ИНСТИТУТ ЭКОНОМИКИ
ПЕРЕХОДНОГО ПЕРИОДА

В.П. Носко

Эконометрика для начинающих

Дополнительные главы

Москва
2005

УДК 330.45:519.862.6
ББК 65в6
Н84

Носко В.П. **Эконометрика для начинающих** (Дополнительные главы). **– М.: ИЭПП, 2005. С. 379.**

Агентство CIP РГБ

В книге рассматриваются методы статистического анализа регрессионных моделей с ограниченной (цензурированной) зависимой переменной, систем одновременных уравнений, панельных данных, а также структурных форм векторных авторегрессий и моделей коррекции ошибок. Предназначена для студентов, освоивших вводный курс эконометрики. Представляет интерес для специалистов в области экономики и финансов.

Nosko V.P. **Econometrics for Beginners (Additional Chapters)**

The monograph deals with methods of statistical analysis of regression models with a limited (censored) depended variable, systems of simultaneous equations, panel data, as well as structural forms of vector autoregressions and error correction models. The monograph is designated for students who have mastered an introductory econometrics, as well as for experts in the area of economics and finance.

JEL Classification: C50, C51, C52, C53

Настоящее издание подготовлено по материалам исследовательского проекта Института экономики переходного периода, выполненного в рамках гранта, предоставленного Агентством международного развития США.

ISBN 5-93255-173-9

Содержание

Предисловие

Настоящая книга является дополнением к ранее изданным публикациям автора “Эконометрика для начинающих: Основные понятия, элементарные методы, границы применимости, интерпретация результатов” (2000), “Эконометрика: Основные понятия и введение в регрессионный анализ временных рядов” (2004). В ней рассматриваются методы статистического анализа моделей с дискретными объясняющими переменными, систем одновременных уравнений, панельных данных, а также структурные и приведенные формы векторных авторегрессий и моделей коррекции ошибок.

В главе 1 обсуждаются особенности статистического анализа моделей, в которых объясняющая переменная имеет лишь конечное количество возможных значений или только частично наблюдаема. При оценивании этих моделей на первый план выступает метод максимального правдоподобия.

Сначала рассматриваются модели бинарного выбора с двумя значениями объясняющей переменной (пробит, логит, гомпит) и модели с несколькими значениями объясняющей переменной (порядковая пробит-модель, мультиномиальная модель), а затем – модели с частично наблюдаемой (цензурированной) объясняющей переменной. При этом цензурирование может определяться как значениями самой объясняемой переменной (модель тобит I), так и значениями некоторой дополнительной функции полезности (модель тобит II).

В главе 2 рассматривается возможность получения подходящих оценок параметров в ситуациях, когда объясняющие переменные, входящие в уравнение регрессии, коррелированы с ошибкой в этом уравнении. Именно такое положение наблюдается в имеющих широкое применение моделях, известных под названием “системы одновременных уравнений”. Это модели, состоящие из нескольких уравнений регрессии и такие, что переменные, являющиеся объясняемыми переменными в одних уравнениях, являются объясняющими переменными в других уравнениях. Здесь основным методом оценивания параметров является метод инструментальных переменных, состоящий в “очистке” объясняющей переменной, коррелированной с ошибкой, от этой коррелированности, и подстановке в правую часть уравнения вместо этой объясняющей переменной ее очищенного варианта. Рассматриваются этот и другие методы оценивания систем одновременных уравнений, связь между различными методами, их недостатки и преимущества.

Глава 3 посвящена методам статистического анализа панельных данных, т.е. данных, содержащих наблюдения за некоторым достаточно большим количеством субъектов в течение некоторого относительно небольшого количества периодов времени. Особенностью многих моделей, используемых для статистического анализа таких данных, является предположение о наличии различий между субъектами исследования, которые постоянны во времени, но которые не удается реально измерить в виде значений некоторой объясняющей переменной. Такие различия специфицируются в этих моделях как фиксированные или случайные эффекты, и в зависимости от пригодности той или иной интерпретации этих эффектов, используются различные методы оценивания параметров модели (обычный или обобщенный метод наименьших квадратов). Метод инструментальных переменных, рассмотренный в главе 2, находит новое применение в динамических моделях панельных данных, в которых в качестве объясняющих переменных в правых частях уравнения могут выступать и запаздывающие значения объясняемой переменной, и реализуется в виде обобщенного метода моментов, ставшего весьма популярным в последние годы. В заключительной части этой главы модели, рассматривавшиеся в главе 1 (пробит, логит, тобит), распространяются на случай панельных данных.

Наконец, глава 4 дополняет материал, содержащийся в главах 11 и 12 ранее изданной книги автора “Эконометрика: Основные понятия и введение в регрессионный анализ временных рядов”, касающийся моделей векторной авторегрессии и моделей коррекции ошибок, сопутствующих системе коинтегрированных временных рядов. Это дополнение связано с рассмотрением возможности построения и оценивания структурной формы модели коррекции ошибок.

Как и в ранее изданных книгах автора по эконометрике, акценты в изложении смещены в сторону разъяснения процедур статистического анализа данных с привлечением большого количества иллюстративных примеров. Предполагается, что читатель владеет методами регрессионного анализа в рамках начального курса эконометрики (в объеме пособия [Носко (2000)] или первой части пособия [Носко (2004)]). Для удобства читателя, при первом упоминании в тексте тех или иных терминов эти термины выделяются жирным курсивом, а в скобках приводятся их англоязычные эквиваленты. Некоторые моменты изложения, требующие привлечения внимания читателя, выделяются подчеркиванием отдельных слов или целых предложений.

Пособие написано на основании курсов лекций, прочитанных автором в Институте экономики переходного периода. Автор считает своим приятным долгом выразить признательность доктору экономических наук Синельникову-Мурылеву С.Г., который инициировал работу по написанию этого учебного пособия.

Глава 1. Модели с дискретными объясняемыми переменными. Метод максимального правдоподобия

1.1 Модели, в которых объясняемая переменная принимает только два различных значения

Ситуации такого рода возникают при исследовании влияния тех или иных субъективных и объективных факторов на наличие или отсутствие некоторого признака у отдельных домашних хозяйств (наличие или отсутствие в семье автомобиля), у отдельных индивидуумов (занятый – безработный), у отдельных фирм (обанкротилась или нет в течение определенного периода) и т.п. Если исследование затрагивает n субъектов, т.е. если мы имеем n наблюдений, то факт наличия или отсутствия такого признака в i-м наблюдении удобно индексировать числами 1 (наличие признака) и 0 (отсутствие признака). Тем самым мы определяем ***индикаторную (дихотомическую, бинарную) переменную*** y, которая принимает в i-м наблюдении значение y_i. При этом $y_i = 1$ при наличии рассматриваемого признака у i-го субъекта и $y_i = 0$ – при отсутствии рассматриваемого признака у i-го субъекта.

Если пытаться объяснить наличие или отсутствие рассматриваемого признака значениями (точнее, сочетанием значений) некоторых факторов (***объясняющих переменных***) , то, следуя идеологии классической линейной модели, мы могли бы расмотреть модель наблюдений

$$y_i = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + \varepsilon_i\,, \quad i = 1, \ldots, n\,,$$

в которой $x_{i1}, \ldots, x_{ip}$ – значения p объясняющих переменных в i-м наблюдении, $\theta_1, \ldots, \theta_p$ – неизвестные параметры, а $\varepsilon_1, \ldots, \varepsilon_n$ – ***случайные ошибки***, отражающие влияние на наличие или отсутствие рассматриваемого признака у i-го субъекта каких-то неучтенных дополнительных факторов. Однако попытка оценить такую модель обычным методом наименьших квадратов (*OLS* – ordinary least squares) наталкивается на определенные трудности.

При обычном предположении $E(\varepsilon_i \mid x_i) = 0$, $i = 1, \ldots, n$, мы получаем

$$E(y_i \mid x_i) = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} = x_i^T \theta,$$

где $\theta = (\theta_1, \ldots, \theta_p)^T$ – вектор-столбец (неизвестных) коэффициентов, (верхний индекс T указывает на транспонирование вектора или матрицы), а $x_i^T = (x_{i1}, \ldots, x_{ip})$ – вектор-строка (известных) значений объясняющих переменных в i-м наблюдении.

Вместе с тем, поскольку y_i – случайная величина, принимающая только два значения 0 и 1, то ее условное математическое ожидание (при заданном значении x_i) равно

$$E(y_i \mid x_i) = 1 \cdot P\{y_i = 1 \mid x_i\} + 0 \cdot P\{y_i = 0 \mid x_i\} = P\{y_i = 1 \mid x_i\}.$$

Таким образом,

$$\theta_1 x_{i1} + \cdots + \theta_p x_{ip} = P\{y_i = 1 \mid x_i\},$$

т.е. $\theta_1 x_{i1} + \cdots + \theta_p x_{ip}$ – <u>вероятность</u>, а значит должно выполняться соотношение

$$0 \le \theta_1 x_{i1} + \cdots + \theta_p x_{ip} \le 1.$$

Это первая из трудностей, с которыми мы сталкиваемся при обращении к таким моделям.

Далее, при $y_i = 1$ получаем $\varepsilon_i = 1 - x_i^T \theta$, а при $y_i = 0$ имеем $\varepsilon_i = -x_i^T \theta$, так что (при фиксированном x_i) ε_i может принимать в i-м наблюдении <u>только два</u> значения, и (условные) вероятности этих значений равны

$$P\{\varepsilon_i = 1 - x_i^T \theta \mid x_i\} = P\{y_i = 1 \mid x_i\} = x_i^T \theta,$$

$$P\{\varepsilon_i = -x_i^T \theta \mid x_i\} = P\{y_i = 0 \mid x_i\} = 1 - x_i^T \theta.$$

Соответственно, случайная величина ε_i имеет условное математическое ожидание

$$E(\varepsilon_i | x_i) = \left(1 - x_i^T\theta\right) \cdot P\left\{\varepsilon_i = 1 - x_i^T\theta \,|\, x_i\right\} + \left(-x_i^T\theta\right) \cdot P\left\{\varepsilon_i = -x_i^T\theta \,|\, x_i\right\} =$$

$$= \left(1 - x_i^T\theta\right) \cdot x_i^T\theta - x_i^T\theta \cdot \left(1 - x_i^T\theta\right) = 0$$

и условную дисперсию

$$D(\varepsilon_i | x_i) = E\left(\varepsilon_i^2 | x_i\right) - \left(E(\varepsilon_i | x_i)\right)^2 = E\left(\varepsilon_i^2 | x_i\right) =$$

$$= \left(1 - x_i^T\theta\right)^2 \cdot x_i^T\theta + \left(-x_i^T\theta\right)^2 \cdot \left(1 - x_i^T\theta\right) =$$

$$= x_i^T\theta \cdot \left(1 - x_i^T\theta\right) \cdot \left[x_i^T\theta + \left(1 - x_i^T\theta\right)\right] = x_i^T\theta \cdot \left(1 - x_i^T\theta\right).$$

Таким образом, здесь возникает также проблема ***гетероскедастичности***, осложненная еще и тем, что в выражения для дисперсий ε_i входит и (неизвестный) вектор параметров θ.

Предположим, что y_i индексирует наличие или отсутствие собственного автомобиля у i-й семьи, а x_i – средний ежемесячный доход, приходящийся на каждого члена этой семьи (в условных единицах). Естественно предполагать, что вероятность наличия автомобиля возрастает с ростом x_i. Если использовать линейную модель

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, \ldots, n,$$

то

$$E(y_i | x_i) = P\left\{y_i = 1 | x_i\right\} = \alpha + \beta x_i,$$

так что если значение x_i увеличить на единицу, то вероятность наличия автомобиля увеличится на величину, равную

$$\left(\alpha + \beta\left(x_i + 1\right)\right) - \left(\alpha + \beta x_i\right) = \beta,$$

<u>независимо от того</u>, сколь большим или малым является среднедушевой доход x_i.

Между тем такое положение вряд ли можно считать оправданным. Скорее можно предположить, что для семей с малыми доходами наличие автомобиля – большая редкость, и некоторое увеличение среднедушевого дохода лишь ненамного увеличит вероятность приобретения автомобиля такой семьей. Для семей с весьма высокими доходами возрастание вероятности наличия автомобиля также не может быть существенным, поскольку такие семьи, как правило, уже обладают автомобилем. Большее влияние увеличения дохода на возрастание вероятности наличия автомобиля должно наблюдаться для семей со “средними” доходами, т.е. в “переходной зоне” от доходов, еще не позволяющих обзавестись собственным автомобилем, к доходам, уже обеспечившим возможность приобретения собственного автомобиля.

Возьмем прямоугольную систему координат, в которой по оси абсцисс будем откладывать размеры среднедушевых семейных доходов. Пусть

$$x_{(1)} = \min\{x_1, \ldots, x_n\}, \quad x_{(n)} = \max\{x_1, \ldots, x_n\},$$

так что $x_{(1)} \le x \le x_{(n)}$ – интервал значений среднедушевых доходов рассматриваемых семей. Разобъем этот интервал на некоторое количество m подинтервалов одинаковой длины $l = \left(x_{(n)} - x_{(1)}\right)/m$. Построим над каждым таким подинтервалом прямоугольник, нижнее основание которого совпадает с этим подинтервалом. Пусть в пределы j-го подинтервала ($j = 1, \ldots, m$) попадают среднедушевые доходы n_j семей, и при этом лишь у $n_{j,1}$ из этих семей имеется автомобиль. (Для определенности, значения x_i, лежащие на границе двух соседних подинтервалов, будем относить к подинтервалу, расположенному левее.) Тогда высоту прямоугольника, построенного над j-м подинтервалом, положим равной

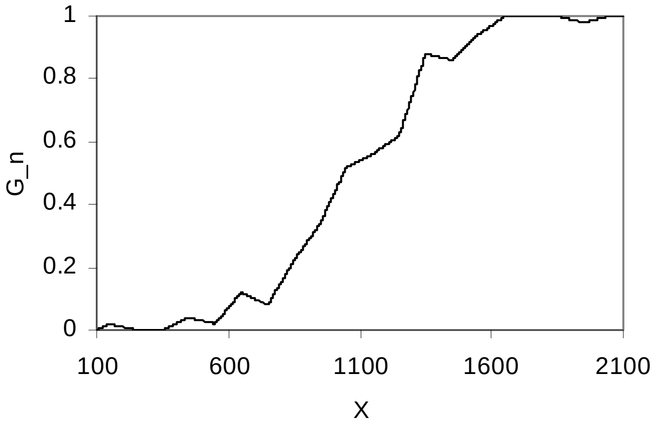
$$h_j = n_{j,1} / n_j .$$

При этом мы предполагаем, что общее количество рассматриваемых семей n достаточно велико, так что можно взять не слишком малое количество подинтервалов m, и при этом все еще иметь достаточное количество значений x_i в каждом подинтервале.

Построим теперь ломаную с концами в точках $(x_{(1)},0)$ и $(x_{(n)},1)$, узлы которой совпадают с серединами верхних сторон построенных прямоугольников. Эта ломаная является графиком некоторой кусочно-линейной функции $G_n(x)$. И если $P\{y_i = 1 | x_i = x\} = G(x)$, то функция $G_n(x)$ в какой-то мере “оценивает” функцию $G(x)$. Правда, если функцию $G(x)$ естественно считать неубывающей (возрастающей) по x, то в силу случайных причин функция $G_n(x)$ вполне может иметь и участки убывания. Тем не менее при большом количестве наблюдений и достаточном количестве подинтервалов график функции $G_n(x)$ отражает в общих чертах форму “истинной” функции $G(x)$, так что по поведению функции $G_n(x)$ можно судить о совместимости или о несовместимости линейной модели с данными наблюдений.

Рассмотрим (искусственно смоделированную) выборку, состоящую из 1000 семей со среднедушевыми месячными доходами от 100 до 2100 условных единиц, среди которых 510 семей имеют собственный автомобиль.

Построенная по этим данным ломаная (график функции $G_n(x)$) имеет следующий вид:



и указывает на то, что “истинная” функция $G(x)$ имеет скорее не линейную, а S-образную форму.

Если, тем не менее, исходить из линейной модели наблюдений, то метод наименьших квадратов дает для параметров такой модели следующие оценки: $\hat{\alpha} = -0.237628$, $\hat{\beta} = 0.000680$, так что условная вероятность $P\{y_i = 1 | x_i\}$ оценивается как

$$\hat{P}\{y_i = 1 | x_i\} = -0.237628 + 0.000680\ x_i\ .$$

При $x_i \le 349$ правая часть принимает отрицательные значения, а при $x_i \ge 1821$ – значения, превышающие единицу, что выходит за пределы интервала возможных значений вероятности.

Заметим теперь, что в число функций, имеющих S-образную форму и значения в пределах от 0 до 1, входит целый ряд функций распределения, используемых в теории вероятностей и математической статистике, например, нормальные функции распределения.

Если использовать функцию нормального распределения $N(\mu,\sigma^2)$, имеющего математическое ожидание μ и дисперсию σ^2, то тогда

$$G(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left\{-\frac{(z-\mu)^2}{2\sigma^2}\right\}dz\,.$$

Замена переменной $(z-\mu)/\sigma=t$ приводит это соотношение к виду

$$G(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(x-\mu)/\sigma}e^{-t^2/2}dt=\Phi\left(\frac{x-\mu}{\sigma}\right),$$

где $\Phi(z)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-t^2/2}dt$ – функция стандартного нормального распределения $N(0,1)$, математическое ожидание которого равно нулю, а дисперсия равна единице.

Соотношение $G(x)=\Phi\left(\frac{x-\mu}{\sigma}\right)$ можно записать также в виде

$$G(x)=\Phi(\alpha+\beta x),$$

где

$$\alpha=-\mu/\sigma,\ \beta=1/\sigma\,.$$

Таким образом, используя для аппроксимации $G(x)$ функцию нормального распределения, мы приходим к модели

$$y_i=\Phi(\alpha+\beta x_i)+\varepsilon_i\,,\quad i=1,\ldots,n\,.$$

Оценив параметры α и β этой модели, мы тем самым получим и оценки параметров функции нормального распределения, аппроксимирующего функцию $G(x)$:

$$\hat{\mu}=-\hat{\alpha}/\hat{\beta},\quad \hat{\sigma}=1/\hat{\beta}\ .$$

Проблема, однако, в том, каким образом производить оценивание.

Заметим, что функция $G(x)=\Phi(\alpha+\beta\, x)$ нелинейна по параметрам, так что мы имеем здесь дело с ***нелинейной моделью регрессии***. Следуя принципу наименьших квадратов, для получения оценок $\hat{\alpha}$ и $\hat{\beta}$ надо минимизировать по α и β сумму квадратов

$$Q(\alpha,\beta)=\sum_{i=1}^{n}(y_i-\Phi(\alpha+\beta\, x_i))^2 .$$

Однако в отличие от линейной модели, получающиеся здесь нормальные уравнения нелинейны, не имеют решения в явном виде, и для получения приближенных значений оценок $\hat{\alpha}$ и $\hat{\beta}$ приходится использовать ***итерационные процедуры***. Как и в рассмотренном ранее случае линейной модели, здесь возникает и проблема ***гетероскедастичности***: условные дисперсии ошибок равны

$$D(\varepsilon_i\,|\,x_i)=\Phi(\alpha+\beta\, x_i)\cdot(1-\Phi(\alpha+\beta\, x_i)).$$

Соответственно, для учета различия этих дисперсий при разных i следует использовать ***взвешенный метод наименьших квадратов***, т.е. минимизировать по α и β сумму квадратов

$$Q(\alpha,\beta)=\sum_{i=1}^{n}w_i\cdot(y_i-\Phi(\alpha+\beta\, x_i))^2 ,$$

где ***веса*** w_i определяются соотношением

$$w_i=1/D(\varepsilon_i\,|\,x_i)=[\Phi(\alpha+\beta\, x_i)\cdot(1-\Phi(\alpha+\beta\, x_i))]^{-1} .$$

К сожалению, эти веса зависят не только от x_i, но и от значений параметров α и β, которые нам не известны и которые как раз и подлежат оцениванию. Поэтому для реализации итерационной процедуры оценивания необходимы некоторые <u>начальные оценки весов</u> $\hat{w}_i^0$, $i=1,\ldots,n$, а для этого необходимы начальные оценки $\hat{G}_i^0$ значений $G_i=G(x_i)=\Phi(\alpha+\beta\, x_i)$, которые дали бы оценки весов в виде

$$\hat{w}_i^0=\left[\hat{G}_i^0\left(1-\hat{G}_i^0\right)\right]^{-1}.$$

Поскольку же у нас $y_i = 0$ или $y_i = 1$, то единственная разумная возможность – положить $\hat{G}_i^0 = 1$, если $y_i = 1$, и $\hat{G}_i^0 = 0$, если $y_i = 0$. Однако в обоих случаях вес $\hat{w}_i^0$ <u>не определен</u> (знаменатель равен нулю).

Ввиду отмеченных выше трудностей в применении метода наименьших квадратов к рассмотренным моделям, мы используем альтернативный метод оценивания, широко применяемый в прикладных исследованиях, а именно – ***метод максимального правдоподобия.***

Однако прежде чем переходить к изложению этого метода, мы должны заметить, что в качестве объясняющих факторов в моделях рассмотренного типа могут выступать несколько переменных, и тогда мы получаем модель вида

$$y_i = G(\theta_1 x_{i1} + \cdots + \theta_p x_{ip}) + \varepsilon_i, \quad i = 1, \ldots, n,$$

которую обычно называют ***моделью бинарного выбора***.

1.2. Использование метода максимального правдоподобия для оценивания моделей бинарного выбора

Итак, пусть наша задача состоит в оценивании параметров модели бинарного выбора

$$y_i = G(\theta_1 x_{i1} + \cdots + \theta_p x_{ip}) + \varepsilon_i, \quad i = 1, \ldots, n,$$

где $G(z)$ – S-образная функция распределения, имеющего плотность $g(z) = G'(z)$. В соответствии с введенными выше обозначениями $\theta_1 x_{i1} + \cdots + \theta_p x_{ip} = x_i^T \theta$, так что $G(\theta_1 x_{i1} + \cdots + \theta_p x_{ip}) = G(x_i^T \theta)$. Мы предполагаем, что <u>при фиксированных</u> значениях объясняющих переменных в n наблюдениях, что соответствует фиксированным значениям векторов x_i, случайные ошибки $\varepsilon_1,\ldots,\varepsilon_n$ статистически независимы и $E(\varepsilon_i \mid x_i)=0$, так что $P\{y_i=1 \mid x_i\}=E(y_i \mid x_i)=G\left(x_i^T\theta\right)$. Тогда при фиксированных x_i статистически независимы и случайные величины $G\left(\theta_1 x_{i1}+\cdots+\theta_p x_{ip}\right)+\varepsilon_i$, $i=1,\ldots,n$, т.е. статистически независимы $y_1,\ldots,y_n$. В силу этого (условная при фиксированных x_i, $i=1,\ldots,n$) совместная вероятность получения конкретного набора наблюдений $y_1,\ldots,y_n$ (конкретного набора нулей и единиц) равна произведению

$$\prod_{i=1}^{n}\left(P\{y_i=1 \mid x_i\}\right)^{y_i}\left(P\{y_i=0 \mid x_i\}\right)^{1-y_i}=\prod_{i=1}^{n}\left(G\left(x_i^T\theta\right)\right)^{y_i}\left(1-G\left(x_i^T\theta\right)\right)^{1-y_i}.$$

Правая часть этого выражения является при фиксированных x_i, $i=1,\ldots,n$, функцией от вектора неизвестных параметров θ,

$$L(\theta)=L\left(\theta \mid x_1,\ldots,x_n\right)=\prod_{i=1}^{n}\left(G\left(x_i^T\theta\right)\right)^{y_i}\left(1-G\left(x_i^T\theta\right)\right)^{1-y_i}$$

и интерпретируется как ***функция правдоподобия*** параметров $\theta_1,\ldots,\theta_p$. Дело в том, что при различных наборах значений $\theta_1,\ldots,\theta_p$ получаются различные $L(\theta)$, т.е. при фиксированных x_i, $i=1,\ldots,n$, вероятность наблюдать конкретный набор значений $y_1,\ldots,y_n$ может быть более высокой или более низкой, в зависимости от значения θ. Метод максимального правдоподобия предлагает в качестве оценки вектора параметров θ использовать значение $\theta=\hat{\theta}$, максимизирующее функцию правдоподобия, так что

$$L\left(\hat{\theta}\right)=\max_{\theta} L\left(\hat{\theta}\right)=\max_{\theta}\prod_{i=1}^{n}\left(G\left(x_i^T\theta\right)\right)^{y_i}\left(1-G\left(x_i^T\theta\right)\right)^{1-y_i}.$$

Использование свойства монотонного возрастания функции $\ln(z)$, позволяет найти <u>то же самое</u> значение $\hat{\theta}$, максимизируя ***логарифмическую функцию правдоподобия*** $\ln L(\theta)$. В нашем случае

$$\ln L(\theta) = \sum_{i=1}^{n} y_i \ln G\left(x_i^T \theta\right) + \sum_{i=1}^{n} (1 - y_i) \ln\left(1 - G\left(x_i^T \theta\right)\right).$$

Мы не будем углубляться в технические детали соответствующих процедур максимизации, имея в виду, что такие процедуры “встроены” во многие прикладные пакеты статистических программ для персональных компьютеров и читатель при необходимости может ими воспользоваться. Заметим только, что если не имеет место чистая мультиколлинеарность объясняющих переменных (т.е. если матрица $X = \left(x_{ij}\right)$ значений p объясняющих переменных в n наблюдениях имеет ранг p, так что ее столбцы линейно независимы), то тогда функция $L(\theta)$ имеет единственный локальный максимум, являющийся и глобальным максимумом, что гарантирует сходимость соответствующих итерационных процедур к оценке максимального правдоподобия.

Мы рассмотрим теперь результаты применения метода максимального правдоподобия для оценивания параметров α и β моделей

$$y_i = G(\alpha + \beta x_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

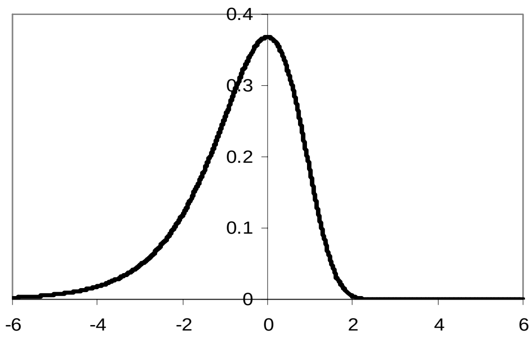
по упомянутым выше смоделированным данным. При этом мы используем предусмотренную в пакете *Econometric Views* (EVIEWS) возможность выбора в качестве $G(z)$ функций

$\Phi(z) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{z} e^{-t^2/2} dt$ – функция ***стандартного нормального распределения*** $N(0,1)$ (***пробит-модель***),

$\Lambda(z) = \dfrac{e^z}{1 + e^z}$ – функция ***стандартного логистического распределения*** (***логит-модель***),

$G(z) = 1 - \exp\left(-e^{z}\right)$ – функция ***стандартного распределения экстремальных значений (минимума) I-го типа*** (***распределение Гомпертца***, ***гомпит-модель***).

Заметим, что функции плотности первых двух распределений являются четными функциями (графики этих плотностей симметричны относительно оси ординат), тогда как функция плотности последнего из трех распределений не обладает таким свойством. Ее график асимметричен и скошен в сторону отрицательных значений аргумента.



Результаты оценивания указанных трех моделей по смоделированным данным (1000 наблюдений) с использованием пакета EVIEWS таковы:[1]

1 В четвертом столбце приведены значения отношений оценок коэффициентов к стандартным ошибкам, рассчитанным по асимптотическому нормальному распределению оценок максимального правдоподобия. В связи с этим, здесь и в последующих таблицах указанное отношение называется не t-статистикой, а z-статистикой. P-значения, приводимые в пятом столбце, соответствуют стандартному нормальному распределению.

Пробит-модель:

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-3.503812	0.200637	-17.46343	0.0000
X	0.003254	0.000178	18.25529	0.0000

Логит-модель:

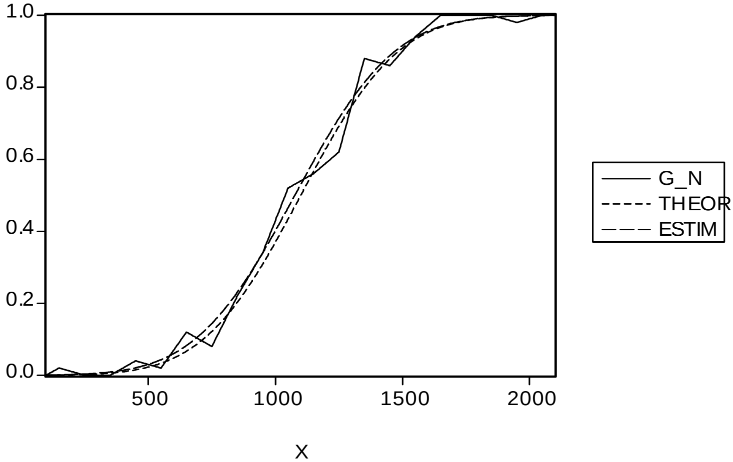
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-6.357013	0.411837	-15.43576	0.0000
X	0.005892	0.000368	16.01461	0.0000

Гомпит-модель:

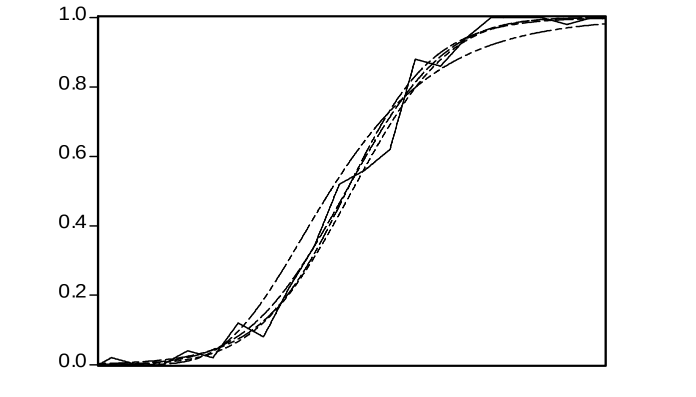
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-3.022612	0.162178	-18.63764	0.0000
X	0.003344	0.000168	19.93322	0.0000

Полученные значения оценок параметров α и β в первой модели ($\hat{\alpha} = -3.503812$, $\hat{\beta} = 0.003254$) соответствуют оценкам $\hat{\mu} = 1076.77$ и $\hat{\sigma} = 307.31$ параметров функции нормального распределения, "сглаживающей" построенную ранее функцию $G_n(x)$, график которой представляет ломаную. Заметим, что в действительности при моделировании данных мы использовали в качестве $G(x)$ функцию нормального распределения с параметрами $\mu = 1100$ и $\sigma = 300$. Следующий график позволяет сравнить поведение

- кусочно-линейной функции $G_n(x)$,
- теоретической функции $G(x)$, соответствующей нормальному распределению $N(1100, 300^2)$,
- оцененной функции $\hat{G}(x)$, соответствующей нормальному распределению $N(1076.77, 307.31^2)$.



На следующем графике добавлены для сравнения также и оцененные функции $\hat{G}(x)$ для логит- и гомпит-моделей



Кривые, получаемые по пробит- и логит-моделям, отличаются очень мало как друг от друга, так и от теоретической кривой. В то же время кривая, полученная по гомпит-модели, представляется менее удовлетворительной. Разумеется, хотелось бы иметь некоторые количественные критерии для сравнения разных моделей и для проверки адекватности каждой из рассматриваемых моделей данным наблюдений. Мы займемся теперь этой проблемой.

1.3. Показатели качества моделей бинарного выбора, критерии согласия с имеющимися данными, сравнение альтернативных моделей

Прежде всего обратим внимание на следующее обстоятельство. Пусть методом наименьших квадратов оценивается обычная линейная модель

$$y_i = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + \varepsilon_i\,, \qquad i = 1, \ldots, n\,,$$

с $x_{i1} \equiv 1$ (модель с константой), в которой объясняемая переменная y может принимать непрерывный ряд значений. В таком случае простейшим показателем качества оцененной модели является ***коэффициент детерминации*** R^2,

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}\,,$$

где $\hat{y}_i = \hat{\theta}_1 x_{i1} + \cdots + \hat{\theta}_p x_{ip}$, $\bar{y} = \left(y_1 + \cdots + y_n\right)/n$. (Здесь TSS – “полная”, а RSS – “остаточная” сумма квадратов.) Если оценивать “тривиальную” модель, в правую часть которой включается единственная объясняющая переменная $x_{i1} \equiv 1$, т.е. модель

$$y_i = \theta_1 + \varepsilon_i\,, \qquad i = 1, \ldots, n\,,$$

то для такой модели $\hat{\theta}_1 = \bar{y}$, $\hat{y}_i = \hat{\theta}_1 = \bar{y}$, так что $RSS = TSS$ и $R^2 = 0$. При добавлении в правую часть модели дополнительных объясняющих переменных коэффициент R^2 возрастает, и этот коэффициент будет тем больше, чем более выраженной является линейная связь объясняемой переменной с совокупностью объясняющих переменных, включенных в правую часть. Своего максимального значения $R^2 = 1$ коэффициент детерминации достигает в предельном случае, когда для всех $i = 1,\ldots,n$ выполняются точные соотношения

$$y_i = \theta_1 x_{i1} + \cdots + \theta_p x_{ip}\,.$$

Поскольку теперь мы имеем дело с нелинейными моделями

$$y_i = G\left(\theta_1 x_{i1} + \cdots + \theta_p x_{ip}\right) + \varepsilon_i\,, \quad i = 1,\ldots,n\,,$$

то не можем пользоваться обычным коэффициентом детерминации R^2, и желательно определить какую-то другую меру качества подобранной модели.

Одна из возможностей в этом отношении – сравнение количеств неправильных предсказаний, получаемых по выбранной модели и по модели, в которой в качестве единственной объясняющей переменной выступает константа (“тривиальная модель”).

Естественным представляется предсказывать значение $y_i = 1$, когда $G\left(x_i^T \hat{\theta}\right) > 1/2$. Для симметричных распределений последнее равносильно условию $x_i^T \hat{\theta} > 0$, так что прогнозные значения равны

$$\hat{y}_i = \begin{cases} 1, & \text{если} \quad x_i^T \hat{\theta} > 0, \\ 0, & \text{если} \quad x_i^T \hat{\theta} \le 0. \end{cases}$$

Количество неправильных предсказаний по выбранной модели равно

$$n_{wrong,1} = \sum_{i=1}^{n} \left| y_i - \hat{y}_i \right| = \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \quad ;$$

доля неправильных предсказаний по выбранной модели равна

$$v_{wrong,1} = \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2\,.$$

В то же время, если рассмотреть тривиальную модель, то для нее значение $y_i = 1$ предсказывается <u>для всех</u> $i = 1, \ldots, n$, когда $\hat{\theta}_1 > 1/2$, т.е. когда $\bar{y} > 1/2$ (значения $y_i = 1$ наблюдаются <u>более, чем в половине</u> наблюдений). Соответственно, значение $y_i = 0$ предсказывается <u>для всех</u> $i = 1, \ldots, n$, когда $\hat{\theta}_1 \le 1/2$, т.е. когда $\bar{y} \le 1/2$ (значения $y_i = 1$ наблюдаются <u>не более, чем в половине</u> наблюдений). При этом доля неправильных предсказаний по тривиальной модели равна

$$\nu_{wrong,0} = \begin{cases} 1 - \bar{y}, & \text{если} \quad \bar{y} > 1/2, \\ \bar{y}, & \text{если} \quad \bar{y} \le 1/2. \end{cases}$$

В качестве показателя качества модели можно было бы взять коэффициент

$$R^2_{predict} = 1 - \frac{\nu_{wrong,1}}{\nu_{wrong,0}} = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\nu_{wrong,0}}.$$

Проблема, однако, в том, что выбранная модель может дать предсказание <u>хуже, чем тривиальная</u>, так что $\nu_{wrong,1} > \nu_{wrong,0}$, и тогда $R^2_{predict} < 0$. Отметим также, что вообще

$$\nu_{wrong,0} \le 0.5,$$

так что тривиальная модель может неправильно предсказать <u>не более половины наблюдений</u>. А если оказывается, что в выборке значения y_i равны 1 для 90% наблюдений, то тогда $\nu_{wrong,0} = 0.1$, и чтобы получить $R^2_{predict} > 0$, необходимо, чтобы альтернативная модель давала <u>более</u> 90% правильных предсказаний. Это означает, что большая доля <u>правильных</u> предсказаний $1 - \nu_{wrong,1}$ сама по себе не говорит еще о качестве модели. Эта доля может быть большой и для плохой модели.

Рассмотрим теперь альтернативный подход к построению аналога коэффициента R^2 для моделей бинарного выбора. Поскольку мы использовали для оценивания таких моделей метод максимального правдоподобия, то естественным представляется сравнение максимумов функций правдоподобия (или максимумов логарифмических функций правдоподобия) для выбранной и тривиальной моделей.

Пусть L_1 – максимум функции правдоподобия для выбранной модели, а L_0 – максимум функции правдоподобия для тривиальной модели. Заметим, что при этом $L_0 \le L_1 \le 1$, так что и $\ln L_0 \le \ln L_1 \le 0$. В рамках этого подхода среди множества других были предложены следующие показатели качества моделей бинарного выбора

$$pseudoR^2 = 1 - \frac{1}{1 + 2(\ln L_1 - \ln L_0)/n}$$ [*Aldrich, Nelson* (*1984*)],

$$McFaddenR^2 = 1 - \frac{\ln L_1}{\ln L_0}.$$

Последний показатель часто обозначают как ***LRI*** – ***индекс отношения правдоподобий*** (*likelihood ratio index*).

Оба этих показателя изменяются в пределах от 0 до 1. Если для выбранной модели $\hat{\theta}_2 = \dots = \hat{\theta}_p = 0$, то $L_0 = L_1$ и оба показателя равны нулю. Второй показатель может оказаться равным единице, если $\ln L_1 = 0$, т.е. $L_1 = 1$. Такая модель дает <u>точное</u> предсказание, так что $\hat{y}_i = y_i$ <u>для всех</u> $i = 1, \dots, n$. Но при этом для рассмотренных выше моделей (пробит, логит и гомпит) оказывается невозможным доведение до конца итерационной процедуры оценивания вектора параметров θ из-за взрывного возрастания абсолютной величины $x_i^T \theta$ в процессе итераций. Это связано с тем, что у таких моделей при конечных значениях $x_i^T \theta$ выполняются <u>строгие</u> неравенства

$0 < G\left(x_i^T \theta\right) < 1$, и поэтому функция правдоподобия не может достигать значения 1.

П р и м е р

Продолжая начатый выше статистический анализ смоделированного множества данных, вычислим значения альтернативных вариантов коэффициента R^2 для трех оцененных моделей бинарного выбора.

Требуемые для вычисления этих значений величины представлены в следующей таблице.

Модель	$\nu_{wrong,1}$	$\ln L_1$
Пробит	0.125	-275.7686
Логит	0.124	-275.4592
Гомпит	0.121	-292.6808
	$\nu_{wrong,0}$	$\ln L_0$
Тривиальная	0.490	-692.9472

(Напомним, что в смоделированной выборке количество семей, имеющих собственный автомобиль, равно 510, что составляет более половины семей. Поэтому тривиальная модель дает для всех 1000 наблюдений прогноз $y_i = 1$, что приводит к 49% ошибок.)

Соответственно, для различных вариантов коэффициента R^2 получаем:

Пробит-модель

$$R^2_{predict} = 1 - \frac{\nu_{wrong,1}}{\nu_{wrong,0}} = 1 - \frac{0.125}{0.490} = 0.745,$$

$$pseudoR^2 = 1 - \frac{1}{1 + 2(\ln L_1 - \ln L_0)/n} =$$

$$= 1 - \frac{1}{1 + 2(-275.7686 + 692.9472)/1000} = 0.4548,$$

$$McFaddenR^2 = 1 - \frac{\ln L_1}{\ln L_0} = 1 - \frac{-275.7686}{-692.9472} = 0.6020.$$

Логит-модель

$$R^2_{predict} = 1 - \frac{\nu_{wrong,1}}{\nu_{wrong,0}} = 1 - \frac{0.124}{0.490} = 0.7470,$$

$$pseudoR^2 = 1 - \frac{1}{1 + 2(\ln L_1 - \ln L_0)/n} =$$

$$= 1 - \frac{1}{1 + 2(-275.4592 + 692.9472)/1000} = 0.4550,$$

$$McFaddenR^2 = 1 - \frac{\ln L_1}{\ln L_0} = 1 - \frac{-275.4592}{-692.9472} = 0.6025.$$

Гомпит-модель

$$R^2_{predict} = 1 - \frac{\nu_{wrong,1}}{\nu_{wrong,0}} = 1 - \frac{0.121}{0.490} = 0.7531,$$

$$pseudoR^2 = 1 - \frac{1}{1 + 2(\ln L_1 - \ln L_0)/n} =$$

$$= 1 - \frac{1}{1 + 2(-292.6808 + 692.9472)/1000} = 0.4446,$$

$$McFaddenR^2 = 1 - \frac{\ln L_1}{\ln L_0} = 1 - \frac{-275.4592}{-692.9472} = 0.5776\,.$$

Сведем полученные значения в общую таблицу.

Модель	$R^2_{predict}$	$pseudoR^2$	$McFaddenR^2$
Пробит	0.7450	0.4548	0.6020
Логит	0.7470	0.4550	0.6025
Гомпит	0.7531	0.4446	0.5776

Отметим близость всех вариантов коэффициента R^2 для пробит- и логит-моделей. Гомпит-модель дает несколько лучшее предсказание, в то время как логит-модель несколько лучше двух других с точки зрения коэффициентов $pseudoR^2$ и $McFaddenR^2$.

Представим теперь, что в нашем примере вместо смоделированных значений y_i наблюдались бы следующие значения:

$y_i = 0$ для $x_i \le 1100$, $y_i = 1$ для $x_i > 1100$.

Тогда 100% точное предсказание этих значений дала бы модель

$$P\{y_i = 1\} = \begin{cases} 0, & \text{если } x_i \le 1100 \\ 1, & \text{если } x_i > 1100 \end{cases}.$$

Вместе с тем, в рамках пробит-, логит- и гомпит-моделей оценки максимального правдоподобия в такой ситуации не определены, т.к. максимум функции правдоподобия не достигается при конечных значениях параметров.

Если речь идет о сравнении нескольких альтернативных моделей бинарного выбора с разным количеством объясняющих переменных, то, как и в случае обычных линейных моделей, сравнивать качество альтернативных моделей можно, опираясь на значения информационных критериев Акаике (AIC) и Шварца (SC):

$$AIC = -2\ln L_k / n + 2p/n,$$

$$SC = -2\ln L_k / n + p\ln n / n,$$

а также информационного критерия Хеннана–Куинна

$$HQ = -2\ln L_k / n + 2p\ln(\ln n)/n.$$

Здесь L_k – максимальное значение функции правдоподобия для k-й из альтернативных моделей, а p – количество объясняющих переменных в этой модели.

При этом среди нескольких альтернативных моделей выбирается та, которая минимизирует значение статистики критерия. Заметим, что эти три критерия различаются размерами “штрафа”, который приходится платить за включение в модель большего количества объясняющих переменных.

В рассмотренном выше примере во всех трех моделях использовались одни и те же объясняющие переменные (константа и среднедушевой доход семьи), так что по каждому информационному критерию в качестве ”наилучшей” будет выбрана модель, для которой максимум функции правдоподобия наибольший. Приведем полученные при оценивании значения информационных критериев:

Модель	AIC	SC	HQ
Пробит	0.555537	0.565353	0.559268
Логит	0.554918	0.564734	0.558649
Гомпит	0.589362	0.599177	0.593092

По всем трем критериям наилучшей признается логит-модель. Эта модель имеет наибольший среди трех моделей максимум функции правдоподобия. Вместе с тем отметим, что преимущество логит-модели над пробит-моделью весьма мало.

Для проверки адекватности подобранной модели имеющимся данным имеется ряд статистических критериев согласия; одним из них является ***критерий Хосмера–Лемешоу***[2]. Мы не будем давать его детальное описание, а воспользуемся тем, что этот критерий реализован в некоторых пакетах статистического анализа, в том числе и в пакете ECONOMETRIC VIEWS. Отметим только, что этот критерий основан на сравнении предсказываемых моделью и действительно наблюдаемых количеств случаев с $y_i = 1$ в нескольких группах, на которые разбивается множество наблюдений.

Сопоставим результаты применения критерия Хосмера–Лемешоу к подобранным выше моделям бинарного выбора. В следующей таблице приведены *P*-значения, соответствующие статистике Хосмера–Лемешоу (рассчитанные по асимптотическому распределению хи-квадрат с соответствующим числом степеней свободы) при разбиении множества наблюдений на 10 групп.

Модель	Пробит	Логит	Гомпит
P-значение	0.1509	0.5511	0.0000

Если ориентироваться на эти *P*-значения, то гомпит-модель следует признать неудовлетворительной.

В заключение рассмотрим пример подбора модели бинарного выбора с несколькими объясняющими переменными. В этом примере мы имеем дело со следующими финансовыми показателями 66 фирм на конец одного и того же года:

$$X_1 = \frac{\text{оборотный капитал}}{\text{общая сумма имущества}} ,$$

[2] Подробнее об этом критерии см., например, в [Hosmer, Lemeshow (1989)]

$$X_2 = \frac{\text{нераспределенная прибыль}}{\text{общая сумма имущества}} ,$$

$$X_3 = \frac{\text{доходы до вычета процентов и налогов}}{\text{общая сумма имущества}} ,$$

$$X_4 = \frac{\text{рыночная стоимость активов за вычетом задолженности}}{\text{балансовая стоимость общей суммы обязательств}} ,$$

$$X_5 = \frac{\text{объем продаж}}{\text{общая сумма имущества}} .$$

В течение последующих двух лет половина из этих фирм обанкротилась. Фирмы занумерованы числами от 1 до 66 так, что первые 33 фирмы в этом списке обанкротились. Введем в рассмотрение индикаторную переменную y_i, полагая

$$y_i = \begin{cases} 0 & \text{для } i = 1,\ldots,33, \\ 1 & \text{для } i = 34,\ldots 66 \end{cases}$$

т.е. $y_i = 1$, если фирма выжила в течение двух лет.

Попробуем сначала подобрать к указанным данным пробит-модель

$$y_i = \Phi(\alpha + \beta_1 x_{i1} + \cdots + \beta_5 x_{i5}) + \varepsilon_i , \quad i = 1,\ldots,66 .$$

При попытке оценить параметры такой модели мы наталкивается на упоминавшееся ранее затруднение, связанное с расходимостью итерационного процесса. Поэтому приходится отказаться от желания включить в правую часть модели сразу все имеющиеся в распоряжении показатели и перейти к рассмотрению редуцированных моделей.

При оценивании большинства моделей, в которых используется только 4 из 5 финансовых показателей, мы наталкиваемся на ту же самую проблему. Итерационный процесс сходится только для двух таких моделей – включающих в качестве объясняющих переменных (помимо константы) наборы показателей (X_1, X_2, X_4, X_5) и (X_1, X_3, X_4, X_5), соответственно. Однако каждый из оцененных коэффициентов этих моделей имеет *P*-значение, превышающее 0.10, что указывает на необходимость дальнейшей редукции моделей.

Среди моделей, использующих только 3 финансовых показателя, лучшей по $McFaddenR^2\,(LRI)$ является модель с набором объясняющих переменных $(1, X_2, X_4, X_5)$, но и в ней все оцененные коэффициенты имеют *P*-значения, превышающие 0.184.

Вообще, множество моделей, в которых оценки коэффициентов при всех включенных в их правые части финансовых показателях статистически значимы (при 5% пороге), исчерпывается 6 моделями, включающими в качестве объясняющих переменных наборы

$(1, X_1, X_4), (1, X_3, X_4),$
$(1, X_1), (1, X_2), (1, X_3), (1, X_4).$

Приведем результаты, характеризующие сравнительное качество этих моделей. В первом столбце указаны финансовые показатели, включенные в модель.

	LRI	AIC	SC	HQ	Кол-во неправ. предсказ.	Хосмер–Лемешоу (5 групп) P-значения
X_1, X_4	0.645	0.582	0.682	0.621	6	0.4955
X_3, X_4	0.785	0.389	0.488	0.427	3	0.6499
X_1	0.441	0.835	0.902	0.861	12	0.4820
X_2	0.829	0.298	0.364	0.324	3	0.6916
X_3	0.668	0.520	0.587	0.547	7	0.0525
X_4	0.460	0.809	0.875	0.835	10	0.0004

Критерий Хосмера–Лемешоу признает неадекватной последнюю модель и близкой к неадекватной предпоследнюю модель. Среди остальных 4 моделей по всем показателям лучшей оказывается модель, использующая единственный финансовый показатель X_2. Она дает следующую оценку вероятности выживания фирмы:

$$\hat{P}\{y_i = 1 | x_i\} = \Phi(-0.6625 + 0.0987 x_{i2}).$$

Оцененная модель правильно предсказывает банкротство 31 из 33 обанкротившихся и выживание 32 из 33 выживших фирм. Это соответствует 95.45% правильных предсказаний, тогда как тривиальная модель дает в рассматриваемом случае только 50% правильных предсказаний.

Таким образом, согласно полученным результатам, вероятность выживания фирмы определяется в основном отношением размера нераспределенной прибыли к общей стоимости имущества фирмы и возрастает с ростом этого отношения.

1.4. Интерпретация коэффициентов

Поскольку модели логит, пробит и гомпит являются нелинейными моделями, то оцененные коэффициенты в этих моделях имеют интерпретацию, отличающуюся от интерпретации коэффициентов в линейной модели.

Все эти модели имеют вид

$$y_i = G(\theta_1 x_{i1} + \cdots + \theta_p x_{ip}) + \varepsilon_i = G(x_i^T \theta), \quad i = 1, \ldots, n;$$

при этом

$$P\{y_i = 1 | x_i\} = E(y_i | x_i) = G(x_i^T \theta).$$

Пусть k-я объясняющая переменная является непрерывной переменной. Тогда ***предельный эффект*** (marginal effect) этой переменной определяется как производная

$$\frac{\partial P\{y_i = 1 | x_i\}}{\partial x_{ik}} = \frac{\partial G(x_i^T \theta)}{\partial x_{ik}},$$

и, в отличие от линейной модели, этот эффект зависит от значений объясняющих переменных для i-го субъекта $x_i = (x_{i1}, \ldots, x_{ip})^T$. Малое изменение Δx_{ik} k-й объясняющей переменной приводит (при неизменных значениях остальных объясняющих переменных) к изменению вероятности $P\{y_i = 1| x_i\}$ на величину, приближенно равную

$$\Delta P\{y_i = 1| x_i\} \cong \frac{\partial P\{y_i = 1| x_i\}}{\partial x_{ik}} \cdot \Delta x_{ik} = \frac{\partial G(x_i^T \theta)}{\partial x_{ik}} \cdot \Delta x_{ik} .$$

Заметим, что поскольку модель нелинейна, при интерпретации значений предельного эффекта надо иметь в виду отклик интересующей нас вероятности именно на малые приращения объясняющей переменной.

В случае, когда сама объясняющая переменная принимает только два значения 0 и 1 (***дамми-переменная*** – dummy variable), указывающие на наличие (1) или отсутствие (0) у субъекта определенного признака, "малые" изменения переменной, о которых говорилось выше, попросту невозможны. В этом случае "предельный эффект" определяют просто как разность

$$P\{y_i = 1| x_i^*, d = 1\} - P\{y_i = 1| x_i^*, d = 0\},$$

где d обозначает рассматриваемую дамми-переменную, а x_i^* – вектор значений остальных объясняющих переменных.

В пробит-модели $P\{y_i = 1| x_i\} = \Phi(x_i^T \theta) = \Phi(\theta_1 x_{i1} + \ldots + \theta_p x_{ip})$.

Малое изменение Δx_{ik} k-й объясняющей переменной приводит здесь (при неизменных значениях остальных объясняющих переменных) к изменению вероятности $P\{y_i = 1| x_i\}$ на величину, приближенно равную

$$\Delta P\{y_i = 1| x_i\} \cong \frac{\partial \Phi(\theta_1 x_{i1} + \cdots + \theta_p x_{ip})}{\partial x_{ik}} \cdot \Delta x_{ik} = \varphi(x_i^T)\, \theta_k \cdot \Delta x_{ik} ,$$

где $\varphi(z) = \dfrac{1}{\sqrt{2\pi}} e^{-t^2/2}$ – функция плотности стандартного нормального распределения $N(0,1)$, математическое ожидание которого равно нулю, а дисперсия равна единице. Предельный эффект k -й объясняющей переменной равен $\varphi\left(x_i^T\right)\theta_k$ (а не θ_k – как в линейной модели).

В логит-модели $P\{y_i = 1 | x_i\} = \Lambda\left(x_i^T\theta\right) = \Lambda\left(\theta_1 x_{i1} + \ldots + \theta_p x_{ip}\right)$ малое изменение Δx_{ik} k -й объясняющей переменной приводит (при неизменных значениях остальных объясняющих переменных) к изменению вероятности $P\{y_i = 1 | x_i\}$ на величину, приближенно равную

$$\Delta P\{y_i = 1 | x_i\} \cong \frac{\partial P\{y_i = 1 | x_i\}}{\partial x_{ik}} \cdot \Delta x_{ik} = \frac{\partial \Lambda\left(x_i^T\theta\right)}{\partial x_{ik}} \cdot \Delta x_{ik} .$$

Учитывая явный вид функции $\Lambda(z)$, находим отсюда:

$$\Delta P\{y_i = 1 | x_i\} \cong \left\{\Lambda\left(x_i^T\theta\right)\left(1 - \Lambda\left(x_i^T\theta\right)\right)\theta_k\right\}\Delta x_{ik} .$$

Выражение, заключенное в фигурные скобки, представляет ***предельный эффект*** для k -й объясняющей переменной в логит-модели.

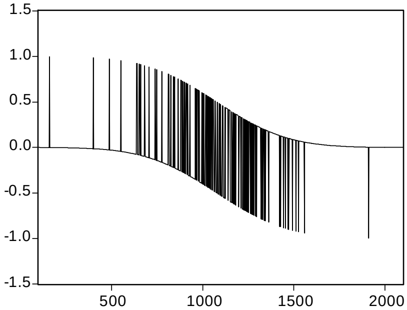
Заметим теперь следующее. Пусть $p = P(A)$ – вероятность некоторого события A, $0 < p < 1$. Отношение $\dfrac{p}{1-p}$ часто называют ***шансами*** (odds) этого события. Например, если $p = 2/3$, то $\dfrac{p}{1-p} = \dfrac{2/3}{1/3} = 2$, и шансы за то, что событие A произойдет, против того, что это событие не произойдет, равны 2:1 (“два к одному”, или “в 2 раза выше”). Логарифм отношения $\dfrac{p}{1-p}$ называют ***логитом*** (logit), $\operatorname{logit}(p) = \ln\left(\dfrac{p}{1-p}\right)$. Если $\operatorname{logit}(p) = 0$, то $p = 1 - p = 0.5$, т.е. шансы для события A равны “50 на 50”. Если $\operatorname{logit}(p) > 0$, то больше шансов, что событие A произойдет. Если $\operatorname{logit}(p) < 0$, то больше шансов, что событие A не произойдет.

Пусть теперь $p = P\{y_i = 1 | x_i\}$. В логит-модели $p = \Lambda(x_i^T\theta) = \dfrac{\exp(x_i^T\theta)}{1+\exp(x_i^T\theta)}$, $1 - p = \dfrac{1}{1+\exp(x_i^T\theta)}$, так что $\operatorname{logit}(p) = x_i^T\theta$, т.е. логит-модель линейна в отношении логита. Отсюда вытекает, что изменение значения k -й объясняющей переменной на величину Δx_{ik} приводит (при неизменных значениях остальных объясняющих переменных) к изменению значения $\ln\left(\dfrac{p}{1-p}\right)$ на $\theta_k \Delta x_{ik}$, что при малых значениях Δx_{ik} означает изменение значения отношения $\dfrac{p}{1-p}$ приблизительно на $100 \cdot \theta_k \Delta x_{ik}$ процентов. Иначе говоря, при этом шансы за то, что $y_i = 1$, против того, что $y_i = 0$, возрастают приблизительно на $100 \cdot \theta_k \Delta x_{ik}$ процентов.

1.5. *Проверка выполнения стандартных предположений*

При анализе обычных линейных моделей регрессии проверка выполнения стандартных предположений осуществляется посредством графического анализа и различных статистических критериев, призванных выявить наличие таких особенностей статистических данных, которые могут говорить не в пользу гипотезы о выполнении стандартных предположений.

Посмотрим, однако, на график остатков для пробит-модели, оцененной по рассматривавшемуся выше множеству данных о наличии (отсутствии) собственных автомобилей у 1000 семей.



Этот график по форме разительно отличается от тех, с которыми приходится сталкиваться при анализе обычных моделей регрессии с непрерывной объясняемой переменной. И это вовсе не должно нас удивлять, если вспомнить свойства случайных ошибок в моделях бинарного выбора: при заданных значениях объясняющих переменных случайная величина ε_i может принимать в i-м наблюдении только два значения. Соответственно, привычный графический анализ остатков не дает здесь полезной информации, и более полезным является непосредственное использование подходящих статистических критериев.

Поскольку мы используем для оценивания модели бинарного выбора метод максимального правдоподобия, естественным представляется сравнение максимумов функций правдоподобия, получаемых при оценивании модели с выполненными стандартными предположениями и при оценивании модели, в которой эти предположения не выполняются. При этом предполагается, что эти две модели – ***гнездовые***, т.е. первая вложена во вторую, так что вторая модель является более сложной, а первая является частным случаем второй модели.

Здесь надо заметить, что сравнением максимумов правдоподобий в двух гнездовых моделях мы фактически уже пользовались выше. Действительно, на таком сравнении основаны определения коэффициентов

$$pseudoR^2 = 1 - \frac{1}{1 + 2(\ln L_1 - \ln L_0)/n}$$

и

$$McFaddenR^2 = 1 - \frac{\ln L_1}{\ln L_0}.$$

В этом случае в качестве гнездовых моделей рассматриваются основная модель (с одной или несколькими объясняющими переменными помимо константы) и вложенная в нее тривиальная модель (в правую часть в качестве объясняющей переменной включается только константа).

Кроме того, если две гнездовые модели сравниваются с использованием информационных критериев (Акаике, Шварца, Хеннана–Куинна), то такое сравнение опять сводится к сравнению максимумов функций правдоподобия в этих моделях.

В этом разделе мы сосредоточимся на некоторых статистических критериях проверки гипотез о выполнении стандартных предположений, но прежде чем перейти к рассмотрению и применению подобных критериев, мы рассмотрим процесс порождения данных, приводящий к пробит-модели.

Предположим, что переменная y_i^* характеризует “полезность” наличия некоторого предмета длительного пользования для i-й семьи, и эта полезность определяется соотношением

$$y_i^* = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \ldots, n,$$

где $x_{i1}, \ldots, x_{ip}$ – значения p объясняющих переменных для i-й семьи, $\varepsilon_1, \ldots, \varepsilon_n$ – ***случайные ошибки***, отражающие влияние на полезность наличия указанного предмета для i-й семьи каких-то неучтенных дополнительных факторов. Пусть i-я семья приобретает этот предмет длительного пользования, если $y_i^* > \gamma_i$, где γ_i – ***пороговое значение***, и индикаторная переменная y_i отмечает наличие ($y_i = 1$) или отсутствие ($y_i = 0$) данного предмета у i -й семьи. Тогда

$$P\{y_i = 1 | x_i\} = P\{y_i^* > \gamma_i | x_i\} = P\{\beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i > \gamma_i | x_i\} =$$
$$= P\{\varepsilon_i > \gamma_i - \beta_1 x_{i1} - \cdots - \beta_p x_{ip} | x_i\},$$

и если $x_{i1} \equiv 1$, то

$$P\{y_i = 1 | x_i\} = P\{\varepsilon_i > (\gamma_i - \beta_1) - (\beta_2 x_{i2} + \cdots + \beta_p x_{ip}) | x_i\}.$$

Если предположить, что ошибки $\varepsilon_1, \ldots, \varepsilon_n$ – независимые в совокупности (и независимые от x_{ij}, $j = 1, \ldots, p$) случайные величины, имеющие одинаковое нормальное распределение $\varepsilon_i \sim N(0, \sigma^2)$, то тогда

$$P\{y_i = 1 | x_i\} = 1 - \Phi\left(\frac{(\gamma_i - \beta_1)}{\sigma} - \frac{(\beta_2 x_{i2} + \cdots + \beta_p x_{ip})}{\sigma}\right) =$$
$$= \Phi\left(\frac{(-\gamma_i + \beta_1)}{\sigma} + \frac{(\beta_2 x_{i2} + \cdots + \beta_p x_{ip})}{\sigma}\right).$$

(Здесь мы использовали вытекающее из симметрии стандартного нормального распределения соотношение $1 - \Phi(x) = \Phi(-x)$.) Обозначая

$$\theta_1 = \frac{(-\gamma_i + \beta_1)}{\sigma}, \quad \theta_j = \frac{\beta_j}{\sigma},$$

получаем:

$$P\{y_i = 1 | x_i\} = \Phi(\theta_1 x_{i1} + \cdots + \theta_p x_{ip}) = \Phi(x_i^T \theta).$$

Но именно таким образом и определяется пробит-модель.

Пусть мы имеем в наличии только значения y_i, $x_{i1}, \ldots, x_{ip}$, а значения y_i^* не доступны наблюдению. В таком случае переменную y_i^* называют ***латентной (скрытой) переменной***. Применяя метод максимального правдоподобия, мы получаем оценки параметров пробит-модели $\hat{\theta}_1,\dots,\hat{\theta}_p$, но не можем однозначно восстановить по ним значения параметров $\beta_1,\dots,\beta_p$, если не известны значения σ и $\gamma_1,\dots,\gamma_n$. Действительно, если оценки $\hat{\sigma}$, $\hat{\gamma}_1,\dots,\hat{\gamma}_n$, $\hat{\beta}_1,\dots,\hat{\beta}_p$ таковы, что

$$\hat{\theta}_1 = \frac{\left(-\hat{\gamma}_i + \hat{\beta}_1\right)}{\hat{\sigma}}, \quad \hat{\theta}_j = \frac{\hat{\beta}_j}{\hat{\sigma}},$$

то к тем же значениям $\hat{\theta}_1,\dots,\hat{\theta}_p$ приводят и оценки $k\hat{\sigma}$, $k\hat{\gamma}_1,\dots,k\hat{\gamma}_n$, $k\hat{\beta}_1,\dots,k\hat{\beta}_p$, где k – произвольное число, $-\infty < k < \infty$.

Таким образом, в рассмотренной ситуации для однозначной идентификации коэффициентов $\beta_1,\dots,\beta_p$ необходима какая-то нормализация функции полезности. В стандартной модели предполагается, что $\sigma = 1$ и $\gamma_1 = \dots = \gamma_n = 0$, так что

$$\hat{\beta}_1 = \hat{\theta}_1,\dots,\hat{\beta}_p = \hat{\theta}_p,$$

и именно такую модель мы будем теперь рассматривать.

Прежде всего заметим, что при получении оценок параметров $\beta_1,\dots,\beta_p$ в такой модели методом максимального правдоподобия мы принципиально опираемся на <u>предположение о нормальности ошибок</u> $\varepsilon_1,\dots,\varepsilon_n$: $\varepsilon_i \sim N(0,1)$. Поэтому важной является задача проверки этого предположения, т.е. проверка гипотезы

$$H_0 : \varepsilon_1,\dots,\varepsilon_n \sim i.i.d., \ \ \varepsilon_i \sim N(0,1).$$

Наряду со стандартной моделью (модель 1) рассмотрим модель 2, отличающуюся от стандартной тем, что в ней

$$P\{\varepsilon_i \le t\} = \Phi\left(t + \omega_1 t^2 + \omega_2 t^3\right),$$

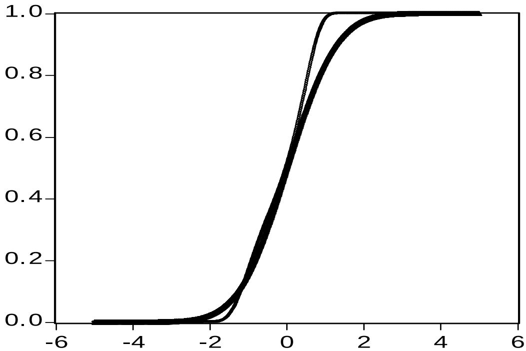
так что

$$P\{y_i = 1 | \, x_i\} = \Phi\left(x_i^T\theta + \omega_1\left(x_i^T\theta\right)^2 + \omega_2\left(x_i^T\theta\right)^3 \right).$$

При этом модель 1 является частным случаем модели 2 (при $\omega_1 = \omega_2 = 0$), так что модель 1 и модель 2 – гнездовые модели, и в рамках более общей модели 2 гипотеза H_0 принимает вид

$$H_0 : \omega_1 = \omega_2 = 0 .$$

Класс распределений вида $P\{\varepsilon_i \le t\} = \Phi\left(t + \omega_1 t^2 + \omega_2 t^3\right)$ допускает асимметрию и положительный эксцесс (островершинность) распределения. Следующий график позволяет сравнить поведение функции стандартного нормального распределения $\Phi(t)$ (толстая линия) и функции $\Phi\left(t + 0.5t^2 + 0.5t^3\right)$ (тонкая линия).



Пусть L_j – максимум функции правдоподобия в модели j, $j = 1, 2$, и $LR = 2(\ln L_2 - \ln L_1)$. ***Критерий отношения правдоподобий*** отвергает гипотезу H_0, если наблюдаемое значение статистики LR превышает критическое значение LR_{crit}, соответствующее выбранному уровню значимости α. Этот критерий асимптотический: критическое значение LR_{crit} вычисляется на основе распределения, к которому стремится при

$n \to \infty$ распределение статистики LR, если гипотеза H_0 верна. Этим предельным распределением является распределение хи-квадрат с двумя степенями свободы. Итак, в соответствии с критерием отношения правдоподобий, гипотеза H_0 отвергается, если

$$LR > \chi^2_{1-\alpha}(2),$$

где $\chi^2_{1-\alpha}(2)$ – квантиль уровня $1-\alpha$ распределения хи-квадрат с двумя степенями свободы.

Обратимся опять к смоделированным данным о наличии или отсутствии собственных автомобилей у 1000 домохозяйств.

Оценивая пробит-модель (модель 1) по этим данным, мы получили следующие результаты:

Коэффициент	Оценка	Std. Error	z-Statistic	Prob.
α	-3.503812	0.200637	-17.46343	0.0000
β	0.003254	0.000178	18.25529	0.0000
$\ln L_1$	-275.7686	Akaike info criterion		0.555537
		Schwarz criterion		0.565353
		Hannan-Quinn criter.		0.559268

Оценивание модели 2 дает следующие результаты:

Коэффициент	Оценка	Std. Error	z-Statistic	Prob.
α	-3.851178	0.324895	-11.85359	0.0000
β	0.003540	0.000292	12.11708	0.0000
ω_1	0.022954	0.025086	0.915039	0.3602
ω_2	-0.017232	0.010178	-1.693097	0.0904
$\ln L_2$	-274.6286	Akaike info criterion		0.557257
		Schwarz criterion		0.576888
		Hannan-Quinn criter.		0.564718

Соответственно, здесь

$$LR = 2\,(\ln L_2 - \ln L_1) = 2\,(275.7686 - 274.6286) = 2.28\,.$$

Поскольку же $\chi^2_{0.95}(2) = 5.99$, то критерий отношения правдоподобий <u>не отвергает</u> гипотезу H_0 при уровне значимости 0.05. Заметим еще, что значению $LR = 2.28$ соответствует (вычисляемое по асимптотическому распределению $\chi^2(2)$) P-значение 0.6802. Таким образом, критерий отношения правдоподобий не отвергает гипотезу H_0 при любом разумном уровне значимости.

Еще одним "стандартным предположением" является предположение об <u>одинаковой распределенности</u> случайных ошибок ε_i в процессе порождения данных. В сочетании с предположением нормальности этих ошибок, данное условие сводится к совпадению дисперсий всех этих ошибок. Нарушение этого условия приводит к гетероскедастичной модели и к <u>несостоятельности оценок максимального правдоподобия, получаемых на основании стандартной модели.</u> Для проверки гипотезы совпадения дисперсий мы можем опять рассмотреть какую-нибудь более общую модель с наличием гетероскедастичности, частным случаем которой является стандартная пробит-модель.

В примере с автомобилями можно допустить, что дисперсии случайных ошибок в процессе порождения данных возрастают с возрастанием значений x_i, например, как

$$D(\varepsilon_i | x_i) = \exp(k\, x_i), \quad k > 0\,,$$

так что (модель 3)

$$P\{y_i = 1 |\, x_i\} = \Phi\left(\frac{\alpha + \beta\, x_i}{\sqrt{\exp(k\, x_i)}}\right).$$

Здесь мы имеем две гнездовые модели – модель 3, допускающую гетероскедастичность в указанной форме, и модель 1 (стандартную пробит-модель) как ее частный случай. В рамках модели 3 выполнение стандартных предположений соответствует гипотезе

$H_0 : k = 0$.

Оценивание модели 3 по смоделированным данным дает следующие результаты:

	Coefficient	Std. Error	z-Statistic	Prob.
α	-3.141966	0.317695	-9.889867	0.0000
β	0.002883	0.000316	9.132687	0.0000
k	-0.000236	0.000186	-1.269192	0.2044
$\ln L_3$	-275.2619		Akaike info criterion	0.556524
			Schwarz criterion	0.571247
			Hannan-Quinn criter.	0.562120

При сравнении с моделью 1 получаем:

$LR = 2(\ln L_3 - \ln L_1) = 2(275.2619 - 274.6286) = 1.27$.

Это значение меньше критического значения 3.84, соответствующего уровню значимости 0.05 и вычисленного как квантиль уровня 0.95 асимптотического распределения хи-квадрат с одной степенью свободы. Следовательно, гипотеза $H_0 : k = 0$ не отвергается.

Отметим, что решения, принятые нами на основании критерия отношения правдоподобий, согласуются с решениями, принимаемыми в рассматриваемом примере на основании информационных критериев:

	AIC	SC	HQ
Модель 1 (пробит)	0.555537	0.565353	0.559268
Модель 2	0.557257	0.576888	0.564718
Модель 3 (гетеро)	0.556524	0.571247	0.562120

По всем трем критериям стандартная пробит-модель предпочтительнее альтернативных моделей.

1.6. Модели, в которых объясняемая переменная принимает несколько различных значений

1.6.1. Порядковая пробит-модель

В том же примере с наличием или отсутствием у семьи собственного автомобиля значение $y_i = 1$ говорило только о том, что i-я семья имеет собственный автомобиль, но не говорило о том, сколько в действительности автомобилей имеет семья – один, два или, быть может, еще больше. Обращаясь к процессу порождения данных, ориентирующемуся на значения функции полезности и сравнение ее с пороговыми значениями, можно предположить наличие не одного, а двух пороговых значений для каждой семьи, так что при превышении первого порога семья имеет в наличии один автомобиль, а при превышении второго (более высокого) порога – два или более автомобилей.

Обобщая эту ситуацию, рассмотрим процесс порождения данных, в котором имеется некоторая ненаблюдаемая (латентная) переменная y_i^*, значения которой связаны со значениями $x_{i1},\dots,x_{ip}$ объясняющих переменных для i-го субъекта исследования следующим образом:

$$y_i^* = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i\,, \quad i = 1,\dots,n\,.$$

Здесь ε_i – случайная ошибка, отражающая влияние на значение y_i^* неучтенных дополнительных факторов. Вместе со значениями $x_{i1},\dots,x_{ip}$ наблюдаются также значения переменной y_i, которая может принимать K различных значений, в соответствии со следующей схемой:

$$y_i = \begin{cases} 1, & \text{если } y_i^* \le \gamma_{i,1}, \\ \dots \\ k, & \text{если } \gamma_{i,k-1} < y_i^* \le \gamma_{i,k}, \\ \dots \\ K, & \text{если } y_i^* > \gamma_{i,K-1}, \end{cases}$$

где $\gamma_{i,1} < \dots < \gamma_{i,k} < \dots < \gamma_{i,K-1}$ – пороговые значения, вообще говоря, ненаблюдаемые.

Предполагая, что ошибки $\varepsilon_1, \dots, \varepsilon_n$ – независимые в совокупности (и независимые от $x_{ij}, j = 1, \dots, p$) случайные величины, имеющие одинаковое нормальное распределение $\varepsilon_i \sim N(0, \sigma^2)$, мы получаем ***порядковую пробит-модель***.

Рассмотрим частный случай, когда $K = 3$ и пороговые значения одинаковы для всех субъектов исследования, так что $\gamma_{i,1} \equiv \gamma_1$, $\gamma_{i,2} \equiv \gamma_2$ и

$$y_i = \begin{cases} 1, & \text{если } y_i^* \le \gamma_1, \\ 2, & \text{если } \gamma_1 < y_i^* \le \gamma_2, \\ 3, & \text{если } y_i^* > \gamma_2. \end{cases}$$

При этом

$$P\{y_i = 1 | x_i\} = P\{y_i^* \le \gamma_1 | x_i\} = P\{\beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i \le \gamma_1 | x_i\} =$$
$$= P\{\varepsilon_i \le (\gamma_1 - \beta_1) - (\beta_2 x_{i2} + \dots + \beta_p x_{ip}) | x_i\} =$$
$$= \Phi\left(\frac{(\gamma_1 - \beta_1)}{\sigma} - \frac{(\beta_2 x_{i2} + \dots + \beta_p x_{ip})}{\sigma} \right).$$

$$P\{y_i = 2| x_i\} = P\{\gamma_1 < y_i^* \le \gamma_2 | x_i\} =$$
$$= P\{\gamma_1 < \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i \le \gamma_2 | x_i\} =$$
$$= \Phi\left(\frac{(\gamma_2 - \beta_1)}{\sigma} - \frac{(\beta_2 x_{i2} + \cdots + \beta_p x_{ip})}{\sigma} \right) -$$
$$- \Phi\left(\frac{(\gamma_1 - \beta_1)}{\sigma} - \frac{(\beta_2 x_{i2} + \cdots + \beta_p x_{ip})}{\sigma} \right),$$
$$P\{y_i = 3| x_i\} = P\{y_i^* > \gamma_2 | x_i\} = P\{\beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i > \gamma_2 | x_i\} =$$
$$= P\{\varepsilon_i > (\gamma_2 - \beta_1) - (\beta_2 x_{i2} + \cdots + \beta_p x_{ip}) | x_i\} =$$
$$= 1 - \Phi\left(\frac{(\gamma_2 - \beta_1)}{\sigma} - \frac{(\beta_2 x_{i2} + \cdots + \beta_p x_{ip})}{\sigma} \right).$$

Пусть мы имеем в наличии только значения y_i, $x_{i1}, \ldots, x_{ip}$, $i = 1, \ldots, n$. Применяя метод максимального правдоподобия, мы, как и в случае пробит-модели с двумя исходами, не можем однозначно восстановить значения параметров $\beta_1, \ldots, \beta_p$, если не известны значения σ, γ_1 и γ_2. Поэтому и здесь для однозначной идентификации коэффициентов $\beta_1, \ldots, \beta_p$ необходима какая-то нормализация. В стандартной модели предполагается, что $\sigma = 1$ и $\gamma_1 = 0$, хотя возможны и другие нормализации.

Используя стандартную нормализацию и обозначая $\gamma_2 = \gamma$, мы получаем в модели с тремя исходами:

$$P\{y_i = 1| x_i\} = P\{y_i^* \le 0 | x_i\} = \Phi(- x_i^T \beta),$$
$$P\{y_i = 2| x_i\} = P\{0 < y_i^* \le \gamma | x_i\} = \Phi(\gamma - x_i^T \beta) - \Phi(- x_i^T \beta),$$
$$P\{y_i = 3| x_i\} = P\{y_i^* > \gamma | x_i\} = 1 - \Phi(\gamma - x_i^T \beta).$$

При этом коэффициент β_j допускает двойное истолкование. В соответствии с моделью для y_i^*, положительное значение этого коэффициента означает, что переменная y_i^* возрастает с возрастанием j-й объясняющей переменной. В соответствии с приведенными выражениями для вероятностей получения значений $y_i = 1$, $y_i = 2$ и $y_i = 3$, последнее приводит к возрастанию вероятности $P\{y_i = 3 | x_i\}$ и к убыванию вероятности $P\{y_i = 1 | x_i\}$. Что же касается вероятности $P\{y_i = 2 | x_i\}$, то здесь возможно как возрастание, так и убывание этой вероятности, в зависимости от конкретной ситуации.

<u>Прогнозирование</u> по оцененной модели производится в соответствии со следующим соглашением. Прогнозное значение $\hat{y}_i$ полагается равным k_0, если $\hat{P}\{y_i = k_0 | x_i\} = \max\limits_{k=1,\ldots,K} \hat{P}\{y_i = k | x_i\}$.

П р и м е р

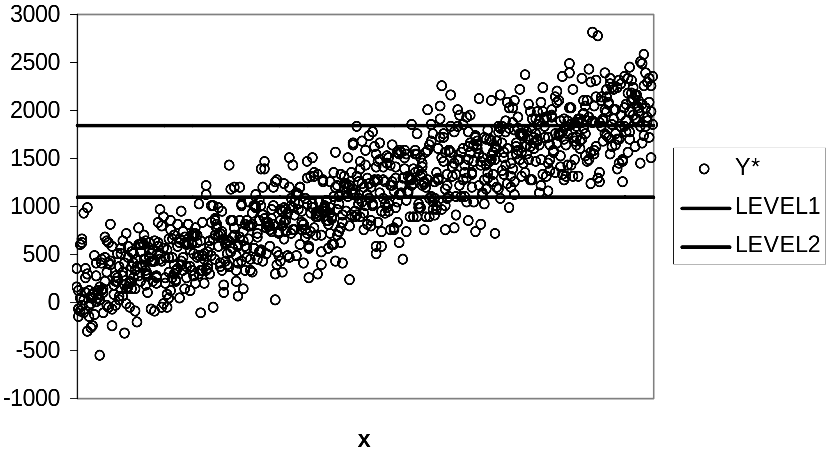
Рассмотрим теперь выборку, состоящую из 1000 семей со среднедушевым месячным доходом от 100 до 2100 условных единиц (у.е.), среди которых 499 семей не имеет собственного автомобиля, 369 семей имеет один автомобиль, 132 семьи имеют два автомобиля. Выборка получена посредством моделирования; при этом был использован процесс порождения данных в виде

$$y_i^* = x_i + \varepsilon_i, \quad i = 1, \ldots, 1000,$$

где ε_i – независимые в совокупности (и независимые от x_i) случайные величины, имеющие одинаковое нормальное распределение $\varepsilon_i \sim N(0, 300^2)$, т.е. $\sigma = 300$. Здесь $K = 3$ и границы γ_1 и γ_2 были выбраны равными, соответственно, $\gamma_1 = 1100$ и $\gamma_1 = 1850$, так что в результате получаем порядковую пробит-модель

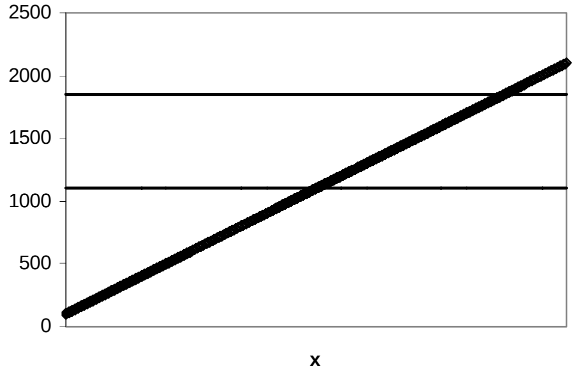
$$y_i = \begin{cases} 1, & \text{если } y_i^* \le \gamma_1, \\ 2, & \text{если } \gamma_1 < y_i^* \le \gamma_2, \\ 3, & \text{если } y_i^* > \gamma_2, \end{cases}$$

где $y_1 = 1$, если i-я семья не имеет автомобиля, $y_1 = 2$, если i-я семья имеет один автомобиль, и $y_1 = 3$, если i-я семья имеет два (или более) автомобиля. На следующем графике показана зависимость полученных значений y_i^* от x_i



Горизонтальные линии соответствуют разделительным порогам LEVEL1=1100 и LEVEL2=1850.

Наблюдения с $y_i^* \le 1100$ встречаются в группе семей с доходами от 200 до 1600 у.е. Наблюдения с $1100 < y_i^* \le 1850$ встречаются в группе семей с доходами от 548 до 2094 у.е. Наблюдения с $y_i^* > 1850$ встречаются в группе семей с доходами от 1318 у.е. и выше. Важно отметить, что эти группы пересекаются, и это связано как раз с наличием случайной составляющей в уравнении полезности. Если бы этой составляющей не было, то мы имели следующую картину.



И тогда мы получили бы разбиение на три непересекающиеся группы. Для всех семей с доходами, не превышающими 1100 у.е., $y_i = 1$. Для всех семей с доходами, превышающими 1100 у.е., но не превышающими 1850 у.е., $y_i = 2$. Для всех семей с доходами, превышающими 1850 у.е., $y_i = 3$.

