МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ УКРАИНЫ
СУМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

А. М. Назаренко

ОСНОВЫ ЭКОНОМЕТРИКИ

РЕКОМЕНДОВАНО УЧЕНЫМ СОВЕТОМ
СУМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
КАК УЧЕБНОЕ ПОСОБИЕ
для студентов экономических специальностей

СУМЫ ИЗД-ВО СУМГУ 2003

ВВЕДЕНИЕ

Экономическая теория изучает причинно-следственные отношения между явлениями и процессами на качественном уровне. Однако существующие явления и процессы связаны между собой, и нахождение зависимости и взаимосвязей между ними путем создания математических моделей и последующего их количественного описания позволяет глубже понять существующие закономерности.

Математическое моделирование получило широкое распространение в различных областях знаний: механике, физике, медицине, биологии, химии, в том числе и в экономике. Различают детерминистическое и стохастическое моделирование. Если в отношении величин, анализируемых детерминистическими моделями, предполагается их стабильность, а случайными отклонениями пренебрегают, относя их только на счет ошибок наблюдений и измерений, то в основе стохастических моделей лежит случайный характер величин, оцениваемый вероятностными методами. Здесь задача заключается в отыскании тенденций, которые проявляются в случайных отклонениях.

Детерминистические модели описывают закономерности, проявляющиеся в одиночном, в каждом отдельно взятом элементе совокупности. Этой закономерности присуща жесткая механическая причинность, которая конкретно определяет поведение каждой единицы совокупности. Такая закономер-ность получила название динамической или закономерности с жесткой детерминацией. Типичными примерами динамических закономерностей могут служить законы классической механики. В динамических закономерностях связь между причиной и следствием может быть выражена вполне точно в виде конкретных математических формул. Здесь каждому набору значений объясняющих переменных всегда соответствует определенное значение объясняемой переменной. Такая связь называется функциональной. Детерминистическая модель служит выражением функциональной связи.

В централизованной плановой экономике вполне обходились детерминистическими моделями. Здесь заранее был известен результат, и теория предназначалась для его обоснования. В основном использовались балансовые или оптимизационные модели: межотраслевого баланса и линейного программирования. Однако в условиях рыночной экономики результат заранее неизвестен, и не учитывать фактор случайности просто невозможно. Экономические явления и процессы представляют собой результат многих одновременно и совокупно действующих причин. При рассмотрении связей между ними главные причины, которые обязательно приводят к данному следствию, необходимо отличать от второстепенных. Последние усложняют и искажают действие существенных в данном аспекте причин. Кроме того, причины могут иметь и непредвиденный характер. Так, например, в банковской деятельности ежедневные денежные потоки формируются под влиянием определенных закономерностей (запланированные платежи), а также необязательных, а порой и непредвиденных поступлений или платежей. Следовательно, экономические процессы имеют вероятностный характер, а развитие исследуемого объекта определяется суммарным влиянием закономерности и случайности.

Чтобы отделить существенные факторы, действующие на исследуемый объект, от второстепенных и случайных, наблюдения должны быть многократными, массовыми. Закономерности, проявляющиеся при массовых наблюдениях, называются *статистическими*. Статистические закономерности также причинно обусловлены, как и динамические, только причин может быть множество, они взаимно переплетены и действуют в разных направлениях. Вероятность получения конкретного результата здесь равна нулю. В подобных ситуациях можно лишь найти интервал, в который попадает значение иссле-дуемого показателя с наперед заданной вероятностью. Выявление статистических закономерностей, определение интервальных оценок неизвестных параметров и проверка различных гипотез осуществляются методами математической статистики.

Стохастические модели описывают закономерности, обусловленные одновременным действием на объект многих факторов и проявляющиеся отчетливо только при массовых наблюдениях. К наиболее распространенным методам построения стохастических моделей относятся методы, объединенные под общим названием – многомерный статистический анализ, в частности, корреляционный и регрессионный анали-зы. Практика показывает, что стохастические модели, получен-ные с помощью корреляционного и регрессионного анализов, имеют преимущество при количественном описании причинно-следственных отношений в экономике и социальной сфере по сравнению с детерминистическими моделями. Выявление количественных соотношений в виде регрессии дает возможность лучше понять природу исследуемого явления. А это, в свою очередь, позволяет воздействовать на выявленные факторы, вмешиваться в соответствующий экономический процесс с целью получения нужных результатов.

Классический регрессионный анализ описывает экономические процессы с помощью одного уравнения регрессии. Это уравнение не функциональное, а стохастическое. В нем каждому набору объясняющих переменных может соответствовать сразу несколько значений объясняемой переменной. В уравнении должны присутствовать только существенные объясняющие переменные. Неконтролируемые или неучтенные факторы, а также ошибки измерения включаются в случайный член (случайное возмущение). Предполагается, что объясняющие переменные не случайны и не коррелируют между собой, а случайная составляющая имеет диагональную дисперсионно-ковариационную матрицу с равными диагональными элементами (дисперсиями).

Как показали дальнейшие исследования, описания экономических процессов с помощью одного уравнения регрессии явно недостаточно в силу многообразного переплетения причин и следствий. Для более адекватного отражения многосторонних реальных взаимоотношений в экономических процессах необходимо использовать систему регрессионных уравнений. Применение разработанных и вполне обоснованных тестов для проверки гипотезы о виде дисперсионно-ковариационной матрицы случайного возмущения показало, что рассчитанные значения случайного члена уравнения регрессии во многих случаях (особенно при анализе временных рядов) отвергают основные предположения классического регрессионного анализа. Идея о взаимосвязях между экономическими переменными, а также предположение об общем виде дисперсионно-ковариационной матрицы случайного возмущения привели к созданию нового типа стохастических моделей, которые стали называться *эконометрическими*.

Впервые термин “*эконометрия*” (Öconometrie) появился в немецкой книге по бухгалтерскому учету, автор которой понимал под ним теорию бухгалтерии. В экономическую науку этот термин ввел в оборот норвежский статистик Рагнар Фриш (впоследствии лауреат Нобелевской премии в области экономики) в докладе, опубликованном им в 1928 г., для обозначения самостоятельной отрасли научных исследований. По иници-ативе Р. Фриша 29 декабря 1930 г. было образовано эконометрическое общество, которое стало издавать ежемесячный журнал «Эконометрика» (“Econometrica”). Основной задачей общество объявило «развитие экономической теории в ее связи со статистикой и математикой». Круг задач, решаемых эконометрикой, непрерывно расширяется. Экономисты, убедившись в достоверности получаемых эконометристами результатов, довольно быстро признали эконометрику в качестве самостоятельной научной дисциплины. Приведем высказывания признанных авторитетов в экономике и эконометрике.

«Эконометрика занимается определением наблюдаемых в экономической жизни конкретных количественных закономерностей, применяя для этой цели статистические методы» (О.Ланге).

«Эконометрика – мощный инструмент для обнаружения наиболее устойчивых характеристик в поведении реальных экономических единиц» (Э.Маленво).

«Эконометрика позволяет проводить количественный анализ реальных экономических явлений, основываясь на современном развитии теории и наблюдениях, связанных с методами получения выводов» (Самуэльсен).

«Предметом эконометрики является экономика в количественном аспекте. Объект изучения – связи между народнохозяйственным ансамблем и его простейшими компонентами и соответственно поведение последних» (Т.Шаттелес).

«Основная задача эконометрики – наполнить эмпирическим содержанием априорные экономические рассуждения» (Л.Клейн).

«Цель эконометрики – эмпирический вывод экономических законов. Эконометрика дополняет теорию, используя реальные данные для проверки и уточнения постулируемых отношений» (Э.Маленво).

«Экономисты используют количественные данные для наблюдения за ходом развития экономики, ее анализа и прогнозов. Набор статистических методов, используемых для этих целей, называется в совокупности эконометрикой» (Ц.Грилихес).

«Современное университетское экономическое образование держится на трех китах: макроэкономике, микроэкономике и эконометрике» (В.Л. Макаров).

Эконометрия, эконометрика – одно из направлений экономико-математических методов анализа, которое заключается в статистическом измерении (оценивании) параметров, характеризующих некоторую экономическую концепцию о взаимосвязи и развитии объекта или явления, и в применении таким путем эконометрических моделей для конкретных экономических выводов.

Данное пособие является вводным в курс эконометрики. В нем рассматриваются модели, описываемые одним регрессионным уравнением. Однако и на простейших примерах можно понять основные идеи эконометрики и трудности, которые возникают при эконометрическом моделировании.

ГЛАВА 1 МОДЕЛИ С ДВУМЯ ПЕРЕМЕННЫМИ

1.1 Однофакторные производственные функции. Основные предположения

Возможности любого производства определяются характером зависимости между объемом выпускаемой продукции и соответствующими ему затратами сырья, энергии, капиталовложений, труда и т. п. Всевозможные виды затрат называют факторами производства или ресурсами. Факторы выражаются в килограммах, метрах, киловатт-часах и других единицах. Общей единицей измерения может служить гривна или любая другая денежная единица. Поэтому удобно иметь дело со стоимостным выражением как факторов производства, так и выпускаемой в результате их использования продукции.

Функцию, выражающую зависимость между стоимостью выпускаемой продукции и стоимостью суммарных затрат на ее производство, называют *однофакторной производственной функцией*. Здесь все виды затрат объединены в единый фактор производства - суммарную стоимость затрат, все разнообразие выпускаемой продукции описано с помощью одной величины - суммарной стоимости выпуска (или уровня выпуска). Функция, в которой роль независимой переменной играют затраты, а зависимая переменная определяет уровень выпуска, называется *функцией выпуска*. *В функции затрат*, наоборот, независимая переменная - выпуск, а зависимая - затраты.

