

ЗАМЕТКИ ПО ЭКОНОМЕТРИКЕ

Часть 1

С.С.ВАЛЛАНДЕР

Содержание

Предисловие	2
1 Введение	4
1.1 Специфика моделей и эмпирических данных в экономике	4
1.2 Начальное описание предмета эконометрики и ее задач	6
2 Линейная регрессионная модель	11
2.1 Спецификация модели. Соглашения об обозначениях и терминологии	11
2.2 Классическая линейная модель — обсуждение предположений	13
2.3 Оценивание коэффициентов регрессии — метод наименьших квадратов	16
2.4 Частный случай — парная регрессия	20
2.5 Свойства оценок наименьших квадратов	22
2.6 Оценивание дисперсии ошибок	24
2.7 Модель с нормально распределенными ошибками . .	26
2.8 Проверка линейных гипотез общего вида	29
2.9 Блочная регрессия	31
2.10 Коэффициент детерминации и качество прогноза . .	35
2.11 Индикаторные величины в линейной модели	40
2.12 Замечания о спецификации модели	42

Предисловие

Предлагаемая читателям первая часть "Заметок по эконометрике" отличается от традиционных учебных пособий в нескольких отношениях. Это связано с тем, что автор склонен рассматривать "Заметки" не как основной учебник, а скорее как дополнительный источник, который может оказаться полезным для различных категорий читателей. Поэтому, с одной стороны, в тексте отсутствуют иллюстративные примеры, содержащие описание результатов конкретных эконометрических работ, а, с другой стороны, нет и традиционных математических дополнений с изложением полезных сведений из линейной алгебры, теории вероятностей и математической статистики. Нет и длинного списка литературы. Все это можно найти у многих авторов, например, в первом современном отечественном учебнике Я.Р.Магнуса, П.К.Катышева и А.А.Пересецкого "Эконометрика. Начальный курс" (изд. 3-е, М., "Дело", 2000).

В то же время особое внимание в "Заметках" уделено описанию некоторых мотивировок, а также изложению деталей, недостаточно, как нам кажется, представленных в учебной литературе. С некоторым разочарованием приходится констатировать, что даже такие признанные учебники, как W.H.Greene "Econometric Analysis" (Fourth ed., Prentice Hall, 2000) и J.Stewart & L.Gill "Econometrics" (Second ed., Prentice Hall, 1998) содержат ряд неточностей, в том числе и в начальных главах. Автор попытался избежать некоторых из них. Разумеется, оставшиеся ошибки ле-

жат исключительно на его совести.

Выбранный способ изложения предполагает достаточную привычку оперирования со стандартными математическими понятиями. Он использовался автором при работе со слушателями магистерской программы по экономике в Европейском университете в Санкт-Петербурге (1997 — 2001 г.г.), а также при чтении лекций в Новгородском государственном университете им. Ярослава Мудрого в рамках гранта ИОО НВС 812 в 1999 и 2000 г.г. Выражаю всем слушателям благодарность за проявленную заинтересованность и стимулирующие вопросы и замечания. Особо благодарю коллектив преподавателей факультета экономики ЕУСПб за ту творческую атмосферу, которая поддерживается на факультете в течение теперь уже многих лет.

Во время написания первой части "Заметок" автор пользовался поддержкой гранта ИОО НВС 812.

Санкт-Петербург, март 2001 г.

Глава 1

Введение

1.1 Специфика моделей и эмпирических данных в экономике

Каждое эконометрическое исследование проводится в рамках некоторой модели — умозрительной конструкции, выделяющей главные, существенные стороны интересующего исследователя фрагмента окружающего экономического мира и отбрасывающей те, которые представляются незначимыми. В процессе исследования модель может претерпевать определенные изменения. Взаимоотношения модели и моделируемого явления могут быть довольно деликатными, и часто именно в них кроется успех (или неудача) исследования. Язык описания модели чаще всего — математика.

Экономическая наука, как одна из наук о человеческом обществе, обладает рядом особенностей, отличающих ее от многих других областей применения математических методов (в частности, от физики, где такие методы развиты в наибольшей степени).

Прежде всего следует отметить, что в экономических исследованиях практически нет места активному эксперименту. Если, скажем, физик-экспериментатор сам создает условия для проведения опыта — готовит аппаратуру, приводит в нужное состояние изучаемую субстанцию и т. д., а физик-теоретик старается

объяснить или предсказать результат такого целенаправленного эксперимента, то экономист-исследователь на первом этапе лишь наблюдает за ходом событий и фиксирует происходящее. (Последующие задачи, конечно, будут, как и в любой другой науке, стандартными — объяснить и предсказать).

Далее, человек, как существо сознательное, способен в той или иной степени влиять на общественные процессы (неважно, опираясь на экономическую теорию, вопреки ей или же вне связи с ней). Некоторые стороны подобного влияния можно условно обозначить как "политические" факторы — бóльшая часть экономических моделей рассматривает их как заданные извне — экзогенно. В других ситуациях возникают так называемые (термин часто используется и в физике) коллективные эффекты. Первый и наиболее известный пример такого эффекта в экономической сфере — "теорема о невидимой руке" Адама Смита.

Коллективные эффекты постоянно в той или иной форме проявляются в эконометрике. Обычно это выражается в присутствии стохастических характеристик (подробнее см. ниже). Заметим, впрочем, что это далеко не единственная причина их появления. Здесь следует отметить одну важную особенность. Статистические методы, развивавшиеся в течение многих десятилетий, начиная со второй половины XIX века (кинетическая теория газов Л.Больцмана), были ориентированы на использование именно в физике, где масштабы "коллективности" явлений выражаются огромными числами — из школьного курса физики известно так называемое число Авогадро — $6 \cdot 10^{23}$ молекул в одном моле вещества. Соответственно, и физические закономерности выполняются с большой точностью. Напротив, коллективные эффекты в экономической области связаны с совсем другими числами, в том числе и весьма скромными. Так, число фирм, работающих на рынке, может исчисляться тысячами, сотнями или быть еще меньше. Число покупателей, принимаемых во внимание в рассма-

триваемой модели, вряд ли будет превышать несколько миллионов (а миллион — это всего лишь 10^6).

Поэтому экономические соотношения, особенно в микроэкономических моделях, выполняются весьма приблизительно, часто даже лучше сказать — в тенденции (скорее качественно, чем количественно). Сами модели, используемые в эконометрических исследованиях, вынужденно (из соображений целесообразности) являются простыми, обычно линейными (см. ниже). Только в редких случаях, как в теории финансовых временных рядов, где исследователю могут оказаться доступными миллионы данных, имеет смысл конструировать замысловатые утонченные модели. Сами статистические методы во многих аспектах приходится переосмысливать и даже менять при переходе от физики к новым областям исследования.

1.2 Начальное описание предмета эконометрики и ее задач

Эконометрика есть ветвь экономической науки, связанная с количественным оцениванием экономических закономерностей. Эконометрическое исследование основывается на экономической теории и на фактах, относящихся к событиям, имевшим место в реальном экономическом мире.

Экономическая теория дает исследователю модель интересующих его явлений. Основные эконометрические модели имеют алгебраический характер, т.е. представляются в виде совокупности уравнений, связывающих принимаемые во внимание характеристики и включающих неопределенные (“свободные”) параметры, которые оцениваются на основе эмпирических данных. Эмпирические данные представляют собой количественно выраженные факты, относящиеся к изучаемой задаче. Как правило, предварительно они подвергаются различным процедурам уточнения, ко-

торых мы здесь не касаемся.

Важной особенностью эконометрических моделей является их стохастический характер — некоторые экономические показатели трактуются как случайные величины. Рассматриваемые ниже модели в большинстве своем являются линейными в двух отношениях. Во-первых, по параметрам, т. е. параметры входят в уравнения модели линейно. Во-вторых, по стохастическим ошибкам (см. ниже) — они включаются в уравнения аддитивно, как слагаемые, описывающие флуктуации вокруг некоторых "главных", например, средних значений. К линейным моделям иногда удается сводить и некоторые другие.

Для оценивания параметров модели, проверки гипотез о них и решения прочих сопутствующих вопросов используется эконометрическая техника, включающая в себя различные методы и приемы математической статистики, во многих случаях специально приспособленные для этих целей.

Оцененная эконометрическая модель может использоваться как для структурного анализа, включая обратное влияние на экономическую теорию, так и для прогнозирования и связанной с ним выработки экономической политики.

Основные величины, входящие в уравнения модели, подразделяются на внутренние (эндогенные) и внешние (экзогенные). Внутренние величины совместно определяются моделью; можно сказать, что в некотором смысле модель объясняет их. Напротив, экзогенные величины, хотя и входят в модель существенным образом, определяются отдельными механизмами вне ее рамок и выступают, в зависимости от ситуации, как объясняющие величины, управляющие величины, начальные или граничные условия и т. д., и т. п.

Стохастические слагаемые, входящие в уравнения линейной модели, отличаются от основных величин прежде всего тем, что они принципиально не наблюдаемы (заметим, что основные ве-

личины также могут быть случайными). Часто их называют ошибками (errors) или возмущениями. Подобные члены обычно включаются во все уравнения модели, кроме условий равновесия и тождеств (тождества можно еще трактовать как определения). Присутствие стохастических ошибок в уравнениях мотивируется комплексом причин — влиянием неучтенных факторов, непредсказуемостью человеческих реакций, неточностями наблюдений и измерений и т. д.

Приведем несколько учебных примеров (подобные примеры в разных вариантах присутствуют практически во всех учебниках). В отличие от реальных эконометрических моделей, которые могут включать значительное (иногда десятки и сотни) число уравнений и величин, упрощенные учебные примеры (часто они называются моделями-прототипами) включают минимальное число уравнений — для понимания основных принципов эконометрического исследования этого достаточно. С точки зрения эконометрической техники значительная часть проблем отчетливо проявляется уже для модели, включающей одно единственное уравнение. Часто таким уравнением оказывается уравнение линейной (множественной) регрессии, которое подробно обсуждается дальше.

Пример 1. Микроэкономическая модель-прототип спроса и предложения. Она задается уравнениями

$$\begin{aligned} q^D &= \beta_1 + \beta_2 p + \gamma_1 I + \varepsilon^D, \\ q^S &= \beta_3 + \beta_4 p + \gamma_2 r + \varepsilon^S, \\ q^D &= q^S. \end{aligned}$$

Можно представлять себе, что речь идет о производстве некоторого сельскохозяйственного продукта. При этом q^D — количество (quantity) продукта, выражающее спрос

(Demand), q^S — количество продукта, выражающее предложение (Supply), p — цена (price), I — доход (Income), r — количество осадков (rainfall). Слагаемые ε^D и ε^S — стохастические ошибки, соответствующие необъясняемым нашими уравнениями частям спроса и предложения. Последнее уравнение — условие равновесия.

Отметим еще, что сельскохозяйственное производство обладает естественной цикличностью, а наша модель предполагает, что в пределах одного цикла устанавливается равновесная цена. Поэтому время нигде явно не присутствует, а модель имеет статический характер. Нет нужды подробно останавливаться на слабых местах выбранного модельного представления — каждый может сделать это самостоятельно. Подчеркнем, однако, что при всей своей простоте модель выражает (если угодно, в карикатурной форме) некоторые теоретические представления: доход входит именно в уравнение спроса, а осадки, влияющие на урожай, — в уравнение предложения. Подобные системы уравнений называются структурными.

