответствующих  $k$  наибольшим собственным числам  $N \times N$  матрицы  $\frac{1}{T} \sum_{t=1}^T X_t X_t'$ . Эти собственные вектора являются первыми  $k$  главными компонентами  $X_t$ .

Однако, есть неопределенность в выборе числа оцениваемых факторов  $k$ . В идеале это число должно быть равно реальному числу факторов  $r$ . Следуя Стоку и Уотсону, будем рассматривать задачу оценивания числа факторов, входящих в уравнение для прогноза, используя информационный критерий в форме

$$BIC_k = \ln \frac{SSR(k)}{T} + \frac{\ln T}{T} k. \quad (17)$$

Здесь  $SSR(k)$  – сумма квадратов остатков оценивания (12) с помощью МНК, используя  $k$  оцениваемых факторов. Информационный критерий оценивает  $r$ , решая задачу  $\min_k BIC_k$ .

Помимо байесовского информационного критерия, который может давать смещенную оценку  $r$ , возможно также использование критерия из Bai & Ng (2002).

### 3 Процедуры оценивания и результаты

#### Данные

Изначальная панель данных, использованная в данной работе, состояла из 24 временных рядов для России. Среди них были переменные из реального сектора, рынка труда и рынка капитала, ценовые и финансовые индексы, денежные агрегаты, обменные курсы и процентные ставки. Использованы два основных источника данных: база Международного Валютного Фонда (МВФ) и база данных Сергея Смирнова из Центра Развития. Выбран временной интервал квартальных данных базы МВФ, начиная с 1-го квартала 1995 г. и заканчивая 1-м кварталом 2007 г. (в общей сложности 49 точек). Месячные данные из базы Сергея Смирнова были трансформированы в квартальные данные на том же интервале времени. Более подробное описание переменных, источников и трансформации данных приведено в таблице 1.

Таблица 1: Описание переменных, источников и трансформации данных

Переменная	Пояснение	Источник	Трансформация
GDP	ВВП	IMF	Разность в логарифмах
Inv	Инвестиции	IMF	Разность в логарифмах
GovCons	Госрасходы	IMF	Разность в логарифмах
PrivCons	Частное потребление	IMF	Разность в логарифмах
IndEmploy	Занятость	IMF	Разность в логарифмах
Wages	Изменения зарплаты (проценты)	IMF	Нет
CPI	Потребительские цены (2000=100)	IMF	Разность в логарифмах
PPI	Изменение в PPI (проценты)	IMF	Нет
M+QM	Деньги + квази-деньги	IMF	Разность в логарифмах
Madj	Деньги без сезонности	IMF	Разность в логарифмах
CapAccounts	Баланс капитала	IMF	Разность в логарифмах
BankReserves	Банковские резервы	IMF	Разность в логарифмах
REER	Реальный эффективный обменный курс	IMF	Разность в логарифмах
NEER	Номинальный эффективный обменный курс	IMF	Разность в логарифмах
MMrate	Межбанковская процентная ставка	IMF	Разность
UralsPrice	Цена на нефть Urals	Smirnov	Разность в логарифмах
ShareNoExc	Доля предприятий без избыточных запасов готовой продукции	Smirnov	Разность в логарифмах
RTS	Индекс RTS	Smirnov	Разность в логарифмах
Agriculture	Сельское хозяйство Y-o-Y	Smirnov	Логарифм
Construction	Строительство Y-o-Y	Smirnov	Логарифм
TransTurnover	Объем грузоперевозок	Smirnov	Разность в логарифмах
Retail	Торговля Y-o-Y	Smirnov	Логарифм
Services	Услуги Y-o-Y	Smirnov	Логарифм
ShareIorS	Доля предприятий с увеличившимся или неизменным спросом	Smirnov	Разность в логарифмах

Замечания: Y-o-Y означает относительное изменение величины по сравнению с прошлым годом.

## Циклические индикаторы для России

Циклический индикатор строится на основе метода Центра Развития как взвешенное среднее семи временных рядов с опережающей динамикой (см. подраздел 2.1). Рис. 1 показывает в сравнении этот индикатор и российский ВВП в логарифмах. Можно увидеть, что циклический индикатор не является хорошим предсказателем будущего ВВП, однако в целом тренды достаточно схожи.