С помощью однофакторных производственных функций описывается также зависимость объема выпускаемой продукции от затрат некоторого специфического вида ресурса. В роли такого ресурса могут выступать трудовые ресурсы, основные производственные фонды, объем капиталовложений, различные виды ресурсов. При этом затраты всех других участвующих в производстве ресурсов считаются постоянными.

Функциями покупательского спроса называются функции, отражающие зависимость объема спроса на отдельные товары и услуги от комплекса факторов, влияющих на него. Такие функции применяются в аналитических моделях спроса и строятся на основе информации о структуре доходов населения, ценах на товары, составе семей и других факторов. Если рассматривать функцию спроса в зависимости от одного факто-ра (дохода или цены), то она называется *однофакторной функцией спроса*.

В настоящее время особенностью математико-статистического и количественно - финансового анализа, а также прогнозирования финансово - экономических показателей является учет временного фактора. Этот фактор, особенно при планировании и прогнозировании долгосрочных операций, играет не меньшую роль, чем размеры денежных сумм. В процессе инвестирования и кредитования это связано, прежде всего, с тем, что теоретически любая сумма денег может быть инвестирована под проценты и принести доход. Таким образом, учет единственного фактора - времени в финансово – экономическом анализе и прогнозировании приводит к *однофактор-ной функции времени*.

В данной главе рассматриваются простейшие однофакторные модели, в которых устанавливается связь в виде одного уравнения результативного признака Y (объясняемой переменной) от одного факторного признака X (объясняющей переменной).

Первое (исходное) предположение состоит в том, что между переменными Y и X постулируется связь

$$Y = F(X). \tag{1.1}$$

Соотношение (1.1) означает, что выполнена идентификация переменной X как воздействующей на переменную Y.

Большая часть традиционных экономических теорий, в которых связи между экономическими категориями отражаются с помощью алгебраических формул, имеет дело с точными функциональными соотношениями. Однако даже самое элементарное знакомство с экономическими данными показывает, что их отдельные значения не могут укладываться на некоторую гладкую линию. Зависимость (1.1) не является функциональной, где каждому возможному значению X ставится в однозначное соответствие определенное значение Y. Она является стохастической и предполагает, что для заданного значения независимой переменной X можно указать ряд значений зависимой переменной Y, случайно рассеянных в некотором интервале. Пример такой диаграммы рассеяния представлен на рис. 1.1.

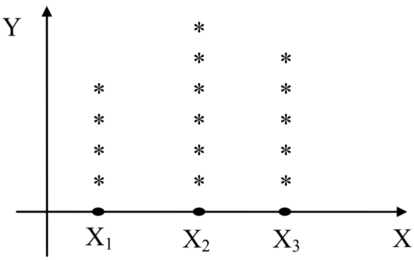


Рисунок 1.1

Стохастическая зависимость между Y и X связана с тем, что зависимая переменная, кроме выделенной переменной X, подвергается влиянию также ряда неконтролируемых или неучтенных факторов, а также тем, что накладываются ошибки измерения. Поскольку значения зависимой переменной подвержены случайному разбросу, они не могут быть предсказаны с достаточной точностью, а только с определенной вероятностью. Появляющиеся значения зависимой переменной Y являются реализациями случайной величины. Случайной называется величина, которая принимает в результате испытания то или иное (но при этом только одно) возможное значение, заранее неизвестное и меняющееся от испытания к испытанию.

Стохастическая зависимость (1.1) является односторонней и называется регрессией. *Регрессия* устанавливает зависимость одной случайной переменной от другой (парная регрессия) или нескольких случайных переменных (множественная регрессия). Односторонняя стохастическая зависимость выражается с помощью функции, которая, для отличия ее от строгой математической функции, называется функцией регрессии или просто регрессией. Существует принципиальное различие между функциональной зависимостью и регрессией. При функциональной зависимости аргумент X полностью определяет значение функции Y. Кроме того, при функциональной зависимости функция может быть обратима (функция $X=\sqrt[3]{Y}$ является обратной к функции $Y = X^3$). Функция регрессии этим свойством не обладает. Учитывая это, различают регрессию Y на X (если исследуют стохастическую зависимость Y от X) и регрессию X на Y (если изучают стохастическую зависимость X от Y). Например, при изучении механизма связи между ценой товара (X) и спросом (Y) практический интерес представляют обе постановки задачи: зависимость цены товара от спроса, а также обратная зависимость - спроса от цены товара, так как измене-ние цен на товары отражается на спросе населения. В этом случае отдельно строятся регрессии Y на X и X на Y. Обе переменные являются здесь случайными. Каждому значению X соответствует множество значений Y и, наоборот, каждому значению Y соответствует множество значений X. Если же логическое истолкование зависимости между двумя переменными возможно только в одном направлении, как, например, при исследовании влияния количества внесенных удобрений или выпавших осадков (X) на урожайность сельскохозяйственных культур (Y), то находится только одна функция регрессии Y на X. Стохастическая зависимость X от Y в данном случае не имеет смысла.

Второе предположение состоит в осуществлении спецификации формы связи между Y и X. Содержательные соображения, на основе которых было принято соотношение (1.1), должны прояснить и его конкретную функциональную форму или подсказать дополнительные условия, которым должны удовлетворять параметры модели. Поскольку одним и тем же условиям могут удовлетворять различные функции, придется обратиться к статистическому анализу и с его помощью осуществить выбор среди возможных альтернативных вариантов. В случае парной регрессии (стохастической зависимости результативной переменной Y от одной объясняющей переменной X) задача заключается в установлении вида функции регрессии

$$\breve{y} = f\,(X, a_0, a_1, ..., a_m), \qquad (1.2)$$

где $a_0, a_1, ..., a_m$ - неизвестные параметры.

Нахождение функции регрессии происходит по эмпирическим данным (табл. 1.1), содержащим случайности и влияние второстепенных причин, которые своей изменчивостью затушевывают и искажают истинную связь. В силу того, что случайности и второстепенные факторы не могут быть исключены из опытных данных, зависимость между Y и X приобретает стохастический характер.

Т а б л и ц а 1.1

Переменная	Наблюдение			
X	X_1	X_2	. . .	X_n
Y	Y_1	Y_2	. . .	Y_n

Случайная величина

$$U = Y - \check{y} \tag{1.3}$$

характеризует отклонение переменной Y от величины $\check{y}$, вычисленной по функции регрессии (1.2), и называется ***возмущением*** (***возмущающей переменной***). Она включает влияние неучтенных факторов - переменных, случайных помех и ошибок наблюдений. Ее значение меняется для каждого наблюдения. Из (1.2) и (1.3) вытекает, что зависимую переменную Y можно представить в виде

$$Y = f\,(X, a_0, a_1, \ldots, a_m) + U, \tag{1.4}$$

где $a_0, a_1, \ldots, a_m$ - неопределенные коэффициенты. Такой вид записи позволяет также интерпретировать случайную переменную U как учитывающую неправильную спецификацию функции регрессии, т. е. неправильный выбор формы уравнения, описывающего искомую зависимость. Благодаря введению случайной переменной U, переменная Y также становится случайной, поскольку при заданных значениях X ей нельзя поставить в соответствие только одно определенное значение. В то время как исследователь располагает значениями объясняемой и объясняющей переменных в результате совместных наблюдений над ними, значения возмущающей переменной U непосредственно получить нельзя, поскольку она представляет собой конгломерат многих трудноучитываемых и случайных влияний. Лишь после количественной оценки функции регрессии можно найти значение возмущающей переменной U**.**

Третье предположение заключается в возможности оценивания параметров $a_0,$ $a_1,\ldots,$ a_m функции регрессии (1.2). Классический регрессионный анализ оценок $\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_m$ параметров $a_0,$ $a_1,\ldots,$ a_m уравнения регрессии (1.4) основан на методе наименьших квадратов (см. § 1.5), который устанавливает некоторые ограничения на вид функции регрессии:

а) параметры a_0, $a_1,\ldots,$ a_m должны входить в выражение для функции регрессии в уравнении (1.4) линейно, т. е.

$$Y = a_0 + a_1 f_1(X) + \ldots + a_m f_m(X) + U\,; \qquad (1.5)$$

б) если в качестве исходной модели выбрана нелинейная относительно параметров модель, то предполагается, что и возмущение U входит в нее нелинейно:

$$Y = f(X, a_0, a_1, \ldots, a_m, U). \qquad (1.6)$$

Кроме того, должно существовать некое преобразование над уравнением (1.6), такое, что

$$g(Y) = g(f(X, a_0, a_1, \ldots, a_m, U)) = a_0^* + a_1^* g_1(X) + \ldots + a_m^* g_m(X) + U.$$

Тогда уравнение

$$Y^* = a_0^* + a_1^* g_1(X) + \ldots + a_m^* g_m(X) + U \qquad (1.7)$$

будет линейным (относительно параметров и возмущения) уравнением модели. Нелинейное уравнение (1.6) идентично исходной модели; линейное регрессионное уравнение (1.7) является важнейшей составной частью преобразованной оцениваемой модели. Часто выполняемыми преобразованиями являются, например, логарифмирование и нахождение обратного значения.