Представим теперь, что мы имеем в распоряжении только выборочные данные, т.е. пары (x_i, y_i), $i = 1,\ldots,1000$. Оценивание методом максимального правдоподобия порядковой пробит-модели с нормализацией $\sigma = 1$, $\alpha = 0$ (именно такая нормализация используется в пакете EVIEWS), дает следующие результаты:

	Coefficient	Std. Error	z-Statistic	Prob.
X	0.003361	0.000158	21.31648	0.0000
	Limit Points			
γ_1	3.693723	0.185109	19.95431	0.0000
γ_2	6.306692	0.279737	22.54510	0.0000

Иначе говоря, нормализованная модель оценивается как

$$y_i^* = 0.003361 x_i + u_i ,$$

где $u_i \sim N(0,1)$, и

$$y_i = \begin{cases} 1, & \text{если } y_i^* \le 3.693723, \\ 2, & \text{если } 3.693723 < y_i^* \le 6.306692, \\ 3, & \text{если } y_i^* > 6.306692. \end{cases}$$

Если учесть, что мы сами смоделировали выборку и поэтому знаем значение σ, то переход к модели с $\sigma = 300$ соответствует оцененной модели

$$y_i^* = 300 \cdot 0.003361\, x_i + 300 \cdot u_i = 1.0083\, x_i + \varepsilon_i ,$$

где $\varepsilon_i \sim N\left(0, 300^2\right)$, и

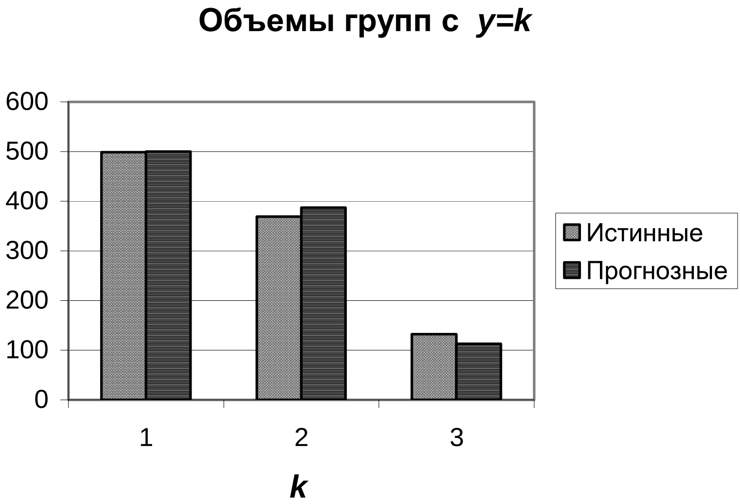
$$\gamma_1 = 300 \cdot 3.693723 = 1108.1169 ,$$

$$\gamma_2 = 300 \cdot 6.306692 = 1892.0076 .$$

Как видим, параметры оцененной модели очень близки к параметрам истинной модели. Результаты прогнозов по оцененной модели приведены в следующей таблице.

y_i	Кол-во набл.	$\hat{y}_i$	Кол-во набл.	Ошибка $y_i - \hat{y}_i$
1	499	1	500	-1
2	369	2	387	-18
3	132	3	113	19

Содержимое таблицы отражает следующая диаграмма.



Для сравнения приведем результаты прогнозов по тривиальной модели, не учитывающей в уравнении для y_i^* влияние доходов *i*-й семьи:

y_i	Кол-во набл.	$\hat{y}_i$	Кол-во набл.	Ошибка $y_i - \hat{y}_i$
1	499	1	1000	-501
2	369	2	0	369
3	132	3	0	132

Приведем также сводную таблицу количеств правильных и неправильных прогнозов для значений $y_i = 1, 2, 3$.

	$\hat{y}_i=1$	$\hat{y}_i=2$	$\hat{y}_i=3$
$y_i=1$	438	61	0
$y_i=2$	62	265	42
$y_i=3$	0	61	71

Таким образом, из 1000 прогнозов правильными оказались 774, т.е. 77.4%. При этом значения $y_i = 1$ правильно прогнозируются в 438 случаях из 499, т.е. в 87.8% случаев; значения $y_i = 2$ правильно прогнозируются в 71.8% случаев; значения $y_i = 3$ правильно прогнозируются в 53.8% случаев.

1.6.2. Мультиномиальная модель

В целом ряде случаев не существует естественного упорядочения альтернатив, благодаря которому и возникает монотонная связь между непрерывной латентной переменной и наблюдаемой переменной, принимающей конечное количество значений.

Пусть мы имеем K таких альтернатив (мы занумеруем их в произвольном порядке числами $1,\ldots,K$) и пусть i-й субъект исследования приписывает k-й альтернативе полезность u_{ik}, так что

$$u_{ik} = \beta_1 x_{i1,k} + \cdots + \beta_p x_{ip,k} + \varepsilon_{ik} = x_{ik}^T \beta + \varepsilon_{ik}\,, \quad i = 1,\ldots,n\,,$$

где $x_{ik} = \left(x_{i1,k},\ldots,x_{ip,k}\right)^T$, а ε_{ik} ($i = 1,\ldots,n$, $k = 1,\ldots,K$) – независимые в совокупности (и независимые от x_{ik}) случайные величины, имеющие одинаковое распределение.

Предположим, что i-й субъект выбирает альтернативу k, если для него эта альтернатива имеет максимальную полезность. В этом случае мы полагаем $y_i = k$. Тогда (условная при заданных значениях x_{ik}, $k = 1,\ldots,K$) вероятность того, что i-й субъект выберет альтернативу k, равна

$$P\{y_i = k\} = P\left\{u_{ik} = \max_{j=1,\ldots,K} u_{ij}\right\} = P\left\{x_{ik}^T\beta + \varepsilon_{ik} > \max_{j=1,\ldots,K,\, j\neq k}\left(x_{ij}^T\beta + \varepsilon_{ij}\right)\right\}.$$

Выразить такую вероятность в явном виде весьма проблематично. Однако если предположить, что общим для всех случайных величин ε_{ik} является стандартное распределение экстремальных значений (максимума) I-го типа с функцией распределения

$$G(z) = \exp\left(-e^{-z}\right),\ -\infty < z < \infty,$$

(это распределение часто называют также распределением Гумбеля), то формула для вычисления вероятности $P\{y_i = k\}$ принимает достаточно простой вид, а именно:

$$P\{y_i = k\} = \frac{\exp\left(x_{ik}^T\beta\right)}{\exp\left(x_{i1}^T\beta\right) + \exp\left(x_{i2}^T\beta\right) + \cdots + \exp\left(x_{iK}^T\beta\right)}.$$

Заметим, однако, что если и числитель и знаменатель правой части последнего выражения разделить на $\exp\left(x_{i1}^T\beta\right)$, то получим

$$P\{y_i = k\} = \frac{\exp\left(x_{ik}^T\beta - x_{i1}^T\beta\right)}{1 + \exp\left(x_{i2}^T\beta - x_{i1}^T\beta\right) + \cdots + \exp\left(x_{iK}^T\beta - x_{i1}^T\beta\right)}.$$

Следовательно, каким бы ни было значение линейной комбинации $x_{i1}^T\beta$, вероятность $P\{y_i = k\}$ будет зависеть <u>только от разностей</u> $\left(x_{i2}^T\beta - x_{i1}^T\beta\right), \ldots, \left(x_{iK}^T\beta - x_{i1}^T\beta\right)$. Это обстоятельство приводит к естественной нормализации, при которой полагают

$$x_{i1}^T\beta = 0,\ i = 1,\ldots,n,$$

так что тогда

$$P\{y_i = k\} = \frac{\exp\left(x_{ik}^T\beta\right)}{1 + \exp\left(x_{i2}^T\beta\right) + \cdots + \exp\left(x_{iK}^T\beta\right)}.$$

Такую модель разные авторы называют по-разному. Так, в книгах [Verbeek (2000)] и [Amemiya (1985)] об этой модели говорится как о ***мультиномиальной логит-модели*** *(multinomial logit model).* В книгах [Green (1993)] и [Davidson, MacKinnon (1993)] эта модель именуется ***условной логит-моделью*** *(conditional logit model)*, а под ***мультиномиальной логит-моделью*** подразумевается модель

$$P\{y_i = k\} = \frac{\exp\left(x_i^T \beta^k\right)}{\exp\left(x_i^T \beta^1\right) + \exp\left(x_i^T \beta^2\right) + \cdots + \exp\left(x_i^T \beta^K\right)},$$

в которой объясняющие переменные специфичны только в отношении <u>самих субъектов исследования</u> (но не в отношении альтернатив), а специфичными в отношении альтернатив являются <u>коэффициенты</u> модели. Соответственно, здесь $\beta^k = \left(\beta_{1,k}, \ldots, \beta_{p,k}\right)^T$ – вектор коэффициентов при объясняющих переменных в представлении функции полезности для k -й альтернативы:

$$u_{ik} = \beta_{1,k} x_{i1} + \cdots + \beta_{p,k} x_{ip} + \varepsilon_{ik} = x_i^T \beta^k + \varepsilon_{ik}, \quad i = 1, \ldots, n.$$

Последняя модель под названием ***мультиномиальной логит-модели*** появляется и в пакете EVIEWS. Поскольку в этой модели x_i не зависят от альтернативы, являясь собственными атрибутами субъекта, то

$$P\{y_i = k\} = \frac{\exp\left(x_i^T \left(\beta^k - \beta^1\right)\right)}{1 + \exp\left(x_i^T \left(\beta^2 - \beta^1\right)\right) + \cdots + \exp\left(x_i^T \left(\beta^K - \beta^1\right)\right)},$$

так что эта вероятность зависит только от разностей $\beta^2 - \beta^1$, $\ldots, \beta^K - \beta^1$, и для нормализации можно положить вектор β^1 равным нулевому вектору. При такой нормализации

$$P\{y_i = k\} = \frac{\exp\left(x_i^T \beta^k\right)}{1 + \exp\left(x_i^T \beta^2\right) + \cdots + \exp\left(x_i^T \beta^K\right)}.$$

В этом случае (условная при фиксированных x_{ij}, $j = 1, \ldots, p$, $i = 1, \ldots, n$) совместная вероятность получения конкретного набора наблюдений $y_1, \ldots, y_n$ (конкретного набора значений $1, \ldots, K$) равна произведению

$$\prod_{i=1}^{n}\prod_{k=1}^{K}\left(P\{y_i=k\}\right)^{d_{ik}}=\prod_{i=1}^{n}\prod_{k=1}^{K}\left(\frac{\exp\left(x_i^T\beta^k\right)}{1+\exp\left(x_i^T\beta^2\right)+\cdots+\exp\left(x_i^T\beta^K\right)}\right)^{d_{ik}},$$

где

$$d_{ik}=\begin{cases}1, \text{ если } \ y_i=k,\\ 0, \text{ если } \ y_i\neq k.\end{cases}$$

Правая часть этого выражения является при фиксированных x_i, $i=1,\ldots,n$, функцией от вектора неизвестных параметров β, $\beta=(\beta^1,\ldots,\beta^K)^T$:

$$L(\beta)=L\left(\beta\middle|x_1,\ldots,x_n\right)=\prod_{i=1}^{n}\prod_{k=1}^{K}\left(\frac{\exp\left(x_i^T\beta^k\right)}{1+\exp\left(x_i^T\beta^2\right)+\cdots+\exp\left(x_i^T\beta^K\right)}\right)^{d_{ik}},$$

и эта функция как функция правдоподобия является объектом максимизации по β. Результатом такой максимизации являются ***оценки максимального правдоподобия*** для векторов коэффициентов $\hat{\beta}^k=\left(\hat{\beta}_{1,k},\ldots,\hat{\beta}_{p,k}\right)^T$, $k=1,\ldots,K$.

П р и м е р

Рассмотрим смоделированную ситуацию, в которой, как и в последней модели, переменные специфичны только в отношении самих субъектов исследования.

Пусть $x_{i1}\equiv 1$, x_{i2} – типичное количество посещений продуктового магазина в неделю i-й семьей (от 1 до 7), x_{i3} – среднемесячный доход на одного члена i-й семьи (от 50 до 250 у.е.). Выбранная модель порождения данных имитирует поведение 1000 семей, проживающих в одном и том же многоэтажном доме и приобретающих продукты в трех продуктовых магазинах, ближайших к этому дому. Каждая семья отдает предпочтение одному из трех магазинов, так что мы имеем здесь 3 альтернативы. Магазины различаются тремя сравнительными характеристиками:

ассортиментом (наименее разнообразный из трех, наиболее разнообразный из трех, промежуточный), удаленностью от дома (наибольшая, наименьшая, средняя) и уровнем цен (максимальный, минимальный, средний). Альтернативы были занумерованы числами $1, 2, 3$ произвольным образом. В итоге была получена следующая нумерация.

k	Характеристики k-го магазина		
	Ассортимент	Удаленность	Уровень цен
1	Богатый	Максимальная	Средний
2	Бедный	Минимальная	Минимальный
3	Промежуточный	Средняя	Максимальный

Предполагается, что i-я семья приписывает k-й альтернативе полезность u_{ik}, где

$$u_{ik} = \beta_{k1}x_{i1} + \beta_{k2}x_{i2} + \beta_{k3}x_{i3} + \varepsilon_{ik}, \quad i = 1, \ldots, 1000,$$

где ε_{ik} ($i = 1, \ldots, 1000$, $k = 1, 2, 3$) – независимые в совокупности (и независимые от x_{ij}) случайные величины, имеющие одинаковое распределение с функцией распределения

$$G(z) = \exp\left(-e^{-z}\right), \ -\infty < z < \infty .$$

При этом мы используем нормализацию

$$\beta_{11} = 0, \ \beta_{12} = 0, \ \beta_{13} = 0 .$$

Остальные коэффициенты выбраны следующим образом:

$$\beta_{21} = -0.8, \ \beta_{22} = 1.0, \ \beta_{23} = -0.0032 ,$$

$$\beta_{31} = -0.4, \ \beta_{32} = 0.3, \ \beta_{33} = 0.0032 ,$$

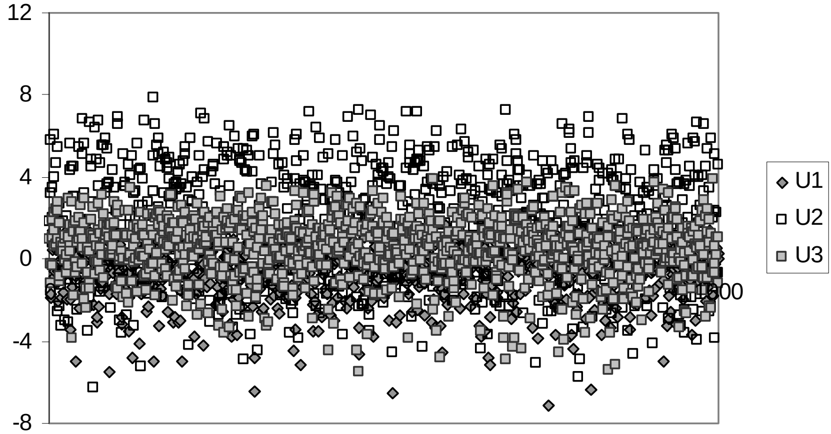
так что функции полезности для трех альтернатив имеют вид

$$u_{i1} = \varepsilon_{i1},$$

$$u_{i2} = -0.8 + x_{i2} - 0.0032x_{i3} + \varepsilon_{i2},$$

$$u_{i3} = -0.4x_{i1} + 0.3x_{i2} + 0.0032x_{i3} + \varepsilon_{i3} .$$

Их поведение иллюстрирует следующий график.



В соответствии с моделью порождения данных, i-я семья выбирает альтернативу k, если для этой семьи альтернатива k имеет максимальную полезность. В этом случае полагаем $y_i = k$.

Результаты оценивания методом максимального правдоподобия:

	Coefficient	Std. Error	z-Statistic	Prob.
β_{21}	-1.655130	0.358914	-4.611496	0.0000
β_{22}	1.270612	0.097636	13.01381	0.0000
β_{23}	-0.001778	0.002134	-0.833304	0.4047
β_{31}	-1.031242	0.327444	-3.149372	0.0016
β_{32}	0.439590	0.087563	5.020273	0.0000
β_{33}	0.006283	0.001957	3.211368	0.0013

Все оцененные коэффициенты, за исключением $\hat{\beta}_{23}$, имеют высокую статистическую значимость.

Сравним истинные и оцененные значения коэффициентов:

	Истинное значение	Оценка
β_{21}	-0.8	-1.655130
β_{22}	1.0	1.270612
β_{23}	-0.0032	-0.001778
β_{31}	-0.4	-1.031242
β_{32}	0.3	0.439590
β_{33}	0.0032	0.006283

Знаки оцененных коэффициентов соответствуют знакам истинных значений коэффициентов. Кроме того, соблюдается упорядочение значений соответственных коэффициентов, имеющих одинаковые знаки:

$\beta_{21} < \beta_{31}$ и $\hat{\beta}_{21} < \hat{\beta}_{31}$,

$\beta_{22} > \beta_{32}$ и $\hat{\beta}_{22} > \hat{\beta}_{32}$.

На основании полученных оценок коэффицентов можно вычислить <u>прогнозные</u> значения вероятностей $P\{y_i = k\}$ предпочтения альтернатив $k = 1, 2, 3$, полагая

$$\hat{P}\{y_i = k\} = \frac{\exp\left(x_i^T \hat{\beta}^k\right)}{1 + \exp\left(x_i^T \hat{\beta}^2\right) + \exp\left(x_i^T \hat{\beta}^3\right)},$$

и, используя эти прогнозные значения, дать предсказание номера альтернативы, которую предпочтет семья из рассматриваемого дома с заданной частотой посещения продуктового магазина и заданным уровнем месячного дохода на одного члена семьи. Можно, например, предсказывать для i-й семьи в качестве предпочтительной альтернативу k, если

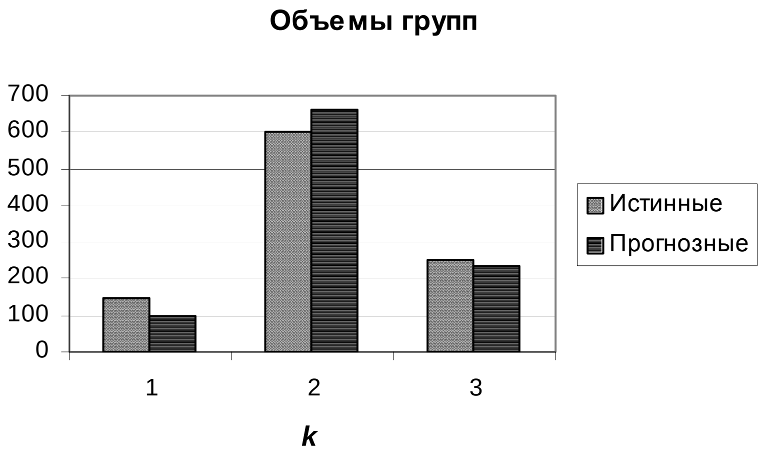
$$\hat{P}\{y_i = k\} > \hat{P}\{y_i = l\}, \ l \neq k.$$

Применяя такое правило к нашему примеру, получаем следующие результаты.

Альтернатива (k)	1	2	3
Истинный объем группы k	146	603	251
Прогноз объема группы k	101	664	235

Здесь под группой k подразумевается группа семей (среди рассматриваемых 1000 семей), отдающих предпочтение альтернативе k .

Следующая диаграмма отображает содержимое таблицы.



Предсказанные объемы групп правильно воспроизводят упорядочение между наблюдаемыми размерами групп: в обоих случаях максимальное количество семей предпочитает альтернативу 2 и минимальное количество семей предпочитает альтернативу 1.

Хотя индивидуальные прогнозы и не являются главной целью в подобных исследованиях, мы все же приведем сводную таблицу количеств правильных и неправильных прогнозов для значений $y_i = 1, 2, 3$.

	$\hat{y}_i = 1$	$\hat{y}_i = 2$	$\hat{y}_i = 3$
$y_i = 1$	48	26	72
$y_i = 2$	11	550	42
$y_i = 3$	42	88	121

Таким образом, из 1000 прогнозов правильными оказались 719, т.е. 71.9%. При этом значения $y_i = 1$ правильно прогнозируются в 48 случаях из 146, т.е. только в 32.9% случаев, тогда как значения $y_i = 2$ правильно прогнозируются в 91.2% случаев; значения $y_i = 3$ правильно прогнозируются в 48.2% случаев.

П р и м е р

В следующей ситуации, в отличие от предыдущих примеров, одна из переменных специфична только в отношении альтернатив, а другая зависит и от альтернативы и от субъекта.

Пусть $stores_k$ – количество магазинов в k-м (из трех) торговом центре, $dist_{ik}$ – расстояние от места проживания i-й семьи до k-го торгового центра. Выбранная модель порождения данных имитирует поведение 1000 семей, предпочитающих совершать покупки в этих трех торговых центрах. Каждая семья отдает предпочтение одному из трех торговых центров, так что мы имеем здесь 3 альтернативы. Альтернативы были занумерованы числами $1, 2, 3$ произвольным образом.

Здесь переменная $stores_k$ специфична только в отношении альтернатив, тогда как значения переменной $dist_{ik}$ зависят и от альтернативы и от конкретной семьи.

Предполагается, что i-я семья приписывает k-й альтернативе полезность u_{ik},

$$u_{ik} = \beta_1 stores_k + \beta_2 dist_{ik} + \varepsilon_{ik}, \quad i = 1, \ldots, 1000,$$

где ε_{ik} ($i=1,\ldots,1000$, $k=1,2,3$) – независимые в совокупности (и независимые от $stores_k$ и $dist_{ik}$) случайные величины, имеющие одинаковое распределение с функцией распределения $G(z)=\exp\left(-e^{-z}\right), -\infty<z<\infty$.

Коэффициенты выбраны следующим образом:

$$\beta_1=0.6,\ \beta_2=-1.0,$$

так что функции полезности для трех альтернатив имеют вид

$$u_{i1} = 0.6stores_1 - dist_{i1} + \varepsilon_{i1},$$

$$u_{i2} = 0.6stores_2 - dist_{i2} + \varepsilon_{i2},$$

$$u_{i3} = 0.6stores_3 - dist_{i3} + \varepsilon_{i3}.$$

В соответствии с моделью порождения данных, i-я семья выбирает альтернативу k, если для этой семьи альтернатива k имеет максимальную полезность. В этом случае полагаем $y_i = k$.

Результаты оценивания методом максимального правдоподобия:

	Coefficient	Std. Error	z-Statistic	Prob.
β_1	0.932414	0.061646	15.12519	0.0000
β_2	-1.521518	0.101902	-14.93120	0.0000

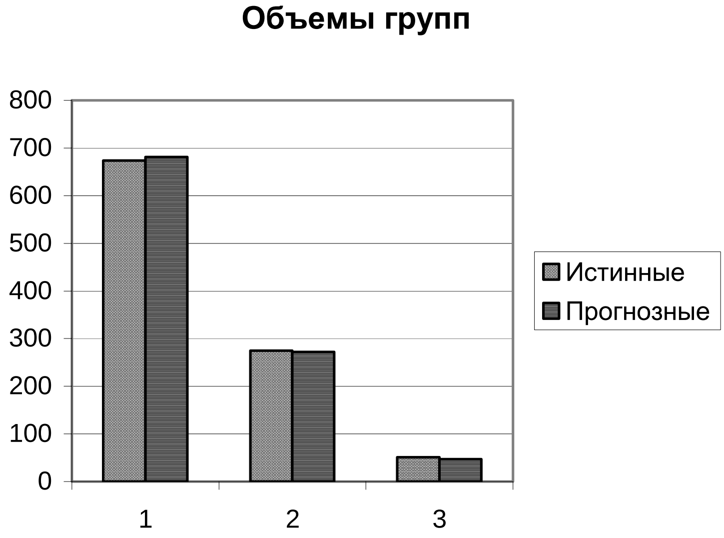
Будем опять предсказывать для i-й семьи в качестве предпочтительной альтернативу k, если

$$\hat{P}\{y_i = k\} > \hat{P}\{y_i = l\},\ l \neq k.$$

Применяя такое правило к нашему примеру, получаем следующие результаты.

Альтернатива (k)	1	2	3
Истинный объем группы k	674	275	51
Прогноз объема группы k	681	272	47

Следующая диаграмма отображает содержимое таблицы.



З а м е ч а н и е 1

Как мы уже отмечали выше, в рассмотренной нами мультиномиальной логит-модели, в которой объясняющие переменные специфичны только в отношении самих субъектов исследования,

$$P\{y_i = k\} = \frac{\exp\left(x_i^T\left(\beta^k - \beta^1\right)\right)}{1 + \exp\left(x_i^T\left(\beta^2 - \beta^1\right)\right) + \cdots + \exp\left(x_i^T\left(\beta^K - \beta^1\right)\right)}.$$

Отсюда вытекает, что

$$\frac{P\{y_i = k\}}{P\{y_i = m\}} = \frac{\exp\left(x_i^T\left(\beta^k - \beta^1\right)\right)}{\exp\left(x_i^T\left(\beta^m - \beta^1\right)\right)} = \exp\left(x_i^T\left(\beta^k - \beta^m\right)\right),$$

т.е. отношение вероятностей выбора альтернатив k и m определяется только параметрами уравнений для полезностей этих двух альтернатив и собственными атрибутами i -го субъекта и не зависит от параметров уравнений для полезностей остальных $K-2$ альтернатив.

Замечание 2

Если рассматривается условная логит-модель (с постоянными значениями коэффициентов во всех K уравнениях полезности), в которой объясняющие переменные специфичны в отношении альтернатив, то, как уже говорилось выше, в такой ситуации

$$P\{y_i = k\} = \frac{\exp\left(x_{ik}^T \beta\right)}{\exp\left(x_{i1}^T \beta\right) + \cdots + \exp\left(x_{iK}^T \beta\right)},$$

так что здесь

$$\frac{P\{y_i = k\}}{P\{y_i = m\}} = \frac{\exp\left(x_{ik}^T \beta\right)}{\exp\left(x_{im}^T \beta\right)} = \exp\left(\left(x_{ik}^T - x_{im}^T\right)\beta\right),$$

т.е. отношение вероятностей выбора альтернатив k и m определяется только общим параметром уравнений для полезностей различных альтернатив и значениями в i-м наблюдении объясняющих переменных, соответствующих k-й и m-й альтернативам. Это отношение не зависит от значений в i-м наблюдении объясняющих переменных, соответствующих остальным $K-2$ альтернативам. Такое свойство независимости оказывается нежелательным во многих ситуациях.

Замечание 3

Пусть среди объясняющих переменных в условной логит-модели (с постоянными значениями коэффициентов во всех K уравнениях полезности) имеются переменные, специфичные только в отношении субъектов (т.е. значения этих переменных для i-го субъекта не зависят от альтернативы). Пусть, соответственно,

$$x_{ik}^T = \left(v_{ik}^T, w_i^T\right),$$

где v_{ik}^T – вектор значений для i-го субъекта переменных, значения которых <u>зависят</u> от альтернативы, а w_i^T - вектор значений для i-го субъекта переменных, значения которых <u>не зависят</u> от альтернативы; соответственно разбивается и вектор коэффициентов:

$$\beta^T = \left(\gamma^T, \delta^T\right).$$

Тогда

$$P\{y_i = k\} = \frac{\exp\left(v_{ik}^T \gamma + w_i^T \delta\right)}{\exp\left(v_{i1}^T \gamma + w_i^T \delta\right) + \cdots + \exp\left(v_{iK}^T \gamma + w_i^T \delta\right)} =$$

$$= \frac{\exp\left(v_{ik}^T \gamma\right)}{\exp\left(v_{i1}^T \gamma\right) + \cdots + \exp\left(v_{iK}^T \gamma\right)},$$

так что эта вероятность <u>не зависит</u> от значений переменных, специфичных <u>только</u> в отношении субъектов.

Чтобы (в рамках модели с постоянным вектором коэффициентов) учесть возможное влияние таких переменных на вероятности $P\{y_i = k\}$, модель надо модифицировать. Одним из возможных способов модификации является создание группы дамми переменных для альтернатив (*DUMMY* для альтернативы k принимает значение 1, если $y_i = k$, и принимает значение 0 в противном случае) и умножение каждой из них на переменные, не зависящие от альтернатив. Тем самым достигается изменение коэффициентов при этих переменных в зависимости от альтернатив.

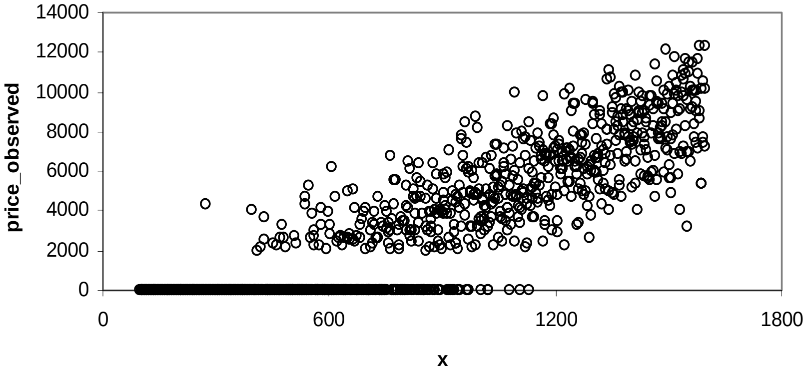
1.7. Цензурированная модель регрессии (тобит–модель)

Развивая пример с наличием или отсутствием у семьи собственного автомобиля, представим, что мы имеем следующие данные. Для семей, имеющих автомобиль, известна стоимость этого автомобиля s_i (если в семье несколько автомобилей, то s_i – суммарная стоимость этих автомобилей). Таким образом, здесь мы наблюдаем пары $(x_i, price\_observed_i)$, где x_i – среднедушевой месячный доход i-й семьи,

$$price\_observed_i = \begin{cases} s_i, & \text{если } i\text{-я семья имеет автомобиль,} \\ 0, & \text{если } i\text{-я семья не имеет автомобиля.} \end{cases}$$

Обратимся к смоделированной выборке, состоящей из 1000 семей со среднедушевым месячным доходом от 100 до 1600 у.е. Для удобства наблюдения переупорядочены в соответствии в возрастанием x_i, так что $x_1 \le x_2 \le \cdots \le x_{1000}$.

Диаграмма рассеяния для этих данных имеет весьма специфический вид:



Обращает на себя внимание большое количество точек, расположенных на оси абсцисс. Таких точек 418, и это означает, что 418 из 1000 рассматриваемых семей не имеет собственного автомобиля. В то же время среди семей, владеющих автомобилем, минимальное значение цены автомобиля равно 2002 у.е., и это может просто означать, что на автомобильном рынке, в том числе и вторичном, просто нет автомобилей с ценой менее 2000 у.е.

Как проводить статистический анализ подобных данных? Можно попытаться, например, использовать все 1000 наблюдений и оценить по этим наблюдениям методом наименьших квадратов линейную статистическую модель

$$price\_observed_i = \alpha + \beta x_i + \varepsilon_i \, .$$

При этом оцененная модель имеет вид

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2427.821	121.0156	-20.06205	0.0000
X	6.915595	0.126948	54.47591	0.0000
R-squared	0.748337			

С другой стороны, можно проигнорировать наблюдения с $price\_observed_i = 0$ и произвести оценивание той же линейной модели только по таким наблюдениям (в количестве 582). При таком подходе получаем

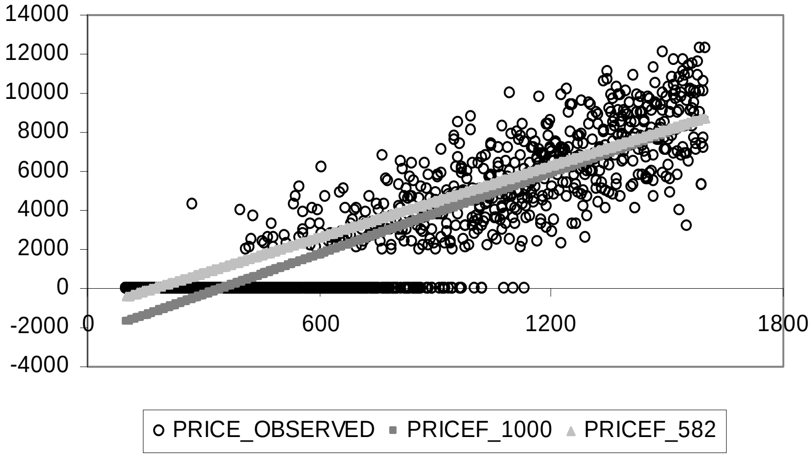
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1037.189	274.4903	-3.778599	0.0002
X	6.119677	0.233812	26.17353	0.0000
R-squared	0.541521	Mean dependent var		5919.670

Следующий график позволяет сравнить значения $price\_observed_i$, прогнозные значения, получаемые по первой модели (по 1000 наблюдениям), т.е.

$$pricef\_1000_i = \hat{\alpha} + \hat{\beta}\, x_i = \text{-}2427.821 + 6.915595\, x_i \, ,$$

и прогнозные значения, получаемые по второй модели (по 582 наблюдениям), т.е.

$$pricef\_582_i = \hat{\alpha} + \hat{\beta}\, x_i = \text{-}1037.189 + 6.119677\, x_i \, .$$



Конечно, имея такую картину, мы вряд ли можем говорить об адекватном представлении данных этими двумя моделями. Желательно было бы построить модель процесса, который мог породить такого рода данные. Для этой цели можно опять использовать идею латентной переменной, но в данной ситуации скорее следовало бы говорить о <u>частично наблюдаемой</u> переменной.

Обращаясь к той же выборке, состоящей из 1000 семей, рассмотрим линейную модель наблюдений

$$price_i^* = \alpha + \beta x_i + \sigma \varepsilon_i\,, \qquad i = 1, \ldots, n\,,$$

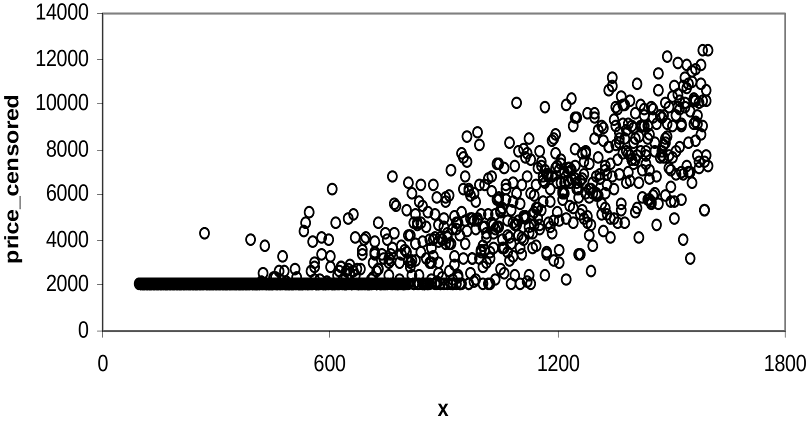
в которой $price_i^*$ – цена, которую <u>уплатила</u> за покупку автомобиля (автомобилей) i-я семья, если эта семья имеет автомобиль, или цена, которую <u>уплатила бы</u> за покупку автомобиля i-я семья, не имеющая автомобиля, если бы эта семья решила приобрести автомобиль. Естественно предполагать, что при этом $\beta > 0$, так что возрастание x_i приводит <u>в среднем</u> к возрастанию $price_i^*$. Однако существенное влияние других ненаблюдаемых факторов, объединяемых в случайную составляющую, может приводить к значительным отклонениям переменной $price_i^*$ от “средней линии” $price^* = \alpha + \beta x$. Возможные отрицательные значения $price_i^*$ свидетельствуют о наличии факторов, в той или иной степени препятствующих планированию каких бы то ни было расходов на покупку автомобиля.

Предположим теперь, что *i*-я семья покупает автомобиль по цене $price_i^*$, если последняя превышает минимально возможную стоимость γ автомобиля на рынке (первичном и вторичном), т.е. если $price_i^* > \gamma$.

В такой модели наблюдений значения переменной $price_i^*$ наблюдаются лишь для части наблюдений – только для семей, имеющих автомобиль. Для остальных семей известно только, что $price_i^* \le \gamma$. Такие данные называют ***цензурированными*** (в данном случае данные ***цензурированы слева на уровне*** γ), а саму модель получения этих данных называют ***цензурированной линейной моделью***. При этом мы наблюдаем цензурированную переменную

$$price\_censored_i = \begin{cases} price_i^*, & \text{если } price_i^* > \gamma, \\ \gamma, & \text{если } price_i^* \le \gamma. \end{cases}$$

В нашем примере диаграмма рассеяния переменных $x_i, price\_censored_i$ принимает вид



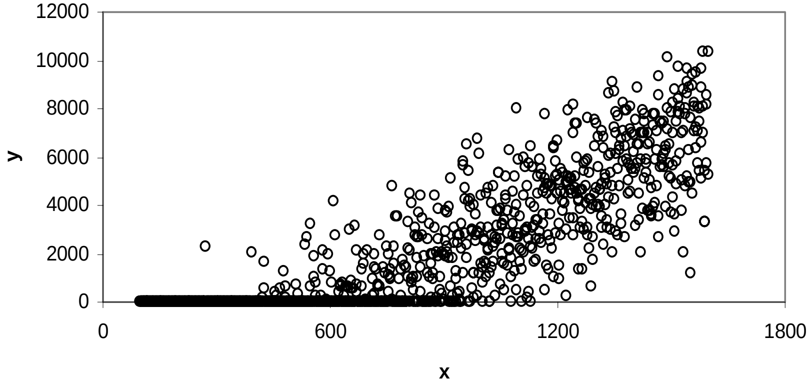
Если значение γ <u>известно</u>, то вместо переменной $price_i^*$ можно рассмотреть переменную

$$y_i = price_i^* - \gamma \,.$$

Значения последней также наблюдаются только для семей, имеющих автомобиль. Для остальных семей положим $y_i = 0$, так что

$$y_i = \begin{cases} price_i^* - \gamma, & \text{если } price_i^* > \gamma, \\ 0, & \text{если } price_i^* \le \gamma. \end{cases}$$

Диаграмма рассеяния переменных x_i, y_i в нашем примере имеет вид



Теперь мы можем поставить вопрос о подходящем методе оценивания параметров цензурированных линейных моделей.

Обычно при рассмотрении подобных ситуаций опираются на предположение ***нормальности распределения ошибок*** ε_i. (Впрочем, имеющиеся пакеты статистических программ позволяют проводить статистический анализ и для других распределений ошибок. Например, в пакете EVIEWS допускается использование вместо нормального распределения ошибок логистического распределения и распределения экстремальных значений первого типа.)

Будем предполагать, что мы имеем дело с некоторым показателем y_i^*, значения которого наблюдаются только при условии $y_i^* > 0$ (в нашем примере в качестве такого показателя выступала переменная $price_i^* - 2000$). Пусть в правую часть модели для этого показателя включаются p объясняющих переменных (показателей, характеризующих i-й субъект), т.е.

$$y_i^* = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + \varepsilon_i, \quad i = 1, \ldots, n,$$

и ошибки $\varepsilon_1,\dots,\varepsilon_n$ – независимые в совокупности (и независимые от $x_{ij}, j=1,\dots,p$) случайные величины, имеющие одинаковое нормальное распределение $\varepsilon_i \sim N\left(0,\sigma^2\right)$. Наблюдаемыми являются значения $x_{ij}, j=1,\dots,p$, $i=1,\dots,n$, и значения переменной y_i,

$$y_i = \begin{cases} y_i^*, & \text{если } y_i^* > 0, \\ 0, & \text{если } y_i^* \le 0. \end{cases}$$

О такой цензурированной модели регрессии говорят как о ***стандартной тобит-модели*** *(tobit model)*.

В стандартной тобит-модели для фиксированных значений $x_{ij}, j=1,\dots,p$, имеем

$$y_i^* \sim N\left(\theta_1 x_{i1} + \cdots + \theta_p x_{ip}, \sigma^2\right),$$

и

$$E\left(y_i^* \middle| x_{ij}, j=1,\dots,p\right) = \theta_1 x_{i1} + \cdots + \theta_p x_{ip},$$

т.е.

$$E\left(y_i^* \middle| x_i\right) = x_i^T \theta,$$

где, как и ранее, обозначено $x_i = (x_{i1},\dots,x_{ip})^T$, $\theta = (\theta_1,\dots,\theta_p)^T$. В нашем примере значение коэффициента θ_j определяет изменение ожидаемой суммы расходов на (возможную) покупку автомобиля для семьи с вектором показателей $x_i = (x_{i1},\dots,x_{ip})^T$ при увеличении на единицу значения j-го показателя.

Если для оценивания коэффициентов θ_j использовать только наблюдения с $y_i > 0$, то получаем ***усеченную модель регрессии***

$$y_i = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + \varepsilon_i, \quad i=1,\dots,n_1,$$

где n_1 – количество семей, имеющих автомобиль (среди всех n рассматриваемых семей). Конечно, при переходе к усеченной модели придется заново перенумеровать используемые n_1 наблюдений. В такой модели для значений $w > 0$ имеем

$$P\{y_i \le w\} = P\{y_i^* \le w | y_i^* > 0\} = \frac{P\{0 < y_i^* \le w\}}{P\{y_i^* > 0\}},$$

где

$$P\{0 < y_i^* \le w\} = P\left\{\frac{-x_i^T\theta}{\sigma} < \frac{y_i^* - x_i^T\theta}{\sigma} \le \frac{w - x_i^T\theta}{\sigma}\right\} =$$

$$= \Phi\left(\frac{w - x_i^T\theta}{\sigma}\right) - \Phi\left(\frac{-x_i^T\theta}{\sigma}\right)$$

и

$$P\{y_i^* > 0\} = 1 - P\left\{\frac{y_i^* - x_i^T\theta}{\sigma} \le \frac{w - x_i^T\theta}{\sigma}\right\} = 1 - \Phi\left(\frac{-x_i^T\theta}{\sigma}\right) = \Phi\left(\frac{x_i^T\theta}{\sigma}\right).$$

Если взять теперь производную $dP\{y_i \le w\}/dw$, то получим функцию плотности распределения случайной величины y_i (условного при заданном x_i):

$$p_{y_i}(w) = \frac{1}{\sigma}\varphi\left(\frac{w - x_i^T\theta}{\sigma}\right) \Big/ \Phi\left(\frac{x_i^T\theta}{\sigma}\right).$$

Отсюда получаем выражение для условного математического ожидания y_i:

$$E(y_i | x_i) = \int_0^\infty w p_{y_i}(w) dw = x_i^T\theta + \sigma\lambda\left(\frac{x_i^T\theta}{\sigma}\right),$$

где обозначено $\lambda(z) = \varphi(z)/\Phi(z)$.

Таким образом, $E(y_i | x_i)$ – <u>нелинейная</u> функция от x_i и φ, причем $E(y_i | x_i) > x_i^T\theta$.

Рассмотрим теперь другой подход к оцениванию коэффициентов исходной модели

$$y_i^* = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + \varepsilon_i\,, \quad i = 1, \ldots, n\,,$$

при котором неполные наблюдения не отбрасываются, а учитываются при оценивании. В рамках этого подхода мы берем в качестве объясняемой переменную

$$y_i = \begin{cases} x_i^T\theta + \varepsilon_i\,, & \text{если} \quad x_i^T\theta + \varepsilon_i > 0 \\ 0, & \text{если} \quad x_i^T\theta + \varepsilon_i \le 0 \end{cases}.$$

В этом случае

$$P\{y_i = 0 |\, x_i\} = P\{\varepsilon_i \le -x_i^T\theta\} = \Phi\left(-\frac{x_i^T\theta}{\sigma}\right) = 1 - \Phi\left(\frac{x_i^T\theta}{\sigma}\right),$$

а для $w > 0$

$$P\{y_i \le w |\, x_i\} = P\left\{\frac{y_i - x_i^T\theta}{\sigma} \le \frac{w - x_i^T\theta}{\sigma}\right\} = \Phi\left(\frac{w - x_i^T\theta}{\sigma}\right).$$

Это приводит к следующему выражению для условного математического ожидания y_i:

$$E(y_i | x_i) = 0 \cdot \left(1 - \Phi\left(\frac{x_i^T\theta}{\sigma}\right)\right) + \int_0^\infty w\varphi\left(\frac{w - x_i^T\theta}{\sigma}\right) dw \cdot \Phi\left(\frac{x_i^T\theta}{\sigma}\right) =$$

$$= \left(x_i^T\theta + \sigma\lambda\left(\frac{x_i^T\theta}{\sigma}\right)\right) \cdot \Phi\left(\frac{x_i^T\theta}{\sigma}\right).$$

Оно отличается от выражения для $E(y_i | x_i)$ в усеченной модели умножением последнего на $\Phi\left(\frac{x_i^T\theta}{\sigma}\right)$, т.е. на величину, меньшую единицы. Раскрывая скобки в правой части, получаем представление

$$E(y_i | x_i) = x_i^T\theta \cdot \Phi\left(\frac{x_i^T\theta}{\sigma}\right) + \sigma\lambda\left(\frac{x_i^T\theta}{\sigma}\right) \cdot \Phi\left(\frac{x_i^T\theta}{\sigma}\right) =$$

$$= x_i^T\theta \cdot \Phi\left(\frac{x_i^T\theta}{\sigma}\right) + \sigma \cdot \varphi\left(\frac{x_i^T\theta}{\sigma}\right).$$

Предельный эффект изменения переменной x_{ij} равен

$$\frac{\partial E(y_i|x_i)}{\partial x_{ij}} = \theta_j \cdot \Phi\left(\frac{x_i^T\theta}{\sigma}\right),$$

т.е. <u>меньше</u> значения коэффициента θ_j в исходной модели: он получается умножением этого коэффициента на вероятность того, что $y_i^* > 0$.

Заметим в связи с этим, что если $\widetilde{E}(y_i|x_i)$ – условное математическое ожидание значения y_i в усеченной модели, то для него

$$\frac{\partial \widetilde{E}(y_i|x_i)}{\partial x_{ij}} = \theta_j \left[1 - z\lambda(z) - \lambda^2(z)\right],$$

где

$$\lambda(z) = \varphi(z)/\Phi(z), \quad z = \frac{x_i^T\theta}{\sigma}.$$

Продолжим рассмотрение смоделированной выборки, состоящей из 1000 семей, 582 из которых имеют автомобиль. Подберем к тем же данным усеченную и цензурированную модели.

Заметим, что если переменная $y_i^* = price_i^* - 2000$ порождается моделью $y_i^* = \alpha + \beta x_i + \varepsilon_i$, $i = 1,\ldots,1000$, то сама переменная $price_i^*$ порождается моделью $price_i^* = (\alpha + 2000) + \beta x_i + \varepsilon_i$. Поэтому достаточно произвести оценивание коэффициентов модели $y_i^* = \alpha + \beta x_i + \varepsilon_i$, опираясь на данные (x_i, y_i). Такое оценивание приводит к следующим результатам.

Усеченная модель:

	Coefficient	Std. Error	z-Statistic	Prob.
C	-5710.678	480.1485	-11.89357	0.0000
X	8.103471	0.376079	21.54728	0.0000
		Error Distribution		
σ	1822.273	66.21537	27.52040	0.0000

Цензурированная модель:

	Coefficient	Std. Error	z-Statistic	Prob.
C	-6041.883	233.5302	-25.87195	0.0000
X	8.363125	0.209276	39.96215	0.0000
		Error Distribution		
σ	1823.565	53.95272	33.79933	0.0000

Это приводит к следующим оцененным моделям для прогноза значений переменной $price_i^*$:

$$price_i^* = -3710.678 + 8.103471\, x_i \quad \text{(усеченная модель)},$$

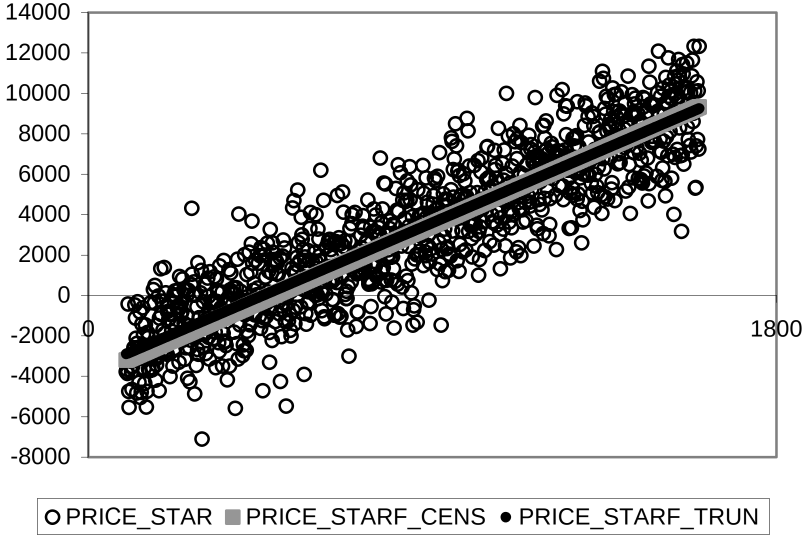
$$price_i^* = -4041.883 + 8.363125\, x_i \quad \text{(цензурированная модель)}.$$

Дисперсии случайных составляющих оцениваются, соответственно, как 1822.273 и 1823.565. Заметим, что “теоретическая” модель, по которой генерировались данные, имела вид

$$price_i^* = -3600 + 8\, x_i + 1800\, u_i\,,$$

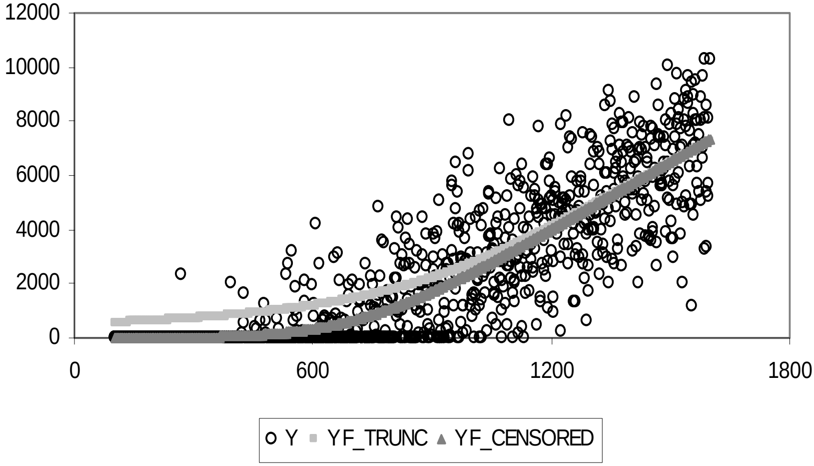
где $u_1, \ldots, u_{1000}$ – независимые случайные величины, имеющие одинаковое стандартное нормальное распределение $N(0,1)$.

На следующем графике для сравнения показаны значения переменной $price_i^*$ и прогнозные значения для этой переменной, полученные по оцененной усеченной модели ($price\_starf\_trun$) и по оцененной цензурированной модели ($price\_starf\_cens$).



Отметим, что прогнозные значения, полученные по двум оцененным моделям, весьма близки.

На следующем графике представлены значения переменной y_i и ожидаемые значения переменной y_i, рассчитанные по двум оцененным моделям.



Отметим, что для значений $x_i \geq 1330$ ожидаемые значения y_i, рассчитанные по цензурированной модели, больше ожидаемых значений y_i, рассчитанных по модели; однако это различие практически незаметно. В то же время, для значений $x_i < 1330$ ожидаемые значения y_i, рассчитанные по цензурированной модели, меньше ожидаемых значений y_i, рассчитанных по усеченной модели, причем это различие становится весьма заметным при уменьшении значений x_i.

Заметим еще, что ожидаемые значения y_i, рассчитанные и по усеченной и по цензурированной модели, положительны для всех 1000 наблюдений, тогда как это не выполняется для линейных моделей, подобранных методом наименьших квадратов

Так, оценивание обычным методом наименьших квадратов модели $y_i = \alpha + \beta x_i + \varepsilon_i$ по всем 1000 наблюдениям дает следующую картину:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2075.806	104.7679	-19.81338	0.0000
X	5.130473	0.109904	46.68158	0.0000

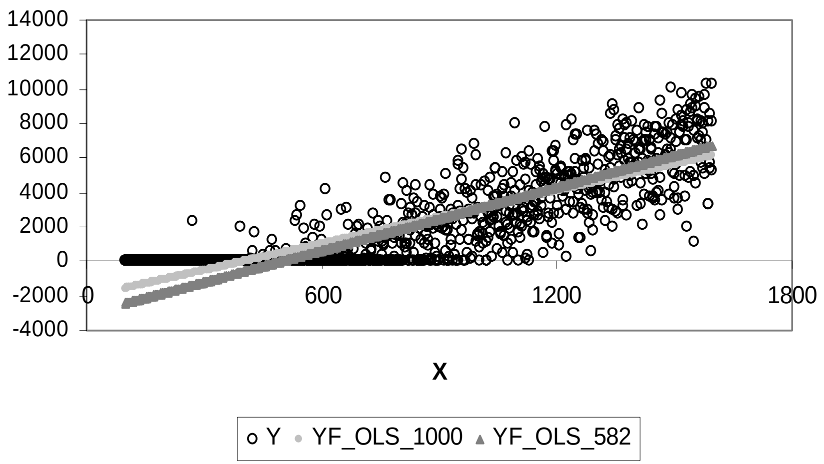
Для значений $x_i \le 470$ подобранная модель прогнозирует отрицательные значения объясняемой переменной.

При подгонке такой модели методом наименьших квадратов по 582 наблюдениям получаем:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3037.189	274.4903	-11.06483	0.0000
X	6.119677	0.233812	26.17353	0.0000

Оцененная модель прогнозирует отрицательные значения объясняемой переменной для значений $x_i \le 498$.

Это положение иллюстрирует следующий график:



Одним из показателей качества прогноза произвольного временного ряда z_i, $i = 1, \ldots, n$, является ***средняя абсолютная процентная ошибка (MAPE –** mean squared absolute error**)***, определяемая следующим образом. Если $\hat{z}_i$ – прогнозное значение для z_i, то

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} 100\left|\frac{z_i - \hat{z}_i}{z_i}\right|.$$

Сравним качество полученных альтернативных прогнозов для y_i с точки зрения средней абсолютной процентной ошибки.

Модель	OLS_582	OLS_1000	Truncated	Censored
MAPE %	118.46	99.86	126.69	71.96

Как видно из этой таблицы, наилучшее качество имеют прогнозы, полученные с использованием цензурированной модели регрессии.

Обратим внимание на еще одно обстоятельство. Мы уже отмечали, что

$$E(y_i|x_i) = \left(x_i^T\theta + \sigma\lambda\left(\frac{x_i^T\theta}{\sigma}\right)\right)\cdot\Phi\left(\frac{x_i^T\theta}{\sigma}\right) = \widetilde{E}(y_i|x_i)\cdot\Phi\left(\frac{x_i^T\theta}{\sigma}\right),$$

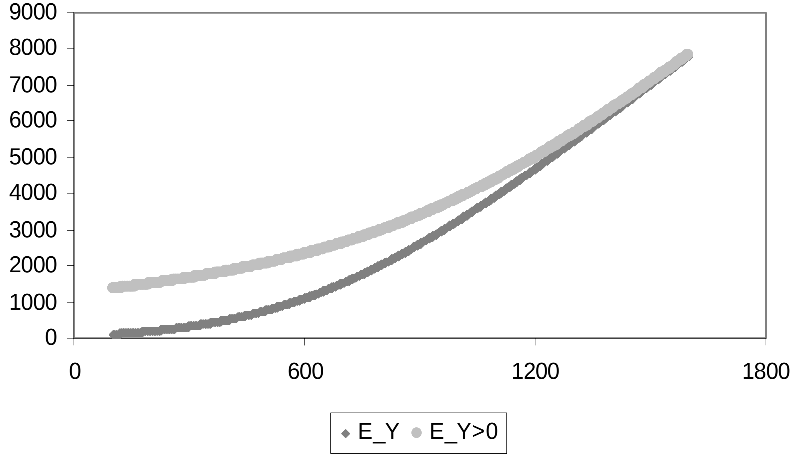
где $\widetilde{E}(y_i|x_i)$ – условное математическое ожидание значения y_i в усеченной модели. Отсюда мы получаем следующее разложение:

$$\frac{\partial E(y_i|x_i)}{\partial x_{ij}} = \Phi\left(\frac{x_i^T\theta}{\sigma}\right)\cdot\frac{\partial\widetilde{E}(y_i|x_i)}{\partial x_{ij}} + \widetilde{E}(y_i|x_i)\cdot\frac{\partial\Phi\left(\frac{x_i^T\theta}{\sigma}\right)}{\partial x_{ij}}.$$

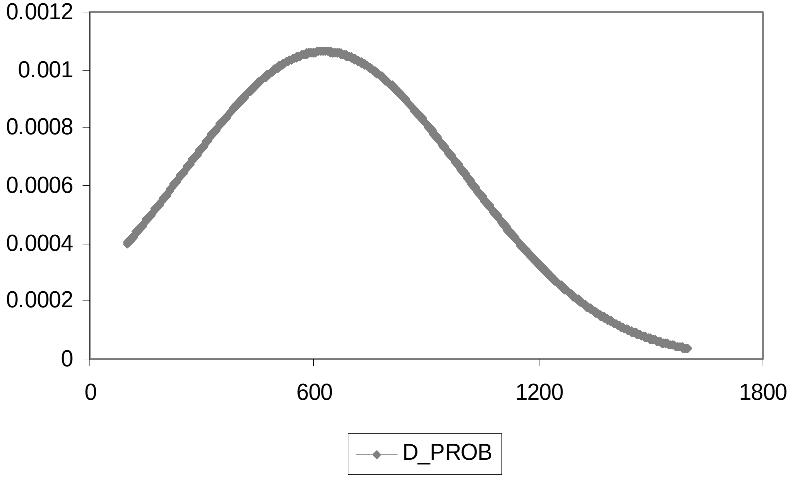
Первое слагаемое отражает изменение в ожидаемых значениях $y_i > 0$, взвешенное с весом $\Phi\left(\frac{x_i^T\theta}{\sigma}\right) = P\{y_i > 0\}$, а второе – изменение вероятности $P\{y_i > 0\}$, взвешенное с весом, равным $\widetilde{E}(y_i|x_i)$. Заметим в этой связи, что

$$\frac{\partial P\{y_i > 0\}}{\partial x_{ij}} = \frac{\partial \Phi\left(\dfrac{x_i^T\theta}{\sigma}\right)}{\partial x_{ij}} = \frac{1}{\sigma}\varphi\left(\frac{x_i^T\theta}{\sigma}\right)\cdot\theta_j\,.$$

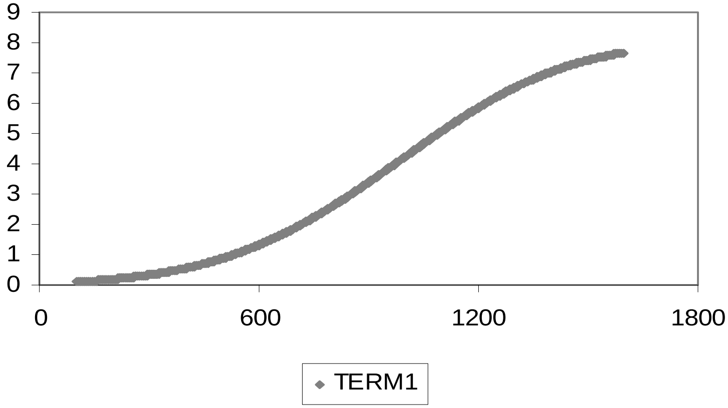
В нашем примере $\widetilde{E}(y_i|x_i)$ изменяется следующим образом (по оси абсцисс на этом и на следующих 5 графиках откладываются значения среднемесячного дохода на одного члена семьи):

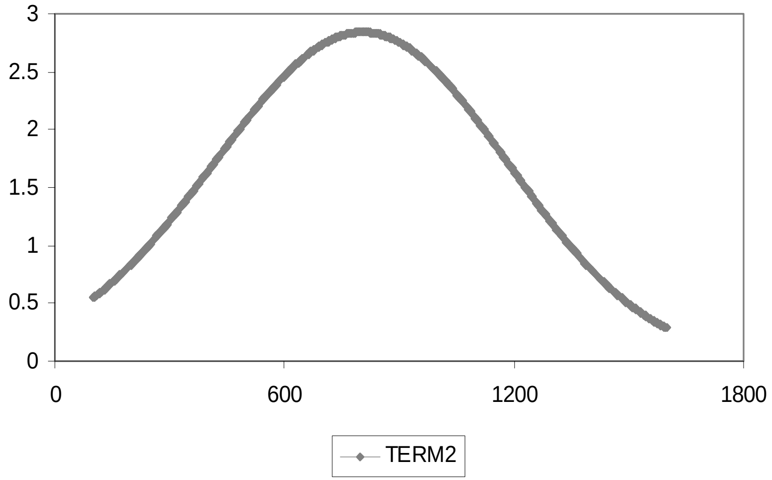


Производная $\dfrac{\partial P\{y_i > 0\}}{\partial x_{ij}}$ изменяется следующим образом:

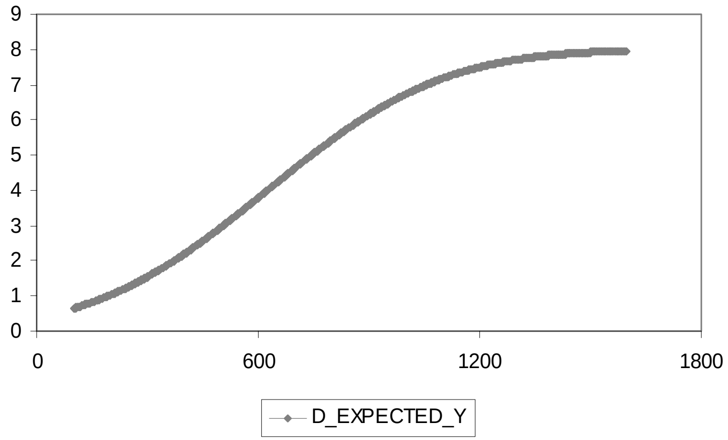


Входящие в разложение для $\dfrac{\partial E\left(y_i \middle| x_i\right)}{\partial x_{ij}}$ слагаемые имеют вид:

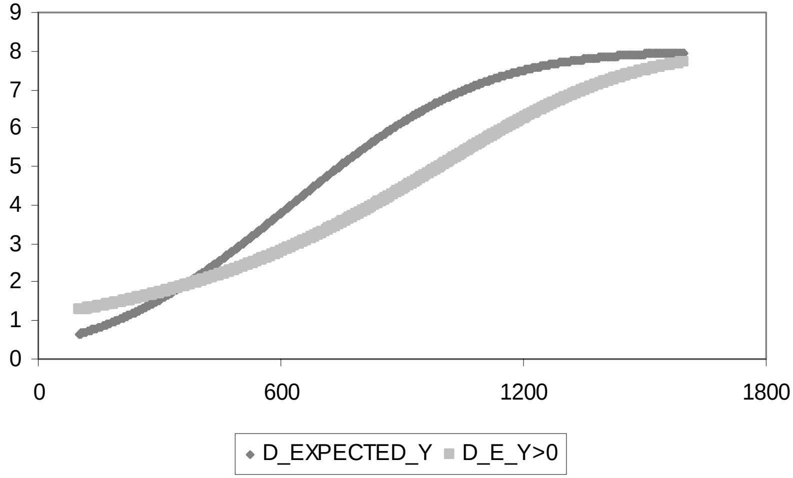




В сумме они дают функцию $\dfrac{\partial E(y_i | x_i)}{\partial x_{ij}} = \theta_j \cdot \Phi\left(\dfrac{x_i^T \theta}{\sigma}\right)$:



Следующий график позволяет сравнить влияние единичного возрастания дохода на ожидаемые значения y_i во всей популяции (D_EXPECTED_Y) и среди семей с $y_i > 0$ (D_E_Y>0).



1.8. Модель Тобит-II

В предыдущем разделе мы рассмотрели линейную модель наблюдений

$$price_i^* = \alpha + \beta x_i + \sigma \varepsilon_i\,, \quad i = 1, \ldots, n\,,$$

в которой $price_i^*$ – цена, которую уплатила за покупку автомобиля (автомобилей) *i*-я семья, если эта семья имеет автомобиль, или цена, которую уплатила бы за покупку автомобиля *i*-я семья, не имеющая автомобиля, если бы эта семья решила приобрести автомобиль. При этом мы предполагали, что i-я семья покупает автомобиль по цене $price_i^*$, если $price_i^* > \gamma$. Таким образом, в этой модели решение о приобретении или неприобретении собственного автомобиля определяется самой ценой, по которой предполагается приобрести автомобиль. В то же время мы могли бы рассмотреть и другую модель, в которой процесс принятия решения о стоимости покупаемого автомобиля отделен от процесса принятия решения о покупке автомобиля.

Пусть мы имеем дело с некоторым показателем y_i^*, значения которого наблюдаются не для всех i. Значение y_i^* наблюдается, если выполнено условие $h_i^* > 0$, где h_i^* – некоторая функция полезности. Мы будем предполагать, что

$$y_i^* = x_{1i}^T \theta_1 + \varepsilon_{1i}, \quad i = 1, \dots, n,$$

$$h_i^* = x_{2i}^T \theta_2 + \varepsilon_{2i}, \quad i = 1, \dots, n,$$

где

$x_{1i} = \left(x_{11,i}, \dots, x_{1p_1,i}\right)^T$ – вектор значений p_1 объясняющих переменных в уравнении для y_i^*,

$\theta_1 = \left(\theta_{11}, \dots, \theta_{1p_1}\right)^T$ – вектор коэффициентов при этих переменных,

$x_{2i} = \left(x_{21,i}, \dots, x_{2p_2,i}\right)^T$ – вектор значений p_2 объясняющих переменных в уравнении для h_i^*,

$\theta_2 = \left(\theta_{21}, \dots, \theta_{2p_2}\right)^T$ – вектор коэффициентов при этих переменных.

Случайные составляющие ε_{1i} и ε_{2i} могут быть коррелированными, так что $Cov(\varepsilon_{1i}, \varepsilon_{2i}) \neq 0$. Следуя обычной практике, мы будем предполагать, что двумерные случайные векторы $(\varepsilon_{1i},\varepsilon_{2i})^T$, $i=1,\ldots,n$, независимы в совокупности и имеют одинаковое двумерное нормальное распределение $N_2(0,\Sigma)$ с нулевым вектором математических ожиданий и ковариационной матрицей

$$\Sigma=\begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix},$$

т.е.

$$\begin{pmatrix}\varepsilon_{1i}\\ \varepsilon_{2i}\end{pmatrix} \sim \ i.i.d.\ N_2\left(0,\begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}\right).$$

Для нормализации функции полезности полагаем

$$\sigma_2=1.$$

Наблюдаемыми являются

- значения объясняющих переменных $x_{1j,i},\, x_{2j,i},\, j=1,\ldots,p$, $i=1,\ldots,n$;
- значения переменной h_i,

$$h_i=\begin{cases}1, & \text{если } h_i^*>0,\\ 0, & \text{если } h_i^*\le 0;\end{cases}$$

- значения переменной y_i,

$$y_i=\begin{cases}y_i^*, & \text{если } h_i=1,\\ 0, & \text{если } h_i=0.\end{cases}$$

Определенную таким образом модель называют ***стандартной Тобит-II моделью***. Соответственно, о модели рассмотренной в предыдущем разделе, в этом контексте говорят как о ***стандартной Тобит-I модели***.

Замечание

Объясняющие переменные в уравнениях для y_i^* и h_i^* могут быть как одинаковыми, так и различными. В ряде ситуаций экономическая аргументация указывает на необходимость включения в правую часть уравнения для h_i^* (уравнение выбора) <u>всех</u> переменных, включенных в правую часть уравнения для y_i^*. При этом коэффициенты при одной и той же переменной в уравнениях для y_i^* и h_i^* могут быть различными.

Если предположить, что $x_{1i}^T\theta_1 = x_{2i}^T\theta_2$ и $\varepsilon_{1i} = \varepsilon_{2i}$, то мы возвращаемся к стандартной Тобит-модели, рассмотренной в предыдущем разделе (модель ***Тобит-I***).

Обращаясь опять к примеру с автомобилями, мы могли бы расмотреть, например, модели, в которых значение $price_i^*$ определяется по той же формуле

$$price_i^* = \alpha + \beta x_i + \sigma \varepsilon_i, \quad i = 1, \ldots, n,$$

но наличие автомобиля соответствует выполнению соотношения

$$h_i^* > 0,$$

в котором

$$h_i^* = \gamma + \delta x_i + u_i,$$

или, например,

$$h_i^* = \gamma + \delta x_i + \kappa d_{man} + u_i,$$

где $d_{man} = 1$, если главой семьи является мужчина, и $d_{man} = 0$, если главой семьи является женщина.

Прежде всего заметим, что (при фиксированных значениях x_{1i}, x_{2i})

$$E\{y_i | h_i = 1\} = x_{1i}^T\theta_1 + E\{\varepsilon_{1i} | h_i = 1\} = x_{1i}^T\theta_1 + E\{\varepsilon_{1i} | \varepsilon_{2i} > -x_{2i}^T\theta_2\} =$$

$$= x_{1i}^T\theta_1 + \frac{\sigma_{12}}{\sigma_2^2} E\left\{\varepsilon_{1i} \middle| \varepsilon_{2i} > -x_{2i}^T\theta_2\right\} = x_{1i}^T\theta_1 + \sigma_{12}\lambda\left(x_{2i}^T\theta_2\right),$$

где, как и ранее,

$$\lambda(z) = \varphi(z)/\Phi(z).$$

Если $\sigma_{12} = 0$, то

$$E\{y_i | h_i = 1\} = x_{1i}^T\theta_1 .$$

Это означает, что если ε_{1i} и ε_{2i} не коррелированы, то можно, игнорируя уравнение для h_i^*, производить непосредственное оценивание уравнения регрессии

$$y_i = x_{1i}^T\theta_1 + \varepsilon_{1i}$$

методом наименьших квадратов по наблюдениям с $h_i = 1$. Это приводит к состоятельному оцениванию значений $x_{1i}^T\theta_1$.

Однако если $\sigma_{12} \neq 0$, то при таком оценивании возникает смещение оценки $x_{1i}^T\theta_1$, пропорциональное величине $\lambda\left(x_{2i}^T\theta_2\right)$, которую называют в этом контексте ***лямбдой Хекмана***.

Получить состоятельные и асимптотически эффективные оценки параметров модели Тобит-II можно, используя метод максимального правдоподобия, при котором соответствующая функция правдоподобия максимизируется по всем возможным значениям параметров модели $\theta_1, \theta_2, \sigma_1, \sigma_{12}$. Однако чаще такую модель оценивают, используя ***двухшаговую процедуру Хекмана***. Она проста в вычислительном отношении и дает хорошие стартовые значения для итерационной процедуры максимизации функции правдоподобия.

Идея Хекмана состоит в использовании уже приводившегося выше соотношения

$$E\{y_i | h_i = 1\} = x_{1i}^T\theta_1 + \sigma_{12}\lambda\left(x_{2i}^T\theta_2\right)$$

и построения на его основе модели регрессии

$$y_i = x_{1i}^T\theta_1 + \sigma_{12}\lambda_i + \nu_i,$$

где λ_i – переменная, определяемая соотношением

$$\lambda_i = \lambda\left(x_{2i}^T\theta_2\right) = \varphi\left(x_{2i}^T\theta_2\right)/\Phi\left(x_{2i}^T\theta_2\right).$$

Если ε_{1i} не коррелирована с x_{1i} и x_{2i}, то ν_i не коррелирована с x_{1i} и λ_i, так что эту модель регрессии можно оценивать методом наименьших квадратов. Проблема, однако, в том, что значения λ_i не наблюдаются, поскольку неизвестен вектор коэффициентов θ_2 в модели выбора.

Оценивание вектора θ_2 производится в рамках пробит-модели бинарного выбора. При этом получаем оцененные значения $\hat{\lambda}_i = \lambda\left(x_{2i}^T\hat{\theta}_2\right)$ (первый шаг процедуры Хекмана). Эти оцененные значения используются затем на втором шаге процедуры вместо λ_i. Модель $y_i = x_{1i}^T\theta_1 + \sigma_{12}\hat{\lambda}_i + \nu_i$ оценивается методом наименьших квадратов; в результате получаем состоятельные (хотя и не эффективные) оценки для θ_1 и σ_{12}. Используя эти оценки, мы получаем оцененное ожидаемое значение y_i при заданных x_{1i}, x_{2i} и $h_i = 1$ в виде

$$\hat{E}\{y_i | x_{1i}, x_{2i}, h_i = 1\} = x_{1i}^T\hat{\theta}_1 + \hat{\sigma}_{12}\lambda\left(x_{2i}^T\hat{\theta}_2\right).$$

Если же нас интересует ожидаемое значение y_i при заданных x_{1i}, x_{2i} <u>без</u> условия $h_i = 1$, то оно оценивается величиной

$$\hat{E}\{y_i | x_{1i}, x_{2i}\} = x_{1i}^T\hat{\theta}_1.$$

Поскольку смещение при оценивании уравнения для y_i^* методом наименьших квадратов вызывается коррелированностью ε_{1i} и ε_{2i}, представляет интерес проверка гипотезы

$$H_0:\ \sigma_{12} = 0$$

об отсутствии такой коррелированности в рамках модели, оцененной на втором шаге. Отметим только, что при проверке этой гипотезы следует производить коррекцию значений стандартных ошибок оценок, учитывающую гетероскедастичность модели и тот факт, что вместо переменной λ_i на втором шаге используется предварительно оцененная переменная $\hat{\lambda}_i$.

Заметим, наконец, что в описанной выше стандартной Тобит-II модели функция правдоподобия имеет вид

$$L(\theta_1,\theta_2,\sigma_1,\sigma_{12}) = \prod_{i=1}^{n} \left(P\{h_i = 0\}\right)^{1-h_i} \left(P\{h_i = 1\} \cdot f\left(y_i \middle| h_i = 1\right)\right)^{h_i},$$

где $f\left(y_i \middle| h_i = 1\right)$ – условная плотность распределения случайной величины y_i при $h_i = 1$. Здесь

$$P\{h_i = 0\} = 1 - \Phi\left(x_{2i}^T\theta_2\right),$$

$$P\{h_i = 1\} \cdot f\left(y_i \middle| h_i = 1\right) = P\{h_i = 1 | y_i\} \cdot f(y_i),$$

$$P\{h_i = 1 | y_i\} = \Phi\left(\frac{x_{2i}^T\theta_2 + \left(\sigma_{12}/\sigma_1^2\right)\cdot\left(y_i - x_{1i}^T\theta_1\right)}{\sqrt{1 - \sigma_{12}/\sigma_1^2}}\right),$$

$$f(y_i) = \frac{1}{\sigma_1\sqrt{2\pi}} \exp\left(-\frac{\left(y_i - x_{1i}^T\theta_1\right)^2}{2\sigma_1^2}\right).$$

Для начала итерационной процедуры в качестве стартовых можно взять значения оценок параметров, полученные в процессе реализации двухшаговой процедуры Хекмана.

Пример 1

Пусть в примере с автомобилями наличие у семьи собственного автомобиля определяется условием $w_i^* > 2000$, где

$$w_i^* = -3600 + 8\,x_i + 1800\,\varepsilon_{2i}\,,\ \ \varepsilon_{21},\ldots,\varepsilon_{2,1000} \sim\ i.i.d.\, N(0{,}1)\,.$$

Обозначив $h_i^* = w_i^* - 2000$, запишем это условие в виде $h_i^* > 0$, где

$$h_i^* = -5600 + 8\,x_i + 1800\,\varepsilon_{2i}\,,$$

и нормализуем функцию полезности, разделив обе части последнего равенства на 1800:

$$h_i^* = -3.111 + 0.00445\,x_i + \varepsilon_{2i}\,.$$

Пусть “потенциальная цена” автомобиля для *i*-й семьи определяется уравнением

$$price_i^* = 4000 + 6x_i + \varepsilon_{1i}\,,\ \ \varepsilon_{11},\ldots,\varepsilon_{1,1000} \sim\ i.i.d.\, N(0{,}1000^2)\,.$$

В смоделированной выборке пары $(\varepsilon_{11},\varepsilon_{21}),\ldots,(\varepsilon_{1,1000},\varepsilon_{2,1000})$ взаимно независимы, но $Cov(\varepsilon_{1i},\varepsilon_{2i}) = 707$, так что коэффициент корреляции случайных величин $\varepsilon_{1i},\varepsilon_{2i}$ равен $\rho_{12} = 0.707$.

В принятых выше общих обозначениях модели Тобит-II получаем:

$$y_i^* = \theta_{11}x_{11,i} + \theta_{12}x_{12,i} + \varepsilon_{1i}\,,\ \ h_i^* = \theta_{21}x_{21,i} + \theta_{22}\,x_{22,i} + \varepsilon_{2i}\,,$$

где $x_{11,i} = x_{21,i} = 1$, $x_{12,i} = x_{22,i} = x_i$, $\theta_{11} = 4000$, $\theta_{12} = 6$, $\theta_{21} = -3.111$, $\theta_{22} = 0.00445$; при этом $\sigma_1 = 1000,\ \sigma_2 = 1,\ \sigma_{12} = 707.$

Применяя к смоделированным данным двухшаговую процедуру Хекмана, получаем на первом шаге оцененное уравнение

$$h_i^* = -3.450 + 0.00476\,x_i\,,$$

а на втором шаге – оцененное уравнение

$$price_i^* = 3936.2 + 5.995x_i\,.$$

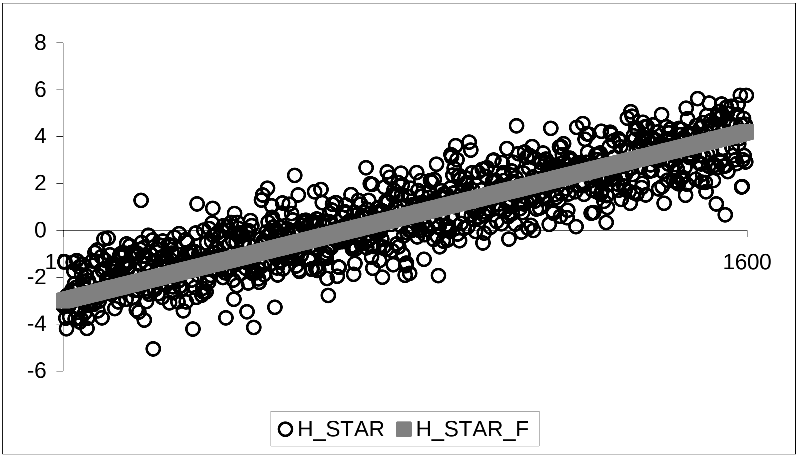
Используя полученные оценки параметров в качестве стартовых значений итерационной процедуры максимального правдоподобия, приходим к уравнениям

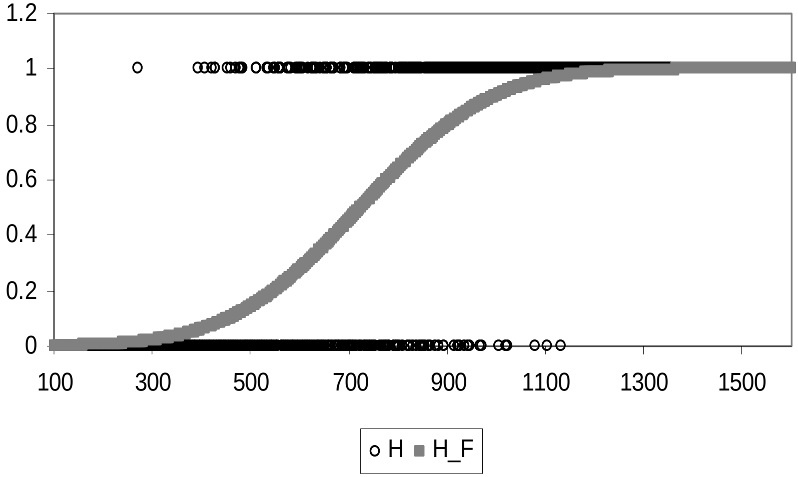
$$h_i^* = -3.483 + 0.00480\,x_i\,,$$

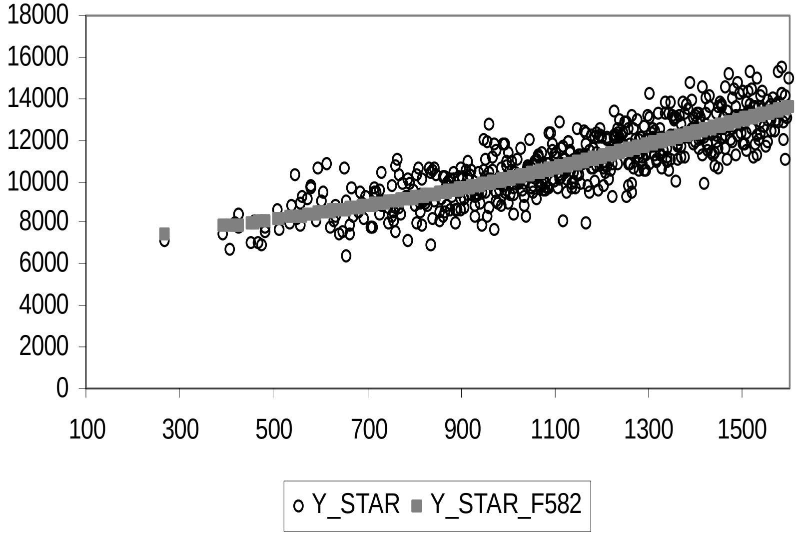
$$price_i^* = 4159.3 + 5.828x_i\,.$$

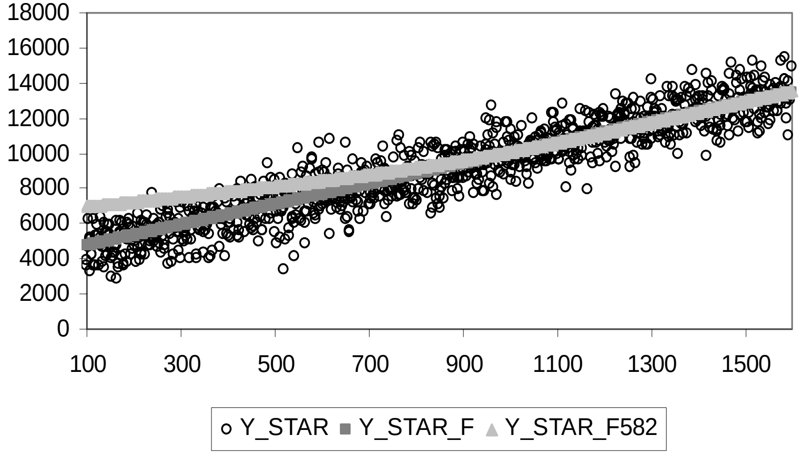
При этом получаем также $\hat{\sigma}_1 = 1010.7,\ \hat{\rho}_{12} = 0.598$.