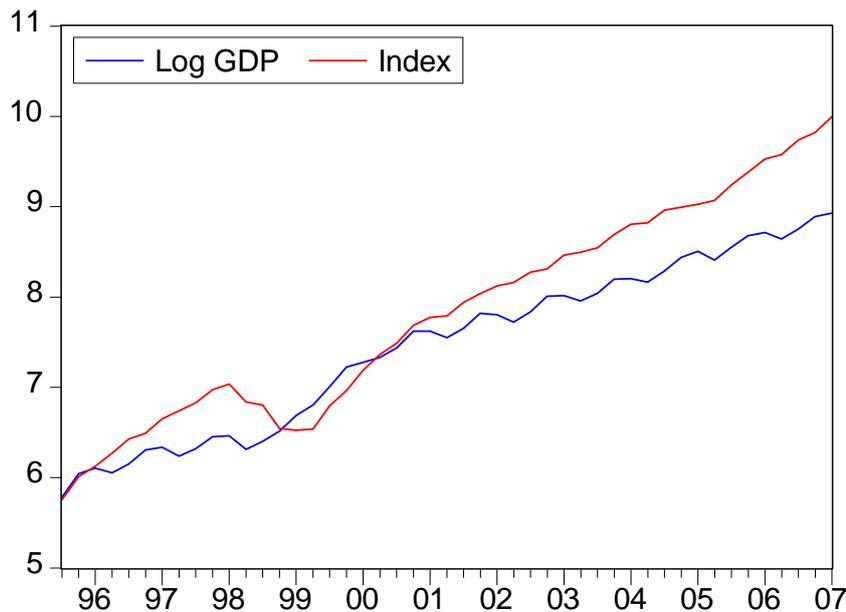


Рис. 1: ВВП в логарифмах и циклический индикатор Центра Развития

Среднеквадратичное отклонение прогноза (СКОП) процентного изменения ВВП для циклического индикатора составляет 0,00888, что соответствует стандартному отклонению прогноза около 0,0942. Такое стандартное отклонение выглядит достаточно большим, поскольку оно предполагает среднюю ошибку прогноза более 9 процентов квартального изменения ВВП. Поэтому необходимо искать другие пути прогнозирования будущей экономической активности для России.

## Авторегрессия

Простейший путь прогнозирования любого временного ряда – это использование одноперенного авторегрессионного прогноза, основанного на модели

$$\Delta \ln y_{t+1} = \mu + \sum_{j=0}^p \gamma_j \Delta \ln y_{t-j} + \epsilon_t.$$

Здесь  $\Delta \ln y_t$  – разность в логарифмах или процентное изменение ВВП в момент  $t$ ,  $\mu$  и  $\gamma_j$  – коэффициенты авторегрессии,  $\epsilon_t$  – инновация, а  $p$  выбирается, исходя из минимизации СКОП.

Вычисления проводились для двух типов преобразования данных – данных с удаленной сезонностью и данных, в которых удалены сезонность и тренд. Сравнение этих двух типов мотивировано наличием структурного сдвига в российских данных в 3-м квартале 1998 г. и, следовательно, двух различных трендов в логарифме ВВП. Поэтому может оказаться неправильным вычитание одинакового тренда из обеих частей наблюдений.

Также вычисления проводились для двух различных интервалов времени – полного интервала (1995Q1:2007Q1 – 49 наблюдений) и укороченного (1998Q3:2007Q1 – 35 наблюдений). В первом интервале больше наблюдений, однако внутри него есть структурный сдвиг, второй – короче, но, возможно, имеет более лучшие структурные характеристики на всем периоде.

Для всех случаев был вычислен СКОП, и лучшим по этому показателю стал прогноз по данным с удаленной сезонностью на полном интервале времени с 7-ю лагами. С точки зрения СКОП, этот прогноз превзошел почти в 10 раз прогноз, основанный на методологии Центра Развития.

### Опережающий и текущий индикаторы

Теперь оценим текущий, опережающий и запаздывающий индикаторы для России, следуя процедуре, описанной в подразделе 2.2. Сначала используем полный интервал наблюдений, начиная с 1-го квартала 1995 г. и заканчивая 1-м кварталом 2007 г. (всего 49 точек).

Используем альтернативную процедуру отбора переменных из раздела 2.2 для получения переменных, которые войдут в финальное оценивание. Выбираем  $\alpha = 0,2$  и находим  $q = 2$ , т.е. используем два общих фактора. Для всех 24 переменных степень общности равна 0,71. Выбираем параметры альтернативной процедуры следующим образом:  $\rho^* = 0,05$ ,  $\mu^* = 0,01$ . Четыре переменные были кандидатами на исключение из-за малой корреляции с общей компонентой ВВП:

- Реальный эффективный обменный курс,
- Межбанковская процентная ставка,
- Доля предприятий с увеличившимся или неизменным спросом,
- Доля предприятий, не имеющих избыточных запасов конечной продукции.