Пример 1 Пусть необходимо охарактеризовать зависимость урожайности Y некоторой сельскохозяйственной культу-ры от количества X внесенных удобрений. В данном случае в качестве однофакторной производственной функции можно выбрать функцию вида

$$\breve{y} = a_0 + a_1 X - a_2 X^2 \quad (a_0>0, \ \ a_1>0, \ \ a_2>0). \qquad (1.8)$$

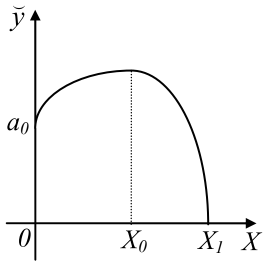


Рисунок 1.2

При отсутствии удобрений урожайность составляет a_0 единиц. С увеличением количества внесенных удобре-ний происходит сначала возрастание урожайности, и в точке $X = X_0$ она достигает наибольшего значения. Дальнейшее наращивание затрат удобрений оказывается неразумным. Оно приводит к снижению урожайности и даже полной ее потере при $X = X_1$ (рис. 1.2).

Функция регрессии (1.8) зависит от неизвестных коэффициентов a_0, a_1, a_2 линейно и, следовательно, согласно (1.5) связь случайных переменных Y и X необходимо задать в виде

$$Y = a_0 + a_1X - a_2X^2 + U, \tag{1.9}$$

где U - случайное возмущение.

Пример 2 При моделировании ситуаций, в которых рост затрат X некоторого ресурса R ведет к неограниченному увеличению выпуска Y, может быть использована степенная производственная функция

$$\breve{y} = a_0 X^{a_1} \quad (a_0 > 0, \quad a_1 > 0). \tag{1.10}$$

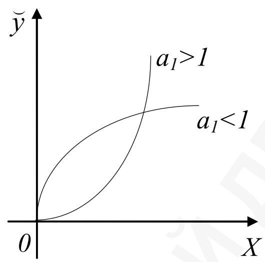


Рисунок 1.3

Насколько быстро растет выпуск Y, зависит, очевидно, от величины параметров a_0 и a_1. При $a_1 > 1$ степенная функция (1.10) является выпуклой, при $a_1 < 1$ - вогнутой (рис. 1.3).

Исходная модель (1.10) – нелинейна относительно параметров a_0 и a_1. В этом случае согласно (1.6) и (1.7) необходимо положить

$$Y = a_0 X^{a_1} \cdot e^U. \tag{1.11}$$

Прологарифмировав (1.11), получаем

$$lnY = ln\, a_0 + a_1\, lnX + U,$$

и уравнение

$$Y^* = a_0^* + a_1\, lnX + U, \qquad Y^* = lnY \tag{1.12}$$

становится линейным относительно параметров $a_0^* = ln\, a_0$ и a_1, а также возмущения U.

Пример 3 Функция

$$\breve{y} = \frac{X}{a_0 + a_1 X} \quad (a_0 > 0,\ a_1 > 0) \tag{1.13}$$

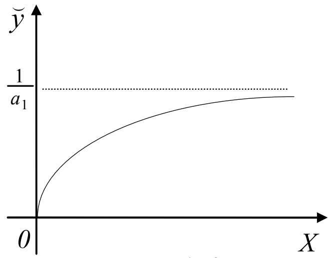


Рисунок 1.4

может моделировать влияние затрат X переменного ресурса R на выпуск Y продукции, если уровень выпуска не может быть больше некоторой предельной величины. Кривая (1.13) является вогнутой. При отсутствии затрат выпуск равен нулю. Если $X \to \infty$, $\breve{y} \to \frac{1}{a_1}$, т.е. при неограниченном возрастании затрат X выпуск Y не может быть больше величины $\frac{1}{a_1}$. Запишем исходное уравнение (1.13) в виде $\breve{y} = \frac{1}{a_1 + \frac{a_0}{X}}$. Тогда если положить

$$Y = \frac{1}{a_1 + \frac{a_0}{X} + U}, \tag{1.14}$$

то преобразованное уравнение

$$Y^* = a_1 + \frac{a_0}{X} + U, \quad Y^* = \frac{1}{Y} \tag{1.15}$$

будет линейным относительно параметров и возмущения.

Четвертое предположение заключается в том, что исходная спецификация взаимосвязи между переменными (1.5) или преобразованная (1.7) должна сопровождаться рядом гипотез о свойствах распределения вероятностей для случайного возмущения. Эти гипотезы относятся к его математическому ожиданию, дисперсии и ковариациям. Простейший из возможных вариантов - принять математическое ожидание, равным нулю, дисперсию - постоянной и не зависящей от *X,* а различные значения U - не зависящими друг от друга. Если i - номер наблюдения, т. е. номер уровня ряда в табл. 1.1, и U_i – соответствующее этому уровню случайное возмущение, то упомянутые гипотезы можно записать в виде

$$M(U_i) = 0, \tag{1.16}$$

$$M(U_i U_j) = \begin{cases} 0, \ \text{если} \ i \neq j, \\ \sigma_u^2, \ \text{если} \ i = j, \quad (i, j = 1,2,\dots,n). \end{cases} \tag{1.17}$$

Здесь σ_u^2 - неизвестная, но постоянная величина.

1.2 Однофакторные функции покупательского спроса. Коэффициенты эластичности

В практике планирования и прогнозирования спроса на отдельные товары и услуги часто используются аналитические модели спроса и потребления, которые строятся в виде уравнений, характеризующих зависимость потребления товаров и услуг от тех или иных факторов. Такие модели могут быть однофакторными и многофакторными. Рассмотрим однофакторные модели зависимости спроса от дохода или цены.

Основным фактором при анализе покупательского спроса является доход населения. Однофакторные функции спроса Y на отдельный товар от дохода Z изучались немецким экономистом

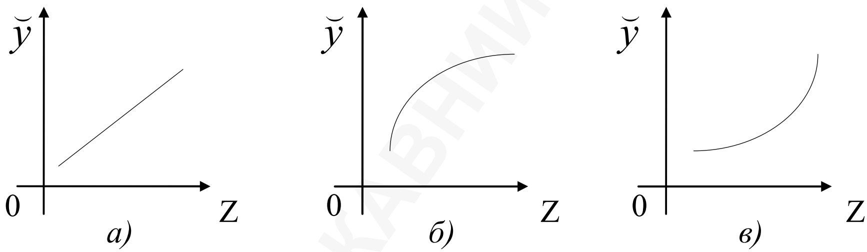


Рисунок 1.5

Энгелем, и соответствующие этим функциям кривые $\breve{y} = f\ (Z)$ называются кривыми Энгеля. Формы этих кривых для различных товаров могут быть различны. Если спрос на данный товар возрастает примерно пропорционально доходу, то функция будет линейной (рис. 1.5*а*). Такой характер имеет, например, спрос на одежду, фрукты. Если рост значений спроса, начиная с определенного момента, по мере насыщения спроса, отстает от роста дохода, то кривая Энгеля имеет вид вогнутой кривой (рис.1.5*б*). Например, таков характер спроса на товары первой необходимости. Наконец, если по мере роста доходов спрос на данную группу товаров возрастает все более высокими темпами, то кривая Энгеля выпуклая (рис. 1.5*в*). Так изменяется спрос на предметы роскоши.

Тот же принцип разграничения групп товаров по типам функций спроса от дохода использовал шведский экономист Л.Торнквист, который предложил специальные виды функций спроса (функции Торнквиста) для трех групп товаров: первой необходимости, второй необходимости и предметов роскоши.

Функция Торнквиста для товаров первой необходимости имеет вид

$$\breve{y} = \frac{a_1 \cdot Z}{Z + C_1} \qquad (1.18)$$

и отражает тот факт, что рост спроса на эти первоочередные товары с ростом дохода постепенно замедляется и имеет предел a_1. График функции является вогнутой кривой. Функция спроса по Торнквисту на товары второй необходимости выражается формулой

$$\breve{y} = \frac{a_2 (Z - b_2)}{Z + C_2}\ , \quad Z \geq b_2. \qquad (1.19)$$

Эта функция также имеет предел a_2, но более высокого уровня, при этом спрос на эту группу товаров появляется лишь после того, как доход достигнет величины b_2; график функции – вогнутая кривая. Наконец, функция Торнквиста для предметов роскоши имеет вид

$$\breve{y} = \frac{a_3 Z (Z - b_3)}{Z + C_3}\ , \quad Z \geq b_3. \qquad (1.20)$$

Эта функция не имеет предела. Спрос на эти товары возникает только после того, как доход превысит b_3 и далее быстро возрастает так, что график функции является выпуклой кривой. Кривые Торнквиста представлены на рис. 1.6.

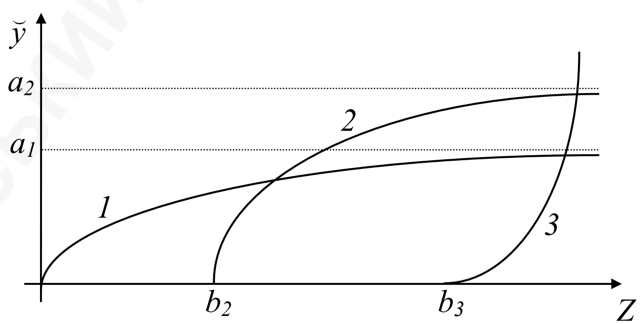


Рисунок 1.6

Важную роль в анализе изменения спроса при небольших изменениях дохода играет коэффициент эластичности. *Коэффициент эластичности спроса от дохода* показывает, на сколько процентов изменятся спрос на товар при изменении дохода на один процент. Следует отметить, что коэффициенты эластичности являются важнейшими характе-ристиками производственных функций. Необходимость выясне-ния, на сколько процентов изменится одна величина, если дру-гая изменилась на 1%, возникает при исследовании диапазона взаимозаменяемости ресурсов производства, определении эффективности тех или иных затрат, прогнозировании изменения прибыли предприятия или фирмы под воздействием различных факторов и решении многих - многих других проблем.