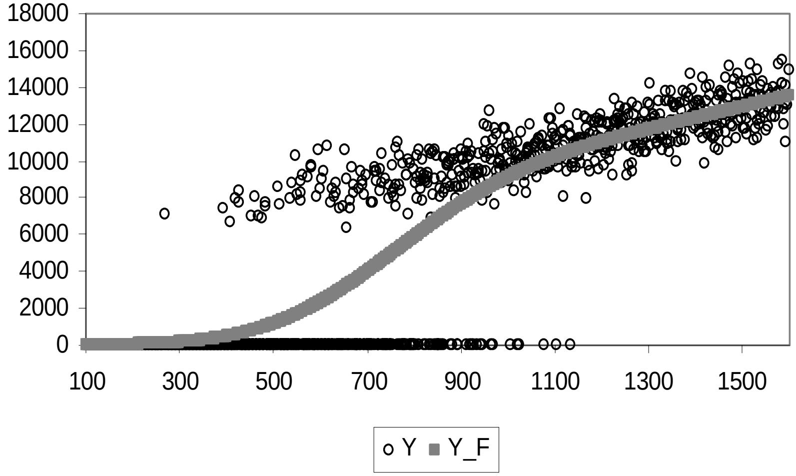
Как видим, оцененные значения параметров достаточно близки к значениям, при которых производилось порождение данных. Приведем теперь графики, иллюстрирующие полученные результаты.




1.2
1
0.8
0.6
0.4
0.2
0
100
300
500
700
900
1100
1300
1500
H
H_F


18000
16000
14000
12000
10000
8000
6000
4000
2000
0
100
300
500
700
900
1100
1300
1500
Y_STAR
Y_STAR_F582


18000
16000
14000
12000
10000
8000
6000
4000
2000
0
100
300
500
700
900
1100
1300
1500
Y_STAR
Y_STAR_F
Y_STAR_F582


18000
16000
14000
12000
10000
8000
6000
4000
2000
0
100
300
500
700
900
1100
1300
1500
Y
Y_F

П р и м е р 2

В условиях Примера 1 перемоделируем данные с измененной функцией полезности, полагая теперь

$$h_i^* = -4 + 0.003\,x_i + 2(d_{man})_i + \varepsilon_{2i}\,,$$

где $d_{man} = 1$, если главой семьи является мужчина, и $d_{man} = 0$, если главой семьи является женщина.

Применяя к новым смоделированным данным двухшаговую процедуру Хекмана, получаем на первом шаге оцененное уравнение

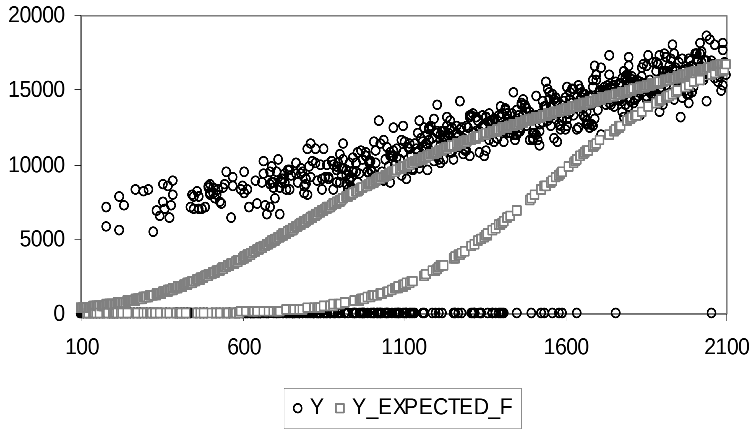
$$h_i^* = -4.280 + 0.00297\,x_i + 2.347\left(d_{man}\right)_i,$$

а на втором шаге – оцененное уравнение

$$price_i^* = 3879.97 + 6.124 x_i\,.$$

При этом получаем также $\hat{\sigma}_1 = 984.2,\ \hat{\rho}_{12} = 0.643$.

Как видим, и здесь оцененные значения параметров достаточно близки к значениям, при которых производилось порождение данных. Приведем для сравнения наблюдаемые значения переменной y_i и оцененные ожидаемые значения этой переменной (Y_EXPECTED_F).



Обратим внимание на две ветви графика оцененных ожидаемых значений y_i. Верхняя ветвь соответствует семьям, которые возглавляют мужчины, а нижняя – семьям, которые возглавляют женщины.

Глава 2. Инструментальные переменные. Системы одновременных уравнений

2.1. Проблема коррелированности случайных ошибок с объясняющими переменными

В главе 1 мы встретили модели наблюдений, в которых естественным образом возникла необходимость использования вместо метода наименьших квадратов другого метода оценивания – метода максимального правдоподобия. (В классической линейной модели с независимыми, нормальными, одинаково распределенными ошибками эти методы совпадают.)

Теперь мы рассмотрим некоторые ситуации, приводящие к еще одному популярному методу оценивания – ***методу инструментальных переменных***. Общим для такого рода ситуаций является наличие коррелированности одной или нескольких объясняющих переменных со случайной ошибкой, входящей в правую часть уравнения. Поскольку случайные ошибки отражают наличие неучтенных факторов, не включенных в уравнение в качестве объясняющих переменных, указанная коррелированность фактически означает наличие корреляции между некоторыми учтенными и неучтенными факторами.

В матрично-векторной форме классическая нормальная линейная модель наблюдений имеет вид

$$y = X\theta + \varepsilon,$$

где

$y = (y_1, y_2, \ldots, y_n)^T$ – вектор-столбец значений объясняемой переменной в n наблюдениях,

X – $(n \times p)$-матрица значений объясняющих переменных в n наблюдениях, $n > p$,

$\theta = (\theta_1, \theta_2, \ldots, \theta_p)^T$ – вектор-столбец коэффициентов,

$\varepsilon = (\varepsilon_1, \varepsilon_2 \ldots, \varepsilon_n)^T$ – вектор-столбец случайных ошибок (возмущений) в n наблюдениях, причем случайный вектор ε имеет n-мерное нормальное распределение с нулевым вектором математических ожиданий

$E(\varepsilon) = (E(\varepsilon_1),\ E(\varepsilon_2),\ \ldots,\ E(\varepsilon_n))^T = (0,\ 0,\ ...,\ 0)^T$ (в краткой записи: $E(\varepsilon) = 0$)

и ковариационной матрицей

$Var(\varepsilon) = (Cov(\varepsilon_i\ ,\ \varepsilon_j)) = \sigma^2\ I_n$,

где I_n – единичная матрица (размера $n \times n$).

Здесь

$Cov(\varepsilon_i\ ,\ \varepsilon_j) = E(\varepsilon_i - E(\varepsilon_i))(\varepsilon_j - E(\varepsilon_j))$ –

ковариация случайных величин ε_i и ε_j .

Предположение о фиксированности значений объясняющих переменных в совокупности со стандартными предположениями об ошибках удобно с чисто математической точки зрения: при таких предположениях оценки параметров, получаемые методом наименьших квадратов, имеют нормальное распределение, что, в свою очередь, дает возможность:

- строить доверительные интервалы для коэффициентов линейной модели, используя квантили t-распределения Стьюдента;
- проверять гипотезы о значениях отдельных коэффициентов, используя квантили t-распределения Стьюдента;
- проверять гипотезы о выполнении тех или иных линейных ограничений на коэффициенты модели, используя квантили F-распределения Фишера;
- строить интервальные прогнозы для “будущих” значений объясняемой переменной, соответствующих заданным будущим значениям объясняющих переменных.

Вместе с тем используемое в классической модели предположение о фиксированности значений объясняющих переменных в n наблюдениях фактически означает, что мы можем повторить наблюдения значений объясняемой переменной при том же наборе значений объясняющих переменных $x_{i1},\dots,x_{ip}$, $i=1,\dots,n$; при этом мы получим другую реализацию (другой набор значений) случайных составляющих ε_i, $i=1,\dots,n$, что приведет к значениям объясняемой переменной, отличающимся от значений $y_1,\dots,y_n$, наблюдавшихся ранее.

С точки зрения моделирования реальных экономических явлений, предположение о фиксированности значений объясняющих пременных можно считать реалистическим лишь в отдельных ситуациях, связанных с проведением контролируемого эксперимента. Между тем в реальных ситуациях по большей части нет возможности сохранять неизменными значения объясняющих переменных. Более того, и сами наблюдаемые значения объясняющих переменных (как и “ошибки”) часто интерпретируются как реализации некоторых случайных величин. В таких ситуациях становится проблематичным использование техники статистических выводов, разработанной для классической нормальной линейной модели.

Поясним последнее, обратившись к матрично-векторной форме классической линейной модели с p объясняющими переменными

$$y = X\theta + \varepsilon$$

и не требуя нормальности распределения вектора ε.

Если матрица X имеет полный ранг p, то матрица X^TX является невырожденной, для нее существует обратная матрица $(X^TX)^{-1}$, и оценка наименьших квадратов для вектора θ неизвестных коэффициентов имеет вид

$$\hat{\theta} = (X^TX)^{-1}X^Ty.$$

Математическое ожидание вектора оценок коэффициентов равно

$$E(\hat{\theta}) = E\,((X^TX)^{-1}X^T(X\theta + \varepsilon)) = E\,((X^TX)^{-1}X^TX\theta) + E\,((X^TX)^{-1}X^T\varepsilon) =$$
$$=\theta + E\,((X^TX)^{-1}X^T\varepsilon).$$

Если матрица X фиксирована, то тогда

$$E\,((X^TX)^{-1}X^T\varepsilon) = (X^TX)^{-1}X^T\,E\,(\varepsilon) = 0,$$

так что $E(\hat{\theta}) = \theta$, т.е. $\hat{\theta}$ – несмещенная оценка для θ.

Если же мы имеем дело со ***стохастическими (случайными, недетерминированными)*** объясняющими переменными, то в общем случае $E\,((X^TX)^{-1}X^T\varepsilon) \neq 0$, так что

$$E(\hat{\theta}) \neq \theta,$$

и $\hat{\theta}$ – смещенная оценка для θ. Кроме того, эта оценка уже не имеет нормального распределения даже если вектор ε имеет нормальное распределение.

Если объясняющие переменные стохастические, то в некоторых случаях все же остается возможным использовать стандартную технику статистических выводов, предназначенную для классической нормальной линейной модели, по крайней мере, в асимптотическом плане (при большом количестве наблюдений).

В этом отношении наиболее благоприятной является

Ситуация А

- случайная величина ε_i не зависит (статистически) от $x_{k1}, \ldots, x_{kp}$ при всех i и k;
- $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ являются независимыми случайными величинами, имеющими одинаковое нормальное распределение с нулевым математическим ожиданием и конечной дисперсией $\sigma^2 > 0$. (Как и ранее, мы кратко обозначаем это как $\varepsilon_i \sim i.i.d.\ N(0, \sigma^2)$. Здесь ***i.i.d.*** – *independent identically distributed*.)

При выполнении таких условий имеем:

$$E\,((X^TX)^{-1}X^T\varepsilon) = E\,((X^TX)^{-1}X^T)\cdot E(\varepsilon) = 0$$

(если, конечно, математическое ожидание $E\,((X^TX)^{-1}X^T)$ существует и конечно), так что оценка наименьших квадратов для θ является несмещенной. Распределение статистик критериев (тестовых статистик) можно найти с помощью двухшаговой процедуры. На первом шаге находим условное распределение при фиксированном значении матрицы X; при этом значения объясняющих переменных рассматриваются как детерминированные (как в классической модели). На втором шаге мы получаем безусловное распределение соответствующей статистики, умножая условное распределение на плотность X и интегрируя по всем возможным значениям X.

Если применить такую процедуру для получения безусловного распределения оценки наименьших квадратов $\hat{\theta}$, то на первом шаге находим:

$$\hat{\theta} \mid X \sim N\left(\theta, \sigma^2 \left(X^T X\right)^{-1}\right).$$

Интегрирование на втором этапе приводит к распределению, являющемуся смесью нормальных распределений $N\left(\theta, \sigma^2 \left(X^T X\right)^{-1}\right)$ по X. Это распределение, в отличие от классического случая, не является нормальным.

В то же время для оценки j-го коэффициента имеем:

$$\hat{\theta}_j \mid X \sim N\left(\theta_j, \sigma^2 \left(X^T X\right)^{-1}_{j\,j}\right),$$

где $\left(X^T X\right)^{-1}_{jj}$ – j-й диагональный элемент матрицы $\left(X^T X\right)^{-1}$,

так что

$$\frac{\hat{\theta}_j - \theta_j}{\sigma\sqrt{(X^T X)^{-1}_{j\,j}}} \mid X \sim N(0, 1).$$

Условным распределением для $(n - p)S^2/\sigma^2$, где $S^2 = RSS/(n-p)$, RSS – остаточная сумма квадратов, является распределение хи-квадрат с $(n-p)$ степенями свободы,

$$(n-p)S^2/\sigma^2 \mid X \sim \chi^2(n-p).$$

Заметим теперь, что t-статистика для проверки гипотезы H_0: $\theta_j = \theta_j^*$ определяется соотношением

$$t = \frac{\hat{\theta}_j - \theta_j^*}{S\sqrt{(X^T X)^{-1}_{j\,j}}} = \frac{\left(\hat{\theta}_j - \theta_j^*\right) \big/ \sigma\sqrt{(X^T X)^{-1}_{j\,j}}}{\sqrt{S^2 \big/ \sigma^2}} \; .$$

Из предыдущего вытекает, что если гипотеза H_0 верна, то условное распределение этой t-статистики имеет t-распределение Стьюдента с $(n-p)$ степенями свободы,

$t \mid X \sim t(n-p)$.

Это условное распределение <u>одно и то же</u> для всех X . Поэтому вне зависимости от того, <u>какое именно</u> распределение имеет X, безусловным распределением t-статистики для $H_0: \theta_j = \theta_j^*$ при выполнении этой гипотезы будет <u>все то же</u> распределение $t(n-p)$.

Аналогичное рассмотрение показывает возможность использования стандартных F-критериев для проверки линейных гипотез о значениях коэффициентов.

Те же самые выводы остаются в силе при замене предположений ситуации А следующим предположением.

<u>Ситуация А′</u>

- $\varepsilon \mid X \sim N(0, \sigma^2 I_n)$, где I_n – единичная матрица (размера $n \times n$) .

<u>Ситуация С</u>

В рассмотренных выше ситуациях, как и в классической модели, предполагалось, что $\varepsilon_i \big| X \sim i.i.d.$ Теперь мы откажемся от этого предположения и предположим, что

- условное распределение случайного вектора ε относительно матрицы X является n-мерным нормальным распределением $N(0, \sigma^2 V)$;
- V – известная <u>положительно определенная</u> симметричная матрица размера $n \times n$.

Поскольку матрица V симметрична и положительно определена, таковой же будет и обратная к ней матрица V^{-1}. Но тогда существует такая невырожденная $(n\times n)$-матрица P, что $V^{-1} = P^TP$. Используя матрицу P, преобразуем вектор ε к вектору

$$\varepsilon^* = P\,\varepsilon\,.$$

При этом $E(\varepsilon^*) = 0$ и условная (относительно X) ковариационная матрица вектора ε^*

$$Cov(\varepsilon^* \mid X) = E(\varepsilon^*\varepsilon^{*T} \mid X) = E(P\,\varepsilon\,(P\,\varepsilon)^T \mid X) =$$
$$= P\,E(\varepsilon\,\varepsilon^T \mid X)\,P^T = P\,\sigma^2\,V\,P^T.$$

Но $V = (V^{-1})^{-1} = (P^TP)^{-1}$, так что

$$Cov(\varepsilon^* \mid X) = P\,\sigma^2\,V\,P^T = \sigma^2\,P(P^TP)^{-1}P^T = \sigma^2 I_n\,.$$

Преобразуя с помощью матрицы P обе части основного уравнения

$$y = X\theta + \varepsilon\,,$$

получаем:

$$Py = PX\theta + P\varepsilon\,,$$

или

$$y^* = X^*\theta + \varepsilon^*\,,$$

где

$$y^* = Py\,,\ X^* = PX\,,\ \varepsilon^* = P\varepsilon\,.$$

В преобразованном уравнении

$$\varepsilon^* \mid X \sim N(0, \sigma^2 I_n)\,,$$

так что преобразованная модель удовлетворяет условиям, характеризующим ситуацию A′. Это означает, что все результаты, полученные в ситуации A, применимы к модели $y^* = X^*\theta + \varepsilon^*$.

В частности, оценка наименьших квадратов

$$\theta^* = (X^{*T}X^*)^{-1}X^{*T}y^* = (X^TP^TPX)^{-1}X^TP^TPy = (X^TV^{-1}X)^{-1}X^TV^{-1}y$$

является несмещенной, т.е. $E(\theta^*) = \theta$, ее условное распределение (относительно X) нормально и имеет ковариационную матрицу

$$Cov(\theta^* \mid X) = \sigma^2(X^{*T}X^*)^{-1} = \sigma^2(X^TV^{-1}X)^{-1}.$$

Получение этой оценки равносильно минимизации по θ суммы

$$\sum_{i=1}^{n}\sum_{k=1}^{n} w_{ik}\left(y_i - \theta_1 x_{i1} - \ldots - \theta_p x_{ip}\right)\left(y_k - \theta_1 x_{k1} - \ldots - \theta_p x_{kp}\right),$$

где $w_{ik} = v_{ik}^{(-1)}$ – элементы матрицы V^{-1}.

Отсюда название метода – ***обобщенный метод наименьших квадратов***. Сама оценка θ^* называется ***обобщенной оценкой наименьших квадратов (GLS** – generalized least squares**)***.

В рамках модели $y^* = X^*\theta + \varepsilon^*$ можно использовать обычные статистические процедуры, основанные на t- и F-статистиках.

Заметим теперь, что во всех трех ситуациях A, A′ и C общим является условие

$$E\left(\varepsilon_i \middle| X\right) = 0, \quad i = 1, \ldots, n,$$

так что

$$E\left(\varepsilon_i \middle| x_{kj}\right) = 0 \quad \text{для } j = 1, \ldots, p \text{ \underline{при всех} } i \text{ и } k.$$

Но тогда

$$E(\varepsilon_i) = 0$$

и

$$Cov(\varepsilon_i, x_{kj}) = E(\varepsilon_i - E(\varepsilon_i))(x_{kj} - E(x_{kj})) = E(\varepsilon_i)(x_{kj} - E(x_{kj})) =$$
$$= E(\varepsilon_i x_{kj}) = 0$$

(конечно, при этом мы предполагаем, что математические ожидания $E(x_{kj})$ существуют и конечны).

Таким образом, если ошибка в i-м уравнении коррелирована хотя бы с одной из случайных величин x_{kj}, то ни одно из условий A, A′, C не выполняется. Например, эти условия не выполняются, если в i-м уравнении какая-нибудь из объясняющих переменных коррелирована с ошибкой в этом уравнении. Последнее характерно для моделей с ошибками в измерении объясняющих переменных и для моделей “одновременных уравнений”, о которых мы будем говорить ниже. Пока же приведем пример, показывающий, к каким последствиям приводит нарушение условия некоррелированности объясняющих переменных с ошибками.

П р и м е р

Смоделированные данные следуют процессу порождения данных (***DGP*** *– data generating process*)

DGP: $y_i = \alpha + \beta x_i + \varepsilon_i$, $\varepsilon_i \sim i.i.d.\ N(0,1)$, $i = 1, \ldots, 100$,

$\alpha = 10,\ \beta = 2$,

$x_i = \varepsilon_i - 0.9\varepsilon_{i-1}$, $i = 2, \ldots, 100$;

при этом $Corr\,(x_i, \varepsilon_i) = 0.743$.

Предположим, что мы имеем в распоряжении значения y_i, x_i, $i = 2, \ldots, 100$, но ничего не знаем о процессе порождения данных. Оценим на основании этих данных статистическую модель $y_i = \alpha + \beta x_i + \varepsilon_i$ методом наименьших квадратов. При этом получаем следующие результаты:

Dependent Variable: Y_FIXED
Method: Least Squares
Sample(adjusted): 2 100

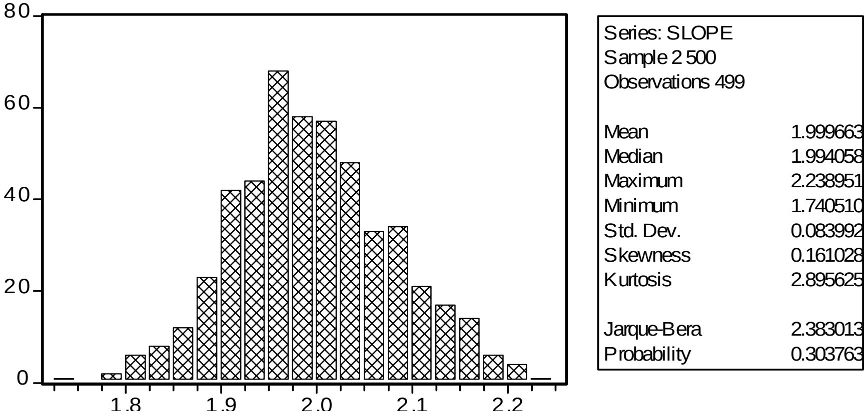
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	10.13984	0.069148	146.6398	0.0000
C(2)	2.553515	0.054971	46.45184	0.0000

Для параметра β получаем оценку $\hat{\beta} = 2.553$, имеющую весьма сильное смещение.

Зафиксировав полученную реализацию $x_2, \ldots, x_{100}$, смоделируем еще 499 последовательностей $\left\{\varepsilon_1^{(k)}, \ldots, \varepsilon_{100}^{(k)}\right\}$, $k = 2, \ldots, 500$, имитирующих реализации независимых случайных величин, имеющих стандартное нормальное распределение, и для каждой такой последовательности построим последовательность $\left\{y_2^{(k)}, \ldots, y_{100}^{(k)}\right\}$ по формуле:

$$y_i^{(k)} = \alpha + \beta x_i + \varepsilon_i^{(k)},\ i = 2, \ldots, 100.$$

Для каждого $k = 2, \ldots, 500$ по “данным” $y_i^{(k)}, x_i$, $i = 2, \ldots, 100$, оцениваем статистическую модель $y_i^{(k)} = \alpha + \beta x_i + \varepsilon_i^{(k)}$ и получаем оценки коэффициентов $\hat{\alpha}^{(k)}, \hat{\beta}^{(k)}$. В результате получаем последовательности оценок $\hat{\alpha}^{(2)}, \ldots, \hat{\alpha}^{(500)}$ и $\hat{\beta}^{(2)}, \ldots, \hat{\beta}^{(500)}$. Приведем статистические характеристики последовательности $\hat{\beta}^{(2)}, \ldots, \hat{\beta}^{(500)}$.



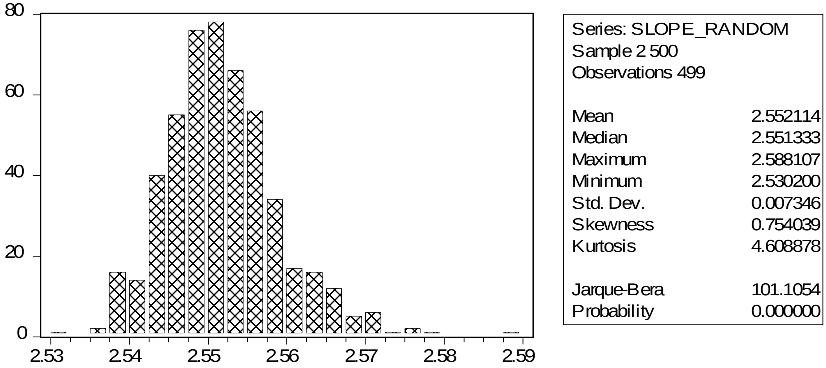
Среднее значение практически совпадает с истинным значением параметра β; гипотеза нормальности распределения оценки $\hat{\beta}$ не отвергается.

Поступим теперь другим образом. Для каждой из смоделированных последовательностей $\left\{\varepsilon_2^{(k)}, \ldots, \varepsilon_{100}^{(k)}\right\}$, $k = 2, \ldots, 500$, сначала построим последовательность $\left\{x_2^{(k)}, \ldots, x_{100}^{(k)}\right\}$, а затем построим последовательность $\left\{y_2^{(k)}, \ldots, y_{100}^{(k)}\right\}$ по формуле:

$$y_i^{(k)} = \alpha + \beta x_i^{(k)} + \varepsilon_i^{(k)}, \; i = 2, \ldots, 100.$$

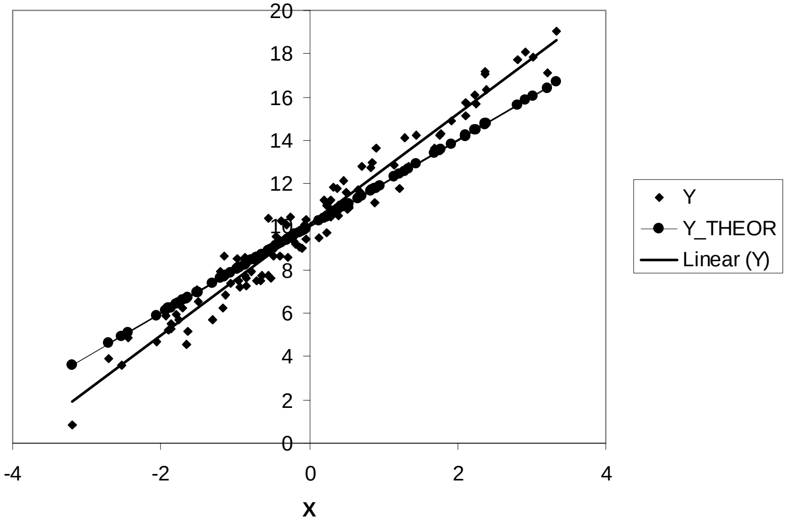
В отличие от предыдущего способа здесь для различных значений k используются <u>различные</u> последовательности $\left\{x_2^{(k)}, \ldots, x_{100}^{(k)}\right\}$, определяемые последовательностью $\left\{\varepsilon_2^{(k)}, \ldots, \varepsilon_{100}^{(k)}\right\}$. После получения последовательностей $\left\{x_2^{(k)}, \ldots, x_{100}^{(k)}\right\}$ и $\left\{y_2^{(k)}, \ldots, y_{100}^{(k)}\right\}$, при каждом $k = 2, \ldots, 500$ производим оценивание статистической модели

$y_i^{(k)} = \alpha + \beta x_i^{(k)} + \varepsilon_i^{(k)}$ и получаем оценки коэффициентов $\hat{\alpha}^{*(k)}, \hat{\beta}^{*(k)}$. Обозначая оценки, полученные в самом начале, как $\hat{\alpha}^{*(1)}$ и $\hat{\beta}^{*(1)}$, так что $\hat{\alpha}^{*(1)} = 10.13984$, $\hat{\beta}^{*(1)} = 2.553515$, получаем последовательности оценок $\hat{\alpha}^{*(1)},\ldots,\hat{\alpha}^{*(500)}$ и $\hat{\beta}^{*(1)},\ldots,\hat{\beta}^{*(500)}$. Приведем сводку статистических характеристик последовательности $\hat{\beta}^{*(1)},\ldots,\hat{\beta}^{*(500)}$.



На этот раз среднее значение полученных значений $\hat{\beta}^{*(k)}$, равное 2.552114, весьма сильно отличается от истинного значения параметра $\beta = 2$, а наблюдаемое значение статистики Харке–Бера говорит о том, что распределение оценки наименьших квадратов параметра $\beta = 2$ не является нормальным.

Заметим еще, что положительная коррелированность x_i и ε_i означает, что значениям x_i, превышающим их среднее значение в выборке, по большей части соответствуют и значения остатков, превышающие их среднее значение в выборке. Но последнее равно нулю при использовании метода наименьших квадратов, так что значения остатков, превышающие их среднее значение в выборке, суть просто положительные значения остатков. В итоге для первоначально смоделированных данных y_i, x_i, $i = 2,\ldots,100$, это приводит к следующей картине:



Здесь Linear(Y) – прямая, подобранная по этим данным методом наименьших квадратов, т.е. прямая $y = 10.13984 + 2.553515\,x$, а Y_THEOR – "теоретическая" прямая $y = 10 + 2\,x$. Как видно из графика, первая прямая "повернута" относительно второй прямой в направлении против часовой стрелки, так что для больших значений x_i наблюдаемые значения y_i смещены вверх по отношению к прямой $y = 10 + 2\,x$.

2.2. Модели, в которых некоторые объясняющие переменные коррелированы с ошибкой

2.2.1. Модели с ошибками в измерении объясняющих переменных

Рассмотрим модель порождения данных

DGP: $y_i = \alpha + \beta z_i + u_i$, $i = 1, \ldots, 100$,

со стохастической объясняющей переменной z, для которой выполнены предположения:

$$E(u_i) = 0,\ D(u_i) = \sigma^2,\ E(u_i | z_i) = 0,$$

так что

$$E(y_i | z_t) = \alpha + \beta z_i .$$

Предположим, что значение z_i невозможно измерить точно, и в результате измерения вместо истинного значения z_i наблюдается значение

$$x_i = z_i + v_i ,$$

где v_i – ошибка измерения. Подобное положение может соответствовать, например, ситуации, в которой y_i – сбережения i-го домохозяйства, а z_i – располагаемый доход домохозяйства. Пусть при этом выполнены следующие условия:

- $E(v_i) = 0,\ D(v_i) = \sigma_v^2$;
- случайные величины u_i и v_i независимы:
- случайная величина v_i не зависит от истинного значения z_i .

(Это означает, что истинный уровень располагаемого дохода не дает какой-либо информации о величине и знаке ошибки измерений.)

Выразим z_i через x_i и подставим $x_i - v_i$ вместо z_i в исходное уравнение. При этом получаем:

$$y_i = \alpha + \beta x_i + \varepsilon_i ,$$

где $\varepsilon_i = v_i - \beta u_i$ и

$$Cov(x_i, \varepsilon_i) = Cov(z_i + v_i, v_i - \beta u_i) = -\beta \sigma_u^2 .$$

Если $\beta > 0$, то x_i и ε_i имеют отрицательную корреляцию; если $\beta < 0$, то x_i и ε_i имеют положительную корреляцию.

Покажем, что оценка наименьших квадратов $\hat{\beta}$ не только имеет смещение при конечных n, но и несостоятельна, т.е. даже при неограниченном увеличении количества наблюдений не сходится к истинному значению β по вероятности. С этой целью обратимся к формуле для $\hat{\beta}$:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2};$$

подставим в нее выражение для y_i. Получаем:

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (\beta x_i - \beta \bar{x} + \varepsilon_i - \bar{\varepsilon})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \beta + \frac{\sum_{i=1}^{n} (\varepsilon_i - \bar{\varepsilon})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2},$$

так что

$$\underset{n \to \infty}{p \lim} \hat{\beta} = \beta + \frac{Cov(x_i, \varepsilon_i)}{D(x_i)} = \beta + \frac{-\beta \sigma_u^2}{\sigma_z^2 + \sigma_u^2} .$$

Таким образом, $\hat{\beta}$ не стремится по вероятности к β, за исключением случая, когда $\sigma_u^2 = 0$, т.е. когда ошибки измерения z_i отсутствуют. Если отношение дисперсий σ_u^2 / σ_z^2 мало, то тогда мало и асимптотическое смещение оценки наименьших квадратов; в противном случае асимптотическое смещение оказывается значительным. В примере со сбережениями $\beta > 0$, так что склонность к сбережению оказывается недооцененной.

2.2.2. Модели одновременных уравнений

Рассмотрим кейнсианскую модель потребления

$$C_t = \alpha + \beta Y_t + \varepsilon_t,$$

где C_t – реальное потребление на душу населения, Y_t – реальный доход на душу населения, и параметр β интерпретируется как склонность к потреблению (норма потребления). Мы могли бы на законных основаниях использовать для оценивания этого параметра метод наименьших квадратов, если бы не одно осложняющее обстоятельство. Если остановиться на модели замкнутой экономики без правительства, то в дополнение к указанному уравнению в этой модели имеется еще и соотношение

$$Y_t = C_t + I_t,$$

где I_t – реальные инвестиции на душу населения, что приводит к системе уравнений

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t \\ Y_t = C_t + I_t \end{cases},$$

о которой говорят как о ***структурной форме модели***.

Выражая из этой системы C_t и Y_t через I_t, получаем ***приведенную форму*** модели в виде:

$$\begin{cases} C_t = \dfrac{\alpha}{1-\beta} + \dfrac{\beta}{1-\beta} I_t + \dfrac{1}{1-\beta} \varepsilon_t, \\ Y_t = \dfrac{\alpha}{1-\beta} + \dfrac{1}{1-\beta} I_t + \dfrac{1}{1-\beta} \varepsilon_t. \end{cases}$$

Предположим, что $\varepsilon_t \sim i.i.d.$, $E(\varepsilon_t) = 0$, $D(\varepsilon_t) = \sigma^2 > 0$ и что для каждого t случайные величины I_t и ε_t независимы. Тогда из второго уравнения приведенной формы находим:

$$Cov(Y_t, \varepsilon_t) = \frac{1}{1-\beta} Cov(\varepsilon_t, \varepsilon_t) = \frac{\sigma^2}{1-\beta} \neq 0,$$

так что в исходном уравнении для C_t объясняющая переменная Y_t коррелирована с ошибкой. При этом для оценки $\hat{\beta}$ коэффициента β, получаемой применением метода наименьших квадратов к исходному уравнению, имеем:

$$\underset{n\to\infty}{p\lim}\, \hat{\beta} = \beta + \frac{Cov(Y, \varepsilon_t)}{D(Y_t)},$$

где

$$D(Y_t) = \frac{1}{(1-\beta)^2}\left(D(I_t) + \sigma^2\right),$$

и

$$\underset{n\to\infty}{p\lim}\, \hat{\beta} = \beta + (1-\beta)\frac{\sigma^2}{D(I_t) + \sigma^2}.$$

Поскольку $\sigma^2 > 0$ и в модели Кейнса $0 < \beta < 1$, то $\hat{\beta}$ <u>переоценивает</u> значение нормы потребления.

Заметим, однако, что получить оценки параметров α и β можно, минуя исходное уравнение и обращаясь только к уравнениям приведенной формы. В каждом из этих двух уравнений объясняющие переменные не коррелированы с ошибкой.

Первое уравнение приведенной формы можно записать в виде:

$$C_t = \tilde{\alpha} + \tilde{\beta} I_t + \tilde{\varepsilon}_t,$$

где $\tilde{\alpha} = \alpha/(1-\beta)$, $\tilde{\beta} = \beta/(1-\beta)$, $\tilde{\varepsilon}_t = \varepsilon_t/(1-\beta)$, $E(\tilde{\varepsilon}_t) = 0$, $D(\tilde{\varepsilon}_t) = \sigma^2_{\tilde{\varepsilon}} = \sigma^2_{\varepsilon}/(1-\beta)^2$. Применяя метод наименьших квадратов к этому уравнению, находим оценки коэффициентов $\tilde{\alpha}$ и $\tilde{\beta}$ и оценку дисперсии $\sigma^2_{\tilde{\varepsilon}}$, после чего можно найти оценки для параметров исходного уравнения, используя соотношения

$$\beta = \tilde{\beta}/(1+\tilde{\beta}),\ \alpha = \tilde{\alpha}/(1+\tilde{\beta}),\ \sigma^2_{\varepsilon} = \sigma^2_{\tilde{\varepsilon}}/(1+\tilde{\beta})^2.$$

Таким образом, структурная форма восстанавливается по первому уравнению приведенной формы. Второе уравнение оказывается в этом плане избыточным. Однако используя одно это уравнение, мы также можем восстановить структурную форму.

Действительно, это уравнение можно записать в виде:

$$Y_t = \tilde{\gamma} + \tilde{\delta}\, I_t + \tilde{\varepsilon}_t\,,$$

где $\tilde{\gamma} = \tilde{\beta} = \beta/(1-\beta)$, $\tilde{\delta} = 1/(1-\beta)$. Применяя метод наименьших квадратов к этому уравнению, находим оценки коэффициентов $\tilde{\gamma}$ и $\tilde{\delta}$ и оценку дисперсии $\sigma^2_{\tilde{\varepsilon}}$, после чего можно найти оценки для параметров исходного уравнения, используя соотношения

$$\beta = \left(\tilde{\delta} - 1\right)/\tilde{\delta}\,,\ \alpha = \tilde{\gamma}/\tilde{\delta}\,,\ \sigma^2_{\varepsilon} = \sigma^2_{\tilde{\varepsilon}}/\tilde{\delta}^2\,.$$

Однако при этом возникает вопрос о том, будут ли совпадать результаты восстановления параметров структурной формы, полученные по двум различным уравнениям приведенной формы.

Если обратиться к выражениям для α, β и σ^2_{ε} через параметры этих уравнений, то нетрудно заметить, что

$$\tilde{\gamma}/\tilde{\delta} = \tilde{\alpha}/\left(1+\tilde{\beta}\right),\ \left(\tilde{\delta}-1\right)/\tilde{\delta} = \tilde{\beta}/\left(1+\tilde{\beta}\right),\ \sigma^2_{\tilde{\tilde{\varepsilon}}}/\tilde{\delta}^2 = \sigma^2_{\tilde{\varepsilon}}/\left(1+\tilde{\beta}\right)^2\,,$$

так что, зная истинные значения параметров уравнений приведенной формы, мы однозначно восстанавливаем по ним значения параметров структурной формы. Однако это возможно, если мы знаем истинные значения параметров уравнений приведенной формы. Последние же нам не известны, и их приходится оценивать по имеющимся статистическим данным.

Поскольку в правых частях обоих уравнений приведенной формы стоят одни и те же объясняющие переменные, то можно показать, что эффективные оценки коэффициентов этих уравнений получаются применением метода наименьших квадратов к каждому из двух уравнений. Но при этом оценки параметров структурной формы, полученные с использованием оценок коэффициентов для разных уравнений приведенной формы, будут в общем случае отличаться друг от друга. И это связано с тем, что количество параметров приведенной формы больше количества, минимально необходимого для восстановления значений параметров структурной формы.

2.3. Метод инструментальных переменных

Прежде, чем перейти к описанию метода инструментальных переменных, обратимся к обычному методу наименьших квадратов, применяемому к простейшей линейной модели

$$y_i = \alpha + \beta x_i + \varepsilon_i, \ \ \varepsilon_i \sim i.i.d., \ E(\varepsilon_i) = 0, \ D(\varepsilon_i) = \sigma^2, \ i = 1, \dots, n.$$

В этом случае оценка наименьших квадратов для коэффициента β удовлетворяет системе нормальных уравнений

$$\sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i) = 0, \quad \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i) x_i = 0,$$

выражающей ортогональность вектора остатков $e = (e_1, \dots, e_n)^T$, где $e_i = y_i - \hat{\alpha} - \hat{\beta} x_i$ – остаток в i-м наблюдении, векторам $1 = (1, \dots, 1)^T$ и $x = (x_1, \dots, x_n)^T$. Эти условия ортогональности, записанные в равносильных формах

$$\frac{1}{n} \sum_{i=1}^{n} e_i 1 = 0, \quad \frac{1}{n} \sum_{i=1}^{n} e_i x_i = 0,$$

являются выборочными аналогами теоретических соотношений

$$Cov(\varepsilon_i, 1) = 0, \ Cov(\varepsilon_i, x_i) = 0.$$

В силу предположения $E(\varepsilon_i) = 0$, первое из двух последних соотношений выполняется автоматически, а второе можно записать в виде:

$$E(\varepsilon_i \cdot x_i) = 0.$$

Если $Cov(\varepsilon_i, x_i) \neq 0$, то $\underset{n \to \infty}{p\lim} \frac{1}{n} \sum_{i=1}^{n} e_i x_i \neq 0$ и соотношение $\frac{1}{n} \sum_{i=1}^{n} e_i x_i = 0$ не является эмпирическим аналогом теоретического соотношения. Можно было бы попытаться найти какую-то другую переменную z_i, для которой выполняется соотношение $Cov(\varepsilon_i, z_i) = E(\varepsilon_i \cdot z_i) = 0$, и заменить второе уравнение нормальной системы выборочным аналогом последнего соотношения, т.е. уравнением

$$\sum_{i=1}^{n}(y_i - \alpha - \beta\, x_i)z_i = 0.$$

Конечно, решение новой системы отличается от решения исходной системы, и мы временно обозначим получаемые оценки коэффициентов как α^* и β^*. Эти оценки удовлетворяют соотношениям

$$\sum_{i=1}^{n}(y_i - \alpha^* - \beta^*\, x_i) = 0, \quad \sum_{i=1}^{n}(y_i - \alpha^* - \beta^*\, x_i)z_i = 0,$$

из которых находим явное выражение для β^*:

$$\beta^* = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(z_i - \bar{z})}{\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})},$$

которое можно также записать в виде

$$\beta^* = \frac{\sum_{i=1}^{n}(\beta\, x_i - \beta\, \bar{x} + \varepsilon_i - \bar{\varepsilon})(z_i - \bar{z})}{\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})} = \beta + \frac{\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})(z_i - \bar{z})}{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z})}.$$

Здесь

$$\operatorname*{p\lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})(z_i - \bar{z}) = Cov(\varepsilon_i, z_i) = 0,$$

$$\operatorname*{p\lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(z_i - \bar{z}) = Cov(x_i, z_i),$$

так что для того, чтобы $\underset{n\to\infty}{p\lim}\,\beta^{*}=0$, необходимо выполнение еще одного условия: $Cov(x_i, z_i) \neq 0$.

Если для переменной z_i выполнены оба условия

$Cov(\varepsilon_i, z_i) = 0,\; Cov(x_i, z_i) \neq 0,$

то такую переменную называют ***инструментальной переменной***, или просто ***инструментом***. Наличие такой переменной позволяет получить состоятельную оценку коэффициента β при переменной x_i в ситуации, когда x_i коррелирована с ε_i. Инструментальная переменная является ***экзогенной переменной***, в том смысле, что она определяется вне связи с рассматриваемым уравнением $y_i = \alpha + \beta x_i + \varepsilon_i$, тогда как переменная x_i в рассматриваемом контексте является ***эндогенной переменной*** – она связана (коррелирована) с ошибкой в этом уравнении, так что значения x_i устанавливаются совместно с ε_i. Следуя обычной практике, мы будем снабжать оценки коэффициентов, полученные с использованием инструментальных переменных, подстрочным (или надстрочным) индексом *IV*: $\hat{\alpha}_{IV}, \hat{\beta}_{IV}$ (или $\hat{\alpha}^{IV}, \hat{\beta}^{IV}$). Здесь IV – аббревиатура от *Instrumental Variables (инструментальные переменные)*. Сам метод получения таких оценок называют ***методом инструментальных переменных***.

Возвратимся к системе, включающей кейнсианскую функцию потребления, т.е. к системе

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t, \\ Y_t = C_t + I_t. \end{cases}$$

При сделанных ранее предположениях относительно этой модели мы имеем: $Cov(Y_t, \varepsilon_t) = \dfrac{\sigma^2}{1-\beta} \neq 0$, так что Y_t – эндогенная переменная. В то же время $Cov(I_t, \varepsilon_t) = 0$ (в силу предположения о независимости этих случайных величин), так что I_t – экзогенная переменная. Используя второе уравнение приведенной формы, находим: $Cov(Y_t, I_t) = Cov\left(\frac{\alpha}{1-\beta} + \frac{1}{1-\beta} I_t + \frac{1}{1-\beta} \varepsilon_t, I_t\right) = \frac{1}{1-\beta} D(I_t) \neq 0$,

так что переменную I_t можно использовать в качестве инструмента для получения состоятельной оценки коэффициента β. Это приводит к оценке

$$\hat{\beta}_{IV} = \frac{\sum_{t=1}^{n}(C_t - \overline{C})(I_t - \bar{I})}{\sum_{t=1}^{n}(Y_t - \overline{Y})(I_t - \bar{I})}.$$

Мы можем получить это же выражение для IV-оценки коэффициента β следующим формальным образом. Возьмем ковариации обеих частей структурного уравнения $C_t = \alpha + \beta Y_t + \varepsilon_t$ с I_t. Это приводит к соотношению:

$$Cov(C_t, I_t) = Cov(\alpha, I_t) + \beta\, Cov(Y_t, I_t) + Cov(\varepsilon_t, I_t).$$

При сделанных предположениях оно сводится к равенству

$$Cov(C_t, I_t) = \beta\, Cov(Y_t, I_t),$$

откуда находим:

$$\beta = \frac{Cov(C_t, I_t)}{Cov(Y_t, I_t)}.$$

Чтобы получить оценку для β по n имеющимся наблюдениям, заменяем теоретические ковариации в правой части их выборочными аналогами:

$$\hat{\beta}_{IV} = \frac{\frac{1}{n}\sum_{t=1}^{n}\left(C_t - \overline{C}\right)\left(I_t - \overline{I}\right)}{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \overline{Y}\right)\left(I_t - \overline{I}\right)} = \frac{\sum_{t=1}^{n}\left(C_t - \overline{C}\right)\left(I_t - \overline{I}\right)}{\sum_{t=1}^{n}\left(Y_t - \overline{Y}\right)\left(I_t - \overline{I}\right)}.$$

П р и м е р

Рассмотрим статистические данные о следующих макроэкономических показателях экономики США [Gujarati (1995), p.651]:

Cons = расходы на личное потребление,

Y = валовый внутренний продукт,

I = валовые внутренние частные инвестиции.

Все показатели даны в млрд долл. 1987 г.

Год	CONS	Y	I
1970	1813.5	2873.9	429.7
1971	1873.7	2955.9	475.7
1972	1978.4	3107.1	532.2
1973	2066.7	3268.6	591.7
1974	2053.8	3248.1	543.0
1975	2097.5	3221.7	437.6
1976	2207.3	3380.8	520.6
1977	2296.6	3533.3	600.4
1978	2391.8	3703.5	664.6
1979	2448.4	3796.8	669.7
1980	2447.1	3776.3	594.4
1981	2476.9	3843.1	631.1
1982	2503.7	3760.3	540.5
1983	2619.4	3906.6	599.5
1984	2746.1	4148.5	757.5
1985	2865.8	4279.8	745.9
1986	2969.1	4404.5	735.1
1987	3052.2	4539.9	749.3
1988	3162.4	4718.6	773.4

1989	3223.3	4838.0	784.0
1990	3260.4	4877.5	739.1
1991	3240.8	4821.0	661.1

Оценивание методом наименьших квадратов уравнения

$$Cons_t = \alpha + \beta Y_t + \varepsilon_t$$

приводит к следующим результатам:

Dependent Variable: CONS
Method: Least Squares
Sample: 1970 1991
Included observations: 22

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-301.1158	38.52693	-7.815722	0.0000
Y	0.734314	0.009844	74.59484	0.0000

Оценивая уравнения приведенной формы, получаем:

$$Cons_t = 216.8 + 3.704 I_t + \varepsilon_t\,,$$

$$Y_t = 667.4 + 5.104 I_t + \varepsilon_t\,,$$

так что в принятых ранее обозначениях

$$\tilde{\alpha} = 216.8\,,\ \tilde{\beta} = 3.704\,,\ \tilde{\gamma} = 667.4\,,\ \tilde{\delta} = 216.8\,.$$

Использование оценок коэффициентов первого уравнения приводит к следующим оценкам для α и β :

$$\hat{\beta} = \tilde{\beta}\big/\big(1 + \tilde{\beta}\big) = 0.787\,,\ \hat{\alpha} = \tilde{\alpha}\big/\big(1 + \tilde{\beta}\big) = 46.1\,.$$

Если использовать оценки коэффициентов второго уравнения, то получаем:

$$\hat{\beta} = \big(\tilde{\delta} - 1\big)\big/\tilde{\delta} = 0.995\,,\ \hat{\alpha} = \tilde{\gamma}\big/\tilde{\delta} = 3.1\,.$$

Различие оказывается весьма существенным.

Вычисляя оценку коэффициента β с привлечением в качестве инструмента переменной I_t, находим:

$$\hat{\beta}_{IV} = \frac{\sum_{t=1}^{n}\left(C_t - \overline{C}\right)\left(I_t - \bar{I}\right)}{\sum_{t=1}^{n}\left(Y_t - \overline{Y}\right)\left(I_t - \bar{I}\right)} = 0.725604\ .$$

Заметим, что ту же самую оценку для β можно получить, используя двухшаговую процедуру, идея которой состоит в построении искусственной инструментальной переменной $\hat{Y}_t$, которой можно подменить эндогенную объясняющую переменную Y_t в структурном уравнении.

На первом шаге методом наименьших квадратов оценивается модель линейной зависимости эндогенной объясняющей переменной Y_t от инструментальной переменной I_t (она соответствует второму уравнению приведенной системы). Используя полученные оценки $\hat{\gamma}$ и $\hat{\delta}$, строим новую переменную

$$\hat{Y}_t = \hat{\gamma} + \hat{\delta}\, I_t\ ,$$

которая интерпретируется как результат “очистки” переменной Y_t от корреляционной связи с ε_t. Фактически, при этом производится “расщепление” переменной Y_t на две составляющие:

$$Y_t = \hat{Y}_t + \left(Y_t - \hat{Y}_t\right),$$

одна из которых затем отбрасывается.

На втором шаге методом наименьших квадратов оценивается модель

$$Cons_t = \alpha + \beta\,\hat{Y}_t + \varepsilon_t\ ,$$

в которой прежняя объясняющая переменная Y_t заменяется ее очищенным вариантом.

Такой метод оценивания параметров структурного уравнения $Cons_t = \alpha + \beta Y_t + \varepsilon_t$ называется ***двухшаговым методом наименьших квадратов***, сокращенно 2SLS (*two-stage least squares*). Оценки $\hat{\alpha}_{2SLS}$ и $\hat{\beta}_{2SLS}$, получаемые этим методом, удовлетворяют соотношениям

$$\sum_{t=1}^{n}(Cons_t - \hat{\alpha}_{2SLS} - \hat{\beta}_{2SLS}\, Y_i) = 0\,,$$

$$\sum_{t=1}^{n}(Cons_t - \hat{\alpha}_{2SLS} - \hat{\beta}_{2SLS}\, Y_t) I_t = 0\,,$$

т.е. являются IV-оценками.

Использование метода инструментальных переменных в форме 2SLS в нашем примере дает на втором шаге:

Dependent Variable: CONS
Method: Least Squares
Sample: 1970 1991
Included observations: 22

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-267.4634	352.8290	-0.758054	0.4573
Y_CLEANED	0.725604	0.090399	8.026691	0.0000

З а м е ч а н и е

В связи с использованием метода инструментальных переменных при наличии коррелированности некоторых объясняющих переменных с ошибками, возникают определенные проблемы:

- этот метод может обеспечить только состоятельность получаемых оценок и, при определенных условиях, асимптотическую нормальность этих оценок, но не обеспечивает несмещенность оценок при небольшом количестве наблюдений;

- для применения метода требуется достаточное количество инструментальных переменных, с помощью которых можно было бы “очистить” эндогенные объясняющие переменные; найти такие переменные удается далеко не всегда.

Первое обстоятельство означает, что ориентироваться на оценки, полученные методом инструментальных переменных, можно только при достаточно большом количестве имеющихся наблюдений, так что приведенный нами пример можно рассматривать только как иллюстрацию. Если наблюдений мало, то IV-оценки могут иметь даже большее смещение, чем OLS-оценки.

Второе обстоятельство значительно затрудняет практическое использование метода инструментальных переменных. Из-за этого, например, на практике обычно игнорируется тот факт, что используемые статистические данные содержат ошибки измерений.

Кроме того, исследования показывают, что если выбранные инструментальные переменные являются “***слабыми инструментами***” (weak instruments), т.е. слабо коррелированы с эндогенными объясняющими переменными, то качество IV-оценок с такими инструментами может быть хуже, чем у OLS-оценок (см., например, [Staiger, Stock (1997)]).

2.4. Проблема идентифицируемости структурной формы системы одновременных уравнений

При рассмотрении примера с кейнсианской функцией потребления мы обнаружили, что оценив коэффициенты приведенной системы, не можем однозначно восстановить с помощью полученных оценок коэффициенты структурного уравнения. Подобное положение встречается на практике довольно часто. Однако возможны и другие ситуации.

Рассмотрим простейшую модель рынка некоторого товара:

$$\begin{cases} Q^d = a_0 + a_1 P, \\ Q^s = b_0 + b_1 P, \\ Q^s = Q^d, \end{cases}$$

в которой Q^s – предложение товара, Q^d – спрос на товар, P – цена единицы товара; $a_1 < 0$, $b_1 > 0$. Если правила определения объемов предложения и спроса известны, т.е. известны коэффициенты a_0, a_1, b_0 и b_1, то при отсутствии флуктуаций равновесное решение для цены P и спроса Q находится без труда. Мы имеем здесь систему двух линейных уравнений с двумя неизвестными P и Q:

$$\begin{cases} Q - a_1 P = a_0, \\ Q - b_1 P = b_0, \end{cases}$$

решениями которой являются значения

$$Q = \frac{a_0 b_1 - a_1 b_0}{b_1 - a_1}, \; P = \frac{a_0 - b_0}{b_1 - a_1}.$$

При $a_0 > b_0$ оба эти значения положительны.

Предположим теперь, что спрос подвержен случайным флуктуациям, изменяющим значение a_0 до значения a_0+u_t в t-м наблюдении, а предложение подвержено флуктуациям, изменяющим в t-м наблюдении значение b_0 до значения b_0+v_t. Тогда каждому t соответствуют свои равновесные значения цены P_t и спроса Q_t, являющиеся решениями системы

$$\begin{cases} Q_t = a_0 + a_1 P_t + u_t, \\ Q_t = b_0 + b_1 P_t + v_t. \end{cases}$$

Поскольку значения P_t и Q_t определяются внутри системы, о переменных P_t и Q_t говорят как об ***эндогенных*** переменных. Их значения в t-м наблюдении определяются коэффициентами a_0, a_1, b_0, b_1 и внешними случайными воздействиями (шоками) u_t, v_t.

Положение выглядит теперь следующим образом:

- агенты наблюдают значения u_t и v_t;
- агенты взаимодействуют на рынке, устанавливая P_t и Q_t в соответствии с указанными правилами;
- статистик-эконометрист наблюдает только значения P_t и Q_t.

Обращаясь к оцениванию модели $Q_t = \alpha + \beta P_t + \varepsilon_t$, статистик даже не знает, что он оценивает: прямую спроса или прямую предложения. Так, при оценивании методом наименьших квадратов линейной модели зависимости потребления свинины на душу населения США от оптовых цен на свинину по годовым данным за период с 1948 по 1961 годы получаются следующие результаты ([Носко (2004), стр. 110]):

Переменная	Коэф-т	Ст. ошибка	t-статист.	P-знач.
1	77.484	13.921	5.566	0.0001
Цена	-24.775	29.794	-0.832	0.4219

Хотя формально отрицательное значение оценки коэффициента при цене говорит о том, что мы имеем дело с уравнением спроса, эта оценка оказывается статистически незначимой при любом разумном выборе уровня значимости, так что в доверительный интервал для данного коэффициента попадают как отрицательные, так и положительные значения.

Получая выражение для P_t из второго уравнения системы и подставляя это выражение вместо P_t в первое уравнение, находим:

$$Q_t = \frac{a_0 b_1 - a_1 b_0}{b_1 - a_1} + \frac{b_1 u_t - a_1 v_t}{b_1 - a_1} = \pi_1 + w_{t1}.$$

Выражение для Q_t из первого уравнения системы подставим во второе уравнение, в результате получаем:

$$P_t = \frac{a_0 - b_0}{b_1 - a_1} + \frac{u_t - v_t}{b_1 - a_1} = \pi_2 + w_{t2}.$$

Пара уравнений

$$\begin{cases} Q_t = \pi_1 + w_{t1} \\ P_t = \pi_2 + w_{t2} \end{cases}$$

представляет приведенную форму системы. Вообще-то говоря, здесь существует корреляция между ошибками в разных уравнениях приведенной формы при одном и том же t, даже если ошибки в структурной системе некоррелированы: в последнем случае

$$Cov(w_{t1}, w_{t2}) = Cov\left(\frac{b_1 u_t - a_1 v_t}{b_1 - a_1}, \frac{u_t - v_t}{b_1 - a_1}\right) = \frac{1}{b_1 - a_1}\left(b_1 D(u_t) + a_1 D(v_t)\right).$$

Однако поскольку в правых частях обоих уравнений приведенной формы находятся одни и те же объясняющие переменные (точнее, одна объясняющая переменная – константа), эффективные оценки коэффициентов приведенной формы получаются раздельным оцениванием обоих уравнений методом наименьших квадратов. Получив таким образом оценки $\hat{\pi}_{t1}, \hat{\pi}_{t1}$, мы тем самым получаем оценки для дробей $\dfrac{a_0 b_1 - a_1 b_0}{b_1 - a_1}$ и $\dfrac{a_0 - b_0}{b_1 - a_1}$. Но этих двух оценок недостаточно для восстановления по ним значений четырех коэффициентов a_0, a_1, b_0, b_1 структурных уравнений, так что здесь мы имеем дело с ***недоидентифицированостью*** структурной формы системы.

Включим в правую часть уравнения спроса доход (например, совокупный располагаемый доход) Y_t, так что система принимает вид:

$$\begin{cases} Q_t = a_0 + a_1 P_t + a_2 Y_t + u_t, \\ Q_t = b_0 + b_1 P_t + v_t. \end{cases}$$

Действуя аналогично предыдущему случаю, находим, что приведенная форма принимает здесь вид:

$$Q_t = \frac{a_0 b_1 - a_1 b_0}{b_1 - a_1} + \frac{a_2 b_1}{b_1 - a_1} Y_t + \frac{b_1 u_t - a_1 v_t}{b_1 - a_1} = \pi_{11} + \pi_{21} Y_t + w_{t1},$$

$$P_t = \frac{a_0 - b_0}{b_1 - a_1} + \frac{a_2}{b_1 - a_1} Y_t + \frac{u_t - v_t}{b_1 - a_1} = \pi_{12} + \pi_{22} Y_t + w_{t2}.$$

В приведенной форме 4 коэффициента, тогда как в структурной форме 5 коэффициентов. Поэтому и здесь нет возможности восстановления всех коэффициентов структурной формы по коэффициентам приведенной формы. Однако кое-что сделать все же можно.

Прежде всего заметим, что

$$\frac{\pi_{21}}{\pi_{22}} = b_1, \quad \pi_{11} - b_1 \pi_{12} = b_0,$$

так что коэффициенты уравнения предложения восстанавливаются по коэффициентам приведенной системы. В то же время для восстановления коэффициентов уравнения спроса остается только два уравнения, так что восстановить однозначно их значения не представляется возможным. Таким образом, здесь уравнение предложения идентифицируемо, а уравнение спроса неидентифицируемо: система ***частично идентифицируема***.

Пополним теперь и уравнение предложения. Если рассматриваемый товар – продукт сельскохозяйственного производства, то в качестве объясняющей переменной в правую часть этого уравнения естественно включить какой-либо подходящий индекс климатических условий, скажем среднее количество осадков в соответствующий период R_t. Тогда мы получаем систему:

$$\begin{cases} Q_t = a_0 + a_1 P_t + a_2 Y_t + u_t, \\ Q_t = b_0 + b_1 P_t + b_2 R_t + v_t \end{cases}$$

с 6 коэффициентами. Найдем приведенную форму этой системы

$$\begin{cases} Q_t = \pi_{11} + \pi_{21} Y_t + \pi_{31} R_t + w_{t1}, \\ P_t = \pi_{12} + \pi_{22} Y_t + \pi_{32} R_t + w_{t2}, \end{cases}$$

применяя матричный подход, обычно используемый для анализа и оценивания систем одновременных уравнений. Для этого заметим, что структурную форму системы можно записать в виде:

$$Q_t - a_1 P_t = a_0 + a_2 Y_t + u_t, \quad Q_t - b_1 P_t = b_0 + b_2 R_t + v_t,$$

или

$$(Q_t, P_t)\begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix} = (1, Y_t, R_t)\begin{pmatrix} a_0 & b_0 \\ a_2 & 0 \\ 0 & b_2 \end{pmatrix} + (u_t, v_t),$$

а приведенную форму – в виде:

$$(Q_t, P_t) = (1, Y_t, R_t)\begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \\ \pi_{31} & \pi_{32} \end{pmatrix} + (w_{t1}, w_{t2}).$$

Приведенную форму системы получаем из структурной формы, умножая обе части предпоследнего уравнения справа на матрицу, обратную к матрице, стоящей в левой части:

$$(Q_t, P_t) = \left\{ (1, Y_t, R_t) \begin{pmatrix} a_0 & b_0 \\ a_2 & 0 \\ 0 & b_2 \end{pmatrix} + (u_t, v_t) \right\} \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}^{-1} =$$

$$= (1, Y_t, R_t) \begin{pmatrix} a_0 & b_0 \\ a_2 & 0 \\ 0 & b_2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}^{-1} + (u_t, v_t) \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}^{-1} =$$

$$= (1, Y_t, R_t)\Pi + (u_t, v_t) \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}^{-1},$$

где Π – матрица коэффициентов приведенной формы,

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \\ \pi_{31} & \pi_{32} \end{pmatrix} = \begin{pmatrix} a_0 & b_0 \\ a_2 & 0 \\ 0 & b_2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}^{-1}.$$

Но

$$\begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}^{-1} = \frac{1}{a_1 - b_1} \begin{pmatrix} -b_1 & a_1 \\ 1 & 1 \end{pmatrix},$$

так что

$$Q_t = \frac{a_0 b_1 - a_1 b_0}{b_1 - a_1} + \frac{a_2 b_1}{b_1 - a_1} Y_t - \frac{a_1 b_2}{b_1 - a_1} R_t + \frac{b_1 u_t - a_1 v_t}{b_1 - a_1},$$

$$P_t = \frac{a_0 - b_0}{b_1 - a_1} + \frac{a_2}{b_1 - a_1} Y_t - \frac{b_2}{b_1 - a_1} R_t + \frac{u_t - v_t}{b_1 - a_1}.$$

Поскольку матрица коэффициентов приведенной формы получается как

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \\ \pi_{31} & \pi_{32} \end{pmatrix} = \begin{pmatrix} a_0 & b_0 \\ a_2 & 0 \\ 0 & b_2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}^{-1},$$

то

$$\begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \\ \pi_{31} & \pi_{32} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix} = \begin{pmatrix} a_0 & b_0 \\ a_2 & 0 \\ 0 & b_2 \end{pmatrix},$$

и это дает 6 уравнений для восстановления 6 коэффициентов структурной формы по коэффициентам приведенной формы:

$$\pi_{11} - \pi_{12}a_1 = a_0,\ \pi_{11} - \pi_{12}b_1 = b_0,$$

$$\pi_{21} - \pi_{22}a_1 = a_2,\ \pi_{21} - \pi_{22}b_1 = 0,$$

$$\pi_{31} - \pi_{32}a_1 = 0,\ \ \pi_{31} - \pi_{32}b_1 = b_2.$$

Из этих уравнений находим:

$$a_1 = \pi_{31} / \pi_{32},\ \ b_1 = \pi_{21} / \pi_{22},$$

$$a_2 = -\pi_{22}(\pi_{31} / \pi_{32} - \pi_{21} / \pi_{22}),\ \ b_2 = \pi_{32}(\pi_{31} / \pi_{32} - \pi_{21} / \pi_{22}),$$

$$a_0 = \pi_{11} - \pi_{12}\pi_{31} / \pi_{32},\ \ b_0 = \pi_{11} - \pi_{12}\pi_{21} / \pi_{22}\ .$$

Таким образом, здесь идентифицируемы и уравнение предложения и предложение спроса.

Рассмотрим теперь систему, в которой доход не включен в уравнение спроса, а уравнение предложения дополнено еще одной объясняющей переменной S_t – пусть это будет, скажем, индекс стоимости горюче-смазочных материалов, используемых при производстве соответствующего продукта сельского хозяйства. Тогда система принимает вид:

$$Q_t - a_1P_t = a_0 + u_t,\ \ Q_t - b_1P_t = b_0 + b_2R_t + b_3S_t + v_t\,,$$

или

$$(Q_t, P_t)\begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix} = (1,\ R_t, S_t)\begin{pmatrix} a_0 & b_0 \\ 0 & b_2 \\ 0 & b_3 \end{pmatrix} + (u_t, v_t)\,.$$

Матрица коэффициентов приведенной формы

$$Q_t = \pi_{11} + \pi_{21}R_t + \pi_{31}S_t + w_{t1},\ P_t = \pi_{12} + \pi_{22}R_t + \pi_{32}S_t + w_{t2}$$

получается как

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \\ \pi_{31} & \pi_{32} \end{pmatrix} = \begin{pmatrix} a_0 & b_0 \\ 0 & b_2 \\ 0 & b_3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}^{-1},$$

так что

$$\begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \\ \pi_{31} & \pi_{32} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix} = \begin{pmatrix} a_0 & b_0 \\ 0 & b_2 \\ 0 & b_3 \end{pmatrix},$$

и здесь мы опять получаем 6 уравнений для восстановления 6 коэффициентов структурной формы:

$$\pi_{11} - \pi_{12}a_1 = a_0,\ \pi_{11} - \pi_{12}b_1 = b_0,$$

$$\pi_{21} - \pi_{22}a_1 = 0,\ \ \pi_{21} - \pi_{22}b_1 = b_2,$$

$$\pi_{31} - \pi_{32}a_1 = 0,\ \ \pi_{31} - \pi_{32}b_1 = b_3.$$

Однако ситуация с идентифицируемостью резко отличается от предыдущего случая.

Для коэффициентов первого структурного уравнения находим:

$$a_0 = \pi_{11} - \pi_{12}a_1,\ \ a_1 = \pi_{21}/\pi_{22},\ \ a_1 = \pi_{31}/\pi_{32},$$

так что для восстановления коэффициента a_1 имеем два соотношения. Поскольку

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \\ \pi_{31} & \pi_{32} \end{pmatrix} = \begin{pmatrix} a_0 & b_0 \\ 0 & b_2 \\ 0 & b_3 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}^{-1} =$$

$$= \frac{1}{a_1 - b_1} \begin{pmatrix} a_0 & b_0 \\ 0 & b_2 \\ 0 & b_3 \end{pmatrix} \begin{pmatrix} -b_1 & a_1 \\ 1 & 1 \end{pmatrix} = \frac{1}{a_1 - b_1} \begin{pmatrix} -a_0b_1 + b_0 & a_0a_1 + b_0 \\ b_2 & b_2 \\ b_3 & b_3 \end{pmatrix},$$

то $\pi_{21}/\pi_{22} = \pi_{31}/\pi_{32}$, так что коэффициент a_1 восстанавливается однозначно, если мы знаем точно коэффициенты приведенной формы. Если же мы производим свободное оценивание уравнений приведенной формы по имеющимся статистическим данным, не принимая во внимание ограничений на их коэффициенты, накладываемые структурной формой, в данном случае ограничения $\pi_{21}/\pi_{22} - \pi_{31}/\pi_{32} = 0,$ то на основании оценок $\hat{\pi}_{21}, \hat{\pi}_{22}, \hat{\pi}_{31}, \hat{\pi}_{32}$ мы получим, как правило, различные значения отношений $\hat{\pi}_{21}/\hat{\pi}_{22}$ и $\hat{\pi}_{31}/\hat{\pi}_{32}$, так что получаются два варианта оценок для коэффициента a_1 и, соответственно, два варианта для коэффициента a_0. Таким образом, уравнение спроса оказывается ***сверхидентифицируемым*** – для восстановления его коэффициентов имеется количество соотношений, большее минимально необходимого.

Для коэффициентов второго структурного уравнения (уравнения предложения) также имеем три соотношения:

$$\pi_{11} - \pi_{12}b_1 = b_0,\ \pi_{21} - \pi_{22}b_1 = b_2,\ \pi_{31} - \pi_{32}b_1 = b_3 .$$

Однако во втором структурном уравнении четыре неизвестных коэффициента b_0, b_1, b_2, b_3, и этих трех соотношений недостаточно для их восстановления – этим соотношениям удовлетворяет бесконечно много наборов значений b_0, b_1, b_2, b_3.

Таким образом, для рассмотренной модели:

- уравнение спроса сверхидентифицировано;
- уравнение предложения недоидентифицировано.

Последний пример показывает, что имеет смысл говорить не только об идентифицируемости или неидентифицируемости системы в целом, а и об идентифицируемости или неидентифицируемости отдельных уравнений системы.

2.5. Проверка выполнения условий идентифицируемости структурных уравнений

При рассмотрении условий идентифицируемости отдельных структурных уравнений, входящих в систему одновременных уравнений[1], прежде всего предполагается, что переменные, задействованные в системе, подразделяются на три типа:

- ***эндогенные*** переменные;
- ***экзогенные*** переменные;
- ***предопределенные*** переменные.

Значения эндогенных переменных определяются внутри рассматриваемой системы; эндогеннная переменная, входящая в i-е уравнение системы, коррелирована с ошибкой в этом уравнении. Значения экзогенных переменных определяются вне рассматриваемой системы; экзогенные переменные не коррелированы с ошибками во всех уравнениях системы для всех моментов времени. Понятие предопределенной переменной относится к системам, в которых наблюдения производятся в последовательные моменты времени. Значения предопределенных переменных, как и значения эндогенных переменных, определяются внутри системы. Однако значение в момент t предопределенной переменной, входящей в i-е уравнение, не должно быть коррелированным со значениями ошибки в этом уравнении, соответствующими моментам t, $t+1$,… Например, в системе

$$\begin{cases} Q_t = a_1 P_t + a_2 Q_{t-1} + u_t, \\ P_t = b_1 Q_{t-1} + v_t \end{cases}$$

переменные Q_t и P_t – эндогенные, а переменная Q_{t-1} – предопределенная.

Предполагается, что

- система состоит из g уравнений, в каждое из которых входит хотя бы одна эндогенная переменная;
- в систему входит g эндогенных переменных;
- в систему входит K экзогенных и предопределенных переменных;

[1] В смысле возможности восстановления коэффициентов структурных уравнений на основании коэффициентов уравнений приведенной формы.

- каждое из g уравнений ***нормировано***, так что коэффициент при одной из эндогенных переменных, входящих в уравнение, равен 1.

(В последнем примере $g = 2$, $K = 1$, уравнения нормированы.)

При выводе условий идентифицируемости можно не различать предопределенные и экзогенные переменные, и мы для краткости будем называть их в контексте проблемы идентифицируемости предопределенными переменными.

Если собрать все эндогенные переменные в левых частях структурных уравнений, то систему одновременных уравнений можно записать в виде:

$$\begin{cases} \gamma_{11} y_{t1} + \ldots + \gamma_{g1} y_{tg} = \beta_{11} x_{t1} + \ldots + \beta_{K1} x_{tK} + u_{t1}, \\ \qquad \ldots \\ \gamma_{1g} y_{t1} + \ldots + \gamma_{gg} y_{tg} = \beta_{1g} x_{t1} + \ldots + \beta_{Kg} x_{tK} + u_{tg}, \end{cases}$$

где $t = 1, \ldots, n$, $y_{t1}, \ldots, y_{tg}$ – эндогенные переменные, $x_{t1}, \ldots, x_{tK}$ – предопределенные переменные, $u_{t1}, \ldots, u_{tg}$ – случайные ошибки. Заметим, что в этой записи γ_{ji} – коэффициент при j-й эндогенной переменной в i-м уравнении, а β_{ji} – коэффициент при j-й предопределенной переменной в i-м уравнении. (Разумеется, часть коэффициентов в конкретных системах равна нулю.) Заметив, что последнюю запись можно также представить как

$$\begin{cases} y_{t1} \gamma_{11} + \ldots + y_{tg} \gamma_{g1} = x_{t1} \beta_{11} + \ldots + x_{tK} \beta_{K1} + u_{t1}, \\ \qquad \ldots \\ y_{t1} \gamma_{1g} + \ldots + y_{tg} \gamma_{gg} = x_{t1} \beta_{1g} + \ldots + x_{tK} \beta_{Kg} + u_{tg}, \end{cases}$$

обозначим:

$$\Gamma = \begin{pmatrix} \gamma_{11} & \dots & \gamma_{1g} \\ \vdots & \ddots & \vdots \\ \gamma_{g1} & \dots & \gamma_{gg} \end{pmatrix}, \quad \mathrm{B} = \begin{pmatrix} \beta_{11} & \dots & \beta_{1g} \\ \vdots & \ddots & \vdots \\ \beta_{K1} & \dots & \beta_{Kg} \end{pmatrix},$$

$$y_t = (y_{t1}, \dots, y_{tg}), \quad x_t = (x_{t1}, \dots, x_{tK}), \; u_t = (u_{t1}, \dots, u_{tg}).$$

(Последние три вектора здесь удобнее представлять как векторы-строки.) Тогда система записывается в компактном виде:

$$y_t \Gamma = x_t \mathrm{B} + u_t, \quad t = 1, \dots, n.$$

Предполагая невырожденность матрицы Γ, так что для этой матрицы существует обратная, умножим обе части последнего уравнения на Γ^{-1}; при этом получаем приведенную форму системы:

$$y_t = x_t \mathrm{B}\Gamma^{-1} + u_t \Gamma^{-1} = x_t \Pi + w_t.$$

Здесь

$$\Pi = \mathrm{B}\Gamma^{-1} = \begin{pmatrix} \pi_{11} & \dots & \pi_{1g} \\ \vdots & \ddots & \vdots \\ \pi_{K1} & \dots & \pi_{Kg} \end{pmatrix}, \; w_t = u_t \Gamma^{-1} = (w_{t1}, \dots, w_{tg}),$$

так что w_{ti} – случайная ошибка в i-м уравнении приведенной формы в момент t. Выше мы уже фактически использовали это представление для получения приведенных форм систем

$$\begin{cases} Q_t = a_0 + a_1 P_t + a_2 Y_t + u_t, \\ Q_t = b_0 + b_1 P_t + b_2 R_t + v_t \end{cases}$$

и

$$\begin{cases} Q_t - a_1 P_t = a_0 + u_t, \\ Q_t - b_1 P_t = b_0 + b_2 R_t + b_3 S_t + v_t. \end{cases}$$

Следует заметить, что даже если векторы $u_t = (u_{t1}, \dots, u_{tg})$, $i = 1, \dots, n$, взаимно независимы и имеют одинаковое g-мерное нормальное распределение с нулевым средним и ковариационной матрицей $\Sigma = \sigma^2 I_g$, где I_g – единичная матрица, векторы

$w_t = (w_{t1}, \ldots, w_{tg})$ могут иметь коррелированные между собой и неодинаково распределенные компоненты. Однако это не препятствует получению эффективных и несмещенных оценок элементов матрицы Π обычным методом наименьших квадратов: достаточно применить этот метод отдельно к каждому уравнению приведенной системы.

Поскольку

$$\Pi = \mathrm{B}\Gamma^{-1} = \begin{pmatrix} \pi_{11} & \ldots & \pi_{1g} \\ \vdots & \ddots & \vdots \\ \pi_{K1} & \ldots & \pi_{Kg} \end{pmatrix},$$

то $\Pi\Gamma = \mathrm{B}$, и мы использовали это соотношение для восстановления коэффициентов структурных уравнений двух последних систем. Вопрос об идентифицируемости структурной формы – это вопрос о возможности однозначного восстановления всех коэффициентов структурной формы, т.е. восстановления матриц Γ и B, на основании матрицы $\Pi = \mathrm{B}\Gamma^{-1}$. Заметим, что в совокупности матрицы Γ и B состоят из $g^2 + Kg$ элементов, тогда как в матрице Π всего Kg элементов. Это означает, что однозначное восстановление коэффициентов структурной формы по коэффициентам приведенной формы невозможно без использования дополнительной информации в виде невключения в отдельные уравнения тех или иных переменных, нормировки коэффициентов, линейных ограничений на параметры структуры[2].

---

[2] Как мы увидим ниже в этом разделе (см. Замечание 5) коэффициенты структурной формы могут не восстанавливаться однозначно по одним только коэффициентам приведенной формы и в то же время однозначно восстанавливаться при привлечении дополнительной информации в виде ограничений на элементы ковариационной матрицы ошибок в правых частях уравнений структурной формы и использовании элементов ковариационной матрицы ошибок в правых частях уравнений приведенной формы.

Если нас интересует i-е структурное уравнение, то идентифицируемость этого уравнения означает возможность однозначного восстановления на основании коэффициентов приведенной формы

- Γ_i – i-го столбца матрицы Γ, который содержит коэффициенты при эндогенных переменных, входящих в i-е структурное уравнение;
- B_i – i-го столбца матрицы B, который содержит коэффициенты при предопределенных переменных, входящих в i-е структурное уравнение.

При этом по-существу достаточно иметь возможность восстановления Γ_i и B_i с точностью до умножения их на один и тот же числовой множитель: единственность достигается в этом случае указанием правила нормировки, в соответствии с которым коэффициент при определенной эндогенной переменной в i-м структурном уравнении полагается равным 1.

Для дальнейшего удобно использовать матрицу A размера $(g+K)\times g$, составленную из матриц Γ и B таким образом, что матрица Γ располагается над матрицей B:

$$A=\begin{bmatrix}\Gamma\\ B\end{bmatrix}.$$

Коэффициенты при g эндогенных и K предопределенных переменных в i-м структурном уравнении составляют i-й столбец α_i матрицы A.

Существенным является то обстоятельство, что коэффициенты i-го структурного уравнения не могут быть восстановлены на основании коэффициентов приведенной формы, если в это уравнение входят все (g) эндогенные и все (K) предопределенные переменные системы.

Поэтому мы будем предполагать далее, что на элементы вектора α_i помимо нормировочного накладываются еще и некоторые дополнительные однородные линейные ограничения в виде уравнений

$$\Phi_i \alpha_i = 0,$$

где Φ_i – матрица размера $R_i \times (g+K)$, R_i – количество этих линейных ограничений. Неспецифицированные коэффициенты i-го уравнения определяются по матрице $\Pi = \mathrm{B}\Gamma^{-1}$ после применения правила нормировки однозначным образом тогда и только тогда, когда выполнено следующее ***ранговое условие идентифицируемости***:

$$\boxed{\operatorname{rank}(\Phi_i \mathrm{A}) = g-1}.$$

(Матрица $\Phi_i \mathrm{A}$ имеет R_i строк и g столбцов.)

Пусть A_i – матрица, получаемая из матрицы A вычеркиванием ее i-го столбца α_i, так что $\mathrm{A} = [\alpha_i : \mathrm{A}_i]$. Тогда

$$\operatorname{rank}(\Phi_i \mathrm{A}) = \operatorname{rank}(\Phi_i [\alpha_i : \mathrm{A}_i]) = \operatorname{rank}(\Phi_i \alpha_i : \Phi_i \mathrm{A}_i),$$

и поскольку $\Phi_i \alpha_i = 0$, то

$$\operatorname{rank}(\Phi_i \mathrm{A}) = \operatorname{rank}(0 : \Phi_i \mathrm{A}_i) = \operatorname{rank}(\Phi_i \mathrm{A}_i).$$

Но матрица $\Phi_i \mathrm{A}_i$ имеет размер $R_i \times (g-1)$, и чтобы ее ранг был равен $g-1$, во всяком случае необходимо, чтобы выполнялось следующее ***порядковое условие идентифицируемости*** i-го структурного уравнения:

$$\boxed{R_i \ge g-1.}$$

Предположим, что все линейные ограничения, накладываемые на элементы столбца α_i (помимо условия нормировки) являются ***исключающими ограничениями*** (т.е. все они состоят в приравнивании определенных элементов столбца α_i нулю) и соответствуют исключению из i-го уравнения g_i^* эндогенных и K_i^* предопределенных переменных. Тогда общее количество исключенных переменных равно $g_i^* + K_i^*$, и необходимое условие идентифицируемости i-го структурного уравнения принимает вид:

$$\boxed{g_i^* + K_i^* \ge g - 1,}$$

или

$$\boxed{K_i^* \ge \left(g - g_i^*\right) - 1.}$$

Иначе говоря, количество предопределенных переменных в системе, не включенных в i-е структурное уравнение, должно быть не меньше количества эндогенных переменных, включенных в i-е уравнение, уменьшенного на единицу. Если в левой части i-го структурного уравнения находится единственная эндогенная переменная, то $\left(g - g_i^*\right) - 1$ есть просто количество эндогенных переменных, включенных в правую часть этого уравнения.

Теперь мы имеем возможность охарактеризовать три ситуации, возникающие при оценивании i-го структурного уравнения:

1. $\operatorname{rank}\left(\Phi_i \mathrm{A}\right) < g - 1$ ➔ i-е уравнение ***неидентифицируемо*** (***недоопределено***);
2. $\operatorname{rank}\left(\Phi_i \mathrm{A}\right) = g - 1$ и $R_i = g - 1$ ➔ i-е уравнение ***идентифицируемо точно***;
3. $\operatorname{rank}\left(\Phi_i \mathrm{A}\right) = g - 1$ и $R_i > g - 1$ ➔ i-е уравнение ***сверхидентифицируемо*** (***переопределено***).

В ситуации 1 просто не выполнено необходимое условие идентифицируемости. В ситуациях 2 и 3 коэффициенты i-го структурного уравнения однозначно восстанавливаются на основании коэффициентов приведенной системы. Однако эти две ситауции различаются существенным образом, если рассматривать задачу восстановления коэффициентов i-го структурного уравнения на основании оценок коэффициентов приведенной формы, полученных методом наименьших квадратов, примененным к каждому отдельному уравнению приведенной системы и не учитывающем ограничения на коэффициенты приведенной формы, накладываемые на них соотношением $\Pi = \mathrm{B}\Gamma^{-1}$. Если $\hat{\Pi}$ – оценка матрицы Π, полученная таким свободным оцениванием, то в ситуации 2 коэффициенты i-го структурного уравнения восстанавливаются по матрице $\hat{\Pi}$ однозначным образом, тогда как в ситуации 3 существует несколько вариантов такого восстановления, приводящих к различным результатам.

Заметим, что разным уравнениям системы могут соответствовать разные ситуации из трех перечисленных.

Пробежимся теперь по уже рассмотренным в этом разделе примерам систем одновременных уравнений.

Первой мы рассмотрели систему

$$\begin{cases} Q_t = a_0 + a_1 P_t + u_t, \\ Q_t = b_0 + b_1 P_t + v_t. \end{cases}$$

Здесь список эндогенных переменных: (Q_t, P_t), а список предопределенных переменных ограничивается переменной, тождественно равной 1, так что полный список переменных в системе: $(Q_t, P_t, 1)$. При этом $g = 2, K = 1$, матрицы Γ, B и A имеют вид:

$$\Gamma = \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}, \ \mathrm{B} = (a_0, b_0),$$

$$\mathrm{A} = \begin{pmatrix} \Gamma \\ \mathrm{B} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \\ a_0 & b_0 \end{pmatrix}.$$

На столбцы матрицы A не накладывается никаких ограничений кроме нормировочных, так что $g_1^* = g_2^* = 0$, $K_1^* = K_2^* = 0$, и ни для одного из двух уравнений не выполнено порядковое условие $g_i^* + K_i^* \geq g - 1$. Следовательно система не идентифицируема.