Среди этих переменных только реальный эффективный обменный курс не был исключен, во-первых, потому что степень общности увеличивалась менее, чем на  $\mu^*$ , при исключении REER, и во-вторых, потому что REER имеет опережающую динамику по отношению к ВВП. Все три исключенные переменные входят в метод построения циклического индикатора Центра Развития; возможно, поэтому такой индикатор не дает удовлетворительного прогноза ВВП.

В итоге, в окончательный массив данных вошла 21 переменная, степень общности составила 0,78.

После построения массива данных необходимо найти параметры процедуры оценивания, оптимальные для прогнозирования ВВП. Делается следующая последовательность шагов. Строится вневыборочный прогноз на один период вперед, основанный на двустороннем фильтре, и находится его СКОП. Сравниваются СКОП прогнозов с различными значениями параметров  $M$  и  $q$  (здесь ширина фильтра  $M$  меняется от 1 до 10, а число общих факторов  $q$  – от 1 до 7). Наконец, выбирается прогноз с наименьшим СКОП. Эта процедура проводится для тех же четырех случаев трансформации данных, что рассматривались в пункте с авторегрессионным прогнозом.

Лучшим прогнозом с точки зрения СКОП является прогноз для данных с удаленными сезонностью и трендом на полном интервале времени с одним общим фактором и шириной фильтра, равной восьми. Это говорит о том, что один общий фактор дает наилучшую аппроксимацию компоненты ВВП, и в данных присутствует сильная зависимость от прошлого. СКОП у этого прогноза немного меньше, чем у авторегрессионного прогноза, однако использование данного метода не дает существенных улучшений по сравнению с обычной авторегрессией.

Также следует отметить, что европроцедура работает гораздо лучше для данных с удаленным трендом, чем без удаления тренда. Возможно, это объясняется тем, что корреляционная теория очень строго требует стационарности используемых временных рядов.

Таблица 2: Результаты оценивания по методике Forni, Hallin, Lippi & Reichlin (2001)

Переменная	Отношение дисперсий	Корреляция с ВВП	Временной сдвиг в месяцах
GDP	0,979	1,000	0,00
Inv	0,751	0,157	2,23
GovCons	0,733	0,143	2,34
PrivCons	0,959	0,312	-1,50
IndEmploy	0,295	-0,246	-1,60
Wages	0,937	0,413	2,07
CPI	1,010	0,437	-2,35
PPI	0,978	0,396	0,73
M+QM	0,845	0,220	0,26
Madj	0,799	0,230	1,78
CapAccounts	0,606	0,138	9,52
BankReserves	0,677	0,144	1,67
REER	0,732	-0,034	17,4
NEER	0,862	-0,187	7,27
UralsPrice	0,540	-0,059	18,6
RTS	0,560	-0,092	16,5
Agriculture	0,689	0,422	2,82
Construction	0,652	0,605	-0,98
TransTurnover	0,726	0,139	-6,51
Retail	0,906	0,640	0,46
Services	0,722	0,563	-1,10

*Замечания:* Объяснения переменных см. в таблице 1.

Таблица 2 представляет результаты процедуры оценивания с оптимальными параметрами для финального массива данных. Она показывает отношение между дисперсией общей компоненты и дисперсией самой переменной, корреляцию переменной с ВВП и временной сдвиг каждой переменной по отношению к ВВП на частоте  $\theta^* = \pi/16$ , соответствующей бизнес-циклу длиной 8 лет. Опережающие переменные определяются как те, у которых временное опережение больше месяца (0,33 квартала), что соответствует фазовому опережению  $\tau = \pi/48$ ; запаздывающие переменные определяются как отстающие более чем на один месяц; остальные переменные являются текущими. Сдвиги представлены в виде временных сдвигов, числа даны в месяцах, положительные числа означают, что переменная опережает ВВП.

Эта таблица показывает, что переменные рынка труда, цены, денежные агрегаты и номинальный обменный курс имеет высокую степень общности с реальным сектором экономики России. В то же время реальный обменный курс, цена на нефть Urals, индекс RTS и процентные ставки имеют более слабую корреляцию с реальным сектором вопреки обычным представлениям.