Если производственная функция, характеризующая связь переменных Y и X, имеет вид $\breve{y} = f(X)$, то коэффициент эластичности вычисляется по формуле

$$\mathrm{E_X} = \lim_{\Delta \mathrm{X} \to 0} \left\lfloor \left(\frac{\Delta \breve{\mathrm{y}}}{\breve{\mathrm{y}}} \cdot 100\%\right) : \left(\frac{\Delta \mathrm{X}}{\mathrm{X}} \cdot 100\%\right) \right\rfloor,$$

т. е.

$$E_X = \frac{X}{\breve{y}} \cdot \frac{d\breve{y}}{dX} = \frac{X}{\breve{y}} \cdot f'(X). \qquad (1.21)$$

Он показывает на сколько процентов изменится значение $\breve{y}$ при увеличении переменной X на 1%. Например, если спрос на товар первой необходимости описывается функцией Торнквиста (1.18), то выражение для коэффициента эластичности спроса от дохода имеет вид

$$E_z = \frac{C_1}{Z + C_1}.$$

Коэффициенты эластичности спроса от дохода различны для разных товаров; они могут быть и отрицательны, если с ростом доходов потребление уменьшается. Принято выделять четыре группы товаров в зависимости от коэффициента эластичности спроса на них от дохода: малоценные товары *($E_Z < 0$)*, товары с малой эластичностью *($0 < E_Z < 1$)*, товары со средней эластичностью *($E_Z \approx 1$)* и товары с высокой эластичностью *($E_Z > 1$)*.

По результатам обследований коэффициенты эластич-ности для основных продуктов питания находятся в интервале от 0,4 до 0,8; по одежде, тканям, обуви - в интервале от 1,1 до 1,3 и так далее. По мере увеличения дохода спрос перемещается с товаров первой и второй группы на товары третьей и четвер-той групп, при этом потребление товаров первой группы по абсолютным размерам сокращается.

Рассмотрим теперь однофакторные функции спроса на отдельный товар от цены этого товара. Предполагаем, что цены на другие товары, доходы потребителей и структура их потребностей - постоянные величины. Зависимость спроса Y от цены P можно описать с помощью однофакторной функции $\breve{y}$ =g (P). Для многих экономических исследований необходимо установить не величину спроса при определенном изменении цены, а эластичность спроса относительно цены, т. е. значение коэффициента

$$E_P = \frac{P}{\breve{y}}\frac{d\breve{y}}{dP} = \frac{P}{\breve{y}}g'(P)\,.$$

Коэффициент эластичности спроса относительно цены определяет на сколько процентов изменится спрос на товар, если цена на него увеличится на 1%. Для большинства товаров действует зависимость: чем выше цена, тем ниже спрос, и наоборот, то есть спрос является убывающей функцией цены: $g'(P) < 0$. Чтобы избежать отрицательных чисел, в этих случаях при изучении эластичности спроса принимают

$$E_P = -\frac{P}{\breve{y}}g'(P).$$

Знак “-” показывает здесь, что спрос уменьшается при увеличении цены.

В зависимости от значений коэффициента эластичности спроса относительно цены, товары разделяют на три группы:

а) товары с неэластичным спросом *(*$0 < E_P < 1$*)*, т.е. спрос неэластичен, если повышение цены на 1% влечет за собой понижение спроса менее чем на 1%. Таковым является спрос на товары первой необходимости (лекарства, обувь, электричество, газ, телефон), на вещи, цена которых мало ощутима для семейного бюджета (карандаш, зубная паста, крем для обуви) и труд-но заменяемые товары (электрические лампочки, хлеб, бензин);

б) товары со средней эластичностью спроса от цены $(E_P \approx 1)$, в частности, если $E_p = 1$, то спрос нейтрален;

в) товары с эластичным спросом $(E_P > 1)$, то есть спрос эластичен, если повышению цены на 1% соответствует снижение спроса более чем на 1%. Другими словами, спрос на товар эластичен, если небольшое изменение цены товара вызывает значительное изменение величины спроса на него. Примерами товаров с эластичным спросом могут служить, например, яблоки, помидоры, персики и т. п. При росте цен на них покупательский спрос может переключаться на другие виды овощей и фруктов.

Исследуем динамику выручки при различных видах спроса. Общие расходы населения на данный товар (выручка от его продажи) при цене P составляют

$$G = P \cdot g(P).$$

Предельная выручка равна

$$\frac{dG}{dP} = g(P) + P \cdot g'(P) = g(P)\left[1 + \frac{P}{g(P)} g'(P)\right] = g(P)(1 - E_P).$$

Рассмотрим возможные ситуации:

а) при неэластичном спросе $(0 < E_P < 1)$ имеем $\frac{dG}{dP} > 0$, т.е. выручка увеличивается с ростом цены;

б) при нейтральном спросе $(E_P = 1)$ $\frac{dG}{dP} = 0$, и выручка практически не зависит от цены. В этом случае $G = C$ (C–постоян-ная величина) и $g(P) = \frac{C}{P}$, т. е. спрос обратно пропорционален цене;

в) если спрос эластичен $(E_P > 1)$, то $\frac{dG}{dP} < 0$, и с повыше-нием цены выручка от продажи товара снижается.

Из сказанного видно, что знание эластичности спроса на данный товар позволяет прогнозировать направление снижения цены. Очевидно, каждой фирме-продавцу выгодно, чтобы спрос на ее продукцию был как можно более неэластичным, ибо в такой ситуации существует возможность назначать сравни-тельно высокие цены. Однако появляющаяся возможность по-вышать цены на неэластичные товары (товары первой необходимости) ведет не только к соблазну производителей-продавцов увеличивать выручку (G) от продажи данных товаров, но и к нежелательному повышению спроса (g) на данные товары. Это может привести к массовым закупкам товаров «на черный день» и, как следствие, исчезновению этих товаров с рынка. Следовательно, цены на такие товары не могут быть свободными и должны контролироваться государством. Последнее должно регулировать цены на товары первой необходимости, проводя разумную политику ценообразования. Вопросами ценообразования на эластичные товары занимается производитель-продавец. Поэтому очень важно уметь вовремя реагировать на изменение эластичности продаваемого товара.

Пусть себестоимость продукции (в денежном выражении) складывается из постоянных (С) и переменных (П) затрат. Можно предположить, что переменные затраты прямо пропор-циональны общему объёму продаваемой продукции (спросу на данный товар). В этом случае прибыль F будет равна разности между товарооборотом (в денежном выражении - выручкой G) и себестоимостью продукции (С + П), т. е.

$$F = pg(p) - C - kg(p).$$

Определим условие, при котором прибыль будет максимальной. Для этого найдём критические точки функции F, приравняв к нулю $F'(p)$. Имеем

$$F'(p) = g + pg' - kg' = g(1 + p\frac{g'}{g} - \frac{k}{p}p\frac{g'}{g}) = g(1 - E_p + \frac{k}{p}E_p).$$

Из условия $F'(p) = 0$ находим

$$p^* = \frac{kE_p}{E_p - 1}.$$

Для эластичных товаров $E_p > 1$, поэтому при переходе через критическую точку p^* первая производная $F'(p)$ меняет знак с «+» на «-». Следовательно, в точке p^* прибыль F достигает максимального значения.

Выводы

1 Цены на товары и услуги могут формироваться на всех уровнях. На базовую для народного хозяйства продукцию (электроэнергия, энергоресурсы, хлеб, соль и т. д.) ценообразование должно происходить на самом верхнем уровне, т. е. на товары первой необходимости (неэластичные товары) необходимо осуществлять правительственный контроль цен. Цены на продукцию предприятий могут устанавливать сами производители или фирмы-продавцы.

2 Решение об объёмах закупок, т. е. формирование спроса осуществляется на нижних уровнях. Ценообразование непосредственно связано с величиной коэффициента эластичности спро-са относительно цены (E_p), который показывает на сколько процентов изменится спрос на товар, если цена на него увеличится (уменьшится) на 1%.

3 Знание эластичности спроса на данный товар позволяет прогнозировать направление изменения товарооборота и получаемой прибыли. При известном коэффициенте эластичности можно рассчитать оптимальную цену, при которой прибыль от реализации продукции достигает максимального значения.

4 Если цена p меньше оптимальной p^*, то есть возможность, увеличив цену при неизменной эластичности, повысить свои доходы. Если же цена находится в закритической зоне ($p>p^*$), то такая завышенная величина цены ведёт к зани-женному значению прибыли. Поэтому следует уменьшить цену и прибыль увеличится.

В реальных экономических условиях спрос на отдельные товары и услуги зависит не только от цены на данный товар, но и от цен на другие товары, доходов различных слоёв населения, состава семей и т. д. Построение функций спроса осуществляется методами эконометрики.