Нетрудно догадаться, что внутренними величинами в модели примера 1 являются цена p и количество продукта $q(= q^D = q^S)$, в то время как доход I и осадки r целесообразно трактовать внешним, экзогенным, образом.

Пример 2. Макроэкономическая модель-прототип определения национального дохода. Она задается уравнениями

$$\begin{aligned} C_t &= \beta_1 + \beta_2 Y_t + \varepsilon_t^C, \\ I_t &= \beta_3 + \beta_4 Y_t + \gamma_1 Y_{t-1} + \varepsilon_t^I, \\ Y_t &= C_t + I_t + G_t. \end{aligned}$$

Здесь внутренними являются величины C_t , I_t , Y_t , описывающие, соответственно, потребление (Consumption), ин-

вестиции (Investment) и доход (Yield) в году t , а внешней — G_t — правительственные расходы (Government spending). Запаздывающее (лаговое, lagged) значение Y_{t-1} национального дохода вместе с G_t составляет набор predetermined) величин. Последнее уравнение является тождеством и не содержит стохастического слагаемого.

Отметим, что пример 2, в отличие от примера 1, имеет отчетливо выраженный динамический характер. При решении этой структурной системы уравнений кроме ”граничных” условий, определяемых правительственными расходами G_t , скорее всего, появится еще и ”начальное” условие (скажем, Y_0 , если время t изменяется, начиная с 1).

Приведенные выше описания моделей в примерах 1 и 2 являются неполными. Следует еще уточнить предположения о характере стохастических слагаемых ε . Анализ и проверка этих предположений — важная часть эконометрического исследования.

Глава 2

Линейная регрессионная модель

2.1 Спецификация модели. Соглашения об обозначениях и терминологии

Спецификацией модели называют ее концептуальную функциональную форму. В этой главе будет рассматриваться модель, имеющая спецификацию

$$Y = \beta_1 X_1 + \cdots + \beta_k X_k + \varepsilon. \quad (2.1)$$

В уравнении (2.1) Y — объясняемая величина, X_1, \dots, X_k — объясняющие величины, или регрессоры, ε — стохастическая ошибка. Коэффициенты β_1, \dots, β_k — неопределенные (свободные) параметры, подлежащие оцениванию.

Спецификация (2.1) подразумевает некоторую теоретическую концепцию — мы считаем, что существуют ”истинные” значения коэффициентов $\beta_{1,true}, \dots, \beta_{k,true}$, но они неизвестны и могут обсуждаться лишь умозрительно. (Конечно, подобное замечание относится к любой задаче оценивания, однако в литературе по статистике об этом обстоятельстве редко говорят прямо.) Следуя установившейся традиции, мы в дальнейшем изложении будем часто использовать обозначение β и для ”истинных” коэффициентов.

С практической точки зрения исследователь располагает дан-

ными N совместных наблюдений величин Y, X_1, \dots, X_k , так что для i -го наблюдения ($i = 1, \dots, N$) может представлять себе соотношение

$$Y_i = \beta_1 X_{i1} + \dots + \beta_k X_{ik} + \varepsilon_i \quad (2.2)$$

(представление данных), вытекающее из спецификации модели. Подчеркнем, что первый индекс из двух в нашей системе обозначений всегда — номер наблюдения. Если же индекс всего один, то он обозначает номер наблюдения у Y и ε , но номер регрессора у X .

Отличие формул (2.1) и (2.2) в том, что спецификация (2.1) может обсуждаться вне всякой связи с эмпирическими данными, т.е. концептуально, при этом $Y, X_1, \dots, X_k, \varepsilon$ оказываются обозначениями для типов объектов. Напротив, $Y_i, X_{ij}, \varepsilon_i$ в формуле (2.2) понимаются как величины, отвечающие i -му наблюдению, т.е. как конкретные объекты, а не типы объектов. С точки зрения пользователя Y_i и X_{ij} можно также трактовать как числа — ”реализовавшиеся” значения соответствующих величин. Для ε_i такого утилитарного понимания быть не может — коэффициенты модели свободны, т.е. неизвестны исследователю, а потому и ошибка не наблюдаема.

Удобно использовать также сокращенные векторно-матричные обозначения. При этом значения Y_i объединяются в вектор-столбец Y размерности N ; аналогично, значения X_{ij} объединяются в матрицу X , имеющую N строк и k столбцов, а ε_i — в вектор-столбец ε . Столбцы матрицы X удобно обозначать X_1, \dots, X_k — они состоят из значений соответствующих регрессоров. В этих обозначениях формула (2.1) приобретает второй смысл — смысл соотношения между N -мерными векторами Y, X_1, \dots, X_k и ε . Полностью сокращенную его запись

$$Y = X\beta + \varepsilon \quad (2.3)$$

мы получим, если введем еще и вектор-столбец β коэффициентов. Размерность вектора β , очевидно, равна k .

Заготовим сразу же еще одно соглашение об обозначениях. Среднее арифметическое компонент некоторого вектора (неважно, случайного или нет) будет обозначаться традиционной для статистики чертой сверху, например,

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i, \bar{X}_j = \frac{1}{N} \sum_{i=1}^N X_{ij},$$

а отклонения от этого среднего значения — соответствующей малой буквой:

$$y_i = Y_i - \bar{Y}, x_{ij} = X_{ij} - \bar{X}_j$$

и т. д. Аналогичные отклонения для вектора ошибок будут записываться подробно: $\varepsilon_i - \bar{\varepsilon}$.

Используя обозначение d^\rightarrow для вектора, все компоненты которого равны d , можно записать отклонения в векторной форме

$$y = Y - \bar{Y}^\rightarrow, \quad x_j = X_j - \bar{X}_j^\rightarrow, \quad \varepsilon = (\bar{\varepsilon})^\rightarrow.$$

2.2 Классическая линейная модель — обсуждение предположений

В этом параграфе мы дополняем спецификацию (2.1) простейшими предположениями о регрессорах и ошибках и получаем полное описание так называемого классического варианта линейной регрессионной модели.

Предположения о регрессорах включают два разноплановых свойства. Во-первых, регрессоры предполагаются неслучайными. Примерами таких регрессоров являются:

1. Константа; этот регрессор обычно включается в модель под первым номером: $X_1 = 1^\rightarrow$ (константу, отличную от единицы,

можно включить множителем в соответствующий коэффициент β_1).

2. "Время": $X_{i2} = i$.

3. Любая "управляющая", т. е. подконтрольная исследователю величина.

С точки зрения экономической теории неслучайность регрессоров (особенно всех!) не очень частое явление, так что сделанное предположение довольно ограничительно. В дальнейшем (глава 3) мы будем обсуждать обобщения классической модели, в которых это предположение заменяется более реалистичными.

Второе предположение о регрессорах имеет прозаический характер: столбцы X_1, \dots, X_k регрессионной матрицы X предполагаются линейно независимыми векторами. Это свойство означает, что нельзя уменьшить количество регрессоров, выразив некоторые из них (хотя бы один) через остальные.

Предположение о линейной независимости столбцов регрессоров может выполняться лишь в случае, когда число наблюдений N не меньше числа регрессоров. Это вполне укладывается в обычные статистические рамки — оценить много параметров по малому числу наблюдений почти никогда не удастся осмысленным образом. Конечно, желательно, чтобы N было значительно больше k .

Перейдем теперь к предположениям об ошибках. В классической модели они формулируются наиболее жестким и не всегда реалистичным образом:

- предполагается, что ошибки ε_i ($i = 1, \dots, N$) образуют так называемый слабый белый шум — последовательность центрированных ($\mathbf{E}\varepsilon_i = 0$) и некоррелированных ($\mathbf{E}(\varepsilon_{i_1}\varepsilon_{i_2}) = 0$ при $i_1 \neq i_2$) случайных величин с одинаковыми дисперсиями $\mathbf{E}(\varepsilon_i^2) = \sigma^2$.

Свойство центрированности практически не является ограничением, т. к. при наличии постоянного регрессора среднее значение ошибки можно было бы включить в соответствующий коэффициент ($\beta_1 + \varepsilon = \beta_1 + \mathbf{E}\varepsilon + (\varepsilon - \mathbf{E}\varepsilon)$).

Обобщения классической модели, включающие автокорреляцию ошибок и/или неоднородность дисперсий, будут рассмотрены дальше (глава 3).

В ряде случаев сделанные предположения об ошибках будут дополняться свойством нормальности (гауссовости) — случайный вектор ε имеет нормальное распределение (гауссовский белый шум). Такую модель мы будем называть классической моделью с нормально распределенными ошибками. Как хорошо известно, многомерное нормальное распределение задается своим вектором математических ожиданий (в нашем случае это нулевой вектор) и матрицей ковариаций — здесь она имеет вид $\sigma^2 \mathbf{1}$, где $\mathbf{1}$ — единичная матрица. Если компоненты нормально распределенного вектора некоррелированы, они автоматически оказываются независимыми, так что в классической модели с нормально распределенными ошибками эти ошибки образуют последовательность независимых одинаково нормально распределенных случайных величин $\mathbf{N}(0, \sigma^2)$.

Отметим еще одну тонкость, касающуюся многомерного нормального распределения — если каждая из величин ε_i нормально распределена, то вектор ε , из них составленный, не обязан быть нормально распределенным (даже если величины ε_i не коррелируют!). К сожалению, в литературе иногда встречаются неаккуратные формулировки, игнорирующие эту тонкость.

2.3 Оценивание коэффициентов регрессии — метод наименьших квадратов

Классическая модель линейной регрессии имеет своими параметрами β_1, \dots, β_k и σ . Подчеркнем, что все они, включая σ , входят в модель линейно (параметр σ можно было бы явным образом выделить, записывая ошибку ε в виде $\sigma \cdot (\varepsilon/\sigma)$ и учитывая, что случайная величина ε/σ стандартизована — имеет нулевое математическое ожидание и единичную дисперсию). Отметим, впрочем, что из наших ”слабых” предположений не следует, что величины ошибок ε_i одинаково распределены — это предполагается лишь на уровне второго порядка, а информация о моментах более высоких порядков не учитывается.

В этом параграфе мы рассматриваем первый этап процедуры оценивания — построение оценок коэффициентов регрессии β_1, \dots, β_k методом наименьших квадратов (МНК; английская аббревиатура OLS — ordinary least squares). Идею этого метода, предложенного К.Гауссом в начале XIX века, удобнее всего излагать геометрически — на языке векторов N -мерного пространства. В ходе этого обсуждения коэффициенты β_1, \dots, β_k будут трактоваться как свободно меняющиеся параметры. ”Истинные” их значения $\beta_{1,true}, \dots, \beta_{k,true}$ в ходе рассуждений явно появляться почти не будут.