Следующий пример:

$$\begin{cases} Q_t = a_0 + a_1 P_t + a_2 Y_t + u_t, \\ Q_t = b_0 + b_1 P_t + v_t, \end{cases}$$

т.е.

$$\begin{cases} Q_t - a_1 P_t = a_0 + a_2 Y_t + u_t, \\ Q_t - b_1 P_t = b_0 + v_t. \end{cases}$$

Здесь список эндогенных переменных тот же: (Q_t, P_t). В список предопределенных переменных входят две переменные: $(1, Y_t)$. Полный список переменных в системе: $(Q_t, P_t, 1, Y_t)$. При этом $g = 2$, $K = 2$, матрицы Γ, B и A имеют вид:

$$\Gamma = \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}, \quad \mathrm{B} = \begin{pmatrix} a_0 & b_0 \\ a_2 & 0 \end{pmatrix},$$

$$\mathrm{A} = \begin{pmatrix} \alpha_1 & \alpha_2 \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \\ \alpha_{41} & \alpha_{42} \end{pmatrix} = \begin{pmatrix} \Gamma \\ \mathrm{B} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \\ a_0 & b_0 \\ a_2 & 0 \end{pmatrix}.$$

На элементы первого столбца матрицы A накладывается только условие нормировки $\alpha_{11} = 1$. Поэтому первое уравнение системы неидентифицируемо. На элементы второго столбца помимо нормировочного накладывается одно исключающее ограничение $\alpha_{42} = 0$, так что для этого столбца $g_2^* = 0$, $K_2^* = 1$, и $g_2^* + K_2^* = g - 1 = 1$, т.е. порядковое условие идентифицируемости выполняется.

Заметим далее, что ограничение $\alpha_{42} = 0$ можно записать в виде $\Phi_2 \alpha_2 = 0$, где $\Phi_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \end{pmatrix}$. Тогда

$$\Phi_2 \mathrm{A}_2 = (0\ 0\ 0\ 1)(1, -a_1, a_0, a_2)^T = (a_2),$$

$$\mathrm{rank}(\Phi_2 \mathrm{A}) = \mathrm{rank}(\Phi_2 \mathrm{A}_2) = \mathrm{rank}(a_2) = 1,$$

так что $\mathrm{rank}(\Phi_2 \mathrm{A}) = g - 1$ и выполнено ранговое условие идентифицируемости. Наконец, поскольку $g_2^* + K_2^* = g - 1$, то второе уравнение идентифицируемо точно.

Следующая система:

$$\begin{cases} Q_t = a_0 + a_1 P_t + a_2 Y_t + u_t, \\ Q_t = b_0 + b_1 P_t + b_2 R_t + v_t, \end{cases}$$

т.е.

$$\begin{cases} Q_t - a_1 P_t = a_0 + a_2 Y_t + u_t, \\ Q_t - b_1 P_t = b_0 + b_2 R_t + v_t. \end{cases}$$

Список эндогенных переменных: (Q_t, P_t). Список предопределенных переменных: $(1, Y_t, R_t)$. Полный список переменных в системе: $(Q_t, P_t, 1, Y_t, R_t)$. При этом $g = 2, K = 3$, матрицы Γ, B и A имеют вид:

$$\Gamma = \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}, \ \mathrm{B} = \begin{pmatrix} a_0 & b_0 \\ a_2 & 0 \\ 0 & b_2 \end{pmatrix},$$

$$\mathrm{A} = \begin{pmatrix} \Gamma \\ \mathrm{B} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \\ a_0 & b_0 \\ a_2 & 0 \\ 0 & b_2 \end{pmatrix}.$$

Соответственно, здесь для каждого из столбцов матрицы A помимо нормирующего ограничения имеется по одному исключающему ограничению на экзогенные переменные, так что $g_1^* = g_2^* = 0$, $K_1^* = K_2^* = 1$, $g_i^* + K_i^* = g - 1$, и порядковое условие выполнено.

Ограничение $\alpha_{41} = 0$ в первом столбце можно записать в виде $\Phi_1 \alpha_1 = 0$, где $\Phi_1 = (0\ 0\ 0\ 0\ 1)$. Тогда

$$\Phi_1 \mathrm{A}_1 = (0\ 0\ 0\ 0\ 1)(1, -b_1, b_0, 0, b_2)^T = (b_2),$$

$$\operatorname{rank}(\Phi_1 \mathrm{A}) = \operatorname{rank}(\Phi_1 \mathrm{A}_1) = \operatorname{rank}(b_2) = 1,$$

так что $\operatorname{rank}(\Phi_1 \mathrm{A}) = g - 1$ и для первого уравнения выполнено ранговое условие идентфицируемости. Наконец, поскольку $g_1^* + K_1^* = g - 1,$ то первое уравнение идентифицируемо точно.

Ограничение $\alpha_{32} = 0$ во втором столбце можно записать в виде $\Phi_2 \alpha_2 = 0$, где $\Phi_2 = (0\ 0\ 0\ 1\ 0)$. Тогда

$$\Phi_2 \mathrm{A}_2 = (0\ 0\ 0\ 1\ 0)(1, -a_1, a_0, a_2, 0)^T = (a_2),$$

$$\operatorname{rank}(\Phi_2 \mathrm{A}) = \operatorname{rank}(\Phi_2 \mathrm{A}_2) = \operatorname{rank}(a_2) = 1,$$

так что $\operatorname{rank}(\Phi_2 \mathrm{A}) = g - 1$ и для второго уравнения также выполнено ранговое условие идентифицируемости. Наконец, поскольку $g_2^* + K_2^* = g - 1,$ то второе уравнение идентифицируемо точно.

Таким образом в данной системе одновременных уравнений оба уравнения идентифицируемы, причем идентифицируемы точно.

Наконец, в системе

$$\begin{cases} Q_t = a_0 + a_1 P_t + u_t, \\ Q_t = b_0 + b_1 P_t + b_2 R_t + b_3 S_t + v_t, \end{cases}$$

т.е.

$$\begin{cases} Q_t - a_1 P_t = a_0 + u_t, \\ Q_t - b_1 P_t = b_0 + b_2 R_t + b_3 S_t + v_t, \end{cases}$$

эндогенные переменные те же, а список предопределенных переменных: $(1, R_t, S_t)$. Полный список переменных в системе:

$(Q_t, P_t, 1, R_t, S_t)$. При этом $g = 2, K = 3$, матрицы Γ, B и A имеют вид:

$$\Gamma = \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}, \ \mathrm{B} = \begin{pmatrix} a_0 & b_0 \\ 0 & b_2 \\ 0 & b_3 \end{pmatrix},$$

$$\mathrm{A} = \begin{pmatrix} \Gamma \\ \mathrm{B} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \\ a_0 & b_0 \\ 0 & b_2 \\ 0 & b_3 \end{pmatrix}.$$

На элементы второго столбца накладывается только условие нормировки. Поэтому второе уравнение системы неидентифицируемо. На элементы первого столбца помимо условия нормировки накладываются два исключающих ограничения: $\alpha_{41} = 0$, $\alpha_{51} = 0$. При этом $g_1^* = 0$, $K_1^* = 2$, $g_1^* + K_1^* = 2 > g - 1$, так что первое уравнение идентифицируемо. Исключающие ограничения можно записать в форме $\Phi_1 \alpha_1 = 0$, где

$$\Phi_1 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Тогда

$$\Phi_1 \mathrm{A} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \\ a_0 & b_0 \\ 0 & b_2 \\ 0 & b_3 \end{pmatrix} = \begin{pmatrix} 0 & b_2 \\ 0 & b_3 \end{pmatrix},$$

$$\Phi_1 \mathrm{A}_1 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} (1, -b_1, b_0, b_2, b_3)^T = \begin{pmatrix} b_2 \\ b_3 \end{pmatrix},$$

$$\operatorname{rank}(\Phi_1 \mathrm{A}) = \operatorname{rank}(\Phi_1 \mathrm{A}_1) = \operatorname{rank}\begin{pmatrix} b_2 \\ b_3 \end{pmatrix} = 1,$$

так что $\operatorname{rank}(\Phi_1 \mathrm{A}) = g - 1$ и для первого уравнения выполнено ранговое условие идентифицируемости. Поскольку $g_1^* + K_1^* > g - 1$, то первое уравнение сверхидентифицируемо.

Приведем теперь пример системы, в которой присутствуют линейные ***ограничения неисключающего типа*** (упрощенный вариант модели мультипликатора-акселератора):

$$\begin{cases} C_t = a_0 + a_1 Y_t + a_2 C_{t-1} + u_{t1}, \\ I_t = b_0 + b_1 (Y_t - Y_{t-1}) + u_{t2}, \\ Y_t = C_t + I_t, \end{cases}$$

где C_t – потребление, I_t – инвестиции, Y_t – доход. Подставляя выражение для Y_t из последнего тождества во второе уравнение, запишем систему в виде:

$$\begin{cases} C_t - a_1 Y_t = a_0 + a_2 C_{t-1} + u_{t1}, \\ -C_t + (1 - b_1) Y_t = b_0 - b_1 Y_{t-1} + u_{t2}. \end{cases}$$

Список эндогенных переменных: (C_t, Y_t). Список предопределенных переменных: $(1, C_{t-1}, Y_{t-1})$. Полный список: $(C_t, Y_t, 1, C_{t-1}, Y_{t-1})$. Матрица A:

$$\mathrm{A} = \begin{pmatrix} 1 & -1 \\ -a_1 & 1 - b_1 \\ a_0 & b_0 \\ a_2 & 0 \\ 0 & -b_1 \end{pmatrix}.$$

В первом столбце одно исключающее ограничение $\alpha_{51} = 0$, т.е. $\Phi_1 \alpha_1 = 0$, где $\Phi_1 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \end{pmatrix}$. При этом

$$\operatorname{rank}\left(\Phi_1 \mathrm{A}\right) = \operatorname{rank}\left(\Phi_1 \mathrm{A}_1\right) = \operatorname{rank}\left(-b_1\right) = 1 = g - 1,$$

так что первое уравнение идентифицируемо. Поскольку

$$g_1^* + K_1^* = 1 = g - 1,$$

это уравнение идентифицируемо точно.

Во втором столбце одно исключающее ограничение $\alpha_{42} = 0$ и одно неисключающее ограничение $\alpha_{12} + \alpha_{22} = \alpha_{52}$. Эту пару ограничений можно записать в виде $\Phi_2 \alpha_2 = 0$, где

$$\Phi_2 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & -1 \end{pmatrix}.$$

Тогда

$$\Phi_2 \mathrm{A} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -a_1 & 1 - b_1 \\ a_0 & b_0 \\ a_2 & 0 \\ 0 & -b_1 \end{pmatrix} = \begin{pmatrix} a_2 & 0 \\ 1 - a_1 & 0 \end{pmatrix},$$

$$\operatorname{rank}\left(\Phi_2 \mathrm{A}\right) = 1,$$

так что $\operatorname{rank}\left(\Phi_2 \mathrm{A}\right) = g - 1$ и для второго уравнения также выполнено ранговое условие идентифицируемости. Поскольку же $R_2 = 2 > g - 1$, то второе уравнение сверхидентифицируемо.

Замечание 1

Константа играет в проблеме идентифицируемости такую же роль, что и остальные предопределенные переменные. Это илюстрирует следующий пример.

Мы уже выяснили ранее, что в системе

$$\begin{cases} Q_t = a_0 + a_1 P_t + u_t, \\ Q_t = b_0 + b_1 P_t + v_t \end{cases}$$

оба уравнения неидентифицируемы. Исключим константу из правой части второго уравнения:

$$\begin{cases} Q_t = a_0 + a_1 P_t + u_t, \\ Q_t = b_1 P_t + v_t. \end{cases}$$

Для измененной системы имеем те же списки эндогенных и предопределенных переменных; полный список переменных в системе: $(Q_t, P_t, 1)$. При этом $g = 2$, $K = 1$, матрица Γ не изменяется, а матрицы B и A принимают вид:

$$\Gamma = \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}, \ \mathrm{B} = (a_0, 0),$$

$$\mathrm{A} = \begin{pmatrix} \Gamma \\ \mathrm{B} \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \\ a_0 & 0 \end{pmatrix}.$$

На первый столбец матрицы A не накладывается никаких ограничений кроме нормировочных, так что $g_1^* = 0$, $K_1^* = 0$, и для первого уравнения не выполнено порядковое условие $g_i^* + K_i^* \geq g - 1$. Следовательно первое уравнение не идентифицируемо. Однако на второй столбец на этот раз накладывается исключающее ограничение $\alpha_{32} = 0$, т.е. $\Phi_2 \alpha_2 = 0$, где $\Phi_2 = (0 \quad 0 \quad 1)$.

При этом

$$\operatorname{rank}(\Phi_2 \mathrm{A}) = \operatorname{rank}(a_0 \quad 0) = \operatorname{rank}(\Phi_2 \mathrm{A}_1) = (a_0) = 1 = g - 1,$$

так что второе уравнение идентифицируемо. Поскольку

$$g_2^* + K_2^* = 1 = g - 1,$$

это уравнение идентифицируемо точно.

З а м е ч а н и е 2

Критерий идентифицируемости дает один и тот же результат в отношении i-го стохастического структурного уравнения (содержащего случайные ошибки в правой части) независимо от того, рассматривается полная система вместе с тождествами или система, в которой тождества учтены и исключены. Это илюстрирует следующий пример.

При исследовании вопроса об идентифицируемости модели

$$\begin{cases} Q^d = a_0 + a_1 P + u_t, \\ Q^s = b_0 + b_1 P + v_t, \\ Q^s = Q^d \end{cases}$$

и различных ее расширений мы, исключая (и учитывая) тождество, сводили эти модели к системам без тождеств, так что в правых частях всех уравнений преобразованных систем присутствовали случайные ошибки. Поступая, например, таким образом с системой трех уравнений

$$\begin{cases} Q_t^d = a_0 + a_1 P_t + a_2 Y_t + u_t, \\ Q_t^s = b_0 + b_1 P_t + v_t, \\ Q_t^d = Q_t^s, \end{cases}$$

мы проверяли условия идентифицируемости системы двух уравнений, полученных на основании этой системы:

$$\begin{cases} Q_t = a_0 + a_1 P_t + a_2 Y_t + u_t, \\ Q_t = b_0 + b_1 P_t + v_t, \end{cases}$$

и обнаружили, что первое уравнение системы неидентифицируемо, а второе идентифицируемо точно.

Попоробуем проверить условия идентифицируемости непосредственно в рамках исходной системы трех уравнений, так что $g = 3$. Для этой системы список эндогенных переменных полнее, чем у преобразованной системы: $\left(Q_t^d, Q_t^s, P_t\right)$, тогда как список предопределенных переменных $(1, Y_t)$ не изменяется. Полный список содержит теперь 5 переменных: $(Q_t^d, Q_t^s, P_t, 1, Y_t)$. Перенесем все эндогенные переменные в левые части уравнений:

$$\begin{cases} Q_t^d - a_1 P_t = a_0 + a_2 Y_t + u_t, \\ Q_t^s - b_1 P_t = b_0 + v_t, \\ Q_t^d - Q_t^s = 0, \end{cases}$$

Матрица A имеет вид:

$$\mathrm{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ -a_1 & -b_1 & 0 \\ a_0 & b_0 & 0 \\ a_2 & 0 & 0 \end{pmatrix}.$$

На элементы первого столбца накладывается исключающее ограничение $\alpha_{21} = 0$, т.е. $\Phi_1 \alpha_1 = 0$, где $\Phi_1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \end{pmatrix}$. При этом

$$\operatorname{rank}(\Phi_1 \mathrm{A}) = \operatorname{rank}\begin{pmatrix} 0 & 1 & -1 \end{pmatrix} = 1 < g - 1 = 2,$$

так что первое уравнение неидентифицируемо. На элементы второго столбца накладывается $K_2^* = 2$ исключающих ограничения: $\alpha_{12} = 0$ и $\alpha_{52} = 0$, т.е. $\Phi_2 \alpha_2 = 0$, где

$$\Phi_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

При этом

$$\operatorname{rank}(\Phi_2 \mathrm{A}) = \operatorname{rank}\begin{pmatrix} 1 & 0 & 1 \\ a_2 & 0 & 0 \end{pmatrix} = 2 = g - 1,$$

так что второе уравнение идентифицируемо, причем идентифицируемо точно, поскольку $g_2^* + K_2^* = 2 = g - 1$. Результаты в отношении каждого из двух стохастических уравнений оказались одинаковыми для систем из трех и из двух уравнений.

До сих пор мы рассматривали только возможность восстановления коэффициентов структурных уравнений по коэффициентам приведенной формы. Однако идентифицируемость i-го стохастического структурного уравнения строго говоря означает не только идентифицируемость коэффициентов этого уравнения, но и идентифицируемость дисперсии случайной составляющей в этом уравнении. ***Идентифицируемость системы структурных уравнений в целом*** (на основании приведенной формы системы) означает не только идентифицируемость всех коэффициентов системы, но и идентифицируемость ковариационной матрицы случайных ошибок, входящих в правые части уравнений системы. При этом при восстановлении коэффициентов и ковариационной матрицы ошибок в структурной форме используются не только коэффициенты приведенной формы, но и ковариационная матрица ошибок в приведенной форме.

Обратимся опять к общей форме системы:

$$y_t \Gamma = x_t \mathrm{B} + u_t\,, \quad t = 1, \ldots, n\,,$$

где

$$\Gamma = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1g} \\ \vdots & \ddots & \vdots \\ \gamma_{g1} & \cdots & \gamma_{gg} \end{pmatrix}, \quad \mathrm{B} = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1g} \\ \vdots & \ddots & \vdots \\ \beta_{K1} & \cdots & \beta_{Kg} \end{pmatrix},$$

$$y_t = \left(y_{t1}, \ldots, y_{tg}\right), \quad x_t = \left(x_{t1}, \ldots, x_{tK}\right), \; u_t = \left(u_{t1}, \ldots, u_{tg}\right)$$

и предполагается невырожденность матрицы Γ. Приведенная форма системы:

$$y_t = x_t \mathrm{B}\Gamma^{-1} + u_t \Gamma^{-1} = x_t \Pi + w_t\,,$$

где

$$\Pi = \mathrm{B}\Gamma^{-1} = \begin{pmatrix} \pi_{11} & \dots & \pi_{1g} \\ \vdots & \ddots & \vdots \\ \pi_{K1} & \dots & \pi_{Kg} \end{pmatrix},\ w_t = u_t\Gamma^{-1} = \left(w_{t1},\dots,w_{tg}\right).$$

Пусть

$$E(u_t) = 0\,,\ Cov\left(u_t^T u_t\right) = \left(Cov\left(u_{ti}, u_{tj}\right)\right) = \Sigma = \left(\sigma_{ij}\right),$$

$$Cov\left(u_t^T u_s\right) = \left(Cov\left(u_{ti}, u_{sj}\right)\right) = 0 \quad \text{для} \quad t \neq s\,,$$

так что ошибки не коррелированы по времени, но для одного и того же момента времени ошибки в разных уравнениях могут быть коррелированными между собой. Тогда $E(w_t) = 0$ и для ковариационной матрицы $\Omega = \left(\omega_{ij}\right) = Cov\left(w_t^T w_t\right) = \left(Cov\left(w_{ti}, w_{tj}\right)\right)$ вектора w_t ошибок в приведенном уравнении имеем:

$$\Omega = Cov(w_t) = Cov\left(u_t\Gamma^{-1}\right) = \left(\Gamma^{-1}\right)^T \Sigma \left(\Gamma^{-1}\right),$$

так что

$$\Sigma = \Gamma^T \Omega\, \Gamma\,.$$

Следовательно, если структурная система идентифицируема (коэффициенты структурной системы однозначно восстанавливаются по коэффициентам приведенной формы), то тогда, восстановив по коэффициентам приведенной формы матрицу Γ, можно, используя эти восстановленные коэффициенты и матрицу Ω, восстановить ковариационную матрицу Σ.

Если структурная форма не восстанавливается целиком, а возможно лишь восстановление некоторых ее уравнений, то тогда для полной идентификации i-го стохастического структурного уравнения надо восстановить все его коэффициенты и дисперсию случайной составляющей этого уравнения. Пусть нас интересует, например, первое уравнение системы. Представим тогда матрицу Γ в виде

$$\Gamma = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1g} \\ \vdots & \ddots & \vdots \\ \gamma_{g1} & \cdots & \gamma_{gg} \end{pmatrix} = [\gamma_1 : \Gamma_1], \text{ где } \gamma_1 = \begin{pmatrix} \gamma_{11} \\ \vdots \\ \gamma_{g1} \end{pmatrix} \; \Gamma_1 = \begin{pmatrix} \gamma_{12} & \cdots & \gamma_{1g} \\ \vdots & \ddots & \vdots \\ \gamma_{g2} & \cdots & \gamma_{gg} \end{pmatrix}.$$

Дисперсия случайной составляющей в первом структурном уравнении равна $\sigma_{11} = \gamma_1^T \Omega \gamma_1$, так что для ее восстановления по приведенной форме достаточно предварительно восстановить только коэффициенты первого уравнения. Аналогично, если нас интересует i-е стохастическое структурное уравнение, то дисперсия случайной составляющей в этом структурном уравнении равна $\sigma_{ii} = \gamma_i^T \Omega \gamma_i$, где γ_i – i-й столбец матрицы Γ, и для восстановления σ_{ii} достаточно предварительно восстановить коэффициенты i-го уравнения.

В качестве примера рассмотрим опять структурную систему

$$\begin{cases} Q_t = a_0 + a_1 P_t + a_2 Y_t + u_t, \\ Q_t = b_0 + b_1 P_t + v_t. \end{cases}$$

Мы установили ранее, что в этой системе коэффициенты первого уравнения неидентифицируемы, а коэффициенты второго идентифицируемы точно. Для этой системы

$$\Gamma = \begin{pmatrix} 1 & 1 \\ -a_1 & -b_1 \end{pmatrix}, \quad \gamma_2 = \begin{pmatrix} 1 \\ -b_1 \end{pmatrix},$$

так что (поскольку ковариационные матрицы симметричны)

$$D(v_t) = \sigma_{22} = \begin{pmatrix} 1 & -b_1 \end{pmatrix} \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix} \begin{pmatrix} 1 \\ -b_1 \end{pmatrix} = \omega_{11} - 2b_1\omega_{12} + b_1^2\omega_{22}.$$

Оценив наряду с коэффициентами приведенной формы элементы ковариационной матрицы ошибок приведенной формы, можно получить оценку для коэффициента b_1, а через нее – и оценку для $D(v_t)$.

З а м е ч а н и е 3

Если посмотреть на все примеры, в которых на уравнения накладывались только исключающие ограничения, то нетрудно заметить, что проверку рангового условия идентифицируемости i-го стохастического структурного уравнения по-существу можно проводить следующим образом.

Составляется таблица, в заголовке которой перечисляются эндогенные и предопределенные переменные, задействованные в системе, а в i-й строке находятся коэффициенты при этих переменных в левой и правой частях i-го уравнения (как они есть, без переносов в левую часть). Например, для системы

$$\begin{cases} Q_t = a_0 + a_1 P_t + a_2 Y_t + u_t, \\ Q_t = b_0 + b_1 P_t + b_2 R_t + v_t \end{cases}$$

такая таблица принимает вид:

i	Q_t	P_t	1	Y_t	R_t
1	1	a_1	a_0	a_2	0
2	1	b_0	b_0	0	b_2

Для исследования i-го уравнения достаточно рассмотреть матрицу, образованную теми столбцами таблицы, элементы которых, стоящие в i-й строке, равны нулю, и всеми строками таблицы кроме i-й. В рассматриваемом примере при исследовании 1-го уравнения такая матрица состоит из единственного элемента b_2, а при исследовании 2-го уравнения – из единственного элемента a_2. В обоих случаях ранг выделенной матрицы равен 1, и поскольку $g-1=1$, оба уравнения идентифицируемы.

Для системы

$$\begin{cases} Q_t = a_0 + a_1 P_t + u_t, \\ Q_t = b_0 + b_1 P_t + b_2 R_t + b_3 S_t + v_t \end{cases}$$

указанная в Замечании 3 таблица имеет вид

i	Q_t	P_t	1	R_t	S_t
1	1	a_1	a_0	0	0
2	1	b_0	b_0	b_2	b_3

Для второго уравнения нет ни исключающих, ни других линейных ограничений – только нормирующее ограничение, так что второе уравнение нединтифицируемо. На коэффициенты первого уравнения помимо нормирующего накладываются только исключающие ограничения. Выделяемая матрица сводится к одной строке с двумя элементами: $(b_2 \quad b_3)$. Ранг этой матрицы равен 1, так что $g-1=1$ и первое уравнение идентифицируемо.

З а м е ч а н и е 4

В реальных ситуациях если порядковое условие выполнено, то, как правило, выполняется и ранговое условие. Приводимые в литературе контрпримеры носят явно искусственный характер. В качестве такого контрпримера выступает, например, система трех стохастических структурных уравнений

$$\begin{cases} a_{11}y_{t1} + a_{12}y_{t2} + a_{13}y_{t3} = a_{14}x_{t1} + a_{15}x_{t2} + u_{t1}, \\ a_{21}y_{t1} + a_{22}y_{t2} + a_{23}y_{t3} = a_{24}x_{t1} + a_{25}x_{t2} + u_{t2}, \\ a_{31}y_{t1} + a_{32}y_{t2} + a_{33}y_{t3} = a_{34}x_{t1} + a_{35}x_{t2} + u_{t3}, \end{cases}$$

в которой на коэффициенты первого уравнения накладываются линейные ограничения $a_{14}=0$, $a_{12}=a_{13}$. Эти ограничения записываются в стандартной форме как $\Phi_1 a_1 = 0$, где

$$\Phi_1 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 \end{pmatrix},$$

так что

$$\Phi_1 A = \begin{pmatrix} a_{14} & a_{24} & a_{34} \\ a_{12}-a_{13} & a_{22}-a_{23} & a_{32}-a_{33} \end{pmatrix} = \begin{pmatrix} 0 & a_{24} & a_{34} \\ 0 & a_{22}-a_{23} & a_{32}-a_{33} \end{pmatrix}.$$

Ранговое условие не выполняется, если строки этой матрицы пропорциональны. Последнее может осуществляться

- если на уравнения системы накладываются одинаковые ограничения;
- если переменная x_{t1} не входит в систему;
- если коэффициенты при y_{t2} и y_{t3} равны во всех уравнениях.

Замечание 5

До сих пор мы не предполагали никаких ограничений на ковариационную матрицу Σ вектора ошибок в структурной форме. Между тем введение ограничений на структуру этой матрицы в некоторых ситуациях может помочь идентификации уравнений, которые без таких ограничений неидентифицируемы. В качестве примера рассмотрим систему

$$\begin{cases} Q_t = a_1 P_t + a_2 Q_{t-1} + u_{t1}, \\ P_t = b_1 Q_{t-1} + u_{t2}. \end{cases}$$

Здесь P_t и Q_t – эндогенные переменные, а единственной предопределенной переменной является Q_{t-1}. При этом

$$\Gamma = \begin{pmatrix} 1 & 1 \\ -a_1 & 0 \end{pmatrix},\ \mathrm{B} = \begin{pmatrix} a_2 & b_1 \end{pmatrix},\ \Pi = \begin{pmatrix} \pi_{11} & \pi_{12} \end{pmatrix},$$

так что соотношение $\Pi\Gamma = \mathrm{B}$ принимает вид:

$$\begin{pmatrix} \pi_{11} & \pi_{12} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -a_1 & 0 \end{pmatrix} = \begin{pmatrix} a_2 & b_1 \end{pmatrix},$$

откуда $\begin{pmatrix} \pi_{11} - a_1\pi_{12} & \pi_{11} \end{pmatrix} = \begin{pmatrix} a_2 & b_1 \end{pmatrix}$, т.е. $a_2 = \pi_{11} - a_1\pi_{12}$, $b_1 = \pi_{11}$. Таким образом, единственный коэффициент второго уравнения восстанавливается по матрице Π однозначно, а для восстановления двух коэффициентов первого уравнения имеется только одно уравнение, и первое уравнение оказывается неидентифицируемым.

Вспомним, однако соотношение между ковариационными матрицами ошибок в приведенной и структурной формах:

$$\Sigma = \Gamma^T \Omega \Gamma .$$

В нашем примере оно принимает вид:

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} 1 & -a_1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \omega_{22} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -a_1 & 0 \end{pmatrix} =$$

$$= \begin{pmatrix} \omega_{11} - 2a_1\omega_{12} + a_1^2\omega_{22} & \omega_{11} - a_1\omega_{21} \\ \omega_{11} - a_1\omega_{12} & \omega_{11} \end{pmatrix}.$$

Если предположить дополнительно, что $\sigma_{12} = \sigma_{21} = 0$, т.е. ошибки в разных уравнениях не коррелированы между собой, то из последнего соотношения получаем:

$$\omega_{11} - a_1\omega_{21} = 0 ,$$

так что коэффициент a_1 первого структурного уравнения восстанавливается по матрице Ω: $a_1 = \omega_{11}/\omega_{21}$. После этого восстанавливается и коэффициент a_2 первого структурного уравнения: $a_2 = \pi_{11} - a_1\pi_{12}$. Тем самым оказывается идентифицируемым все первое уравнение структурной формы.

Замечание 6

Рассмотренная в Замечании 5 система

$$\begin{cases} Q_t = a_1 P_t + a_2 Q_{t-1} + u_{t1}, \\ P_t = b_1 Q_{t-1} + u_{t2} \end{cases}$$

при выполнении условия $\sigma_{12} = \sigma_{21} = 0$ принадлежит классу ***рекурсивных систем***. Благодаря последовательному определению переменных в таких системах при переходе от уравнения к уравнению в правых частях каждого из уравнений системы не оказывается переменных, значения которых коррелированы со значением ошибки в этом уравнении при одном и том же t. Во втором уравнении рассматриваемой системы $Cov(Q_{t-1}, u_{t2}) = 0$, т. к. значение Q_{t-1} определяется ранее момента t. В правой части первого уравнения $Cov(Q_{t-1}, u_{t1}) = 0$ по той же причине и

$$Cov(P_t, u_{t1}) = Cov(b_1 Q_{t-1} + u_{t2}) = b_1 Cov(Q_{t-1}, u_{t1}) + Cov(u_{t2}, u_{t1}) = 0,$$

так что при выполнении условия $\sigma_{12} = \sigma_{21} = 0$ переменная P_t не является эндогенной. Если же $\sigma_{12} \neq 0$, то P_t становится эндогенной переменной, а система перестает быть рекурсивной.

2.6. Оценивание систем одновременных уравнений

В этом разделе мы рассматриваем некоторые методы оценивания систем одновременных уравнений. Выбор того или иного метода оценивания связан с идентифицируемостью (неидентифицируемостью) системы в целом, идентифицируемостью отдельных уравнений системы, а также с имеющимися предположениями о вероятностной структуре случайных ошибок в правых частях структурных уравнений.

2.6.1. Косвенный метод наименьших квадратов

Если i-е стохастическое уравнение структурной формы идентифицируемо точно, то параметры этого уравнения (коэффициенты уравнения и дисперсия случайной ошибки) восстанавливаются по параметрам приведенной системы однозначно. Поэтому для оценивания параметров такого уравнения достаточно оценить методом наименьших квадратов коэффициенты каждого из уравнений приведенной формы методом наименьших квадратов (отдельно для каждого уравнения) и получить оценку ковариационной матрицы Ω ошибок в приведенной форме, после чего воспользоваться соотношениями $\Pi\Gamma = \mathrm{B}$ и $\Sigma = \Gamma^T \Omega \Gamma$, подставляя в них вместо Π оцененную матрицу коэффициентов приведенной формы $\hat{\Pi}$ и оцененную ковариационную матрицу ошибок в приведенной форме $\hat{\Omega}$. Такая процедура называется

косвенным методом наименьших квадратов (***ILS*** – *indirect least squares*). Полученные в результате оценки коэффициентов i-го стохастического уравнения структурной формы наследуют свойство состоятельности оценок приведенной формы. Однако они не наследуют таких свойств оценок приведенной формы как несмещенность и эффективность из-за того, что получаются в результате некоторых нелинейных преобразований. Соответственно, при небольшом количестве наблюдений даже у этих естественных оценок может возникать заметное смещение. В связи с этим при рассмотрении различных методов оценивания коэффициентов структурных уравнений в первую очередь заботятся об обеспечении именно состоятельности получаемых оценок.

2.6.2. Двухшаговый метод наименьших квадратов

Мы фактически уже воспользовались этим методом в разделе 2.3 при рассмотрении системы

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t, \\ Y_t = C_t + I_t. \end{cases}$$

Там мы подменили переменную Y_t в первом структурном уравнении искусственной инструментальной переменной $\hat{Y}_t = \hat{\gamma} + \hat{\delta}\, I_t$, где $\hat{\gamma}$ и $\hat{\delta}$ – оценки наименьших квадратов, получаемые при оценивании модели $Y_t = \gamma + \delta\, I_t + \varepsilon_t$. После такой подмены уравнение $C_t = \alpha + \beta \hat{Y}_t + \varepsilon_t$ состоятельно оценивается обычным методом наименьших квадратов, поскольку “объясняющая переменная” $\hat{Y}_t$ в этом уравнении не коррелирована с ε_t.

Пусть мы имеем систему g одновременных уравнений

$$y_t \Gamma = x_t \mathrm{B} + u_t\,, \quad t = 1, \ldots, n\,,$$

где

$$\Gamma = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1g} \\ \vdots & \ddots & \vdots \\ \gamma_{g1} & \cdots & \gamma_{gg} \end{pmatrix}, \quad \mathrm{B} = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1g} \\ \vdots & \ddots & \vdots \\ \beta_{K1} & \cdots & \beta_{Kg} \end{pmatrix},$$

$y_t = (y_{t1}, \ldots, y_{tg})$ – вектор значений эндогенных переменных в t-м наблюдении,

$x_t = (x_{t1}, \ldots, x_{tK})$ – вектор значений предопределенных переменных в t-м наблюдении,

$u_t = (u_{t1}, \ldots, u_{tg})$ – вектор значений случайных ошибок в t-м наблюдении, и при этом предполагается невырожденность матрицы Γ. Пусть наибольший интерес представляет первое уравнение системы. (Это не уменьшает общности, поскольку уравнения всегда можно надлежащим образом перенумеровать.) Считая, что первое уравнение нормировано на коэффициент при переменной y_{t1}, уединим эту переменную в левой части, преобразуя уравнение к виду:

$$y_{t1} = \alpha_{11} y^*_{t1} + \ldots + \alpha_{g_1,1} y^*_{tg_1} + \theta_{11} x^*_{t1} + \ldots + \theta_{K_1,1} x^*_{tK_1} + u_{t1},$$

или

$$y_{t1} = Y_{t1}\alpha_1 + X_{t1}\theta_1 + u_{t1},$$

где $Y_{t1} = (y^*_{t1}, \ldots, y^*_{tg_1})$ – вектор значений g_1 эндогенных переменных, включенных в правую часть первого уравнения, $X_{t1} = (x^*_{t1}, \ldots, x^*_{tK_1})$ – вектор значений K_1 предопределенных переменных, включенных в правую часть первого уравнения, α_1 и θ_1 – векторы коэффициентов при эндогенных и предопределенных переменных, включенных в первое уравнение. Состоятельному оцениванию коэффициентов уравнения мешает эндогенность переменных $y^*_{t1}, \ldots, y^*_{tg_1}$. Это затруднение преодолевается за два шага (отсюда название метода: ***двухшаговый метод наименьших квадратов***, **2SLS –** *two-step least squares*, *two-stage least squares*).

1. Производится оценивание уравнений регрессии каждой из этих переменных на все предопределенные переменные, включенные в систему. В качестве очищенных вариантов переменных $y_{t1}^*, \ldots, y_{tg_1}^*$ берутся предсказанные значения $\hat{y}_{t1}^*, \ldots, \hat{y}_{tg_1}^*$ этих переменных. (В этом контексте предопределенные переменные понимаются как инструменты для очистки эндогенных переменных.)
2. Значения $y_{t1}^*, \ldots, y_{tg_1}^*$ эндогенных переменных в первом уравнении заменяются значениями $\hat{y}_{t1}^*, \ldots, \hat{y}_{tg_1}^*$. Полученное уравнение оценивается методом наименьших квадратов.

Как и в случае оценивания обычным методом наименьших квадратов (OLS) единственного уравнения в линейной множественной регрессии, процедуру 2SLS можно представить в матричном виде. Для этого будем предполагать, что в левой части i-го стохастического структурного уравнения находится единственная эндогенная переменная y_{ti}, и обозначим:

$y_i = (y_{1i}, \ldots, y_{ni})^T$ – вектор-столбец значений i-й эндогенной переменной в n наблюдениях,

$$Y_i = \begin{pmatrix} y_{11}^* & \cdots & y_{1g_i}^* \\ \vdots & \ddots & \vdots \\ y_{n1}^* & \cdots & y_{ng_i}^* \end{pmatrix}$$ – матрица значений в n наблюдениях g_i

эндогенных переменных, включенных в правую часть i-го уравнения,

$$X_i = \begin{pmatrix} x_{11}^* & \dots & x_{1K_i}^* \\ \vdots & \ddots & \vdots \\ x_{n1}^* & \dots & x_{nK_i}^* \end{pmatrix}$$ – матрица значений в n наблюдениях K_i предопределенных переменных, включенных в правую часть i-го структурного уравнения,

$u_i = (u_{1i}, \dots, u_{ni})^T$ – вектор-столбец значений ошибки в i-м структурном уравнении в n наблюдениях,

$$X = \begin{pmatrix} x_{11} & \dots & x_{1K} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nK} \end{pmatrix}$$ – матрица значений в n наблюдениях всех K предопределенных переменных, включенных в систему,

$$\Pi = \begin{pmatrix} \pi_{11} & \dots & \pi_{1g} \\ \vdots & \ddots & \vdots \\ \pi_{K1} & \dots & \pi_{Kg} \end{pmatrix}$$ – матрица коэффициентов приведенной формы системы,

$$W = \begin{pmatrix} w_{11} & \dots & w_{1g} \\ \vdots & \ddots & \vdots \\ w_{n1} & \dots & w_{ng} \end{pmatrix}$$ – матрица значений в n наблюдениях ошибок в g уравнениях приведенной формы. (Заметим, во избежание путаницы, что $y_i = (y_{1i}, \dots, y_{ni})^T$ и $u_i = (u_{1i}, \dots, u_{ni})^T$ – векторы-столбцы, содержащие значения объясняемой переменной и случайных ошибок в i-м структурном уравнении в моменты времени $t = 1, \dots, n$. Их не следует путать с ранее использовавшимися векторами-строками $y_t = (y_{t1}, \dots, y_{tg})$ и $u_t = (u_{t1}, \dots, u_{tg})$, содержащими значения объясняемой переменной и случайных ошибок в g уравнениях, относящиеся к одному и тому же моменту времени t.)

Тогда ***первый шаг процедуры 2SLS*** оценивания i-го структурного уравнения состоит в оценивании методом наименьших квадратов отдельных уравнений системы

$$Y_i = X\Pi_i + W_i,$$

где Π_i – подматрица матрицы Π коэффициентов приведенной формы, образованная теми ее столбцами, которые соответствуют эндогенным переменным, включенным в правую часть i-го структурного уравнения, а W_i – подматрица матрицы W, образованная столбцами матрицы W, соответствующими тем же эндогенным переменным. Оценив отдельные уравнения методом наименьших квадратов, мы получаем оценку $\hat{\Pi}_i$ матрицы Π_i и на ее основе – оценку матрицы Y_i в виде $\hat{Y}_i = X\hat{\Pi}_i$. Матрица $\hat{Y}_i$ содержит значения эндогенных переменных, включенных в правую часть i-го структурного уравнения, "очищенных" от корреляции с ошибкой в этом уравнении.

Обозначая α_i и θ_i – векторы коэффициентов при эндогенных и предопределенных переменных, включенных в i-е структурное уравнение, запишем это уравнение в виде:

$$y_i = Y_i\alpha_i + X_i\theta_i + u_i = Z_i\delta_i + u_i,$$

где

$$\delta_i = \begin{pmatrix} \alpha_i \\ \theta_i \end{pmatrix}, \quad Z_i = [Y_i \quad X_i].$$

Мы также можем записать это уравнение в следующей форме:

$$y_i = Y_i\alpha_i + X_i\theta_i + u_i = \hat{Y}_i\alpha_i + X_i\theta_i + \left(\left(Y_i - \hat{Y}_i\right)\alpha_i + u_i\right),$$

или

$$y_i = \hat{Z}_i\delta_i + \varepsilon_i,$$

где

$$\hat{Z}_i = \left[\hat{Y}_i \quad X_i\right].$$

Второй шаг процедуры 2SLS состоит в вычислении оценки наименьших квадратов вектора δ_i в последнем уравнении. Эта оценка вычисляется по обычной формуле:

$$\hat{\delta}_i^{2SLS} = \left(\hat{Z}_i^T \hat{Z}_i^T\right)^{-1} \hat{Z}_i^T y_i \,.$$

При этом, естественно, требуется, чтобы матрица $\hat{Z}_i^T \hat{Z}_i^T$ была невырожденной. Заметим, что $\operatorname{rank}\left(\hat{Z}_i^T \hat{Z}_i^T\right) = \operatorname{rank} \hat{Z}_i^T \le \operatorname{rank} X$. Но матрица $\hat{Z}_i^T \hat{Z}_i^T$ имеет порядок, равный $g_i + K_i$, а $\operatorname{rank} X = K$, так что необходимо, чтобы $K \ge g_i + K_i$, т.е. $K - K_i \ge g_i = \left(g - g_i^*\right) - 1$, а это есть не что иное, как известное нам порядковое условие идентифицируемости. Для состоятельности $\hat{\delta}_i^{2SLS}$ требуется еще, чтобы предельная матрица

$$\underset{n\to\infty}{p\lim}\left(\frac{1}{n}\hat{Z}_i^T \hat{Z}_i^T\right) = Q_i$$

имела конечные элементы и была невырожденной, а для этого матрица Π_i должна иметь полный столбцовый ранг. Последнее же выполняется в случае идентифицируемости i-го уравнения (см. [Schmidt (1976), p. 150-151]).

Хотя на втором шаге используется OLS, непосредственное использование для построения t-статистик и доверительных интервалов вычисленных (по формулам OLS) значений стандартных ошибок коэффициентов невозможно. Эти значения должны быть скорректированы. Например, если дело касается первого уравнения, то на втором шаге оценивается уравнение

$$y_{t1} = \alpha_{11}\hat{y}_{t1}^* + \ldots + \alpha_{g_1,1}\hat{y}_{tg_1}^* + \theta_{11}x_{t1}^* + \ldots + \theta_{K_1,1}x_{tK_1}^* + \tilde{u}_{t1}\,,$$

где $\hat{y}_{t1}^*, \ldots, \hat{y}_{tg_1}^*$ – очищенные на первом шаге значения эндогенных переменных. При вычислении стандартных ошибок коэффициентов по обычным формулам OLS используется оценка дисперсии случайной ошибки в правой части уравнения в виде:

$$\tilde{S}^2 = \frac{\sum_{t=1}^{n}\left(y_{t1} - \hat{\alpha}_{11}\hat{y}_{t1}^* - \ldots - \hat{\alpha}_{g_1,1}\hat{y}_{tg_1}^* - \hat{\theta}_{11}x_{t1}^* - \ldots - \hat{\theta}_{K_1,1}x_{tK_1}^*\right)^2}{n - g_1 - K_1};$$

вместо этого следует использовать другую оценку:

$$S^2 = \frac{\sum_{t=1}^{n}\left(y_{t1} - \hat{\alpha}_{11}y_{t1}^* - \ldots - \hat{\alpha}_{g_1,1}y_{tg_1}^* - \hat{\theta}_{11}x_{t1}^* - \ldots - \hat{\theta}_{K_1,1}x_{tK_1}^*\right)^2}{n - g_1 - K_1},$$

в которой вместо “очищенных” значений $\hat{y}_{t1}^*, \ldots, \hat{y}_{tg_1}^*$ используются “сырые” значения $y_{t1}^*, \ldots, y_{tg_1}^*$.

В обеих формулах оценки параметров получены методом 2SLS; для сокращения записи в обозначениях этих оценок опущен верхний индекс, указывающий на 2SLS (например, $\hat{\alpha}_{11}$ вместо $\hat{\alpha}_{11}^{2SLS}$). В специализированных программах статистического анализа систем одновременных уравнений такая коррекция предусмотрена.

Замечание 1

На втором шаге 2SLS не следует особенно полагаться на указываемые в распечатках значения t-статистик, поскольку если данных мало, то асимптотическая теория неприменима и эти статистики не имеют ни нормального, ни t-распределения. Напротив, статистики, получаемые на первом шаге, имеют t-распределения при нормальном распределении ошибок. Однако они не предназначены для выяснения значимости отдельных коэффициентов.

Замечание 2

Вернемся к рассмотренной ранее системе

$$\begin{cases} Q_t = a_0 + a_1 P_t + u_t, \\ Q_t = b_0 + b_1 P_t + b_2 R_t + b_3 S_t + v_t, \end{cases}$$

в которой, как мы установили, первое уравнение сверхидентифицируемо. Следуя методу инструментальных переменных, мы могли бы произвести “очистку” эндогенной переменной P_t в правой части первого уравнения, используя в качестве инструмента только одну из переменных R_t и S_t. В 2SLS в качестве инструментов для очистки P_t используются сразу обе эти переменные. Очистка с использованием только одной из этих переменных также приводит к состоятельной оценке, но эта оценка менее эффективна (ее асимптотическая дисперсия больше, чем у оценки, полученной с применением пары инструментальных переменных). В этом смысле 2SLS оценка является ***наилучшей инструментальной оценкой*** среди этих трех альтернатив.

З а м е ч а н и е 3 (***Проверка адекватности***)

Получив в результате применения двухшагового метода наименьших квадратов к i-му структурному уравнению оценку $\hat{\delta}_i^{2SLS}$, мы получаем далее оцененное значение $\hat{y}_i^{2SLS} = Z_i \hat{\delta}_i^{2SLS}$ вектора y_i и вектор остатков $\hat{u}_i^{2SLS} = y_i - \hat{y}_i^{2SLS}$. (Заметим, что эти остатки отличаются от остатков, непосредственно получаемых на втором шаге 2SLS и равных $y_i - \hat{Z}_i \hat{\delta}_i^{2SLS}$.) Опираясь на эти остатки, можно обычным образом проверять адекватность этого уравнения, используя критерии:

- нормальности (Харке–Бера, по оцененным асимметрии и куртозису последовательности остатков);
- линейности (добавляя степени и перекрестные произведения предопределенных переменных и проверяя гипотезу зануления коэффициентов при “лишних” составляющих);
- гомоскедастичности (Уайта, Пагана–Холла);
- независимости – против автокоррелированности остатков (Бройша–Годфри).

При обнаружении нарушений стандартных предположений необходимо соответствующим образом изменить спецификацию модели или, не изменяя спецификации, скорректировать статистические выводы.

Замечание 3a

Критерий Пагана–Холла предназначен специально для выявления гетероскедастичности в отдельном уравнении системы. В отличие от других критериев гомоскедастичности, он не предполагает отсутствия гетероскедастичности в других уравнениях системы. Этот критерий реализован, например, в пакете Stata.

2.6.3. GLS-оценивание систем одновременных уравнений. Трехшаговый метод наименьших квадратов

Если нас интересует оценивание коэффициентов всех g структурных уравнений, и каждое из уравнений идентифицируемо (идентифицируемо точно или сверхидентифицируемо), то тогда мы можем сразу получить OLS-оценку $\hat{\Pi}$ матрицы коэффициентов приведенной формы системы и на ее основе сформировать подматрицы $\hat{\Pi}_1,\ldots,\hat{\Pi}_g$, соответствующие эндогенным переменным, включенным в отдельные уравнения структурной формы

$$y_i = Y_i\alpha_i + X_i\theta_i + u_i = Z_i\delta_i + u_i\,,\ \ i = 1,\ldots,g\,.$$

После этого можно вычислить $\hat{Y}_i = X\hat{\Pi}_i$ и получить 2SLS-оценку $\hat{\delta}_i^{2SLS}$ для δ_i, применяя OLS к уравнению $y_i = \hat{Z}_i\delta_i + \varepsilon_i$.

Совокупность оценок $\hat{\delta}_i^{2SLS}$, $i = 1,\ldots,g$, в форме вектора

$$\hat{\delta}^{2SLS} = \begin{pmatrix} \hat{\delta}_1^{2SLS} \\ \vdots \\ \hat{\delta}_g^{2SLS} \end{pmatrix}$$

фактически получается как OLS-оценка уравнения регрессии

$$\begin{pmatrix} y_1 \\ \vdots \\ y_g \end{pmatrix} = \begin{pmatrix} \hat{Z}_1 & \mathrm{O} & \dots & \mathrm{O} \\ \mathrm{O} & \hat{Z}_2 & \dots & \mathrm{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{O} & \mathrm{O} & \dots & \hat{Z}_g \end{pmatrix} \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_g \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_g \end{pmatrix},$$

в сокращенной форме

$$y = \hat{Z}\delta + \varepsilon,$$

где

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_g \end{pmatrix},\ \hat{Z} = \begin{pmatrix} \hat{Z}_1 & \mathrm{O} & \dots & \mathrm{O} \\ \mathrm{O} & \hat{Z}_2 & \dots & \mathrm{O} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{O} & \mathrm{O} & \dots & \hat{Z}_g \end{pmatrix},\ \delta = \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_g \end{pmatrix},\ \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_g \end{pmatrix}.$$

Соответственно, вектор $\hat{\delta}^{2SLS}$ находится по обычной формуле:

$$\hat{\delta}^{2SLS} = \left(\hat{Z}^T \hat{Z}\right)^T \hat{Z}^T y.$$

Эта оценка неэффективна вследствие коррелированности ошибок $\varepsilon_{t1}, \dots, \varepsilon_{tg}$ (имеющей место даже при некоррелированности ошибок $u_{t1}, \dots, u_{tg}$ в структурных уравнениях) и различия матриц $\hat{Z}_1, \dots, \hat{Z}_g$. Эффективную оценку можно было бы получить, используя здесь вместо OLS *обобщенный метод наименьших квадратов*, *(GLS)*, но для этого надо знать ковариационную матрицу вектора ε. Поскольку же эта матрица неизвестна, мы можем довольствоваться только ее оценкой, и такая оценка должна быть состоятельной, если мы хотим получить в итоге асимптотически эффективную оценку вектора δ.

Заметим, что ковариационная матрица вектора ε при наших предположениях имеет весьма специфический вид:

$$Cov(\varepsilon) = Cov\begin{pmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{n1} \\ \vdots \\ \varepsilon_{1g} \\ \vdots \\ \varepsilon_{ng} \end{pmatrix} = \begin{pmatrix} \lambda_{11}I_g & \lambda_{12}I_g & \dots & \lambda_{1g}I_g \\ \lambda_{21}I_g & \lambda_{22}I_g & \dots & \lambda_{2g}I_g \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{g1}I_g & \lambda_{g2}I_g & \dots & \lambda_{gg}I_g \end{pmatrix},$$

или

$$Cov(\varepsilon) = \Lambda \otimes I_g ,$$

где

I_g – единичная матрица порядка g ,

$\lambda_{ij} = Cov(\varepsilon_{ti}, \varepsilon_{tj})$,

$\Lambda = (\lambda_{ij})$ – ковариационная матрица вектора $(\varepsilon_{t1}, \dots, \varepsilon_{tg})$,

$\Lambda \otimes I_g$ – формализованное обозначение матрицы, являющейся ***кронекеровским произведением матриц*** Λ и I_g .

Таким образом, для реализации доступного GLS необходимо состоятельно оценить ковариации λ_{ij} . Это можно сделать, используя для λ_{ij} естественную оценку

$$\hat{\lambda}_{ij} = \frac{1}{n}\left(\hat{u}_i^{2SLS}\right)^T \hat{u}_j^{2SLS} .$$

Заменяя в выражении для $Cov(\varepsilon)$ истинные значения λ_{ij} их оценками $\hat{\lambda}_{ij}$, получаем состоятельную оценку ковариационной матрицы $Cov(\varepsilon)$ в виде $\hat{\Lambda} \otimes I_g$. В результате приходим к доступной обобщенной оценке наименьших квадратов (***FGLS*** – *feasible generalized least squares*)

$$\hat{\delta}^{3SLS} = \left(\hat{Z}^T\left(\hat{\Lambda}^{-1} \otimes I_g\right)\hat{Z}\right)^{-1} \hat{Z}^T\left(\hat{\Lambda}^{-1} \otimes I_g\right)y ,$$

известной под названием ***трехшаговой оценки наименьших квадратов***, или ***3SLS-оценки*** (*three-stage least squares*).

2.6.4. Оценивание систем одновременных уравнений с использованием метода максимального правдоподобия

Запишем совокупность g одновременных уравнений для n наблюдений в виде:

$$Y\,\Gamma = X\,\mathrm{B} + U\ ,$$

где

$$Y = \begin{pmatrix} y_{11} & \cdots & y_{1g} \\ \vdots & \ddots & \vdots \\ y_{n1} & \cdots & y_{ng} \end{pmatrix}$$ – матрица значений в n наблюдениях всех g эндогенных переменных, включенных в систему,

$$X = \begin{pmatrix} x_{11} & \dots & x_{1K} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nK} \end{pmatrix}$$ – матрица значений в n наблюдениях всех K предопределенных переменных, включенных в систему,

$$U = \begin{pmatrix} u_{11} & \dots & u_{1g} \\ \vdots & \ddots & \vdots \\ u_{n1} & \cdots & u_{ng} \end{pmatrix}$$ – матрица значений в n наблюдениях случайных ошибок в g структурных уравнениях системы,

$$\Gamma = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1g} \\ \vdots & \ddots & \vdots \\ \gamma_{g1} & \dots & \gamma_{gg} \end{pmatrix}, \qquad \mathrm{B} = \begin{pmatrix} \beta_{11} & \cdots & \beta_{1g} \\ \vdots & \ddots & \vdots \\ \beta_{K1} & \dots & \beta_{Kg} \end{pmatrix}$$ – матрицы коэффициентов.

Сделаем следующие предположения относительно векторов $u_t = \left(u_{t1}, \dots, u_{tg}\right)$ ошибок в g уравнениях в t-м наблюдении:

- векторы $u_1,\ldots,u_n$ имеют одинаковое g-мерное нормальное распределение $N_g(0,\Sigma)$ с нулевым вектором математических ожиданий и положительно определенной ковариационной матрицей Σ;
- векторы $u_1,\ldots,u_n$ независимы между собой.

При таких предположениях совместная плотность распределения случайных векторов $u_1,\ldots,u_n$ имеет вид:

$$p_U(u_1,\ldots,u_n)=\prod_{t=1}^{n}\left[\frac{1}{\left(\sqrt{2\pi}\right)^g\sqrt{\det\Sigma}}\cdot\exp\left\{-\frac{1}{2}u_t\Sigma^{-1}u_t^T\right\}\right].$$

Поскольку $u_t=y_t\Gamma-x_t\mathrm{B}$, то переходя от переменных $u_1,\ldots,u_n$ к переменным $y_1,\ldots,y_n$, получаем выражение для совместной плотности значений векторов $y_1,\ldots,y_n$ в виде:

$$p_Y(y_1,\ldots,y_n)=|J|\cdot\left(\left(\sqrt{2\pi}\right)^g\sqrt{\det\Sigma}\right)^{-n}\times$$
$$\times\exp\left\{-\frac{1}{2}(y_t\,\Gamma-x_t\,\mathrm{B})\Sigma^{-1}(y_t\,\Gamma-x_t\,\mathrm{B})^T\right\},$$

где $|J|$ – якобиан перехода от переменных $u_1,\ldots,u_n$ к переменным $y_1,\ldots,y_n$, $|J|=|\det\Gamma|^n$. Для взаимной однозначности такого перехода требуется, чтобы $|\det\Gamma|\neq 0$. Рассматривая правую часть последнего выражения как функцию от неизвестных параметров, составляющих матрицы Γ,B,Σ, при известных значениях Y,X, получаем функцию правдоподобия

$$L(\Gamma,\mathrm{B},\Sigma\,|\,Y,X)=|\det\Gamma|^n\cdot\left(\left(\sqrt{2\pi}\right)^g\sqrt{\det\Sigma}\right)^{-n}\times$$
$$\times\exp\left\{-\frac{1}{2}\sum_{t=1}^{n}(y_t\,\Gamma-x_t\,\mathrm{B})\Sigma^{-1}(y_t\,\Gamma-x_t\,\mathrm{B})^T\right\},$$

логарифм которой равен

$$\ln L\left(\Gamma, \mathrm{B}, \Sigma \,\middle|\, Y, X\right) = n \ln\left|\det \Gamma\right| - \frac{n}{2}\ln\left(\det \Sigma\right) - \frac{ng}{2}\ln\left(2\pi\right) -$$

$$-\frac{1}{2}\sum_{t=1}^{n}\left(y_t\,\Gamma - x_t\,\mathrm{B}\right)\Sigma^{-1}\left(y_t\,\Gamma - x_t\,\mathrm{B}\right)^T .$$

Продифференцировав последнее выражение по элементам матрицы Σ^{-1} и приравнивая найденные производные нулю, приходим к соотношению:

$$\Sigma = n^{-1}\left(Y\,\Gamma - X\,\mathrm{B}\right)^T\left(Y\,\Gamma - X\,\mathrm{B}\right).$$

Подставляя полученное выражение для Σ в правую часть выражения для $\ln L\left(\Gamma, \mathrm{B}, \Sigma \,\middle|\, Y, X\right)$ и отбрасывая слагаемые, не зависящие от неизвестных параметров, получаем ***концентрированную*** логарифмическую функцию правдоподобия

$$\ln L^*\left(\Gamma, \mathrm{B}\right) = n \ln\left|\det \Gamma\right| - \frac{n}{2}\ln\left|n^{-1}\left(Y\,\Gamma - X\,\mathrm{B}\right)^T\left(Y\,\Gamma - X\,\mathrm{B}\right)\right|.$$

Оценки матриц Γ и B находятся максимизацией $\ln L^*$ с учетом всех ограничений, накладываемых на коэффициенты структурных уравнений. В частности, если используются только исключающие и нормировочные ограничения, то максимизация проводится только по неспецифицированным элементам этих матриц. В процессе такой максимизации матрица $\left(Y\,\Gamma - X\,\mathrm{B}\right)^T\left(Y\,\Gamma - X\,\mathrm{B}\right)$ должна иметь полный ранг для всех допустимых значений Γ и B. Необходимым условием для этого является условие $n \ge g + K$. Такое условие может быть ограничительным для систем с большим количеством переменных.

Матрицы $\hat{\Gamma}$ и $\hat{\mathrm{B}}$, получаемые в результате максимизации $\ln L^*$, и матрица $\hat{\Sigma} = n^{-1}\left(Y\,\hat{\Gamma} - X\,\hat{\mathrm{B}}\right)^T\left(Y\,\hat{\Gamma} - X\,\hat{\mathrm{B}}\right)$ вместе образуют оценку максимального правдоподобия, учитывающую полную информацию о структуре модели одновременных уравнений. Поэтому такую оценку называют ***оценкой максимального правдоподобия с полной информацией*** (***FIML*** – *full information maximum likelihood*).

Пусть

- выполнены сделанные выше предположения,
- ранговое условие идентифицируемости выполняется для всех структурных уравнений системы,
- матрица X имеет полный ранг,
- предельная матрица $\underset{n\to\infty}{p\lim}\left(\frac{1}{n}X^T X\right)=Q$ имеет конечные элементы и положительно определена.

Тогда FIML-оценка состоятельна и асимптотически нормальна. При этом требование нормальности распределения векторов $u_1,\ldots,u_n$ не обязательно. (См., например, [Amemiya (1985), глава 7].)

Замечание 4

Если при выводе FIML отправляться не от структурной, а от приведенной формы системы, учитывающей ограничения, накладываемые на коэффициенты структурной формы, то тогда дело сводится к максимизации концентрированной функции правдоподобия

$$\ln L^{**}(\Gamma,\mathrm{B})=-\frac{n}{2}\ln\left|n^{-1}\left(Y-X\,\mathrm{B}\Gamma^{-1}\right)^T\left(Y-X\,\mathrm{B}\Gamma^{-1}\right)\right|,$$

т.е., с учетом соотношения $\Pi=\mathrm{B}\Gamma^{-1}$, к минимизации целевой функции

$$Q(\Gamma,\mathrm{B})=\widetilde{Q}(\Pi)=(Y-X\,\Pi)^T(Y-X\,\Pi)$$

по элементам матрицы Π с учетом ограничений, накладываемых на коэффициенты этой матрицы выбранной спецификацией матриц Γ и B.

Если не учитывать эти ограничения при минимизации целевой функции $\widetilde{Q}(\Pi)$, то при сверхидентифицируемости отдельных уравнений системы возникает неоднозначность восстановления Γ и B по полученной оценке $\hat{\Pi}$. Если же все уравнения системы идентифицируемы точно, то значения Γ и B восстанавливаются однозначно и при этом восстановленные значения $\hat{\Gamma}$ и $\hat{\mathrm{B}}$ совпадают с оценками, полученными при минимизации $Q(\Gamma, \mathrm{B})$ по Γ и B.

Замечание 5

При практической реализации метода FIML приходится использовать итерационные процедуры. Для обеспечения состоятельности и асимптотической нормальности FIML-оценки в качестве начальных значений параметров необходимо использовать их состоятельные оценки. Их можно получить двухшаговым методом наименьших квадратов. Если система неидентифицируема, то итерационный процесс может не сходиться.

Замечание 6

Перед применением FIML обычно производят исключение из системы тождеств и недоидентифицируемых уравнений.

Замечание 7

В рекурсивной системе с диагональной ковариационной матрицей Σ оценка FIML получается применением OLS отдельно к каждому уравнению.

Пусть первое стохастическое структурное уравнение

$$y_1 = Y_1\alpha_1 + X_1\theta_1 + u_1$$

идентифицируемо, а другие уравнения либо неидентифицируемы либо имеются сомнения в правильности их спецификации. Пусть при этом известен список всех предопределенных переменных, включаемых в систему, и значения этих предопределенных переменных в n наблюдениях, так что известна матрица X этих значений и можно говорить о приведенной форме системы:

$$Y = X\Pi + W .$$

Удалим из приведенной формы часть, относящуюся к y_1; оставшаяся часть принимает вид:

$$Y_1 = X\Pi_1 + W_1.$$

Вместо полной системы структурных уравнений или полной приведенной системы рассмотрим теперь смешанную систему:

$$\begin{cases} y_1 = Y_1\alpha_1 + X_1\theta_1 + u_1, \\ Y_1 = X\Pi_1 + W_1. \end{cases}$$

Эту систему можно записать в виде

$$\begin{bmatrix} y_1 & Y_1 \end{bmatrix} \begin{pmatrix} 1 & \mathrm{O} \\ -\alpha_1 & I_{g-1} \end{pmatrix} = \begin{bmatrix} X_1 & X_1^* \end{bmatrix} \left(\begin{matrix} \theta_1 \\ 0 \end{matrix} \middle| \Pi_1 \right) + \begin{bmatrix} u_1 \\ W_1 \end{bmatrix}$$

и применить к ней метод FIML. Такая процедура приводит к оценке параметров α_1, θ_1, Π, называемой ***оценкой максимального правдоподобия с ограниченной информацией*** (***LIML*** – *limited information maximum likelihood*).

При практической реализации этой процедуры итерационными методами в качестве начальных значений следует использовать состоятельные оценки параметров $\alpha_1, \theta_1, \Pi_1$. Состоятельную оценку матрицы Π_1 можно получить непосредственным применением метода наименьших квадратов к уравнениям системы $Y_1 = X\Pi_1 + W_1$. Состоятельные оценки параметров α_1 и θ_1 можно получить, применяя двухшаговый метод наименьших квадратов.

2.6.5. Связь между различными оценками систем одновременных уравнений

Пусть интерес представляет оценивание ***отдельного уравнения*** (*single equation*) структурной формы системы. Тогда:

- если оно идентифицируемо без запаса (точно), то достаточно применить косвенный метод наименьших квадратов ILS;
- если оно сверхидентифицируемо, то можно применить 2SLS, LIML, 3SLS, FIML.

Для применимости 3SLS и FIML необходимо знать структуру всех уравнений системы и убедиться в идентифицируемости всех этих уравнений. Для применимости 2SLS и LIML достаточно знать только структуру рассматриваемого уравнения и список (и значения) всех предопределенных переменных, включенных в систему. В 2SLS и LIML ошибка спецификации одного структурного уравнения системы локализуется в пределах этого уравнения; в 3SLS и FIML такая ошибка влияет на оценку всех уравнений.

Предположим теперь, что выполнены указанные ранее условия состоятельности оценок. Тогда:

- если i-е структурное уравнение идентифицируемо точно, то оценки 2SLS, LIML и ILS совпадают; если же оно сверхидентифицируемо, то тогда оценки 2SLS и LIML имеют одинаковое асимптотическое распределение, но оценка LIML предпочтительнее при малом количестве наблюдений;
- если i-е структурное уравнение идентифицируемо, то 2SLS дает состоятельные оценки параметров и $\sqrt{n}\left(\hat{\delta}_i^{2SLS} - \delta_i\right) \xrightarrow{d} N\left(0, C_2\right)$;
- если все структурные уравнения идентифицируемы, то 3SLS дает состоятельные оценки параметров и $\sqrt{n}\left(\hat{\delta}_i^{3SLS} - \delta_i\right) \xrightarrow{d} N\left(0, C_3\right)$, причем матрица $C_2 - C_3$ неотрицательно определена (положительно полуопределена), так что 3SLS приводит к более эффективным оценкам;
- если $\Sigma = I_g$ и все структурные уравнения идентифицируемы точно, то $\hat{\delta}_i^{3SLS} = \hat{\delta}_i^{2SLS}$;
- оценки FIML и 3SLS имеют одинаковое асимптотическое распределение; при конечных n предпочтительнее FIML.

З а м е ч а н и е 8

Если в i-м структурном уравнении системы $Y\Gamma = X\,\mathrm{B} + U$ ошибки автокоррелированы, то для учета этой автокоррелированности можно использовать комбинацию 2SLS и GLS, не прибегая к 3SLS. Пусть, например, ошибки в i-м уравнении следуют процессу авторегрессии первого порядка,

$$u_{ti} = \rho u_{t-1,i} + \eta_t\,,\ |\rho| < 1\,.$$

Тогда естественно применить к i-му уравнению авторегрессионное преобразование (Кохрейна–Оркатта). Состоятельную оценку для ρ можно получить, оценивая обычным методом наименьших квадратов (OLS) уравнение

$$\hat{u}_{ti}^{IV} = \rho\,\hat{u}_{t-1,i}^{IV} + \nu_t\,,$$

где $\hat{u}_{ti}^{IV}$ – остатки, полученные в результате применения к i-му уравнению метода инструментальных переменных. При этом для повышения эффективности оценивания к используемым в качестве инструментов в 2SLS переменным $x_t = (x_{t1},\ldots,x_{tK})$ можно добавить $y_{t-1} = (y_{t-1,1},\ldots,y_{t-1,g})$ и $x_{t-1} = (x_{t-1,1},\ldots,x_{t-1,K})$.

2.6.6. Проверка правильности спецификации системы одновременных уравнений

Мы уже говорили выше (Замечание 3 в разд. 2.6.2) о возможности проверки адекватности i-го структурного уравнения, опираясь на остатки $\hat{u}_i^{2SLS} = y_i - Z_i\hat{\delta}_i^{2SLS}$, полученные в результате применения двухшагового метода наименьших квадратов к этому уравнению, в отношении таких стандартных предположений как линейность уравнения, нормальность, гомоскедастичность и некоррелированность ошибок.

Между тем не менее важным является вопрос о правильности подразделения включенных в систему переменных на эндогенные и экзогенные переменные, произведенного на основании соответствующих экономических и логических представлений о связях между переменными. Еще одна проблема спецификации структурных уравнений состоит в том, что сверхидентифицируемость i -го уравнения системы

$$y_i = Y_i\alpha_i + X_i\theta_i + u_i = Z_i\delta_i + u_i$$

может быть просто следствием того, что на коэффициенты этого уравнения наложены ограничения, которых в действительности нет. Например, из i -го уравнения могут быть ошибочно исключены некоторые предопределенные переменные, включенные в другие уравнения системы. Хотелось бы иметь какой-то статистический инструментарий, позволяющий ответить на такие вопросы. Ряд статистических критериев, служащих этой цели, использует следующую идею Хаусмана [Hausman (1978)].

Пусть для $(p\times 1)$ -вектора параметров θ имеются две различные оценки $\hat{\theta}$ и $\tilde{\theta}$, причем оценка $\tilde{\theta}$ состоятельна и при гипотезе H_0 и при альтернативной гипотезе H_A, а оценка $\hat{\theta}$ состоятельна и асимптотически эффективна при гипотезе H_0, но не является состоятельной при гипотезе H_A. Рассмотрим разность этих двух оценок $\hat{q} = \hat{\theta} - \tilde{\theta}$. Поскольку при гипотезе H_0 обе оценки состоятельны, т.е. сходятся по вероятности к истинному значению θ, то их разность $\hat{q}$ сходится по вероятности к нулю. Следовательно, если гипотеза H_0 верна, то мы не ожидаем больших отклонений значения $\hat{q}$ от нуля, и наличие таковых может трактоваться как указание на невыполнение гипотезы H_0.

Критерий Хаусмана для ***проверки правильности спецификации системы одновременных уравнений*** ([Hausman (1978)]) использует в качестве $\hat{\theta}$ трехшаговую оценку наименьших квадратов, а в качестве $\tilde{\theta}$ – двухшаговую оценку наименьших квадратов. Если все структурные уравнения специфицированы правильно, то 3SLS состоятельна и эффективна; если же хотя бы одно из уравнений специфицировано неправильно, то 3SLS перестает быть состоятельной оценкой.

Однако, как было отмечено в [Spencer, Berk (1981)], для применения этого критерия необходима спецификация всех структурных уравнений системы, тогда как на практике чаще представляет интерес правильность спецификации какого-то отдельного структурного уравнения. В таком случае речь идет о проверке правильности спецификации i-го структурного уравнения при ограниченной информации об остальной части системы (как при построении LIML оценки).

Статистика критерия Хаусмана определяется как

$$H = n\,\hat{q}^T\,[as\hat{Cov}(\hat{q})]^{-1}\,\hat{q}\,,$$

где $as\hat{Cov}(\hat{q})$ – состоятельная оценка асимптотической ковариационной матрицы $asCov(\hat{q})$ разности $\hat{q} = \hat{\theta} - \tilde{\theta}$, и при выполнении достаточно общих условий гипотезе H_0 имеет асимптотическое распределение хи-квадрат. Некоторые трудности при использовании этой статистики вызывает то обстоятельство, что компоненты вектора $\hat{q} = \hat{\theta} - \tilde{\theta}$ в общем случае линейно зависимы, вследствие чего матрица $asCov(\hat{q})$ может быть вырожденной и не иметь обратной в обычном смысле. В связи с этим, в формуле для статистики H следует использовать не обычную обратную матрицу, а так называемую обобщенную обратную матрицу.

Указанные трудности можно обойти, используя различные асимптотически эквивалентные версии критерия Хаусмана, основанные на оценивании тех или иных уравнений регрессии. В таких вариантах этого критерия дело сводится к проверке значимости оцененных коэффициентов соответствующих уравнений.

Версия критерия Хаусмана, приведенная в [Davidson, MacKinnon (1993)] и называемая там ***критерием Дарбина–Ву–Хаусмана*** (Durbin–Wu–Hausman test), состоит в следующем.

Пусть $y_i = Y_i\alpha_i + X_i\theta_i + u_i = Z_i\delta_i + u_i$, X – матрица значений инструментальных переменных, Y_i – матрица значений тех объясняющих переменных в i-м уравнении, которые не входят в состав инструментальных переменных и не являются линейными комбинациями последних. Гипотеза H_0: в i-м уравнении отсутствует проблема эндогенности, т.е. все объясняющие переменные в составе Y_i не коррелированы с u_i. Иначе говоря, это ***гипотеза экзогенности*** (предопределенности) переменных, входящих в состав Y_i. Если эта гипотеза выполнена, то оценивание i-го уравнения можно производить обычным методом наименьших квадратов (OLS). В противном случае надо применять метод инструментальных переменных.

Сначала производится OLS оценивание уравнений регрессии объясняющих переменных, входящих в состав Y_i, на инструментальные переменные:

$$Y_i = X\,\Pi_i + W_i$$

и вычисляются прогнозные значения $\hat{Y}_i$ этих переменных. Затем эти прогнозные значения добавляются в качестве дополнительных объясняющих переменных в правую часть i-го уравнения, что приводит к расширенному уравнению

$$y_i = Z_i\delta_i + \hat{Y}_i\,\gamma + \eta_i\,,$$

производится OLS оценивание расширенного уравнения и проверяется гипотеза H_0: $\gamma = 0$. Для проверки этой гипотезы используется обычный F-критерий, хотя, вообще говоря, он является в этой ситуации только приближенным критерием.

Вместо $\hat{Y}_i$ в расширенном уравнении можно использовать остатки $\hat{W}_i = Y_i - \hat{Y}_i$, т.е. оценивать уравнение

$$y_i = Z_i\delta_i + \hat{W}_i\,\gamma + \eta_i$$

и проверять гипотезу $H_0:\ \gamma = 0$ в рамках этого уравнения. В любом случае отклонение гипотезы H_0 трактуется как наличие проблемы эндогенности, вызывающей несостоятельность OLS оценок параметров i-го структурного уравнения.

Еще один вариант критерия Хаусмана для проверки той же гипотезы состоит в следующем.

Наряду с остатками $\hat{W}_i = Y_i - \hat{Y}_i$, определенными выше, рассмотрим остатки $\hat{u}_i$, получаемые при оценивании i-го уравнения обычным методом наименьших квадратов (OLS). Пусть R^2 – коэффициент детерминации, получаемый при OLS оценивании уравнения $\hat{u}_i = Z_i\delta_i + \hat{W}_i\,\gamma + \xi_i$. Тогда при выполнении гипотезы экзогенности статистика nR^2 имеет асимптотическое (при $n \to \infty$) распределение $\chi^2(g_i)$, где g_i – количество переменных в составе Y_i. Эта гипотеза отвергается при $nR^2 > \chi^2_{1-\alpha}(g_i)$, где α – выбранный уровень значимости критерия.

Можно указать и некоторые другие варианты реализации критерия Хаусмана для проверки гипотезы об отсутствии проблемы эндогенности в i-м уравнении. Но как бы там ни было, прежде чем производить проверку тех или иных переменных, включенных в структурное уравнение, на эндогенность, рекомендуется предварительно провести ***проверку пригодности самих выбранных инструментов***. Такую проверку можно провести в том случае, когда количество имеющихся инструментов превышает их необходимое количество, и сделать это можно, используя, например, ***J-статистику***, предложенную в работе [Godfrey, Hutton (1994)].

Пусть для очистки эндогенных переменных, входящих в правую часть i-го уравнения системы

$$y_i = Y_i\alpha_i + X_i\theta_i + u_i = Z_i\delta_i + u_i$$

используется уравнение

$$Y_i = X\,\Pi_i + W_i\,,$$

где X – матрица значений инструментальных переменных.

Применив к i-му уравнению двухшаговый метод наименьших квадратов, получим 2SLS-остатки в виде

$$\hat{u}_i^{2SLS} = y_i - Z_i\hat{\delta}_i^{2SLS}\,.$$

После этого оценим линейную модель регрессии $\hat{u}_i^{2SLS}$ на переменные, входящие в состав X. Пусть R^2 – полученное при этом значение коэффициента детерминации. Указанная J-статистика равна $J = nR^2$ и имеет асимптотическое (при $n \to \infty$) распределение хи-квадрат с числом степеней свободы, равным разности между количеством переменных в составе X и количеством объясняющих переменных в первом уравнении.

Гипотеза пригодности выбранного множества инструментов отвергается при значениях J-статистики, превышающих критическое значение, рассчитанное по указанному хи-квадрат распределению (т.е. при значениях J-статистики, для которых P-значение оказывается меньше заданного уровня значимости). Если это происходит, то тогда нет смысла заниматься IV оцениванием коэффициентов рассматриваемого уравнения с выбранным множеством инструментов, поскольку в этом случае или сами эти инструменты непригодны или уравнение неправильно специфицировано.

Если указанная гипотеза не отвергается J-критерием, то тогда переходят ко второму шагу, на котором используется критерий Хаусмана (в том или ином его варианте) для проверки переменных в i-м уравнении системы на эндогенность/экзогенность.

В работе [Godfrey, Hutton (1994)] показано, что статистики, используемые в такой двухступенчатой процедуре, асимптотически независимы, так что вероятность ошибочного решения в этой процедуре приближенно равна $1-(1-\alpha_J)(1-\alpha_H) = \alpha_J + \alpha_H - \alpha_J\alpha_H$, где α_J – уровень значимости J-критерия, а α_H – уровень значимости критерия Хаусмана, используемого на втором шаге.

З а м е ч а н и е 8

Отклонение нулевой гипотезы при применении критериев экзогенности означает только, что проблема эндогенности существует. Однако степень влияния обнаруженной эндогенности на смещение обычных оценок наименьших квадратов остается при этом неизвестной. Вместе с тем, мощность критериев типа Хаусмана становится довольно низкой, если инструменты слабо коррелированы с эндогенными переменными. И это означает, что нулевая гипотеза экзогенности может быть не отвергнута, а смещение OLS оценок в то же время велико. Поэтому во многих практических исследованиях авторы сообщают и результаты IV-оценивания и результаты OLS-оценивания.

2.6.7. Примеры оценивания систем одновременных уравнений.

П р и м е р 1

Рассмотрим модель спроса-предложения в виде:

$$\begin{cases} P = \alpha_1 Q + \theta_{11} + \theta_{12} DPI + u_1, \\ Q = \alpha_2 P + \theta_{21} + \theta_{22} Weather + \theta_{23} Invest + u_2, \end{cases}$$

где

P – розничная цена свежих фруктов, выраженная в постоянных ценах с использованием индекса розничных цен,

Q – потребление свежих фруктов на душу населения,

DPI – располагаемый доход на душу населения, дефлированный на индекс потребительских цен (CPI),

$Weather$ – климатическая характеристика, отражающая размер потенциальных потерь урожая из-за неблагоприятных погодных условий,

$Invest$ – дефлированный на CPI объем на душу населения чистых инвестиций производителей свежих фруктов, отражающий издержки производства.

Первое уравнение является уравнением спроса, а второе – уравнением предложения.

Всего имеется 30 наблюдений; все переменные выражены в индексной форме с одним и тем же базовым периодом.

	Price (P)	Quantity (Q)	DPI	Weather	Invest
1	108.9	127.4	97.6	99.1	142.9
2	100.6	105.1	98.2	98.9	123.8
3	109.7	76.7	99.8	110.8	111.9
4	111.6	93.8	100.5	108.2	121.4
5	109.8	88.3	96.6	108.7	92.9
6	104.4	78.4	88.9	100.6	97.6
7	89.6	89.6	84.6	70.9	64.3
8	117.2	75.3	96.4	110.5	78.6
9	109.3	109.1	104.4	92.5	109.5
10	114.9	121.3	110.7	89.3	128.6
11	112.0	106.3	99.1	90.3	95.8
12	112.9	129.1	105.6	95.2	130.9
13	121.0	118.6	116.8	98.6	125.7
14	112.8	94.3	105.3	105.7	109.8
15	102.9	81.0	85.6	107.8	88.4
16	86.0	104.9	84.8	80.4	96.9
17	95.7	94.6	89.8	90.7	90.8
18	104.9	102.9	93.2	88.9	101.7
19	114.0	110.6	105.9	96.9	110.8
20	121.9	111.7	110.8	101.9	117.9
21	127.2	117.6	115.3	104.9	134.8
22	128.3	125.1	120.6	103.6	140.2
23	125.0	87.4	105.7	106.2	78.3
24	117.1	84.6	103.5	100.8	94.7
25	122.7	107.8	110.6	110.5	135.9
26	111.6	120.7	109.3	86.7	126.8
27	114.1	102.8	99.5	93.8	90.5
28	110.4	99.2	105.9	99.9	134.8
29	109.2	107.1	102.7	104	123.8
30	108.9	127.4	97.6	99.1	142.9

Переходя к обозначениям, использованным ранее при рассмотрении систем одновременных уравнений, запишем систему в виде:

$$\begin{cases} y_{t1} = \alpha_{11} y_{t2} + \theta_{11} x_{t1} + \theta_{21} x_{t2} + u_{t1}, \\ y_{t2} = \alpha_{12} y_{t1} + \theta_{12} x_{t1} + \theta_{22} x_{t3} + \theta_{32} x_{t4} + u_{t2}, \end{cases}$$

где $y_{t1} = P_t$, $y_{t2} = Q_t$, $x_{t1} \equiv 1$, $x_{t2} = (DPI)_t$, $x_{t3} = (Weather)_t$, $x_{t4} = (Invest)_t$. Список эндогенных переменных: (y_{t1}, y_{t2}). Список экзогенных переменных: $(1, x_{t2}, x_{t3}, x_{t4})$. Полный список переменных, включенных в систему: $(y_{t1}, y_{t2}, 1, x_{t2}, x_{t3}, x_{t4})$. Соответственно, $g = 2$, $K = 4$,

$$\Gamma = \begin{pmatrix} 1 & -\alpha_{12} \\ -\alpha_{11} & 1 \end{pmatrix}, \quad \mathrm{B} = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & 0 \\ 0 & \theta_{22} \\ 0 & \theta_{32} \end{pmatrix}, \quad \mathrm{A} = \begin{pmatrix} \Gamma \\ \mathrm{B} \end{pmatrix} = \begin{pmatrix} 1 & -\alpha_{12} \\ -\alpha_{11} & 1 \\ \theta_{11} & \theta_{12} \\ \theta_{21} & 0 \\ 0 & \theta_{22} \\ 0 & \theta_{32} \end{pmatrix}.$$

На элементы первого столбца матрицы A помимо нормировочного накладывается два исключающих ограничения $\alpha_{51} = 0$, $\alpha_{61} = 0$, так что для этого столбца $g_1^* = 0$, $K_1^* = 2$, и $g_1^* + K_1^* > g - 1$, т.е. порядковое условие идентифицируемости выполняется. На элементы второго столбца помимо нормировочного накладывается одно исключающее ограничение $\alpha_{42} = 0$, так что для этого столбца $g_2^* = 0$, $K_2^* = 1$, и $g_2^* + K_2^* = g - 1$, т.е. порядковое условие идентифицируемости выполняется.