1.3 Предварительный анализ данных

Составление математической модели, отражающей разви-тие того или иного экономического процесса, начинается с оценки данных (табл. 1.1). Все методы регрессионного анализа используют аппарат математической статистики, который требует от исходных данных, чтобы они были сопоставимы и однородны, а для выявления закономерности, кроме этого, устойчивы и количество наблюдений достаточно велико. Невыполнение одного из этих требований делает бессмысленным применение математического анализа.

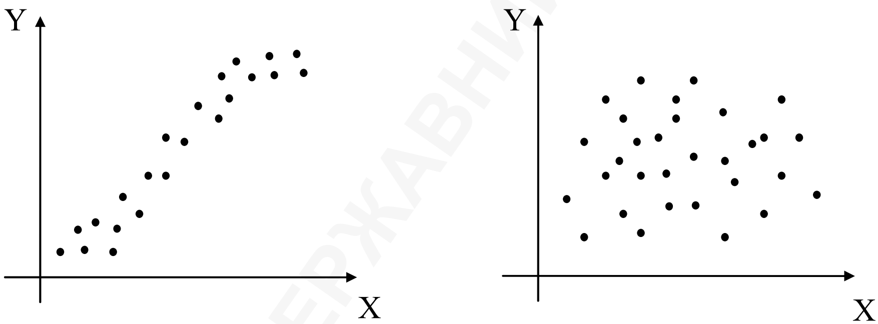
Сопоставимость предполагает формирование всех уровней ряда наблюдений по одной и той же методике, использование одинаковой единицы измерения и по возможности шага наблюдений. Во временных рядах (функция регрессии – однофакторная функция времени) требование одинакового шага по времени является обязательным.

Однородность достигается отсутствием сильных изломов, а также нетипичных аномальных наблюдений. При поиске закономерностей бывает целесообразно отбросить часть прошлых данных, если они отражают уже утратившую силу закономерность прошлого развития. Наличие аномальных (резко выделяющихся) наблюдений приводит к искажению результатов. Формально аномальность проявляется как сильный скачок с последующим приблизительным восстановлением предыдущего уровня. Примером такого наблюдения может служить значение курса доллара, зафиксированного в “черный вторник”.

Требование *полноты данных* обуславливается тем, что закономерность может обнаружиться лишь при наличии минимально допустимого объема наблюдений. Достаточное число наблюдений определяется в зависимости от цели приводимого исследования. Во временных рядах, например, если цель исследования - построение модели динамики с целью последующего прогноза, число уровней исходного динамического ряда должно быть не меньше семи.

Устойчивость характеризует преобладание закономерности над случайностью в изменении уровней ряда. На графиках устойчивых рядов даже визуально прослеживается закономерность (тенденция развития процесса), а на графиках неустойчивых рядов изменения последовательных уровней ряда представляются хаотичными, и поэтому поиск закономерностей в формировании значений уровней таких рядов лишен смысла. На рис. 1.7 представлены диаграммы рассеяния, соответствующие устойчивому (*а*) и неустойчивому рядам (*б*). Вывод об устойчивости или неустойчивости уровней исходных рядов данных можно сделать по ширине разброса точек (X, Y) на плоскости. Если точки расположены близко друг к другу в виде узкой полоски (рис. 1.7*а*), то можно утверждать наличие устойчивости и, следовательно, между переменными Y и X наблюдается относительно тесная связь. Если же точки разбросаны широко по диаграмме (рис. 1.7*б*), то устойчивости нет, и связь между переменными Y и X слабая или вообще отсутствует.

Диаграмма рассеяния позволяет произвести визуальный анализ эмпирических данных. Однако существуют и более точные, теоретически обоснованные методы обнаружения закономерной



а) *б)*

Рисунок 1.7

связи между случайными переменными Y и X. Наиболее распространенным из них является ***метод Фостера - Стюарта***. Он позволяет не только установить наличие тенденции в связи количественных признаков Y и X, но и проверить гипотезу (1.17) о постоянстве дисперсии случайного возмущения. Сущность метода заключается в следующем:

1 Сравнивается каждый уровень ряда со всеми предыдущими, при этом

$$f_i = 1, \quad e_i = 0, \quad \text{если } Y_i > Y_k, \quad k=1, 2, \ldots, i-1;$$

$$f_i = 0, \quad e_i = 1, \quad \text{если } Y_i < Y_k, \quad k=1, 2, \ldots, i-1;$$

$f_i = 0, \quad e_i = 0$ в остальных случаях.

2 Вычисляются значения величин

$$d = \sum_{i=2}^{n} (f_i - e_i), \quad s = \sum_{i=2}^{n} (f_i + e_i).$$

Показатели d и s характеризуют тенденции в связях Y и X и дисперсии σ_u^2 и X соответственно.

3 С помощью t-критерия Стьюдента проверяется гипотеза о том, можно ли считать случайными разности d - 0 и $s - \mu$. С этой целью находятся величины

$$t_d = \frac{d-0}{\sigma_1}, \quad t_s = \frac{s-\mu}{\sigma_2}, \tag{1.22}$$

где μ - среднее значение величины s; σ_1 и σ_2 - стандартные ошибки величин d и s соответственно. Значения величин μ, σ_1 и σ_2 табулированы и приведены в табл. 1.2.

4 При заданном уровне α сравниваются расчетные значения t_d и t_s с табличным. Если $t_d < t_{табл}$ и $t_s < t_{табл}$, то гипотеза об отсутствии тенденций в связи Y и X, и σ_u^2 и X подтверждается.

Т а б л и ц а 1.2 - *Значения средней* μ *и стандартных ошибок* σ_1 *и* σ_2 *для n от* 10 *до* 55

n	μ	σ_1	σ_2
10	3,858	1,964	1,288
15	4,636	2,153	1,521
20	5,195	2,279	1,677
25	5,632	2,373	1,791
30	5,990	2,447	1,882
35	6,294	2,509	1,956
40	6,557	2,561	2,019
45	6,790	2,606	2,072
50	6,998	2,645	2,121
55	7,187	2,681	2,163

В качестве примера рассмотрим определение наличия тенденций в ряду динамики производства реализованной продукции на производственном объединении за 1977 - 1996гг. (табл. 1.3). Данные приведены в сопоставимых ценах 1996 г.

Находим $d = 5,\ s = 7$. По данным табл. 1.2 при $n = 20$ имеем $\mu = 5{,}195$, $\sigma_1 = 2{,}279$, $\sigma_2 = 1{,}677$. Подставляя полученные значения в формулу (1.22), рассчитываем значения t_d и t_s, т. е.

$$t_d = \frac{5-0}{2{,}279} = 2{,}194, \quad t_s = \frac{7-5{,}195}{1{,}677} = 1{,}076.$$

Ближайшее табличное значение $t_{табл}$ для двустороннего критерия при n = 20 и уровне значимости $\alpha = 0{,}1$ равно $t_{табл}$=1,725 (см. приложение А), т. е. $|t_d| > t_{табл}$, $|t_s| < t_{табл}$. Следовательно, гипотеза об отсутствии тенденции в связи σ_u^2 и X подтвердилась, а в связи Y и X - отвергнута, т. е. в ряду динамики реализованной продукции прослеживается некоторая закономерность.

Т а б л и ц а 1.3 - *Реализованная продукция производственного объединения. Определение f_i и e_i*

Годы X	Млн. грн. Y_i	f_i	e_i	Годы X	Млн. грн. Y_i	f_i	e_i
1977	37,4	0	0	1987	36,5	0	0
1978	36,2	0	1	1988	34,1	0	1
1979	35,7	0	1	1989	35,2	0	0
1980	35,2	0	1	1990	35,0	0	0
1981	35,5	0	0	1991	34,2	0	0
1982	35,2	0	0	1992	27,1	0	1
1983	36,3	0	0	1993	25,2	0	1
1984	38,6	1	0	1994	25,4	0	0
1985	38,2	0	0	1995	26,2	0	0
1986	38,0	0	0	1996	26,8	0	0

1.4 Построение регрессионных однофакторных функций

Наглядной формой представления информации является диаграмма рассеяния, и ее обычно используют для предварительного анализа эмпирических данных.

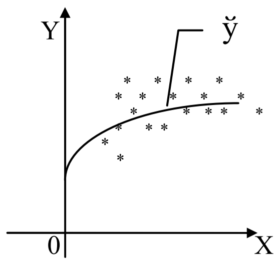


Рисунок 1.8

Случайное возмущение U в уравнении регрессии (1.4) может принимать как положительные, так и отрицательные значения. Поэтому функцию регрессии (1.2) можно интерпретировать на плоскости как линию, вокруг которой (сверху и снизу) разбросаны точки (X,Y) диаграммы рассеяния (рис. 1.8).

Предположим, что проведен качественный анализ возможной спецификации формы связи между переменными Y и X и установлен конкретный вид функции регрессии (1.2). Пусть далее каким-то способом, например, методом наименьших квад-ратов, получены значения оценок $\hat{a}_0,\ \hat{a}_1,...,\ \hat{a}_m$ параметров $a_0, a_1, ...$, a_{m}. Тогда функция

$$\hat{y} = f(X,\ \hat{a}_0,\ \hat{a}_1, \ldots,\ \hat{a}_m) \tag{1.23}$$

лучше других из выбранного множества функций (1.2) характеризует изучаемую закономерность связи между количественными признаками Y и X. Для каждого значения X_{i} объясняющей переменной X по уравнению (1.23) может быть вычислено значение $\hat{y}_i$, которое будем называть предсказанным или расчетным значением переменной Y_i для фиксированного X_i.