Все переменные, кроме занятости, обменных курсов, цены на нефть и индекса RTS, цикличны. Имеется 11 опережающих переменных, среди которых самые очевидные – это инвестиции, госрасходы, обменные курсы, цена на нефть и индекс RTS. Также присутствуют 5 запаздывающих переменных, среди которых частное потребление, занятость и объем грузоперевозок. В целом, большая часть результатов интуитивно достаточно понятна.

Рис. 2 показывает результирующие текущий, опережающий и запаздывающий индикаторы для России по сравнению с логарифмом ВВП. С одной стороны, можно увидеть, что

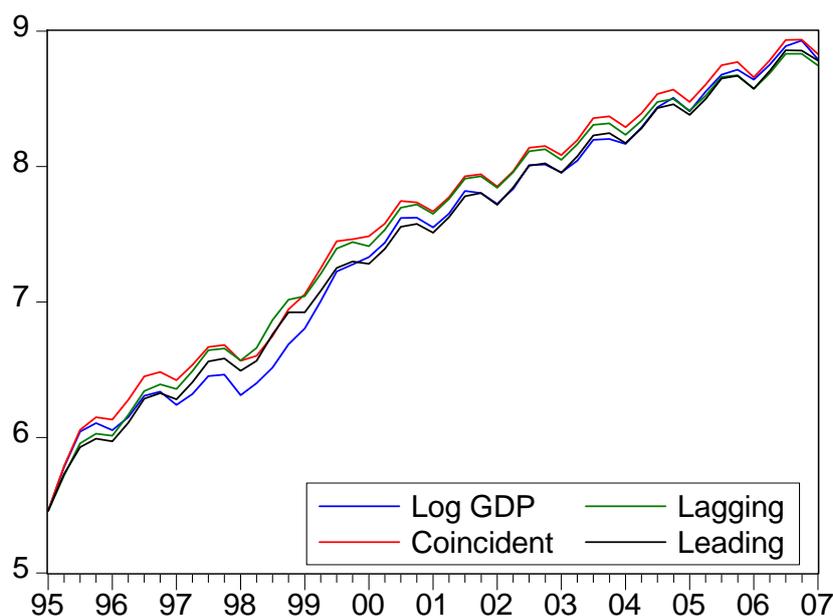


Рис. 2: ВВП в логарифмах и текущий, опережающий и запаздывающий индикаторы европроцедуры.

все индикаторы гораздо лучше коррелируют с ВВП, чем циклический индикатор Центра Развития. Поэтому прогнозные характеристики европроцедуры гораздо лучше. Однако, с другой стороны, из Рис. 2 видно, что текущий индикатор далеко не полностью повторяет логарифм ВВП. Когда мы делаем прогноз с помощью двустороннего фильтра, то прогнозируется текущий индикатор, а не сам логарифм ВВП. Поэтому точность прогнозирования данным методом не такая высокая, как если бы мы строили прогноз самого ВВП.

Таким образом, методология Forni, Hallin, Lippi & Reichlin (2001) дает очень полезную информацию о связи различных переменных с ВВП. Она также дает неплохой метод прогнозирования ВВП, который, однако, имеет определенный недостаток, описанный выше. Достаточно естественная идея, которая здесь возникает, – это соединить два метода: европроцедуру и регрессионный анализ. Это означает, что регрессия ВВП на общие факторы и дальнейшее прогнозирование на основе этой регрессии могут дать существенные улучшения точностных характеристик прогноза ВВП. Данный подход используется Стоком и Уотсоном в построении их диффузионных индексов.

### Методология Стока и Уотсона

Будем рассматривать прогноз диффузионного индекса в виде

$$\Delta \ln y_{t+1} = \mu + \sum_{i=1}^q \beta_i F_{i,t} + \sum_{j=0}^p \gamma_j \Delta \ln y_{t-j} + \epsilon_t,$$

где  $F_{i,t}$  – оцениваемые факторы в момент времени  $t$ ,  $q$  – число факторов,  $p$  – число лагов авторегрессионной компоненты, и  $q$  и  $p$  выбираются из условия минимизации СКОП.

При поиске наилучшего прогноза  $q$  пробегает значения от 1 до 7,  $p$  – от 0 до 8. Здесь также производится оценивание для всех четырех типов трансформации данных. Наилучший прогноз получается с одним общим фактором и семью лагами авторегрессионной компоненты для данных с удаленными сезонностью и трендом на полном временном интервале. Этот прогноз ощутимо превосходит прогноз, основанный на европроцедуре, поскольку в данном случае мы прогнозируем сам логарифм ВВП, а не его общую компоненту.