Для проверки выполнения ранговых условий идентифицируемости воспользуемся Замечанием 3 из разд. 2.5. В соответствии с этим замечанием построим таблицу коэффициентов:

i	y_{t1}	y_{t2}	1	x_{t2}	x_{t3}	x_{t4}
1	1	α_{11}	θ_{11}	θ_{21}	0	0
2	α_{12}	1	θ_{12}	0	θ_{22}	θ_{32}

При рассмотрении первого уравнения выделяемая матрица сводится к одной строке с двумя элементами: $(\theta_{22} \quad \theta_{32})$. Ранг этой матрицы равен 1, что совпадает со значением $g-1=1$, так что первое уравнение идентифицируемо. При рассмотрении второго уравнения выделяемая матрица сводится к одному элементу: (θ_{21}). Ранг этой матрицы равен также равен 1, так что и второе уравнение идентифицируемо. Разница только в том, что для первого уравнения $g_1^* + K_1^* > g-1$, а для второго $g_2^* + K_2^* = g-1$, т.е. первое уравнение сверхидентифицируемо, а второе идентифицируемо точно. Соответственно, для оценивания второго уравнения можно использовать косвенный метод наименьших квадратов, а для оценивания первого уравнения этот метод не годится.

Чтобы применить косвенный метод наименьших квадратов, сначала раздельно оценим методом наименьших квадратов уравнения приведенной формы

$$y_{t1} = \pi_{11} + \pi_{21}x_{t2} + \pi_{31}x_{t3} + \pi_{41}x_{t4} + w_{t1},$$
$$y_{t2} = \pi_{12} + \pi_{22}x_{t2} + \pi_{32}x_{t3} + \pi_{42}x_{t4} + w_{t2}.$$

Это дает следующие результаты (в пакете EVIEWS):

Dependent Variable: Y1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-12.40954	8.192675	-1.514712	0.1419
X2	1.030854	0.090988	11.32951	0.0000
X3	0.361564	0.066508	5.436388	0.0000
X4	-0.152442	0.040203	-3.791820	0.0008

R-squared	0.902361	Mean dependent var	111.1445
Adjusted R-squared	0.891094	S.D. dependent var	9.877858
S.E. of regression	3.259777	Akaike info criterion	5.324760
Sum squared resid	276.2797	Schwarz criterion	5.511587
Log likelihood	-75.87140	F-statistic	80.09524
Durbin-Watson stat	2.016289	Prob(F-statistic)	0.000000

Dependent Variable: Y2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	81.78495	17.56752	4.655463	0.0001
X2	0.581396	0.195106	2.979896	0.0062
X3	-0.924096	0.142613	-6.479734	0.0000
X4	0.475229	0.086207	5.512656	0.0000

R-squared	0.824120	Mean dependent var	101.8111
Adjusted R-squared	0.803826	S.D. dependent var	15.78165
S.E. of regression	6.989927	Akaike info criterion	6.850383
Sum squared resid	1270.336	Schwarz criterion	7.037209
Log likelihood	-98.75575	F-statistic	40.60943
Durbin-Watson stat	2.084533	Prob(F-statistic)	0.000000

Используем теперь соотношение $\Pi\Gamma = \mathrm{B}$, которое в нашем примере принимает вид

$$\begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \\ \pi_{31} & \pi_{32} \\ \pi_{41} & \pi_{42} \end{pmatrix} \begin{pmatrix} 1 & -\alpha_{12} \\ -\alpha_{11} & 1 \end{pmatrix} = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & 0 \\ 0 & \theta_{22} \\ 0 & \theta_{32} \end{pmatrix}$$

и приводит к уравнениям:

$$\pi_{11} - \pi_{12}\alpha_{11} = \theta_{11}, \quad \pi_{12} - \pi_{11}\alpha_{12} = \theta_{12},$$
$$\pi_{21} - \pi_{22}\alpha_{11} = \theta_{21}, \quad \pi_{22} - \pi_{21}\alpha_{12} = 0,$$
$$\pi_{31} - \pi_{32}\alpha_{11} = 0, \quad \pi_{32} - \pi_{31}\alpha_{12} = \theta_{22},$$
$$\pi_{41} - \pi_{42}\alpha_{11} = 0, \quad \pi_{42} - \pi_{41}\alpha_{12} = \theta_{32}.$$

Поскольку точно идентифицируемо только второе структурное уравнение системы, интерес для применения косвенного метода наименьших квадратов представляют только коэффициенты этого уравнения $\alpha_{12}, \theta_{12}, \theta_{22}$ и θ_{32}. Это означает, что из восьми приведенных уравнений достаточно рассмотреть только четыре уравнения, стоящие в правом столбце. Решая эти уравнения, находим:

$$\alpha_{12} = \pi_{22}/\pi_{21},$$

$$\theta_{12} = \pi_{11} - \pi_{12}(\pi_{22}/\pi_{21}),$$
$$\theta_{22} = \pi_{31} - \pi_{32}(\pi_{22}/\pi_{21}),$$
$$\theta_{32} = \pi_{41} - \pi_{42}(\pi_{22}/\pi_{21}).$$

Подставляя в правые части оцененные значения коэффициентов π_{ki}, находим оценки для коэффициентов второго структурного уравнения. Например, $\hat{\alpha}_{12} = \hat{\pi}_{22}/\hat{\pi}_{21} = 0.581396/1.030854 = 0.5639945$. В отношении трех остальных коэффициентов получаем: $\hat{\theta}_{12} = 88.78386$, $\hat{\theta}_{22} =$ -1.128017, $\hat{\theta}_{32} = 0.561206$.

Замечание

В таблицах результатов применения косвенного метода наименьших квадратов обычно не приводятся значения стандартных ошибок коэффициентов, поскольку из-за нелинейности соотношений между коэффициентами структурной и приведенной форм вычисление стандартных ошибок оценок коэффициентов при конечных n затруднительно. В то же время при применении двухшагового метода наименьших квадратов для вычисления этих ошибок имеются соответствующие формулы. Поэтому мы могли бы вычислить искомые стандартные ошибки оценок коэффициентов первого уравнения рассматриваемой системы, используя 2SLS и имея в виду, что в случае точно идентифицируемого уравнения результаты оценивания его коэффициентов методами ILS и 2SLS совпадают. Проблема, однако, в том, что (см., например, [Sawa (1969)]) в этой ситуации у 2SLS оценки не существует конечных выборочных моментов. Соответственно, по сравнению с нормальным распределением, оценки более часто далеко отклоняются от истинных значений параметров, и это затрудняет интерпретацию полученных результатов.

Имея в виду сделанное замечание, применим все же двухшаговый метод наименьших квадратов для оценивания обоих структурных уравнений. Результаты применения этого метода таковы:

System: FRU
Estimation Method: Two-Stage Least Squares

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.361831	0.081657	-4.431122	0.0000
C(2)	18.16639	9.760145	1.861283	0.0683
C(3)	1.279190	0.126018	10.15089	0.0000
C(4)	0.563994	0.175307	3.217171	0.0022
C(5)	88.78386	15.06578	5.893080	0.0000
C(6)	-1.128017	0.158217	-7.129565	0.0000
C(7)	0.561206	0.065442	8.575557	0.0000
Determinant residual covariance		479.5253		

Equation: Y1=C(1)*Y2+C(2)+C(3)*X2
Observations: 30

R-squared	0.781184	Mean dependent var	111.1445
Adjusted R-squared	0.764975	S.D. dependent var	9.877858
S.E. of regression	4.788722	Sum squared resid	619.1603
Durbin-Watson stat	2.036078		

Equation: Y2=C(4)*Y1+C(5)+C(6)*X3+C(7)*X4
Observations: 30

R-squared	0.849107	Mean dependent var	101.8111
Adjusted R-squared	0.831696	S.D. dependent var	15.78165
S.E. of regression	6.474401	Sum squared resid	1089.865
Durbin-Watson stat	2.152230		

Оценки всех коэффициентов кроме постоянной составляющей в первом уравнении имеют высокую статистическую значимость. Отрицательное значение оценки коэффициента при переменной y_{t2} в первом уравнении согласуется с тем, что первое уравнение является уравнением спроса. Положительное значение оценки при переменной y_{t1} во втором уравнении согласуется с тем, что второе уравнение является уравнением предложения. Также соответствуют априорным предположениям знаки оцененных коэффициентов при переменных x_{t2}, x_{t3} и x_{t4} (увеличение спроса при возрастании дохода, уменьшение предложения при усилении неблагоприятных погодных факторов и увеличение предложения при возрастании инвестиций).

Вместе с тем, если обратиться к оцениванию корреляционной матрицы ошибок в структурных уравнениях, то оцененная на основании векторов остатков $\hat{u}_i^{2SLS} = y_i - Z_i\hat{\delta}_i^{2SLS}$ ковариационная матрица имеет вид:

$$\begin{pmatrix} 20.638675 & 16.439387 \\ 16.439387 & 36.328821 \end{pmatrix}.$$

Ей соответствует оцененная корреляционная матрица

$$\begin{pmatrix} 1 & 0.600370 \\ 0.600370 & 1 \end{pmatrix},$$

указывающая на наличие заметной корреляции между ошибками в разных уравнениях. Это означает, что потенциально имеется возможность повысить эффективность оценивания, учитывая такую коррелированность и применяя трехшаговый метод наименьших квадратов или метод максимального правдоподобия с полной информацией.

При применении 3SLS получаем:

System: FRU
Estimation Method: Iterative Three-Stage Least Squares
Convergence achieved after: 2 weight matricies, 3 total coef iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.361831	0.077466	-4.670812	0.0000
C(2)	18.16639	9.259287	1.961964	0.0550
C(3)	1.279190	0.119551	10.69998	0.0000
C(4)	0.592909	0.150576	3.937596	0.0002
C(5)	89.91484	13.79846	6.516294	0.0000
C(6)	-1.159822	0.130071	-8.916811	0.0000
C(7)	0.550297	0.056107	9.808057	0.0000
Determinant residual covariance		486.0481		

Equation: Y1=C(1)*Y2+C(2)+C(3)*X2
Observations: 30

R-squared	0.781184	Mean dependent var	111.1445
Adjusted R-squared	0.764975	S.D. dependent var	9.877858
S.E. of regression	4.788722	Sum squared resid	619.1603
Durbin-Watson stat	2.036078		

Equation: Y2=C(4)*Y1+C(5)+C(6)*X3+C(7)*X4
Observations: 30

R-squared	0.849356	Mean dependent var	101.8111
Adjusted R-squared	0.831974	S.D. dependent var	15.78165
S.E. of regression	6.469047	Sum squared resid	1088.063
Durbin-Watson stat	2.139746		

Сравним оцененные коэффициенты и оцененные стандартные ошибки оценок коэффициентов, полученные двумя методами:

	Coefficients		Std. Errors	
	2SLS	3SLS	2SLS	3SLS
C(1)	-0.361831	-0.361831	0.081657	0.077466
C(2)	18.16639	18.16639	9.760145	9.259287
C(3)	1.279190	1.279190	0.126018	0.119551
C(4)	0.563994	0.592909	0.175307	0.150576
C(5)	88.78386	89.91484	15.06578	13.79846
C(6)	-1.128017	-1.159822	0.158217	0.130071
C(7)	0.561206	0.550297	0.065442	0.056107

Оцененные значения коэффициентов практически не изменились. При этом произошло некоторое уменьшение всех оцененных стандартных ошибок коэффициентов.

Результаты применения метода FIML:

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.363540	0.097355	-3.734157	0.0005
C(2)	18.18156	13.65099	1.331886	0.1886
C(3)	1.280771	0.167147	7.662534	0.0000
C(4)	0.593625	0.221715	2.677418	0.0099
C(5)	89.82550	19.84547	4.526248	0.0000
C(6)	-1.159165	0.151139	-7.669539	0.0000
C(7)	0.549823	0.065905	8.342675	0.0000
Log Likelihood		-172.0962		
Determinant residual covariance		486.9397		

Equation: Y1=C(1)*Y2+C(2)+C(3)*X2
Observations: 30

R-squared	0.780336	Mean dependent var	111.1445
Adjusted R-squared	0.764065	S.D. dependent var	9.877858
S.E. of regression	4.797990	Sum squared resid	621.5590
Durbin-Watson stat	2.030828		

Equation: Y2=C(4)*Y1+C(5)+C(6)*X3+C(7)*X4
Observations: 30

R-squared	0.849367	Mean dependent var	101.8111
Adjusted R-squared	0.831986	S.D. dependent var	15.78165
S.E. of regression	6.468814	Sum squared resid	1087.984
Durbin-Watson stat	2.137757		

Оцененные значения коэффициентов незначительно отличаются от полученных ранее.

В связи с рассматриваемым примером обратим внимание на следующее обстоятельство. Попробуем оценить уравнения структурной формы обычным методом наименьших квадратов, игнорируя наличие в правых частях этих уравнений эндогенных переменных. Это приводит к следующим результатам:

System: SYS_FULL_OLS
Estimation Method: Least Squares

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.213716	0.063693	-3.355431	0.0015
C(2)	18.15974	8.908678	2.038432	0.0465
C(3)	1.130662	0.108276	10.44241	0.0000
C(4)	0.614038	0.159561	3.848293	0.0003
C(5)	86.61460	14.71164	5.887487	0.0000
C(6)	-1.154140	0.153404	-7.523517	0.0000
C(7)	0.553841	0.064458	8.592225	0.0000

Determinant residual covariance 517.6508

Equation: Y1=C(1)*Y2+C(2)+C(3)*X2
Observations: 30

R-squared	0.817697	Mean dependent var	111.1445
Adjusted R-squared	0.804193	S.D. dependent var	9.877858
S.E. of regression	4.370958	Sum squared resid	515.8424
Durbin-Watson stat	2.442191		

Equation: Y2=C(4)*Y1+C(5)+C(6)*X3+C(7)*X4
Observations: 30

R-squared	0.849675	Mean dependent var	101.8111
Adjusted R-squared	0.832330	S.D. dependent var	15.78165
S.E. of regression	6.462188	Sum squared resid	1085.757
Durbin-Watson stat	2.138293		

Сравним оценки коэффициентов, полученные при применении системного метода оценивания FIML и несистемного OLS:

Coefficient	
FIML	OLS
-0.363540	-0.213716
18.18156	18.15974
1.280771	1.130662
0.593625	0.614038
89.82550	86.61460
-1.159165	-1.154140
0.549823	0.553841

За исключением коэффициента при переменной y_{t2} в первом уравнении, оценки всех остальных коэффициентов структурных уравнений, полученные двумя методами, довольно близки друг к другу. И это характерно для ситуаций, в которых при оценивании уравнений приведенной формы получаются высокие значения коэффициентов детерминации. В нашем примере это значения 0.902361 для первого уравнения и 0.824120 для второго уравнения приведенной формы.

Рассмотрим теперь, что получится, если во второе уравнение не включена одна из переменных x_{t3}, x_{t4}. (Система остается при этом идентифицируемой.) При этом будем ориентироваться на результаты применения 3SLS. Сначала исключим переменную x_{t4} – издержки производства. Это приводит к следующему результату:

System: SYS_EXACT_2_3
Estimation Method: Iterative Three-Stage Least Squares
Convergence achieved after: 1 weight matrix, 2 total coef iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.393021	0.106679	-3.684137	0.0005
C(2)	18.16779	9.612186	1.890079	0.0641
C(3)	1.310468	0.142630	9.187882	0.0000
C(4)	1.603271	0.324620	4.938909	0.0000
C(5)	68.85896	29.34434	2.346584	0.0226
C(6)	-1.469719	0.308619	-4.762247	0.0000
Determinant residual covariance		3322.482		

Equation: Y1=C(1)*Y2+C(2)+C(3)*X2
Observations: 30

R-squared	0.764186	Mean dependent var	111.1445
Adjusted R-squared	0.746719	S.D. dependent var	9.877858
S.E. of regression	4.971235	Sum squared resid	667.2557
Durbin-Watson stat	1.946538		

Equation: Y2=C(4)*Y1+C(5)+C(6)*X3
Observations: 30

R-squared	0.360021	Mean dependent var	101.8111
Adjusted R-squared	0.312615	S.D. dependent var	15.78165
S.E. of regression	13.08436	Sum squared resid	4622.411
Durbin-Watson stat	1.199501		

Сравним коэффициенты, полученные при исключении x_{t4} из второго уравнения, с коэффициентами, полученными 3SLS без исключения этой переменной из второго уравнения:

	X4 excluded	X4 included
C(1)	-0.393021	-0.361831
C(2)	18.16779	18.16639
C(3)	1.310468	1.279190
C(4)	**1.603271**	**0.592909**
C(5)	68.85896	89.91484
C(6)	-1.469719	-1.159822

Произошло более чем двукратное возрастание коэффициента при переменной y_{t1} во втором уравнении. Обозревая таблицу результатов, можно заметить довольно низкое значение статистики Дарбина–Уотсона для второго уравнения, что может указывать на пропуск существенной объясняющей переменной в этом уравнении. Отметим также на довольно большое значение оцененной стандартной ошибки случайной составляющей во втором уравнении – 13.08436 против 4.971235, что, конечно, вполне возможно, но также может указывать на пропуск объясняющей переменной во втором уравнении.

Если вместо переменной x_{t4} исключить из второго уравнения переменную x_{t3} – погодный фактор, то это приводит к следующему результату:

System: SYS_EXACT_2_4
Estimation Method: Three-Stage Least Squares

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.318323	0.116740	-2.726758	0.0086
C(2)	18.16444	8.863637	2.049321	0.0453
C(3)	1.235561	0.145848	8.471570	0.0000
C(4)	0.274776	0.265732	1.034037	0.3057
C(5)	16.26065	26.12243	0.622478	0.5362
C(7)	0.499415	0.113699	4.392423	0.0001
Determinant residual covariance		2157.560		

Equation: Y1=C(1)*Y2+C(2)+C(3)*X2
Observations: 30

R-squared	0.799484	Mean dependent var	111.1445
Adjusted R-squared	0.784631	S.D. dependent var	9.877858
S.E. of regression	4.584099	Sum squared resid	567.3771
Durbin-Watson stat	2.166955		

Equation: Y2=C(4)*Y1+C(5)+C(7)*X4
Observations: 30

R-squared	0.490954	Mean dependent var	101.8111
Adjusted R-squared	0.453247	S.D. dependent var	15.78165
S.E. of regression	11.66938	Sum squared resid	3676.710
Durbin-Watson stat	1.624247		

Сравним коэффициенты, полученные при исключении x_{t3} из второго уравнения с коэффициентами, полученными 3SLS без исключения этой переменной из второго уравнения:

	X3 excluded	X3 included
C(1)	-0.318323	-0.361831
C(2)	18.16444	18.16639
C(3)	1.235561	1.279190
C(4)	**0.274776**	**0.592909**
C(5)	16.26065	89.91484
C(6)		-1.159822
C(7)	0.499415	0.550297

На этот раз произошло более чем двукратное уменьшение коэффициента при переменной y_{t1} во втором уравнении. Опять наблюдается довольно большое значение оцененной стандартной ошибки случайной составляющей во втором уравнении – 11.66938 против 4.584099 в первом уравнении, что может указывать на пропуск объясняющей переменной во втором уравнении. Заметим, что при включении в правую часть второго уравнения обеих переменных x_{t3}, x_{t4} такого большого различия оцененных стандартных ошибок случайных составляющих не наблюдается: в первом уравнении оцененное значение равно 4.788722, а во втором 6.469047; при этом вполне удовлетворительными выглядят и значения статистик Дарбина–Уотсона: 2.036078 в первом и 2.139746 во втором уравнениях.

П р и м е р 2

Чтобы составить некоторое представление о возможности более или менее точного восстановления параметров структурной формы, сгенерируем 30 новых “наблюдений”, используя процесс порождения данных в виде:

$$\begin{cases} y_{t1} = -0.5y_{t2} + 80 + 0.7x_{t2} + u_{t1}, \\ y_{t2} = 0.75y_{t1} + 10 - 1.5x_{t3} + 1.5x_{t4} + u_{t2}, \end{cases}$$

где переменные имеют ту же интерпретацию и значения переменных x_{t2}, x_{t3} и x_{t4} те же, что и в предыдущем примере, а случайные составляющие u_{t1} и u_{t2} не коррелированы во времени, имеют нулевые ожидания, одинаковые дисперсии $D(u_{t1}) = D(u_{t2}) = 36$ и $Corr(u_{t1}, u_{t2}) = 0.7$. Полученные “данные” приведены в таблице:

Y1	Y2	X2	X3	X4
88.1	152.4	97.6	99.1	142.9
87.1	114.6	98.2	98.9	123.8
101.1	74.1	99.8	110.8	111.9
98.6	99.2	100.5	108.2	121.4
109.4	74.0	96.6	108.7	92.9

105.2	72.4	88.9	100.6	97.6
100.6	69.2	84.6	70.9	64.3
121.7	51.3	96.4	110.5	78.6
96.0	109.7	104.4	92.5	109.5
90.3	135.5	110.7	89.3	128.6
105.7	100.2	99.1	90.3	95.8
91.3	144.5	105.6	95.2	130.9
95.5	126.1	116.8	98.6	125.7
101.0	90.6	105.3	105.7	109.8
110.2	66.7	85.6	107.8	88.4
87.1	104.0	84.8	80.4	96.9
97.4	85.3	89.8	90.7	90.8
99.8	102.6	93.2	88.9	101.7
99.9	110.8	105.9	96.9	110.8
102.8	114.8	110.8	101.9	117.9
99.8	131.2	115.3	104.9	134.8
95.8	141.2	120.6	103.6	140.2
123.4	62.6	105.7	106.2	78.3
110.4	73.2	103.5	100.8	94.7
98.7	121.1	110.6	110.5	135.9
88.2	134.5	109.3	86.7	126.8
110.0	91.8	99.5	93.8	90.5
88.3	115.1	105.9	99.9	134.8
93.6	114.5	102.7	104	123.8
103.5	76.9	96.8	108.4	104.5

Оценим по сгенерированным данным уравнения структурной формы обычным методом наименьших квадратов, игнорируя наличие в правых частях этих уравнений эндогенных переменных. Это приводит к следующим результатам:

System: SYS_FULL_OLS
Estimation Method: Least Squares

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.363218	0.041816	-8.686139	0.0000
C(2)	88.69833	10.49401	8.452281	0.0000
C(3)	0.476608	0.122697	3.884420	0.0003
C(4)	1.019773	0.297976	3.422330	0.0012
C(5)	-11.07961	28.30156	-0.391484	0.6970

C(6)	-1.752307	0.214654	-8.163409	0.0000
C(7)	1.672787	0.127403	13.12991	0.0000

Determinant residual covariance 576.6877

Equation: Y1=C(1)*Y2+C(2)+C(3)*X2
Observations: 30

R-squared	0.743523	Mean dependent var	100.0167
Adjusted R-squared	0.724524	S.D. dependent var	9.334551
S.E. of regression	4.899311	Sum squared resid	648.0877
Durbin-Watson stat	2.154319		

Equation: Y2=C(4)*Y1+C(5)+C(6)*X3+C(7)*X4
Observations: 30

R-squared	0.950061	Mean dependent var	102.0033
Adjusted R-squared	0.944299	S.D. dependent var	27.23976
S.E. of regression	6.428860	Sum squared resid	1074.586
Durbin-Watson stat	2.151388		

Применим двухшаговый метод наименьших квадратов. Результаты применения этого метода:

System: SYS_FULL_OLS
Estimation Method: Two-Stage Least Squares

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.424633	0.045795	-9.272577	0.0000
C(2)	83.95884	10.96205	7.659048	0.0000
C(3)	0.585040	0.130033	4.499177	0.0000
C(4)	0.901844	0.420366	2.145379	0.0365
C(5)	-1.113488	37.81363	-0.029447	0.9766
C(6)	-1.684157	0.274838	-6.127827	0.0000
C(7)	1.628247	0.169687	9.595614	0.0000

Determinant residual covariance 522.6707

Equation: Y1=C(1)*Y2+C(2)+C(3)*X2
Observations: 30

R-squared	0.723032	Mean dependent var	100.0167
Adjusted R-squared	0.702516	S.D. dependent var	9.334551
S.E. of regression	5.091263	Sum squared resid	699.8659
Durbin-Watson stat	1.952744		

Equation: Y2=C(4)*Y1+C(5)+C(6)*X3+C(7)*X4
Observations: 30

R-squared	0.949761	Mean dependent var	102.0033
Adjusted R-squared	0.943964	S.D. dependent var	27.23976
S.E. of regression	6.448196	Sum squared resid	1081.060
Durbin-Watson stat	2.162986		

При применении 3SLS получаем:

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.424633	0.043445	-9.774155	0.0000
C(2)	83.95884	10.39951	8.073345	0.0000
C(3)	0.585040	0.123360	4.742549	0.0000
C(4)	0.889189	0.369342	2.407495	0.0196
C(5)	-0.653169	34.88671	-0.018723	0.9851
C(6)	-1.671481	0.220626	-7.576096	0.0000
C(7)	1.624186	0.152419	10.65606	0.0000
Determinant residual covariance		518.5160		

Equation: Y1=C(1)*Y2+C(2)+C(3)*X2
Observations: 30

R-squared	0.723032	Mean dependent var	100.0167
Adjusted R-squared	0.702516	S.D. dependent var	9.334551
S.E. of regression	5.091263	Sum squared resid	699.8659
Durbin-Watson stat	1.952744		

Equation: Y2=C(4)*Y1+C(5)+C(6)*X3+C(7)*X4
Observations: 30

R-squared	0.949688	Mean dependent var	102.0033
Adjusted R-squared	0.943883	S.D. dependent var	27.23976
S.E. of regression	6.452837	Sum squared resid	1082.617
Durbin-Watson stat	2.161325		

Сравним оцененные этими методами коэффициенты:

	Coefficient			
	OLS	2SLS	3SLS	True
	Первое уравнение			
C(1)	-0.363218	-0.424633	-0.424633	-0.5
C(2)	88.69833	83.95884	83.95884	80.0
C(3)	0.476608	0.585040	0.585040	0.7
	Второе уравнение			
C(4)	1.019773	0.901844	0.889189	0.75
C(5)	-11.07961	-1.113488	-0.653169	10.0
C(6)	-1.752307	-1.684157	-1.671481	-1.5
C(7)	1.672787	1.628247	1.624186	1.5

Приведем для иллюстрации результаты применения двухступенчатой процедуры Godfrey–Hutton к рассмотренной системе уравнений:

$$\begin{cases} y_{t1} = \alpha_{11} y_{t2} + \theta_{11} x_{t1} + \theta_{21} x_{t2} + u_{t1}, \\ y_{t2} = \alpha_{12} y_{t1} + \theta_{12} x_{t1} + \theta_{22} x_{t3} + \theta_{32} x_{t4} + u_{t2}, \end{cases}$$

используя данные, приведенные в Примере 2. В качестве инструментальных переменных используются $x_1 = 1, x_2, x_3, x_4$.

Мы уже произвели выше оценивание обоих уравнений двухшаговым методом наименьших квадратов. Полученные при этом 2SLS-остатки обозначим, соответственно, $\hat{u}_{t1}^{2SLS}$, $\hat{u}_{t2}^{2SLS}$. Оценим уравнение регрессии $\hat{u}_{t1}^{2SLS}$ на $x_1 = 1, x_2, x_3, x_4$. Полученное значение коэффициента детерминации равно 0.000319, так что $J = nR^2 = 0.00957$. Число степеней свободы равно $4 - 3 = 1$. Поскольку P-значение равно 0.922, гипотеза пригодности использованных инструментов не отвергается. (Если оценить уравнение регрессии $\hat{u}_{t2}^{2SLS}$ на $x_1 = 1, x_2, x_3, x_4$, то в этом случае $J = 0$, число степеней свободы равно $4 - 4 = 0$, и J-критерий неприменим.)

Поскольку гипотеза пригодности инструментов не отвергнута, перейдем ко второму этапу и используем на этом этапе критерий Дарбина-Ву-Хаусмана.

Сначала оцениваем уравнение

$$y_{t2} = \pi_{12} + \pi_{22} x_{t2} + \pi_{32} x_{t3} + \pi_{42} x_{t4} + w_{t2}$$

и получаем ряд остатков $\hat{w}_{t2} = y_{t2} - \hat{y}_{t2}$. Затем оцениваем расширенное первое уравнение

$$y_{t1} = \alpha_{11} y_{t2} + \theta_{11} x_{t1} + \theta_{21} x_{t2} + \gamma \hat{w}_{t2} + \eta_{t1}$$

и проверяем гипотезу $H_0 : \gamma = 0$. Поскольку здесь γ – скалярная величина, то для проверки этой гипотезы можно использовать t-критерий. Результаты оценивания расширенной модели:

Dependent Variable: Y1
Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y2	-0.424633	0.023668	-17.94140	0.0000
C	83.95884	5.665468	14.81940	0.0000
X2	0.585040	0.067204	8.705405	0.0000
W2	0.616608	0.074993	8.222165	0.0000
R-squared	0.928759	Mean dependent var		100.0167
Durbin-Watson stat	2.089751	Prob(F-statistic)		0.000000

Оцененный коэффициент $\hat{\gamma} = 0.617$ имеет очень высокую статистическую значимость, что говорит о наличии проблемы эндогенности в первом уравнении.

Применяя тот же критерий ко второму уравнению, получаем для расширенного уравнения

$$y_{t2} = \alpha_{12} y_{t1} + \theta_{12} x_{t1} + \theta_{22} x_{t3} + \theta_{32} x_{t4} + \gamma \hat{w}_{t1} + \eta_{t2}$$

следующие результаты:

Dependent Variable: Y2
Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
Y1	0.901844	0.426087	2.116571	0.0444
C	-1.113488	38.32829	-0.029051	0.9771
X3	-1.684157	0.278578	-6.045545	0.0000
X4	1.628247	0.171996	9.466768	0.0000
W1	0.238481	0.605917	0.393586	0.6972
R-squared	0.950369	Mean dependent var		102.0033
Durbin-Watson stat	2.187028	Prob(F-statistic)		0.000000

В этом случае оцененный коэффициент при дополнительной переменной статистически незначим, так что эндогенность во втором уравнении не выявляется. Между тем, OLS оценки всех параметров имеют большее смещение по сравнению с другими методами оценивания и в первом и во втором уравнении, а во втором уравнении оценка постоянной имеет большое смещение при всех методах оценивания.

Заметим, что используемые на втором шаге процедур 2SLS и 3SLS очищенные значения $\tilde{y}_{t1}, \tilde{y}_{t2}$ эндогенных переменных y_{t1}, y_{t2} являются линейными комбинациями экзогенных переменных x_{t2}, x_{t3} и x_{t4}, в число которых входят переменные, находящиеся в правых частях соответствующих структурных уравнений. В правой части первого уравнения оказываются переменные $\tilde{y}_{t1}$ и x_{t2}, а в правой части второго уравнения – переменные $\tilde{y}_{t2}$, x_{t3} и x_{t4}. Однако при этом опасной мультиколлинеарности переменных в правых частях каждого из уравнений не возникло, и это происходит благодаря тому, что:

- в состав $\tilde{y}_{t1}$ входят не только x_{t2}, но также и переменные x_{t3} и x_{t4}, не слишком сильно коррелированные с x_{t2}: $Corr(x_3, x_2) = 0.308$, $Corr(x_4, x_2) = 0.680$,

- в состав $\tilde{y}_{t2}$ входят не только x_{t3} и x_{t4}, но также и переменная x_{t2}, не слишком сильно коррелированная с x_{t3} и x_{t4}.

Попробуем смоделировать ситуацию, в которой происходит критическая потеря точности оценивания. Для этого реализуем процесс порождения данных, соответствующий системе

$$\begin{cases} y_{t1} = -0.5y_{t2} + 50 + 1.1x_{t2} + u_{t1}, \\ y_{t2} = 0.5y_{t1} - 45 + 1.1x_{t4} + u_{t2}, \end{cases}$$

с теми же значениями переменных x_{t2}, x_{t4} и ошибок u_{t1}, u_{t2}, что и в только что рассмотренном примере. Эта система идентифицируема точно, и ее оценивание методом 3SLS дает следующий результат:

System: SYS_EXACT_2_4
Estimation Method: Three-Stage Least Squares

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.412458	0.062834	-6.564206	0.0000
C(2)	53.69121	10.66377	5.034918	0.0000
C(3)	0.955053	0.151915	6.286768	0.0000
C(4)	0.545426	0.213593	2.553572	0.0135
C(5)	-56.93625	25.48195	-2.234375	0.0296
C(7)	1.167493	0.064218	18.18025	0.0000

Determinant residual covariance 506.2606

Equation: Y1=C(1)*Y2+C(2)+C(3)*X2
Observations: 30

R-squared	0.489024	Mean dependent var	98.82331
Adjusted R-squared	0.451174	S.D. dependent var	6.911911
S.E. of regression	5.120537	Sum squared resid	707.9373
Durbin-Watson stat	1.741253		

Equation: Y2=C(4)*Y1+C(5)+C(7)*X4
Observations: 30

R-squared	0.929722	Mean dependent var	125.5639
Adjusted R-squared	0.924517	S.D. dependent var	23.42729
S.E. of regression	6.436470	Sum squared resid	1118.560
Durbin-Watson stat	2.049311		

Все оцененные коэффициенты имеют высокую статистическую значимость.

Заменим во втором уравнении переменную x_{t4} новой переменной x_{t5}, порождаемой соотношением

$$x_{t5} = x_{t2} + 2\eta_t, \quad \eta_t \sim i.i.d.\ N(0,1), \ t = 1,\ldots,30,$$

так что система принимает вид:

$$\begin{cases} y_{t1} = -0.5y_{t2} + 50 + 1.1x_{t2} + u_{t1}, \\ y_{t2} = 0.5y_{t1} - 45 + 1.1x_{t5} + u_{t2}. \end{cases}$$

С точки зрения критериев идентифицируемости формально ничего не изменилось: первое уравнения сверхидентифицируемо, а второе идентифицируемо точно. Посмотрим, однако, на результаты оценивания.

System: SYS_2_5
Estimation Method: Three-Stage Least Squares

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-2.109257	3.560708	-0.592370	**0.5561**
C(2)	9.837905	90.42874	0.108792	**0.9138**
C(3)	3.365924	4.983926	0.675356	**0.5020**
C(4)	0.940376	0.427263	2.200928	0.0320
C(5)	-82.48076	30.85206	-2.673428	0.0099
C(8)	1.023874	0.193137	5.301291	0.0000
Determinant residual covariance		3003.720		

Equation: Y1=C(1)*Y2+C(2)+C(3)*X2
Observations: 30

R-squared	-7.828599	Mean dependent var	102.6282
Adjusted R-squared	-8.482569	S.D. dependent var	5.264654
S.E. of regression	16.21186	Sum squared resid	7096.260
Durbin-Watson stat	1.700910		

Equation: Y2=C(4)*Y1+C(5)+C(8)*X5
Observations: 30

R-squared	0.825653	Mean dependent var	117.9538
Adjusted R-squared	0.812738	S.D. dependent var	14.61378
S.E. of regression	6.323932	Sum squared resid	1079.787
Durbin-Watson stat	2.085919		

Здесь оценки стандартных ошибок оцененных коэффициентов значительно выросли, особенно в первом уравнении, так что оказались статистически незначимыми все три оцененных коэффициента этого уравнения; значительно изменились и оценки коэффициентов, особенно в первом уравнении:

	Coefficient		Std. Error	
	Equation 1	Equation 2	Equation 1	Equation 2
C(1)	-0.412458	-2.109257	0.062834	3.560708
C(2)	53.69121	9.837905	10.66377	90.42874
C(3)	0.955053	3.365924	0.151915	4.983926
C(4)	0.545426	0.940376	0.213593	0.427263
C(5)	-56.93625	-82.48076	25.48195	30.85206
C(7)	1.167493		0.064218	
C(8)		1.023874		0.193137

Столь драматические изменения произошли по той причине, что переменная x_{t5} имеет очень высокую корреляцию с x_{t2} в выборке: $Corr(x_{t5}, x_{t2}) = 0.969$. И хотя формально эти переменные различны и оба уравнения идентифицируемы, со статистической точки зрения эти переменные “слишком близки”, что порождает проблему опасной мультиколлинеарности и приводит к практической неидентифицируемости коэффициентов структурных уравнений, поскольку теперь:

- в состав $\tilde{y}_{t1}$ помимо x_{t2} и константы входит только переменная x_{t5}, сильно коррелированная с x_{t2},
- в состав $\tilde{y}_{t2}$ помимо x_{t5} и константы входит только переменная x_{t2}, сильно коррелированная с x_{t5}.

2.6.8. Прогнозирование по оцененной системе одновременных уравнений

Если нас интересует только предсказание значений $y_{t+1,1},\ldots,y_{t+1,g}$ эндогенных переменных в новом наблюдении по заданным (планируемым) значениям экзогенных переменных $x_{t+1,1},\ldots,x_{t+1,K}$, то это можно осуществить непосредственно с использованием оцененной матрицы $\hat{\Pi}$ приведенной формы:

$$\left(\hat{y}_{t+1,1},\ldots,\hat{y}_{t+1,g}\right)=\left(x_{t+1,1},\ldots,x_{t+1,K}\right)\hat{\Pi}\,.$$

Возьмем смоделированные в примере 2 предыдущего раздела данные и на основании оцененной приведенной формы построим точечные прогнозы значений $\left(y_{30+t,1},y_{30+t,2}\right)$, $t=1,\ldots,30$, для новой последовательности значений экзогенных переменных $\left(x_{30+t,2},x_{30+t,3},x_{30+t,4}\right)$, $t=1,\ldots,30$, воспроизводящей основные характеристики последовательности $\left(x_{t,2},x_{t,3},x_{t,4}\right)$, $t=1,\ldots,30$. Оценивание уравнений редуцированной формы дает следующие результаты:

Dependent Variable: Y1

Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	61.35073	7.478051	8.204107	0.0000
X2	0.428163	0.083052	5.155376	0.0000
X3	0.511247	0.060707	8.421559	0.0000
X4	-0.502120	0.036696	-13.68320	0.0000

Dependent Variable: Y2
Method: Least Squares

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	54.21529	18.22556	2.974685	0.0063
X2	0.386136	0.202414	1.907652	0.0675
X3	-1.223092	0.147955	-8.266637	0.0000
X4	1.175413	0.089436	13.14250	0.0000

R-squared	0.936459	Mean dependent var	102.0033
Adjusted R-squared	0.929127	S.D. dependent var	27.23976
S.E. of regression	7.251752	Akaike info criterion	6.923929
Sum squared resid	1367.286	Schwarz criterion	7.110755
Log likelihood	-99.85894	F-statistic	127.7280
Durbin-Watson stat	2.066287	Prob(F-statistic)	0.000000

Соответственно, прогнозы для $y_{t,1}, y_{t,2}$, $t = 31,\ldots,60$, вычисляются по формулам:

$$\hat{y}_{t,1} = 61.351 + 0.428x_{t2} + 0.511x_{t3} - 0.502x_{t4},$$

$$\hat{y}_{t,2} = 54.215 + 0.386x_{t2} - 1.223x_{t3} + 1.175x_{t4}.$$

Параллельно вычислим для $t = 31,\ldots,60$ ”теоретические” значения $y_{t,1}\_true, y_{t,2}\_true$, порожденные системой без случайных ошибок:

$$\begin{cases} y_{t1} = -0.5y_{t2} + 80 + 0.7x_{t2}, \\ y_{t2} = 0.75y_{t1} + 10 - 1.5x_{t3} + 1.5x_{t4}. \end{cases}$$

Приведем полученные таким образом прогнозные и “теоретические” значения (округленные):

Y1_PREDICT	Y1_TRUE	Y2_PEDICT	Y2_TRUE
116.3	119.2	49.4	53.3
92.1	98.4	101.2	101.2
95.6	94.6	106.7	106.4
77.9	117.9	70.3	73.0
87.4	91.1	107.2	106.6
98.3	94.6	91.5	91.9
102.0	75.9	145.6	141.9
97.0	85.8	114.1	113.1
122.0	98.6	121.6	120.8
108.9	103.0	116.0	114.3
98.6	96.6	105.2	105.2
92.6	124.2	64.4	67.7
88.2	109.4	73.2	75.6
91.8	99.1	117.3	115.7
107.1	92.1	119.4	117.9
115.2	87.4	132.9	130.3
98.8	91.2	124.1	122.9
100.1	108.3	93.1	92.7
92.8	116.5	68.0	70.9
99.4	98.0	79.7	80.5
86.5	100.1	100.6	100.9
110.2	91.8	107.2	106.8
106.2	99.1	94.1	94.5
83.8	85.4	133.1	130.4
93.0	111.4	89.0	90.5
89.7	106.0	61.9	64.5
88.3	82.6	140.6	137.4
116.3	92.3	117.7	116.9
92.1.	89.9	155.3	151.4
95.6	87.8	141.7	138.5

Как видим, прогнозные значения оказались весьма близкими к “истинным”. Если использовать при сравнении этих значений среднюю абсолютную процентную ошибку прогноза (MAPE – Mean Absolute Percent Error), вычисляемую по формуле

$$\text{MAPE}(i) = \frac{1}{30}\sum_{t=1}^{30} 100\left|\frac{y_{t,i}\_predict - y_{t,i}\_true}{y_{t,i}\_true}\right|,$$

то получаются такие результаты:

$$\text{MAPE}(1) = 0.881\%, \quad \text{MAPE}(2) = 1.797\%.$$

Подойдем теперь к вопросу прогнозирования с другой стороны: попробуем получить прогнозные значения на основании оцененных структурных уравнений. При этом будем ориентироваться на оценки, полученные ранее методом 3SLS:

	Coefficient
C(1)	-0.424633
C(2)	83.95884
C(3)	0.585040
C(4)	0.889189
C(5)	-0.653169
C(6)	-1.671481
C(7)	1.624186

Иначе говоря, будем опираться на оцененные структурные уравнения

$$\begin{cases} y_{t1} = -0.425 y_{t2} + 83.959 + 0.585 x_{t2}, \\ y_{t2} = 0.889 y_{t1} - 0.653 - 1.671 x_{t3} + 1.624 x_{t4}. \end{cases}$$

Это дает следующие прогнозные значения (Y1F, Y2F):

Y1F	Y1_PREDICT	Y2F	Y2_PREDICT
117.9539	116.3	49.43060	49.4
98.62752	92.1	101.2295	101.2
95.21508	95.6	106.7986	106.7
116.2262	77.9	70.28356	70.3
92.14986	87.4	107.2833	107.2
95.65463	98.3	91.58455	91.5
77.93844	102.0	145.6529	145.6
87.49209	97.0	114.3175	114.1
98.28987	122.0	121.6912	121.6
101.7954	108.9	115.6271	116.0
97.07454	98.6	105.3218	105.2

121.9508	92.6	64.30229	64.4
108.9582	88.2	73.29769	73.2
98.49096	91.8	117.0801	117.3
92.60201	107.1	119.3591	119.4
88.14985	115.2	132.8158	132.9
91.87218	98.8	124.2472	124.1
106.9471	100.1	92.66480	93.1
115.1834	92.8	68.00794	68.0
98.80164	99.4	79.70102	79.7
100.0887	86.5	100.6958	100.6
92.81281	110.2	107.3434	107.2
99.37046	106.2	94.12591	94.1
86.43539	83.8	133.1007	133.1
110.1443	93.0	89.02851	89.0
106.2692	89.7	62.01926	61.9
83.80259	88.3	140.5781	140.6
93.01326	116.3	117.8567	117.7
89.61098	92.1.	155.1329	155.3
88.19077	95.6	141.5809	141.7

Во втором и четвертом столбцах таблицы приведены для сравнения прогнозные значения, полученные ранее на основании оцененной приведенной формы.

Для прогнозов, полученных по структурным уравнениям, получаем:

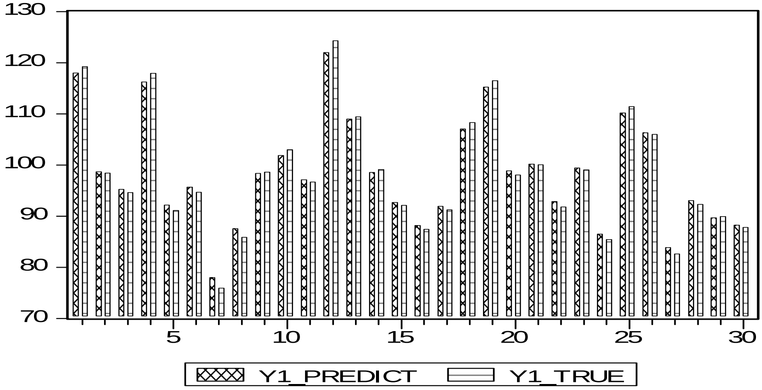
$$\mathrm{MAPE}(1) = 0.906\%, \quad \mathrm{MAPE}(2) = 1.773\%,$$

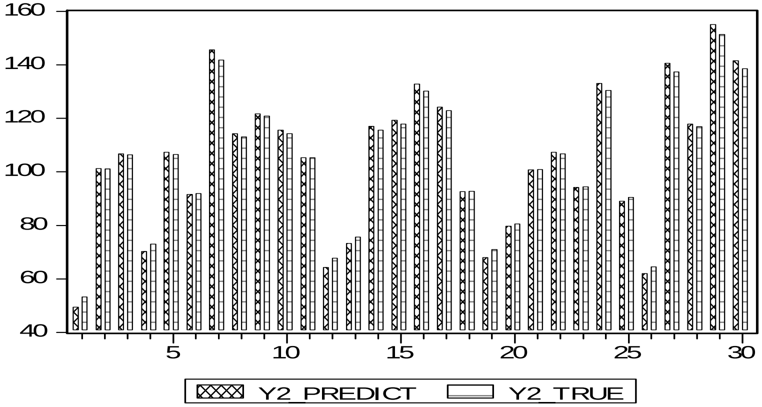
в сравнении со значениями

$$\mathrm{MAPE}(1) = 0.881\%, \quad \mathrm{MAPE}(2) = 1.797\%,$$

полученными для прогнозов на основании оцененных редуцированных уравнений. В целом, прогнозы обоих типов близки по указанной характеристике качества.

Следующие графики иллюстрируют качество прогнозов в сравнении с “теоретическими” значениями:


130
120
110
100
90
80
70
5
10
15
20
25
30
Y1_PREDICT
Y1_TRUE


160
140
120
100
80
60
40
5
10
15
20
25
30
Y2_PREDICT
Y2_TRUE

Глава 3. Панельные данные

3.1. Модель кажущихся несвязанными регрессий, модель ковариационного анализа

Пусть мы имеем данные $\{y_{it},\ x_{it};\ i = 1,\ldots,N,\ t = 1,\ldots,T\}$ о значениях переменных y и x для N субъектов (индивидов, фирм, стран, регионов и т. п.) в T последовательных моментов (периодов) времени (в этом случае говорят, что мы имеем дело с ***панельными данными***) и хотим оценить модель линейной связи между переменными y и x, считая y объясняемой, а x – объясняющей переменной. В общем случае x является вектором конечной размерности p, и наиболее общей формой линейной модели наблюдений для такой ситуации являлась бы спецификация

$$y_{it} = x_{it}^T \theta_{it} + u_{it}\ ,\ \ i = 1,\ldots,N,\ t = 1,\ldots,T\ ,$$

где θ_{it} измеряет частное влияние x_{it} в момент (период) t для субъекта i. Однако такая модель слишком обща, чтобы быть полезной, и приходится накладывать какую-ту структуру на коэффициенты θ_{it}.

Простейшей в этом отношении является модель ***пула*** (*pool*) с $\theta_{it} \equiv \theta$:

$$y_{it} = x_{it}^T \theta + u_{it}\ ,$$

в которой предполагается, что

$$u_{it} \sim \textbf{\textit{i.i.d.}}\ N\left(0, \sigma_u^2\right),\ \ i = 1,\ldots,N,\ t = 1,\ldots,T\ ,$$

и что

$$E\left(x_{it} u_{js}\right) = 0\ \underline{\text{для любых}}\ \ i, j = 1,\ldots,N,\ t, s = 1,\ldots,T\ ,$$

так что x является **экзогенной** переменной. В этом случае мы имеем дело с обычной линейной регрессией с NT наблюдениями, удовлетворяющей предположениям классической нормальной линейной модели. Для получения эффективных оценок вектора коэффициентов θ достаточно использовать обычный метод наименьших квадратов *(OLS)* . Полученная при этом оценка $\hat{\theta}$ является **наилучшей линейной несмещенной оценкой** (***BLUE*** **–** *best linear unbiased estimate*) вектора θ. При соответствующих предположениях о поведении значений объясняющих переменных, когда $N \to \infty$ или/и $T \to \infty$, эта оценка является также и ***состоятельной*** оценкой этого вектора.

П р и м е р

Рассмотрим один популярный объект статистических исследований – данные о размере инвестиций (*invest*), рыночной цене (*mvalue*) и акционерном капитале (*kstock*) 10 крупных компаний США за период с 1935 по 1954 г.г. При анализе этих и других данных мы по большей части используем пакет статистического анализа STATA8 и приводим протоколы оценивания, иногда с некоторыми сокращениями. Оценивая по указанным статистическим данным модель пула, получаем следующие результаты.

Cross-sectional time-series FGLS regression

Coefficients: generalized least squares
Panels: **homoskedastic**
Correlation: **no autocorrelation**
Estimated covariances = 1 Number of obs = 200
Estimated autocorrelations = 0 Number of groups = 10
Estimated coefficients = 3 Time periods = 20
Wald chi2(2)= 866.14, Prob > chi2 =0.0000
Log likelihood = -1191.802

invest	Coef.	Std. Err.	z	P>z
mvalue	.1155622	.0057918	19.95	0.00
kstock	.2306785	.025284	9.12	0.00
cons	-42.71437	9.440069	-4.52	0.00

Приведенный протокол оценивания показывает, что мы оцениваем модель пула, в которой отсутствуют гетероскедастичность и автокоррелированность ошибок. Для проверки гипотезы о незначимости регрессии в целом (т.е. гипотезы о нулевых значениях коэффициентов при объясняющих переменных *mvalue* и *kstock*) здесь используется критерий Вальда, основанный на статистике $Wald = qF$, где F – обычная F-статистика для проверки этой гипотезы, а q - количество линейных ограничений на параметры модели (в данном примере $q = 2$). Статистика критерия Вальда имеет асимптотическое распределение хи-квадрат с q степенями свободы. Вычисленный на основе этого асимптотического распределения наблюденный уровень значимости (P-значение), соответствующий наблюдаемому значению 866.14, равен Prob > chi2 =0.0000, так что гипотеза о нулевых значениях коэффициентов при объясняющих переменных *mvalue* и *kstock* отвергается. Во втором столбце таблицы приведены оценки коэффицентов, в третьем – оценки для стандартных ошибок этих оценок. В четвертом столбце приведены значения t-статистик для раздельной проверки гипотез о равенстве нулю отдельных коэффициентов, а в пятом столбце – соответствующие им P-значения, вычисляемые на основании нормального приближения распределения Стьюдента (отсюда обозначение Z вместо обычного t в заголовке четвертого столбца). Полученные P-значения говорят о высокой статистической значимости оценок коэффициентов.

В такой упрощенной модели, собственно, и не возникает никаких особенностей статистического анализа, связанных с панельным представлением данных. Положение, однако, изменится, если предположить, что в той же модели $y_{it} = x_{it}^T\theta + u_{it}$ ошибки u_{it}, оставаясь статистически независимыми между собой, имеют <u>разные дисперсии для разных субъектов</u>: $D(u_{it}) = \sigma_{ui}^2$. В этом случае OLS-оценки коэффициентов остаются несмещенными, но возникает смещение при оценивании дисперсий этих оценок, что отражается на оцененных значениях стандартных ошибок оценок, используемых при построении доверительных интервалов для коэффициентов и при проверке гипотез о значениях коэффициентов (например, при проверке их статистической значимости). Хотя и здесь особенность панельного характера данных отражается лишь в структуре весов: при применении ***взвешенного метода наименьших квадратов*** (***WLS*** *– weighted least squares*) веса, приписываемые различным наблюдениям, не изменяются в пределах наблюдений одного субъекта.

П р и м е р (продолжение)

Cross-sectional time-series FGLS regression

Coefficients: generalized least squares
Panels: **heteroskedastic**
Correlation: **no autocorrelation**
Wald chi2(2) = 669.69, Prob > chi2 = 0.0000
Log likelihood = -1037.152

invest	Coef.	Std. Err.	z	P>z
mvalue	.1116328	.0049823	22.41	0.00
kstock	.1537718	.0125707	12.23	0.00
cons	-21.44348	3.901219	-5.50	0.00

Здесь протокол оценивания указывает на применение взвешенного метода наименьших квадратов (Panels: heteroscedastic). По сравнению с предыдущим результатом, существенно снизилось значение оцененного коэффициента при переменной *kstock* и произошло двукратное уменьшение оцененной стандартной ошибки для этого коэффициента. Соответственно изменился и вычисленный 95% интервал для данного коэффициента: теперь это интервал (0.129, 0.178), тогда как ранее это был интервал (0.181, 0.280)

Заметим, что как и в обычной модели регрессии, вместо применения взвешенного метода наименьших квадратов можно использовать оценки коэффициентов, полученные обычным методом наименьших квадратов (OLS), но при этом следует произвести ***коррекцию стандартных ошибок*** этих оценок. Использование такого подхода в пакете Stata8 дает в рассматриваемом примере следующие результаты.

П р и м е р (продолжение)

. xtpcse invest mvalue kstock, hetonly casewise

Linear regression, heteroskedastic panels corrected standard errors

Estimated covariances = 10, R-squared = 0.8124

Wald chi2(2) = 567.87, Prob > chi2 = 0.000

Het-corrected

invest	Coef.	Std. Err.	z	P>z
mvalue	.1155622	.0070863	16.31	0.00
kstock	.2306785	.029747	7.75	0.00
cons	-42.71437	7.131515	-5.99	0.00

Следующим шагом в усложнении модели может быть дополнительное предположение о наличии корреляционной связи между ошибками в уравнениях для разных субъектов в один и тот же момент времени (***cross-sectional correlation***):

$$Cov(u_{it}, u_{js}) = \begin{cases} 0 & \text{для } t \neq s \\ \sigma_{ij}\ (\neq 0) & \text{для } t = s \end{cases},$$

так что матрица ковариаций $\Sigma = (\sigma_{ij})$ не является диагональной.

В этом случае приходится использовать обобщенный метод наименьших квадратов (***GLS** – generalized least squares*), учитывающий это обстоятельство. Мы коснемся деталей несколько позже, при рассмотрении моделей “кажущихся несвязанными регрессий”, а сейчас только посмотрим, что дает применение этого метода к анализируемым данным.

П р и м е р (продолжение)

Cross-sectional time-series FGLS regression

Coefficients: generalized least squares

Panels: heteroskedastic with cross-sectional correlation

Correlation: no autocorrelation
Estimated covariances = 55 Number of obs = 200
Estimated autocorrelations = 0 Number of groups =10
Estimated coefficients = 3 Time periods = 20
Wald chi2(2) =3738.07, Prob > chi2 = 0.0000
Log likelihood = -879.4274

invest	Coef.	Std. Err.	z	P>z
mvalue	.1127515	.0022364	50.42	0.00
kstock	.2231176	.0057363	38.90	0.00
cons	-39.84382	1.717563	-23.20	0.00

Заметим, что в этом случае помимо собственно трех коэффициентов модели приходится оценивать еще и 10 дисперсий случайных ошибок в уравнениях для 10 предприятий, а также 45 ковариаций σ_{ij}, $i \neq j$. Если не накладывать никаких дополнительных ограничений на структуру матрицы ковариаций $\Sigma = (\sigma_{ij})$, то оценивание каждой ковариации (или дисперсии) производится на основании всего 20 наблюдений и потому может быть весьма неточным.

И здесь можно оставить OLS-оценки для коэффициентов, скорректировав оценки стандартных ошибок (correlated panels corrected standard errors). При этом получаем:
.xtpcse invest mvalue kstock, casewise
Linear regression, correlated panels corrected standard errors (PCSEs)
Estimated covariances = 55, R-squared = 0.8124
Wald chi2(2) = 637.41, Prob > chi2 = 0.0000
Panel-corrected

invest	Coef.	Std. Err.	z	P>z
mvalue	.1155622	.0072124	16.02	0.00
kstock	.2306785	.0278862	8.27	0.00
cons	-42.71437	6.780965	-6.30	0.00

Следующим шагом в усложнении модели является снятие предположения о взаимной независимости ошибок в пределах одного субъекта, например, путем предположения о том, что последовательность ошибок при наблюдении i-го субъекта следует процессу авторегрессии первого порядка AR(1) с нулевым средним. Поясним это на примере модели $y_{it} = x_{it}^T\theta + u_{it}$, в которой

$$u_{it} = \rho_i\, u_{i,t-1} + \varepsilon_{it}\,,$$

где $|\rho_i| < 1$, а случайные величины $\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT}$ являются гауссовскими ***инновациями***, так что они взаимно независимы и имеют одинаковое распределение $N\left(0, \sigma_{\varepsilon i}^2\right)$ и, кроме того, ε_{it} не зависит от значений $u_{i,t-k}$, $k \ge 1$. Коэффициент ρ_i можно оценить различными способами. Можно, например, оценить (методом наименьших квадратов) модель $y_{it} = x_{it}^T\theta + u_{it}$ без учета автокоррелированности ошибок, получить последовательность остатков $\hat{u}_{i1}, \hat{u}_{i2}, \ldots, \hat{u}_{iT}$, вычислить значение статистики Дарбина–Уотсона

$$d_i = \frac{\sum_{t=2}^{T}\left(\hat{u}_{it} - \hat{u}_{i,t-1}\right)^2}{\sum_{t=1}^{T}\hat{u}_{it}^2}$$

и, используя приближенное соотношение $\rho_i \cong 1 - d_i/2$, получить оценку $\hat{\rho}_{i,DW} \cong 1 - d_i/2$.

Можно поступить иначе: получив последовательность остатков $\hat{u}_{i1}, \hat{u}_{i2}, \ldots, \hat{u}_{iT}$, использовать оценку наименьших квадратов, получаемую при оценивании уравнения регрессии

$$\hat{u}_{it} = \rho_i\, \hat{u}_{i,t-1} + \eta_{it}\,.$$

Искомая оценка вычисляется по формуле:

$$\hat{\rho}_i = \frac{\sum_{t=2}^{T} \hat{u}_{it}\hat{u}_{i,t-1}}{\sum_{t=1}^{T} \hat{u}_{it}^2}.$$

[В пакете Stata8 эта оценка обозначена как $\hat{\rho}_{tscorr}$.]

После получения оценок для ρ_i, $i = 1,\ldots,N$, в уравнениях для каждого субъекта производится известное преобразование Прайса–Уинстена переменных, призванное получить модель с независимыми ошибками. Объединяя преобразованные уравнения, получаем возможность произвести в ней OLS-оценивание коэффициентов.

Если предполагается, что уравнения имеют общий AR-параметр, т.е. $\rho_1 = \rho_2 = \cdots = \rho_N = \rho$, то это общее значение ρ оценивается величиной $\hat{\rho} = (\rho_1 + \rho_2 + \cdots + \rho_N)/N$, так что преобразование Прайса–Уинстена использует одну эту оценку.

П р и м е р (продолжение)

Будем предполагать, что дисперсии ошибок для разных субъектов могут быть различными. Если при этом предполагается отсутствие перекрестной коррелированности ошибок между уравнениями и если ошибки в уравнениях для разных субъектов следуют одинаковым AR(1)-моделям (с общим ρ), то оценивание такой модели дает следующие результаты.

```
. xtpcse invest mvalue kstock, correlation(ar1) hetonly rhotype(dw)
(note: estimates of rho outside [-1,1] bounded to be in the range [-1,1])
```

Prais-Winsten regression, heteroskedastic panels corrected standard errors

Autocorrelation: common AR(1)

```
Estimated covariances  =   10        R-square  =  0.5468
Estimated autocorrelations = 1
Estimated coefficients  = 3
```

Wald chi2(2) = 91.72,   Prob > chi2 = 0.0000

Het-corrected

invest	Coef.	Std. Err.	z	P>z
mvalue	.0972395	.0126259	7.70	0.00
kstock	.306441	.0561245	5.46	0.00
cons	-42.07116	21.02442	-2.00	0.045

rho   .8678619

Здесь использовалась оценка, вычисляемая через посредство статистики Дарбина–Уотсона.

Если использовать второй вариант оценивания коэффициента ρ, описанный выше, то это приводит к следующим результатам.

. xtpcse invest mvalue kstock, correlation(ar1) hetonly rhotype(tscorr) casewise

Prais-Winsten regression, heteroskedastic panels corrected standard errors

Autocorrelation: common AR(1)

Estimated covariances = 10   R-squared = 0.6904

Estimated autocorrelations = 1

Estimated coefficients   = 3

Wald chi2(2) = 192.41, Prob > chi2 = 0.0000

Het-corrected

invest	Coef.	Std. Err.	z	P>z
mvalue	.1032102	.0112252	9.19	0.000
kstock	.2947986	.0459298	6.42	0.000
cons	-45.78767	13.97367	-3.28	0.001

rho   .7563511

Оцененное значение ρ существенно изменилось.

Если допускается перекрестная коррелированность ошибок между уравнениями и ошибки в уравнениях для разных субъектов следуют одинаковым AR(1)-моделям (с общим ρ), то оценивание такой модели (по DW-варианту) дает следующие результаты.

. xtpcse invest mvalue kstock, correlation(ar1) rhotype(dw)
(note: estimates of rho outside [-1,1] bounded to be in the range [-1,1])
Prais-Winsten regression, heteroskedastic panels corrected standard errors
Panels: heteroskedastic (balanced)
Estimated covariances = 55, R-squared = 0.5468
Estimated autocorrelations = 1
Estimated coefficients = 3
Wald chi2(2) = 120.05, Prob > chi2 = 0.0000
Panel-corrected

	Coef.	Std. Err.	z	P>z
mvalue	.0972395	.0124362	7.82	0.000
kstock	.306441	.054533	5.62	0.000
cons	-42.07116	24.09387	-1.75	0.081

rho .8678619

И опять использование второго варианта оценивания коэффициента ρ приводит к несколько отличным результатам:. xtpcse invest mvalue kstock, correlation(ar1) rhotype(tscorr) casewise
Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)
Panels: correlated (balanced
Estimated covariances = 55, R-squared = 0.6904
Estimated autocorrelations = 1
Estimated coefficients = 3
Wald chi2(2)= 215.52, Prob > chi2 = 0.0000
Panel-corrected

	Coef.	Std. Err.	z	P>z
mvalue	.1032102	.0108656	9.50	0.000
kstock	.2947986	.0432809	6.81	0.000
cons	-45.78767	15.24513	-3.00	0.003

rho .7563511

Посмотрим, что дает оценивание модели с перекрестной коррелированностью ошибок между уравнениями, когда ошибки в уравнениях для разных субъектов могут следовать разным AR(1)-моделям (с разными ρ_i). При оценивании такой модели (по DW-варианту) получаем:

. xtpcse invest mvalue kstock, correlation(psar1) rhotype(dw) casewise
Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)
Panels: correlated (balanced)
Autocorrelation: panel-specific AR(1)
Estimated covariances = 55, R-squared = 0.7570
Estimated autocorrelations = 10,
Estimated coefficients = 3,
Wald chi2(2) = 211.38, Prob > chi2 = 0.0000
Panel-corrected

	Coef.	Std. Err.	z	P>z
mvalue	.1013946	.0108632	9.33	0.000
kstock	.3449446	.0478113	7.21	0.000
cons	-41.18685	19.33078	-2.13	0.033

rhos = .7427231 .8831453 .9741851 .7277056 .9564705 ... .9343119

А при оценивании по второму варианту –
.xtpcse invest mvalue kstock, correlation(psar1) rhotype(tscorr) casewise
Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)
Panels: correlated (balanced)
Autocorrelation: panel-specific AR(1)
Estimated covariances = 55, R-squared = 0.8670
Estimated autocorrelations =10
Estimated coefficients = 3
Wald chi2(2) = 444.53, Prob > chi2 = 0.0000
Panel-corrected

	Coef.	Std. Err.	z	P>z
mvalue	.1052613	.0086018	12.24	0.000
kstock	.3386743	.0367568	9.21	0.000
cons	-58.18714	12.63687	-4.60	0.000

rhos = .5135627 .87017 .9023497 .63368 .8571502 ... .8752707

Заметим, что и здесь приходится оценивать значительное количество дисперсий, ковариаций и автокорреляций, используя всего 20 наблюдений, растянутых во времени.

Для удобства объединим полученные результаты в одну таблицу, дополнительно указав в последнем столбце 95% доверительные интервалы для коэффициентов.

invest	Coef.	Std.Err.	z	P>z	95% Conf. Int.
Независимые одинаково распределенные ошибки					
mvalue	.116	.0058	19.95	0.00	.104 .127
kstock	.231	.0253	9.12	0.00	.181 .280
Гетероскедастичность – WLS					
mvalue	.112	.0050	22.41	0.00	.102 .122
kstock	.154	.0126	12.23	0.00	.129 .178
SUR – GLS					
mvalue	.113	.0022	50.42	0.00	.108 .117
kstock	.223	.0057	38.90	0.00	.212 .234
AR(1) – common rho (Durbin – Watson) : est rho = .8678619					
mvalue	.097	.0126	7.70	0.00	.072 .122
kstock	.306	.0561	5.46	0.00	.196 .416
AR(1) – common rho (OLS) : est rho = .7563511					
mvalue	.103	.0112	9.19	0.00	.081 .125
kstock	.295	.0460	6.42	0.00	.205 .385
SUR&AR(1) – common rho (D–W) : est rho = .8678619					
mvalue	.097	.0124	7.82	0.00	.073 .122
kstock	.306	.0545	5.62	0.00	.200 .413
SUR&AR(1) – common rho (OLS) : est rho = .7563511					
mvalue	.103	.0109	9.50	0.00	.082 .125

kstock	.295	.0433	6.81	0.00	.210 .388
SUR&AR(1) – different rho (D–W)					
mvalue	.101	.0109	9.33	0.00	
kstock	.345	.0478	7.21	0.00	
rhos = .7427231 .8831453 .9741851 .7277056 ... .9343119					
SUR&AR(1) – different rho (TSCORR)					
mvalue	.105	.0086	12.24	0.00	
kstock	.339	.0368	9.21	0.00	
rhos = .5135627 .87017 .9023497 .63368 ... .8752707					

Здесь **SUR&AR(1)** означает наличие корреляции между ошибками в разных уравнениях в совпадающие моменты времени и AR(1)-модель для ошибок в пределах каждого предприятия.

Обратимся теперь к модели ***ковариационного анализа***

$$\mathbf{M_0:}\quad y_{it} = \alpha_i + \beta_i x_{it} + u_{it}\,,\quad i = 1,\ldots,N,\ t = 1,\ldots,T\ ;$$

α_i и β_i – неизвестные постоянные, u_{it} – случайные ошибки. (Для простоты переменную x будем рассматривать пока как скалярную переменную.)

Если предполагать, что

$$u_{it} \sim i.i.d.\ N\left(0, \sigma_u^2\right),\quad i = 1,\ldots,N,\ t = 1,\ldots,T\ ,$$

и что

$$E\left(x_{it} u_{js}\right) = 0 \ \underline{\text{для любых}}\ \ i, j = 1,\ldots,N,\ t, s = 1,\ldots,T\ ,$$

так что x является экзогенной переменной, то мы имеем дело с N ***не связанными между собой*** линейными регрессиями, удовлетворяющими предположениям классической нормальной линейной регрессии. Для получения оценок параметров α_i и β_i эти регрессии можно оценивать в этом случае порознь, так что оценки наименьших квадратов для α_i и β_i имеют вид:

$$\hat{\beta}_i = \frac{\sum_{t=1}^{T} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)}{\sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2},$$

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta}_i \bar{x}_i, \qquad i = 1, \ldots, N,$$

где

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{it}, \quad \bar{x}_i = \frac{1}{T} \sum_{t=1}^{T} x_{it}.$$

Эти оценки имеют при фиксированных значениях y_{it} и x_{it} нормальное распределение и являются наилучшими линейными несмещенными оценками (*BLUE*).

Если ошибки u_{it} независимы между собой и имеют нормальные распределения с нулевыми средними, но дисперсии их различны для разных субъектов, так что

$$u_{it} \sim \textbf{\textit{i.i.d.}}\ N\left(0, \sigma_{ui}^2\right), \ i = 1, \ldots, N, \ t = 1, \ldots, T,$$

то тогда следует использовать ***взвешенный метод наименьших квадратов***, приписывая каждому наблюдению для i-го субъекта вес $w_i = 1/\sigma_{ui}$. Поскольку же дисперсии σ_{ui}^2 в реальных исследованиях не известны, приходится использовать ***доступную версию*** этого метода, в которой вместо весов $w_i = 1/\sigma_{ui}$ берутся их оценки $w_i = 1/\hat{\sigma}_{ui}$, где $\hat{\sigma}_{ui}^2$ – подходящие оценки неизвестных дисперсий. В качестве таковых можно брать, например, несмещенные оценки дисперсий σ_{ui}^2:

$$\hat{\sigma}_{ui}^2 = RSS^{(i)} / (N - p).$$

Здесь $RSS^{(i)}$ – сумма квадратов остатков, получаемых при оценивании модели регрессии для i-го субъекта, а p – количество объясняющих переменных в уравнениях регрессии (для парной регрессии с константой $p = 2$).

Несколько более сложная модель возникает, если предположить ***коррелированность ошибок*** для разных субъектов в ***совпадающие*** моменты времени. Это ***модель кажущихся несвязанными регрессий***[1] (***SUR, SURE*** – *seemingly unrelated regressions*). При наличии такой коррелированности следует использовать уже не взвешенный, а ***обобщенный метод наименьших квадратов***. Если представить уравнение для i-го субъекта в векторно-матричной форме

$$y^{(i)} = X^{(i)}\theta^{(i)} + u^{(i)},$$

где

$$y^{(i)} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}, \; X^{(i)} = \begin{pmatrix} 1 & x_{i1} \\ \vdots & \vdots \\ 1 & x_{iT} \end{pmatrix}, \; \theta^{(i)} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \; u^{(i)} = \begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix},$$

то модель SUR можно записать в следующем виде:

$$\begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(N)} \end{pmatrix} = \begin{pmatrix} X^{(1)} & 0 & \cdots & 0 \\ 0 & X^{(2)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X^{(N)} \end{pmatrix} \begin{pmatrix} \theta^{(1)} \\ \vdots \\ \theta^{(N)} \end{pmatrix} + \begin{pmatrix} u^{(1)} \\ \vdots \\ u^{(N)} \end{pmatrix},$$

или (в очевидных обозначениях)

$$y = X\theta + u.$$

При сделанных предположениях ковариационная матрица $NT \times 1$-вектора u равна

$$\Omega = Cov(u) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1N} \\ \Sigma_{12} & \Sigma_{22} & \cdots & \Sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{1N} & \Sigma_{2N} & \cdots & \Sigma_{NN} \end{pmatrix},$$

[1] В некоторых руководствах по эконометрике такую модель называют системой внешне не связанных между собой уравнений.

где $(T \times T)$-матрица Σ_{ij} имеет вид

$$\Sigma_{ij} = \begin{pmatrix} \sigma_{ij} & 0 & \cdots & 0 \\ 0 & \sigma_{ij} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{ij} \end{pmatrix}, \ \sigma_{ij} = Cov(u_{it}, u_{jt}).$$

Обобщенная оценка наименьших квадратов (***GLS-оценка***) вектора θ находится по формуле

$$\hat{\theta}_{SUR} = \hat{\theta}_{GLS} = \left(X^T \Omega^{-1} X\right)^{-1} X^T \Omega^{-1} y .$$

Заметим, что подлежащая обращению матрица Ω имеет размер $NT \times NT$. Однако такого обращения можно избежать вследствие наличия следующего соотношения:

$$\Omega^{-1} = \begin{pmatrix} \Sigma_{11}^{(-1)} & \Sigma_{12}^{(-1)} & \cdots & \Sigma_{1N}^{(-1)} \\ \Sigma_{12}^{(-1)} & \Sigma_{22}^{(-1)} & \cdots & \Sigma_{2N}^{(-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{1N}^{(-1)} & \Sigma_{2N}^{(-1)} & \cdots & \Sigma_{NN}^{(-1)} \end{pmatrix},$$

где

$$\Sigma_{ij}^{(-1)} = \begin{pmatrix} \sigma_{ij}^{(-1)} & 0 & \cdots & 0 \\ 0 & \sigma_{ij}^{(-1)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{ij}^{(-1)} \end{pmatrix},$$

а $\sigma_{ij}^{(-1)}$ – элементы матрицы Σ^{-1}, обратной к матрице

$$\Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{1N} & \cdots & \sigma_{NN} \end{pmatrix}.$$

Благодаря этому достаточно произвести обращение матрицы размера $N \times N$.

Учет коррелированности ошибок в различных уравнениях позволяет ожидать определенного выигрыша в точности оценивания каждого из $\theta^{(i)}$ за счет информации, идущей от других уравнений через указанную коррелированность. Однако реальный выигрыш зависит от целого ряда факторов. Например, если $\sigma_{ij} = \sigma^2 \rho$ для $i \neq j$, то предпочтительность SUR-оценки возрастает с ростом $\rho \to 1$, когда T велико. С другой стороны, если $\rho = 0$, то SUR- и OLS-оценки совпадают. Кроме того, непосредственная реализация SUR-оценивания на практике невозможна из-за того, что значения σ_{ij} не известны исследователю. ***Доступный*** (*feasible*) вариант SUR-оценивания предусматривает использование адаптивной оценки $\hat{\theta}_{FGLS}$ вектора θ, при вычислении которой неизвестные значения σ_{ij} заменяются их состоятельными оценками $\hat{\sigma}_{ij}$.

Пусть $e^{(i)} = y^{(i)} - X^{(i)}\hat{\theta}^{(i)}$ – вектор остатков, получаемый при OLS-оценивании уравнения для i-го субъекта. Тогда естественной оценкой для σ_{ij} является

$$\hat{\sigma}_{ij} = \frac{\left(e^{(i)}\right)^T e^{(j)}}{T}.$$

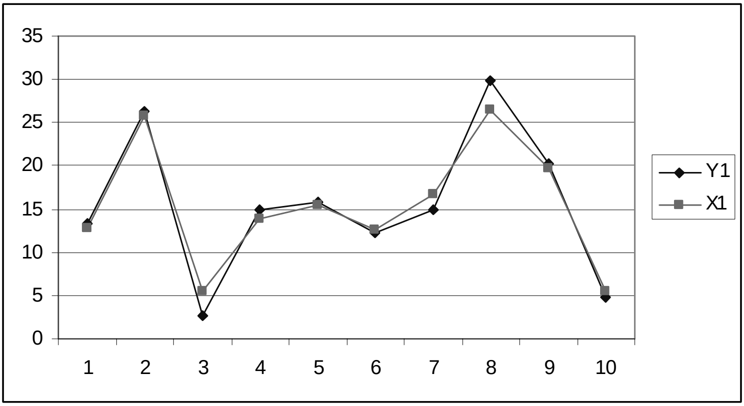
При $j = i$ это есть просто $RSS^{(i)}/T$, и, как известно, такая оценка для дисперсии ошибки в i-м уравнении имеет смещение, а несмещенной оценкой для этой дисперсии является $RSS^{(i)}/(T-p)$, где p – количество объясняющих переменных в уравнении регрессии. (Конечно, при этом должно выполняться условие $T > p$.) При соответствующих условиях на матрицу X, требующихся и в классической модели линейной регрессии, обе оценки $\hat{\theta}_{GLS}$ и $\hat{\theta}_{FGLS}$ при $T \to \infty$ <u>состоятельны</u>.

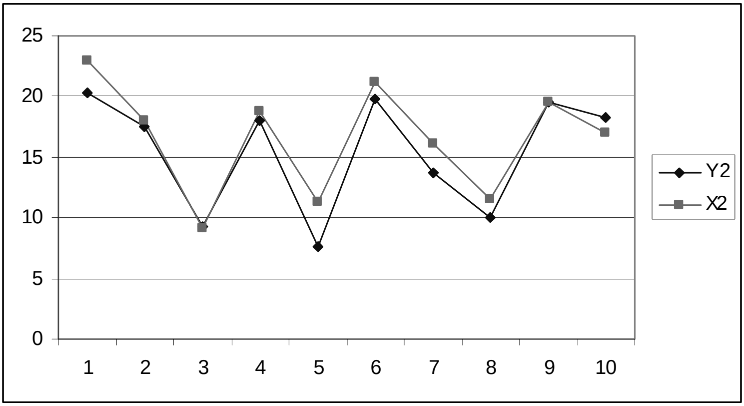
П р и м е р

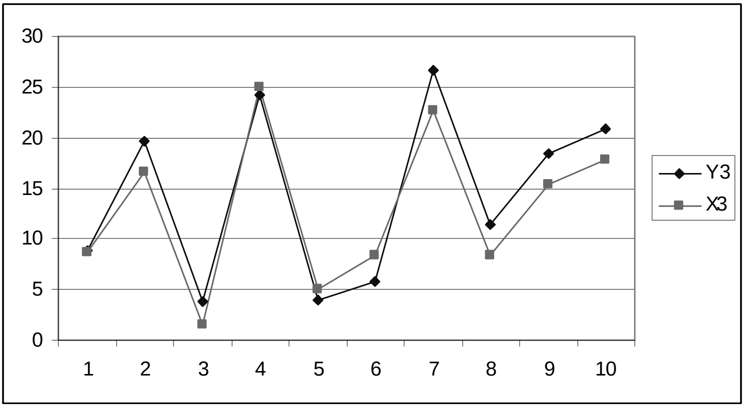
Рассмотрим для иллюстрации приведенные в [Greene (1993), стр. 481] ежегодные данные об объемах инвестиций y и прибыли x трех предприятий ($N = 3$) за десятилетний период ($T = 10$).

t	Y1	X1		Y2	X2		Y3	X3
1	13.32	12.85		20.30	22.93		8.85	8.65
2	26.30	25.69		17.47	17.96		19.60	16.55
3	2.62	5.48		9.31	9.160		3.87	1.47
4	14.94	13.79		18.01	18.73		24.19	24.91
5	15.80	15.41		7.63	11.31		3.99	5.01
6	12.20	12.59		19.84	21.15		5.73	8.34
7	14.93	16.64		13.76	16.13		26.68	22.70
8	29.82	26.45		10.00	11.61		11.49	8.36
9	20.32	19.64		19.51	19.55		18.49	15.44
10	4.77	5.43		18.32	17.06		20.84	17.87

Здесь столбцы Y_i, X_i содержат данные по i-му предприятию, $i = 1, 2, 3$. Ниже приведены графики изменения объемов инвестиций и прибыли по каждому из трех предприятий.




25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
Y2
X2


30
25
20
15
10
5
0
1
2
3
4
5
6
7
8
9
10
Y3
X3

Раздельное оценивание уравнений регрессии (в пакете Eviews) дает следующие результаты.

Первое предприятие:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.468913	0.980426	-2.518205	0.0359
X1	1.167170	0.058250	20.03737	0.0000
R-squared	0.9805			

Второе предприятие:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.384797	1.972680	-0.701988	0.5026
X2	1.014542	0.115314	8.798102	0.0000
R-squared	0.9063			

Третье предприятие:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.455479	1.491604	0.305362	0.7679
X3	1.076374	0.100360	10.72516	0.0000
R-squared	0.9350			

Матрица $\Sigma = (\sigma_{ij})$ оценивается как

$$\hat{\Sigma} = \begin{pmatrix} 1.2549 & -0.0099 & -0.9101 \\ -0.0099 & 1.9628 & 1.0351 \\ -0.9101 & 1.0351 & 4.3279 \end{pmatrix};$$

соответствующая корреляционная матрица имеет вид

$$\begin{pmatrix} 1 & -0.0063 & -0.3905 \\ -0.0063 & 1 & 0.3552 \\ -0.3905 & 0.3552 & 1 \end{pmatrix}.$$

Использование доступного GLS приводит к следующим результатам.

Первое предприятие:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.857213	0.812548	-3.52	0.002
X1	1.192389	0047494	25.11	0.000

R-squared 0.9800

Второе предприятие:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.11701	1.66034	-1.28	0.214
X2	1.05876	0.09663	10.96	0.000

R-squared 0.9046

Третье предприятие:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.721196	1.199687	0.60	0.553
X3	1.055824	0.077589	13.61	0.000

R-squared 0.9346

Оцененные коэффициенты несколько отличаются от результатов раздельного оценивания уравнений, и вопрос в том, сколь значимым является это различие. В связи с этим представляет интерес проверка гипотезы

$H_0 : \sigma_{ij} = 0$ для $i \neq j$.

В предположении нормальности ошибок эта гипотеза соответствует ***статистической независимости ошибок в разных уравнениях***. Для проверки этой гипотезы можно использовать ***критерий Бройша–Пагана***. Статистика этого критерия равна

$$\lambda = T \sum_{i=2}^{N} \sum_{j=1}^{i-1} r_{ij}^2 ,$$

где

$$r_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii} \hat{\sigma}_{jj}}}$$ – оцененная корреляция между ошибками в i-м и j-м уравнениях. При гипотезе H_0 эта статистика имеет

асимптотическое распределение хи-квадрат с числом степеней свободы, равным $N(N-1)/2$ (заметим, что гипотеза H_0 накладывает именно столько ограничений, поскольку $\sigma_{ij}=\sigma_{ji}$).

В нашем примере накладывается 3 ограничения, статистика критерия принимает значение 2.787. Соответствующее ему *P*-значение, вычисленное на основании распределения $\chi^2(3)$, равно 0.4256, так что если ориентироваться на это *P*-значение, то гипотеза независимости не отвергается.