Итак, в нашем распоряжении имеются векторы значений регрессоров X_1, \dots, X_k и вектор значений объясняемой величины Y . Мы стремимся найти такую линейную комбинацию $X\beta = \beta_1 X_1 + \dots + \beta_k X_k$ регрессоров, которая ”лучше всего” объясняла бы Y , т.е. ”с наименьшим отклонением”. Естественнее всего представляется измерять отклонение $Y - X\beta$ длиной соответствующего вектора и подбирать коэффициенты β так, чтобы эта длина (или, что равносильно, ее квадрат) была минимальна. Квадрат

длины отклонения $Y - X\beta$ равен

$$(Y - X\beta)'(Y - X\beta) = \sum_{i=1}^N (Y_i - \beta_1 X_{i1} - \cdots - \beta_k X_{ik})^2, \quad (2.4)$$

так что предложение Гаусса сводится к поиску точки минимума $\hat{\beta}$ этой квадратичной функции коэффициентов и объявлению ее оценкой вектора "истинных" коэффициентов β_{true} .

Хотя возможны и другие меры отклонения, например, сумма модулей вместо суммы квадратов, однако они не получили широкого распространения. Отчасти это связано с наличием у суммы квадратов ряда удобных свойств (см. ниже), а отчасти, по-видимому, с тем, что мы привыкли к евклидову способу измерения расстояний, и он нам кажется самым естественным. Определенную роль играют и установившиеся традиции.

Для нахождения точки минимума $\hat{\beta}$ мы снова воспользуемся геометрическими рассуждениями. Рассмотрим в N -мерном пространстве \mathbb{R}^N взаимное положение вектора Y и подпространства $\mathcal{L}(X_1, \dots, X_k)$, порожденного векторами X_1, \dots, X_k регрессоров (его размерность, очевидно, равна k). Пусть \hat{Y} — ортогональная проекция вектора Y на подпространство $\mathcal{L}(X_1, \dots, X_k)$. Тогда вектор-разность $Y - \hat{Y}$ перпендикулярен этому подпространству. Если $X\beta = \beta_1 X_1 + \cdots + \beta_k X_k$ — какая-то другая точка подпространства $\mathcal{L}(X_1, \dots, X_k)$, то разность $Y - X\beta$ можно трактовать как наклонную, в то время как $Y - \hat{Y}$ — перпендикуляр. Так как перпендикуляр короче наклонной, получаем

$$(Y - \hat{Y})'(Y - \hat{Y}) < (Y - X\beta)'(Y - X\beta).$$

Поэтому \hat{Y} доставляет минимум сумме квадратов (2.4).

Поскольку векторы регрессоров X_1, \dots, X_k линейно независимы, проекция \hat{Y} единственным образом разлагается в линейную комбинацию их:

$$\hat{Y} = \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k = X\hat{\beta}.$$

Вектор $\hat{\beta}$ коэффициентов — искомый.

От геометрической интерпретации точки минимума перейдем к соответствующим формулам. Запишем условие ортогональности

$$Y - \hat{Y} \perp \mathcal{L}(X_1, \dots, X_k)$$

в виде

$$(X\beta)'(Y - X\hat{\beta}) = 0. \quad (2.5)$$

Здесь $X\beta$ — произвольный вектор пространства $\mathcal{L}(X_1, \dots, X_k)$.

Перепишем теперь равенство (2.5) в виде

$$\beta' \cdot X'(Y - X\hat{\beta}) = 0$$

и заметим, что геометрически оно может быть истолковано как еще одно условие ортогональности

$$\beta \perp X'(Y - X\hat{\beta})$$

(теперь уже для векторов k -мерного пространства \mathbb{R}^k). Таким образом, k -мерный вектор $X'(Y - X\hat{\beta})$ ортогонален произвольному вектору β пространства \mathbb{R}^k . Отсюда следует (даже равносильно), что он нулевой:

$$X'(Y - X\hat{\beta}) = 0.$$

Записывая это равенство в виде

$$X'X\hat{\beta} = X'Y, \quad (2.6)$$

получаем для $\hat{\beta}$ так называемое нормальное уравнение МНК. Легко сообразить, что оно имеет единственное решение. Действительно, по предположению, ранг матрицы X равен k . Из свойств ранга матрицы следует, что тогда и ранг $X'X$ равен k . Поскольку $X'X$ — квадратная матрица порядка k , заключаем, что она обратима.

Окончательно, получаем выражение для оценок метода наименьших квадратов

$$\hat{\beta} = (X'X)^{-1}X'Y. \quad (2.7)$$

Важно подчеркнуть, что вектор оценок $\hat{\beta}$ получается линейным преобразованием случайного вектора Y .

Образованный с помощью этих оценок вектор $\hat{Y} = X\hat{\beta}$ можно назвать вектором прогнозных (предсказываемых моделью) значений величины Y (английский термин — predicted values или fitted values).

Обозначим через P оператор ортогонального проектирования на подпространство регрессоров $\mathcal{L}(X_1, \dots, X_k)$ (и соответствующую матрицу). Из формулы (2.7) следует, что

$$P = X(X'X)^{-1}X'. \quad (2.8)$$

Эта матрица, а также матрица $P^\perp = \mathbf{1} - P$, соответствующая проектированию на подпространство $\mathcal{L}^\perp(X_1, \dots, X_k)$ векторов, ортогональных регрессорам, будут часто использоваться в последующих обсуждениях. Выпишем некоторые их свойства, легко вытекающие как из геометрического смысла проекций, так и из формального определения (2.8). Проверка этих свойств оставляется читателю.

$$\begin{aligned} P &= P', & P^\perp &= (P^\perp)', & (\text{симметричность}) \\ P &= P^2, & P^\perp &= (P^\perp)^2, & (\text{идемпотентность}) \\ PP^\perp &= P^\perp P = 0, & P + P^\perp &= \mathbf{1}, \\ PX_j &= X_j, & P^\perp X_j &= 0, \\ PX &= X, & P^\perp X &= 0. \end{aligned}$$

Вектор

$$\hat{\varepsilon} = Y - \hat{Y} = P^\perp Y$$

называется вектором остатков (residuals). Для него можно записать также другое выражение

$$\hat{\varepsilon} = P^\perp(X\beta + \varepsilon) = P^\perp\varepsilon$$

($P^\perp X = 0$, как указано ранее). Остатки можно интерпретировать как "оцененные ошибки". Очевидно, $P\hat{\varepsilon} = 0$.

Подставляя в формулу (2.7) спецификацию (2.3), получаем еще одну полезную формулу

$$\hat{\beta} = (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon. \quad (2.9)$$

Если формула (2.7) содержит лишь наблюдаемые значения и поэтому может использоваться для расчетов, формула (2.9) играет важную теоретическую роль (см. дальше параграф 2.5).

2.4 Частный случай — парная регрессия

Полезно выписать явно два простейших случая формулы (2.7).

Случай 1 ($k = 1$). Очевидно, имеем

$$X'X = \sum_{i=1}^N X_{i1}^2, \quad X'Y = \sum_{i=1}^N X_{i1}Y_i, \\ \hat{\beta}_1 = \frac{\sum_{i=1}^N X_{i1}Y_i}{\sum_{i=1}^N X_{i1}^2} = \frac{\overline{X_1Y}}{\overline{X_1^2}}.$$

Если дополнительно предположить, что $X_1 = 1^\rightarrow$ (регрессия на константу), получаем

$$\hat{\beta}_1 = \bar{Y},$$

так что прогнозные значения \hat{Y}_i равны \bar{Y} при всех i , что можно записать также в виде $\hat{Y} = \bar{Y}^\rightarrow$.

Случай 2 ($k = 2$). Аналогично предыдущему случаю получаем

$$\frac{1}{N}X'X = \begin{pmatrix} \overline{X_1^2} & \overline{X_1X_2} \\ \overline{X_1X_2} & \overline{X_2^2} \end{pmatrix}, \quad \frac{1}{N}X'Y = \begin{pmatrix} \overline{X_1Y} \\ \overline{X_2Y} \end{pmatrix},$$

$$\hat{\beta}_1 = \frac{\overline{X_2^2} \cdot \overline{X_1 Y} - \overline{X_1 X_2} \cdot \overline{X_2 Y}}{\overline{X_1^2} \cdot \overline{X_2^2} - \overline{X_1 X_2}^2},$$

$$\hat{\beta}_2 = \frac{\overline{X_1^2} \cdot \overline{X_2 Y} - \overline{X_1 X_2} \cdot \overline{X_1 Y}}{\overline{X_1^2} \cdot \overline{X_2^2} - \overline{X_1 X_2}^2}.$$

При дополнительном предположении $X_1 = 1 \rightarrow$ (модель парной регрессии) формулы можно несколько упростить:

$$\hat{\beta}_1 = \frac{\overline{X_2^2} \cdot \bar{Y} - \bar{X}_2 \cdot \overline{X_2 Y}}{\overline{X_2^2} - \bar{X}_2^2} = \bar{Y} - \bar{X}_2 \hat{\beta}_2,$$

$$\hat{\beta}_2 = \frac{\overline{X_2 Y} - \bar{X}_2 \bar{Y}}{\overline{X_2^2} - \bar{X}_2^2} = \frac{\overline{x_2 y}}{\overline{x_2^2}}. \quad (2.10)$$

Для вектора \hat{Y} прогнозных значений из формул (2.10) получаем

$$\hat{Y} = \bar{Y} \rightarrow + \frac{\overline{x_2 y}}{\overline{x_2^2}} (X_2 - \bar{X}_2 \rightarrow) = \bar{Y} \rightarrow + \frac{\overline{x_2 y}}{\overline{x_2^2}} x_2. \quad (2.11)$$

Очевидно, $\bar{x}_2 = 0$, поэтому, усредняя (2.11), находим

$$\overline{\hat{Y}} = \bar{Y}.$$

Переносим теперь в (2.11) вектор $\bar{Y} \rightarrow$ в левую часть, находим

$$\hat{y} = \frac{\overline{x_2 y}}{\overline{x_2^2}} x_2 \quad (2.11')$$

— прогнозный вектор в отклонениях.

Сопоставляя между собой полученные формулы, можно обнаружить еще и такую двухступенчатую процедуру построения оценки коэффициента парной регрессии $\hat{\beta}_2$ (см. (2.10)): сначала строятся регрессии величин Y и X_2 на константу и находятся векторы остатков y и x_2 . Затем строится регрессия величины y на

x_2 — формула (2.11'). Сходная процедура для линейной модели с произвольным числом регрессоров будет обсуждаться в §2.9.

Упражнение. Показать, что регрессия с двумя произвольными регрессорами может быть получена аналогичной двухступенчатой процедурой.

2.5 Свойства оценок наименьших квадратов

В этом параграфе рассматриваются статистические свойства оценок МНК, поэтому предположение о том, что регрессоры неслучайны, будет играть важную роль (до сих пор оно не использовалось).