Как уже упоминалось, из-за искажающего влияния посторонних факторов для каждого значения X_i может наблюдаться несколько эмпирических значений Y_i, т. е. каждому значению X_i соответствует распределение значений переменной Y_i. Значение $\hat{y}_i$ функции регрессии (1.23), таким образом, можно рассматривать как оценку среднего значения переменной Y_{i} (точнее говоря, величины ее математического ожидания) для каждого фиксированного значения X. Процесс нахождения функции регрессии (1.23) называется выравниванием отдельных значений зависимой переменной, а изображение этой функции на графике - выровненной кривой. Говорят, что на этой кривой лежат усредненные значения результативной переменной Y, соответствующие отдельным значениям переменной X.

Значения $\hat{y}_i (i = 1, \ 2, \ ..., \ n)$ - оценки средних значений переменных Y_i, найденные по формуле (1.23), зависят от оценок $\hat{a}_0, \ \hat{a}_1, \ ..., \ \hat{a}_m$ параметров a_0, $a_1, ..., a_m$. Так как значения $\hat{a}_0$, $\hat{a}_1$, ..., $\hat{a}_m$ определяются видом функции регрессии (1.2), то выбор оптимальной спецификации формы связи между переменными Y и X приобретает здесь решающее значение. Подбор такой функции регрессии, которая как можно лучше бы характеризовала изучаемую закономерность связи между результативной и объясняющей переменными, является основной задачей регрессионного анализа. Поэтому на этапе, предшествующем построению функции регрессии, необходим обстоятельный качественный анализ исследуемой зависимости. На основе этого анализа формируется гипотеза о типе функции, правдоподобие которой затем статистически проверяется по эмпирическим данным.

Для диаграммы рассеяния, представленной на рис. 1.9, очевидно, функцию регрессии можно искать в виде

$$\breve{y} = a_0 + a_1 X \, . \qquad (1.24)$$

Коэффициент a_1 называется коэффициентом регрессии. Он характеризует наклон прямой к оси OX. Если γ - угол, который прямая регрессии образует с осью абсцисс, то a_1 = tgγ. Коэффициент a_1 является мерой влияния, оказываемого изменением переменной X на переменную $\breve{y}$. Согласно уравнению (1.24) a_1 указывает величину изменения $\breve{y}$

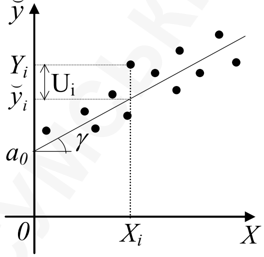


Рисунок1.9

при изменении аргумента X на одну единицу.

Если X обозначает объем производства некоторой продукции, а Y - суммарные затраты или издержки производства, то производная производственной функции $\frac{d\bar{y}}{dX}$ называется предельными издержками производства. В случае (1.24) $\frac{d\bar{y}}{dX} = a_1$, т. е. значение коэффициента a_1 определяет величину предельных издержек производства.

Представим себе такую ситуацию. В производстве продукции используется несколько видов сырья. Однако затраты всех ресурсов строго регламентированы технологией производства. Только один ресурс (например, затраты труда) может изменяться, оказывая влияние на объем производства. Производственная функция (1.24) характеризует здесь зависимость выпуска продукции Y от затрат X этого специфического ресурса. Скорость изменения этой функции выражается ее производной $\frac{d\bar{y}}{dX}$ и называется предельной производительностью ресурса. Если речь идет о затратах труда, то $\frac{d\bar{y}}{dX}$ – предельная производительность труда. Следовательно, в предположении, что производственная функция имеет вид (1.24), коэффициент регрессии a_1 в данном случае определяет предельную производительность ресурса.

На практике для оценки влияния факторного признака X на результативный признак Y обычно используют коэффициент эластичности (1.21). В случае линейной регрессии (1.24) коэффициент эластичности вычисляют по формуле

$$E_X = a_1 \cdot \frac{\bar{X}}{\bar{Y}}, \qquad (1.25)$$

где $\bar{X}$ и $\bar{Y}$ - средние признаков X и Y, найденные по результатам выборки. Параметр E_X показывает, на сколько процентов в среднем изменится результативный признак Y, если факторный признак X увеличится на 1 % .

В реальных экономических условиях связь между переменными адекватно представима, как правило, в нелинейной форме. Самый простой способ в таких ситуациях, на первый взгляд, - аппроксимация переменной Y с помощью многочлена от

X. Если предположить , что все табличные значения X_i (i = 1, 2, ..., n) различны, то, как известно из теории интерполяции, через n точек можно единственным способом провести многочлен степени n-1. Однако поскольку число наблюдений n сравнительно велико, мы будем иметь дело с многочленами большой степени. Ясно, что нас интересуют многочлены, степень которых ниже возможной:

$$\breve{y} = a_0 + a_1 X + a_2 X^2 + ... + a_m X^m, \quad m < n\text{-}1. \qquad (1.26)$$

При моделировании немонотонных процессов многочлены вида (1.26) получили широкое распространение. В статистике при изучении временных рядов выработано правило выбора степени многочлена (1.26), основанное на определении величин конечных разностей уровней заданных рядов. Согласно этому правилу многочлен первой степени (прямая) применяется как модель такого ряда, у которого первые разности

$$\Delta_t^{(1)} = Y_{t+1} - Y_t, \quad t = 1, 2, ..., n\text{-}1$$

постоянны. Если первые разности не равны, но варьируют с незначительными отклонениями друг от друга, а средняя арифметическая вторых разностей

$$\Delta_t^{(2)} = \Delta_{t+1}^{(1)} - \Delta_t^{(1)}, \quad t = 1, 2, ..., n\text{-}2$$

настолько мала, что ею можно пренебречь, то первые разности считают практически равными. Аналогично, если анализируя вторые разности, мы придем к выводу, что они практически равны, то для отражения ряда эмпирических данных используют многочлен второй степени (параболу) и т.д.

В качестве первого приближения при расчете выров-ненных значений уровней ряда динамики можно использовать формулу

$$\hat{y}_t = \overline{Y} + \overline{\Delta}^{(1)} \cdot \tau, \; \tau = t - \bar{t} \qquad (1.27)$$

при равных или почти равных первых разностях и

$$\hat{y}_t = \overline{Y} - \tfrac{n^2-1}{24} \overline{\Delta}^{(2)} + \overline{\Delta}^{(1)} \cdot \tau + \tfrac{1}{2} \overline{\Delta}^{(2)} \cdot \tau^2 \qquad (1.28)$$

в случае равных или почти равных вторых разностей. Здесь $\hat{y}_t$ - выровненное значение ряда динамики; $\overline{Y}$ - средний уровень ряда;

$\overline{\Delta}^{(1)}$ и $\overline{\Delta}^{(2)}$ - средние арифметические первых и вторых разностей соответственно; $\bar{t}$ - среднее время.

Рассмотрим следующий пример. Даны уровни производства продукции на фирме за первые 9 месяцев 1997 года (см.табл. 1.4). Расчеты показывают, что средняя арифме-тическая третьих разностей настолько мала, что ею можно пренебречь и, как следствие, вторые разности можно считать практически равными. По формуле (1.28) получаем следующее уравнение тренда (тенденции), отображающего развитие рас-сматриваемого процесса:

$$\hat{y}_t = 20302{,}86 + 43{,}75\tau + 73{,}57\tau^2.$$

В табл. 1.4 приведены выровненные значения $\hat{y}_t$, рассчитанные по уравнению тренда.

Т а б л и ц а 1.4 - *Сглаживание уровней производства готовой продукции на фирме за первые 9 месяцев 1997года методом конечных разностей*

Номер месяца t	Готовая продукция фирмы Y_t, грн	Условное время τ	Разности			Выровненные значения $\hat{y}_t$
			$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(3)}$	
1	21300	-4				21305
2	20840	-3	-460			20834
3	20530	-2	-310	150		20510
4	20350	-1	-180	130	-20	20333
5	20310	0	- 40	140	10	20303
6	20410	1	100	140	0	20420
7	20670	2	260	160	20	20684
8	21080	3	410	150	-10	21096
9	21650	4	570	160	10	21655
Итого	187140	0	350	1030	10	187140
Средние	20793,33		43,75	147,14	1,67	

Остановимся теперь на моделировании монотонных процессов, т.е. когда при $X_{i+1} > X_i$ разность $Y_{i+1} - Y_i$ сохраняет постоянный знак (i =1, 2, ..., n - 1). Если $Y_{i+1} > Y_i$, то функцию регрессии (1.2) называют кривой роста, в случае $Y_{i+1} < Y_i$ - кривой спада. Такие кривые позволяют описывать процессы трех основных типов: без предела роста (спада), с пределом роста (спада) без точки перегиба, с пределом роста (спада) и точкой перегиба.