Следует также обратить внимание на то, что различие между оценками коэффициентов при переменных x_1, x_2, x_3 довольно невелико, так что возникает вопрос о проверке гипотезы ***совпадения этих коэффициентов***:

$H_0 : \beta_1=\beta_2=\beta_3$.

В рамках модели SUR для проверки этой гипотезы используются две формы критерия Вальда: одна основана на F-статистике и *P*-значении, рассчитанном исходя из соответствующего F-распределения, а другая основана на статистике qF **(**q – количество линейных ограничений) и *P*-значении, рассчитанном исходя из асимптотического распределения $\chi^2(q)$ этой статистики**.** Использование этих двух форм дает следующие результаты:

Wald Test:

F-statistic	1.342317	Probability	0.278120
Chi-square	2.684634	Probability	0.261240

Разница в двух *P*-значениях весьма мала и не приводит к различию в статистических выводах: гипотеза $H_0 : \beta_1=\beta_2=\beta_3$ не отвергается.

Поскольку ранее на основании применения критерия Бройша–Пагана мы не отвергли гипотезу независимости ошибок в разных уравнениях, естественно попытаться проверить гипотезу $H_0 : \beta_1=\beta_2=\beta_3$ в условиях такой независимости.

Заметим, что модель SUR в нашем примере записывается как

$$\begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{2T} \\ y_{31} \\ \vdots \\ y_{3T} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1T} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & x_{21} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_{2T} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & x_{31} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & x_{3T} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \beta_1 \\ \alpha_2 \\ \beta_2 \\ \alpha_3 \\ \beta_3 \end{pmatrix} + \begin{pmatrix} u_{11} \\ \vdots \\ u_{1T} \\ u_{21} \\ \vdots \\ u_{2T} \\ u_{31} \\ \vdots \\ u_{3T} \end{pmatrix}$$

Это означает, что ее можно рассматривать как модель линейной регрессии переменной y, принимающей значения $y_{11}, y_{12}, \ldots, y_{1T}$, $y_{21}, y_{22}, \ldots, y_{2T}$, $y_{31}, y_{32}, \ldots, y_{3T}$, на следующие 6 переменных: три ***дамми***-переменных (*dummy variables*)

d_1 со значениями $\underbrace{1,\ldots,1}_{T}, \underbrace{0,\ldots,0}_{T}, \underbrace{0,\ldots,0}_{T}$;

d_2 со значениями $\underbrace{0,\ldots,0}_{T}, \underbrace{1,\ldots,1}_{T}, \underbrace{0,\ldots,0}_{T}$;

d_3 со значениями $\underbrace{0,\ldots,0}_{T}, \underbrace{0,\ldots,0}_{T}, \underbrace{1,\ldots,1}_{T}$;

и три комбинированных переменных d_1x, d_2x и d_3x, построенных на основании указанных дамми-переменных и переменной x, принимающей значения $x_{11}, x_{12}, \ldots, x_{1T}$, $x_{21}, x_{22}, \ldots, x_{2T}$, $x_{31}, x_{32}, \ldots, x_{3T}$, так что d_1x, d_2x и d_3x принимают значения

d_1x: $\underbrace{x_{11}, x_{12}, \ldots, x_{1T}}_{T}, \underbrace{0,0,\ldots,0}_{T}, \underbrace{0,0,\ldots,0}_{T}$

d_2x: $\underbrace{0,0,\ldots,0}_{T}, \underbrace{x_{21}, x_{22}, \ldots, x_{2T}}_{T}, \underbrace{0,0,\ldots,0}_{T}$

d_3x: $\underbrace{0,0,\ldots,0}_{T}, \underbrace{0,0,\ldots,0}_{T}, \underbrace{x_{31}, x_{32}, \ldots, x_{3T}}_{T}$.

Если случайные ошибки в разных уравнениях статистически независимы, то мы можем получить эффективные оценки коэффициентов, применяя OLS-оценивание, и проверить интересующую нас гипотезу $H_0 : \beta_1 = \beta_2 = \beta_3$ с использованием обычного F-критерия. При этом получаются следующие результаты:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D1	-2.468913	1.388033	-1.778714	0.0880
D2	-1.384797	2.233064	-0.620133	0.5410
D3	0.455479	1.137109	0.400559	0.6923
D1*X	1.167170	0.082467	14.15324	0.0000
D2*X	1.014542	0.130535	7.772208	0.0000
D3*X	1.076374	0.076508	14.06874	0.0000

R-squared 0.950532

Wald Test:

F-statistic 0.592788 Probability 0.560676

Таким образом, гипотеза $H_0 : \beta_1 = \beta_2 = \beta_3$ не отвергается и в предположении независимости ошибок.

Мы будем говорить о модели

M₀ : $y_{it} = \alpha_i + \beta_i\, x_{it} + u_{it}$, $i = 1,\ldots,N,\; t = 1,\ldots,T$,

как о “модели без ограничений”. Обозначим остаточную сумму квадратов (RSS) в этой модели как S_0,

$$S_0 = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \hat{\alpha}_i - \hat{\beta}_i\, x_{it}\right)^2 .$$

В рамках модели без ограничений рассмотрим две гипотезы:

$H_1 :\ \beta_i$ одинаковы для всех i,

$H_2 :\ \beta_i$ и α_i одинаковы для всех i.

Гипотеза $H_1 :\ \beta_i$ одинаковы для всех i. Этой гипотезе соответствует модель

M₁ : $y_{it} = \alpha_i + \beta x_{it} + u_{it}$, $i = 1,\ldots,N$, $t = 1,\ldots,T$.
Остаточную сумму квадратов (RSS) в модели M_1 обозначим через S_1,

$$S_1 = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \hat{\alpha}_i - \hat{\beta}\, x_{it}\right)^2 .$$

Модель M_1 можно записать в виде

$$y_{it} = \sum_{j=1}^{N} \alpha_i d_{ij} + \beta x_{it} + u_{it} ,$$

где $d_{ij} = 1$, если $j = i$, и $d_{ij} = 0$ в противном случае, так что мы имеем здесь в правой части N ***дамми***-переменных. Обозначим:

$$y = \left(y_{11}, y_{12}, \ldots, y_{1T}, y_{21}, y_{22}, \ldots, y_{2T}, \ldots, y_{N1}, y_{N2}, \ldots, y_{NT}\right)^T,$$

$$x = \left(x_{11}, x_{12}, \ldots, x_{1T},\ x_{21}, x_{22}, \ldots, x_{2T}, \ldots, x_{N1}, x_{N2}, \ldots, x_{NT}\right)^T,$$

$$u = \left(u_{11}, u_{12}, \ldots, u_{1T},\ u_{21}, u_{22}, \ldots, u_{2T}, \ldots, u_{N1}, u_{N2}, \ldots, u_{NT}\right)^T,$$

$$d_1 = \left(\underbrace{1,1,\ldots,1}_{T}, \underbrace{0,0,\ldots,0}_{NT-T}\right)^T,$$

$$d_2 = \left(\underbrace{0,0,\ldots,0}_{T}, \underbrace{1,1,\ldots,1}_{T}, \underbrace{0,0,\ldots,0}_{NT-2T}\right)^T,$$

$$\ldots$$

$$d_N = \left(\underbrace{0,0,\ldots,0}_{NT-T}, \underbrace{1,1,\ldots,1}_{T}\right)^T,$$

и пусть $X = [d_1\, d_2 \ldots d_N\ x]$ – матрица размера $NT \times (N+1)$, столбцами которых являются векторы $d_1, d_2, \ldots, d_N, x$. В этих обозначениях модель M_1 принимает вид

$$y = X\theta + u ,$$

где

$$\theta = (\alpha_1, \alpha_2, \ldots, \alpha_N, \beta)^T .$$

Соответственно, оценка наименьших квадратов $\hat{\theta} = \left(\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_N, \hat{\beta}\right)^T$ для вектора θ вычисляется по формуле

$$\hat{\theta} = \left(X^T X\right)^{-1} X^T y .$$

Будем предполагать далее, что

$$u_{it} \sim i.i.d.\ N\left(0, \sigma_u^2\right), \quad i = 1, \ldots, N, \ t = 1, \ldots, T ,$$

и что

$$E\left(x_{it} u_{js}\right) = 0 \ \text{для любых} \ \ i, j = 1, \ldots, N, \ t, s = 1, \ldots, T ,$$

так что x является ***экзогенной*** переменной.

При этих предположениях и при фиксированной матрице X оценка $\hat{\theta}$ имеет $(N+1)$-мерное нормальное распределение, причем

$$E\left(\hat{\theta}\right) = \theta ,$$

т.е. $\hat{\theta}$ является несмещенной оценкой для θ, а ковариационная матрица случайного вектора $\hat{\theta}$ имеет вид

$$Cov\left(\hat{\theta}\right) = \sigma_u^2 \left(X^T X\right)^{-1} .$$

Интересно, что численно *те же самые* значения оценок параметров $\alpha_1, \alpha_2, \ldots, \alpha_N, \beta$ можно получить иным способом. Именно, пусть

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{it} , \quad \bar{x}_i = \frac{1}{T} \sum_{t=1}^{T} x_{it} , \quad \bar{u}_i = \frac{1}{T} \sum_{t=1}^{T} u_{it}$$

– средние по времени значения переменных y, x и ошибки для i-го субъекта исследования. Усредняя по времени обе части уравнений

$$y_{it} = \alpha_i + \beta x_{it} + u_{it} ,$$

получаем

$$\bar{y}_i = \alpha_i + \beta \bar{x}_i + \bar{u}_i , \quad i = 1, \ldots, N .$$

Из двух последних уравнений находим:

$$y_{it} - \bar{y}_i = \beta \left(x_{it} - \bar{x}_i\right) + \left(u_{it} - \bar{u}_i\right), \ i = 1, \ldots, N, \ t = 1, \ldots, T$$

("модель, скорректированная на индивидуальные средние").

В правой части полученной модели оказались исключенными параметры $\alpha_1, \alpha_2, \ldots, \alpha_N$. Это приводит к меньшей стандартной ошибке оценки для параметра β, который обычно представляет первоочередной интерес. В рамках последней модели эта оценка вычисляется по формуле

$$\hat{\beta} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)}{\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x}_i)^2},$$

и о ней говорят как о ***"внутригрупповой"*** оценке ("*within-group*" *estimate*), имея в виду, что она строится только на основании отклонений значений переменных от их средних по времени и тем самым принимает во внимание только изменчивость в пределах каждого субъекта, не обращая внимание на изменчивость между субъектами. Точнее, конечно, следовало бы говорить о ***"внутрисубъектной"*** оценке, но используемая здесь терминология исторически вышла из теории дисперсионного анализа, где субъекты исследования часто объединяются в некоторое количество групп, так что индекс i относится не к отдельному субъекту, а к группе субъектов. Впрочем, в последнее время в эконометрической литературе чаще стали говорить об указанной оценке просто как о "***within***"-оценке. Соответственно, мы часто будем использовать термин ***"внутри"*** -оценка.

Получив в последней модели оценку наименьших квадратов $\hat{\beta}$, оценки для параметров $\alpha_1, \alpha_2, \ldots, \alpha_N$ можно вычислить следующим образом:

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta}\, \bar{x}_i, \quad i = 1, \ldots, N .$$

Полученные в итоге этих двух шагов оценки $\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_N, \hat{\beta}$ численно совпадают с оценками наименьших квадратов, получаемыми в модели $y_{it} = \sum_{j=1}^{N} \alpha_i d_{ij} + \beta x_{it} + u_{it}$. Следует только учитывать, что стандартные ошибки оценок $\hat{\beta}$ <u>отличаются</u> в этих двух моделях, а стандартные ошибки оценок $\hat{\alpha}_i$, получаемые в результате двухшаговой процедуры, <u>нельзя</u> вычислять по формулам для стандартных ошибок оценок наименьших квадратов.

<u>Гипотеза</u> H_2 : β_i и α_i одинаковы для всех i. Ей соответствует модель

M₂ : $y_{it} = \alpha + \beta x_{it} + u_{it}$, $i = 1,\ldots,N$, $t = 1,\ldots,T$ $(пул - pool)$.

Оценки наименьших квадратов для параметров α и β вычисляются по формулам

$$\hat{\beta} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x})(y_{it} - \bar{y})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x})^2},$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\,\bar{x},$$

где

$$\bar{y} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}, \quad \bar{x} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}.$$

Обозначим остаточную сумму квадратов в модели M_2 через S_2,

$$S_2 = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \hat{\alpha} - \hat{\beta}\,x_{it}\right)^2.$$

Рассмотрим задачу проверки гипотез H_1 и H_2 в рамках модели M_0 при сделанных ранее предположениях.

Проверка гипотезы H_2. Используем *F*-статистику

$$F_2 = \frac{(S_2 - S_0)/2(N-1)}{S_0/(NT-2N)} \;;$$

если гипотеза H_2 верна, то $F_2 \sim F_{2(N-1),\,NT-2N}$.

Если значение F_2 статистически незначимо, то следует объединить данные в пул. Если же значение F_2 статистически значимо, то следует искать источник гетерогенности параметров.

Проверка гипотезы H_1. Используем *F*-статистику

$$F_1 = \frac{(S_1 - S_0)/(N-1)}{S_0/(NT-2N)} \;;$$

если гипотеза H_1 верна, $F_1 \sim F_{N-1,\,NT-2N}$.

Если значение F_1 статистически значимо, то проверка прекращается. Если же значение F_1 статистически незначимо, то гипотеза H_1 (о совпадении всех β_i) не отвергается.

Можно также применить <u>условный</u> критерий гетерогенности α_i, а именно, проверить гипотезу

$$H_3:\ \alpha_1 = \ldots = \alpha_N$$

при условии $\beta_1 = \ldots = \beta_N$, т.е. в рамках модели

$$\mathrm{M}_1:\ y_{it} = \alpha_i + \beta x_{it} + u_{it}\,.$$

Для этой цели используем статистику

$$F_3 = \frac{(S_2 - S_1)/(N-1)}{S_1/(NT-N-1)} \,,$$

которая при гипотезе H_3 имеет распределение $F_{N-1,\,NT-N-1}$.

П р и м е р

Продолжим расмотрение данных об инвестициях и прибыли трех предприятий.

Выше мы уже проверили гипотезу H_1 : “β_i одинаковы для всех i ” в рамках модели M_0 и не отвергли эту гипотезу. Проверим в рамках той же модели M_0 гипотезу H_2: “β_i и α_i одинаковы для всех i”:

Wald Test:			
F-statistic	3.595209	Probability	0.019644

Эта гипотеза отвергается. Проверим теперь гипотезу H_3: $\alpha_1 = \ldots = \alpha_N$ в рамках модели M_1 : $y_{it} = \alpha_i + \beta x_{it} + u_{it}$:

Wald Test:			
F-statistic	6.810977	Probability	0.004183

Эта гипотеза также отвергается.

Полученные результаты говорят в пользу модели M_1 с одинаковыми β_i, но различными α_i.

3.2. Фиксированные эффекты

Начиная с этого раздела, мы обращаемся к методам анализа панельных данных, предназначенным в основном для анализа данных $\{y_{it},\ x_{it};\ i = 1,\ldots,N,\ t = 1,\ldots,T\}$, в которых количество субъектов исследования N велико, а количество наблюдений T над каждым субъектом мало. Вследствие малости T в таких ситуациях затруднительно использовать технику, интерпретирующую $y_{1t}, y_{2t},\ldots, y_{Nt}$ как N временных рядов длины T (например, технику векторных авторегрессий и моделей коррекции ошибок для нестационарных временных рядов). Основная направленность методов, предполагающих малость T , – получение по возможности наиболее эффективных оценок коэффициентов.

Сначала мы сфокусируем внимание на модели, соответствующей гипотезе H_1 со скалярной объясняющей переменной x:

$$y_{it} = \alpha_i + \beta x_{it} + u_{it}\,, \quad i = 1,\dots,N,\ t = 1,\dots,T\,,$$

т.е.

$$y_{it} = \sum_{j=1}^{N} \alpha_i d_{ij} + \beta x_{it} + u_{it}\,,$$

где $d_{ij} = 1$, если $i = j$ и $d_{ij} = 0$ в противном случае, так что мы имеем здесь в правой части N дамми-переменных. Здесь α_i трактуются как неизвестные фиксированные параметры (***фиксированные эффекты,*** *fixed effects*). Как и ранее, будем предполагать , что в этой модели

$$u_{it} \sim i.i.d.\ N\left(0, \sigma_u^2\right), \quad i = 1,\dots,N,\ t = 1,\dots,T\,,$$

и что

$$E\left(x_{it} u_{js}\right) = 0 \ \underline{\text{для любых}} \ \ i, j = 1,\dots,N,\ t, s = 1,\dots,T\,,$$

так что x является **экзогенной** переменной. Альтернативные названия этой модели:

1. ***OLS – дамми модель*** (***LSDV*** – *least squares dummy variables*);
2. ***модель с фиксированными эффектами*** (***FE*** – *fixed effects model*);
3. ***модель ковариационного анализа*** (***CV*** – *covariance analysis*).

В этой модели оценка наименьших квадратов, как мы уже отмечали выше, имеет вид:

$$\hat{\beta}_{CV} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)\left(y_{it} - \bar{y}_i\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)^2}\,;$$

при этом

$$D\left(\hat{\beta}_{CV}\right)=\frac{\sigma_u^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-\bar{x}_i\right)^2}.$$

Альтернативные названия для этой оценки

1. ***“внутригрупповая”*** оценка (***“внутри”-оценка,*** *within-estimator*),
2. оценка ***фиксированных эффектов***,
3. ***ковариационная*** оценка.

Часто для этой оценки используют также обозначения $\hat{\beta}_W$ (индекс W – от *within*) и $\hat{\beta}_{FE}$. Как мы уже отмечали выше, эта оценка теоретически имеет одно и то же значение при двух альтернативных методах ее получения: в рамках статистической модели $y_{it}=\sum_{j=1}^{N}\alpha_i d_{ij}+\beta x_{it}+u_{it}$ с дамми-переменными и в рамках модели в отклонениях от групповых средних $y_{it}-\bar{y}_i=\beta\left(x_{it}-\bar{x}_i\right)+\left(u_{it}-\bar{u}_i\right)$, $i=1,\ldots,N,\ t=1,\ldots,T$. Однако если количество субъектов анализа N велико, то в первой модели приходится обращать матрицу весьма большого размера ($N+1$), тогда как во второй модели такой проблемы не возникает.

Оценки для фиксированных эффектов вычисляются по формуле:

$$\hat{\alpha}_i=\bar{y}_i-\hat{\beta}\,\bar{x}_i,\quad i=1,\ldots,N.$$

При сделанных предположениях $\hat{\beta}_{CV}$ является наилучшей линейной несмещенной оценкой *(BLUE)* для коэффициента β,

$$\underset{T\to\infty}{p\lim}\,\beta_{CV}=\beta,\quad \underset{N\to\infty}{p\lim}\,\beta_{CV}=\beta,\quad \underset{T\to\infty}{p\lim}\,\hat{\alpha}_i=\alpha_i,$$

но

$$\underset{N\to\infty}{p\lim}\,\hat{\alpha}_i\neq\alpha_i,\ \text{хотя}\quad E\left(\hat{\alpha}_i\right)=\alpha_i.$$

Таким образом, $\hat{\beta}_{CV}$ является состоятельной оценкой и когда $N\to\infty$ и когда $T\to\infty$, в то время как $\hat{\alpha}_i$ состоятельна только, когда $T \to \infty$. Последнее есть следствие того, что оценивание каждого α_i производится фактически лишь по T наблюдениям, так что при фиксированном T с ростом N происходит лишь увеличение количества параметров α_i, но это не приводит к возрастанию точности оценивания каждого конкретного α_i.

Заметим, что если нас интересует только состоятельность оценки $\hat{\beta}_{CV}$, но не ее эффективность (т.е. свойство BLUE), то для этого не требуется строгая экзогенность x (т.е. не требуется , чтобы $E(x_{it}u_{js}) = 0$ для любых $i, j = 1,\ldots,N,\ t,s = 1,\ldots,T$). В этом случае достаточно выполнения соотношений $E(x_{it}u_{is}) = 0$ для любых $t,s = 1,\ldots,T$ и $i = 1,\ldots,N$ (т.е. требуется лишь экзогенность x в рамках каждого отдельного субъекта исследования).

В модели с фиксированными эффектами получаемые выводы – **условные** по отношению к значениям эффектов α_i **в выборке.** Такая интерпретация наиболее подходит для случаев, когда субъектами исследования являются страны, крупные компании или предприятия, т.е. каждый субъект ”имеет свое лицо”.

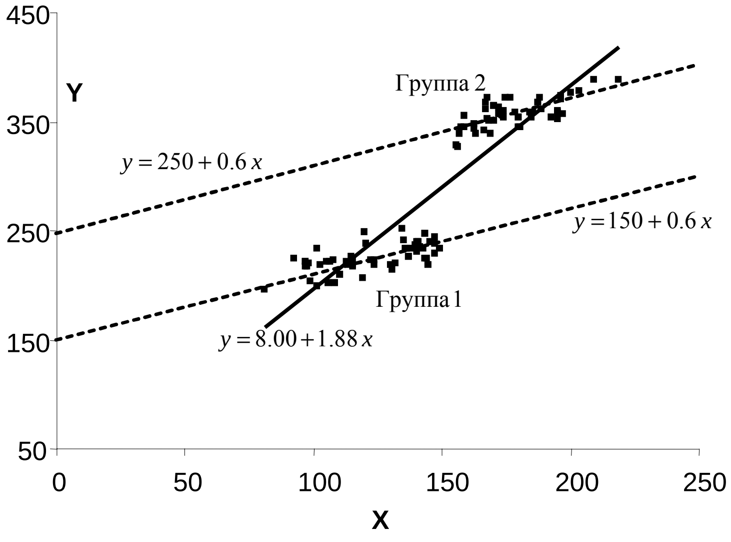
Сами эффекты α_i по-существу отражают наличие у субъектов исследования некоторых индивидуальных характеристик, не изменяющихся со временем в процессе наблюдений, которые трудно или даже невозможно наблюдать или измерить. Если значения таких характеристик не наблюдаются, то эти характеристики невозможно непосредственно включить в правые части уравнений регрессии в качестве объясняющих переменных. Но тогда мы имеем дело с “пропущенными переменными” – с ситуацией, которая может приводить к смещению оценок наименьших квадратов. Чтобы исключить такое смещение, в правые части уравнений вместо значений ненаблюдаемых индивидуальных характеристик как раз и вводятся ненаблюдаемые эффекты α_i. Проиллюстрируем возникновение указанного смещения следующим примером.

Пример

На следующем графике представлено облако рассеяния точек (x_{it}, y_{it}), порожденных моделью

$$y_{it} = \alpha_i + \beta x_{it} + u_{it}, \quad i = 1, 2, \; t = 1, \ldots, 100,$$

в которой $\alpha_1 = 150$, $\alpha_2 = 250$, $\beta = 0.6$, $u_{it} \sim i.i.d. \, N(0, 10^2)$. Значения x_{it} заданы (неслучайны); при $i = 1$ значения x_{1t} меньше 150, а при $i = 2$ значения x_{2t} больше 150.



Облако точек распадается на две группы точек: в группе 1 объединяются точки, соответствующие $i = 1$, а в группе 2 – точки, соответствующие $i = 2$. Точки первой группы располагаются вдоль (теоретической) прямой $y = 150 + 0.6\,x$ (нижняя пунктирная линия на графике), точки второй группы – вдоль (теоретической) прямой $y = 250 + 0.6\,x$.

Если по имеющимся 100 наблюдениям оценивать статистическую модель

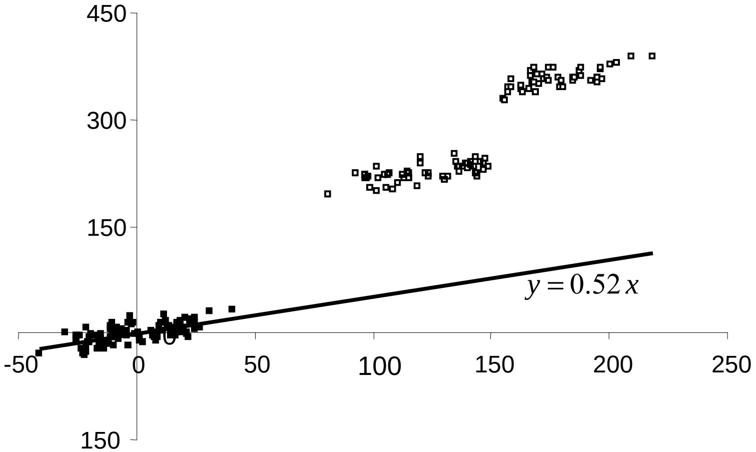
$$y_{it} = \alpha + \beta\, x_{it} + u_{it}\,,\;\; i = 1, 2,\;\; t = 1, \ldots, 100\,,$$

(пул), не принимающую во внимание возможное наличие индивидуальных эффектов, то оцененная модель принимает вид

$$\hat{y}_{it} = 8.00 + 1.88\, x_{it}\,,$$

так что оценка коэффициента β оказывается завышенной втрое по сравнению со значением, использованным при моделировании.

В то же время, если перейти от переменных x_{it}, y_{it} к отклонениям от средних значений в группах $x_{it} - \bar{x}_i$ и $y_{it} - \bar{y}_i$, то для новых переменных облако рассеяния концентрируется вокруг начала координат (на следующем графике соответствующие точки изображены черными квадратами) и вытянуто в правильном направлении:



Оцененная модель в отклонениях от средних в группах имеет вид:

$$\hat{y}_{it} = 0.517\, x_{it} \,,$$

и на этот раз оценка коэффициента β оказывается близкой к значению $\beta = 0.6$, использованному при моделировании.

Если оценивать модель с дамми-переменными

$$y_{it} = \alpha_1 d_{i1} + \alpha_2 d_{i2} + \beta\, x_{it} + u_{it} \,,$$

где $d_{ij} = 1$, если $j = i$, и $d_{ij} = 0$ в противном случае, то результаты оценивания таковы:

Dependent Variable: Y

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D1	161.4615	7.260153	22.23940	0.0000
D2	264.8593	10.46430	25.31074	0.0000
X	0.517319	0.058229	8.884227	0.0000

Полученные оценки $\hat{\alpha}_1 = 161.46,\ \hat{\alpha}_2 = 264.86,\ \hat{\beta} = 0.517$ близки к значениям параметров, использованным при моделировании.

3.3. Случайные эффекты

Запишем модель

$$y_{it} = \alpha_i + \beta\, x_{it} + u_{it} \,,\quad i = 1,\ldots,N,\ t = 1,\ldots,T \,,$$

соответствующую гипотезе H_1, в равносильном виде:

$$y_{it} = \mu + \alpha_i + \beta\, x_{it} + u_{it} \,,$$

где $\sum_{i=1}^{N} \alpha_i = 0$ (при таком условии α_i называют ***дифференциальными эффектами***). В ряде ситуаций N субъектов, для которых имеются статистические данные, рассматриваются как случайная выборка из некоторой более широкой совокупности (популяции), и исследователя интересуют не конкретные субъекты, попавшие в выборку, а обезличенные субъекты, имеющие заданные характеристики. Соответственно, в таких ситуациях предполагается, что α_i являются <u>*случайными величинами*</u>, и мы говорим тогда о модели

$$y_{it} = \mu + \alpha_i + \beta x_{it} + u_{it}$$

как о ***модели со случайными эффектами*** *(random effects)*. В такой модели α_i уже не интерпретируются как значения некоторых фиксированных параметров и не подлежат оцениванию. Вместо этого оцениваются параметры распределения случайных величин α_i.

Обозначая $\nu_{it} = \alpha_i + u_{it}$, получаем другую запись такой модели:

$$y_{it} = \mu + \beta x_{it} + (\alpha_i + u_{it}) = \mu + \beta x_{it} + \nu_{it} .$$

В такой форме модели ошибка ν_{it} состоит из двух компонент α_i и u_{it}. Как и в модели с фиксированными эффектами, случайные эффекты α_i также отражают наличие у субъектов исследования некоторых индивидуальных характеристик, не изменяющихся со временем в процессе наблюдений, которые трудно или даже невозможно наблюдать или измерить. Однако теперь значения этих характеристик встраиваются в состав случайной ошибки, как это делается в классической модели регрессии, в которой наличие случайных ошибок интерпретируется как недостаточность включенных в модель объясняющих переменных для полного объяснения изменений объясняемой переменной.

К прежним предположениям о том , что

$$u_{it} \sim i.i.d.\ N(0, \sigma_u^2),\quad i = 1,\ldots,N,\ t = 1,\ldots,T ,$$

и $E(x_{it} u_{js}) = 0$ для любых $i, j = 1,\ldots,N,\ t, s = 1,\ldots,T$,

добавим также следующие предположения:

$E(\alpha_i) = 0$ (так что и $E(\nu_{it}) = 0$),

$$E(\alpha_i \alpha_j) = \begin{cases} \sigma_\alpha^2, \text{если } i = j, \\ 0, \text{ если } i \neq j \end{cases}$$

(это означает, что последовательность значений $\alpha_1,\ldots,\alpha_N$ представляет ***случайную выборку*** из распределения $N\left(0,\sigma_u^2\right)$),

$$E\left(x_{it}\alpha_j\right)=0\,,\, i,\, j=1,\ldots,N,\;\; t=1,\ldots,T\,,$$

(так что $E\left(x_{it}\nu_{js}\right)=0$, и в модели со случайными ошибками ν_{it} переменная x является экзогенной переменной).

Если предположить еще, что

$$E\left(u_{it}\alpha_i\right)=0\,,$$

то тогда условная относительно x_{it} дисперсия случайной величины y_{it} равна

$$D\left(y_{it}\middle|x_{it}\right)=D\left(\nu_{it}\middle|x_{it}\right)=D\left(\nu_{it}\right)=D\left(\alpha_i+u_{it}\right)=\sigma_\alpha^2+\sigma_u^2\,.$$

Таким образом, дисперсия y_{it} складывается из двух некоррелированных компонент; их называют ***компонентами дисперсии***, а саму модель называют

1. ***моделью компонент дисперсии***;
2. ***однофакторной моделью компонент дисперсии***;
3. ***стандартной моделью со случайными эффектами*** (***RE модель*** – *random effects model*).

В векторной форме эта модель имеет вид

$$y_i=[e\,x_i]\delta+\nu_i\,,$$

где

$$y_i=\begin{pmatrix}y_{i1}\\ y_{i2}\\ \vdots\\ y_{iT}\end{pmatrix},\; e=\begin{pmatrix}1\\ 1\\ \vdots\\ 1\end{pmatrix},\; x_i=\begin{pmatrix}x_{i1}\\ x_{i2}\\ \vdots\\ x_{iT}\end{pmatrix},\; \delta=\begin{pmatrix}\mu\\ \beta\end{pmatrix},\; \nu_i=\begin{pmatrix}\nu_{i1}\\ \nu_{i2}\\ \vdots\\ \nu_{iT}\end{pmatrix}.$$

Заметим, что

$$E\left(\nu_{it}\nu_{is}\right)=E\left(\alpha_i+u_{it},\alpha_i+u_{is}\right)=\begin{cases}\sigma_\alpha^2+\sigma_u^2, \text{ если } t=s,\\ \sigma_\alpha^2, \text{ если } t\neq s,\end{cases}$$

так что случайные величины ν_{it} и ν_{is} коррелированы даже если некоррелированы ошибки u_{it}, и ковариационная матрица случайного вектора ν_i имеет вид

$$V = E\left(\nu_i \nu_i^T\right) = \sigma_u^2 I_T + \sigma_\alpha^2 \, ee^T .$$

Например, при $T = 3$

$$V = \begin{bmatrix} \sigma_u^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_u^2 + \sigma_\alpha^2 & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_u^2 + \sigma_\alpha^2 \end{bmatrix}.$$

При этом

$$Corr\left(\nu_{it}, \nu_{is}\right) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_u^2} = \rho \quad \text{для всех} \quad t \neq s$$

(предположение *равной коррелированности* в модели компонент дисперсии).

Оценивание

В RE-модели оценка

$$\hat{\beta}_{CV} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{it} - \bar{x}_i\right)\left(y_{it} - \bar{y}_i\right)}{\sum_{i=1}^{N} \sum_{t=1}^{T} \left(x_{it} - \bar{x}_i\right)^2}$$

остается несмещенной и состоятельной оценкой для β. Однако она перестает быть эффективной оценкой (BLUE), как это было в модели с фиксированными эффектами, поскольку не учитывает коррелированность ν_{it} во времени для субъекта i.

Мы можем ожидать, что обобщенная оценка наименьших квадратов (***GLS***-оценка), учитывающая такую коррелированность, будет более эффективной. Заметим, что ***GLS***-оценка для δ имеет вид

$$\hat{\delta}_{GLS} = \left[\sum_{i=1}^{N}\begin{bmatrix} e^T \\ x_i^T \end{bmatrix} V^{-1}\left[e\, x_i\right]\right]^{-1}\left[\sum_{i=1}^{N}\begin{bmatrix} e^T \\ x_i^T \end{bmatrix} V^{-1} y_i\right]$$

и что

$$V^{-1} = \frac{1}{\sigma_u^2}\left[I_T - \frac{ee^T}{T} + \Psi\frac{ee^T}{T}\right],$$

где $\Psi = \dfrac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}$. Чтобы не возникало путаницы с другими GLS-оценками, для GLS-оценки в стандартной модели со случайными эффектами используется также обозначение $\hat{\beta}_{RE}$. Заметим, что диагональные элементы матрицы V^{-1} равны $\left(1 - \dfrac{1-\Psi}{T}\right)\Big/\sigma_u^2$, а недиагональные элементы равны $-\dfrac{1-\Psi}{T\sigma_u^2}$.

Практически получить $\hat{\beta}_{RE}$ можно следующим образом. Усредняя по t обе части уравнения

$$y_{it} = \mu + \beta x_{it} + \nu_{it},$$

получаем соотношение

$$\bar{y}_i = \mu + \beta \bar{x}_i + \bar{\nu}_i.$$

Обозначив $\theta = 1 - \sqrt{\Psi}$, произведем преобразование

$$y_{it}^* = y_{it} - \theta\bar{y}_i,\ x_{it}^* = x_{it} - \theta\bar{x}_i,\ \nu_{it}^* = \nu_{it} - \theta\bar{\nu}_i,\ \mu^* = (1-\theta)\mu.$$

В результате получаем преобразованную модель

$$y_{it}^* = \mu^* + \beta x_{it}^* + \nu_{it}^*$$

с экзогенной переменной x_{it}^*, в которой ковариационная матрица вектора ошибок ν_{it}^* имеет вид:

$$Cov(\nu_{it}^*) = \sigma_u^2 I_T.$$

Поэтому применение OLS к преобразованной модели дает BLUE-оценку

$$\hat{\beta}_{GLS} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}^{*} - \bar{x}^{*}\right)\left(y_{it}^{*} - \bar{y}^{*}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}^{*} - \bar{x}^{*}\right)^{2}},$$

где

$$\bar{y}^{*} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}^{*}, \quad \bar{x}^{*} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}^{*}.$$

Это выражение можно представить в виде

$$\hat{\beta}_{GLS} = w\hat{\beta}_{b} + (1 - w)\hat{\beta}_{CV},$$

где

$$\hat{\beta}_{b} = \frac{\sum_{i=1}^{T}(\bar{x}_{i} - \bar{x})(\bar{y}_{i} - \bar{y})}{\sum_{i=1}^{T}(\bar{x}_{i} - \bar{x})^{2}}$$

– ***“межгрупповая”*** оценка (***между-оценка***, *between-group estimate*), соответствующая регрессии средних значений $\bar{y}_i$ на константу и средние значения $\bar{x}_i$, т.е.

$$\bar{y}_{i} = \mu + \beta\,\bar{x}_{i} + \bar{v}_{i}$$

(“модель для групповых средних”), и игнорирующая внутригрупповую изменчивость,

$$w = \frac{\Psi\sum_{i=1}^{N}(\bar{x}_{i} - \bar{x})^{2}}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_{i})^{2} + \Psi\sum_{i=1}^{N}(\bar{x}_{i} - \bar{x})^{2}}.$$

Таким образом, обобщенная оценка наименьших квадратов $\hat{\beta}_{GLS}$ в RE-модели учитывает и внутригрупповую и межгрупповую изменчивость. Она является взвешенным средним “межгрупповой” оценки $\hat{\beta}_{b}$ (учитывающей только межгрупповую изменчивость) и “внутригрупповой” оценки $\hat{\beta}_{CV}$ (учитывающей только внутригрупповую изменчивость), а w измеряет вес, придаваемый межгрупповой изменчивости. При сделанных предположениях обе оценки $\hat{\beta}_b$ и $\hat{\beta}_{CV}$ состоятельны, так что состоятельна и сама $\hat{\beta}_{GLS}$.

Если $T \to \infty$, то $\Psi \to 0$, $w \to 0$ и

$$\hat{\beta}_{GLS} \to \hat{\beta}_{CV},$$

так что при больших T оценки для β, получаемые в рамках моделей фиксированных и случайных эффектов, эквивалентны.

Если $\sigma_\alpha^2 \to 0$, то $\Psi \to 1$ и $V = E\left(v_i v_i^T\right) = \sigma_u^2 I_T + \sigma_\alpha^2 ee^T \to \sigma_u^2 I_T$. Соответственно, при этом GLS-оценка переходит в OLS-оценку, т.е.

$$\hat{\beta}_{GLS} \to \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x})(y_{it}-\bar{y})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x})^2} = \hat{\beta}_{OLS}$$

(в пределе нет никаких эффектов).

Заметим далее, что

$$D\left(\hat{\beta}_{GLS}\right) = \frac{\sigma_u^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2 + \Psi\sum_{i=1}^{N}(\bar{x}_i-\bar{x})^2}.$$

В то же время

$$D\left(\hat{\beta}_{CV}\right) = \frac{\sigma_u^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it}-\bar{x}_i)^2}.$$

Поскольку $\Psi > 0$, то из двух последних соотношений следует, что

$$D\left(\hat{\beta}_{GLS}\right) < D\left(\hat{\beta}_{CV}\right),$$

т.е. *GLS-оценка эффективнее*. Она эффективнее оценки $\hat{\beta}_{CV}$ именно потому, что использует как информацию о внутригрупповой изменчивости, так и информацию о межгрупповой изменчивости.

Чтобы реализовать эту GLS, т.е. получить ***доступную GLS-оценку*** (***FGLS*** – *feasible GLS*, или ***EGLS*** – *estimated GLS*), надо подставить в выражения для Ψ (и θ) подходящие оценки для σ_u^2 и $\sigma_u^2 + T\sigma_\alpha^2$.

Оценить σ_u^2 можно, используя ***внутригрупповые остатки***

$$(y_{it} - \bar{y}_i) - \hat{\beta}_{CV}(x_{it} - \bar{x}_i),$$

полученные при оценивании модели, скорректированной на индивидуальные средние:

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(y_{it} - \bar{y}_i) - \hat{\beta}_{CV}(x_{it} - \bar{x}_i)\right]^2}{N(T-1)-1}$$

(в знаменателе число степеней свободы равно количеству наблюдений NT минус количество оцениваемых параметров $N+1$).

Оценить дисперсию σ_α^2 случайных эффектов $\sigma_\alpha^2 = D(\alpha_i)$ можно, заметив, что при оценивании модели $\bar{y}_i = \mu + \beta\bar{x}_i + \bar{v}_i$, приводящей к межгрупповой оценке

$$\boxed{\hat{\beta}_b = \frac{\sum_{i=1}^{T}(\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\sum_{i=1}^{T}(\bar{x}_i - \bar{x})^2}},$$

дисперсия остатка для i-й группы равна

$$D\left(\bar{y}_i - \hat{\mu}_b - \hat{\beta}_b\bar{x}_i\right) = \sigma_u^2/T + \sigma_\alpha^2 .$$

Состоятельной оценкой для $\sigma_u^2/T + \sigma_\alpha^2$ является

$$\frac{\sum_{i=1}^{N}\left(\bar{y}_i - \hat{\mu}_b - \hat{\beta}_b \bar{x}_i\right)^2}{N-2}.$$

Поэтому состоятельной оценкой для σ_α^2 служит

$$\hat{\sigma}_\alpha^2 = \frac{\sum_{i=1}^{N}\left(\bar{y}_i - \hat{\mu}_b - \hat{\beta}_b \bar{x}_i\right)^2}{N-2} - \hat{\sigma}_u^2 / T;$$

$$\hat{\sigma}_u^2 + T\hat{\sigma}_\alpha^2 = T\,\frac{\sum_{i=1}^{N}\left(\bar{y}_i - \hat{\mu}_b - \hat{\beta}_b \bar{x}_i\right)^2}{N-2}$$

является состоятельной оценкой для $\sigma_u^2 + T\sigma_\alpha^2$. Эти две оценки используют межгрупповые остатки. Они являются также оценками максимального правдоподобия соответствующих дисперсий.

Следует только отметить, что, особенно при небольших значениях N и T, значение вычисленной указанным образом оценки дисперсии σ_α^2 может оказаться отрицательным.

Замечание

Внутригруповую и GLS-оценки можно получить в пакете STATA, используя команду ***xtreg*** (с опциями ***fe*** или ***re***).

Как мы уже отмечали выше, $\hat{\beta}_{GLS}$ можно представить в виде

$$\hat{\beta}_{GLS} = w\hat{\beta}_b + (1-w)\hat{\beta}_{CV},$$

так что $\hat{\beta}_{GLS}$ является линейной комбинацией “внутри” и “между” оценок. Эта линейная комбинация оптимальна. Поэтому, например, оценка $\hat{\beta}_{OLS}$, также являющаяся линейной комбинацией этих двух оценок (при $\Psi = 1$), хотя и состоятельна, но менее эффективна.

Критерий Бройша–Пагана для индивидуальных эффектов.

Это критерий для проверки в рамках RE-модели (со стандартными предположениями) гипотезы

$H_0 : \sigma_\alpha^2 = 0$ (сведение к модели пула).

Идея критерия основана на тождестве

$$\sum_{i=1}^{N}\left(\sum_{t=1}^{T} u_{it}\right)^2 = \sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}^2 + \sum_{i=1}^{N}\sum_{s\neq t} u_{is}u_{it} ,$$

из которого следует , что

$$\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T} u_{it}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}^2} - 1 = \frac{\sum_{i=1}^{N}\sum_{s\neq t} u_{is}u_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{T} u_{it}^2} .$$

При отсутствии автокоррелированности случайных ошибок u_{it} правая часть последнего равенства мала. Поэтому статистику критерия можно основывать на выражении, стоящем в левой части, в которое вместо ненаблюдаемых значений u_{it} подставляются остатки $\hat{u}_{it}$, полученные при OLS-оценивании модели пула. Против гипотезы H_0 говорят “слишком большие” значения

$$\left[\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T} \hat{u}_{it}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{u}_{it}^2} - 1\right]^2 .$$

Статистика критерия Бройша–Пагана

$$BP=\frac{NT}{2(T-1)}\left[\frac{\sum\limits_{i=1}^{N}\left(\sum\limits_{t=1}^{T}\hat{u}_{it}\right)^2}{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T}\hat{u}_{it}^2}-1\right]^2=\frac{NT}{2(T-1)}\left[1-\frac{T^2\sum\limits_{i=1}^{N}\bar{\hat{u}}_i^2}{\sum\limits_{i=1}^{N}\sum\limits_{t=1}^{T}\hat{u}_{it}^2}\right]^2$$

при гипотезе H_0 имеет асимптотическое распределение $\chi^2(1)$. Соответственно, гипотеза H_0 отвергается, если наблюдаемое значение статистики BP превышает критическое значение, рассчитанное по распределению $\chi^2(1)$.

3.4. Коэффициенты детерминации, разложение полной суммы квадратов

При анализе панельных данных возникают некоторые проблемы с определением коэффициента детерминации R^2, так что во многих руководствах по эконометрике и монографиях, специально посвященных анализу панельных данных, вообще не упоминается о коэффициенте детерминации. В то же время в некоторых пакетах статистического анализа предусмотрено вычисление коэффициентов детерминации и для панельных данных.

Проблема с определением коэффициента детерминации в случае панельных данных связана с неопределенностью в отношении того, что считать полной суммой квадратов, подлежащей разложению на объясненную регрессией и остаточную суммы квадратов. Здесь мы имеем соотношение

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-\bar{y})^2=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it}-\bar{y}_i)^2+\frac{1}{N}\sum_{i=1}^{N}(\bar{y}_i-\bar{y})^2,$$

и в качестве полной суммы квадратов может использоваться каждая из входящих в это выражение трех сумм квадратов. Соответствующие этим полным суммам регрессионные модели объясняют:

- отклонения наблюдаемых значений y_{it} от их среднего по всем NT наблюдениям;
- отклонения наблюдаемых значений y_{it} в группах от их средних по группе;
- отклонения средних по группам от среднего по всем NT наблюдениям.

Если мы используем оценку “пул”, то она получается в результате применения метода наименьших квадратов к уравнению

$$y_{it} = \alpha + \beta x_{it} + u_{it},\ i = 1,\ldots,N,\ t = 1,\ldots,T\ .$$

При этом коэффициент детерминации равен квадрату (выборочного) коэффициента корреляции между переменными y_{it} и

$$\hat{y}_{it} = \hat{\alpha} + \hat{\beta}_{OLS}\ x_{it},$$

где $\hat{\beta}_{OLS}$ – OLS-оценка коэффициента β в модели пула. Об этом коэффициенте детерминации говорят как о “R^2***-полном***” (R^2-overall),

$$R^2_{overall} = corr^2\left(y_{it}, \hat{\alpha} + \hat{\beta}_{OLS}\ x_{it}\right) = corr^2\left(y_{it}, \hat{\beta}_{OLS}\ x_{it}\right).$$

Если мы используем оценку “между”, то она получается в результате применения метода наименьших квадратов к уравнению

$$\bar{y}_i = \mu + \beta \bar{x}_i + \bar{v}_i,\ i = 1,\ldots,N\ .$$

При этом коэффициент детерминации равен квадрату (выборочного) коэффициента корреляции между переменными $\bar{y}_{it}$ и

$$\hat{\bar{y}}_i = \hat{\mu} + \hat{\beta}_b\ \bar{x}_i,$$

где $\hat{\beta}_b$ – “между”-оценка для коэффициента β. Об этом коэффициенте детерминации говорят как о “R^2***-между***” (R^2-between),

$$R^2_{between} = corr^2\left(\hat{\mu} + \hat{\beta}_b\ \bar{x}_i, \bar{y}_i\right) = corr^2\left(\hat{\beta}_b\ \bar{x}_i, \bar{y}_i\right).$$

Если мы используем оценку “внутри”, то она получается в результате применения метода наименьших квадратов к уравнению

$$y_{it} - \bar{y}_i = \beta\left(x_{it} - \bar{x}_i\right) + \left(u_{it} - \bar{u}_i\right),\ i = 1,\ldots,N,\ t = 1,\ldots,T\ .$$

В правой части последнего уравнению отсутствует константа. А при OLS-оценивании уравнений вида $z_i = \beta w_i + v_i$ коэффициент детерминации в общем случае не равен квадрату выборочного коэффициента корреляции между переменными $\hat{z}_i = \hat{\beta} w_i$ и z_i. Однако если переменные z_i и w_i <u>центрированы</u>, так что $\bar{z} = \overline{w} = 0$, то такое равенство обеспечивается. В нашем случае переменные $y_{it} - \bar{y}_i$ и $x_{it} - \bar{x}_i$ центрированы, так что коэффициент детерминации, получаемый при оценивании уравнения в отклонениях от средних по группам равен квадрату (выборочного) коэффициента корреляции между переменными $\tilde{y}_{it} = y_{it} - \bar{y}_i$ и

$$\hat{\tilde{y}}_{it} = \hat{\beta}_{CV}\left(x_{it} - \bar{x}_i\right),\ i = 1,\ldots,N,\ t = 1,\ldots,T\ ,$$

где $\hat{\beta}_{CV}$ – “внутри”-оценка для коэффициента β. Об этом коэффициенте детерминации говорят как о “R^2***-внутри***” (R^2-*within*),

$$R^2_{within} = corr^2\left(\hat{\beta}_{CV}\left(x_{it} - \bar{x}_i\right), y_{it} - \bar{y}_i\right).$$

Каждый из этих трех вариантов R^2 является обычным коэффициентом детерминации в соответствующей модели регрессии. В то же время при анализе различных моделей панельных данных часто сообщаются вычисленные значения всех трех вариантов R^2, несмотря на то, что в модели с фиксированными эффектами используется оценка $\hat{\beta}_{CV}$, в модели со случайными эффектами – оценка $\hat{\beta}_{GLS}$, а в модели пула – оценка $\hat{\beta}_{OLS}$.

Более точно, при анализе панельных данных принято сообщать под названиями R^2_{within}, $R^2_{between}$, $R^2_{overall}$ значения

$$R^2_{within} = corr^2\left(y_{it} - \bar{y}_i, \hat{\beta}\left(x_{it} - \bar{x}_i\right)\right),$$

$$R^2_{between} = corr^2\left(\bar{y}_i, \hat{\beta}\,\bar{x}_i\right),$$

$$R^2_{overall} = corr^2\left(y_{it}, \hat{\beta}\,x_{it}\right),$$

вне зависимости от того, каким образом была получена оценка $\hat{\beta}$. Если β является p-мерным вектором, то, соответственно,

$$R^2_{within} = corr^2\left(y_{it} - \bar{y}_i,\ (x_{it} - \bar{x}_i)^T\,\hat{\beta}\right),$$

$$R^2_{between} = corr^2\left(\bar{y}_i,\ \bar{x}_i^T\,\hat{\beta}\right),$$

$$R^2_{overall} = corr^2\left(y_{it},\ x_{it}^T\,\hat{\beta}\right).$$

При этом приводимое значение R^2_{within} является коэффициентом детерминации в обычном смысле, если $\hat{\beta} = \hat{\beta}_{within}$; приводимое значение $R^2_{between}$ является коэффициентом детерминации в обычном смысле, если $\hat{\beta} = \hat{\beta}_{between}$; приводимое значение $R^2_{overall}$ является коэффициентом детерминации в обычном смысле, если $\hat{\beta} = \hat{\beta}_{OLS}$.

П р и м е р (продолжение)

Приведем результаты оценивания в пакете Stata8 моделей с фиксированными и случайными эффектами для данных о трех предприятиях. При этом заметим, что в рамках этого пакета приняты обозначения, отличающиеся от используемых нами: индивидуальные эффекты обозначаются как u_i, а случайные ошибки – как e_{it}. Чтобы избежать путаницы, в приводимые далее протоколы оценивания внесены соответствующие изменения.

Fixed-effects (within) regression

R-sq:
within = 0.9478
between = 0.8567
overall = 0.9209

Проверка значимости регрессии в целом:
F(1,26) = 472.26, Prob > F = 0.0000

Оцененное значение коэффициента корреляции между индивидуальным эффектом и переменной *x* corr(α_i, X) = -0.2311

VARIABLE	Coef.	Std. Err.	t	P>t
x	1.102192	.0507186	21.73	0.000
cons	-1.394325	.8230266	-1.69	0.102

sigma_αlfa	1.480319
sigma_u	1.745136
rho	.4184474 (fraction of variance due to α_i)

F test that all alfa_i=0:
F(2, 26) = 6.81, Prob > F = 0.0042

Последний критерий соответствует гипотезе с двумя линейными ограничениями: поскольку в модель включена постоянная составляющая, то одно линейное ограничение накладывается заранее как идентифицирующее и не подлежащее проверке.

Random-effects GLS regression

R-sq:
within=0.9478
between=0.8567
overall= 0.9209

Random effects: u_i ~ Gaussian

corr(α_i, X) = 0 (предполагается)

Критерий значимости регрессии в целом:

Wald chi2(1) = 325.94, Prob>chi2 = 0.0000

sigma_alfa	0
sigma_u	1.7451362
rho	0 (fraction of variance due to α _i)

Здесь полученная оценка для σ_α^2 оказывается <u>отрицательной</u>; поэтому ее значение полагается равным нулю. Однако тогда модель со случайными эффектами редуцируется к модели “пула”:

y	Coef.	Std. Err.	z	P>z
x	1.058959	.0586557	18.05	0.000
cons	-.7474755	.955953	-0.78	0.434

В то же время, если в рамках модели со случайными эффектами применить критерий Бройша–Пагана для проверки гипотезы об отсутствии таковых эффектов, т.е. гипотезы $H_0 : \sigma_\alpha^2 = 0$, то полученное значение статистики критерия равно 8.47. Этому значению соответствует рассчитанное по асимптотическому распределению хи-квадрат с 1 степенью свободы *P*-значение 0.0036. Но это говорит против гипотезы H_0. И опять это можно объяснить малым количеством наблюдений – ведь распределение хи-квадрат здесь только асимптотическое.

В пакете Stata8 есть возможность оценить модель со случайными эффектами, не прибегая к GLS-оцениванию, а используя метод максимального правдоподобия. Это дает следующие результаты:

Random-effects ML regression
Random effects: α_i ~ Gaussian
Log likelihood = -61.09837
Критерий значимости регрессии в целом:
LR chi2(1) = 121.60, Prob > chi2 = 0.0000

y	Coef.	Std. Err.	z	P>z
x	1.092893	.0501518	21.79	0.000
cons	-1.255205	1.019264	-1.23	0.218
sigma_α	1.064836	.5552752	1.92	0.055
sigma_u	1.713621	.2334960	7.34	0.000
rho	.2785682	.2205921		

Likelihood-ratio test of sigma_alfa=0:
chibar2(01) = 4.70, Prob>=chibar2 = 0.015

По этому критерию гипотеза $H_0 : \sigma_\alpha^2 = 0$ отвергается.

Between regression (regression on group means)

R-sq:
within = 0.9478
between = 0.8567
overall = 0.9209

Проверка значимости регрессии в целом:
F(1,1) = 5.98, Prob > F = 0.2471

y	Coef.	Std. Err.	t	P>t
x	.3137715	.1283133	2.45	0.247
cons	10.40202	1.929616	5.39	0.117

Здесь регрессия оказывается статистически незначимой, а близкое к 1 значение коэффициента детерминации $R^2_{between}$ не должно вводить в заблуждение: для оценивания двух коэффициентов имеется всего 3 наблюдения.

3.5. Выбор между моделями с фиксированными или случайными эффектами.

Прежде всего напомним уже отмеченные ранее особенности моделей с фиксированными или случайными эффектами.

- **FE**: получаемые выводы – **условные** по отношению к значениям эффектов α_i **в выборке;** это соответствует ситуациям, в которых эти значения нельзя рассматривать как случайную выборку из некоторой более широкой совокупности (популяции). Такая интерпретация наиболее подходит для случаев, когда субъектами исследования являются страны, крупные компании или предприятия, т.е. каждый субъект ”имеет свое лицо”.
- **RE**: получаемые выводы – **безусловные** относительно **популяции** всех эффектов α_i. Исследователя интересуют не

конкретные субъекты, а обезличенные субъекты, имеющие заданные характеристики.

Заметим в этой связи, что

- в FE-модели
$$E\left(y_{it} \mid x_{it}\right) = E\left(\alpha_i + \beta x_{it} + u_{it} \middle| x_{it}\right) = \alpha_i + \beta x_{it} \, .$$

- в RE-модели $E\left(y_{it} \mid x_{it}\right) = E\left(\mu + \alpha_i + \beta x_{it} + u_{it} \middle| x_{it}\right) = \mu + \beta x_{it}$

Напомним, теперь, что RE модель предполагает, в частности, что $E\left(\alpha_i x_{it}\right) = 0$. Чтобы избавиться от этого условия ортогональности, предположим, что

$$\alpha_i = a\,\bar{x}_i + \varepsilon_i,\ \varepsilon_i \sim N\left(0, \sigma_\varepsilon^2\right),\ \ E\left(\varepsilon_i x_{it}\right) = 0\,.$$

Это приводит к ***“модели Мундлака”***:

$$y_{it} = \mu + \beta x_{it} + a\,\bar{x}_i + \varepsilon_i + u_{it}\,,$$

которая также является моделью компонент ошибки. Она отличается от предыдущей модели тем, что в правую часть добавляется переменная $\bar{x}_i$, изменяющаяся только от субъекта к субъекту и отражающая неоднородность субъектов. Эта переменная, в отличие от α_i, наблюдаема. Заметим, что в модели Мундлака

$$E\left(\alpha_i x_{it}\right) = E\left(\left(a\,\bar{x}_i + \varepsilon_i\right)x_{it}\right) = aE\left(\bar{x}_i x_{it}\right) =$$
$$= \frac{a}{T}\sum_{s=1}^{T} E\left(x_{is}x_{it}\right) = \frac{a}{T}\left(E\left(x_{it}^2\right) + \sum_{s=1,\, s\neq t}^{T} E\left(x_{is}x_{it}\right)\right),$$

так что если $a \neq 0$, то условие $E\left(\alpha_i x_{it}\right) = 0$ в общем случае не выполняется ни для одного $i = 1, \ldots, N$.

Применение GLS к этой модели дает BLUE оценки для β и a, имеющие вид:

$$\hat{\beta}_{GLS}^{*} = \hat{\beta}_{CV}\,,\ \ a_{GLS}^{*} = \hat{\beta}_b - \hat{\beta}_{CV}\,,$$

и

$$\hat{\mu}^*_{GLS} = \bar{y} - \bar{x}\hat{\beta}_b .$$

Иначе говоря, BLUE оценкой для β в этой модели является ковариационная (внутригрупповая) оценка, и из $E\left(\hat{a}^*_{GLS}\right) = a$, $E\left(\hat{a}^*_{GLS}\right) = E\left(\hat{\beta}_b\right) - E\left(\hat{\beta}_{CV}\right)$ и $E\left(\hat{\beta}_{CV}\right) = \beta$, получаем:

$$E\left(\hat{\beta}_b\right) = a + \beta .$$

Как было показано выше, в RE-модели (предполагающей выполнение условия $E(\alpha_i x_{it}) = 0$)

$$\hat{\beta}_{RE} = w\hat{\beta}_b + (1-w)\hat{\beta}_{CV} .$$

Если использовать эту же оценку в модели Мундлака, то для нее получим:

$$E\left(\hat{\beta}_{RE}\right) = wE\left(\hat{\beta}_b\right) + (1-w)E\left(\hat{\beta}_{CV}\right) = w(\beta + a) + (1-w)\beta = \beta + wa ,$$

так что если $a \neq 0$, то $\hat{\beta}_{RE}$ – смещенная оценка.

Критерии спецификации

Речь здесь идет о том, совпадает ли условное распределение α_i при заданном x_i с безусловным распределением α_i (в более слабой форме: выполняется ли соотношение $E\left(\alpha_i \mid x_i\right) = 0$ – тогда и $E\left(\alpha_i x_i\right) = 0$). Если нет, то наилучшей оценкой является $\hat{\beta}_{CV}$ (FE). Если да, то наилучшей оценкой является $\hat{\beta}_{GLS}$ (RE).

Критерий 1

Используя формулировку Мундлака, проверяем гипотезу $H_0 : a = 0$ против альтернативы $H_1 : a \neq 0$.

Критерий 2 – критерий Хаусмана (Hausman)

Проверяемая гипотеза: $H_0 : E\left(\alpha_i \middle| x_i\right) = 0$, альтернативная гипотеза: $H_1 : E\left(\alpha_i \middle| x_i\right) \neq 0$.

Идея этого критерия основывается на следующих фактах:

- При гипотезе H_0 и оценка $\hat{\beta}_{GLS}$, соответствующая RE-модели, и оценка $\hat{\beta}_{CV}$, соответствующая FE-модели, – состоятельны.
- При гипотезе H_1 оценка $\hat{\beta}_{GLS}$ несостоятельна, а оценка $\hat{\beta}_{CV}$ состоятельна.

Соответственно, если гипотеза H_0 верна, то между оценками $\hat{\beta}_{GLS}$ и $\hat{\beta}_{CV}$ не должно наблюдаться систематического расхождения, и эта гипотеза должна отвергаться при “слишком больших” по абсолютной величине значениях разности $\hat{\beta}_{CV} - \hat{\beta}_{GLS}$ (больших в сравнении со стандартной ошибкой этой разности).

Пусть $\hat{q} = \hat{\beta}_{CV} - \hat{\beta}_{GLS}$; тогда из общей формулы для дисперсии суммы двух случайных величин следует, что

$$D(\hat{q}) = D\left(\hat{\beta}_{CV} - \hat{\beta}_{GLS}\right) =$$
$$= D\left(\hat{\beta}_{CV}\right) + D\left(\hat{\beta}_{GLS}\right) - 2Cov\left(\hat{\beta}_{CV}, \hat{\beta}_{GLS}\right).$$

Если выполняются предположения стандартной RE-модели, то как было указано выше, $\hat{\beta}_{GLS}$ является эффективной оценкой, а $\hat{\beta}_{CV}$ – неэффективной оценкой. Хаусман показал, что эффективная оценка не коррелирована с разностью ее и неэффективной оценки, так что если гипотеза H_0 верна, то

$$Cov\left(\hat{\beta}_{GLS}, \hat{\beta}_{GLS} - \hat{\beta}_{CV}\right) = D\left(\hat{\beta}_{GLS}\right) - Cov\left(\hat{\beta}_{GLS}, \hat{\beta}_{CV}\right) = 0,$$
$$Cov\left(\hat{\beta}_{GLS}, \hat{\beta}_{CV}\right) = D\left(\hat{\beta}_{GLS}\right),$$

и

$$D(\hat{q}) = D\left(\hat{\beta}_{CV}\right) - D\left(\hat{\beta}_{GLS}\right).$$

Как уже говорилось выше,

$$D\left(\hat{\beta}_{GLS}\right) = \frac{\sigma_u^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)^2 + \Psi\sum_{i=1}^{N}\left(\bar{x}_i - \bar{x}\right)^2},$$

$$D\left(\hat{\beta}_{CV}\right)=\frac{\sigma_u^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(x_{it}-\bar{x}_i\right)^2}.$$

Заменяя в этих выражениях неизвестные параметры их состоятельными оценками, указанными ранее, получаем состоятельную оценку $\hat{D}(\hat{q})$ для $D(\hat{q})$.

Статистика критерия Хаусмана

$$m=\frac{\hat{q}^2}{\hat{D}(\hat{q})}$$

имеет при гипотезе H_0 асимптотическое (N $\to \infty$) распределение $\chi^2(1)$.

Для K регрессоров при гипотезе H_0 статистика

$$m=\hat{q}^T\,[C\hat{o}v(\hat{q})]^{-1}\,\hat{q}$$

имеет асимптотическое распределение $\chi^2(K)$.

Численно идентичный критерий для проверки гипотезы H_0 получается с использованием расширенной модели регрессии

$$y_{it}^*=x_{it}^*\beta+\left(x_{it}-\bar{x}_i\right)\gamma+\varepsilon_{it},$$

где

$$y_{it}^*=y_{it}-\theta\,\bar{y}_i,\quad x_{it}^*=x_{it}-\theta\,\bar{x}_i,$$

$$\theta=1-\sqrt{\Psi}=1-\sqrt{\frac{\sigma_u^2}{\sigma_u^2+T\sigma_\alpha^2}}.$$

Гипотеза H_0 означает в этой регрессии, что $\gamma=0$. Можно показать, что здесь

$$\hat{\beta}_{OLS}=\hat{\beta}_b,\quad \hat{\gamma}_{OLS}=\hat{\beta}_{CV}-\hat{\beta}_b.$$

Гипотезу H_0 можно также проверить, используя любую из следующих разностей:

$$\hat{q}_1 = \hat{\beta}_{GLS} - \hat{\beta}_{CV},$$
$$\hat{q}_2 = \hat{\beta}_{GLS} - \hat{\beta}_b,$$
$$\hat{q}_3 = \hat{\beta}_{CV} - \hat{\beta}_b,$$
$$\hat{q}_4 = \hat{\beta}_{GLS} - \hat{\beta}_{OLS}.$$

Это вытекает из соотношения

$$\hat{\beta}_{GLS} = w\hat{\beta}_b + (1-w)\hat{\beta}_{CV}.$$

З а м е ч а н и е

Все входящие в эти разности оценки параметра β состоятельны при гипотезе H_0; поэтому все эти разности должны сходиться при гипотезе H_0 к нулю.

П р и м е р (продолжение примера с тремя предприятиями) Применим критерий Хаусмана.

	Coefficients			
	(b) fix	(B)	(b-B)	S.E.
x	1.102192	1.058959	.0432328	

Здесь b = $\hat{\beta}_{CV}$, B = $\hat{\beta}_{GLS}$, (b-B) = $\hat{q} = \hat{\beta}_{CV} - \hat{\beta}_{GLS}$.

Статистика критерия Хаусмана равна

chi2(1) = (b-B)'[(V_b-V_B)^(-1)](b-B) =-2.15,

где V_b и V_B – состоятельные оценки дисперсий оценок $\hat{\beta}_{CV}$ и $\hat{\beta}_{GLS}$, соответственно. Поскольку значение этой статистики оказалось отрицательным, критерий Хаусмана применить не удается.

П р и м е р (объяснение размеров заработной платы)

Статистические данные (из National Longitudinal Survey, Youth Sample, США) содержат сведения о 545 полностью занятых мужчинах, окончивших школу до 1980 г. и наблюдавшихся в течение 1980–1987 гг. В 1980 г. эти мужчины имели возраст в пределах от 17 до 23 лет и включились в рынок труда совсем недавно, так что на начало периода их трудовой стаж составлял в среднем около 3 лет. Логарифмы заработной платы объясняются здесь длительностью школьного обучения, трудовым стажем и его квадратом, а также дамми-переменными, указывающими на членство в профсоюзе, работу в государственном секторе, семейный статус (состоит ли в браке), а также на цвет кожи (чернокожий или нет) и испаноязычность.

Результаты оценивания:

Переменная	Оценка			
	Between	FE	OLS	RE
Const	0.490 (0.221)	–	-0.034 (0.065)	-0.104 (0.111)
scooling	0.095 (0.011)	–	0.099 (0.005)	0.101 (0.009)
experience	-0.050 (0.050)	0.116 (0.008)	0.089 (0.010)	0.112 (0.008)
experience^2	0.005 (0.003)	-0.0043 (0.0006)	-0.0028 (0.0007)	-0.0041 (0.0006)
Union member	0.274 (0.047)	0.081 (0.019)	0.180 (0.017)	0.106 (0.018)
married	0.145 0.041	0.045 (0.018)	0.108 (0.016)	0.063 (0.017)
black	-0.139 (0.049)	–	-0.144 (0.024)	-0.144 (0.048)
hispanic	0.005 (0.043)	–	0.016 (0.021)	0.020 (0.043)
Public sector	-0.056 (0.109)	0.035 (0.039)	0.004 (0.037)	0.030 (0.036)
Within R^2 Between R^2 Overall R^2	0.0470 0.2196 0.1371	0.1782 0.0006 0.0642	0.1679 0.2027 0.1866	0.1776 0.1835 0.1808

RE=EGLS (FGLS) соответствует θ = 0.64
FE=CV соответствует θ = 1
OLS соответствует θ = 0
Значения RE и OLS заключены между Between и FE.

Если выполнены предположения модели со <u>случайными</u> эффектами, то все четыре оценки состоятельны. Если, однако, индивидуальные эффекты α_i <u>коррелированы</u> хотя бы с одной из объясняющих переменных, то состоятельной остается только FE-оценка. Поэтому встает вопрос о проверке гипотезы H_0 о том, что модель является RE-моделью. Для этого можно сравнивать оценки “внутри” (FE) и “между” или оценки “внутри” (FE) и RE (соответствующие критерии равносильны). Проще сравнивать вторую пару, используя критерий Хаусмана, описанный ранее.

	Coefficients			
	(b)	(B)	(b-B)	sqrt(diag(V_b-V_B))
	fix			S.E.
EXPER	.116457	.1117851	.0046718	.0016345
EXPER2	-.0042886	.0040575	-.0002311	.0001269
MAR	.0451061	.0625465	-.0174403	.0073395
PUB	.0349267	.0301555	.0047713	.0126785
UNION	.081203	.1064134	-.0252104	.0073402

chi2(5) = (b-B)'[(V_b-V_B)^(-1)](b-B) = 31.75
Prob>chi2 = 0.0000

Вычисленное значение статистики этого критерия равно 31.75 и отражает различия в FE- и RE-оценках коэффициентов при 5 переменных: *experience, experience*^{2}*, Union member, married, Public sector*. Для распределения $\chi^2(5)$ значение 31.75 соответствует P-значению $6.6\cdot10^{-6}$, так что нулевая гипотеза (RE-модель) заведомо отвергается. Чем это может объясняться?

Ненаблюдаемая гетерогенность α_i может, например, коррелировать с семейным положением, что подтверждается данными приведенной таблицы. Если из модели выметаются индивидуальные эффекты и используется FE-оценка, то влияние этого фактора уменьшается до 4.5%, тогда как для “between”-оценки это влияние составляет 14.5%. Заметим, что влияние этого фактора идентифицируется только благодаря наличию в выборке индивидов, изменивших свое семейное положение в период наблюдений. Аналогичные замечания справедливы и в отношении членства в профсоюзе.

Следует, впрочем, заметить, что все использованные оценки предполагают некоррелированность объясняющих переменных с u_{it}, и при нарушении этого условия даже FE-оценка оказывается несостоятельной.

Приведенные в конце таблицы значения трех различных мер согласия показывает, что наибольшее значение *within*-R^2 достигается для FE-оценки, т.е. именно она наилучшим образом объясняет внутригрупповую изменчивость. OLS-оценка максимизирует обычный *overall*-R^2. RE-оценка дает близкие значения всем трем мерам согласия. Отметим еще, что стандартные ошибки OLS-оценки вообще-то говоря не соответствуют действительности, т.к. не принимают во внимание коррелированность ошибок между собой. Правильные их значения должны быть больше стандартных ошибок EGLS-оценок, принимающих во внимание эту коррелированность.

3.6. Автокоррелированные ошибки

Во всех рассмотренных выше ситуациях предполагалось, что случайные составляющие u_{it} – взаимно независимые случайные величины, имеющие одинаковое распределение $N\left(0,\sigma_u^2\right)$. Между тем, вполне возможно, что для i-го субъекта последовательные ошибки $u_{i1},u_{i2},\ldots,u_{iT}$ не являются независимыми, а следуют, скажем, стационарному процессу авторегрессии первого порядка с нулевым средним.

Точнее говоря, пусть мы имеем дело с моделью

$$y_{it} = \mu + \alpha_i + \beta x_{it} + u_{it}\,,\ i = 1,\ldots,N,\ t = 1,\ldots,T\,,$$

в которой

$$u_{it} = \rho u_{i,t-1} + \varepsilon_{it}\,,$$

где $|\rho| < 1$, а случайные величины $\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT}$ являются гауссовскими ***инновациями***, так что они взаимно независимы и имеют одинаковое распределение $N(0, \sigma_\varepsilon^2)$ и, кроме того, ε_{it} не зависит от значений $u_{i,t-k}$, $k \ge 1$. Общий для всех субъектов коэффициент ρ можно оценить различными способами. При этом в большинстве случаев сначала переходят к модели, скорректированной на групповые средние:

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i)\,,$$

т.е.

$$\tilde{y}_{it} = \beta \tilde{x}_{it} + \tilde{u}_{it}\,,$$

а затем поступают по-разному.

Можно, например, оценить (методом наименьших квадратов) последнюю модель без учета автокоррелированности ошибок, получить последовательность остатков $\hat{\tilde{u}}_{i1}, \hat{\tilde{u}}_{i2}, \ldots, \hat{\tilde{u}}_{iT}$, вычислить значение статистики Дарбина–Уотсона

$$d = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\hat{\tilde{u}}_{it} - \hat{\tilde{u}}_{i,t-1}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\tilde{u}}_{it}^2}$$

и, используя приближенное соотношение $\rho \cong 1 - d/2$, получить оценку $\hat{\rho}_{DW} \cong 1 - d/2$.

Можно поступить иначе: получив последовательность остатков $\hat{\tilde{u}}_{i1}, \hat{\tilde{u}}_{i2}, \ldots, \hat{\tilde{u}}_{iT}$, использовать оценку наименьших квадратов, получаемую при оценивании уравнения регрессии

$$\hat{\tilde{u}}_{it} = \rho \hat{\tilde{u}}_{i,t-1} + \eta_{it} .$$

Искомая оценка вычисляется по формуле:

$$\hat{\rho} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \hat{\tilde{u}}_{it} \, \hat{\tilde{u}}_{i,t-1}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\tilde{u}}_{it}^2} .$$

[В пакете Stata8 эта оценка обозначена как $\hat{\rho}_{tscorr}$.]

После получения оценки для ρ производится преобразование переменных, призванное получить модель с независимыми ошибками. Наконец, в рамках преобразованной модели производится обычный анализ на фиксированные или случайные эффекты.

Пример

В примере с тремя предприятиями для модели с фиксированными эффектами получаем следующие результаты.

<u>При использовании DW-оценки:</u>

. xtregar y x, fe rhotype(dw) lbi -

FE (within) regression with AR(1) disturbances

Number of obs = 27

R-sq:
within = 0.9569
between = 0.1111
overall = 0.9252

F(1,23) = 510.54, Prob > F = 0.0000

corr(α_i, Xb) = -0.1625

y	Coef.	Std. Err.	t	P>t
x	1.105796	.0489398	22.60	0.000
cons	-1.317209	.6964634	-1.89	0.071

rho_ar	.170171
sigma_α	1.423608
sigma_u	1.773845
rho_fov	.3917622 (доля дисперсии, соответствующей индивидуальным эффектам α_i)

F test that all alfa_i=0:
F(2,23)= 3.82, Prob > F = 0.0370

modified Bhargava et al. Durbin-Watson = 1.664958
Baltagi-Wu LBI = 1.8196862

Оценка коэффициента β по сравнению со значением 1.102192, полученным ранее без учета возможной автокоррелированности ошибок, практически не изменилась. И это согласуется со значением статистики Дарбина–Уотсона. Вывод об отсутствии индивидуальных эффектов также не изменяется.

При использовании *tscorr*-оценки:

. xtregar y x, fe rhotype(tscorr)
FE (within) regression with AR(1) disturbances
Number of obs = 27

R-sq: within = 0.9540
between = 0.1111
overall = 0.9252

F(1,23)= 476.47, Prob > F = 0.0000
corr(α_i, Xb) = -0.1626

y	Coef.	Std. Err.	t	P>t
x	1.109267	.050818	21.83	0.000
cons	-1.392025	.7698403	-1.81	0.084

rho_ar	.09213053
sigma_α	1.4281087
sigma_u	1.7701594
rho_fov	.39426073 (доля дисперсии, соответствующей индивидуальным эффектам α_i))

F test that all alfa_i=0:
F(2,23) = 4.66, Prob > F = 0.0199

Оцененное значение коэффициента автокорреляции ρ на этот раз почти в два раза меньше. Оценка коэффициента β практически не изменилась.

3.7. Двухфакторные (двунаправленные) модели

3.7.1. Фиксированные эффекты

Здесь мы рассматриваем модель, в которую помимо ***индивидуальных эффектов*** α_i включаются также и ***временные эффекты*** λ_t:

$$y_{it} = \mu + \alpha_i + \lambda_t + \beta x_{it} + u_{it}, \quad i = 1, \ldots, N, \; t = 1, \ldots, T,$$

где $\sum_{i=1}^{N} \alpha_i = 0$, $\sum_{t=1}^{T} \lambda_t = 0$, так что α_i и λ_t – ***“дифференциальные эффекты”***. При этом и α_i и λ_t интерпретируются как неизвестные постоянные.

Обозначая

$$\bar{y}_t = \frac{1}{N}\sum_{i=1}^{N} y_{it} \quad \text{и т.д. (средние по субъектам),}$$

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it} \quad \text{и т.д. (средние по времени),}$$

$$\bar{y} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}$$ (среднее по всем наблюдениям),

получаем:

$$(y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}) = (x_{it} - \bar{x}_i - \bar{x}_t + \bar{x})\beta + (u_{it} - \bar{u}_i - \bar{u}_t + \bar{u}).$$

Оценка наименьших квадратов для коэффициента β в этом уравнении (***двухфакторная внутригрупповая оценка***) имеет вид

$$\hat{\beta}_{CV} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{it} - \bar{y}_i - \bar{y}_t + \bar{y})(x_{it} - \bar{x}_i - \bar{x}_t + \bar{x})}{\sum_{i=1}^{N}\sum_{t=1}^{T}(x_{it} - \bar{x}_i - \bar{x}_t + \bar{x})^2} = \frac{W_{xy}}{W_{xx}}.$$

На основании соотношений

$$(\bar{y}_i - \bar{y}) = \alpha_i + (\bar{x}_i - \bar{x})\beta + (\bar{u}_i - \bar{u}),$$

$$(\bar{y}_t - \bar{y}) = \lambda_t + (\bar{x}_t - \bar{x})\beta + (\bar{u}_t - \bar{u})$$

можно получить оценки для α_i и λ_t:

$$\hat{\alpha}_i = (\bar{y}_i - \bar{y}) - \hat{\beta}_{CV}(\bar{x}_i - \bar{x}),$$

$$\hat{\lambda}_t = (\bar{y}_t - \bar{y}) - \hat{\beta}_{CV}(\bar{x}_t - \bar{x}).$$

Для оценивания двунаправленной модели с фиксированными эффектами в пакете Stata применяется процедура *xtreg, fe* с включением дамми-переменных для временных периодов.

П р и м е р

(объяснение размеров заработной платы, продолжение)

Fixed-effects (within) regression

R-sq:
within = 0.1808
between = 0.0005
overall = 0.0638

F(11,3804) = 76.30, Prob > F = 0.0000

corr(α_i, Xb) = -0.1203

WAGE_LN	Coef.	Std. Err.	t	P>t
BLACK (dropped)				
EXPER	.1317361	.0098356	13.39	0.000
EXPER2	-.0051704	.0007047	-7.34	0.000
HISP (dropped)				
MAR	.0464781	.0183123	2.54	0.011
PUB	.0347278	.0385989	0.90	0.368
SCHOOL (dropped)				
UNION	.0791253	.0193354	4.09	0.000
Y80 (dropped)				
Y81	.0193045	.0203652	0.95	0.343
Y82	-.0112773	.0202281	-0.56	0.577
Y83	-.0419533	.0203211	-2.06	0.039
Y84	-.0383904	.0203151	-1.89	0.059
Y85	-.0428743	.0202506	-2.12	0.034
Y86	-.0275581	.0203878	-1.35	0.177
cons	1.028383	.0299620	34.32	0.000

Заметим, что значимыми оказываются здесь оцененные коэффициенты при тех же переменных что и в однонаправленной модели с фиксированными эффектами.

sigma_α .40078197
sigma_u .3509988
rho .56593121 (fraction of variance due to α_i)

F test that all α_i=0:
F(544, 3804) = 7.97 Prob > F = 0.0000

. test Y81=Y82=Y83=Y84=Y85=Y86=0

(Здесь Y81 – дамми-переменная, равная 1 в 1981 г. и равная 0 в остальные годы; аналогично определяются Y82,…,Y86.)

F(6, 3804) =	1.96
Prob > F =	0.0680

Результаты последнего теста не выявляют значимого влияния временных эффектов.

3.7.2. Случайные эффекты

В этой двухфакторной модели

$$y_{it} = \mu + \alpha_i + \lambda_t + \beta x_{it} + u_{it}, \quad i = 1, \ldots, N, \; t = 1, \ldots, T$$

предполагается, что α_i и λ_t – случайные величины и что

$$E(\alpha_i) = E(\lambda_t) = E(u_{it}) = 0,$$

$$E(\alpha_i \alpha_j) = \begin{cases} \sigma_\alpha^2, & \text{если } i = j, \\ 0, & \text{если } i \neq j, \end{cases}$$

$$E(\lambda_t \lambda_s) = \begin{cases} \sigma_\lambda^2, & \text{если } t = s, \\ 0, & \text{если } t \neq s, \end{cases}$$

$$E(u_{it} u_{js}) = \begin{cases} \sigma_u^2, & \text{если } i = j \text{ и } t = s, \\ 0, & \text{в противном случае,} \end{cases}$$

$$E(\alpha_i \lambda_t) = 0, \; E(\alpha_i u_{it}) = 0, \; E(\lambda_t u_{it}) = 0,$$

$$E(\alpha_i x_{it}) = E(\lambda_t x_{it}) = E(u_{it} x_{it}) = 0.$$

Определим $\nu_{it} = \alpha_i + \lambda_t + u_{it}$. Тогда

$$D(\nu_{it}) = D(y_{it} | x_{it}) = \sigma_\alpha^2 + \sigma_\lambda^2 + \sigma_u^2,$$

так что ошибка состоит из трех компонент. При этом

$$Cov(\nu_{it}, \nu_{js}) = \begin{cases} \sigma_\alpha^2, & \text{если } i = j, t \neq s, \\ \sigma_\lambda^2, & \text{если } t = s, i \neq j, \end{cases}$$

$$Corr\left(\nu_{it}, \nu_{js}\right) = \begin{cases} \dfrac{\sigma_\alpha^2}{D(\nu_{it})}, & \text{если } i = j,\ t \neq s, \\ \dfrac{\sigma_\lambda^2}{D(\nu_{it})}, & \text{если } t = s,\ i \neq j, \\ 1, & \text{если } t = s,\ i = j, \\ 0, & \text{если } t \neq s,\ i \neq j. \end{cases}$$

"Межсубъектная" оценка определяется как

$$\hat{\beta}_{bi} = \frac{\sum_{i=1}^{N} (\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})}{\sum_{i=1}^{N} (\bar{x}_i - \bar{x})^2} = \frac{B_{xy}}{B_{xx}},$$

а "межвременная" оценка определяется как

$$\hat{\beta}_{bt} = \frac{\sum_{t=1}^{T} (\bar{x}_t - \bar{x})(\bar{y}_t - \bar{y})}{\sum_{t=1}^{T} (\bar{x}_t - \bar{x})^2} = \frac{C_{xy}}{C_{xx}}.$$

GLS-оценка равна

$$\hat{\beta}_{GLS} = \omega_1 \hat{\beta}_{CV} + \omega_2 \hat{\beta}_{bi} + \omega_3 \hat{\beta}_{bt},$$

где

$$\omega_1 = \frac{W_{xx}}{W_{xx} + \phi_2^2 B_{xx} \phi_3^2 C_{xx}} = \frac{W_{xx}}{T_{xx}},$$

$$\omega_2 = \frac{\phi_2^2 B_{xx}}{T_{xx}}, \quad \omega_3 = \frac{\phi_3^2 C_{xx}}{T_{xx}},$$

$$\phi_2^2 = \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}, \quad \phi_3^2 = \frac{\sigma_u^2}{\sigma_u^2 + N\sigma_\lambda^2},$$

$$T_{xx} = W_{xx} + \phi_2^2 B_{xx} + \phi_3^2 C_{xx}.$$

Иначе говоря, GLS-оценка является взвешенным средним одной “внутри”- и двух “между”-оценок, с весами, отражающими источники изменчивости.