Первое свойство — несмещенность вектора оценок $\hat{\beta}$. Оно является, как сейчас будет видно, следствием линейности по Y . Действительно, с помощью формулы (2.9) получаем

$$\begin{aligned}\mathbf{E}\hat{\beta} &= \beta + \mathbf{E}(X'X)^{-1}X'\varepsilon = \\ &= \beta + (X'X)^{-1}X'\mathbf{E}\varepsilon = \beta.\end{aligned}$$

Здесь мы в чистом виде пользуемся линейностью — постоянные множители, в том числе и матричные, выносятся за знак математического ожидания. Сходное вычисление дает нам матрицу ковариаций вектора $\hat{\beta}$:

$$\begin{aligned}\text{cov}(\hat{\beta}) &= \mathbf{E}[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = \mathbf{E}[(X'X)^{-1}X'\varepsilon \cdot ((X'X)^{-1}X'\varepsilon)'] = \\ &= \mathbf{E}[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}] = (X'X)^{-1}X'\mathbf{E}(\varepsilon\varepsilon')X(X'X)^{-1} = \\ &= \sigma^2(X'X)^{-1}X' \cdot X(X'X)^{-1} = \sigma^2(X'X)^{-1}.\end{aligned}$$

Нелишним будет подчеркнуть, что в матричных вычислениях порядок сомножителей должен выдерживаться (левый множитель — налево, правый — направо).

ТЕОРЕМА ГАУССА-МАРКОВА. Оценка $\hat{\beta}$ метода наименьших квадратов является эффективной в классе линейных несмещенных оценок.

Уточним сначала, что понимается под эффективностью векторной несмещенной оценки. Пусть $\tilde{\beta}$ — другая линейная несмещенная оценка вектора β . Тогда матрица

$$\text{cov}(\tilde{\beta}) - \text{cov}(\hat{\beta})$$

должна быть неотрицательно определена. Это означает, что для любого вектора $\gamma \in \mathbb{R}^k$ величина

$$\gamma' [\text{cov}(\tilde{\beta}) - \text{cov}(\hat{\beta})] \gamma \quad (= \text{var}(\gamma' \tilde{\beta}) - \text{var}(\gamma' \hat{\beta}))$$

неотрицательна.

Доказательство теоремы. Запишем линейную оценку $\tilde{\beta}$ в виде

$$\tilde{\beta} = CY.$$

Тогда условие несмещенности $\mathbf{E}\tilde{\beta} = \beta$ записывается в виде $CX\beta = \beta$, причем последнее равенство должно выполняться тождественно по β (ведь β — это неизвестный параметр). Таким образом, матрица C должна удовлетворять условию $CX = \mathbf{1}$. Представим ее в виде

$$C = (X'X)^{-1}X' + D.$$

Через вспомогательную матрицу D условие несмещенности записывается как $DX = 0$. Матрица ковариаций $\text{cov}(\tilde{\beta})$ выражается формулой

$$\begin{aligned} \text{cov}(\tilde{\beta}) &= \mathbf{E}[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'] = \\ &= \mathbf{E}[C\varepsilon(C\varepsilon)'] = \sigma^2 CC' = \\ &= \sigma^2[(X'X)^{-1} + DD' + (X'X)^{-1}X'D' + D((X'X)^{-1}X')'] = \\ &= \sigma^2[(X'X)^{-1} + DD']. \end{aligned}$$

Здесь мы воспользовались условием несмещенности $DX = 0$. Остается проверить неотрицательную определенность матрицы DD' :

$$\gamma' DD' \gamma = (D' \gamma)' (D' \gamma) \geq 0$$

как квадрат длины вектора $D' \gamma$. Теорема доказана.

Из теоремы Гаусса-Маркова вытекает, в частности, что $\text{var}(\tilde{\beta}_j) \geq \text{var}(\hat{\beta}_j)$, так что скалярные оценки $\hat{\beta}_j$ эффективны в аналогичном классе линейных несмещенных оценок.

Повторяя почти дословно доказательство теоремы Гаусса-Маркова, можно доказать, что для любой матрицы Γ , имеющей k столбцов, эффективной линейной несмещенной оценкой вектора $\Gamma\beta$ является оценка $\Gamma\hat{\beta}$. Это утверждение оставляется читателю для самостоятельной проверки.

В частности, линейные комбинации оценок МНК эффективно оценивают аналогичные линейные комбинации коэффициентов регрессии.

2.6 Оценивание дисперсии ошибок

Дисперсия σ^2 является квадратичной характеристикой ошибок — моментом второго порядка, поэтому оценивать ее, видимо, следует также квадратичным образом. При этом естественным эмпирическим объектом, ассоциирующимся с ошибками, является вектор остатков $\hat{\varepsilon} = P^\perp \varepsilon$. Очевидно, $\mathbf{E}\hat{\varepsilon} = 0$. Найдем матрицу ковариаций

$$\text{cov}(\hat{\varepsilon}) = \mathbf{E}[P^\perp \varepsilon (P^\perp \varepsilon)'] = P^\perp \mathbf{E}(\varepsilon \varepsilon') P^\perp = \sigma^2 P^\perp.$$

Рассмотрим теперь сумму квадратов

$$\hat{\varepsilon}' \hat{\varepsilon} = \text{tr}(\hat{\varepsilon} \hat{\varepsilon}').$$

Соответствующее математическое ожидание равно

$$\mathbf{E}(\hat{\varepsilon}' \hat{\varepsilon}) = \mathbf{E} \text{tr}(\hat{\varepsilon} \hat{\varepsilon}') = \text{tr} \mathbf{E}(\hat{\varepsilon} \hat{\varepsilon}') = \sigma^2 \text{tr} P^\perp.$$

Остается вспомнить, что P^\perp — ортогональный проектор на подпространство $\mathcal{L}^\perp(X_1, \dots, X_k)$, имеющее размерность $N - k$, дополнительную к размерности подпространства регрессоров, и его след (как и любого проектора) равен этой размерности.

Альтернативное доказательство равенства $\text{tr} P^\perp = N - k$ можно провести прямым вычислением

$$\begin{aligned}\text{tr} P^\perp &= \text{tr}[\mathbf{1}_N - X(X'X)^{-1}X'] = N - \text{tr}[X(X'X)^{-1}X'] = \\ &= N - \text{tr}[(X'X)^{-1}X'X] = N - \text{tr} \mathbf{1}_k = N - k\end{aligned}$$

(мы пользуемся тем, что при циклической перестановке сомножителей след произведения матриц не меняется).

Из проведенных вычислений следует, что статистика

$$s^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{N - k} \quad (2.12)$$

является несмещенной оценкой дисперсии σ^2 . Этот результат эвристически объясняется тем, что после оценивания k коэффициентов регрессии в эмпирических данных остается $N - k$ неиспользованных степеней свободы.

В модели со слабым белым шумом, оперирующей только с моментами первого и второго порядка, обсуждать эффективность оценки s^2 (в каком-либо подходящем классе) невозможно, т.к. отсутствуют предположения о старших моментах. Единственное, что остается еще получить в рамках этого подхода — это матрицу перекрестных ковариаций векторов $\hat{\beta}$ и $\hat{\varepsilon}$:

$$\begin{aligned}\text{cov}(\hat{\beta}, \hat{\varepsilon}) &= \mathbf{E}((\hat{\beta} - \beta)\hat{\varepsilon}') = (X'X)^{-1}X'\mathbf{E}(\varepsilon\varepsilon')P^\perp = \\ &= \sigma^2(X'X)^{-1}X'P^\perp = 0\end{aligned} \quad (2.13)$$

(опять используем равенство $P^\perp X = 0$ из параграфа 2.3).

Оценка s^2 позволяет оценить и матрицу ковариаций вектора $\hat{\beta}$. В выражении

$$\text{cov}(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

надо лишь заменить σ^2 на s^2 :

$$\overline{\text{cov}}(\hat{\beta}) = s^2(X'X)^{-1}.$$

Эта матричная оценка, очевидно, оказывается несмещенной.

2.7 Модель с нормально распределенными ошибками

Предположение о нормальности распределения вектора ошибок позволяет уточнить и усилить ряд свойств, выведенных в предыдущих параграфах. Во-первых, появляется возможность включить оценки наименьших квадратов в общую схему метода максимального правдоподобия и сравнивать их не только с линейными оценками. Во-вторых, с нормальным распределением связаны другие, хорошо известные в статистике распределения — хи-квадрат, Стюдента, Фишера, которые сразу начинают работать.

Начнем с обсуждения метода максимального правдоподобия. В сделанных предположениях наблюдаемый вектор Y имеет нормальное распределение $N(X\beta, \sigma^2 \mathbf{1})$. Соответствующая функция правдоподобия имеет вид

$$\begin{aligned} L(\beta, \sigma^2) &= \prod_{i=1}^N \left[\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(Y_i - (X\beta)_i)^2}{2\sigma^2}} \right] = \\ &= (2\pi)^{-N/2} \sigma^{-N} \exp\left[-\frac{1}{2\sigma^2} (Y - X\beta)'(Y - X\beta)\right]. \end{aligned}$$

Поэтому максимизировать ее по β — то же самое, что минимизировать сумму квадратов $(Y - X\beta)'(Y - X\beta)$. Таким образом, оценка $\hat{\beta}$ метода наименьших квадратов оказывается одновременно и оценкой максимального правдоподобия. Далее,

$$L(\hat{\beta}, \sigma^2) = (2\pi)^{-N/2} \sigma^{-N} \exp\left[-\frac{\hat{\varepsilon}'\hat{\varepsilon}}{2\sigma^2}\right].$$

Отсюда находится оценка максимального правдоподобия для σ^2 :

$$\sigma_{M.L.}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{N}.$$

Как и следовало ожидать, она смещенная (см. предыдущий параграф). Ее исправление дает несмещенную оценку s^2 , обсуждавшуюся выше.

С помощью многомерного неравенства Рао–Крамера можно доказать, что $\hat{\beta}$ — эффективная оценка в классе всех (не обязательно линейных) несмещенных оценок вектора β . Утверждение о том, что s^2 — эффективная несмещенная оценка дисперсии σ^2 , тоже верно, но для его доказательства приходится применять более сложные методы — теорию достаточных статистик (достаточная статистика в нашей ситуации имеет вид $(Y'Y, X'Y)$). Мы не приводим деталей соответствующих рассуждений, оставляя их для самостоятельного исследования наиболее подготовленными читателями.

Перейдем теперь к свойствам оценок $\hat{\beta}$ и s^2 . Прежде всего, заметим, что они независимы. Действительно, случайный вектор $(\hat{\beta}', \hat{\varepsilon}')'$ нормально распределен. Согласно формуле (2.13) подвекторы $\hat{\beta}$ и $\hat{\varepsilon}$ не коррелируют. Следовательно, они независимы. А тогда и $s^2 = \hat{\varepsilon}'\hat{\varepsilon}/(N - k)$ не зависит от $\hat{\beta}$.