Например, для описания процессов без предела роста служат функции: линейная (1.24), степенная (1.10), показательная (рис. 1.10):

$$\breve{y} = a_0 \cdot a_1^X, \quad a_0 > 0, \quad a_1 > 1; \tag{1.29}$$

логарифмическая (рис. 1.11):

$$\breve{y} = a_0 + a_1 \ln X, \quad a_0 > 0, \quad a_1 > 0 \tag{1.30}$$

и другие. Для степенной функции $\breve{y} = a_0 \cdot X^{a_1}$ коэффициент эластичности $E_X = \frac{X}{\breve{y}} \cdot \frac{d\breve{y}}{dX} = a_1$, и она часто используется в экономических исследованиях, поскольку соответствует предположению о постоянной эластичности между переменными *Y* и *X*. Показательная функция времени $\breve{y} = a_0 \cdot a_1^t$ характеризует временные ряды с приблизительно постоянным темпом прироста в единицу времени (g = a_1 - 1). Действительно, в данном случае

$$\frac{Y_t}{Y_{t-1}} - 1 \approx \frac{\breve{y}_t}{\breve{y}_{t-1}} - 1 = \frac{a_0 \cdot a_1^t}{a_0 \cdot a_1^{t-1}} - 1 = a_1 - 1 = g\ .$$

Средний темп прироста для временного ряда здесь также приблизительно будет равен *g:*

$$\sqrt[n-1]{\frac{Y_n}{Y_1}} - 1 \approx \sqrt[n-1]{\frac{\breve{y}_n}{\breve{y}_1}} - 1 = \sqrt[n-1]{\frac{a_0 \cdot a_1^n}{a_0 \cdot a_1}} - 1 == g\ .$$

Для логарифмической функции (1.30) $\frac{d\breve{y}}{dX} = \frac{a_1}{X}$, и она может применяться в тех случаях, в которых тангенс угла наклона касательной к функции регрессии уменьшается пропорционально росту переменной *X*.

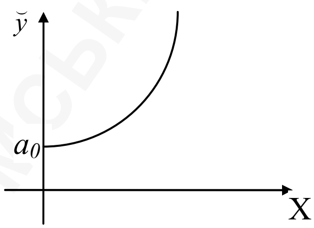


Рисунок1.10

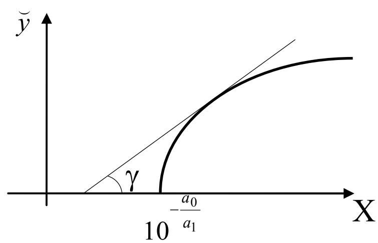


Рисунок 1.11

Процессы с пределом роста характерны для многих относительных показателей (душевое потребление продуктов питания, внесение удобрений на единицу площади, затраты на одну гривну произведенной продукции и т. п.). В таких случаях принимается гипотеза о существовании асимптотического уровня и могут быть использованы следующие функции: первая (1.18) и вторая (1.19) функции Торнквиста, гиперболическая функция (рис. 1.12)

$$\breve{y} = a_0 - \tfrac{a_1}{X}, \qquad a_0 > 0, \quad a_1 > 0\,. \tag{1.31}$$

Для функции (1.31) $\frac{d\breve{y}}{dX} = \frac{a_1}{X^2}$ так, что тангенс угла наклона касательной к этой кривой всегда положителен и убывает с возрастанием *X*. При $X \to \infty$ $\breve{y} \to a_0$, т. е. оценка параметра a_0 служит оценкой асимптотического уровня (рис. 1.12).

Кривые с пределом роста и точкой перегиба широко используются при статистическом анализе спроса на некоторые новые товары. Наиболее простой и удобной для практических приложений является здесь кривая Джонсона

$$\breve{y} = e^{a_0 - \frac{a_1}{X}}, \quad a_0 > 0,\ a_1 > 0\,. \tag{1.32}$$

Значение $\breve{y}$ в (1.32) не определено для $X = 0$, однако при $X \to 0$ имеем $\breve{y} \to 0$, что позволяет положить $\breve{y}(0) = 0$ и получить функцию, которая непрерывна справа в точке $X = 0$:

$$\frac{d\breve{y}}{dX} = \frac{a_1}{X^2} \cdot e^{a_0 - \frac{a_1}{X}}, \qquad \frac{d^2\breve{y}}{dX^2} = \left(\frac{a_1^2}{X^4} - \frac{2a_1}{X^3}\right) \cdot e^{a_0 - \frac{a_1}{X}}\,.$$

Так как $\frac{d\breve{y}}{dX} > 0$ при $X > 0$, то кривая Джонсона возрастает при положительных значениях *X*. Приравнивая нулю вторую производную, находим единственную точку перегиба при $X = \frac{a_1}{2}$.

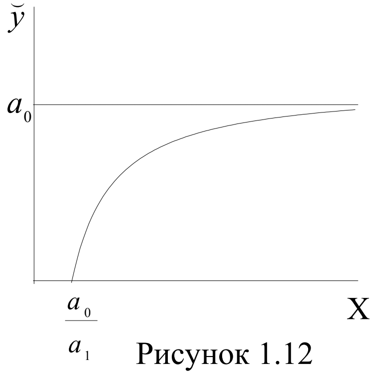


Рисунок 1.12

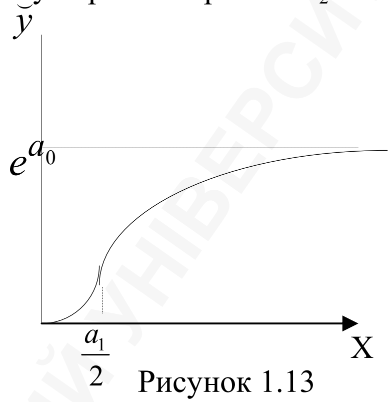


Рисунок 1.13

Слева от этой точки наблюдается рост функции с положительным ускорением ($\breve{y}''(X) > 0$), а справа от точки $X = a_1/2$ функция растет с замедлением ($\breve{y}''(X) < 0$). Если $X \to \infty$, то $\breve{y} \to e^{a_0}$. Таким образом, кривая Джонсона имеет вид, изображенный на рис. 1.13. Если положить

$$Y = e^{a_0 - \frac{a_1}{X} + U}$$

и прологарифмировать последнее равенство, то получим линейную регрессионную модель

$$Y^* = a_0 - \frac{a_1}{X} + U, \qquad Y^* = \ln Y,$$

удобную для нахождения оценок параметров a_0 и a_1.

При моделировании монотонных процессов (возрастающих или убывающих) , когда число наблюдений *n* невелико и неясно, есть ли асимптотический уровень и перегиб в тенден-ции изменения результативной переменной Y с ростом объясня-ющей переменной X может быть использована одна из следую-щих функций регрессии, зависящих от двух параметров:

$$а)\ \breve{y} = a_0 + a_1 X;\quad б)\ \breve{y} = a_0 + a_1 \ln X;\quad в)\ \breve{y} = a_0 + \frac{a_1}{X};$$

$$г)\ \breve{y} = a_0 \cdot a_1^{X};\quad д)\ \breve{y} = a_0 \cdot X^{a_1};\quad е)\ \breve{y} = e^{a_0 + \frac{a_1}{x}};$$

$$ж)\ \breve{y} = \frac{1}{a_0 + a_1 X};\quad з)\ \breve{y} = \frac{1}{a_0 + a_1 \ln X};\quad и)\ \breve{y} = \frac{X}{a_0 + a_1 X}.$$

Первые производные этих функций при $X > 0$ имеют постоянный знак и, следовательно, сами функции либо возрастают, либо убывают при положительных X (в зависимости от значения параметров).

Приведенные девять зависимостей примечательны тем, что если все табличные точки (X_i, Y_i) удовлетворяют одному из этих уравнений, то и средние значения $\overline{X}$ и $\overline{Y}$ также ему удовлетворяют. При этом в качестве средних значений $\overline{X}$ и $\overline{Y}$ могут быть средние арифметические, средние геометрические или средние гармонические. Вспомним, что среднее арифметическое n положительных чисел $Z_1, Z_2, \ldots, Z_n$ определяется по формуле

$$\overline{Z}_{ар} = \frac{Z_1 + Z_2 + \ldots + Z_n}{n},$$

среднее геометрическое есть величина, равная

$$\overline{Z}_{геом} = \sqrt[n]{Z_1 \cdot Z_2 \ldots Z_n},$$

среднее гармоническое

$$\overline{Z}_{гарм} = \frac{n}{\frac{1}{Z_1} + \frac{1}{Z_2} + \ldots + \frac{1}{Z_n}}.$$

Характерные средние $\overline{X}$ и $\overline{Y}$ для каждой из девяти возможных функций приведены в таблице 1.5.