- Если $\sigma_\alpha^2 = \sigma_\lambda^2 = 0$ (так что все α_i и λ_t равны нулю), то

$$\phi_2^2 = \phi_3^2 = 1, \quad \omega_2 = \frac{B_{xx}}{T_{xx}}, \quad \omega_3 = \frac{C_{xx}}{T_{xx}}, \quad T_{xx} = W_{xx} + B_{xx} + C_{xx},$$

$$\hat{\beta}_{GLS} = \frac{W_{xx}}{T_{xx}} \cdot \frac{W_{xy}}{W_{xx}} + \frac{B_{xx}}{T_{xx}} \cdot \frac{B_{xy}}{B_{xx}} + \frac{C_{xx}}{T_{xx}} \cdot \frac{C_{xy}}{C_{xx}} = \frac{W_{xy} + B_{xy} + C_{xy}}{W_{xx} + B_{xx} + C_{xx}} =$$

$$= \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x})(y_{it} - \bar{y})}{\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} - \bar{x})^2} = \hat{\beta}_{OLS} \text{ (как для пула).}$$

- При $T \to \infty$ и $N \to \infty$ имеем $\phi_2^2 \to 0$ и $\phi_3^2 \to 0$, и
  $\hat{\beta}_{GLS} = \hat{\beta}_{CV}$ (как для модели с фиксированными эффектами).

Можно также комбинировать фиксированные временные эффекты и случайные индивидуальные эффекты в рамках процедуры *xtreg, re* с включением dummy-переменных для временных периодов.

3.7.3. Критерии для индивидуальных и временных эффектов

Критерий Бройша–Пагана. Это критерий для проверки гипотезы

$H_0 : \sigma_\alpha^2 = \sigma_\lambda^2 = 0$ (сведение к модели пула).

Здесь опять можно использовать OLS-оценки параметров для получения OLS-остатков $\hat{u}_{it}$.

Статистика критерия $BP = LM_1 + LM_2$, где

$$LM_1 = \frac{NT}{2(T-1)}\left[\frac{T^2 \sum_{i=1}^{N} \bar{\hat{u}}_i^2}{\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{u}_{it}^2} - 1\right]^2,$$

$$LM_2 = \frac{NT}{2(T-1)}\left[\frac{N^2 \sum_{t=1}^{T} \bar{\hat{u}}_t^2}{\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{u}_{it}^2} - 1\right]^2.$$

При гипотезе H_0 статистика BP имеет асимптотическое распределение $\chi^2(2)$. При гипотезе $H_0^a : \sigma_\alpha^2 = 0 \quad LM_1 \sim \chi^2(1)$; при гипотезе $H_0^b : \sigma_\lambda^2 = 0 \quad LM_2 \sim \chi^2(1)$.

В качестве альтернативных могут быть использованы F-критерии, как при проверке гипотез для фиксированных эффектов.

- Для проверки гипотезы $H_0^\alpha : \sigma_\alpha^2 = 0$ (при условии $\lambda_t = 0$) используем уже применявшуюся ранее статистику

$$F_3 = \frac{(S_2 - S_1)/(N-1)}{S_1/(NT - N - 1)},$$

где S_2 – RSS от OLS регрессии (пул), а S_1 – RSS от однофакторной “внутригрупповой” регрессии (основанной на однофакторной CV-оценке).

- Для проверки гипотезы

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_{N-1} = 0, \;\; \lambda_1 = \lambda_2 = \cdots = \lambda_{T-1} = 0$$

используем статистику

$$F_{2-way} = \frac{(S_3 - S_{2w})/(N + T - 2)}{S_{2w}/(N-1)(T-1)-1},$$

где S_{2w} – RSS от двухфакторной “внутригрупповой” регрессии.

- Для проверки гипотезы

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_{N-1} = 0 \ \text{ при } \lambda_t \neq 0$$

используем статистику

$$F_4 = \frac{(S_4 - S_{2w})/(N-1)}{S_{2w}/((N-1)(T-1)-1)},$$

где S_4 – RSS от регрессии

$$(y_{it} - \bar{y}_t) = (x_{it} - \bar{x}_t)\beta + (u_{it} - \bar{u}_t)$$

- Для проверки гипотезы

$$H_0 : \lambda_1 = \lambda_2 = \cdots = \lambda_{T-1} = 0 \ \text{ при } \ \alpha_i \neq 0$$

используем статистику

$$F_5 = \frac{(S_1 - S_{2w})/(T-1)}{S_{2w}/((N-1)(T-1)-1)}.$$

Может иметь смысл трактовка α_i как случайных, а λ_t как фиксированных эффектов: STATA *xtreg, re* включает dummies как фиксированные эффекты.

Для рассматриваемого примера:

В рамках модели с фиксированными индивидуальными и временными эффектами проверка гипотезы об отсутствии индивидуальных эффектов дает следующие результаты:

F test that all ALFA_i=0:		
F(544, 3804) =	7.97	Prob > F = 0.0000

Полученное значение F-статистики приводит к отвержению этой гипотезы на любом разумном уровне значимости.

Проверка в рамках той же модели гипотезы об отсутствии временных эффектов приводит к следующему:

test Y81=Y82=Y83=Y84=Y85=Y86=0
F(6, 3804) = 1.96
Prob > F = 0.0680

Гипотеза отсутствия фиксированных временных эффектов не отвергается F-критерием на 5% уровне значимости.

3.8. Несбалансированные панели

Мы сосредоточимся здесь на однофакторной модели. Такая модель могла бы включать временные дамми как фиксированные эффекты.

До сих пор мы предполагали, что

$i = 1, \ldots, N$, $t = 1, \ldots, T$,

так что в каждый из T моментов времени имеются данные о всех N субъектах, участвующих в анализе. В таких случаях говорят о ***сбалансированной панели***. Теперь мы рассмотрим модель

$$y_{it} = \mu + \beta x_{it} + \alpha_i + u_{it}\,,\;\; i = 1, \ldots, N,\, t = 1, \ldots, T_i,\;\; \sum_{i=1}^{N} \alpha_i = 0\,,$$

в которой количество наблюдений для разных субъектов может быть различным. В этом случае говорят о ***несбалансированной панели***.

Основные результаты для несбалансированных панелей.

1. OLS-оценка та же, что и раньше, и она является BLUE, если $\sigma_\alpha^2 = 0$.
2. Внутригрупповая (CV) оценка в основных чертах та же, что и ранее, хотя $\bar{y}_i$ и $\bar{x}_i$ вычисляются по периодам времени разной длины для разных субъектов.
3. “Между”-оценка также в основном сохраняется, только средние вычисляются по T_i наблюдениям для субъекта i.

4. GLS-оценка по-существу сохраняется, но преобразование переменных имеет вид $y_{it}^{*} = y_{it} - \theta_i \, \bar{y}_i$, $x_{it}^{*} = x_{it} - \theta_i \, \bar{x}_i$, где

$$\bar{y}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} y_{it}\,,\quad \bar{x}_i = \frac{1}{T_i}\sum_{t=1}^{T_i} x_{it}\,,\ \theta_i = 1 - \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T_i\sigma_\alpha^2}}\,,$$

т. е. θ_i <u>изменяется</u> от субъекта к субъекту.

3.9. Эндогенные объясняющие переменные

Здесь мы рассмотрим модель с несколькими объясняющими переменными, некоторые из которых являются эндогенными. Именно, мы рассматриваем модель

$$y_{2it} = y_{1it}\gamma + x_{it}\beta + \alpha_i + u_{it}\,,\quad i = 1,\ldots,N,\ t = 1,\ldots,T,$$

где

y_{1it} – вектор-строка g_1 эндогенных переменных,

x_{it} – вектор-строка k_1 экзогенных переменных,

γ и β – векторы-столбцы размерностей g_1 и k_1.

Пусть

z_{it} – вектор-строка k_2 инструментальных переменных, $k_2 \geq g_1$,

так что

$E(z_{it}u_{it}) = 0$.

Обозначим:

$$\bar{y}_{2i} = \frac{1}{T}\sum_{t=1}^{T} y_{2it}\,,\ \bar{y}_{1i} = \frac{1}{T}\sum_{t=1}^{T} y_{1it}\,,\ \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} \bar{x}_{it}\,,$$

$$\bar{y}_2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{2it}\,,\ \bar{y}_1 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{1it}\,,\ \bar{x} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}\,.$$

Оценивая методом инструментальных переменных (метод IV) “внутри”-регрессию

$$(y_{2it} - \bar{y}_{2i}) = (y_{1it} - \bar{y}_{1i})\gamma + (x_{it} - \bar{x}_i)\beta + (u_{it} - \bar{u}_i)\,,$$

получим ***IV-“внутри”-оценки*** $\hat{\gamma}_{IVW}$, $\hat{\beta}_{IVW}$. Оценивая методом инструментальных переменных “между”-регрессию

$$(\bar{y}_{2i}-\bar{y}_2)=(\bar{y}_{1i}-\bar{y}_1)\gamma+(\bar{x}_i-\bar{x})\beta+(\bar{u}_i-\bar{u}),$$

получим ***IV-“между”-оценку*** $\hat{\gamma}_{IVB}$.

Пример: $k_2=g_1=1$, $k_1=0$, так что $y_{2it}=y_{1it}\gamma+\alpha_i+u_{it}$.
Здесь

$$\hat{\gamma}_{IVW}=\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{2it}-\bar{y}_{2i})(z_{it}-\bar{z}_i)}{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{1it}-\bar{y}_{1i})(z_{it}-\bar{z}_i)},$$

$$\hat{\gamma}_{IVB}=\frac{\sum_{i=1}^{N}\bar{y}_{2i}\bar{z}_i}{\sum_{i=1}^{N}\bar{y}_{1i}\bar{z}_i}.$$

Если $E(Y_{1it}\alpha_i)=0$, то тогда более эффективна оценка со случайными эффектами. Чтобы получить ее, используем преобразование переменных

$$\tilde{y}_{2it}=y_{2it}-\hat{\theta}\,\bar{y}_{2i}\ ,\ \tilde{y}_{1it}=y_{1it}-\hat{\theta}\,\bar{y}_{1i}\,,$$

$$\tilde{u}_{it}=u_{it}-\hat{\theta}\,\bar{u}_i\,,\ \tilde{z}_{it}=z_{it}-\hat{\theta}\,\tilde{z}_i\,,$$

где

$$\hat{\theta}=1-\sqrt{\frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2+T\hat{\sigma}_\alpha^2}},$$

$$\hat{\sigma}_u^2=\frac{\sum_{i=1}^{N}\sum_{t=1}^{T}((y_{2it}-\bar{y}_{2i})-(y_{1it}-\bar{y}_{1i})\hat{\gamma}_{IVW})^2}{NT-N},$$

$$\hat{\sigma}_u^2 + T\hat{\sigma}_\alpha^2 = \frac{T\sum_{i=1}^{N}\left(\bar{y}_{2i} - \bar{y}_{1i}\hat{\gamma}_{IVB}\right)^2}{N}.$$

Применим теперь метод IV к уравнению $\tilde{y}_{2it} = \tilde{y}_{1it}\gamma + \tilde{u}_{it}$, используя в качестве инструментов $z_{it} - \bar{z}_i$ и $\bar{z}_i$ или $\tilde{z}_{it}$.

Более общим образом, пусть $\delta = [\gamma\beta]$ – вектор-строка размерности $g_1 + k_1$. На практике приходится применять метод IV (2SLS) трижды:

1. для получения $\hat{\delta}_{IVW}$;
2. для получения $\hat{\delta}_{IVB}$;

в результате этих двух шагов получают оценки $\hat{\sigma}_u^2$ и $\hat{\sigma}_u^2 + T\hat{\sigma}_\alpha^2$, которые используются для преобразования модели;

3. реализуя метод IV для преобразованной модели.

Оценивание можно выполнить в пакете STATA, используя команду *xtivreg* с опцией *re*. Если $E(Y_{1it}\alpha_i) \neq 0$, то используется опция *fe*.

Пример

Исследование зависимости заработной платы женщин от различных факторов. Для исследования были взяты данные (из National Longitudinal Survey, Youth Sample, США) по N=4134 молодым женщинам, имевшим в 1968 г. возраст от 14 до 26 лет. Наблюдения проводились с 1968 по 1988 гг. Однако данные неполные: по отдельным субъектам количество наблюдений изменялось от 1 до 12 (в среднем 4.6 наблюдения для одного субъекта).

Рассмотрим сначала модель с фиксированными эффектами

$$\ln w_{it} = \beta_1\, tenure + \beta_2\, age + \beta_3 age2 + \beta_4 notsmsa +$$
$$+ \beta_5\, union + \beta_6\, south + \mu + \alpha_i + u_{it}\,,$$

$i = 1,\ldots,N,\ t = 1,\ldots,T,$ где w_{it} – размер заработной платы,

tenure – продолжительность (стаж) работы на наблюдаемом рабочем месте,

age – возраст,

notsmsa – проживание вне столичных регионов,

union – принадлежность к профсоюзу,

south – проживание на юге страны.

Оценивание указанной модели с использованием “внутри”-оценки дает следующие результаты:

. xtreg ln_w tenure age age_2 not_smsa union south, fe i(idcode)
Fixed-effects (within) regression Number of obs =19007
Group variable (i): idcode Number of groups = 4134
R-sq: within = 0.1333 Obs per group: min = 1
between = 0.2375 avg = 4.6
overall = 0.2031 max = 12
F(6,14867) = 381.19 Prob > F = 0.0000
corr(α_i, Xb) = 0.2074

ln_wage	Coef.	Std. Err.	z	P>z
tenure	.0176205	.0008099	21.76	0.000
age	.0311984	.0033902	9.20	0.000
age_2	-.0003457	.0000543	-6.37	0.000
not_smsa	-.0972535	.0125377	-7.76	0.000
union	.0975672	.0069844	13.97	0.000
south	-.0620932	.0133270	-4.66	0.000
cons	1.091612	.0523126	20.87	0.000

sigma_α | .3910683
sigma_u | .25545969
rho | .70091004 (fraction of variance due to α_i)
F test that all α_i=0:
F(4133, 14867) = 8.31 Prob > F = 0.0000

Все оцененные коэффициенты имеют высокую статистическую значимость и ожидаемые знаки. Значительная часть изменчивости (70%) объясняется индивидуальными эффектами.

В то же время, если считать, например, что стаж работы на наблюдаемом рабочем месте зависит от принадлежности к профсоюзу и от региона проживания (юг – не юг) и что ошибки в уравнении для такой связи коррелированы с ошибками в уравнении для логарифма заработной платы, то тогда переменная $tenure_{it}$ в уравнении

$$\ln w_{it} = \beta_1 tenure_{it} + \beta_2 age_{it} + \beta_3 age2_{it} + \beta_4 notsmsa_{it} + \\ + \mu + \alpha_i + u_{it}$$

коррелирована с ошибкой u_{it}, и для получения состоятельных оценок коэффициентов этого уравнения приходится прибегать к методу инструментальных переменных. При сделанных предположениях в качестве инструментов для $tenure_{it}$ можно использовать $union$ и $south$. Полный список инструментов, обеспечивающий однозначную идентификацию коэффициентов μ и $\beta_1, \beta_2, \beta_{3,} \beta_4$, включает 5 переменных: $union$, $south$, age, $age2$, $notsmsa$. Использование этих переменных в качестве инструментов приводит к следующему результату:

```
. xtivreg ln_w age age_2 not_smsa (tenure=union south), fe i(idcode)
Fixed-effects (within) IV regression
Number of obs   =  19007
Number of groups   =  4134
Obs per group: min = 1    avg =4.6    max = 12
Wald chi2(4) = 147926.58   Prob > chi2  = 0.0000
corr(α_i, Xb)  = -0.6843
```

ln_wage	Coef.	Std. Err.	z	P>z
tenure	.2403531	.0373419	6.44	0.000
age	.0118437	.0090032	1.32	0.188
age_2	-.0012145	.0001968	-6.17	0.000
not_smsa	-.0167178	.0339236	-0.49	0.622
cons	1.678287	.1626657	10.32	0.000

sigma_ α | .70661941
sigma_ u | .63029359
rho | .55690561 (fraction of variance due to α_i)

F test that all α_i=0:
F(4133,14869) = 1.44 Prob > F = 0.0000

Instrumented: tenure
Instruments: age age_2 not_smsa union south

Оцененные коэффициенты и здесь имеют ожидаемые знаки. Однако на этот раз оказались статистически незначимыми оцененные коэффициенты при переменных *age* и *notsmsa* .

Применение “внутри”-оценки предпочтительно при интерпретации индивидуальных эффектов как фиксированных эффектов. Если рассматривать эти эффекты как случайные и некоррелированные с остальными объясняющими переменными, то предпочтительнее использовать инструментальное GLS-оценивание, как это было описано выше. Такой подход приводит к следующим результатам:

. xtivreg ln_w age age2 not_smsa black (tenure = union birth_yr south black) , re i(idcode)
G2SLS random-effects IV regression
Number of obs = 19007
Number of groups = 4134

R-sq: within = 0.0664 Obs per group: min = 1
between = 0.2098 avg = 4.6
overall = 0.1463 max = 12
Wald chi2(5) = 1446.37 Prob > chi2 = 0.0000
corr(α_i, X) = 0 (assumed)

ln_wage	Coef.	Std. Err.	z	P>z
tenure	.1391798	.0078756	17.67	0.000
age	.0279649	.0054182	5.16	0.000
age_2	-.0008357	.0000871	-9.60	0.000
not_smsa	-.2235103	.0111371	-20.07	0.000
black	-.2078613	.0125803	-16.52	0.000
cons	1.337684	.0844988	15.83	0.000

sigma_α | .36582493
sigma_u | .63031479
rho | .25197078 (fraction of variance due to α_i)

Instrumented: tenure
Instruments: age age2 not_smsa black union birth_yr south

3.10. Модели с индивидуально-специфическими переменными

3.10.1. Оценивание в RE- и FE-моделях

До сих пор в модели с фиксированными эффектами неоднородность субъектов исследования характеризовалась наличием ненаблюдаемых характеристик, влияние которых отражалось в модели посредством параметров α_i. Однако неоднородность субъектов может выражаться также в различных значениях для этих субъектов некоторых наблюдаемых характеристик, не изменяющихся для каждого субъекта в процессе наблюдений. Например, в исследованиях, касающихся зависимости размера заработной платы индивида от различных факторов, такими характеристиками могут быть пол, базовое образование и т.п. В связи с этим мы рассмотрим теперь модель

$$y_{it} = \mu + \alpha_i + \beta x_{it} + \gamma z_i + u_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T,$$

где z_i – переменная, специфическая только в отношении субъекта.

Если эта модель трактуется как RE-модель, в которой эффекты не коррелированы с x_{it} и z_i, то проблем с оцениванием коэффициентов β и γ не возникает: в этом случае BLUE являются GLS-оценки для β и γ.

Если же эта модель трактуется как FE-модель или если α_i случайны и $E(\alpha_i z_i) = 0$, но $E(\alpha_i x_{it}) \neq 0$, то тогда GLS-оценки (строящиеся, как в RE-модели) несостоятельны, и приходится искать другие оценки. Усредним обе части уравнения по t:

$$\bar{y}_i = \mu + \alpha_i + \beta \bar{x}_i + \gamma z_i + \bar{u}_i, \quad i = 1, \ldots, N.$$

Тогда

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}), \quad i = 1, \ldots, N.$$

В применении к последнему уравнению метод наименьших квадратов приводит опять к оценке с фиксированными эффектами $(\hat{\beta}_{CV})$. Но при таком подходе из исходного уравнения выметаются не только α_i, но и z_i. Однако если α_i фиксированы или $E(\alpha_i x_{it}) \neq 0$, но $E(\alpha_i z_i) = 0$, то тогда все же можно построить состоятельную оценку коэффициента γ.

Для этого заметим, что $\bar{y}_i - \beta \bar{x}_i = \mu + \gamma z_i + (\alpha_i + \bar{u}_i)$. Если считать значение β известным, то мы можем оценить эту модель, минимизируя $\sum_{i=1}^{N} (\alpha_i + \bar{u}_i)^2$, и получить *between*-оценки

$$\hat{\gamma} = \frac{\sum_{i=1}^{N} \left((\bar{y}_i - \bar{y}) - (\bar{x}_i - \bar{x})\beta \right)(z_i - \bar{z})}{\sum_{i=1}^{N} (z_i - \bar{z})^2} ,$$

$$\hat{\mu} = \bar{y} - \bar{x}\beta - \bar{z}\hat{\gamma} .$$

Подставляя $\hat{\beta}_{CV}$ вместо β в эти два выражения, получаем оценки $\tilde{\gamma}$ и $\tilde{\mu}$ для γ и μ; при этом

$$\operatorname*{plim}_{N\to\infty} \tilde{\gamma} = \gamma ,$$

т.е. $\tilde{\gamma}$ – состоятельная оценка.

[В пакете STATA используем *predict residfe, ue* после *xtreg…, fe* и затем *xtreg…, fe* Z_i *, be*]

П р и м е р (объяснение размеров заработной платы, продолжение)

При оценивании модели с фиксированными эффектами оказались выметенными переменные *scooling, black, hispanic*. Попробуем все же получить оценки коэффициентов при этих переменных. Использование процедуры, предусмотренной Stata8, приводит к следующему результату.

Between regression (regression on group means)

R-sq:
within = 0.0000
between = 0.2119
overall = 0.1264

F(3,541) = 48.47, Prob > F = 0.0000

	Coef.	Std. Err.	t	P>t
SCHOOL	.1015825	.0089372	11.37	0.000
BLACK	-.1442396	.0484007	-2.98	0.003
HISP	.0210173	.0435069	0.48	0.629

Полученные оценки практически совпадают с оценками для соответствующих коэффициентов в модели со случайными коэффициентами (при GLS-оценивании).

3.10.2. Модель Хаусмана–Тейлора

Рассмотрим модель

$$y_{it} = X_{it}\beta + Z_i\gamma + \alpha_i + u_{it}\,,\; i = 1,\ldots,N,\; t = 1,\ldots,T\,,$$

где X_{it} – строка k наблюдаемых переменных, изменяющихся и от субъекта к субъекту и во времени, а Z_i – строка g наблюдаемых переменных, инвариантных относительно времени на периоде наблюдений. Предполагается, что выполнены обычные предположения RE-модели (в частности, что все объясняющие переменные экзогенны в обычном смысле), за исключением того, что теперь ненаблюдаемые индивидуальные эффекты α_i могут быть коррелироваными с одними объясняющими переменными и некоррелироваными с другими объясняющими переменными.

Хаусман и Тейлор предложили в таком случае производить разбиение:

$$X_{it} = \left[X_{1it}\;\; X_{2it}\right],\;\; Z_i = \left[Z_{1i}\;\; Z_{2i}\right],$$

где k_1 переменных, входящих в состав X_{1it} , и g_1 переменных, входящих в состав Z_{1i} , <u>экзогенны</u> по отношению к α_i в том смысле, что

$$E\left(X_{1it}\alpha_i\right) = E\left(Z_{1i}\alpha_i\right) = 0\,,$$

а k_2 переменных, входящих в состав X_{2it} , и g_2 переменных, входящих в состав Z_{2i} , <u>эндогенны</u> по отношению к α_i в том смысле, что

$$E\left(X_{2it}\alpha_i\right) \neq 0\,,\; E\left(Z_{2i}\alpha_i\right) \neq 0\,.$$

При таком разбиении модель записывается в виде

$$y_{it} = X_{1it}\beta_1 + X_{2it}\beta_2 + Z_{1i}\gamma_1 + Z_{2i}\gamma_2 + \alpha_i + u_{it}\,,$$

и переход к модели, скорректированной на индивидуальные средние,

$$y_{it} - \bar{y}_i = \left(X_{1it} - \bar{X}_{1i}\right)\beta_1 + \left(X_{2it} - \bar{X}_{2i}\right)\beta_2 + u_{it},$$

приводит к возможно неэффективной, но состоятельной оценке $\hat{\beta}_{CV}$ для $\beta = \left(\beta_1^T, \beta_2^T\right)^T$. Однако при этом опять вместе с α_i выметаются переменные, инвариантные относительно времени.

Получив оценку $\hat{\beta}_{CV} = \left(\hat{\beta}_{1,CV}^T, \hat{\beta}_{2,CV}^T\right)^T$ для β, вычисляем для каждого субъекта остатки от оцененной “внутри”-регрессии:

$$d_{it} = \left(y_{it} - \bar{y}_i\right) - \left(X_{1it} - \bar{X}_{1i}\right)\hat{\beta}_{1,CV} - \left(X_{2it} - \bar{X}_{2i}\right)\hat{\beta}_{2,CV}$$

и получаем состоятельную оценку для дисперсии случайных ошибок $D(u_{it}) = \sigma_u^2$:

$$\hat{\sigma}_u^2 = \frac{RSS_{CV}}{NT - \left(k_1 + k_2\right)}.$$

Далее, по аналогии с уже делавшимся ранее, замечаем, что

$$\bar{y}_i - \bar{X}_{1i}\beta_1 - \bar{X}_{2i}\beta_2 = Z_{1i}\gamma_1 + Z_{2i}\gamma_2 + \left(\alpha_i + u_{it}\right).$$

Только на этот раз в правой части $E\left(Z_{2i}\alpha_i\right) \neq 0$, так что OLS-оценки для γ_1 и γ_2, получаемые по этой модели (*between*-оценки) смещены и несостоятельны. Поэтому здесь для получения состоятельных оценок для γ_1 и γ_2 применяем метод инструментальных переменных (2SLS), используя инструменты $\left[Z_{1i}\ X_{1it}\right]$. При этом количество экзогенных переменных в X_{1it} (k_1) должно быть не меньше числа эндогенных переменных в Z_{2i} (g_2). Полученные оценки для γ_1 и γ_2 обозначим $\hat{\gamma}_{1,IV}$ и $\hat{\gamma}_{2,IV}$.

Используя теперь все четыре полученные оценки, образуем “остатки”

$$e_{it} = y_{it} - X_{1it}\hat{\beta}_{1,CV} - X_{2it}\hat{\beta}_{2,CV} - Z_{1i}\hat{\gamma}_{1,IV} - Z_{2i}\hat{\gamma}_{2,IV}$$

и на их основе определим статистику

$$S^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2 ,$$

которая является состоятельной оценкой для суммы $\sigma_u^2 + T\sigma_\alpha^2$. Тогда состоятельная оценка для σ_α^2 получается как

$$\hat{\sigma}_\alpha^2 = \frac{S^2 - \hat{\sigma}_u^2}{T} .$$

Следующим шагом является выполнение стандартного GLS-преобразования всех переменных, используемого в RE-модели:

$$y_{it}^* = y_{it} - \theta\,\bar{y}_i \quad \text{и т.п.}$$

Для реализации этого преобразования в качестве оценки параметра θ берется $\hat{\theta} = 1 - \sqrt{\hat{\Psi}}$, где

$$\hat{\Psi} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + T\hat{\sigma}_\alpha^2} ,$$

а $\hat{\sigma}_u^2$ и $\hat{\sigma}_\alpha^2$ – полученные выше состоятельные оценки для σ_u^2 и σ_α^2. Это приводит к преобразованному уравнению

$$y_{it}^* = X_{1it}^*\beta_1 + X_{2it}^*\beta_2 + Z_{1i}^*\gamma_1 + Z_{2i}^*\gamma_2 + v_{it}^* .$$

К этому уравнению применяем метод IV (2SLS), используя инструменты $\left[X_{1it} - \overline{X}_{1i}, X_{2it} - \overline{X}_{2i}, \overline{X}_{1i}, Z_{1i}\right]$, и получаем в результате для вектора $\delta = \left[\beta^T, \gamma^T\right]^T$ инструментальную оценку Хаусмана–Тейлора $\hat{\delta}_{HT} = \left[\hat{\beta}_{HT}^T, \hat{\gamma}_{HT}^T\right]^T$.

Замечания

1. В процедуре Хаусмана–Тейлора инструменты берутся внутри самой модели.
2. X_{1it} используется как инструмент дважды: как среднее и как отклонение от среднего.
3. Если $k_1 < g_2$, то параметр γ не идентифицируем. В этом случае $\hat{\beta}_{HT} = \hat{\beta}_{CV}$ и $\hat{\gamma}_{HT}$ не существует.

4. Если $k_1 = g_2$, то $\hat{\beta}_{HT} = \hat{\beta}_{CV}$ и $\hat{\gamma}_{HT} = \hat{\gamma}_{IV} = \left(\hat{\gamma}_{1,IV}^T, \hat{\gamma}_{2,IV}^T\right)^T$. (Случай точной идентифицируемости.)
5. Если $k_1 > g_2$, то уравнение сверхидентифицируемо. В этом случае матрица $Cov\left(\hat{\beta}_{CV}\right) - Cov\left(\hat{\beta}_{HT}\right)$ положительно определена и оценка Хаусмана–Тейлора более эффективна, чем "внутри"-оценка.

Влияние метода Хаусмана–Тейлора на прикладные исследования относительно мало из-за трудности нахождения экзогенных переменных X_1, которые можно было бы уверенно рассматривать как не коррелированные с α_i (так что $E\left(X_{1it}\alpha_i\right) = 0$).

3.11. Динамические модели

Рассмотрим динамическую модель.

$y_{it} = \gamma y_{i,t-1} + \alpha_i + u_{it}$, $i = 1, \ldots, N$, $t = 1, \ldots, T$,

в которой $|\gamma| < 1$, $u_{it} \sim i.i.d.N\left(0, \sigma_u^2\right)$ – инновации (так что $E\left(u_{it} y_{i,t-k}\right) = 0$ для $k > 0$). Будем предполагать, что значения y_{it} наблюдаются для $t = 0, 1, \ldots, T$. "Внутри"-оценка для γ имеет вид:

$$\hat{\gamma}_{CV} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \bar{y}_i\right)\left(y_{i,t-1} - \bar{y}_{i,-1}\right)}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t-1} - \bar{y}_{i,-1}\right)^2},$$

где

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{y}_{i,-1} = \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}.$$

В соответствии с определением модели,

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T}\left(\gamma y_{i,t-1} + \alpha_i + u_{it}\right) = \gamma \bar{y}_{i,-1} + \alpha_i + \bar{u}_i ,$$

так что

$$y_{it} - \bar{y}_i = \gamma\left(y_{i,t-1} - \bar{y}_{i,-1}\right) + \left(u_{it} - \bar{u}_i\right)$$

и

$$\hat{\gamma}_{CV} = \gamma + \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(u_{it} - \bar{u}_i\right)\left(y_{i,t-1} - \bar{y}_{i,-1}\right)/NT}{\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t-1} - \bar{y}_{i,-1}\right)^2/NT} .$$

Рассмотрим предел по вероятности числителя дроби в правой части последнего выражения:

$$\underset{N\to\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(u_{it} - \bar{u}_i\right)\left(y_{i,t-1} - \bar{y}_{i,-1}\right) =$$

$$= \underset{N\to\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\left(\sum_{t=1}^{T}u_{it}y_{i,t-1} - \bar{u}_i\sum_{t=1}^{T}y_{i,t-1} - \bar{y}_{i,-1}\sum_{t=1}^{T}u_{it} + T\,\bar{u}_i\bar{y}_{i,-1}\right) =$$

$$= \underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\left(-\bar{u}_i\bar{y}_{i,-1}\right).$$

Заметим, что

$$y_{i1} = \gamma y_{i0} + \alpha_i + u_{i1} ,$$

$$y_{i2} = \gamma y_{i1} + \alpha_i + u_{i2} = \gamma^2 y_{i0} + (\gamma+1)\alpha_i + \gamma u_{i1} + u_{i2} ,$$

$$\ldots$$

$$y_{it} = \gamma^t y_{i0} + \left(1-\gamma^t\right)\alpha_i/(1-\gamma) + \gamma^{t-1} u_{i1} + \cdots + \gamma u_{i,t-1} + u_{it} .$$

Отсюда, в частности, получаем, что при $T = 2$ указанный предел по вероятности равен

$$-\underset{N\to\infty}{p\lim}\frac{1}{4N}\sum_{i=1}^{N}\left((\gamma+1)y_{i0} + \alpha_i + u_{i1}\right)\left(u_{i1} + u_{i2}\right) =$$

$$-\underset{N\to\infty}{p\lim}\frac{1}{4N}\sum_{i=1}^{N}u_{i1}^2 = -\frac{\sigma_u^2}{4} \neq 0 ,$$

так что оценка $\hat{\gamma}_{CV}$ несостоятельна.

При произвольном T получаем:

$$\underset{N\to\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(u_{it}-\bar{u}_i\right)\left(y_{i,t-1}-\bar{y}_{i,-1}\right)=$$

$$=\underset{N\to\infty}{p\lim}\frac{1}{N}\sum_{i=1}^{N}\left(-\bar{u}_i\bar{y}_{i,-1}\right)=-\sigma_u^2\left[\frac{T-1-T\gamma+\gamma^T}{T^2(1-\gamma)^2}\right],$$

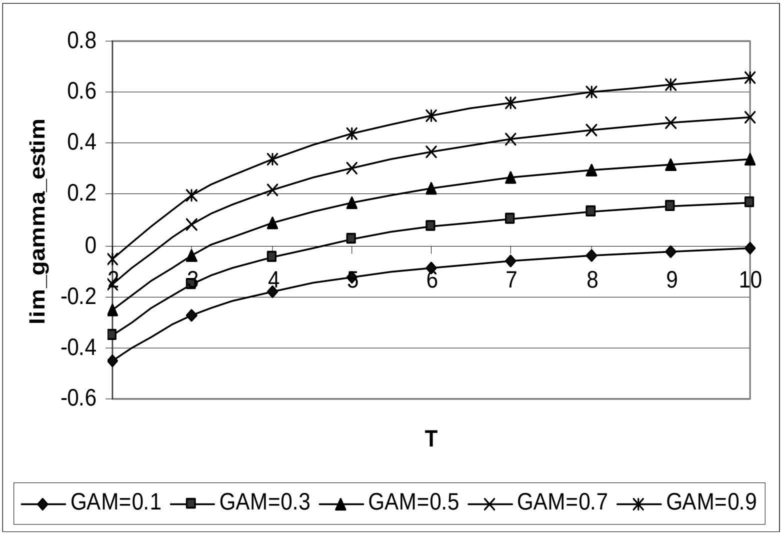
$$\underset{N\to\infty}{p\lim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{i,t-1}-\bar{y}_{i,-1}\right)^2=\frac{\sigma_u^2}{1-\gamma^2}\left[1-\frac{1}{T}-2\gamma\frac{T-1-T\gamma+\gamma^T}{T^2(1-\gamma)^2}\right].$$

Отсюда вытекает, что если не только $N\to\infty$, но и $T\to\infty$, то первый предел по вероятности стремится к нулю, а второй – к

$$\underset{N\to\infty,T\to\infty}{p\lim}\ \hat{\gamma}_{CV}=\frac{\sigma_u^2}{1-\gamma^2}\neq 0,$$

так что $\underset{N\to\infty}{p\lim}\hat{\gamma}_{CV}=\gamma$. Если же значение T фиксировано, то тогда первый предел не равен нулю и $\underset{N\to\infty}{p\lim}\hat{\gamma}_{CV}\neq\gamma$, так что оценка $\hat{\gamma}_{CV}$ несостоятельна.

Асимптотическое смещение вызывается “внутри”-преобразованием, исключающим индивидуальные эффекты α_i из каждого наблюдения, что порождает корреляцию между остатками $\left(u_{it}-\bar{u}_i\right)$ в преобразованной модели и “объясняющей переменной” $\left(y_{i,t-1}-\bar{y}_{i,-1}\right)$. Когда T велико, эта корреляция близка к нулю. Для малых значений T смещение отрицательно, если $\gamma>0$. Например, для T асимптотическое смещение равно $-(1-\gamma)/2$. Следующий график иллюстрирует, как асимптотическое смещение оценки $\hat{\gamma}_{CV}$ изменяется с ростом T.



”Внутри”-оценка остается несостоятельной при малых T и когда в правую часть уравнения модели добавляются экзогенные объясняющие переменные.

Для преодоления этой проблемы можно воспользоваться другим преобразованием, “выметающим” α_i : вместо вычитания средних по времени перейти к первым разностям временных рядов для каждого субъекта. При этом получаем:

$$y_{i,t} - y_{i,t-1} = \gamma\left(y_{i,t-1} - y_{i,t-2}\right) + \left(u_{it} - u_{i,t-1}\right),$$

или

$$\Delta y_{i,t} = \gamma \Delta y_{i,t-1} + \Delta u_{it},$$

где обозначено

$$\Delta y_{i,t} = y_{i,t} - y_{i,t-1},\ \Delta u_{it} = u_{it} - u_{i,t-1}.$$

Здесь

$$Cov\left(\Delta y_{i,t-1},\, \Delta u_{it}\right) = Cov\left(y_{i,t-1} - y_{i,t-2},\, u_{it} - u_{i,t-1}\right) =$$
$$= -Cov\left(y_{i,t-1}, u_{i,t-1}\right) \neq 0\ .$$

Поэтому OLS-оценка для γ в преобразованном (“продифференцированном”) уравнении оказывается несостоятельной даже если $T \to \infty$. Однако к преобразованному уравнению можно применить ***метод инструментальных переменных*** (***IV*** – *instrumental variables*). Для этого достаточно заметить, что если взять переменную $y_{i,t-2}$, то для нее

$$Cov(y_{i,t-2}, \Delta u_{it}) = Cov(y_{i,t-2}, u_{it} - u_{i,t-1}) = 0,$$

$$Cov(y_{i,t-2}, \Delta y_{i,t-1}) = Cov(y_{i,t-2}, y_{i,t-1} - y_{i,t-2}) \neq 0,$$

а это означает, что эта переменная может использоваться в качестве инструмента, и это приводит к оценке

$$\hat{\gamma}_{IV} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t} - y_{i,t-1})y_{i,t-2}}{\sum_{i=1}^{N}\sum_{t=2}^{T}(y_{i,t-1} - y_{i,t-2})y_{i,t-2}}.$$

Необходимое условие состоятельности этой оценки:

$$\underset{N\to\infty}{p\lim}\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}(u_{it} - u_{i,t-1})y_{i,t-2} = 0$$

при $T \to \infty$ или/и $N \to \infty$.

В качестве инструмента вместо $y_{i,t-2}$ можно использовать, например, разность $y_{i,t-2} - y_{i,t-3}$, что приводит к другой оценке

$$\hat{\gamma}_{IV} = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t} - y_{i,t-1})(y_{i,t-2} - y_{i,t-3})}{\sum_{i=1}^{N}\sum_{t=3}^{T}(y_{i,t-1} - y_{i,t-2})(y_{i,t-2} - y_{i,t-3})},$$

для состоятельности которой необходимо, чтобы

$$\underset{N\to\infty}{p\lim}\frac{1}{N(T-2)}\sum_{i=1}^{N}\sum_{t=3}^{T}(u_{it} - u_{i,t-1})(y_{i,t-2} - y_{i,t-3}) = 0.$$

Состоятельность обеих оценок гарантируется отсутствием автокоррелированности $u_{i,t}$.

Заметим теперь, что

$$\underset{N\to\infty}{p\lim}\frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=2}^{T}\left(u_{it}-u_{i,t-1}\right)y_{i,t-2}=E\left[\left(u_{it}-u_{i,t-1}\right)y_{i,t-2}\right]=0,$$

$$\underset{N\to\infty}{p\lim}\frac{1}{N(T-2)}\sum_{i=1}^{N}\sum_{t=3}^{T}\left(u_{it}-u_{i,t-1}\right)\left(y_{i,t-2}-y_{i,t-3}\right)=$$

$$=E\left[\left(u_{it}-u_{i,t-1}\right)\left(y_{i,t-2}-y_{i,t-3}\right)\right]=0\,.$$

Это условия на моменты совместных распределений пар случайных величин $\left(u_{it}-u_{i,t-1}\right)$, $y_{i,t-2}$ и $\left(u_{it}-u_{i,t-1}\right)$, $\left(y_{i,t-2}-y_{i,t-3}\right)$. Если оба эти условия (***условия ортогональности***) выполняются, то использование сразу двух инструментов приводит к повышению эффективности оценок (используется большее количество информации). Заметим, что можно найти и другие подходящие инструменты. Например, каждая из переменных $y_{i,t-2-j}$, $j=0,1,\ldots$ удовлетворяет условиям

$$E\left[\left(u_{it}-u_{i,t-1}\right)y_{i,t-2-j}\right]=0,$$

$$E\left[\left(y_{i,t-1}-y_{i,t-2}\right)y_{i,t-2-j}\right]\neq 0,$$

так что и эти переменные годятся в качестве инструментов.

Arellano, Bond (1991) предложили расширить список инструментов, поступая следующим образом. Пусть, например, T = 4. Соотношение

$$E\left[\left(u_{i2}-u_{i1}\right)y_{i0}\right]=0$$

рассматривается как моментное условие для $t=2$. Для $t=3$ имеем

$$E\left[\left(u_{i3}-u_{i2}\right)y_{i1}\right]=0\,,$$

но при этом справедливо и соотношение

$$E\left[\left(u_{i3}-u_{i2}\right)y_{i0}\right]=0\,.$$

Для $t=4$ имеется три аналогичных моментных условия:

$E[(u_{i4}-u_{i3})y_{i0}]=0$,
$E[(u_{i4}-u_{i3})y_{i1}]=0$,
$E[(u_{i4}-u_{i3})y_{i2}]=0$.

Всю эту совокупность моментных условий можно использовать рамках ***обобщенного метода моментов*** (***GMM*** – *generalized method of moments*).

Для произвольного T определим $(T-1)\times 1$-вектор

$$\Delta u_i = \begin{pmatrix} u_{i2}-u_{i1} \\ \vdots \\ u_{i,T}-u_{i,T-1} \end{pmatrix}$$

и $(T-1)\times(T(T-1)/2)$-матрицу

$$Z_i = \begin{pmatrix} [y_{i0}] & 0 & \cdots & 0 \\ 0 & [y_{i0}, y_{i1}] & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & [y_{i0},\ldots,y_{i,T-2}] \end{pmatrix};$$

каждая строка матрицы Z_i содержит инструменты, подходящие для соответствующего момента времени. В этих обозначениях, указанная совокупность $T(T-1)/2$ моментных условий записывается в виде:

$E[Z_i^T \Delta u_i]=0$.

Если определить еще

$$\Delta y_i = \begin{pmatrix} y_{i2}-y_{i1} \\ \vdots \\ y_{i,T}-y_{i,T-1} \end{pmatrix}, \quad \Delta y_{i,-1} = \begin{pmatrix} y_{i1}-y_{i0} \\ \vdots \\ y_{i,T-1}-y_{i,T-2} \end{pmatrix},$$

то последнее соотношение записывается также в виде:

$E[Z_i^T(\Delta y_i - \gamma \Delta y_{i,-1})]=0$.

В отличие от метода наименьших квадратов здесь мы имеем количество моментных условий больше необходимого для определения с их помощью значения γ, так что использование разных условий приводит к различным оценкам. Следовательно, нет возможности получения значения γ, при котором выполняются все указанные моментные условия. Вместо этого приходится ограничиваться требованием в каком-то смысле "наилучшего" приближения ко всем моментным условиям сразу. Чтобы использовать всю совокупность моментных условий, в GMM минимизируется квадратичная форма от выборочных аналогов моментных условий

$$Q(\gamma)=\left[\frac{1}{N}\sum_{i=1}^{N} Z_i^T\left(\Delta y_i-\gamma\Delta y_{i,-1}\right)\right]^T W_N\left[\frac{1}{N}\sum_{i=1}^{N} Z_i^T\left(\Delta y_i-\gamma\Delta y_{i,-1}\right)\right],$$

где W_N – симметричная положительно определенная весовая матрица. Искомый минимум достигается при значении, равном

$$\hat{\gamma}_{GMM}=\left(\left(\sum_{i=1}^{N}\Delta y_{i,-1}^T Z_i\right)W_N\left(\sum_{i=1}^{N} Z_i^T\ \Delta y_{i,-1}\right)\right)^{-1}\times\left(\sum_{i=1}^{N}\Delta y_{i,-1}^T Z_i\right)W_N\left(\sum_{i=1}^{N} Z_i^T\ \Delta y_i\right).$$

Это и есть GMM-оценка параметра γ. Свойства этой оценки зависят от выбора взвешивающей матрицы W_N. Хотя она состоятельна при положительной определенности матрицы W_N, в частности, для единичной матрицы $W_N=I_N$, желательно выбирать матрицу W_N таким образом, чтобы GMM-оценка была по возможности наиболее эффективна – о такой матрице говорят как об ***оптимальной взвешивающей матрице***. Такая матрица должна удовлетворять условию:

$$p\lim_{N\to\infty} W_N=\left(Cov\left(Z_i^T\Delta u_i\right)\right)^{-1}=\left[E\left(Z_i^T\Delta u_i\left(\Delta u_i\right)^T Z_i\right)\right]^{-1}.$$

И если на ковариационную матрицу вектора ошибок наблюдений для i-го субъекта не накладывается никаких ограничений, то можно поступить следующим образом.

Сначала полагаем $W_N=I_N$ и производим GMM-оценивание коэффициента γ в модели $\Delta y_{i,t}=\gamma\Delta y_{i,t-1}+\Delta u_{it}$ с такой взвешивающей матрицей, получая предварительную оценку $\hat{\gamma}^{(1)}$ для γ. При этом получаем остатки

$$\Delta\hat{u}_{it} = \Delta y_{i,t} - \hat{\gamma}^{(1)}\,\Delta y_{i,t-1}$$

и составляем из них векторы

$$\Delta\hat{u}_i = \begin{pmatrix} \Delta\hat{u}_{i2} \\ \vdots \\ \Delta\hat{u}_{i,T} \end{pmatrix}.$$

Искомая матрица определяется после этого соотношением

$$W_N^{opt} = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i^T \Delta\hat{u}_i \left(\Delta\hat{u}_i\right)^T Z_i \right)^{-1}.$$

Если $u_{it} \sim\ i.i.d.$, то положение значительно упрощается. В этом случае

$$E\left(\Delta u_i \left(\Delta u_i\right)^T\right) = \sigma_u^2\, G = \sigma_u^2 \begin{pmatrix} 2 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 \\ 0 & -1 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & 0 & \dots & -1 & 2 \end{pmatrix}$$

и поэтому не требуется предварительного оценивания γ. Оптимальная матрица определяется соотношением

$$W_N^{opt} = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i^T G\, Z_i \right)^{-1}$$

и GMM-оценивание производится за один шаг.

В общем случае GMM-оценка $\hat{\gamma}_{GMM}$ имеет асимптотически нормальное распределение с ковариационной матрицей

$$\underset{N\to\infty}{p\lim}\left(\left(\frac{1}{N}\sum_{i=1}^{N}\Delta y_{i,-1}^{T}Z_i\right)\left(\frac{1}{N}\sum_{i=1}^{N}Z_i^T\,\Delta u_i\left(\Delta u_i\right)^T Z_i\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}Z_i^T\,\Delta y_{i,-1}\right)\right)^{-1}.$$

Если $u_{it}\sim\,i.i.d.$, то средняя составляющая редуцируется к

$$\sigma_u^2\,W_N^{opt}=\sigma_u^2\left(\frac{1}{N}\sum_{i=1}^{N}Z_i^T\,G\,Z_i\right)^{-1}.$$

Рассмотрим теперь динамическую модель с экзогенными переменными

$$y_{it}=\gamma\,y_{i,t-1}+x_{it}^T\beta+\alpha_i+u_{it}\,,\quad i=1,\ldots,N,\,t=1,\ldots,T\,.$$

Здесь дифференцирование приводит к модели

$$\Delta y_{i,t}=\gamma\Delta y_{i,t-1}+\Delta x_{it}^T\beta+\Delta u_{it}\,.$$

Если предполагать, что все p объясняющих переменных, входящих в состав вектора x_{it}, <u>строго экзогенны</u>, в том смысле, что они не коррелированы с каждым u_{is}, то тогда

$$E\left(x_{it}\Delta u_{is}\right)=0\quad\text{для всех}\quad t,s\,,$$

так что в указанные ранее списки инструментов для каждого момента (периода) времени можно добавить $x_{i1},\ldots,x_{iT}$. Тогда для момента t список инструментов принимает вид: $[\ y_{i0},y_{i1},\ldots,y_{i,t-2},x_{i1},\ldots,x_{iT}\ ]$. Такой список может быть весьма длинным, если p и T не очень малы. В то же время при строгой экзогенности x_{it} имеем также:

$$E\left(\Delta x_{it}\Delta u_{it}\right)=0\ \text{ для каждого }\ t\,,$$

так что Δx_{it} могут выступать в качестве инструментов для самих себя. При таком подходе список инструментов для момента t имеет вид: $\left[y_{i0}, y_{i1}, \ldots, y_{i,t-2}, \Delta x_{i1}, \ldots, \Delta x_{it}\right]$. Этот список существенно короче, если панель достаточно длинна.

Предположим теперь, что переменные в x_{it} не являются строго экзогенными, но являются <u>предетерминированными</u>, в том смысле, что

$$E(x_{it}u_{is}) = 0 \quad \text{для всех} \quad s > t .$$

В этом случае уже не все $x_{i1}, \ldots, x_{iT}$ годятся в качестве инструментов для продифференцированного уравнения в момент t, а только $x_{i1}, \ldots, x_{i,t-1}$, и, соответственно, накладываются моментные условия

$$E(x_{i,t-j}\Delta u_{it}) = 0 \text{ для } \quad j = 1, \ldots, t-1 .$$

Разумеется, если в состав x_{it} входят как строго эндогенные, так и предетерминированные переменные, то списки инструментов соответствующим образом корректируются.

З а м е ч а н и е

Указанная выше “оптимальная” взвешивающая матрица является оптимальной в отношении выбранного множества инструментов. В то же время возникает вопрос об “оптимальном” выборе самих инструментов.

Привлечение большего количества инструментов подразумевает получение более эффективных оценок. Однако здесь возникают две опасности:

- некоторые из переменных, привлеченных в качестве инструментов, в действительности могут быть коррелированными с ошибками; для предотвращения таких ситуаций необходимо проверять гипотезу о выполнении соответствующих условий ортогональности;
- оценки коэффициентов могут иметь значительное смещение вследствие оценивания взвешивающей матрицы W_N .

Проверка гипотез о правильности спецификации динамической модели

Пример

Вернемся к рассмотренным ранее данным об объемах инвестиций y и прибыли x трех предприятий ($N=3$) за десятилетний период ($T=10$) и рассмотрим на этот раз динамическую модель первого порядка

$$y_{it}=\mu+\alpha_i+\gamma y_{i,t-1}+\beta_1 x_{it}+\beta_2 x_{i,t-1}+u_{it}, \quad i=1,2,3, \quad t=2,\ldots,10 .$$

Дифференцирование приводит к модели для разностей:

$$\Delta y_{it}=\gamma\Delta y_{i,t-1}+\beta_1\Delta x_{it}+\beta_2\Delta x_{i,t-1}+\Delta u_{it}, \quad i=1,2,3, \quad t=3,\ldots,10 .$$

Если следовать описанию программы ***xtabond*** в пакете Stata8, то в этой программе будут использованы в качестве инструментов:

t	Инструменты	Кол-во
3	y_{i1}	1
4	y_{i1}, y_{i2}	2
5	y_{i1}, y_{i2}, y_{i3}	3
6	$y_{i1}, y_{i2}, y_{i3}, y_{i4}$	4
7	$y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}$	5
8	$y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}, y_{i6}$	6
9	$y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}, y_{i6}, y_{i7}$	7
10	$y_{i1}, y_{i2}, y_{i3}, y_{i4}, y_{i5}, y_{i6}, y_{i7}, y_{i8}$	8
Всего:		36

а также $\Delta x_{i3}+\cdots+\Delta x_{i,10}$ и $\Delta x_{i2}+\cdots+\Delta x_{i9}$. Это дает всего $36+2=38$ моментных условий. Поскольку модель в разностях содержит 3 неизвестных коэффициента, для оценивания этих коэффициентов достаточно использовать только 3 моментных условия, а остальные $38-3=35$ условий – избыточные. Их можно было бы не использовать вовсе, но это снизило бы эффективность получаемых оценок.

Вместе с тем наличие избыточных условий позволяет проверять адекватность сделанных в отношении модели предположений. Точнее говоря, возникает возможность проверки гипотезы H_0 о том, что избыточные условия (выведенные на основании исходных предположений о рассматриваемой модели) действительно выполняются. Для проверки этой гипотезы используется ***статистика Саргана*** (Sargan):

$$S = N\,Q\left(\hat{\theta}_{GMM}\right),$$

где $\hat{\theta}_{GMM}$ – GMM-оценка вектора коэффициентов θ модели (в нашем примере $\theta = (\gamma, \beta_1, \beta_2)^T$), а $Q\left(\hat{\theta}_{GMM}\right)$ – значение при $\theta = \hat{\theta}_{GMM}$ минимизируемой в методе GMM квадратичной формы. Если гипотеза H_0 справедлива, то статистика Саргана имеет асимптотическое (при $N \to \infty$) распределение хи-квадрат с числом степеней свободы, равным количеству избыточных моментных условий (в нашем примере оно равно 35).

Приведем теперь результаты применения программы ***xtabond*** при использовании взвешивающей матрицы

$$W_N^{opt} = \left(\frac{1}{N} \sum_{i=1}^{N} Z_i^T G\, Z_i \right)^{-1}$$

(так что оценивание производится за один шаг).

. xtabond y x l1(x), lags(1)

Arellano-Bond dynamic panel-data estimation

Number of obs = 24

Number of groups = 3

Obs per group: min = 8

One-step results

	Coef.	Std. Err.	z	P>z
Δy_{-1}	.0830295	.2100682	0.40	0.693
Δx	1.132349	.0606134	18.68	0.000
Δx_{-1}	-.0423772	.2375232	-0.18	0.858
cons	.1032841	.1361505	0.76	0.448

Sargan test of over-identifying restrictions:
chi2(35) = 21.81 Prob > chi2 = 0.9600

Результаты применения критерия Саргана говорят в пользу гипотезы о выполнении избыточных предположений. В то же время коэффициенты при запаздывающей разности значений объясняемой переменной и запаздывающей разности объясняющей переменной статистически незначимы, что возвращает нас к статической модели регрессии.

В программе ***xtabond*** используется еще один критерий проверки адекватности модели. Он основан на следующем обстоятельстве. Если ошибки $u_{i1},\ldots,u_{iT}$ взаимно независимы, то

- “соседние” разности $\Delta u_{it},\Delta u_{i,t-1}$ коррелированы, т.к. $Corr\left(\Delta u_{it},\Delta u_{i,t-1}\right)=Corr\left(u_{it}-u_{i,t-1},u_{i,t-1}-u_{i,t-2}\right)=-\sigma_u^2$;
- отстоящие на большее количество периодов времени разности Δu_{it}, $\Delta u_{i,t-s}$, $s=2,3,\ldots$ напротив, не являются коррелированными.

Соответственно, с целью дополнительной проверки адекватности оцененной модели проверяется “наличие автокоррелированности первого порядка” и “отсутствие автокоррелированности второго порядка”. Приведем результаты такой проверки для только что оцененной модели:

Arellano-Bond test that average autocovariance in residuals of order 1 is 0: H0: no autocorrelation z = -2.56, Pr > z = 0.0106.

Arellano-Bond test that average autocovariance in residuals of order 2 is 0: H0: no autocorrelation z = 0.77, Pr > z = 0.4427.

Полученные результаты подтверждают правильность сделанных предположений об ошибках

Пример

Здесь мы обратимся к исследованию, предпринятому в работе Konigs, Roodhooft [*De Economist*, 145 (1997)] и касающемуся эластичности спроса на труд со стороны предприятий Бельгии. Данные относились к более, чем 3000 крупных фирм в период с 1986 по 1994 г.

Статическое уравнение спроса на труд бралось в форме

$$\ln L_{it} = \beta_1 + \beta_2 \ln w_{it} + \beta_3 \ln r_{it} + \beta_4 \ln Y_{it} + \beta_5 \ln w_{jt} + u_{it} \,,$$

где w_{it} – стоимость единицы труда,

r_{it} – стоимость единицы капитала,

Y_{it} – объем выпуска,

w_{it} – средняя реальная заработная плата по отрасли.

Динамические модели брались в разных формах, из которых простейшей была модель, в правую часть которой включалось запаздывающее на один период значение объясняемой переменной. При этом r_{it} заменялась акционерным капиталом K_{it}, а в качестве Y_{it} выступала условно-чистая продукция. Соответствующая модель имела вид

$$\ln L_{it} = \beta_1 + \beta_2 \ln w_{it} + \beta_3 \ln K_{it} + \beta_4 \ln Y_{it} + \beta_5 \ln w_{jt} + \ln L_{i,t-1} + \alpha_i + \varepsilon_{it} \,.$$

Здесь α_i отражает наличие ненаблюдаемой переменной, специфичной в отношении фирм и не изменяющейся во времени.

Переходя к первым разностям, выметаем α_i, но полученное уравнение нельзя состоятельно оценить OLS из-за наличия корреляции между $\Delta \ln L_{i,t-1}$ и $\Delta\varepsilon_{i,t}$. Кроме того, может существовать коррелированность между $\Delta\varepsilon_{i,t}$ и $\Delta \ln w_{i,t}$ вследствие наличия договоров нанимателей с профсоюзами по вопросу о занятости и оплате труда, так что помимо $\Delta \ln L_{i,t-1}$ надо инструментировать и $\Delta \ln w_{i,t}$. В качестве инструментов можно использовать запаздывающие разности $\Delta \ln w_{i,t-2}$, $\Delta \ln w_{i,t-3}$ и т.д. Приводимые ниже результаты оценивания относятся к модели, в правую часть которой дополнительно включались временные и региональные дамми-переменные.

Переменные	Оценка коэффициента	Стандартная ошибка
$\ln L_{i,t-1}$	0.60	0.045
$\ln Y_{it}$	0.008	0.005
$\ln w_{i,t}$	-0.66	0.19
$\ln w_{j,t}$	0.054	0.033
$\ln K_{it}$	0.078	0.006

Значимость коэффициента при $\ln L_{i,t-1}$ говорит в пользу динамической модели. Значение -0.66 характеризует “краткосрочную” эластичность спроса по заработной плате, а значение $-0.60/(1-0.60)=-1.6$ характеризует “долговременную” эластичность спроса по заработной плате.

Заметим, что если производить оценивание “долговременной” эластичности в рамках статической модели, интерпретируя такую модель как модель долговременной связи между переменными, то тогда долговременная эластичность оценивается величиной -1.78. Такого рассогласования оценок можно было бы избежать в рамках модели коррекции ошибок.

Впрочем, к самим полученным результатам можно отнестись критически хотя бы по следующим обстоятельствам. В процессе инструментирования было использовано 29 “лишних” инструментов. Соответственно, имеется возможность проверить гипотезу о выполнении соответствующих 29 условий ортогональности с помощью статистики Саргана $S = N\,Q\left(\hat{\theta}_{GMM}\right)$, где $\hat{\theta}_{GMM}$ – оценка вектора кэффициентов модели обобщенным методом моментов. Статистика Саргана принимает здесь значение 51.66, что соответствует Р-значению, рассчитанному по распределению хи-квадрат с 29 степенями свободы, равному 0.006. Следовательно, гипотеза о выполнении дополнительных моментных условий отвергается.

3.12. Модели бинарного выбора

Мы уже рассматривали модели такого рода в Главе 1, но только для данных, относящихся к одному-единственному моменту (периоду) времени (*cross-section* data – данные в сечениях). Если исследование затрагивало n субъектов, то мы отмечали факт наличия или отсутствия некоторого интересующего нас признака у i-го субъекта при помощи индикаторной (бинарной) переменной y, которая принимает в i-м наблюдении значение y_i. При этом $y_i = 1$ при наличии рассматриваемого признака у i-го субъекта и $y_i = 0$ – при отсутствии рассматриваемого признака у i-го субъекта.

Если пытаться объяснить наличие или отсутствие рассматриваемого признака значениями (точнее, сочетанием значений) некоторых факторов (***объясняющих переменных***), то, следуя идеологии классической линейной модели, мы могли бы расмотреть модель наблюдений

$$y_i = \theta_1 x_{i1} + \cdots + \theta_p x_{ip} + \varepsilon_i\,,\quad i = 1, \ldots, n\,,$$

в которой $x_{i1}, \ldots, x_{ip}$ – значения p объясняющих переменных в i-м наблюдении, $\theta_1, \ldots, \theta_p$ – неизвестные параметры, а $\varepsilon_1, \ldots, \varepsilon_n$ – случайные ошибки, отражающие влияние на наличие или отсутствие рассматриваемого признака у i-го субъекта каких-то неучтенных дополнительных факторов. При обычных предположениях регрессионного анализа в такой модели получаем:

$$P\{y_i = 1 | \, x_i\} = \theta_1 x_{i1} + \cdots + \theta_p x_{ip}\,,$$

т.е. должно выполняться соотношение

$$0 \le \theta_1 x_{i1} + \dots + \theta_p x_{ip} \le 1 .$$

Это первая из трудностей, с которыми мы сталкиваемся при обращении к таким моделям.

Далее, случайная величина ε_i имеет условную дисперсию

$$D(\varepsilon_i | x_i) = x_i^T \theta \cdot \left(1 - x_i^T \theta\right) .$$

Таким образом, здесь возникает также проблема ***гетероскедастичности***, осложненная еще и тем, что в выражения для дисперсий ε_i входит и (неизвестный) вектор параметров θ.

Кроме того, если, скажем,

$$y_i = \alpha + \beta x_i + \varepsilon_i , \quad i = 1, \dots, n ,$$

то

$$E(y_i | x_i) = P\{y_i = 1 | x_i\} = \alpha + \beta x_i ,$$

так что если значение x_i увеличить на единицу, то вероятность наличия представляющего интерес признака у i-го субъекта увеличится на величину, равную

$$(\alpha + \beta (x_i + 1)) - (\alpha + \beta x_i) = \beta ,$$

<u>независимо от того</u>, сколь большим или малым является значение x_i.

Более подробное рассмотрение этой ситуации привело нас к необходимости использования нелинейных моделей вероятности вида

$$P\{y_i = 1 | x_i\} = G(\theta_1 x_{i1} + \dots + \theta_p x_{ip}),$$

где $G(z)$ – S-образная функция распределения, имеющего плотность $g(z) = G'(z)$, что соответствует нелинейной модели регрессии

$$y_i = G(\theta_1 x_{i1} + \dots + \theta_p x_{ip}) + \varepsilon_i , \quad i = 1, \dots, n .$$

Предположим, что <u>при фиксированных</u> значениях объясняющих переменных в n наблюдениях, что соответствует фиксированным значениям векторов x_i, случайные ошибки $\varepsilon_1, \dots, \varepsilon_n$ статистически независимы. Тогда при фиксированных x_i статистически независимы и случайные величины $G\left(\theta_1 x_{i1} + \cdots + \theta_p x_{ip}\right) + \varepsilon_i$, $i = 1, \ldots, n$, т.е. статистически независимы $y_1, \ldots, y_n$. В силу этого (условная при фиксированных x_i, $i = 1, \ldots, n$) совместная вероятность получения конкретного набора наблюдений $y_1, \ldots, y_n$ (конкретного набора нулей и единиц) равна произведению

$$\prod_{i=1}^{n} \left(P\{y_i = 1 | x_i\}\right)^{y_i} \left(P\{y_i = 0 | x_i\}\right)^{1-y_i} = \prod_{i=1}^{n} \left(G\left(x_i^T \theta\right)\right)^{y_i} \left(1 - G\left(x_i^T \theta\right)\right)^{1-y_i} .$$

Правая часть этого выражения является при фиксированных x_i, $i = 1, \ldots, n$, функцией от вектора неизвестных параметров θ,

$$L(\theta) = L\left(\theta | x_1, \ldots, x_n\right) = \prod_{i=1}^{n} \left(G\left(x_i^T \theta\right)\right)^{y_i} \left(1 - G\left(x_i^T \theta\right)\right)^{1-y_i}$$

и интерпретируется как ***функция правдоподобия*** параметров $\theta_1, \ldots, \theta_p$. Дело в том, что при различных наборах значений $\theta_1, \ldots, \theta_p$ получаются различные $L(\theta)$, т.е. при фиксированных x_i, $i = 1, \ldots, n$, вероятность наблюдать конкретный набор значений $y_1, \ldots, y_n$ может быть более высокой или более низкой, в зависимости от значения θ. Метод максимального правдоподобия предлагает в качестве оценки вектора параметров θ использовать значение $\theta = \hat{\theta}$, максимизирующее функцию правдоподобия, так что

$$L\left(\hat{\theta}\right) = \max_{\theta} L\left(\hat{\theta}\right) = \max_{\theta} \prod_{i=1}^{n} \left(G\left(x_i^T \theta\right)\right)^{y_i} \left(1 - G\left(x_i^T \theta\right)\right)^{1-y_i} .$$

Использование свойства монотонного возрастания функции $\ln(z)$, позволяет найти <u>то же самое</u> значение $\hat{\theta}$, максимизируя ***логарифмическую функцию правдоподобия*** $\ln L(\theta)$. В нашем случае

$$\ln L(\theta) = \sum_{i=1}^{n} y_i \ln G\left(x_i^T \theta\right) + \sum_{i=1}^{n} (1 - y_i) \ln\left(1 - G\left(x_i^T \theta\right)\right).$$

Если не имеет место чистая мультиколлинеарность объясняющих переменных (т.е. если матрица $X = (x_{ij})$ значений p объясняющих переменных в n наблюдениях имеет ранг p, так что ее столбцы линейно независимы), то тогда функция $L(\theta)$ имеет единственный локальный максимум, являющийся и глобальным максимумом, что гарантирует сходимость соответствующих итерационных процедур к оценке максимального правдоподобия.

Модель бинарного выбора обычно связывается с наличием некоторой ненаблюдаемой (латентной) переменной y_i^*, часто трактуемой как “полезность” интересующего нас признака для i-го субъекта, и такой, что

$$y_i = \begin{cases} 1, & \text{если} \quad y_i^* > \gamma, \\ 0 & - \text{ в противном случае.} \end{cases}$$

Если эта “полезность” определяется соотношением

$$y_i^* = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, \quad i = 1, \ldots, n,$$

и случайные ошибки имеют нормальное распределение, то

$$P\{y_i = 1 | \, x_i\} = G(x_i^T \beta) = \Phi(\beta_1 x_{i1} + \cdots + \beta_p x_{ip}).$$

В стандартной модели полагается $\gamma = 0$.

Нарушение предположения об <u>одинаковой распределенности</u> случайных ошибок ε_i в процессе порождения данных приводит к гетероскедастичной модели и к <u>несостоятельности оценок максимального правдоподобия, получаемых на основании стандартной модели.</u>

Перейдем теперь к панельным данным. И здесь модель бинарного выбора обычно связывается с наличием некоторой ненаблюдаемой (латентной) переменной y_{it}^* и наблюдаемой индикаторной переменной y_{it} такой, что

$$y_{it} = \begin{cases} 1, & \text{если} \quad y_{it}^* > 0, \\ 0 & \text{– в противном случае.} \end{cases}$$

Например, y_{it} может указывать на то, работает ли i-й индивид в период t или нет. Типичной является модель, в которой

$$y_{it}^* = x_{it}^T \beta + \alpha_i + \varepsilon_{it} \,,$$

где x_{it}^T – вектор-строка значений объясняющих переменных для i-го субъекта в период t .

Предположим, что случайные ошибки ε_{it} независимы и одинаково распределены в обоих направлениях и имеют симметричное распределение с функцией распределения G и что объясняющие переменные строго экзогенны.

Если трактовать α_i как фиксированные неизвестные параметры, то это соответствует включению в модель N дамми-переменных. Логарифмическая функция правдоподобия принимает тогда вид:

$$\ln L(\beta, \alpha_1, \ldots, \alpha_N) =$$
$$= \sum_{i,t} y_{it} \ln G(\alpha_i + x_{it}^T \beta) + \sum_{i,t} (1 - y_{it}) \ln\left(1 - G\left(\alpha_i + x_{it}^T \beta\right)\right).$$

И здесь возникает ситуация, когда оценка максимального правдоподобия для β и α_i даже при выполнении стандартных предположений состоятельна только если $T \to \infty$, а при конечном T и $N \to \infty$ она несостоятельна. Мы встречались уже с такой ситуацией в рамках OLS-оценивания линейной модели с фиксированными эффектами. Только там при несостоятельности оценок для α_i оценка для β оставалась все же состоятельной, тогда как здесь несостоятельность оценок для α_i в общем случае переносится и на оценку для β . Одним из исключений является *линейная модель вероятности*, в которой вероятность $P\{y_{it} = 1 | \, x_{it}\}$ моделируется как линейная функция от объясняющих переменных.

Конечно, по подобию проделанного раньше, возникает желание вымести α_i непосредственно из уравнения $y_{it}^* = x_{it}^T \beta + \alpha_i + \varepsilon_{it}$. Однако это не дает практической пользы, поскольку, например, если выметание производится путем перехода в этом уравнении к разностям $y_{it}^* - y_{i,t-1}^*$, то не удается указать переход к наблюдаемым переменным типа $y_{it} - y_{i,t-1}$.

Альтернативный подход состоит в использовании ***условного максимального правдоподобия*** *(conditional FE)*.

В этом случае рассматривается функция правдоподобия, условная относительно некоторого множества статистик t_i, которое ***достаточно*** для α_i. Это означает, что при заданных значениях этих статистик вклад i-го субъекта в правдоподобие уже не зависит от α_i. Таким образом, если $f\left(y_{i1},\ldots,y_{iT}\middle|\alpha_i,\beta\right)$ – совместная вероятность значений индикаторной переменной для i-го субъекта, то

$$f\left(y_{i1},\ldots,y_{iT}\middle|t_i,\alpha_i,\beta\right) = f\left(y_{i1},\ldots,y_{iT}\middle|t_i,\beta\right).$$

Соответственно, мы можем получить состоятельную оценку для β, максимизируя условную функцию правдоподобия, основанную на $f\left(y_{i1},\ldots,y_{iT}\middle|t_i,\beta\right)$.

Наличие указанных достаточных статистик зависит от вида распределения ε_{it}. В ***линейной модели вероятности с нормальными*** ошибками такой достаточной статистикой является $\bar{y}$. Максимизация соответствующей условной функции правдоподобия приводит к ***FE-оценке*** для α_i. Однако, например, в случае ***пробит-модели***, в которой нормальными являются ошибки в уравнении

$$y_{it}^* = x_{it}^T \beta + \alpha_i + \varepsilon_{it},$$

достаточной статистики для α_i просто не существует. А это означает, что мы не можем состоятельно оценить пробит-модель с фиксированными эффектами при конечном *T*.

3.12.1. Логит-модель с фиксированными эффектами

Такой модели соответствует латентное уравнение

$$y_{it}^{*} = x_{it}^{T}\beta + \alpha_i + \varepsilon_{it},$$

в котором ε_{it} имеют логистическое распределение $G(z) = \Lambda(z) = \dfrac{e^z}{1+e^z}$. При этом

$$P\{y_{it} = 1 | x_{it}, \alpha_i, \beta\} = \frac{\exp\{x_{it}^T\beta + \alpha_i\}}{1+\exp\{x_{it}^T\beta + \alpha_i\}},$$

$$P\{y_{it} = 0 | x_{it}, \alpha_i, \beta\} = 1 - \frac{\exp\{x_{it}^T\beta + \alpha_i\}}{1+\exp\{x_{it}^T\beta + \alpha_i\}} = \frac{1}{1+\exp\{x_{it}^T\beta + \alpha_i\}}.$$

Рассмотрим для иллюстрации логит-модель с $T = 2$. Заметим, что в этом случае сумма $y_{i1} + y_{i2}$ есть просто общее количество периодов безработицы (суммарная продолжительность пребывания в состоянии безработицы) для i-го субъекта, и что для каждого субъекта есть 4 возможных последовательности состояний (y_{i1}, y_{i2}): $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$. Соответственно, сумма $y_{i1} + y_{i2}$ может принимать только 3 возможных значения: $0, 1, 2$ (у i-го субъекта не было периодов безработицы, был один период безработицы, был безработным в течение обоих периодов).

Вычислим условные вероятности указанных четырех последовательностей при различных значениях суммы $y_{i1} + y_{i2}$.

Если $y_{i1} + y_{i2} = 0$, то возможна только последовательность $(0,0)$, так что

$$P\{(0,0)|y_{i1}+y_{i2}=0,\alpha_i,\beta\}=1,$$

а условные вероятности трех других последовательностей при $y_{i1}+y_{i2}=0$ равны 0.

Если $y_{i1}+y_{i2}=2$, то возможна только последовательность $(1,1)$, так что

$$P\{(1,1)|y_{i1}+y_{i2}=1,\alpha_i,\beta\}=1,$$

а условные вероятности трех других последовательностей при $y_{i1}+y_{i2}=1$ равны 0.

Если же $y_{i1}+y_{i2}=1$, то возможны две последовательности $(0,1)$ и $(1,0)$, и они имеют условные вероятности, отличные от 0 и 1, тогда как условные вероятности последовательностей $(0,0)$ и $(1,1)$ равны 0. При этом

$$P\{(0,1)|y_{i1}+y_{i2}=1,x_{it},\alpha_i,\beta\}=\frac{P\{(0,1)|x_{it},\alpha_i,\beta\}}{P\{y_1+y_2=1|x_{it},\alpha_i,\beta\}}=$$

$$=\frac{P\{y_{i1}=0|x_{it},\alpha_i,\beta\}P\{y_{i2}=1|x_{it},\alpha_i,\beta\}}{P\{y_{i1}=0|x_{it},\alpha_i,\beta\}P\{y_{i2}=1|x_{it},\alpha_i,\beta\}+P\{y_{i1}=1|x_{it},\alpha_i,\beta\}P\{y_{i2}=0|x_{it},\alpha_i,\beta\}}$$

$$=\frac{\dfrac{1}{1+\exp\{x_{i1}^T\beta+\alpha_i\}}\cdot\dfrac{\exp\{x_{i2}^T\beta+\alpha_i\}}{1+\exp\{x_{i2}^T\beta+\alpha_i\}}}{\dfrac{1}{1+\exp\{x_{i1}^T\beta+\alpha_i\}}\cdot\dfrac{\exp\{x_{i2}^T\beta+\alpha_i\}}{1+\exp\{x_{i2}^T\beta+\alpha_i\}}+\dfrac{\exp\{x_{i1}^T\beta+\alpha_i\}}{1+\exp\{x_{i1}^T\beta+\alpha_i\}}\cdot\dfrac{1}{1+\exp\{x_{i2}^T\beta+\alpha_i\}}}=$$

$$=\frac{\exp\{x_{i2}^T\beta+\alpha_i\}}{\exp\{x_{i1}^T\beta+\alpha_i\}+\exp\{x_{i2}^T\beta+\alpha_i\}}=\frac{\exp\{x_{i2}^T\beta\}}{\exp\{x_{i1}^T\beta\}+\exp\{x_{i2}^T\beta\}},$$

так что индивидуальные эффекты выметаются, и

$$P\{(0,1)|y_{i1}+y_{i2}=1,x_{it},\alpha_i,\beta\}=\frac{\exp\{(x_{i2}-x_{i1})^T\beta\}}{1+\exp\{(x_{i2}-x_{i1})^T\beta\}}.$$

Соответственно,

$$P\{(1,0)|y_{i1}+y_{i2}=1,x_{it},\alpha_i,\beta\}=1-P\{(0,1)|y_{i1}+y_{i2}=1,x_{it},\alpha_i,\beta\}=$$

$$=\frac{1}{1+\exp\left\{(x_{i2}-x_{i1})^T\beta\right\}}.$$

Таким образом, все условные вероятности

$$P\{(y_{i1},y_{i2})=(r,s)|y_{i1}+y_{i2}=l,\ x_{it},\alpha_i,\beta\},$$

где $r,s=0,1,\ \ l=0,1,2$, не зависят от индивидуальных эффектов, так что продолжительность периодов безработицы является достаточной статистикой для α_i , и использование условного правдоподобия приводит к состоятельной оценке для β .

Полученные результаты означают, что при $T=2$ мы можем оценивать логит-модель с фиксированными эффектами, используя стандартную логит модель, в которой в качестве объясняющей переменной выступает $x_{i2}-x_{i1}$, а в качестве наблюдаемой бинарной переменной – переменная, отражающая изменение значения переменной y_{it} при переходе от первого ко второму наблюдению (1 – при возрастании y_{it} , 0 – при убывании y_{it}). Соответственно, функцию правдоподобия можно записать в виде:

$$L=\prod_{i=1}^{N}P\{(0,0)|\cdot\}P\{(0,1)|\cdot\}P\{(1,0)|\cdot\}P\{(1,1)|\cdot\}=$$

$$=\prod_{i=1}^{N}1^{S_{00}}\cdot\left(\frac{\exp\left\{(x_{i2}-x_{i1})^T\beta\right\}}{1+\exp\left\{(x_{i2}-x_{i1})^T\beta\right\}}\right)^{S_{01}}\cdot\left(\frac{1}{1+\exp\left\{(x_{i2}-x_{i1})^T\beta\right\}}\right)^{S_{10}}\cdot 1^{S_{11}},$$

где $S_{rs}=1$, если $y_{i1}=r$, $y_{i2}=s$, и $S_{rs}=0$ в противном случае.

Мы видим, что субъекты, для которых $y_{i1}=y_{i2}=0$ или $y_{i1}=y_{i2}=1$, не вносят дополнительный вклад в функцию правдоподобия и фактически игнорируются при оценивании. Для оценивания β существенны только те субъекты, которые хотя бы однажды изменяют свой статус в отношении y_{it} .

Параметр β идентифицируется здесь по одному только “within”-измерению данных. Вклад в оценивание вносят только те субъекты, которые изменили свой статус. При значениях $T > 2$ вычисление условных вероятностей более хлопотно, но также возможно.

Проверку гипотезы H_0 об отсутствии индивидуальных эффектов можно осуществить с помощью критерия типа Хаусмана, основанного на разности между оценкой $\hat{\beta}_{CML}$ для β, полученной с использованием условного правдоподобия и обычной логит ML-оценкой $\hat{\beta}_{ML}$, игнорирующей индивидуальные эффекты (при вычислении последней константа исключается). Логит MLE состоятельна и эффективна только при гипотезе H_0 и несостоятельна при альтернативе. Условная же MLE состоятельна и при гипотезе H_0 и при альтернативе, но не эффективна, так как использует не все данные. Таким образом, положение здесь соответствует тому, при котором реализуется схема Хаусмана (см. Главу 2, разд. 2.6.5.). Статистика критерия

$$H = \left(\hat{\beta}_{CML} - \hat{\beta}_{ML}\right)^T \left(C\hat{o}v\left(\hat{\beta}_{CML}\right) - C\hat{o}v\left(\hat{\beta}_{ML}\right)\right)^{-1} \left(\hat{\beta}_{CML} - \hat{\beta}_{ML}\right)$$

имеет при гипотезе H_0 асимптотическое распределение хи-квадрат с числом степеней свободы, равным размерности вектора β.

З а м е ч а н и е

Существенным недостатком метода условного правдоподобия является то, что он предполагает (условную) независимость $y_{i1},\dots,y_{iT}$ при фиксированных x_{it},α_i,β. Метод неприменим к моделям с запаздывающими значениями объясняющей переменной в правой части.

3.12.2. Пробит-модель со случайными эффектами

Если α_i статистически независимы от объясняющих переменных, то модель со случайными эффектами представляется более подходящей, и в этом случае проще рассматривать не логит, а пробит-модель. (Хотя можно использовать и логит-модель, что и делается в приводимом далее примере.)

Здесь мы отправляемся от спецификации

$$y_{it} = \begin{cases} 1, & \text{если} \quad y_{it}^* > 0, \\ 0 & - \text{в противном случае,} \end{cases}$$

$$y_{it}^* = x_{it}^T \beta + u_{it} ,$$

$$u_{it} = \alpha_i + \varepsilon_{it} ,$$

где

α_i – случайные индивидуальные эффекты, $\alpha_i \sim i.i.d.\ N\left(0, \sigma_\alpha^2\right)$,

ε_{it} – случайные ошибки, $\varepsilon_{it} \sim i.i.d.\ N\left(0, \sigma_\varepsilon^2\right)$,

причем индивидуальные эффекты и ошибки статистически независимы. В этих условиях совместное распределение случайных величин $u_{i1}, \ldots, u_{iT}$ нормально, причем $E(u_{it}) = 0$, $D(u_{it}) = \sigma_\alpha^2 + \sigma_\varepsilon^2$, $Cov(u_{it}, u_{is}) = Cov(\alpha_i + \varepsilon_{it}, \alpha_i + \varepsilon_{is}) = \sigma_\alpha^2$ для $s \neq t$, так что значение коэффициента корреляции $\rho = \dfrac{Cov(u_{it}, u_{is})}{\sqrt{D(u_{it})}\sqrt{D(u_{is})}} = \dfrac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_\varepsilon^2}$ между ошибками u_{it} и u_{is} внутри одной группы (субъекта) одинаково для любых $s \neq t$. В “стандартной” спецификации $D(u_{it}) = 1$ и $\rho = \sigma_\alpha^2$. Это соответствует предположению о том, что $\sigma_\varepsilon^2 = 1 - \sigma_\alpha^2$.