### Объединенная методология

Вполне возможно, что лучшая точность предсказания процедуры Стока и Уотсона может быть достигнута, если учитывать фазовый сдвиг переменных относительно ВВП. Действительно, запаздывающие переменные могут приносить с собой запаздывающую динамику в общие факторы, которые используются как регрессоры для будущего ВВП. Однако очевидно, что качество регрессоров тем лучше, чем больше в них опережающей динамики. Поэтому возникает идея, что исключение запаздывающих переменных из массива данных может привести к более качественному прогнозу.

Здесь рассмотрены два случая. В первом случае в процедуру оценивания включены только опережающие переменные, во втором – к опережающим переменным добавлены еще и текущие. В обоих случаях были рассмотрены все типы трансформации данных. Типы переменных для полного интервала времени определялись из таблицы 2, для укороченного интервала – из европроцедуры с оптимальными параметрами на укороченном интервале. Стоит отметить, что почти все переменные определялись одинаково на обоих интервалах времени.

Использование текущих переменных мотивировано тем, что они также могут нести часть ведущей динамики. Однако результаты оценивания показали, что лучше использовать только опережающие переменные. Наилучший прогноз получается для данных с удаленной сезонностью на полном интервале с тремя общими факторами и шестью лагами в авторегрессионной части. СКОП этого прогноза почти в 20 раз лучше СКОП, полученного по методологии Центра Развития.

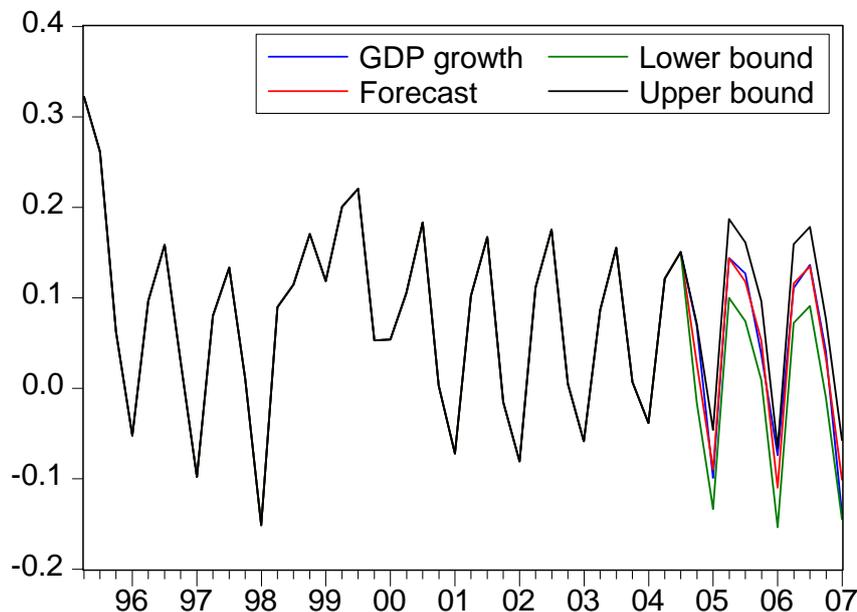


Рис. 3: Разность в логарифмах ВВП, прогноз, нижняя и верхняя границы 95%-ного доверительного интервала объединенной процедуры.

Рис. 3 показывает в сравнении наилучший прогноз и процентное изменение ВВП. Также показаны верхняя и нижняя границы 95%-ного доверительного интервала, и можно видеть, что логарифм ВВП остается внутри этого интервала в течение всего времени.

### Результаты

Результаты для всех процедур прогнозирования представлены в таблице 3. В ней показаны отношения СКОП каждого прогноза к СКОП прогноза по методу Центра Развития. Это

означает, что относительное СКОП прогноза Смирнова равно 1, и если СКОП прогноза меньше 1, то это прогноз превосходит прогноз по методу Центра Развития.