Докажем теперь, что случайная величина $\hat{\varepsilon}'\hat{\varepsilon}/\sigma^2$ распределена по хи-квадрат с $N - k$ степенями свободы. Мы уже проверяли в параграфе 2.3, что $\hat{\varepsilon} = P^\perp \varepsilon$. Выберем ортогональный нормированный базис e_1, \dots, e_{N-k} в подпространстве $\mathcal{L}^\perp(X_1, \dots, X_k)$, где принимает значения $\hat{\varepsilon}$. Пусть e — матрица, составленная из столбцов e_1, \dots, e_{N-k} . Тогда $e'\hat{\varepsilon}$ — вектор размерности $N - k$, составленный из координат вектора $\hat{\varepsilon}$ в базисе e_1, \dots, e_{N-k} . Очевидно, $e'\hat{\varepsilon}$ нормально распределен и центрирован. Вычислим его матрицу

ковариаций

$$\begin{aligned}\text{cov}(e' \hat{\varepsilon}) &= \mathbf{E}[e' \hat{\varepsilon} (e' \hat{\varepsilon})'] = e' \mathbf{E}(\hat{\varepsilon} \hat{\varepsilon}') e = \\ &= \sigma^2 e' P^\perp e = \sigma^2 e' e = \sigma^2 \mathbf{1}_{N-k}\end{aligned}$$

(мы воспользовались вычисленным в параграфе 2.6 значением $\mathbf{E}(\hat{\varepsilon} \hat{\varepsilon}') = \sigma^2 P^\perp$, а также тем, что P^\perp действует тождественно на векторы базиса e_1, \dots, e_{N-k}). Заметим теперь, что суммы квадратов $\hat{\varepsilon}' \hat{\varepsilon}$ и $(e' \hat{\varepsilon})' \cdot (e' \hat{\varepsilon})$ дают одну величину — квадрат длины вектора $\hat{\varepsilon}$. Отсюда получаем, что

$$\sigma^{-2} \hat{\varepsilon}' \hat{\varepsilon} = \sigma^{-2} \sum_{j=1}^{N-k} (e'_j \hat{\varepsilon})^2$$

имеет распределение χ^2_{N-k} . Действительно, величины $\sigma^{-1} e'_j \hat{\varepsilon}$ имеют стандартное нормальное распределение и независимы.

Теперь мы получаем возможность построения доверительных интервалов для коэффициентов регрессии β_j и совместных доверительных областей для них. Ограничимся пока описанием конструкции доверительных интервалов. Мы знаем, что

$$\hat{\beta}_j \in \mathbf{N}(\beta_j, \sigma^2 [(X'X)^{-1}]_{jj}),$$

$$\frac{(N-k)s^2}{\sigma^2} \in \chi^2_{N-k},$$

и эти величины независимы. Поэтому

$$\sqrt{N-k} \frac{\frac{\hat{\beta}_j - \beta_j}{\sqrt{\text{var}(\hat{\beta}_j)}}}{\sqrt{\frac{(N-k)s^2}{\sigma^2}}} = \frac{\hat{\beta}_j - \beta_j}{s \sqrt{[(X'X)^{-1}]_{jj}}}$$

имеет распределение Стьюдента \mathbf{t}_{N-k} . Выбирая по доверительной вероятности $1 - \alpha$ соответствующее табличное значение z_α ($(1 - \alpha/2)$ -квантиль распределения Стьюдента), мы получаем доверительный интервал вида $\hat{\beta}_j \pm z_\alpha s \sqrt{[(X'X)^{-1}]_{jj}}$ для коэффици-

ента β_j . При большом числе степеней свободы распределение Стьюдента, как обычно, может быть заменено нормальным.

Доверительный интервал позволяет проверять гипотезу вида $\beta_j = \beta_{j0}$. Для этого достаточно лишь выяснить, попадает ли гипотетическое значение β_{j0} в построенный доверительный интервал. Гипотеза отвергается на уровне α , если гипотетическое значение β_{j0} не попадает в доверительный интервал.

Проверка более сложных гипотез, включающих линейные комбинации коэффициентов регрессии, обсуждается в следующем параграфе.

Доверительный интервал для σ^2 строится непосредственно по χ^2 -распределенной дроби $(N - k)s^2/\sigma^2$. Мы предполагаем, что читатель может проделать это самостоятельно.

Без предположения о нормальности ошибок оба специальных распределения — Стьюдента и хи-квадрат — исчезают, однако часто предполагают, что при больших N изложенные рецепты дают "приближенные" доверительные интервалы.

2.8 Проверка линейных гипотез общего вида

Простейшие гипотезы вида $\beta_j = \beta_{j0}$ о коэффициентах регрессии, рассмотренные выше, составляют лишь малую часть содержательных линейных гипотез. Обозначим на уровне идей ряд примеров, в которых появляются гипотезы другого вида.

Гипотеза $\beta_2 + \beta_3 = 1$ появляется в связи с производственной функцией Кобба–Дугласа.

Гипотеза $\beta_2 + \beta_3 = 0$ может проверяться в модели, где X_2 — ставка банковского процента, а X_3 — уровень инфляции.

Гипотеза $\beta_2 = \beta_3 = \dots = \beta_k = 0$ появляется при выяснении вопроса о значимости всей регрессионной связи.

Общая формулировка линейной гипотезы о коэффициентах

имеет следующий вид:

$$H_0 : \quad R\beta = \gamma.$$

Здесь R — матрица коэффициентов, имеющая k столбцов. Каждая ее строка (будем считать, что число строк равно r) задает линейное ограничение

$$R_{l1}\beta_1 + \cdots + R_{lk}\beta_k = \gamma_l, \quad l = 1, \dots, r.$$

Без ограничения общности можно считать, что строки матрицы ограничений R линейно независимы, так что $r \leq k$ (как правило число ограничений значительно меньше k).

Как и в предыдущем параграфе, мы будем предполагать, что ошибки нормально распределены. Для построения теста проверки гипотезы H_0 воспользуемся тем, что случайный вектор $R\hat{\beta}$ распределен по нормальному закону с математическим ожиданием $R\beta$ и матрицей ковариаций

$$\begin{aligned} \text{cov}(R\hat{\beta}) &= \mathbf{E}(R\hat{\beta} - R\beta)(R\hat{\beta} - R\beta)' = \\ &= R\text{cov}(\hat{\beta})R' = \sigma^2 R(X'X)^{-1}R'. \end{aligned}$$

Легко проверить, что эта матрица невырождена. Действительно, она представляется в виде $R_*R'_*$, где $R_* = R(X'X)^{-1/2}$ — матрица полного ранга r . Отсюда следует, что нормально распределенный вектор

$$(R(X'X)^{-1}R')^{-1/2}(R\hat{\beta} - R\beta)$$

центрирован и имеет матрицу ковариаций $\sigma^2\mathbf{1}_r$. Поэтому нормализованная сумма квадратов его компонент

$$\sigma^{-2}(R\hat{\beta} - R\beta)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - R\beta)$$

распределена по закону χ_r^2 . В предыдущем параграфе установлено, что случайная величина

$$\frac{(N - k)s^2}{\sigma^2}$$

также распределена по хи-квадрат (с $N - k$ степенями свободы) и что она не зависит от вектора оценок $\hat{\beta}$. Вспоминая, что отношение независимых хи-квадрат величин, деленных на соответствующие числа степеней свободы, имеет \mathbf{F} -распределение Фишера, получаем, что в предположении H_0 дробь

$$\frac{(R\hat{\beta} - \gamma)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - \gamma)/r}{s^2}$$

распределена по закону $\mathbf{F}_{r, N-k}$. Большие значения этой дроби образуют критическую область искомого теста. Точно так же, неравенства вида

$$(R\hat{\beta} - \gamma)'(R(X'X)^{-1}R')^{-1}(R\hat{\beta} - \gamma) \leq \text{const}$$

задают совместные доверительные области для компонент вектора $R\beta$, ограниченные эллипсоидами (поверхностями второго порядка). В обоих случаях используются процентные точки \mathbf{F} -распределения.

Описанные в предыдущем параграфе доверительные интервалы укладываются в нашу теперешнюю схему в качестве частного случая, т.к. имеет место "символическое" равенство:

$$(\mathbf{t}_{N-k})^2 = \mathbf{F}_{1, N-k}.$$

Тестирование вызывающей особый интерес гипотезы $\beta_2 = \dots = \beta_k = 0$ детально обсуждается в параграфе 2.10.

2.9 Блочная регрессия

Рассмотрим модель, в которой регрессоры разбиты на два непесекающихся блока:

$$X = (X_{(1)}, X_{(2)}),$$

содержащих, соответственно, k_1 и k_2 регрессоров ($k_1 + k_2 = k$). Для определенности будем предполагать, что $X_{(1)}$ состоит из первых k_1 регрессоров.

Вектор коэффициентов β при этом также разбивается на подвекторы $\beta_{(1)}$ и $\beta_{(2)}$. Мы получим двухэтапную процедуру построения подвектора $\hat{\beta}_{(2)}$ оценок наименьших квадратов, обобщающую схему, изложенную в параграфе 2.4. Важнейший частный случай (ср. с §2.4) — $X_{(1)} = X_1 = \mathbf{1}^{\rightarrow}$, $X_{(2)} = (X_2, \dots, X_k)$, однако мы увидим в дальнейшем, что блочная структура оказывается полезной и совсем в других контекстах.

Запишем формулу (2.6) для оценок наименьших квадратов в блочной форме:

$$\begin{pmatrix} X'_{(1)}X_{(1)} & X'_{(1)}X_{(2)} \\ X'_{(2)}X_{(1)} & X'_{(2)}X_{(2)} \end{pmatrix} \begin{pmatrix} \hat{\beta}_{(1)} \\ \hat{\beta}_{(2)} \end{pmatrix} = \begin{pmatrix} X'_{(1)}Y \\ X'_{(2)}Y \end{pmatrix},$$

так что

$$\begin{aligned} X'_{(1)}X_{(1)}\hat{\beta}_{(1)} + X'_{(1)}X_{(2)}\hat{\beta}_{(2)} &= X'_{(1)}Y, \\ X'_{(2)}X_{(1)}\hat{\beta}_{(1)} + X'_{(2)}X_{(2)}\hat{\beta}_{(2)} &= X'_{(2)}Y. \end{aligned}$$

Поскольку регрессоры первой группы линейно независимы, матрица $X'_{(1)}X_{(1)}$ обратима. Выражая $\hat{\beta}_{(1)}$ из первого уравнения и подставляя во второе, получаем

$$X'_{(2)}X_{(1)}(X'_{(1)}X_{(1)})^{-1}[X'_{(1)}Y - X'_{(1)}X_{(2)}\hat{\beta}_{(2)}] + X'_{(2)}X_{(2)}\hat{\beta}_{(2)} = X'_{(2)}Y.$$

Производя перегруппировку, запишем это равенство в виде

$$\begin{aligned} [X'_{(2)}X_{(2)} - X'_{(2)}X_{(1)}(X'_{(1)}X_{(1)})^{-1}X'_{(1)}X_{(2)}]\hat{\beta}_{(2)} &= \\ &= X'_{(2)}Y - X'_{(2)}X_{(1)}(X'_{(1)}X_{(1)})^{-1}X'_{(1)}Y. \end{aligned}$$

Вводя естественные обозначения

$$P_{(1)} = X_{(1)}(X'_{(1)}X_{(1)})^{-1}X'_{(1)}, \quad P_{(1)}^{\perp} = \mathbf{1} - P_{(1)},$$

получаем

$$X'_{(2)}P_{(1)}^{\perp}X_{(2)}\hat{\beta}_{(2)} = X'_{(2)}P_{(1)}^{\perp}Y,$$

откуда

$$(P_{(1)}^\perp X_{(2)})' (P_{(1)}^\perp X_{(2)}) \hat{\beta}_{(2)} = (P_{(1)}^\perp X_{(2)})' (P_{(1)}^\perp Y). \quad (2.14)$$

Вектор $P_{(1)}^\perp Y$ можно рассматривать как вектор остатков от проектирования Y на подпространство $\mathcal{L}(X_{(1)}) = \mathcal{L}(X_1, \dots, X_{k_1})$. Обозначим его Y_* . Точно так же, столбцы матрицы $P_{(1)}^\perp X_{(2)}$ можно рассматривать как остатки от проектирования регрессоров второй группы на $\mathcal{L}(X_{(1)})$. Обозначим эту матрицу остатков X_* . Тогда (2.14) приобретает вид, сходный с (2.6):

$$X_*' X_* \hat{\beta}_{(2)} = X_*' Y_*. \quad (2.14')$$

Матрица X_* имеет линейно независимые столбцы, в чем легко убедиться, выражая эти столбцы через первоначальные регрессоры X_1, \dots, X_k . Действительно,

$$X_* = X_{(2)} - P_{(1)} X_{(2)} = X_{(2)} - X_{(1)} L,$$

т. к. столбцы матрицы $P_{(1)} X_{(2)}$ — линейные комбинации регрессоров первой группы, т. е. представляются в виде $X_{(1)} L_j$, где L_j — некоторые векторы коэффициентов — столбцы матрицы L . Рассмотрим линейную комбинацию $X_* \gamma$ столбцов матрицы X_* . Она представляется в виде $X_{(2)} \gamma - X_{(1)} L \gamma$ и равна нулю только при $\gamma = 0$ (регрессоры X_1, \dots, X_k линейно независимы).