Совместная вероятность получения набора $y_{i1}, \ldots, y_{iT}$ при заданных $x_{i1}, \ldots, x_{iT}$ определяется как

$$f\left(y_{i1},\ldots,y_{iT}\middle|x_{i1},\ldots,x_{iT},\beta\right)=\int_{-\infty}^{\infty}f\left(y_{i1},\ldots,y_{iT}\middle|x_{i1},\ldots,x_{iT},\alpha_i,\beta\right)f\left(\alpha_i\right)d\alpha_i =$$

$$=\int_{-\infty}^{\infty}\left[\prod_{t=1}^{T}f\left(y_{it}\middle|x_{it},\alpha_i,\beta\right)\right]f\left(\alpha_i\right)d\alpha_i.$$

В рассматриваемой пробит-модели

$$f\left(y_{it}\middle|x_{it},\alpha_i,\beta\right)=\begin{cases}\Phi\left(\dfrac{x_{it}^T\beta+\alpha_i}{\sqrt{1-\sigma_\alpha^2}}\right), & \text{если}\quad y_{it}=1,\\ 1-\Phi\left(\dfrac{x_{it}^T\beta+\alpha_i}{\sqrt{1-\sigma_\alpha^2}}\right), & \text{если}\quad y_{it}=0,\end{cases}$$

$$f\left(\alpha_i\right)=\frac{1}{\sqrt{2\pi\sigma_\alpha^2}}\exp\left\{-\frac{1}{2}\frac{\alpha_i^2}{\sigma_\alpha^2}\right\}.$$

Итоговый интеграл вычисляется численными методами. Соответствующие программы имеются в некоторых пакетах (например, в Stata).

При гипотезе H_0 об отсутствии индивидуальных эффектов

$$\rho=\sigma_\alpha^2=0,$$

так что проверка отсутствия индивидуальных эффектов сводится к проверке гипотезы $\rho=0$.

З а м е ч а н и е

В контексте панельных данных построение порядковых пробит моделей с несколькими категориями отклика довольно затруднительно.

3.12.3. Пример

Рассмотрим результаты исследования панельных данных по Германии (*German Socio-Economic Panel*) за 5-летний период с 1985 по 1989 гг., проводившихся с целью выяснения влияния состояния безработицы на степень удовлетворенности жизнью (“well-being”). Исходные данные были представлены в порядковой шкале и индексировались числами от 0 до 10. Ввиду сложности построения порядковых логит- и пробит-моделей для панельных данных, данные были затем сжаты до бинарных. Индивиды, у которых удовлетворенность жизнью индексировалась числами от 0 до 4, считались неудовлетворенными жизнью, тогда как индивиды с индексами 5 и выше считались более или менее удовлетворенными жизнью.

В качестве объясняющих переменных первоначально были привлечены следующие характеристики индивидов:

Непрерывная переменная:

LOGINCOME – логарифм (располагаемого) дохода – см. ниже.

Дамми-переменные:

UNEMP (безработный), *NOPARTIC* (не включен в рынок труда), *SELFEMP* (самозанятый), *PARTTIME* (частично занятый) – указывают текущий сатус на рынке труда (категория “полностью занятый” не снабжается дамми-переменнной),

MALE – индивид мужского пола,

OKHEALTH – хорошее состояние здоровья,

AGE – возраст и *AGESQUARED* – квадрат возраста,

VOCATIONAL D. – наличие профессиональной степени,

UNIVERSITY D. – наличие университетской степени,

MARRIED – состояние в браке.

Для правильной интерпретации влияния безработицы на удовлетворенность жизнью надо учитывать целый ряд обстоятельств.

Например, безработица обычно ведет к сокращению дохода, что, в свою очередь, может уменьшить удовлетворенность. Однако, если доход включается в число объясняющих переменных, то коэффициент при безработице фактически измеряет влияние безработицы *при прочих равных*, т.е. при сохранении дохода постоянным. Это могло быть в реальности, если бы страховое возмещение при безработице достигало 100%. При высоком (отрицательном) коэффициенте корреляции между безработицей и доходом оценки остаются несмещенными, но их точность может значительно уменьшаться.

Удовлетворенность индивида может уменьшаться вследствие безработицы также и потому, что доля его вклада в общий семейный бюджет уменьшается. Поскольку, в отличие от доходов домохозяйства, имеющиеся в панели данные в отношении индивидуального дохода ограничены прошлыми доходами, в уравнение включается не индивидуальный доход, а доход домохозяйства.

Включение в модель взаимодействий факторов позволяет выделять дифференциальные эффекты влияния безработицы в группах, имеющих разные атрибуты, например, в разных возрастных группах. В связи с этим в модель включены помимо собственно UNEMP также и переменные

*UNEMP*AGE<29*

*UNEMP*AGE30-49*.

Другим важным взаимодействием является взаимодействие между безработицей и продолжительностью текущего периода пребывания в состоянии безработного. Влияние продолжительности безработицы на психическое состояние индивида достаточно хорошо документировано в литературе по психологии. В соответствии с этим, в уравнение включается переменная

UNEMPDUR (продолжительность периода безработицы, продолжающегося в настоящее время) и, для учета возможной нелинейности, переменная

*UNEMPDURSQ*10*$^{-2}$.

К сожалению, в данных полностью отсутствовал региональный аспект, так что не удалось выяснить влияние локального уровня безработицы.

Следует отметить очень существенное взаимодействие безработицы и пола индивида. По этой причине обработка данных производилась также раздельно по мужчинам и женщинам.

Наличие протяженных наблюдений позволяет, в принципе, включать в модель элементы динамики. Мы, однако, уже говорили о проблемах с оцениванием динамических моделей бинарного выбора, в частности, с включением в правую часть запаздывающих значений объясняемой переменной. В данном анализе предполагается устойчивость индивидуальных откликов во времени, и она как раз учитывается введением в модель индивидуальных эффектов. Для проверки этого свойства в модель помимо переменной *UNEMPDUR* включается также переменная *PREVDUR*, указывающая на суммарную продолжительность периодов безработицы за 10 лет, предшествующих анализируемому периоду.

Кроме того в модели используется в качестве объясняющей еще и переменная *CHANGEINC* – относительное изменение дохода домохозяйства по сравнению с предыдущим годом. Это связано с некоторой неясностью в отношении того, что именно влияет на удовлетворенность – собственно уровень доходов или его изменение.

Наконец, нельзя исключать возможного наличия и обратной причинной связи между безработицей и уровнем удовлетворенности, т.е. того, что внутренне менее удовлетворенные индивиды скорее и оказываются безработными. Если это так, то переменная *UNEMP* коррелирует со случайным эффектом в модели случайных эффектов и оценка максимального правдоподобия, не учитывающая этого, оказывается несостоятельной. Поскольку далее приводятся результаты статистического анализа и по модели с фиксированными эффектами и по модели со случайными эффектами, имеется возможность оценить устойчивость результатов к этой возможной эндогенности.

Основная задача состоит в проверке гипотезы о негативном влиянии безработицы на удовлетворенность при очистке от влияния прочих факторов, включая влияние уменьшения доходов.

В приводимых ниже *табл.* A1 и A2 указаны результаты оценивания порядковой пробит-модели, не учитывающей панельной структуры данных (данные усреднены по времени), но включающей временные дамми. В рассмотренной порядковой пробит-модели предполагается, что уравнение полезности имеет вид

$$S_{it}^* = x_{it}^T \beta + \varepsilon_{it} \,,$$

и наблюдаемое значение индекса удовлетворенности определяется соотношениями

$$S_{it} = \begin{cases} 0, & \text{если } S_{it}^* < 0, \\ 1, & \text{если } 0 < S_{it}^* < \gamma_1, \\ & \vdots \\ 10, & \text{если } S_{it}^* > \gamma_9. \end{cases}$$

Таблицы B1 и B2 содержат результаты оценивания FE логит-модели и RE пробит-модели с бинарным откликом (1 – если уровень удовлетворенности равен 5 или выше, 0 – если уровень удовлетворенности ниже 5).

Таблица A1 (объединенная выборка)

	Pooled	
	coeff.	s.e.
Constant	1.661	(0.168)
MALE	-0.059	(0.015)
UNEMP	-0.265	(0.050)

*UNEMP*AGE<29*	-0.164	(0.078)
*UNEMP*AGE30-49*	-0.221	(0.065)
UNEMPDUR	-0.011	(0.007)
*UNEMPDURSQ**10^{-2}	0.013	(0.020)
PREVDUR	-0.009	(0.001)
NOPARTIC	-0.042	(0.018)
SELFEMP	-0.107	(0.033)
PARTTIME	-0.060	(0.028)
MARRIED	0.140	(0.017)
OKHEALTH	0.428	(0.014)
AGE	-0.050	(0.004)
*AGESQ**10^{-2}	0.062	(0.005)
VOCATIONAL D.	0.083	(0.013)
UNIVERSITY D.	0.106	(0.033)
LOGINCOME	0.208	(0.017)
CHANGEINC	0.018	(0.022)
1986	-0.027	(0.020)
1987	-0.127	(0.020)
1988	-0.191	(0.020)
1989	-0.219	(0.021)
Observations	24055	

Таблица A2 (раздельное оценивание по мужчинам и женщинам)

	Male		Female	
	coeff.	s.e.	coeff.	s.e.
Constant	2.273	(0.246)	1.259	(0.234)
UNEMP	-0.402	(0.063)	-0.118	(0.089)
*UNEMP*AGE<29*	-0.321	(0.105)	-0.064	(0.126)
*UNEMP*AGE30-49*	-0.245	(0.085)	-0.175	(0.109)
UNEMPDUR	-0.003	(0.010)	-0.010	(0.018)
UNEMPDURSQ$*10^{-2}$	-0.013	(0.028)	0.042	(0.064)
PREVDUR	-0.010	(0.001)	-0.008	(0.001)
NOPARTIC	-0.273	(0.035)	0.056	(0.024)
SELFEMP	-0.087	(0.038)	-0.155	(0.069)
PARTTIME	-0.180	(0.129)	-0.001	(0.031)
MARRIED	0.133	(0.024)	0.110	(0.026)
OKHEALTH	0.409	(0.021)	0.426	(0.021)
AGE	-0.063	(0.006)	-0.049	(0.006)
AGESQ$*10^{-2}$	0.078	(0.007)	0.062	(0.007)
VOCATIONAL D.	0.066	(0.019)	0.093	(0.019)
UNIVERSITY D.	0.086	(0.040)	0.174	(0.059)
LOGINCOME	0.165	(0.024)	0.253	(0.025)
CHANGEINC	0.014	(0.030)	0.015	(0.031)
1986	-0.018	(0.028)	-0.037	(0.029)
1987	-0.104	(0.028)	-0.152	(0.030)
1988	-0.175	(0.028)	-0.206	(0.030)
1989	-0.178	(0.029)	-0.258	(0.030)
Observations	12605		11450	

Таблица В1 (мужчины)

	MALE			
	Random Effects		Fixed Effects	
	coef.	s.e.	coef.	s.e.
Constant	1.849	(0.682)		
UNEMP	-0.849	(0.145)	-1.257	(0.358)
*UNEMP*AGE<29*	-0.089	(0.225)	0.115	(0.518)
UNEMP+AGE30-49	-0.013	(0.187)	0.368	(0.419)
UNEMPDUR	-0.026	(0.025)	-0.030	(0.047)
UNEMPDURSQ$*10^{-2}$	0.027	(0.072)	0.109	(0.145)
PREVDUR	-0.011	(0.004)		
NOPARTIC	-0.511	(0.084)	-0.287	(0.243)
SELFEMP	-0.185	(0.107)	0.643	(0.418)
PARTTIME	-0.274	(0.274)	-0.162	(0.554)
MARRIED	0.294	(0.068)	0.575	(0.254)
OKHEALTH	0.448	(0.052)	0.409	(0.128)
AGE	-0.073	(0.019)	0.009	(0.114)
AGESQ$*10^{-2}$	0.086	(0.023)	-0.112	(0.131)
VOCATIONAL D.	0.098	(0.060)		
UNIVERSITY D.	0.057	(0.115)		
LOGINCOME	0.135	(0.068)	0.676	(0.245)
CHANGEINC	0.158	(0.072)	0.086	(0.174)
ρ	0.382	(0.040)		
No. of individuals	2521		556	

Таблица B2 (женщины)

	FEMALE			
	Random Effects		Fixed Effects	
	coef.	s.e.	coef.	s.e.
Constant	-0.160	(0.639)		
UNEMP	0.003	(0.209)	-0.192	(0.484)
*UNEMP*AGE<29*	-0.255	(0.299)	-0.069	(0.641)
*UNEMP*AGE30-49*	-0.550	(0.240)	-0.821	(0.548)
UNEMPDUR	-0.017	(0.039)	0.014	(0.074)
UNEMPDURSQ$*10^{-2}$	0.047	(0.144)	-0.022	(0.270)
PREVDUR	-0.009	(0.003)		
NOPARTIC	0.007	(0.069)	-0.180	(0.225)
SELFEMP	-0.057	(0.191)	-0.268	(0.497)
PARTTIME	0.022	(0.087)	-0.206	(0.241)
MARRIED	0.220	(0.074)	0.561	(0.261)
OKHEALTH	0.529	(0.053)	0.620	(0.126)
AGE	-0.035	(0.019)	-0.179	(0.108)
AGESQ$*10^{-2}$	0.042	(0.023)	0.066	(0.125)
VOCATIONAL D.	0.149	(0.063)		
UNIVERSITY D.	0.280	(0.185)		
LOGINCOME	0.278	(0.068)	0.356	(0.233)
CHANGEINC	0.159	(0.073)	0.265	(0.160)
ρ	0.411	(0.039)		
No. of individuals	2290		538	

Результаты, приведенные в таблицах, подтверждают сделанные ранее выводы. В частности, влияние безработицы представляется отрицательным. Однако следует помнить, что в нелинейных моделях оцененные коэффициенты не представляют предельные эффекты[2]. Более того, они не сравнимы для разных спецификаций модели.

Тем не менее в рамках оценивания порядковой пробит модели без разделения по полу, можно отметить следующие моменты:

- Влияние пола оказывается весьма значимым (коэффициент –0.059 при стандартной ошибке 0.015), что говорит в пользу раздельного анализа данных по мужчинам и женщинам.
- Подтверждаются разультаты более ранних исследований, указывающие на U-образную форму степени негативного влияния безработицы на удовлетворенность: оцененный коэффициент выше по абсолютной величине в наиболее активных возрастных группах.
- Уровень доходов имеет существенный положительный эффект при измерении его в уровнях, но не в приращениях.
- Положительное влияние имеют наличие профессионального или университетского диплома, а также состояние в браке.

При раздельном оценивании отдельно по мужчинам и по женщинам обнаруживаются большие дифференциальные эффекты, так что публикация результатов только по объединенной выборке, как это делают некоторые авторы, может вводить в заблуждение. Наиболее существенное различие: на мужчин состояние безработного действует в значительно большей степени. Так, для мужчин среднего возраста эффект безработицы равен

$$-0.402-0.245 = -0.65,$$

тогда как для женщин он равен только

$$-0.118-0.175 = -0.29.$$

[2] См. главу 1, разд. 1.4.

Это соответствует росту неудовлетворенности на 12% для мужчин и только на 4% для женщин.

Отметим также, что пребывание вне рынка рабочей силы негативно влияет на мужчин, но не на женщин.

При переходе к анализу собственно панельными методами следует заметить, что в FE-модели выметаются все регрессоры, значения которых не изменяются во времени (*Const*, *PREVDUR*, *VOCATIONAL D.*, *UNIVERSITY D.*) Кроме того, количество индивидов, по которым производится оценивание, сокращается, поскольку в выборку в этом случае надо включать только тех индивидов, у которых значение бинарного отклика изменялось на 5-летнем периоде исследования. Переход к FE- и RE-моделям не очень сильно повлиял на выводы, сделанные ранее на основе порядковой модели.

3.13. Тобит-модели

Тобит-модель со случайными эффектами отличается от пробит-модели со случайными эффектами механизмом наблюдений. В тобит-модели

$$y_{it} = \begin{cases} y_{it}^*, & \text{если} \quad y_{it}^* > 0, \\ 0 \; - \text{в противном случае.} \end{cases}$$

В остальном методика та же:

$$f\left(y_{i1},\ldots,y_{iT} \middle| x_{i1},\ldots,x_{iT},\beta\right) = \int_{-\infty}^{\infty}\left[\prod_{t=1}^{T} f\left(y_{it} \middle| x_{it},\alpha_i,\beta\right)\right] f(\alpha_i)\,d\alpha_i\,,$$

где

$$f(\alpha_i) = \frac{1}{\sqrt{2\pi\sigma_\alpha^2}}\exp\left\{-\frac{1}{2}\frac{\alpha_i^2}{\sigma_\alpha^2}\right\},$$

$$f\left(y_{it}\middle|x_{it},\alpha_i,\beta\right)=\begin{cases}\dfrac{1}{\sqrt{2\pi\sigma_\varepsilon^2}}\exp\left\{-\dfrac{1}{2}\dfrac{\left(y_{it}-x_{it}^T\beta-\alpha_i\right)^2}{\sigma_\varepsilon^2}\right\}, & \text{если} \quad y_{it}>0,\\ 1-\Phi\left(\dfrac{x_{it}^T\beta+\alpha_i}{\sigma_\varepsilon}\right), & \text{если} \quad y_{it}=0.\end{cases}$$

П р и м е р

Аукционы РЕПО (кредит под залог ценных бумаг) являются основным инструментом управления рынком со стороны Европейского Центрального Банка (ЕЦБ). На этих аукционах отдельные банки подают заявки – “биды” (bids) на получение кредита на определенный срок (maturity), в которых одновременно указываются величина кредита и проценты по кредиту, под которые покупатель кредита предполагает получить соответствующий кредит. Центральный банк на каждом аукционе выделяет для предоставления кредитов некоторую сумму и последовательно удовлетворяет полученные заявки банков, начиная с тех, в которых указан максимальный среди поданных заявок процент по кредиту. Отличительной чертой аукционов РЕПО, проводимых ЕЦБ, является предварительное объявление Центральным банком минимального процента по бидам – нижней границы процентной ставки по кредиту, по которой еще принимаются заявки. Эта граница весьма определенно указывает на направленность политики ЕЦБ и обычно устанавливает нижнюю границу для краткосрочных процентных ставок в зоне евро. Однако, когда банки ожидают снижения процентных ставок, текущая минимальная ставка представляется им завышенной, и банки начинают воздерживаться от участия в аукционах. Такое поведение банков серьезно затрудняет управление ликвидностью со стороны ЕЦБ, повышает волатильность процентных ставок. В этой связи представляет интерес выяснение роли минимальной ставки. Для этой цели можно использовать данные об индивидуальных заявках банков на аукционах без установления минимальной ставки. Такие аукционы РЕПО проводились Бундесбанком Германии. Важно, что в остальном аукционы РЕПО, проводимые Бундесбанком и ЕЦБ, следовали одним и тем же правилам. Это дает уникальную возможность исследования действительной роли минимальной ставки.

Используя панельные данные, можно оценить модели для вероятности неучастия банка в аукционе и для размера индивидуальной заявки банка. Последняя переменная цензурирована слева, поскольку она наблюдается только в случае участия банка в аукционе, и это обстоятельство можно учесть, используя панельную тобит-модель.

Для анализа были взяты данные о заявках 275 банков, поданных в Центробанке земли Гессен. Эта выборка достаточно репрезентативна, поскольку в этой земле находится Франкфурт – финансовый центр Германии. РЕПО аукционы проводились еженедельно. Но фактически для анализа подходит только один достаточно долгий период, когда Бундесбанк не изменял формат аукциона. Это период с апреля по ноябрь 1995 года. Соответственно, за этот период было проведено 33 аукциона со стандартным сроком кредита около двух недель.

Во многих отношениях поведение банков на этих аукционах сходно с поведением банков на аукционах, проводимых ЕЦБ. Во-первых, более крупные покупатели участвуют в аукционах чаще, но их заявки относительно малы по сравнению с общим объемом заявок. Во-вторых, банки обычно подают на одном аукционе не более трех заявок (разные заявки ориентируются на разные проценты). В-третьих довольно много покупателей участвует в аукционах редко. Из 275 банков только 175 хотя бы раз участвовали в аукционах на рассматриваемом периоде. Подобная картина имеет место и в РЕПО аукционах ЕЦБ.

Заметим, что на указанном периоде наблюдалось падение процентных ставок. В частности, в августе 1995 г. Бундесбанк понизил ломбардную процентную ставку на 50 базисных пунктов. И это дает возможность посмотреть, как повлияло это на поведение банков на аукционах.

На спрос банков на РЕПО аукционах должна влиять стоимость альтернативных возможностей рефинансирования. В связи с этим в число объясняющих переменных включается переменная *spread*, определяемая как разность между ожидаемой ставкой отсечения на аукционе и ставкой овернайт. Ожидаемая ставка отсечения была рассчитана на основании уравнения коррекции ошибок[3], соответствующего выявленной коинтеграционной связи между ставкой отсечения и ставкой овернайт на межбанковском рынке.

Следующая объясняющая переменная *term spread* (временной спред) определяется как разность между одномесячной процентной ставкой и овернайт ставкой. Отрицательное значение временного спреда означает ожидание уменьшения процентных ставок. Для аукционов РЕПО ЕЦБ ожидаемые изменения ключевых процентных ставок ЕЦБ оказывают весьма сильное влияние на поведение банков. В частности, когда банки ожидают снижения процентных ставок, неучастие их в аукционах затрудняет управление ликвидностью со стороны центрального банка. Ориентируясь на движение временного спреда, банки также должны были ожидать снижения Бундесбанком ломбардной ставки в августе 1995 г. Для учета возможного "забастовочного" поведения банков, в число объясняющих переменных включается дамми-переменная *underbidding*, принимающая значение 1 для последнего аукциона, предшествовавшего понижению Бундесбанком ломбардной ставки.

Неопределенность, ощущаемая в день аукциона, отражает переменная *volatility*, которая оценивается на основе модели[4] EGARCH(1,1) для ежедневных наблюдений значений ставки овернайт, в которой условная дисперсия ε_t относительно прошлого удовлетворяет соотношению

---

[3] См., например, [Носко (2004)].
[4] См., например, [Verbeek (2000)].

$$\ln\sigma_t^2 = \omega + \beta \ln\sigma_{t-1}^2 + \alpha\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right| + \gamma\frac{\varepsilon_{t-1}}{\sigma_{t-1}}.$$

Влияние такой неопределенности может быть двояким. С одной стороны, при возрастании неопределенности банки становятся более осторожными. С другой стороны, если дело касается возможного проигрыша в аукционе, то чем больше неопределенность, тем большей может оказаться заявка банка и под более высокие проценты.

Переменная *reserve fulfillment* измеряет потребность банковского сектора в ликвидности. Она определялась как отношение резервных остатков (reserve holdings) всех банков Германии перед проведением аукциона к агрегированному требованию минимального резерва. Если резервные остатки малы, то банки имеют более сильные побуждения к участию в аукционе. Поскольку Бундесбанк допускает усреднение по (обычно месячному) периоду между проверками выполнения требований в отношении резервов (maintenance period), этот эффект может быть особенно выражен на последнем аукционе этого периода, и для учета этого вводится дамми-переменная *end of period*, равная единице, если аукцион является последним на указанном периоде.

Еще одна переменная *maturing allotment* учитывает тот факт, что банки часто используют двухнедельный РЕПО кредит на возобновляемой основе. Эта переменная определяется как логарифм размера РЕПО, полученного банком двумя неделями ранее.

Дамми переменные *large*, *medium*, and *small* характеризуют размер банка, отражающийся в среднем размере заявки. Эти переменные комбинируются со всеми другими объясняющими переменными для изучения поведения банков в зависимости от их размеров.

<u>Эмпирические результаты</u>

Решение банка об участии в аукционе

Здесь объясняемая переменная y_{it} равна 1, если i-й банк участвует в t-м аукционе, проводимом в течение рассматриваемого периода, и равна 0 в противном случае. В *табл. 1* приведены результаты оценивания панельной логит-модели

$$P\{y_{it}=1|\,x_{it}\}=\Lambda\left(x_{it}^{T}\beta\right),$$

в которую включены все указанные выше объясняющие переменные.

Таблица 1

	Random Effects		**Conditional FE** Coef. Estimate
	Coef. Estimate	Marginal Effect	
term spread	8.19 (5.16)	0.879	8.70 (5.48)
underbidding	0.36 (1.20)	-	0.30 (0.97)
spread	-11.57 (-6.82)	-1.242	-12.14 (-7.14)
volatility	-0.19 (-2.08)	-0.020	- 0.20 (-2.18)
reserve fulfillment	0.68 (0.55)	-	1.11 (0.90)
maturing allotment	0.13 (22.24)	0.014	0.11 (19.15)
end of period	0.74 (6.28)	0.060	0.76 (6.41)
Size dummies			
large	2.51 (1.57)		
medium	-8.98 (-3.43)		
small	-12.70 (-4.82)		
Pseudo-R^2	0.1142		0.1048
No. of observations	8525		4495
No. of groups	275		145

В скобках приведены значения *t*-статистик. Для статистически значимых оценок приведены значения предельных эффектов. При оценивании FE-модели методом условного максимального правдоподобия учитываются только те банки, которые принимают участие в аукционах хотя бы дважды. Псевдо-R^2 вычислен по формуле, приведенной ранее в главе 1, разд. 1.3.

Результаты в *табл. 1* приводятся и для RE- и для FE-моделей. Однако RE-модель предпочтительнее, поскольку критерий типа

Хаусмана не отвергает гипотезу о некоррелированности индивидуальных эффектов с объясняющими переменными.

Напомним (см. разд. 1.4. Главы 1), что поскольку логит-модель является нелинейной моделью, то оцененные коэффициенты имеют интерпретацию, отличающуюся от интерпретации коэффициентов в линейной модели. В связи с этим, в третьем столбце *табл. 1* приведены значения предельного эффекта для переменных со статистически значимыми оценками коэффициентов, вычисленные при средних значениях объясняющих переменных на рассмотренном периоде. Так, значение 0.060 предельного эффекта для дамми переменной *end of period* означает, что если аукцион проводится в конце периода между проверками выполнения требований в отношении резервов, то (при неизменных значениях остальных объясняющих переменных) шансы за то, что банк примет участие в аукционе, против того, что банк не примет участие в аукционе, возрастают в среднем приблизительно на 6%.

Таблица показывает, что ожидания в отношении процентных ставок имеют статистически значимое влияние на решение банка об участии в аукционе. В полном согласии с тем, что наблюдается на аукционах ЕЦБ, вероятность участия банка в РЕПО аукционах Бундесбанка уменьшается, когда отрицательное значение переменной *term spread* указывает на ожидание уменьшения процентных ставок.

В то же время статистическая незначимость оцененного коэффициента при переменной *underbidding* подчеркивает, что ожидаемый уровень отсечки на аукционе Бундесбанка не имеет определяющего влияния на решения банков об участии в аукционе. Оцененный коэфффициент при переменной *spread* имеет высокую статистическую значимость и отрицательное значение. Это означает, что если ожидаемая ставка РЕПО центрального банка выше ставки на рынке денег, то количество банков, принимающих решение об участии в аукционе РЕПО, сокращается.

Оцененный коэффициент при переменной *maturing allotment* имеет высокую статистическую значимость и положительное значение, подтверждая то, что банки используют РЕПО на возобновляемой основе.

Высокая статистическая значимость и положительность оцененного коэффициента при переменной *end of period* указывают на возрастание вероятности участия банка в аукционе, проводимом на последней неделе периода между проверками выполнения требований в отношении резервов.

Коэффициенты при дамми переменных, связанных с размерами банков, отражают тот очевидный факт, что большие банки участвуют в аукционах чаще, чем малые банки. Большие банки используют РЕПО аукционы не только для своей собственной потребности в ликвидности, но и для перепродажи и активной торговли резервами на вторичном рынке.

Влияние переменной *reserve fulfillment* не выявлено: оцененный коэффициент при этой переменной статистически незначим. Что касается переменной *volatility*, то для нее значение t -статистики лишь ненамного превосходит 5% критический уровень.

Для выяснения вопроса о том, зависит ли отклик банка на изменение того или иного фактора от размера банка, в правую часть уравнения добавляются взаимодействия факторов с размером банка, т.е. переменные, являющиеся произведениями объясняющих переменных на дамми, соответствующие возможным размерам банка. Результаты оценивания расширенной RE-модели представлены во втором столбце *табл. 2*.

В третьем столбце этой таблицы приведены P-значения статистик критерия для проверки гипотезы об отсутствии эффекта размера для отдельных объясняющих переменных.

Таблица 2

	Coef. Estimate	H_0: нет эффекта размера в группе (p-значение)	H_0: равны нулю все коэффициенты в группе (p-значение)
term spread			
large banks	20.36(0.98)		
medium banks	0.74 (0.31)	0.0002	0.0000
small banks	14.02 (6.45)		
underbidding			
large banks	0.19 (0.08)		
medium banks	0.12 (0.26)	0.7285	0.5392
small banks	0.60 (1.45)		
spread			
large banks	-20.71 (-0.92)		
medium banks	-4.65 (-1.84)	0.0018	0.0000
small banks	-16.81 (-7.17)		
volatility			
large banks	-1.24 (-1.06)		
medium banks	-0.13 (-0.96)	0.5788	0.1260
small banks	-0.24 (-1.92)		
reserve fulfillment			
large banks	1.60 (0.12)		
medium banks	-1.23 (-0.67)	0.3998	0.5626
small banks	2.17 (1.26)		
maturing allotment			
large banks	0.02 (0.40)		
medium banks	0.12 (13.67)	0.0187	0.0000
small banks	0.15 (17.27)		
end of period			
large banks	1.07 (0.71)		
medium banks	0.66 (3.59)	0.8521	0.0000
small banks	0.78 (5.00)		
Size dummies			
large	-6.23 (-0.40)		
medium	-1.75 (-0.79)	0.1123	0.0000
small	-8.12 (-3.89)		
Pseudo-R^2	0.1094		
No. of observations	8525		
No. of groups	275		

В скобках приведены значения *t*-статистик.

Отметим значимое влияние размера банка на отклик банка в отношении ожидаемых процентных ставок (*term spread*) и ожидаемой альтернативной стоимости (*spread*). Для обеих переменных влияние оказывается наиболее слабым для банков среднего размера. Коэффициенты при переменной *maturing allotment* показывают, что сезонный характер участия в аукционах более выражен для банков малого и среднего размера.

Что касается остальных объясняющих переменных, то здесь не обнаруживается значимого влияния размера банка.

В последнем столбце *табл. 2* приведены P-значения критериев совместной значимости для каждой группы переменных. Отметим, что результаты, полученные для расширенной логит- модели, очень похожи на результаты, полученные для модели без взаимодействий (*табл.1*). В частности, расширенная модель подтверждает сомнения в значимости волатильности для принятия банком решения об участии в аукционе.

Размер заявки отдельного банка

Исследуем теперь, каким образом перечисленные выше факторы влияют на размер заявки отдельного банка. Поскольку эту переменную можно наблюдать только если банк принимает решение об участии в аукционе, то она цензурирована слева (при неучастии банка в аукционе размер его заявки считаем равной нулю) и игнорирование этого обстоятельства может выразиться в смещении получаемых оценок. Соответственно, исследование здесь проводится с привлечением панельной тобит-модели.

Предельный эффект k-й объясняющей переменной в тобит-модели вычисляется по формуле

$$\frac{\partial E\{y_{it} \mid x_{it}\}}{\partial x_{k,it}} = \Phi\left(\frac{x_{it}^T \beta}{\sigma}\right)\beta_k .$$

В *табл. 3* приведены оцененные коэффициенты и вычисленные значения предельных эффектов.

Для сравнения в последнем столбце *табл. 3* приведены результаты GLS-оценивания RE-модели, которая пренебрегает информацией, содержащейся в нулевых заявках.

Таблица3

	Tobit Coef. Estimate	**Tobit** Marginal Effect	**Random Effects GLS** Coef. Estimate
term spread	11.59 (4.23)	1.00	1.43 (4.27)
underbidding	0.32 (0.61)	-	0.05 (1.97)
spread	-16.55 (-7.82)	-1.43	-1.94 (-5.40)
volatility	-0.24 (-1.50)	-	-0.01 (0.38)
reserve fulfillment	1.32 (0.61)	-	-0.67 (2.39)
maturing allotment	0.23 (20.41)	0.02	0.01 (7.39)
end of period	0.93 (4.61)	0.08	-0.08 (-1.13)
Size dummies			
large	3.95 (1.50)		-
medium	-8.98 (-3.43)		-3.46 (-5.10)
small	-12.70 (-4.82)		-4.80 (-7.14)
constant	-		21.01 (28.61)
Pseudo-R^2 *No. of observations* *No. of groups*	0.161 8525 275		0.062 2625 275

В скобках приведены значения *t*-статистик. Для статистически значимых оценок приведены значения предельных эффектов. При RE GLS-оценивании все 5900 цензурированных слева наблюдений отбрасывались.

Можно отметить определенное сходство оценок предельных эффектов в тобит-модели с оценками коэффициентов в указанной RE-модели.

Исследование не обнаруживает значимого влияния на размер заявок переменных *volatility* и *underbidding*, но обнаруживает значимое влияние переменных *term spread*, *spread*, *maturing allotment* и дамми-переменной *end of period*. Незначимость оцененного коэффициента при переменной *underbidding* показывает, что даже за неделю до ожидаемого понижения ставок Бундесбанк не имел трудностей с поставкой на РЕПО аукцион соответствующего объема резервов. Таким образом, в отличие от аукционов ЕЦБ, ожидаемые процентные ставки не препятствуют Бундесбанку в управлении денежным рынком.

В *табл. 4* приведены результаты оценивания расширенной тобит-модели, которая включает взаимодействия объясняющих переменных и дамми-переменных, относящихся к размеру банка.

Таблица 4

	Coef. Estimate	H_0: нет эффекта размера в группе (р-значение)	H_0: равны нулю все коэффициенты в группе (р-значение)
term spread			
large banks	5.19 (0.74)		
medium banks	1.82 (0.89)	0.0004	0.0000
small banks	12.91 (6.63)		
underbidding			
large banks	-0.70 (-0.55)		
medium banks	-0.01 (-0.04)	0.5847	0.7014
small banks	0.42 (1.06)		
spread			
large banks	-5.64 (-0.75)		
medium banks	-6.15 (-2.82)	0.0076	0.0000
small banks	-15.30 (-7.37)		
volatility			
large banks	-0.23 (-0.57)		
medium banks	-0.36 (-0.30)	0.2357	0.0428
small banks	-0.31 (-2.78)		
reserve fulfillment			
large banks	0.09 (0.02)		
medium banks	-0.57 (-0.36)	0.5056	0.6187
small banks	2.02 (1.29)		
maturing allotment			
large banks	0.03 (1.48)		
medium banks	0.13 (17.90)	0.0000	0.0000
small banks	0.20 (23.57)		
period end dummy			
large banks	0.09 (0.18)		

medium banks	0.59 (3.91)	0.5445	0.0000
small banks	0.67 (4.77)		
Size dummies			
large	4.78 (0.74)	0.0103	0.0001
medium	-1.99 (-1.04)		
small	-8.34 (-4.40)		
Pseudo-R^2	0.228		
No. of observations	8525		
No. of groups	275		

В скобках приведены значения *t*-статистик. При оценивании учитывались все 5900 цензурированных слева наблюдений.

Полученные оценки явно указывают на эффект размера банка в отношении влияния ожидаемых процентных ставок (*term spread*), ожидаемой альтернативной стоимости (*spread*) и *maturing allotment*.

Глава 4. Структурные и приведенные формы векторных авторегрессий и моделей коррекции ошибок

Здесь мы предполагаем, что читатель знаком с материалом, относящимся к построению и статистическому анализу моделей коррекции ошибок для коинтегрированных временных рядов, изложенным, например, в книге автора [Носко (2004)]. Кратко напомним необходимые для последующего изложения факты.

Пусть мы имеем N временных рядов $y_{1t},\ldots,y_{Nt}$,[1] каждый из которых является ***интегрированным порядка* 1**, так что, в общепринятых обозначениях, $y_{jt} \sim \mathrm{I}(1),\ j = 1,\ldots,N$. Если существует такой вектор $\beta = (\beta_1,\ldots,\beta_N)^T$, отличный от нулевого, для которого

$$\beta_1 y_{1t} + \ldots + \beta_N y_{Nt} \sim \mathrm{I}(0) - \text{ стационарный ряд,}$$

то ряды ***коинтегрированы (в узком смысле)***[2]; такой вектор β называется ***коинтегрирующим*** вектором. Если при этом

$$c = E\left(\beta_1 y_{1t} + \ldots + \beta_N y_{Nt}\right),$$

то тогда можно говорить о ***долговременном положении равновесия системы*** в виде

$$\beta_1 y_{1t} + \ldots + \beta_N y_{Nt} = c\,.$$

---

[1] Здесь в подстрочных индексах номер (момент) наблюдения стоит на втором месте – *после* номера ряда. Это отличается от системы обозначений, принятой в гл. 2, где номер наблюдения стоял на первом месте, а за ним следовал номер уравнения.

[2] Такое положение называют еще *детерминистской* коинтеграцией.

В каждый конкретный момент времени t существует некоторое ***отклонение*** системы от этого положения равновесия, характеризующееся величиной

$$z_t = \beta_1 y_{1t} + \ldots + \beta_N y_{Nt} - c .$$

Ряд z_t, в силу сделанных предположений, является стационарным рядом, имеющим нулевое математическое ожидание, так что он достаточно часто пересекает нулевой уровень, т.е. система колеблется вокруг указанного выше положения равновесия.

Положение, однако, осложняется тем, что у коинтегрированной системы I(1) рядов может быть несколько линейно независимых коинтегрирующих векторов. Если максимальное количество линейно независимых коинтегрирующих векторов для заданных рядов $y_{1t},\ldots,y_{Nt}$ равно r, то это число r называется ***рангом коинтеграции***. Для коинтегрированной системы, состоящей из N рядов, ранг коинтеграции может принимать значения $r = 1,\ldots,N-1$. (Формально, если ряды не коинтегрированы, то $r = 0$. Если же имеется $r = N$ линейно независимых коинтегрирующих векторов, то все N рядов стационарны.) Совокупность всех возможных коинтегрирующих векторов для коинтегрированной системы I(1) рядов образует r-мерное линейное векторное пространство, которое называют ***коинтеграционным пространством***. Любой набор r линейно независимых коинтегрирующих векторов образует ***базис*** этого пространства, и если зафиксировать этот набор в качестве базиса, то тогда любой коинтегрирующий вектор является линейной комбинацией векторов, составляющих базис.

Пусть коинтегрированная система I(1) рядов $y_{1t},\ldots,y_{Nt}$ имеет ранг коинтеграции r и может быть представлена в форме VAR($p+1$) – ***векторной авторегрессии*** порядка $p+1$ (VAR – *vector autoregression*). Тогда существует представление этой VAR в форме ECM (модели коррекции ошибок)

$$\Delta y_{1t} = \mu_1 + \alpha_{11} z_{1,t-1} + \ldots + \alpha_{1r} z_{r,t-1} +$$
$$+ \sum_{j=1}^{p} \left(\gamma_{11,j} \Delta y_{1,t-j} + \ldots + \gamma_{1N,j} \Delta y_{N,t-j} \right) + \varepsilon_{1t} ,$$

……………………………………..

$$\Delta y_{Nt} = \mu_N + \alpha_{N1} z_{1,t-1} + \ldots + \alpha_{Nr} z_{r,t-1} +$$
$$+ \sum_{j=1}^{p} \left(\gamma_{N1,j} \Delta y_{1,t-j} + \ldots + \gamma_{NN,j} \Delta y_{N,t-j} \right) + \varepsilon_{Nt} ,$$

где

$z_{1t}, \ldots, z_{rt}$ – стационарные I(0) ряды, соответствующие r линейно независимым коинтегрирующим векторам $\beta_{(1)}, \ldots, \beta_{(r)}$,

$(\alpha_{11}, \ldots, \alpha_{N1})^T, \ldots, (\alpha_{1r}, \ldots, \alpha_{Nr})^T$ – линейно независимые векторы корректирующих коэффициентов (коэффициентов адаптации).

Такую модель коррекции ошибок можно записать в компактном виде как

$$\Delta y_t = \mu + \alpha \beta^T y_{t-1} + \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_p \Delta y_{t-p} + \varepsilon_t$$

где $\Gamma_1, \ldots, \Gamma_p$ – матрицы размера $N \times N$, а α и β – $(N \times r)$-матрицы полного ранга r. При этом столбцы $\beta_{(1)}, \ldots, \beta_{(r)}$ матрицы β являются линейно независимыми коинтегрирующими векторами, а элементы α_{ij} матрицы α являются коэффициентами при стационарных линейных комбинациях

$$z_{1,t-1} = \beta_{(1)}^T y_{t-1}, \ldots, z_{r,t-1} = \beta_{(r)}^T y_{t-1}$$

(представляющих отклонения в момент $t-1$ от r долговременных соотношений между рядами $y_{1t}, \ldots, y_{Nt}$) в правых частях уравнений для $\Delta y_{1t}, \ldots, \Delta y_{Nt}$. Мы будем говорить о такой модели коррекции ошибок как о ***модели коррекции ошибок без ограничений*** (***UECM* – *unrestricted error correction model***).

Представление коинтегрированной VAR в форме модели коррекции ошибок не единственно, поскольку в качестве набора $\beta_{(1)},\ldots,\beta_{(r)}$ можно взять любой базис коинтеграционного пространства. Соответственно, неоднозначность имеется и в отношении матрицы α. В практических задачах на первый план (наряду с определением ранга коинтеграции) выходит ***идентификация коинтегрирующих векторов***, выражающих осмысленные с экономической точки зрения (экономической теории) долговременные связи между рассматриваемыми переменными (например, паритет покупательной способности, спрос на деньги и т.п.). Это, в свою очередь, требует наложения на коинтегрирующие векторы соответствующих ***идентифицирующих ограничений***, позволяющих различать эти векторы, выделяя их из всего множества линейных комбинаций базисных векторов.

Заметим теперь, что в правую часть уравнений стандартной формы VAR включаются только запаздывающие значения переменных $y_{1t},\ldots,y_{Nt}$. Поэтому в правой части ECM, соответствующей такой VAR, оказываются только запаздывающие значения приращений $\Delta y_{1t},\ldots,\Delta y_{Nt}$. Между тем в практических исследованиях при рассмотрении оцененной корреляционной матрицы вектора приращений $(\Delta y_{1t},\ldots,\Delta y_{Nt})$ часто наблюдаются достаточно удаленные от нуля значения недиагональных элементов этой матрицы. Последнее указывает на возможную коррелированность приращений, соответствующих одному и тому же моменту времени. Непосредственно учесть такого рода коррелированность можно, переходя к модели ***структурной VAR*** (***SVAR – ***_structural vector autoregression_) и соответствующей ей модели ***структурной ECM*** (***SECM – ***_structural error correction model_).

Рассмотрим пару рядов y_{1t}, y_{2t}, составляющих векторный случайный процесс $y_t = (y_{1t}, y_{2t})^T$, порождающийся структурной ECM:

$$\Delta y_{1t} = \varphi_{12}\Delta y_{2t} + \gamma_{11}^{*}\Delta y_{1,t-1} + \gamma_{12}^{*}\Delta y_{2,t-1} +$$
$$+ a_{11}\left(\beta_{11}\Delta y_{1,t-1} + \beta_{21}\Delta y_{2,t-1}\right) + a_{12}\left(\beta_{12}\Delta y_{1,t-1} + \beta_{22}\Delta y_{2,t-1}\right) + \zeta_{1t},$$
$$\Delta y_{2t} = \varphi_{21}\Delta y_{1t} + \gamma_{21}^{*}\Delta y_{1,t-1} + \gamma_{22}^{*}\Delta y_{2,t-1} +$$
$$+ a_{21}\left(\beta_{11}\Delta y_{1,t-1} + \beta_{21}\Delta y_{2,t-1}\right) + a_{22}\left(\beta_{12}\Delta y_{1,t-1} + \beta_{22}\Delta y_{2,t-1}\right) + \zeta_{2t},$$

Обозначая

$$\beta = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix},\ \Gamma_1^{*} = \begin{pmatrix} \gamma_{11}^{*} & \gamma_{12}^{*} \\ \gamma_{21}^{*} & \gamma_{22}^{*} \end{pmatrix},\ a = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

$$\Phi = \begin{pmatrix} 1 & -\varphi_{12} \\ -\varphi_{21} & 1 \end{pmatrix},\ \zeta_t = \begin{pmatrix} \zeta_{1t} \\ \zeta_{2t} \end{pmatrix},$$

получаем для этой структурной ЕСМ компактное выражение:

$$\Phi\Delta y_t = \Gamma_1^{*}\Delta y_{t-1} + a\beta^T y_{t-1} + \zeta_t .$$

Умножая обе части последнего выражения слева на матрицу Φ^{-1}, находим ***приведенную форму*** ЕСМ:

$$\Delta y_t = \Gamma_1\Delta y_{t-1} + \alpha\beta^T y_{t-1} + \varepsilon_t ,$$

где

$$\Gamma_1 = \Phi^{-1}\Gamma_1^{*},\ \alpha = \Phi^{-1}a,\ \varepsilon_t = \Phi^{-1}\zeta_t .$$

Но

$$\Phi^{-1} = \frac{1}{\left(1-\varphi_{12}\varphi_{21}\right)}\begin{pmatrix} 1 & \varphi_{12} \\ \varphi_{21} & 1 \end{pmatrix},$$

так что, обозначая, $\delta = 1-\varphi_{12}\varphi_{21}$, получаем:

$$\gamma_{11} = \left(\gamma_{11}^{*} + \varphi_{12}\gamma_{21}^{*}\right)/\delta,$$
$$\gamma_{12} = \left(\gamma_{12}^{*} + \varphi_{12}\gamma_{22}^{*}\right)/\delta,$$
$$\gamma_{21} = \left(\gamma_{21}^{*} + \varphi_{21}\gamma_{11}^{*}\right)/\delta,$$
$$\gamma_{22} = \left(\gamma_{22}^{*} + \varphi_{21}\gamma_{12}^{*}\right)/\delta,$$

$$\alpha = \frac{1}{\delta}\begin{pmatrix} 1 & \varphi_{12} \\ \varphi_{21} & 1 \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

$$\alpha_{11} = (a_{11} + \varphi_{12}a_{21})/\delta,\ \alpha_{12} = (a_{12} + \varphi_{12}a_{22})/\delta,$$

$$\alpha_{21} = (a_{21} + \varphi_{21}a_{11})/\delta,\ \alpha_{22} = (a_{22} + \varphi_{21}a_{11})/\delta.$$

Рассматриваемая структурная ECM не имеет ограничений в том смысле, что в ней не приравниваются нулю:

- никакие внедиагональные элементы матрицы Φ^{-1};
- никакие элементы матрицы Γ_1^*;
- никакие элементы вектора a;
- никакие элементы вектора β.

В конкретных же примерах некоторые из перечисленных элементов зануляются. Например, если $y_{1t}, y_{2t} \sim$ I(1) и коинтегрированы, то коинтегрирующий вектор единствен. Если это вектор $(\beta_{11}, \beta_{21})^T$, то тогда можно положить $\beta_{12} = \beta_{22} = 0,\ \ \alpha_{12} = \alpha_{22} = 0,$ что уменьшает количество неизвестных коэффициентов.

В общем случае структурная ECM имеет вид:

$$\Phi\Delta y_t = \Gamma_1^*\Delta y_{t-1} + \ldots + \Gamma_p^*\Delta y_{t-p} + a\beta^T y_{t-1} + \zeta_t,$$

где Φ – недиагональная невырожденная квадратная матрица размера $N \times N$. Умножая здесь обе части уравнения на Φ^{-1}, мы приходим к приведенной форме ECM:

$$\Delta y_t = \Gamma_1\Delta y_{t-1} + \ldots + \Gamma_p\Delta y_{t-p} + \alpha\beta^T y_{t-1} + \varepsilon_t,$$

где

$$\Gamma_j = \Phi^{-1}\Gamma_j^*,\ \alpha = \Phi^{-1}a,\ \varepsilon_t = \Phi^{-1}\zeta_t.$$

В этих двух формах общим является только $\beta^T y_{t-1}$, т.е. долговременное соотношение, тогда как коэффициенты адаптации (в приведенной форме ECM) являются функциями от коэффициентов структурной ECM: $\alpha = \Phi^{-1}a.$ Последнее приводит к тому, что отсутствие некоторой корректирующей составляющей в одном из уравнений SECM отнюдь не означает, что эта составляющая будет отсутствовать и в соответствующем уравнении приведенной ECM. Соответственно, коррекция ошибок в одном уравнении SECM может распространяться и на все остальные уравнения приведенной ECM.

При рассмотрении вопроса об идентифицируемости параметров структурной ECM естественно выделить отдельно идентификацию коинтегрирующих векторов и идентификацию коэффициентов, связанных с динамической адаптацией, т.е. элементов матриц $\Phi, \Gamma_1^*, \ldots, \Gamma_p^*, a$. Поскольку коинтегрирующие соотношения в структурной и приведенной ECM одни и те же, можно использовать результаты, касающиеся идентифицируемости коинтегрирующих векторов в приведенной ECM.

Вопрос об идентификации коинтегрирующих векторов естественно возникает при наличии двух и более коинтегрирующих векторов и связан с возможностью различения таких векторов. В процедуре Йохансена, реализованной в пакете EVIEWS, такое различение производится исходя из того, что если ранг коинтеграции равен $1 < r < N$, то тогда существует $N - r$ некоинтегрированных между собой (в совокупности) переменных – “общих трендов” (common trends), таких, что добавление к этой совокупности любой из оставшихся r переменных приводит к коинтегрированности пополненного множества из $N - r + 1$ переменных. Это означает, что в качестве линейно независимых коинтегрирующих векторов можно взять любой набор из r векторов вида

$$\begin{pmatrix} \beta_{11} \\ 0 \\ \vdots \\ 0 \\ \beta_{1,r+1} \\ \vdots \\ \beta_{1k} \end{pmatrix}, \begin{pmatrix} 0 \\ \beta_{22} \\ \vdots \\ 0 \\ \beta_{2,r+1} \\ \vdots \\ \beta_{2k} \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \beta_{rr} \\ \beta_{r,r+1} \\ \vdots \\ \beta_{rk} \end{pmatrix}.$$

Конечно, при этом предполагается, что переменные занумерованы так, что “общие тренды” соответствуют переменным с номерами $r+1,\dots,N$. Выделение из этого множества возможных наборов единственного набора производится в EVIEWS нормализацией каждого из этих векторов, так что j-й вектор нормализуется делением всех его элементов на β_{jj}, вследствие чего получаем набор из r векторов:

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ \beta^*_{1,r+1} \\ \vdots \\ \beta^*_{1N} \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ \beta^*_{2,r+1} \\ \vdots \\ \beta^*_{2N} \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \beta^*_{r,r+1} \\ \vdots \\ \beta^*_{r,N} \end{pmatrix}.$$

Такой набор образует базис r-мерного линейного пространства коинтегрирующих векторов при ранге коинтеграции r.

Проблема, однако, в том, что с точки зрения экономической теории нас могут интересовать коинтегрирующие векторы и какого-то другого вида (являющиеся, конечно, линейными комбинациями векторов, принадлежащих базису). При этом на соответствующие коинтегрирующие векторы накладываются определенные ограничения, вытекающие из экономической теории: невхождение в коинтегрирующую линейную комбинацию тех или иных переменных, равенство некоторых компонент коинтегрирующего вектора или наличие у них противоположных знаков при одинаковой абсолютной величине и т.п. В такой ситуации возникает вопрос о необходимости и достаточности множества накладываемых ограничений для идентификации, т.е. различения коинтегрирующих векторов.

Обычно ограничиваются рассмотрением линейных ограничений, в том числе исключающих появление отдельных переменных в коинтегрирующей линейной комбинации. При этом ограничения могут быть представлены как в явной, так и в неявной форме. Если векторы уже нормализованы, то тогда ***необходимым условием идентифицируемости** r **коинтегрирующих векторов*** является наложение на каждый из r векторов не менее $r-1$ линейных ограничений. Об этом условии говорят как о ***порядковом условии идентифицируемости***.

Порядковое условие, вообще говоря, не является достаточным для идентифицируемости, поскольку при его выполнении полученные r векторов могут все же оказаться линейно зависимыми, так что, скажем, вектор β_1 нельзя отличить от некоторой линейной комбинации векторов $\beta_2,\ldots,\beta_r$. Поэтому, в принципе, следует производить еще и проверку линейной независимости полученных r векторов. Для этого можно воспользоваться достаточными условиями идентифицируемости, формулируемыми в терминах матриц, участвующих в формировании явной и неявной форм линейных ограничений.

Если на i-й коинтегрирующий вектор накладывается r_i линейных ограничений, то их можно записать в двух формах: явной и неявной. Под ***неявной формой*** понимается представление этих ограничений в виде:

$$R_i\beta_i = 0,$$

где R_i – матрица размера $r_i \times N$ ранга r_i. Ту же самую совокупность ограничений можно представить в ***явной форме*** в виде:

$$\beta_i = H_i \vartheta_i,$$

где H_i – матрица размера $N \times (N - r_i)$ ранга $N - r_i$, ϑ_i – вектор размера $(N - r_i) \times 1$. При этом выполняется соотношение

$$R_i H_i = 0,$$

т.е. строки матрицы R_i ортогональны столбцам матрицы H_i. Поясним это на примере модели IS/LM, связывающей следующие макроэкономические параметры:

$m_t = \ln M_t$, где M_t – номинальная денежная масса,

$inc_t = GDP_t$, где GDP_t – реальный валовый внутренний продукт,

$p_t = \ln P_t$, где P_t – дефлятор GDP,

r_t^s – краткосрочная процентная ставка,

r_t^b – долгосрочная процентная ставка.

Пусть все эти ряды интегрированные порядка 1, ранг коинтеграции этих временных рядов равен $r = 3$ и в коинтеграционное соотношение приходится включать еще и временной тренд t. Тогда речь идет об идентификации трех коинтегрирующих векторов

$$\beta_1 = \begin{pmatrix} \beta_{11} \\ \beta_{21} \\ \beta_{31} \\ \beta_{41} \\ \beta_{51} \\ \beta_{61} \end{pmatrix}, \ \beta_2 = \begin{pmatrix} \beta_{12} \\ \beta_{22} \\ \beta_{32} \\ \beta_{42} \\ \beta_{52} \\ \beta_{62} \end{pmatrix}, \ \beta_3 = \begin{pmatrix} \beta_{13} \\ \beta_{23} \\ \beta_{33} \\ \beta_{43} \\ \beta_{53} \\ \beta_{63} \end{pmatrix},$$

обеспечивающих стационарность линейных комбинаций

$$z_{1t} = \beta_{11} m_t + \beta_{21} inc_t + \beta_{31} p_t + \beta_{41} r_t^s + \beta_{51} r_t^b + \beta_{61} t,$$

$$z_{2t} = \beta_{12} m_t + \beta_{22} inc_t + \beta_{32} p_t + \beta_{42} r_t^s + \beta_{52} r_t^b + \beta_{62} t,$$

$$z_{3t} = \beta_{13} m_t + \beta_{23} inc_t + \beta_{33} p_t + \beta_{43} r_t^s + \beta_{53} r_t^b + \beta_{63} t.$$

Ограничения на коэффициенты этих векторов могут быть получены, например, из следующих соображений.

Если спрос на реальные деньги рассматривается как функция от реального дохода, краткосрочной ставки и тренда, т.е. $m_t - p_t = f_1(inc_t, r_t^s, t)$, то это подразумевает наличие долгосрочной связи между переменными $m_t - p_t$, inc_t, r_t^s, t без включения в нее переменной r_t^b, так что стационарной является линейная комбинация $z_{1t} = \beta_{11} m_t + \beta_{21} inc_t - \beta_{11} p_t + \beta_{41} r_t^s + \beta_{61} t.$ Ограничения на вектор β_1 принимают вид: $\beta_{31} = -\beta_{11}$, $\beta_{51} = 0.$ Эти ограничения можно записать в виде $R_1 \beta_1 = 0$ (неявная форма), где

$$R_1 = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

или в виде $\beta_1 = H_1 \vartheta_1$, где

$$H_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \ \vartheta_1 = \begin{pmatrix} \vartheta_{11} \\ \vartheta_{21} \\ \vartheta_{31} \\ \vartheta_{41} \end{pmatrix}, \ \text{так что} \ \ \beta_1 = \begin{pmatrix} \vartheta_{11} \\ \vartheta_{21} \\ -\vartheta_{11} \\ \vartheta_{31} \\ 0 \\ \vartheta_{41} \end{pmatrix}.$$

Нетрудно проверить, что $R_1 H_1 = 0.$

Если дифференциал процентных ставок $r_t^s - r_t^b$ определяется через остальные переменные без включения тренда, т.е. $r_t^s - r_t^b = f_1(m_t, inc_t, p_t)$, то это подразумевает наличие долговременной связи между переменными $r_t^s - r_t^b, m_t, inc_t, p_t$ без включения в нее переменной t, так что стационарной является линейная комбинация $z_{2t} = \beta_{12} m_t + \beta_{22} inc_t + \beta_{32} p_t + \beta_{42} r_t^s - \beta_{42} r_t^b.$ Ограничения на вектор β_2 принимают вид: $\beta_{52} = -\beta_{42}$, $\beta_{62} = 0.$ Эти ограничения можно записать в виде $R_2 \beta_2 = 0$ (неявная форма), где

$$R_2 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

или в виде $\beta_2 = H_2 \vartheta_2$, где

$$H_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad \vartheta_2 = \begin{pmatrix} \vartheta_{12} \\ \vartheta_{22} \\ \vartheta_{32} \\ \vartheta_{42} \end{pmatrix}, \text{ так что } \beta_2 = \begin{pmatrix} \vartheta_{12} \\ \vartheta_{22} \\ \vartheta_{32} \\ \vartheta_{42} \\ -\vartheta_{42} \\ 0 \end{pmatrix}.$$

Наконец, если долгосрочная процентная ставка r_t^b определяется как функция только от m_t, p_t и t, то это подразумевает наличие долговременной связи между переменными r_t^b, m_t, p_t и t без включения в нее переменных inc_t и r_t^s, так что стационарной является линейная комбинация $z_{3t} = \beta_{13} m_t + \beta_{33} p_t + \beta_{53} r_t^b + \beta_{63} t.$ Ограничения на вектор β_3 принимают вид: $\beta_{23} = 0,\ \beta_{43} = 0.$ Эти ограничения можно записать в виде $R_3 \beta_3 = 0$ (неявная форма), где

$$R_3 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix},$$

или в виде $\beta_3 = H_3 \vartheta_3$, где

$$H_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \vartheta_3 = \begin{pmatrix} \vartheta_{13} \\ \vartheta_{23} \\ \vartheta_{33} \\ \vartheta_{43} \end{pmatrix}, \text{ так что } \beta_3 = \begin{pmatrix} \vartheta_{13} \\ 0 \\ \vartheta_{23} \\ 0 \\ \vartheta_{33} \\ \vartheta_{43} \end{pmatrix}.$$

Необходимое и достаточное условие идентифицируемости $1 < r < N$ коинтегрирующих векторов имеет в общем случае довольно сложный вид. Однако при $r = 2,3$ оно достаточно просто для проверки (см., например, [Patterson (2000)]).

При $r = 2$ коинтегрирующие векторы β_1, β_2 идентифицируемы тогда и только тогда, когда выполняются соотношения:

$\operatorname{rank}(R_1 H_2) \geq 1$, $\operatorname{rank}(R_2 H_1) \geq 1$.

При $r = 3$ коинтегрирующие векторы $\beta_1, \beta_2, \beta_3$ идентифицируемы тогда и только тогда, когда выполняются соотношения:

$\operatorname{rank}(R_i H_j) \geq 1,\ i \neq j,\ i, j = 1,2,3,$ (6 соотношений)

$\operatorname{rank}(R_1[H_2, H_3]) \geq 2,\ \operatorname{rank}(R_2[H_1, H_3]) \geq 2,\ \operatorname{rank}(R_3[H_1, H_2]) \geq 2.$

Проверим выполнение этого условия в только что рассмотренном примере, где $r = 3$. Имеем:

$$R_1 H_2 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}, \quad R_1 H_3 = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

$$R_2 H_1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad R_2 H_3 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$R_3 H_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad R_3 H_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Все 6 матриц имеют ранг 2>1, так что первая группа условий выполняется. Далее,

$$R_1[H_2,H_3]=\begin{pmatrix}1&0&1&0&0&0\\0&0&0&0&1&0\end{pmatrix}\begin{pmatrix}1&0&0&0&1&0&0&0\\0&1&0&0&0&0&0&0\\0&0&1&0&0&1&0&0\\0&0&0&1&0&0&0&0\\0&0&0&-1&0&0&1&0\\0&0&0&0&0&0&0&1\end{pmatrix}=$$

$$=\begin{pmatrix}1&0&1&0&1&1&0&0\\0&0&0&-1&0&0&1&0\end{pmatrix},$$

$$R_2[H_1,H_3]=\begin{pmatrix}0&0&0&1&1&0\\0&0&0&0&0&1\end{pmatrix}\begin{pmatrix}1&0&0&0&1&0&0&0\\0&1&0&0&0&0&0&0\\-1&0&0&0&0&1&0&0\\0&0&1&0&0&0&0&0\\0&0&0&0&0&0&1&0\\0&0&0&1&0&0&0&1\end{pmatrix}=$$

$$=\begin{pmatrix}0&0&1&0&0&0&1&0\\0&0&0&1&0&0&0&1\end{pmatrix},$$

$$R_3[H_1,H_2]=\begin{pmatrix}0&1&0&0&0&0\\0&0&0&1&0&0\end{pmatrix}\begin{pmatrix}1&0&0&0&1&0&0&0\\0&1&0&0&0&1&0&0\\-1&0&0&0&0&0&1&0\\0&0&1&0&0&0&0&1\\0&0&0&0&0&0&0&-1\\0&0&0&1&0&0&0&0\end{pmatrix}=$$

$$=\begin{pmatrix}0&1&0&0&0&1&0&0\\0&0&1&0&0&0&0&1\end{pmatrix}.$$

Ранги всех трех матриц равны 2, так что и эта группа условий выполнена. Таким образом, коинтегрирующие векторы $\beta_1, \beta_2, \beta_3$ идентифицируемы.

Если r коинтегрирующих векторов идентифицируемы, то на каждый их них накладывается не менее $r-1$ линейных ограничений. В случае, когда на каждый из этих векторов накладывается ровно $r-1$ ограничений, мы имеем дело с ***точной идентифицируемостью***. Если же хотя бы для одного из векторов количество ограничений превышает $r-1$, то мы имеем дело со ***сверхидентифицируемостью***. В последнем случае имеются “лишние” ограничения и имеется возможность проверки гипотезы о том, что заявленные дополнительные ограничения на векторы $\beta_1, \beta_2, \ldots, \beta_r$ действительно выполняются.

В рассмотренном примере можно, например, следуя работе [Johansen, Juselius (1994)], наложить следующие более строгие ограничения на векторы $\beta_1, \beta_2, \beta_3$.

Вектор β_1 нормализуется на inc_t, коэффициенты при m_t и p_t равны по абсолютной величине и противоположны по знаку, а коэффициенты при обеих процентных ставках равны нулю. Соответствующая равновесная связь интерпретируется как прокси для совокупного дохода относительно линейного тренда. Это дает матрицу

$$R_1 = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

У вектора β_2 равны нулю все коэффициенты кроме коэффициентов при r_t^s и r_t^b, которые равны по абсолютной величине и противоположны по знаку; вектор нормализуется на один из них. Интерпретация: стационарность дифференциала процентных ставок. Матрица R_2 принимает вид:

$$R_2 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Наконец, у вектора β_3 равны нулю коэффициенты при m_t, inc_t и r_t^s; вектор нормализуется на r_t^b. Интерпретация: долгосрочная процентная ставка определяется как функция от цены и временного тренда. Матрица R_3 имеет вид:

$$R_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}.$$

В итоге мы получаем следующую картину:

Вектор	Кол-во ограничений	Кол-во “лишних” ограничений
β_1	3	1
β_2	5	3
β_3	3	1
Всего	11	5

Проверка гипотезы H_0 о выполнении совокупности всех сверхидентифицирующих ограничений производится с использованием асимптотического критерия хи-квадрат, указанного в работе [Johansen, Juselius (1994)]. При справедливости гипотезы H_0 статистика этого критерия имеет асимптотическое распределение хи-квадрат с q степенями свободы, где q –

суммарное количество дополнительных ограничений. Гипотеза H_0 отвергается при слишком больших значениях этой статистики. Однако неотвержение гипотезы H_0 отнюдь не означает, что именно указанную в гипотезе совокупность сверхидентифицирующих ограничений следует использовать, поскольку вообще-то можно сформировать не один, а несколько различных наборов сверхидентифицирующих ограничений, которые также могут оказаться неотвергнутыми. Приемлемость некоторого конкретного множества ограничений должна подкрепляться еще и другими соображениями, среди которых можно выделить следующие два:

- являются ли осмысленными с точки зрения экономической теории числовые значения оценок коэффициентов, получаемых при выбранных ограничениях;
- являются ли осмысленными с точки зрения экономической теории коэффициенты ***адаптации*** α_{ij} .

В нашем примере наблюдаемое значение статистики этого критерия равно 3.5; асимптотическое критическое значение, соответствующее уровню значимости 0.05, равно $\chi^2_{0.95}(5)=11.07$, так что нулевая гипотеза о выполнении всех 5 сверхидентифицирующих ограничений не отвергается.

После уточнения идентифицирующих ограничений, которые накладываются на коинтегрирующие векторы, производится оценивание параметров приведенной формы ECM

$$\Delta y_t = \alpha\beta^T y_{t-1} + \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_p \Delta y_{t-p} + \varepsilon_t,$$

в результате чего находятся, в частности, и оценки коинтегрирующих векторов $\hat\beta_1, \hat\beta_2, \ldots, \hat\beta_r$. Эти оценки подставляются в уравнение структурной ECM

$$\Phi\Delta y_t = \Gamma_1^*\Delta y_{t-1} + \ldots + \Gamma_p^*\Delta y_{t-p} + a\beta^T y_{t-1} + \zeta_t$$

вместо неизвестных “истинных” коинтегрирующих векторов $\beta_1, \beta_2, \ldots, \beta_r$. Возможность получения состоятельных оценок для параметров $\Phi, \Gamma_1^*, \ldots, \Gamma_p^*, a$ структурной формы связана с идентифицируемостью этой системы; ее можно записать также виде

$$\left(\Phi, -\Gamma_1^*, \ldots, -\Gamma_p^*, -a\right)\begin{pmatrix} \Delta y_t \\ \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p} \\ \beta^T y_{t-1} \end{pmatrix} = \zeta_t,$$

или

$$A^T Z_t = \zeta_t,$$

где

$A^T = \left(\Phi, -\Gamma_1^*, \ldots, -\Gamma_p^*, -a\right)$ – матрица размера $N \times \left(N + pN + rN\right)$,

$Z_t = \begin{pmatrix} \Delta y_t \\ \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p} \\ \beta^T y_{t-1} \end{pmatrix}$ – $\left(N + pN + rN\right) \times 1$-вектор стационарных переменных.

Каждая строка матрицы A^T, т.е. каждый столбец матрицы A, относится к отдельному уравнению. Соответственно, для идентифицируемости i-го уравнения необходимо наложить на i-й столбец A_i матрицы A не менее $N - 1$ ограничений в виде

$R_i A_i = 0$ (неявная форма)

или

$A_i = H_i \vartheta_i$ (явная форма), $R_i H_i = 0$.

Как и в главе 2, гарантией идентифицируемости i-го структурного уравнения служит выполнение рангового условия идентифицируемости; само оценивание должно проводиться системными методами (например, FIML). На основании оцененной структурной ECM (SECM) строится приведенная форма – ***ECM с ограничениями***, соответствующая этой SECM.

Проиллюстрируем процесс построения ECM с ограничениями примером для модели IS/LM, который мы уже начали рассматривать ранее. В работе [Johansen, Juselius (1994)] такое построение проводилось по статистическим данным для Австралии (квартальные данные, период с III кв. 1975 г. по I кв. 1991 г.).

Оценивание ECM с учетом указанных выше 11 линейных ограничений на коинтегрирующие векторы и того, что по результатам предварительного анализа модель VAR в уровнях имеет порядок 2, приводит к следующим оценкам для коинтегрирующих векторов β_i и векторов коэффициентов адаптации α_i :

Переменная	β_1	β_2	β_3
m_t	– 0.193	0	0
inc_t	1	0	0
p_t	0.193	0	–0.488
r_t^s	0	1	0
r_t^b	0	–1	1
t	– 0.005	0	0.009

Уравнение для	α		
Δm_t	0.030	0.159	–0.569
Δinc_t	–0.458	– 0.001	0.405
Δp_t	0.325	– 0.039	0.054
Δr_t^s	0.337	– 0.008	–0.168
Δr_t^b	0.109	0.023	–0.213

Для того чтобы коинтегрирующие линейные комбинации флуктуировали вокруг нулевого уровня, к ним были добавлены константа и дамми-переменная $D84_t$, равная 1 в период с I кв. 1984 г. по I кв. 1991 г., и равная нулю на остальной части периода наблюдений, отражающая изменение правил регулирования в банковском секторе. В результате в правых частях уравнений ECM вместо переменных $z_{1,t-1}$, $z_{2,t-1}$, $z_{3,t-1}$ используются переменные $ecm1_{t-1}$, $ecm2_{t-1}$, $ecm3_{t-1}$, где

$$ecm1_t = inc_t - 0.193(m_t - p_t) - 0.005t - 0.027 \cdot D84_t - 8.43,$$
$$ecm2_t = (r_t^s - r_t^b) + 0.00967 \cdot D84_t + 0.03,$$
$$ecm3_t = r_t^b - 0.488(p_t - 0.019t) - 0.008 \cdot D84_t - 0.52\,.$$

Как мы уже говорили ранее, необходимость в построении структурной ECM возникает из-за наличия коррелированности между переменными, стоящими в левых частях ECM. В рассматриваемом примере в левых частях ECM находятся переменные Δm_t, Δinc_t, Δp_t, Δr_t^s и Δr_t^b. Следовательно, надо выяснить, имеется ли между ними заметная корреляция. Вычисленные выборочные коэффициенты корреляции между этими переменными приведены в следующей таблице:

	Δm_t	Δinc_t	Δp_t	Δr_t^s	Δr_t^b
Δm_t	1	0.29	0.20	–0.10	–0.10
Δinc_t	0.29	1	0.35	–0.18	0.00
Δp_t	0.20	0.35	1	–0.12	–0.10
Δr_t^s	–0.10	–0.18	–0.12	1	0.65
Δr_t^b	–0.10	0.00	–0.10	0.65	1

Ориентируясь на эту таблицу, Johansen и Juselius делают вывод о наличии коррелированности переменных, что требует системного оценивания. Если исходить из того, что модель VAR в уровнях имеет порядок 2, то в правую часть ECM могут входить значения приращений переменных, запаздывающие не более, чем на один шаг. Соответственно, в правых частях уравнений структурной ECM потенциально могут присутствовать следующие переменные: Δm_t, $\Delta inc_t, \Delta p_t, \Delta r_t^s, \Delta r_t^b, \Delta m_{t-1}, \Delta inc_{t-1}, \Delta p_{t-1}, \Delta r_{t-1}^s, \Delta r_{t-1}^b, ecm1_{t-1}, ecm2_{t-1}$, $ecm3_{t-1}$, так что в каждом из 5 уравнений (для Δm_t, $\Delta inc_t, \Delta p_t, \Delta r_t^s$ и Δr_t^b) потенциально может быть 12 коэффициентов, подлежащих оцениванию. Однако при таком количестве неизвестных (неспецифицированных) коэффициентов уравнения не могут быть идентифицируемыми. Для их идентифицируемости необходимо наложить на коэффициенты каждого уравнения как минимум $5-1=4$ ограничения. И здесь, в отличие от выбора идентифицирующих ограничений на коинтегрирующие векторы, ориентирующегося главным образом на экономические представления, приходится опираться по большей части на статистическую информацию, содержащуюся в самих данных, т.е. на эмпирическую картину адаптации, а не на строгое априорное обусловливание.

Первоначально Johansen и Juselius берут ровно по 4 ограничения на каждое уравнение, что обеспечивает точную идентифицируемость системы:

Уравнение	Зануляются коэффициенты при переменных:			
Δm_t	Δr_t^s	Δinc_t	Δr_t^b	Δinc_{t-1}
Δinc_t	Δr_t^s	Δp_t	Δr_t^b	Δr_{t-1}^b
Δp_t	Δr_t^b	Δp_{t-1}	Δr_{t-1}^s	Δr_{t-1}^b
Δr_t^s	Δr_{t-1}^s	Δinc_{t-1}	Δp_{t-1}	Δr_{t-1}^b
Δr_t^b	Δinc_t	Δm_t	Δp_t	Δr_{t-1}^b

Однако в оцененной с такими ограничениями модели оказались статистически незначимыми (по *t*-статистикам) почти все оцененные коэффициенты. В связи с этим производились эксперименты с различными выборами ограничений, и в итоге Johansen и Juselius пришли к разбиению системы 5 уравнений на два блока, один из которых содержит уравнения для $\Delta m_t, \Delta inc_t, \Delta r_t^s$, а другой – уравнения для Δp_t и Δr_t^b. При этом первый блок носит системный характер, а второй – характер приведенной формы, т.е. в правых частях уравнений первого блока имеются эндогенные переменные, порождаемые в рамках системы трех уравнений, а в рамках второго блока – нет. Специфицированные коэффициенты уравнений приведены в следующей таблице:

	Δm_t	Δinc_t	Δr_t^s	Δp_t	Δr^b_t	Δm_{t-1}	Δinc_{t-1}	Δr^s_{t-1}	Δp_{t-1}	Δr^b_{t-1}
Δm_t	–1	0	0		0		0	0		0
Δinc_t		–1	0		0	0	0		0	0
Δr_t^s	0	0	–1	0			0			
Δp_t	0	0	0	–1	0		0	0		0
Δr^b_t	0	0	0	0	–1		0	0	0	0

	$ecm1_{t-1}$	$ecm2_{t-1}$	$ecm3_{t-1}$
Δm_t	0		
Δinc_t		0	
Δr_t^s			0
Δp_t			
Δr^b_t		0	

В системе из трех первых уравнений помимо нормализующих накладывается еще 7+7+5=19 ограничений (зануляется 19 коэффициентов). Однако необходимым минимумом для каждого из трех уравнений является наличие двух ($3-1=2$) ограничений. Таким образом, “избыточными” здесь являются: 5 ограничений в первом уравнении, 5 ограничений во втором уравнении и 3 ограничения в третьем уравнении – всего 13 ограничений.

Оценивание SECM с такими ограничениями дает следующие результаты (в скобках приведены значения *t*-статистик для оцененных коэффициентов):

	Δm_t	Δinc_t	Δr_t^s	Δp_t	Δr^b_t	Δm_{t-1}	Δr^s_{t-1}	Δp_{t-1}	Δr^b_{t-1}
Δm_t	–1			0.35 (2.6)		0.31 (2.9)		0.41 (3.4)	
Δinc_t	0.25 (1.1)	–1		0.31 (2.0)			0.17 (1.3)		
Δr_t^s			–1		1.10 (6.5)	0.21 (3.0)	0.34 (3.2)	– 0.24 (2.8)	– 0.45 (2.1)
Δp_t				–1		– 0.08 (0.8)		– 0.13 (1.0)	
Δr^b_t					–1	0.08 (1.7)			

	$ecm1_{t-1}$	$ecm2_{t-1}$	$ecm3_{t-1}$
Δm_t		0.20 (2.3)	– 0.55 (3.6)
Δinc_t	–0.44 (4.1)		0.28 (1.3)
Δr_t^s	0.19 (2.8)	– 0.28 (4.7)	
Δp_t	0.20 (2.2)	– 0.12 (1.4)	0.48 (3.4)
Δr^b_t	0.12 (3.0)		– 0.09 (1.3)

Иначе говоря, оцененная SECM имеет вид:

$$\Delta m_t = 0.35\Delta p_t + 0.31\Delta m_{t-1} + 0.41\Delta p_{t-1} + 0.20ecm2_{t-1} - 0.55ecm3_{t-1}\,,$$

$$\Delta inc = 0.25m_t + 0.31\Delta p_t + 0.17\Delta r^s_{t-1} - 0.44ecm1_{t-1} + 0.28ecm3_{t-1}\,,$$

$$\Delta r_t^s = 1.10\Delta r_t^b + 0.21\Delta m_{t-1} + 0.34\Delta r^s_{t-1} - 0.24\Delta p_{t-1} - 0.45r^b_{t-1} +$$

$$+ 0.19ecm1_{t-1} - 0.28ecm2_{t-1}\,,$$

$$\Delta p_t = -0.08\Delta m_{t-1} - 0.13\Delta p_{t-1} + 0.20ecm1_{t-1} - 0.12ecm2_{t-1} +$$
$$+ 0.48ecm3_{t-1}$$
$$\Delta r^b_{t-1} = 0.08\Delta m_{t-1} + 0.12ecm1_{t-1} - 0.09ecm3_{t-1}.$$

При этом оцененные коэффициенты по большей части статистически значимы. Поскольку система оценивалась с наложением количества ограничений больше необходимого для идентифицируемости уравнений, имеется возможность проверки гипотезы о действительном выполнении “лишних” ограничений. Статистика соответствующего критерия принимает значение 4.82, что намного меньше критического значения $\chi^2_{0.95}(13) = 22.36$, так что указанная гипотеза не отвергается.

Заметим теперь,что оцененной SECM соответствуют следующие матрицы Φ и a:

$$\Phi = \begin{pmatrix} 1 & 0 & 0 & -0.35 & 0 \\ -0.25 & 1 & 0 & -0.31 & 0 \\ 0 & 0 & 1 & 0 & -1.1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$a = \begin{pmatrix} 0 & 0.2 & -0.55 \\ -0.44 & 0 & 0.28 \\ 0.19 & -0.28 & 0 \\ 0.20 & -0.12 & 0.48 \\ 0.12 & 0 & -0.09 \end{pmatrix}.$$

При переходе от структурной формы к приведенной матрица коэффициентов адаптации находится по формуле $\alpha = \Phi^{-1}a$. В результате получаются следующие коэффициенты адаптации:

Уравнение	в приведенной ЕСМ с ограничениями			в ЕСМ (для сравнения)		
Δm_t	0.070	0.158	–0.382	0.030	0.159	–0.569
Δinc_t	–0.360	0.002	0.333	–0.458	–0.001	0.405
Δp_t	0.322	–0.280	–0.100	0.325	–0.039	0.054
Δr_t^s	0.200	–0.120	0.480	0.337	– 0.008	–0.168
Δr_t^b	0.120	0.000	–0.090	0.109	0.023	–0.213

Хотя в первых трех уравнениях SECM были нулевые коэффиценты при некоторых из $ecm1_t, ecm2_t, ecm3_t$, в приведенной форме соответствующие им коэффициенты α_{ij} отличны от нуля и показывают, как влияние отклонений от равновесия распространяется на все переменные системы.

В цитированной работе [Johansen, Juselius (1994)] проведен детальный экономический анализ результирующих уравнений, который говорит об удовлетворительной аппроксимации изучаемой экономической структуры, согласованной с имеющимися наблюдениями.

Литература

1. **Носко В.П. (2000)** *Эконометрика для начинающих*. М., ИЭПП.
2. **Носко В.П. (2004)** *Эконометрика. Элементарные методы и введение в регрессионный анализ временных рядов*. М., ИЭПП.
3. **Носко В.П. (2004)** *Эконометрика*. М., ЛОГОС.
4. **Aldrich J.H., F.D. Nelson (1984)** *Linear Probability, Logit, and Probit Models*. Beverly Hills: Sage.
5. **Amemiya T. (1985)** *Advanced Econometrics*. Blackwell, Oxford.
6. **Arellano M., S. Bond (1991)** "Some Tests of Specification for Panel Data: Monte-Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58, 277–294.
7. **Davidson R., J.G. MacKinnon (1993)** *Estimation and Inference in Econometrics*. Oxford University Press, Oxford.
8. **Godfrey L.G., J. Hutton (1994)** “Discriminating between errors-in-variables/simultaneity and misspecification in linear regression models”, *Economic Letters*, 44, 359–364.
9. **Green W.H. (1993)** *Econometric Analysis*. 2rd edition, Macmillan, New York.
10. **Hausman J. A. (1978) “**Specification tests in econometrics”, *Econometrica*, 46, 1251–1271.
11. **Hosmer D.W., S. Lemeshow (1989)** *Applied Logistic Regression*. Wiley, New York.
12. **Johansen S., K. Juselius (1994)** “Identification of the Long-run and Short-run Structure. An Application to the IS/LM Model.”, *J. of Econometrics*, 63, 7–36.
13. **Patterson K. (2000)** *An Introduction to Applied Econometrics: A Time Series Approach*. New York: St’s Martin Press.

14. **Sawa, T. (1969)** "The Exact Finite Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators," *Journal of the American Statistical Association*, 64, 923–936.
15. **Schmidt P. (1976)** *Econometrics*. New York, Marcel Dekker.
16. **Spencer D., K. Berk (1981)** "A Limited Information Specification Test", *Econometrica*, 49, 1079–1085.
17. **Staiger, D., J.H. Stock (1997)** "Instrumental Variables Regression with Weak Instruments", *Econometrica*, 65, N3, 557–586.
18. **Verbeek M. (2000)** *Modern Econometrics*. Wiley, Chichester.

Предметный указатель

О

Носко Владимир Петрович

Эконометрика для начинающих
Дополнительные главы

Носко Владимир Петрович – кандидат физико-математических наук, старший научный сотрудник механико-математического факультета Московского государственного университета им. М.В.Ломоносова. Автор более 50 научных работ и таких учебных пособий как: “Эконометрика для начинающих: Основные понятия, элементарные методы, границы применимости, интерпретация результатов”, “Эконометрика: Основные понятия и введение в регрессионный анализ временных рядов”, “Эконометрика”, соавтор учебного пособия “Основные понятия и задачи математической статистики”. Преподает эконометрику с 1994 года. В настоящее время читает курсы лекций по эконометрике на механико-математическом факультете МГУ, в Институте экономики переходного периода и в Академии народного хозяйства при Правительстве РФ.

Редактор: Н. Главацкая
Корректор: Н. Андрианова
Компьютерный дизайн: В. Юдичев

Подписано в печать 10.11.05
Тираж 500 экз.

125993, Москва, Газетный пер., 5
Тел. (095) 229–6736,
FAX (095) 203–8816
E-MAIL – info@iet.ru,
WEB Site – http://www.iet.ru