Таблица 3: Результаты расчетов по прогнозированию

СКОП	SA DM	SA DR	SA DM S	SA DR S
AR	0,1665 (10)	0,1047 (7)	0,2170 (4)	0,2101 (6)
Euro	0,1012 (1,8)	0,7068 (1,8)	0,1560 (1,8)	0,7867 (1,4)
Stock & Watson Full	0,0823 (1,7)	0,0834 (2,7)	0,1490 (2,5)	0,1479 (1,4)
Stock & Watson Leading	0,0578 (3,6)	<b>0,0557 (3,6)</b>	0,0894 (5,5)	0,0906 (5,5)
Stock & Watson Lead&Coin	0,0701 (2,7)	0,0704 (4,6)	0,0927 (6,5)	0,0900 (6,5)

*Замечания:* Результаты представлены для четырех типов трансформации данных: с удаленными сезонностью и трендом для полного периода наблюдений (1995Q1:2007Q1), с удаленной сезонностью для полного периода (1995Q1:2007Q1), с удаленными сезонностью и трендом для укороченного периода (1998Q3:2007Q1), с удаленной сезонностью для укороченного периода (1998Q3:2007Q1). Числа без скобок показывают отношение СКОП каждого прогноза к СКОП прогноза по методу Центра Развития. Для авторегрессионного прогноза (AR) число в скобках означает используемое число лагов ВВП. Для европроцедуры (Euro) числа в скобках соответствуют числу общих факторов и ширине двустороннего фильтра. Для прогноза по Стоку и Уотсону числа в скобках показывают выбранные числа общих факторов и лагов ВВП. Full обозначает то, что процедура сделана по всем переменным, Leading – только по опережающим, Lead&Coin – по опережающим и текущим.

Можно видеть, что для прогнозирования лучше использовать полный интервал наблюдений, чем укороченный. Видимо, это связано с тем, что проблема короткой выборки более серьезная, чем проблема, связанная со структурным сдвигом в 1998Q3.

Первый основной результат состоит в том, что установлен порядок среди существующих статистических методов с точки зрения точности прогноза: самый лучший – это Stock & Watson (2002), затем европроцедура и авторегрессия, имеющие почти одинаковые свойства, и, наконец, метод Центра Развития, обладающий наименьшей точностью.

Второй результат состоит в том, что использование метода, который объединяет методологию Стока и Уотсона с европроцедурой, может дать значительное улучшение точности прогноза. Этот объединенный метод дает прогноз со стандартным отклонением около 0,0222, что соответствует средней ошибке прогноза процентного изменения ВВП на следующий квартал порядка двух процентов.

#### 4 Заключение

В данной статье рассмотрены различные типы индексов для прогнозирования экономической активности в России. В частности, обсуждены три методики: циклический индикатор Центра Развития, методология, предложенная в Forni, Hallin, Lippi & Reichlin (2001), и диффузионные индексы Стока и Уотсона (Stock & Watson, 2002).

Результаты, основанные на дальнейшем развитии процедуры Forni, Hallin, Lippi & Reichlin (2001), показали, что текущий индикатор российского бизнес-цикла сильно коррелирован с некоторыми переменными рынка труда, ценами, денежными агрегатами и номинальным обменным курсом, в то время как он слабее коррелирован с реальным обменным курсом, ценой на нефть Urals, индексом RTS и процентными ставками. Текущий индикатор достаточно хорошо прогнозируется взвешенным средним опережающих переменных, таких как инвестиции, госрасходы, обменные курсы, цена на нефть и индекс РТС.

Все методы прогнозирования будущего ВВП были сравнены, используя среднее квадратичное отклонение прогноза. Среди существующих методов наиболее точный прогноз дает метод Стока и Уотсона. Главное достижение данной работы – это построение методологии, которая

объединила в себе диффузионные индексы Стока и Уотсона с процедурой Forni, Hallin, Lippi & Reichlin (2001). Она дала наилучший прогноз с точки зрения СКОП среди всех опробованных методов.

## Благодарности

Автор благодарит Олег Замулина, Кирилла Сосунова, Константина Стырина и Станислава Анатольева за полезные обсуждения и комментарии. Автор также благодарит Сергея Смирнова, Кирилла Сосунова и Олега Замулина за предоставленные данные.