Из доказанной линейной независимости столбцов X_* следуют обратимость матрицы $X_*' X_*$ и возможность разрешить уравнение (2.14'):

$$\hat{\beta}_{(2)} = (X_*' X_*)^{-1} X_*' Y_*. \quad (2.15)$$

В неявном виде эта разрешимость, конечно, следует из разрешимости системы (2.6) для полного набора оценок $\hat{\beta}$.

Теперь, подводя итог, мы можем интерпретировать изложенную схему следующим образом. На первом шаге процедуры строятся регрессии Y на $X_{(1)}$ и каждого столбца матрицы $X_{(2)}$ на $X_{(1)}$.

На втором шаге строится регрессия остатков Y_* регрессии первого шага на X_* — матрицу остатков остальных регрессий первого шага. Полученные на втором шаге оценки $\hat{\beta}_{(2)}$ — искомые оценки коэффициентов регрессии из второй группы.

Возвращаясь к первой группе коэффициентов, мы можем теперь написать

$$\hat{\beta}_{(1)} = (X'_{(1)}X_{(1)})^{-1}X'_{(1)}(Y - X_{(2)}\hat{\beta}_{(2)}) \quad (2.15_1)$$

Рассмотрим теперь частный случай, упомянутый в начале параграфа — $X_{(1)} = X_1 = 1^{\rightarrow}$. Тогда на первом шаге строятся регрессии на константу, остатками от которых являются векторы отклонений $y = Y - \bar{Y}^{\rightarrow}$, $x_j = X_j - \bar{X}_j^{\rightarrow}$ ($j = 2, \dots, k$). На втором шаге строится регрессия вектора y на укороченный набор новых регрессоров x_2, \dots, x_k . Формулу (2.15) можно записать в виде (x — матрица, составленная из столбцов x_2, \dots, x_k)

$$\hat{\beta}_{(2)} = (x'x)^{-1}x'y \quad (2.16)$$

— оценка коэффициентов линейной регрессии в отклонениях. Для оставшегося коэффициента β_1 теперь легко получаем

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}_2 - \dots - \hat{\beta}_k\bar{X}_k. \quad (2.16_1)$$

Очевидно, (2.15) и (2.16) обобщают ранее полученные формулы (2.10).

Из формул (2.16) получаем также

$$\hat{Y} = \hat{\beta}_1^{\rightarrow} + \hat{\beta}_2X_2 + \dots + \hat{\beta}_kX_k = \bar{Y}^{\rightarrow} + \hat{\beta}_2x_2 + \dots + \hat{\beta}_kx_k. \quad (2.17)$$

Отсюда следует, что $\overline{\hat{Y}} = \bar{Y}$ (для парной регрессии это было получено в параграфе 2.4). Действительно, нужное соотношение непосредственно вытекает из очевидных равенств $\bar{x}_2 = \dots = \bar{x}_k = 0$.

Мы будем использовать блочную регрессию при обсуждении проблем спецификации (см. §2.12).

2.10 Коэффициент детерминации и качество прогноза

В этом параграфе мы предполагаем, что $X_1 = 1^{\rightarrow}$.

Наиболее короткое определение коэффициента детерминации — квадрат выборочного коэффициента корреляции между фактическими (Y) и прогнозными (\hat{Y}) значениями объясняемой величины. Отсюда происходят обозначение R^2 и соответствующая формула. Для вычисления, впрочем, используется несколько иная формула

$$R^2 = \frac{\hat{y}'\hat{y}}{y'y} = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{y'y}, \quad (2.18)$$

которая получается несложными преобразованиями.

Запишем сначала по определению

$$R^2 = \frac{(y'\hat{y})^2}{y'y \cdot \hat{y}'\hat{y}}.$$

Поскольку $\overline{\hat{Y}} = \bar{Y}$, имеем

$$y = Y - \bar{Y}^{\rightarrow} = \hat{Y} + \hat{\varepsilon} - \bar{Y}^{\rightarrow} = \hat{y} + \hat{\varepsilon}.$$

Поэтому

$$y'\hat{y} = (\hat{\varepsilon} + \hat{y})'\hat{y} = \hat{y}'\hat{y}$$

(мы воспользовались ортогональностью остатков $\hat{\varepsilon}$ с прогнозным вектором \hat{Y} и регрессором $X_1 = 1^{\rightarrow}$). Теперь из определения коэффициента детерминации получаем

$$R^2 = \frac{\hat{y}'\hat{y}}{y'y} = \frac{(y - \hat{\varepsilon})'(y - \hat{\varepsilon})}{y'y} = \frac{y'y - \hat{\varepsilon}'\hat{\varepsilon}}{y'y} = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{y'y},$$

что и требовалось доказать.

Если вспомнить, что разложение $Y = \hat{Y} + \hat{\varepsilon}$ определяется не набором регрессоров, а порожденным ими подпространством $\mathcal{L}(X_1, \dots, X_k)$, определение коэффициента детерминации (в любой форме) без изменения переносится на чуть более общий случай

— когда 1^{\rightarrow} лежит в этом подпространстве (но не обязательно является регрессором).

Из определения R^2 непосредственно вытекает неравенство

$$0 \leq R^2 \leq 1.$$

Можно еще отметить, что коэффициент корреляции R между Y и \hat{Y} неотрицателен и сам по себе (без возведения в квадрат), т. к. прогноз \hat{Y} не хуже прогноза без использования регрессоров — посредством \bar{Y}^{\rightarrow} . Крайнее значение $R^2 = 1$ означает совпадение $Y = \hat{Y}$, ожидать этого равенства вряд ли целесообразно. Другое крайнее значение $R^2 = 0$ свидетельствует о незначимом вкладе регрессоров X_2, \dots, X_k в объяснение — см. ниже обсуждение проверки соответствующей гипотезы.

При добавлении в модель новых регрессоров коэффициент детерминации может лишь увеличиться — сумма квадратов остатков уменьшается.

Принято считать, что выражение

$$y'y = \sum_{i=1}^N (Y_i - \bar{Y})^2$$

(оно иногда называется вариацией) характеризует изменчивость величины Y . В этих терминах R^2 показывает, какую часть вариации $y'y$ составляет объясненная моделью часть вариации $\hat{y}'\hat{y}$. Хотя традиционная эконометрика считает коэффициент детерминации достаточно важной характеристикой модели (скажем, его значение вычисляется эконометрическими пакетами), роль коэффициента R^2 не следует преувеличивать. Все авторы учебников подробно объясняют проблемы, возникающие в связи с его использованием.

Во-первых, различные варианты определения перестают совпадать, если константа не лежит в подпространстве регрессоров. Приемлемого определения в этом случае дать не удастся.

Во-вторых, R^2 не инвариантен относительно выбора объясняемой величины. Действительно, возьмем в качестве новой объясняемой величины $Y_* = Y - X\alpha$, где α — некоторый (известный) вектор коэффициентов. Тогда наша модель приобретет вид

$$Y_* = X\beta_* + \varepsilon,$$

причем, очевидно, $\beta_* = \beta - \alpha$. Вектор остатков $\hat{\varepsilon} = P^\perp \varepsilon$ в обоих случаях один и тот же (матрица P^\perp не связана с выбором объясняемой величины). Однако вектор $y_* = y - x\alpha$ совсем не обязан иметь ту же длину, что и y . Поэтому и

$$R_*^2 = 1 - \frac{\hat{\varepsilon}'\hat{\varepsilon}}{y_*'y_*}$$

не обязан совпадать с R^2 . В то же время прогнозные свойства обеих моделей одинаковы:

$$\hat{Y}_* = PY_* = PY - PX\alpha = \hat{Y} - X\alpha.$$

По-существу, мы имеем дело с двумя представлениями одной модели, а не с двумя моделями.

В-третьих, несмотря на кажущуюся объективность этой характеристики качества модели (мы имеем в виду безразмерность R^2), коэффициент детерминации можно сделать сколь угодно близким к единице (или даже равным ей), если присоединить к модели дополнительные регрессоры в достаточном числе. При этом совершенно не требуется, чтобы эта операция имела какой-нибудь содержательный экономический смысл, главное — линейная независимость регрессоров. В учебной литературе обсуждается так называемый подправленный или скорректированный (adjusted) на число регрессоров коэффициент:

$$1 - R_{adj}^2 = \frac{N-1}{N-k}(1 - R^2),$$

который далее использоваться не будет. Убедительного объяснения именно такой формулы для R_{adj}^2 мы не нашли.

Наиболее важным применением коэффициента детерминации является использование его при тестировании значимости регрессионной модели в целом — при проверке гипотезы $H_0 : \beta_2 = \dots = \beta_k = 0$. Опишем это применение более подробно.

Как уже было отмечено выше, о малой значимости регрессии свидетельствуют малые значения R^2 . Остается (предполагая ошибки нормально распределенными) связать с R^2 одно из традиционных шаблонных распределений. Формулы (2.18) позволяют сделать это без труда. Действительно,

$$R^2 = \frac{\hat{y}'\hat{y}}{y'y}, \quad 1 - R^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{y'y}.$$

Деля первое равенство на второе, получаем

$$\frac{R^2}{1 - R^2} = \frac{\hat{y}'\hat{y}}{\hat{\varepsilon}'\hat{\varepsilon}} = \frac{\hat{y}'\hat{y}/\sigma^2}{\hat{\varepsilon}'\hat{\varepsilon}/\sigma^2}. \quad (2.19)$$

При этом для модели с нормально распределенными ошибками числитель и знаменатель последней дроби независимы и распределены по закону χ^2 . Действительно, мы уже проверяли в параграфе 2.7 независимость $\hat{\beta}$ и $\hat{\varepsilon}$, откуда следует независимость \hat{y} и $\hat{\varepsilon}$, а, тем самым, и желаемая независимость числителя и знаменателя. Там же установлено, что $\hat{\varepsilon}'\hat{\varepsilon}/\sigma^2$ распределена по закону χ_{N-k}^2 . Остается разобраться с числителем.

Заметим сначала, что согласно формуле (2.17)

$$\hat{y} = \hat{\beta}_2 x_2 + \dots + \hat{\beta}_k x_k,$$

так что $\mathbf{E}\hat{y} = \beta_2 x_2 + \dots + \beta_k x_k$. Последнее выражение равно нулю в предположении справедливости H_0 . Кроме того, очевидно, вектор \hat{y} нормально распределен. Вычислим, снова в предположении справедливости H_0 , его матрицу ковариаций

$$\text{cov}(\hat{y}) = \mathbf{E}(\hat{y}\hat{y}').$$

Будем при этом использовать обозначение $P_{(2)} = x(x'x)^{-1}x'$ в духе параграфа 2.9. Геометрический смысл матрицы $P_{(2)}$ фактически уже был получен в 2.9 — это — матрица проектирования на $(k-1)$ -мерное подпространство в $\mathcal{L}(X_1, \dots, X_k)$, состоящее из векторов, ортогональных $X_1 = 1^\rightarrow$.

Заметим еще, что согласно формулам (2.17) и (2.16) из §2.9

$$\hat{y} = x\hat{\beta}_{(2)} = x(x'x)^{-1}x'y = P_{(2)}y.$$

Кроме того,

$$y - \mathbf{E}y = (Y - \mathbf{E}Y) - (\bar{Y} - \mathbf{E}\bar{Y})^\rightarrow = \varepsilon - (\bar{\varepsilon})^\rightarrow.$$

Легко сообразить, что $P_{(2)}(\bar{\varepsilon})^\rightarrow = 0$. Поэтому

$$\begin{aligned} \text{cov}(\hat{y}) &= \mathbf{E}[P_{(2)}(\varepsilon - (\bar{\varepsilon})^\rightarrow)(\varepsilon - (\bar{\varepsilon})^\rightarrow)'P_{(2)}] = \\ &= \mathbf{E}[P_{(2)}\varepsilon\varepsilon'P_{(2)}] = P_{(2)}\mathbf{E}[\varepsilon\varepsilon']P_{(2)} = \sigma^2 P_{(2)}. \end{aligned}$$

Теперь утверждение о том, что величина $\hat{y}'\hat{y}/\sigma^2$ распределена по закону χ_{k-1}^2 , доказывается тем же рассуждением, что и аналогичное утверждение для $\hat{\varepsilon}'\hat{\varepsilon}/\sigma^2$ в параграфе 2.7 (напомним, что мы рассуждаем в предположении справедливости гипотезы H_0 , так что $\mathbf{E}\hat{y} = 0$).

Возвращаясь, наконец, к (2.19), заключаем, что дробь

$$\frac{R^2/(k-1)}{(1-R^2)/(N-k)} = \frac{N-k}{k-1} \frac{R^2}{1-R^2}$$

имеет распределение Фишера $\mathbf{F}_{k-1, N-k}$. Остается взять нужную процентную точку \mathbf{F} -распределения и зафиксировать критическую область теста вида

$$\frac{R^2}{1-R^2} \geq \text{const.}$$

Упражнение. Используя блочную регрессию общего вида, обобщить проведенное рассуждение и доказать, что, в предположении

справедливости гипотезы $\beta_{(2)} = 0$, дробь

$$\frac{(R^2 - R_{(1)}^2)/k_2}{(1 - R^2)/(N - k)} = \frac{(\hat{\varepsilon}'_{(1)}\hat{\varepsilon}_{(1)} - \hat{\varepsilon}'\hat{\varepsilon})/k_2}{\hat{\varepsilon}'\hat{\varepsilon}/(N - k)}$$

имеет распределение Фишера $\mathbf{F}_{k_2, N-k}$.

2.11 Индикаторные величины в линейной модели

Индикаторными или сигнальными мы называем величины, принимающие только два значения — 0 и 1 (английский термин — *dummy*; в русскоязычных текстах можно встретить крайне неудачный перевод ”фиктивная” — и неверно по сути, и бессмысленно). Величины такого сорта появляются во многих случаях, когда неоднородность эмпирических данных имеет ”групповой” характер, и мы пытаемся учесть ее, не выходя за рамки классической модели. Рассмотрим несколько стандартных примеров.

Пример 1. Индикатор военного времени. Если эмпирические данные представляют собой временной ряд (например, годовые данные), включающий, скажем, показатели, относящиеся к промежутку между двумя мировыми войнами, к периоду второй мировой войны и к послевоенному периоду, то может оказаться важным выделение военного времени. Это можно сделать следующим образом. Рассмотрим индикаторную величину I , принимающую значение $I_i = 1$ для военных лет, и значение $I_i = 0$ для остальных. С ее помощью каждый регрессор X_j , для которого различия мирного и военного времени кажутся нам существенными, порождает парную величину IX_j , которая включается в линейную модель со своим коэффициентом γ_j . Таким образом, модель включает слагаемые $\beta_j X_j$ и $\gamma_j IX_j$, которые учитывают различия мирного и военного времени на уровне коэффициентов. Для мирных лет в модели

присутствует слагаемое $\beta_j X_j$, а для военных — слагаемое $(\beta_j + \gamma_j) X_j$. Тем самым, некоторым образом показатель X_j ”переключается” с одного режима на другой.

Пример 2. Сезонные колебания. Аналогично примеру 1 можно учесть колебания коэффициентов по месяцам или другим естественным периодам. Для каждого месяца можно ввести свой индикатор: I_1, I_2, \dots, I_{12} . По очевидным причинам сумма этих двенадцати индикаторов тождественно равна единице, так что они линейно зависимы. Поэтому, вводя величины $I_1 X_j, \dots, I_{12} X_j$, мы должны опустить исходную величину X_j . Конечно, в примере 1 можно было бы поступить аналогичным образом.

Общая черта рассмотренных примеров — моменты переключения режимов известны. В примере 1 это не вполне очевидно, т.к. определенные факторы могут иметь последствие. Попытки обобщения вывели бы нас за рамки классической модели, и мы не будем сейчас их обсуждать.

Дискретные величины более чем с двумя значениями, обобщающие индикаторы, практически не используются, т.к. их удобнее заменять более простыми индикаторами, увеличивая при необходимости их число (как в примере 2). Выигрыша в числе параметров, заменяя один способ другим, не добиться.

В качестве иллюстрации использования индикаторных величин рассмотрим так называемый тест Chow проверки совпадения моделей. Предположим, что мы имеем дело с двумя сериями из N_1 и N_2 однотипных наблюдений:

$$Y_{(1)} = X_{(1)}\beta_{(1)} + \varepsilon_{(1)}, \quad Y_{(2)} = X_{(2)}\beta_{(2)} + \varepsilon_{(2)}.$$

Однотипность понимается как совпадение множеств регрессоров в двух сериях (в содержательном смысле — если в первой серии X_2 — процентная ставка, то и во второй серии X_2 — процентная ставка). Будем предполагать также, что дисперсии ошибок

одинаковы (это предположение, вообще говоря, сомнительно, но отказ от него снова выведет нас за рамки классической модели, и потому это обобщение сейчас обсуждаться не будет). Рассмотрим проверку гипотезы $\beta_{(1)} = \beta_{(2)}$. Для этого введем индикатор второй серии I и рассмотрим объединенную систему данных

$$Y = \begin{pmatrix} Y_{(1)} \\ Y_{(2)} \end{pmatrix}, \quad X = \begin{pmatrix} X_{(1)} & IX_{(1)} \\ X_{(2)} & IX_{(2)} \end{pmatrix} = \begin{pmatrix} X_{(1)} & 0 \\ X_{(2)} & X_{(2)} \end{pmatrix}.$$

Соответствующая спецификация имеет вид

$$Y = X\gamma + \varepsilon,$$

где

$$\varepsilon = \begin{pmatrix} \varepsilon_{(1)} \\ \varepsilon_{(2)} \end{pmatrix}, \quad \gamma = \begin{pmatrix} \beta_{(1)} \\ \beta_{(2)} - \beta_{(1)} \end{pmatrix}.$$

Наша гипотеза $\beta_{(1)} = \beta_{(2)}$, или, эквивалентно, $\gamma_{(2)} = 0$, имеет вид, обсуждавшийся ранее, и проверяется (это и есть тест Chow) с использованием распределения $\mathbf{F}_{k, N_1 + N_2 - 2k}$ — см. упражнение в конце параграфа 2.10. При этом коэффициент детерминации R^2 и вектор остатков $\hat{\varepsilon}$ вычисляются по полной регрессионной матрице X , а коэффициент детерминации $R_{(1)}^2$ и вектор остатков $\hat{\varepsilon}_{(1)}$ — по уменьшенной (restricted) матрице

$$X_{rest} = \begin{pmatrix} X_{(1)} \\ X_{(2)} \end{pmatrix}.$$

2.12 Замечания о спецификации модели

На практике исследователь **выбирает** спецификацию модели. Сделать сразу окончательный выбор, как правило, не удастся.

Так, если речь идет о прогнозировании спроса на депозитные сертификаты, можно предполагать, что среди регрессоров окажутся ставка процента по этим сертификатам, ставка процента по каким-либо конкурирующим ценным бумагам и т. д. С уверенностью включать или не включать тот или иной регрессор в модель вряд ли возможно. Поэтому рассматриваются различные варианты модели, с тем чтобы в конечном итоге остановиться на одном из них. В примере с депозитными сертификатами можно попытаться учесть, скажем, разность между ставками процента по краткосрочным и долгосрочным вложениям. Но целесообразно ли это — является ли соответствующий фактор существенным (статистически значимым)? Ответы на подобные вопросы можно получить только после анализа эмпирических данных и сравнения разных модификаций модели. При этом может оказаться, что некоторые регрессоры — лишние, а некоторые, наоборот, пропущены. Мы обсудим в этом параграфе часть подобных вопросов, связанных с выбором спецификации модели.

Начнем с замечаний концептуально-философского характера. Как понимать высказывание о том, что данная модель правильна (true model)? И существует ли вообще таковая? Вопросы "взаимоотношений" между моделью и моделируемым явлением достаточно деликатны. Обсуждаемые нами линейные регрессионные модели включают стохастическую ошибку ε , концентрирующую в себе всю совокупность неучтенных факторов, и потому в самом линейном представлении $Y = X\beta + \varepsilon$ еще нет потенциальных трудностей. Проблемы появляются, когда мы начинаем постулировать какие-либо свойства стохастической ошибки. Проверить (тестировать) постулируемые свойства удастся не всегда, надежность соответствующего вывода может быть невысокой. Надежный же вывод, скорее всего, окажется отрицательным. Таким образом, представление о том, что имеется некоторая "правильная" модель, является (еще одной) идеализацией, появляющейся в процессе мо-

делирования. В этом параграфе мы только начинаем обсуждение проблем спецификации, поэтому будем, все-таки, считать, что "правильную" модель можно представить себе, и для нее выполнены классические предположения.