## Список литературы

- Bai, J. & S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica* 70, 191–221.
- Brillinger, D.R. (1981). *Time Series Data Analysis and Theory*. New York: Holt, Rinehart and Winston.
- Forni, M., M. Hallin, M. Lippi & L. Reichlin (2001). Coincident and leading indicators for the Euro area. *Economic Journal* 111, C62–85.
- Forni, M., M. Hallin, M. Lippi & L. Reichlin (2000). The generalized factor model: Identification and estimation. *Review of Economic and Statistics* 82, 540–554.
- Forni, M. & M. Lippi (2001). The generalized dynamic factor model: Representation theory. *Econometric Theory* 17, 1113–1141.
- Geweke, J. (1977). The dynamic factor analysis of economic time series. Глава 19 в *Latent Variables in Socio-Economic Models* под редакцией D.J. Aigner & A.S. Golberger. Amsterdam: North-Holland.
- Sargent, T.J. & C.A. Sims (1977). Business cycle modeling without pretending to have too much a priori economic theory. Глава 19 в *New Methods in Business Research* под редакцией C.A. Sims. Minneapolis: Federal Reserve Bank of Minneapolis.
- Smirnov, S. (2000). A system of leading indicators for Russia. Working paper, Центр Развития (Development Center, <http://dcenter.ru>).
- Smirnov S. (2006). A new system of cyclical indicators for Russia. 28th CIRET Conference, Rome.
- Stock, J.H. & M.H. Watson (1989). New indexes of coincident and leading economic indicators. *NBER Macroeconomic Annual* 1989, 351–394.
- Stock, J.H. & M.H. Watson (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics* 20, 147–162.

## Приложение

Данное приложение содержит некоторые исправления и дополнения к процедуре оценивания Forni, Hallin, Lippi & Reichlin (2001). Здесь показывается, как оцениваются общие компоненты  $\chi_{nt}$  и матрица их спектральной плотности  $\Sigma_n^{\chi^T}(\theta)$ . Используется выражение (9), то есть опускается индекс  $q$  и принимается явная зависимость от  $n$ .

Процедура оценивания состоит из четырех шагов. Во-первых, вычисляется матрица спектральной плотности  $\Sigma(\theta)$  вектора  $x_{nt}$  на нескольких частотах, используя окно Бартлетта размера  $M = M(T)$ . Точнее, вычисляется выборочная ковариационная матрица  $\Gamma_k^T$  векторов  $x_{nt}$  и  $x_{n,t-k}$  для  $k = 0, \dots, M$ :

$$\Gamma_k^T = \frac{1}{T} \sum_{t=k+1}^T x_t x_{t-k}.$$

Затем вычисляются  $2M + 1$  точек дискретного преобразования Фурье усеченной двусторонней последовательности  $\Gamma_{-M}^T, \dots, \Gamma_M^T$ , где  $\Gamma_{-k} = \Gamma_k'$ , то есть вычисляется

$$\Sigma_n^T(\theta_s) = \frac{1}{2\pi} \sum_{k=-M}^M \Gamma_k^T \omega_k e^{-ik\theta_s},$$

где

$$\theta_s = \pi s/S, \quad s = -S + 1, \dots, S, \quad \theta_s \in (-\pi, \pi],$$

и  $\omega_k = 1 - |k|/(M + 1)$  – веса, соответствующие окну Бартлета размера  $M$ . Стоит отметить, что  $\Sigma_n^T(\theta_s)$  – это комплексная эрмитова матрица по построению. Состоятельное оценивание  $\Sigma(\theta_s)$  обеспечивается, когда  $M(T) \rightarrow \infty$  и  $M(T)/T \rightarrow 0$  при  $T \rightarrow \infty$ . Правило  $M = \lfloor \sqrt{T}/4 \rfloor$  действует достаточно хорошо для большинства MA и AR моделей малого порядка в симуляциях (см. Forni, Hallin, Lippi & Reichlin, 2000).

Во-вторых, вычисляются собственные числа  $\lambda_h(\theta_s)$  матрицы  $\Sigma_n^T(\theta_s)$  в убывающем порядке для всех  $s$ , и формируются

$$\lambda_h = \frac{1}{2S} \sum_{s=-S+1}^S \lambda_h(\theta_s).$$

Затем используется критерий (10) с некоторым  $\alpha$  для определения  $q$ , и фиксируется соответствующая степень общности (8). Используя эту степень, решается вопрос о том, какие переменные включать в панель для оценивания.

В-третьих, вычисляются первые  $q$  собственных векторов  $p_{hn}^T(\theta_s)$ ,  $h = 1, \dots, q$ , матрицы  $\Sigma_n^T(\theta_s)$  для  $s = -S + 1, \dots, S$ . В случае, когда  $M = 0$ ,  $p_{hn}^T(\theta_s)$  является просто  $h$ -м собственным вектором ковариационной матрицы  $x_{nt}$ : динамические главные компоненты редуцируются к статическим главным компонентам. С помощью собственных векторов вычисляются

$$\Phi_n^T(\theta_s) = \sum_{h=1}^q p_{hn}^T(\theta_s)' p_{hn}^T(\theta_s),$$

где  $'$  обозначает сопряжение и транспонирование.