Будем записывать правильную модель в виде

$$Y = X_t\beta_t + \varepsilon_t; \quad (2.20)$$

здесь индекс t является сокращением от true. Помимо модели (2.20), имеющей только умозрительный характер, исследователь имеет дело с фактической спецификацией $Y = X\beta + \varepsilon$, которая меняется в процессе работы.

Рассмотрим сначала относительно безобидный (как будет видно дальше) случай, когда в спецификацию включены дополнительные ("лишние") регрессоры, так что

$$X = (X_t, X_c),$$

и

$$Y = X_t\beta_{(1)} + X_c\beta_{(2)} + \varepsilon,$$

где $\beta_{(1)}$ и $\beta_{(2)}$ — частичные векторы коэффициентов. Отметим, что правильная модель получается при $\beta_{(2)} = 0$, но нам это неизвестно. Мы, надо думать, считаем, что вектор β , подразумеваемый нашей спецификацией, и есть правильный вектор коэффициентов β_t , что не совсем точно (они имеют разные размерности), и что вектор ошибок ε есть правильный вектор ошибок ε_t — это похоже на истину, впрочем, с оговоркой, что ошибки все-таки не наблюдаемы.

С практической точки зрения мы можем оценить коэффициенты β нашей спецификации стандартным образом, т.е. найти по выборке их оценки $\hat{\beta}$, а также соответствующие остатки $\hat{\varepsilon}$. На самом-то деле наша спецификация ошибочна (точнее, избыточна), так что таковы же и выражения для $\hat{\beta}$ и $\hat{\varepsilon}$. Точнее, частичный вектор $\hat{\beta}_{(1)}$ оценивает вектор β_t правильных коэффициентов, а $\hat{\beta}_{(2)}$

”оценивает” нулевой вектор. При обсуждении блочной регрессии в параграфе 2.9 мы получили формулы (2.15), из которых следует

$$\begin{aligned}\hat{\beta}_{(2)} &= (X_c' P_t^\perp X_c)^{-1} X_c' P_t^\perp Y = (X_c' P_t^\perp X_c)^{-1} X_c' P_t^\perp \varepsilon_t, \\ \hat{\beta}_{(1)} &= \beta_t + (X_t' P_c^\perp X_t)^{-1} X_t' P_c^\perp \varepsilon_t.\end{aligned}$$

Эти оценки несмещенные —

$$\mathbf{E} \hat{\beta}_{(1)} = \beta_t, \quad \mathbf{E} \hat{\beta}_{(2)} = 0,$$

но неправильный выбор спецификации привел к потере в эффективности:

$$\begin{aligned}\text{cov}(\hat{\beta}_{(1)}) &= \sigma^2 (X_t' P_c^\perp X_t)^{-1} \geq \sigma^2 (X_t' X_t)^{-1} = \text{cov}(\hat{\beta}_t), \\ \text{cov}(\hat{\beta}_{(2)}) &= \sigma^2 (X_c' P_t^\perp X_c)^{-1} \geq 0 = \text{cov}(0).\end{aligned}$$

Первое неравенство вытекает из того, что

$$X_t' X_t - X_t' P_c^\perp X_t = X_t' P_c X_t \geq 0,$$

а второе — самоочевидно.

Эффективность — это важное свойство, так что злоупотреблять включением в модель лишних регрессоров не следует. Выявить наличие их поможет проверка гипотезы вида $\hat{\beta}_{(2)} = 0$ — она обсуждалась в параграфе 2.10.

Рассмотрим теперь оценку дисперсии σ_t^2 в рамках выбранной спецификации. Такой оценкой является

$$s^2 = \hat{\varepsilon}' \hat{\varepsilon} / (N - k),$$

где $k = k_1 + k_2$ — полное число коэффициентов. Она, естественно, отличается от

$$s_t^2 = \hat{\varepsilon}_t' \hat{\varepsilon}_t / (N - k_1),$$

но, как это ни парадоксально, обе оценки s_t^2 и s^2 являются несмещенными. Это следует из общих соображений — обе они получаются одной и той же процедурой, только в разных спецификациях.

Первая — в правильной спецификации (2.20), а вторая — в фактически выбранной. Отметим, впрочем, что несмещенность s^2 можно проверить и непосредственно, используя несложно проверяемые соотношения

$$\begin{aligned}\hat{\varepsilon}_t &= P_t^\perp Y = P_t^\perp X_c \hat{\beta}_{(2)} + \hat{\varepsilon}, \\ \hat{\varepsilon}_t' \hat{\varepsilon}_t &= \hat{\varepsilon}' \hat{\varepsilon} + \hat{\beta}_{(2)}' X_c' P_t^\perp X_c \hat{\beta}_{(2)}, \\ \mathbf{E}[\hat{\beta}_{(2)}' X_c' P_t^\perp X_c \hat{\beta}_{(2)}] &= \sigma_t^2 k_2.\end{aligned}$$

Поскольку оценки различаются, а s_t^2 в модели с нормально распределенными ошибками — эффективная несмещенная оценка (см. параграф 2.7), и здесь происходит потеря в эффективности.

Перейдем теперь к более печальной ситуации, когда выбранная спецификация не включает часть регрессоров из правильной модели, т.е. $X_t = (X, X_c)$. Теперь вектор β_t правильных коэффициентов разбивается на два подвектора $\beta_{t(1)}$ и $\beta_{t(2)} \neq 0$. Коэффициенты $\beta_{t(1)}$ отвечают регрессорам, включенным в нашу спецификацию $Y = X\beta + \varepsilon$. Оценки $\hat{\beta}$, которые мы можем построить, предназначаются для оценивания $\beta_{t(1)}$, что же касается коэффициентов $\beta_{t(2)}$, то мы, видимо, и не подозреваем о соответствующих объясняющих факторах, или не считаем их важными.

К сожалению, оценка $\hat{\beta}$, вообще говоря, смещена:

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta_{t(1)} + X_c\beta_{t(2)} + \varepsilon_t) = \\ &= \beta_{t(1)} + (X'X)^{-1}X'X_c\beta_{t(2)} + (X'X)^{-1}X'\varepsilon_t, \\ \mathbf{E}\hat{\beta} &= \beta_{t(1)} + (X'X)^{-1}X'X_c\beta_{t(2)}.\end{aligned}$$

Несмещенной оценка $\hat{\beta}$ оказывается в исключительном случае $X'X_c = 0$, когда столбцы дополнительных регрессоров ортогональны столбцам использованных регрессоров (исключение и есть исключение).

Рассмотрим теперь оценку дисперсии σ_t^2 . Имеем

$$\begin{aligned}\hat{\varepsilon} &= Y - X\hat{\beta} = X\beta_{t(1)} + X_c\beta_{t(2)} + \varepsilon_t - \\ &- X(\beta_{t(1)} + (X'X)^{-1}X'X_c\beta_{t(2)} + (X'X)^{-1}X'\varepsilon_t) = \\ &= P^\perp X_c\beta_{t(2)} + P^\perp \varepsilon_t.\end{aligned}$$

Поэтому

$$\hat{\varepsilon}'\hat{\varepsilon} = \beta_{t(2)}'X_c'P^\perp X_c\beta_{t(2)} + \varepsilon_t'P^\perp \varepsilon_t + 2\beta_{t(2)}'X_c'P^\perp \varepsilon_t.$$

Второе слагаемое имеет требуемое математическое ожидание $(N - k)\sigma_t^2$, третье — вклада не дает, т.к. $\mathbf{E}\varepsilon_t = 0$. Наконец, первое слагаемое, очевидно, практически всегда положительно (даже в случае $X'X_c = 0$, когда оно обращается в $(X_c\beta_{t(2)})'X_c\beta_{t(2)}$). Таким образом,

$$s^2 = \hat{\varepsilon}'\hat{\varepsilon}/(N - k)$$

— смещенная вправо оценка дисперсии.

Обсудим, в завершение параграфа, вопрос о том, как выявить пропуски регрессоров в модели с нормально распределенными ошибками. Для этого заметим, что

$$\mathbf{E}\hat{\varepsilon} = P^\perp X_c\beta_{t(2)}.$$

Предположим сначала, что этот вектор отличен от нуля. В этом случае можно (как и в параграфе 2.7) выбрать некоторый ортонормированный базис e_1, \dots, e_{N-k} в подпространстве $\mathcal{L}^\perp(X_1, \dots, X_k)$, где лежит $\hat{\varepsilon}$, составить из этих векторов-столбцов матрицу e и рассмотреть вектор $e'\hat{\varepsilon}$ с координатами $e'_j\hat{\varepsilon}$. Очевидно,

$$\mathbf{E}(e'\hat{\varepsilon}) = e'P^\perp X_c\beta_{t(2)} = e'X_c\beta_{t(2)}.$$

Кроме того, доказанная в параграфе 2.7 формула

$$\text{cov}(e'\hat{\varepsilon}) = \sigma^2 \mathbf{1}_{N-k},$$

очевидно, справедлива и сейчас (единственное отличие, нецентрированность $e'\hat{\varepsilon}$, не играет роли при вычислении ковариаций).

Среднее арифметическое случайных величин $e'_j \hat{\varepsilon}$

$$\frac{1}{N-k} (1^{\rightarrow})' e' \hat{\varepsilon} \quad (2.21)$$

представляется в виде суммы своего математического ожидания

$$\frac{1}{N-k} (1^{\rightarrow})' e' X_c \beta_{t(2)} \quad (2.22)$$

и среднего арифметического $N-k$ независимых величин с распределением $N(0, \sigma^2)$. Поэтому можно построить доверительный интервал для (2.22) — среднего значения нормально распределенной с дисперсией $\sigma^2/(N-k)$ случайной величины (2.21).

Конечно, дисперсия σ^2 нам не известна, но если воспользоваться завышенной (см. выше) оценкой s^2 , то мы будем лишь несколько реже отвергать гипотезу $\beta_{t(2)} = 0$ и следствия из нее, но, все-таки, при удачном стечении обстоятельств сможем выявить отсутствие центрированности для вектора $e' \hat{\varepsilon}$.

Если математическое ожидание (2.22) обращается в 0, наш прием непригоден. В этом случае можно попытаться сменить базисную матрицу e .

Большим недостатком указанного метода является необходимость строить матрицу e — это трудоемкая вычислительная задача. В некоторых случаях можно не доводить построение базиса $\{e_j\}$ до конца и ограничиться несколькими первыми базисными векторами.

Вернемся теперь к случаю $E\hat{\varepsilon} = 0$, когда не поможет никакое изменение матрицы e . В этой ситуации можно попытаться уменьшить выборку, отбрасывая одно или несколько наблюдений. В любом случае, можно надеяться, что возможный пропуск регрессоров вскрыется после нескольких попыток. А на "нет", как говорят в статистике, и суда нет.

И, наконец, последнее замечание. Предположим, что мы выявили нечто, похожее на пропуск регрессоров. Ведь это всего-

лишь сигнал о том, что "что-то не в порядке". Предположений в нашей модели довольно много, и, может быть, нарушается одно из свойств ошибок. На этой вопросительной ноте мы заканчиваем параграф и главу.