Наконец, оценивается  $n \times n$  матрица фильтра  $K_n(L)$  как обратное дискретное преобразование Фурье от

$$(\Phi_n^T(\theta_{-S+1}), \dots, \Phi_n^T(\theta_S)).$$

Точнее, вычисляются

$$K_{kn}^T = \frac{1}{2S} \sum_{s=-S+1}^S \Phi_n^T(\theta_s) e^{ik\theta_s}$$

для  $k = -M, \dots, M$ . Оценка фильтра задается выражением

$$K_n^T(L) = \sum_{k=-M}^M K_{kn}^T L^k.$$

Оценка общих компонент

$$\chi_{nt}^T = K_n^T(L) x_t = \sum_{k=-M}^M K_{kn}^T x_{t-k}.$$

Матрица спектральной плотности общих компонент может быть оценена как

$$\Sigma_n^{\chi^T}(\theta_s) = \Phi_n^T(\theta_s) \Sigma_n^T(\theta_s) \Phi_n^T(\theta_s).$$

## Different indexes for forecasting economic activity in Russia

Oleg Demidov

*Droege & Comp., Moscow, Russia*

This paper considers different ways of computing indexes for forecasting economic activity in Russia. The first is the methodology used by the Russian Development Centre based on the concept of “growth cycles”. The second combines the dynamic principal components and dynamic factor analyses. The third approach is the NBER methodology based of diffusion indexes constructed using a dynamic factor model. This paper is an attempt to reveal strengths and weaknesses of the three methods in application to Russian data and to develop a better methodology for forecasting economic activity in Russia.

*Keywords: index of economic activity, leading and coincident indicators, dynamic principal components, factor model, Russia*

*JEL Classification: C32, E37, O11*

# Quantile

**#5, September 2008**

English page in the world wide web: <http://quantile.ru/eng>

Electronic mail address: [quantile@quantile.ru](mailto:quantile@quantile.ru)

Access to the journal is free and unlimited

## EDITOR

Stanislav Anatolyev  
New Economic School (Moscow, Russia)

## EDITORIAL COUNSEL

Victoria Zinde-Walsh  
McGill University (Montréal, Canada)

Rustam Ibragimov  
Harvard University (Cambridge, USA)

Anna Mikusheva  
Massachusetts Institute of Technology (Cambridge, USA)

Alexey Onatsky  
Columbia University (New York, USA)

Vladimir Pavlov  
Queensland University of Technology (Brisbane, Australia)

Konstantin Tyurin  
Indiana University (Bloomington, USA)

Alexander Tsyplakov  
Novosibirsk State University (Novosibirsk, Russia)

Victor Chernozhukov  
Massachusetts Institute of Technology (Cambridge, USA)

## GUIDE TO AUTHORS

Manuscripts for publication in the “Articles” section should be submitted by electronic mail to the address [submit@quantile.ru](mailto:submit@quantile.ru). Submitted work may be drawn from any applied field within the economics profession. The main requirement is correct usage of adequate econometric methodology. The manuscript should be written in Russian (for Russian-speaking persons) or in English (for all others) in the *Microsoft Word* or (preferably) *LaTeX* formats, and not exceed 30 double-spaced A4 pages. All submissions are subject to quality control by the editorial counsel and independent referees. A promising manuscript may be returned to the author(s) for polishing or rewriting. The editor also invites econometrics experts worldwide to contribute to the methodological sections of the journal.

Articles and methodological material published in “Quantile” do not transfer original copyright, neither in full, nor in part.

# *Quantile*

*international econometric journal  
in Russian language*

**#5  
September 2008**

## **IN THIS ISSUE**

### **Econometric literacy: selected topics in microeconometrics**

Rodríguez, Germán. Survival models	1
McFadden, Daniel. Semiparametric analysis	29

### **Advice to econometrics students**

Tsyplakov, Alexander. A mini-dictionary of English econometric terminology II	41
Anatolyev, Stanislav. Review of English textbooks in time series analysis	49

### **Problems and solutions**

Problems 5.1, 5.2, 5.3	57
------------------------	----

### **Articles: applied econometrics**

Nazrullaeva, Eugenia. Measurement of technological progress in Russia	59
Demidov, Oleg. Различные индексы прогнозирования экономической активности в России	83