Т а б л и ц а 1.5 - *Характерные средние для монотонных функций*

Номер функции	$\overline{X}$	$\overline{Y}$	Вид эмпирической функции
1	$\overline{X}_{ар}$	$\overline{Y}_{ар}$	$\breve{y} = a_0 + a_1 X$
2	$\overline{X}_{геом}$	$\overline{Y}_{ар}$	$\breve{y} = a_0 + a_1 \ln X$
3	$\overline{X}_{гарм}$	$\overline{Y}_{ар}$	$\breve{y} = a_0 + \frac{a_1}{X}$
4	$\overline{X}_{ар}$	$\overline{Y}_{геом}$	$\breve{y} = a_0 \cdot a_1^X$
5	$\overline{X}_{геом}$	$\overline{Y}_{геом}$	$\breve{y} = a_0 \cdot X^{a_1}$
6	$\overline{X}_{гарм}$	$\overline{Y}_{геом}$	$\breve{y} = e^{a_0 + \frac{a_1}{x}}$
7	$\overline{X}_{ар}$	$\overline{Y}_{гарм}$	$\breve{y} = \frac{1}{a_0 + a_1 X}$
8	$\overline{X}_{геом}$	$\overline{Y}_{гарм}$	$\breve{y} = \frac{1}{a_0 + a_1 \ln X}$
9	$\overline{X}_{гарм}$	$\overline{Y}_{гарм}$	$\breve{y} = \frac{X}{a_0 + a_1 X}$

Для проверки пригодности выбранной эмпирической функции, используя исходные данные табл. 1.1, находят значения $\overline{X}$ и $\overline{Y}$. Затем сравнивают Y_i, соответствующее исходным данным, со значением $\overline{Y}$. Если $\overline{X}$ не находится среди исходных данных X_i, то соответствующее значение Y можно определить с помощью линейной интерполяции, проведя через точки (X_i, Y_i), (X_{i+1}, Y_{i+1}) прямую. Здесь X_i и X_{i+1} - промежуточные значения, между которыми содержится $\overline{X}$ $\left(X_i < \overline{X} < X_{i+1}\right)$. Из уравнения прямой получаем $\breve{Y} = Y_i + \frac{Y_{i+1} - Y_i}{X_{i+1} - X_i}\left(\overline{X} - X_i\right)$. В качестве критерия выбора наилучшей функциональной зависимости можно выбрать следующий:

$$\left|\frac{\breve{Y} - \overline{Y}}{\breve{Y}}\right| \to \min . \qquad (1.33)$$

Т а б л и ц а 1.6 - *Данные о курсе акций за 9 недель*

t	1	2	3	4	5	6	0	8	9
Y_t	25	34	42	51	55	67	73	76	81

Т а б л и ц а 1.7 - *Выбор оптимальной эмпирической функции при моделировании монотонных процессов на основании данных табл.1.6*

Номер функции	Вид эмпирической функции	$\bar{t}$	$\bar{Y}$	$\breve{Y}$	$\left\lvert\frac{\breve{Y}-\bar{Y}}{\breve{Y}}\right\rvert$
1	$\breve{y}=a_0+a_1t$	5	56	55	0,018
2	$\breve{y}=a_0+a_1\ln t$	4,147	56	51,588	0,086
3	$\breve{y}=a_0+a_1/t$	3,181	56	43,629	0,284
4	$\breve{y}=a_0\cdot a_1^t$	5	52,471	55	0,046
5	$\breve{y}=a_0\cdot t^{a_1}$	4,147	52,471	51,588	0,017
6	$\breve{y}=e^{a_0+\frac{a_1}{t}}$	3,181	52,471	43,629	0,169
7	$\breve{y}=\frac{1}{a_0+a_1t}$	5	48,612	55	0,116
8	$\breve{y}=\frac{1}{a_0+a_1\ln t}$	4,147	48,612	51,588	0,061
9	$\breve{y}=\frac{t}{a_0+a_1t}$	3,181	48,612	43,629	0,114

В качестве примера рассмотрим данные о курсе акций за девять недель (табл. 1.6), и с помощью критерия (1.33) выберем оптимальную функцию из предложенных в табл. 1.5.

Анализ полученных результатов (табл. 1.7) показывает, что лучшей эмпирической функцией является степенная функция $\breve{y}=a_0\cdot t^{a_1}$. Немного уступает ей линейная функция $\breve{y}=a_0+a_1t$, и ее также можно выбрать в качестве функции регрессии. Критерий (1.33) указывает также на то, что гиперболические функции $\breve{y} = a_0 + \frac{a_1}{t}$ и $\breve{y} = \frac{1}{a_0 + a_1 t}$ вряд ли могут адекватно отображать рассматриваемый процесс.

1.5 Метод наименьших квадратов оценивания параметров функции регрессии

Рассмотрим сущность метода наименьших квадратов на примере прямой линии регрессии (1.24). Из графика рис. 1.9 видно, что $U_i = Y_i - \breve{y}_i$ (i = 1, 2, ..., n) есть отклонение опытной точки от оцениваемой линии, измеренное по вертикали. Это отклонение может быть как положительным, так и отрицательным, в зависимости от того, по какую сторону от линии лежит конкретная точка. При подборе прямой можно было бы выдвинуть требование, чтобы сумма отклонений всех точек от линии регрессии была равна нулю, т. е.

$$\sum \left(Y_i - \breve{y}_i\right) = \sum U_i = 0. \tag{1.34}$$

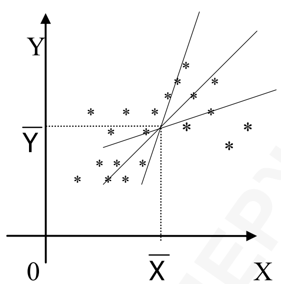


Рисунок 1.14

Условие (1.34) означает, что сумма положительных отклонений равна сумме отрицательных отклонений. Однако условию (1.34) будет удовлетворять бесконечно много прямых, а именно: это будут все те прямые, которые проходят через точку $(\overline{X}, \overline{Y})$ – центр рассеяния (рис. 1.14).

Для определения однозначной прямой используют такой показатель рассеяния случайной величины, как дисперсию. Если все отклонения возвести в квадрат и сложить, то результат будет непосредственно зависеть от разброса точек около искомой прямой. Из всех возможных прямых должна быть выбрана такая, для которой мера рассеяния опытных точек (X_i , Y_i) будет минимальна, т. е.

$$S = \sum \left(Y_i - \breve{y}\right)^2 = \sum U_i^2 \to \min. \tag{1.35}$$

Метод, основанный на требовании минимизации суммы квадратов отклонений (1.35), называется ***методом наименьших квадратов (МНК)***. Он является основой регрессионного анализа. С его помощью отыскиваются такие оценки $\hat{a}_0$ и $\hat{a}_1$ параметров a_0 и a_1 прямой линии регрессии (1.24), которые сводят к минимуму выбранную меру разброса. В случае линейной регрессии $\breve{y} = a_0 + a_1 X$ функция ошибок (1.35) принимает вид

$$S(a_0, a_1) = \sum \left[Y_i - (a_0 + a_1 X_i)\right]^2 \qquad (1.36)$$

и является функцией двух переменных - параметров a_0 и a_1. Таким образом, задача сводится к минимизации функции двух переменных. Известно, что если функция $S(a_0, a_1)$ имеет в точке ($\hat{a}_0$, $\hat{a}_1$) минимум, то обе частные производные первого порядка в этой точке равны нулю (необходимое условие экстремума). В данном случае имеем

$$\frac{\partial S}{\partial a_0} = -2\sum \left[Y_i - (a_0 + a_1 X_i)\right],$$

$$\frac{\partial S}{\partial a_1} = -2\sum \left[Y_i - (a_0 + a_1 X_i)\right] X_i \,. \qquad (1.37)$$

Если $a_0 = \hat{a}_0$, $a_1 = \hat{a}_1$, то $\frac{\partial S}{\partial a_0} = 0$ и $\frac{\partial S}{\partial a_1} = 0$. С учетом (1.37) для определения оценок $\hat{a}_0$ и $\hat{a}_1$ получаем следующую систему двух линейных уравнений с двумя неизвестными:

$$\begin{cases} n\hat{a}_0 + \hat{a}_1 \sum X_i = \sum Y_i \,, \\ \hat{a}_0 \sum X_i + \hat{a}_1 \sum X_i^2 = \sum X_i Y_i \,. \end{cases} \qquad (1.38)$$

Покажем, что определитель системы (1.38)

$$\Delta = n\sum X_i^2 - \left(\sum X_i\right)^2 > 0 \,, \qquad (1.39)$$

тогда система уравнений (1.38) имеет единственное решение. С этой целью рассмотрим квадратный трехчлен

$$P(z) = \sum \left(X_i z + 1 \right)^2 = z^2 \sum X_i^2 + 2z \sum X_i + n.$$

Равенство $P(z)=0$ выполняется лишь в случае $z = -\frac{1}{X_1} = -\frac{1}{X_2} = \cdots = -\frac{1}{X_n}$, что невозможно. Следовательно, $P(z) > 0$ для всех значений z и дискриминант

$$D = \left(\sum X_i \right)^2 - n \sum X_i^2 < 0,$$

что и требовалось доказать. Последнее неравенство можно записать также в виде

$$\frac{1}{n} \sum X_i < \sqrt{\frac{1}{n} \sum X_i^2}, \qquad (1.40)$$

что означает: среднее арифметическое n чисел, хотя бы два из которых не равны между собой, меньше среднего квадратичного этих чисел.

Так как определитель $\Delta \neq 0$ (1.39), то решение системы уравнений (1.38) может быть получено, например, по формулам Крамера. Имеем

$$\hat{a}_0 = \frac{\Delta_1}{\Delta}, \quad \hat{a}_1 = \frac{\Delta_2}{\Delta}, \qquad (1.41)$$

$$\Delta_1 = \sum Y_i \cdot \sum X_i^2 - \sum X_i \cdot \sum X_i Y_i, \quad \Delta_2 = n \sum X_i Y_i - \sum X_i \cdot \sum Y_i .$$

Осталось показать, что значения $\hat{a}_0$ и $\hat{a}_1$ (1.41) обеспечивают минимум функции ошибок (1.36). Для этого вычислим вторые производные функции (1.36) и воспользуемся достаточным условием минимума функции двух переменных. Используя (1.37), находим

$$A = \frac{\partial^2 S}{\partial a_0^2} = 2n, \quad B = \frac{\partial^2 S}{\partial a_0 \partial a_1} = 2\sum X_i, \quad C = \frac{\partial^2 S}{\partial a_1^2} = 2\sum X_i^2.$$

Теперь $A > 0$, $AC - B^2 = 4\Delta > 0$, т. е., действительно, при $a_0 = \hat{a}_0$ и $a_1 = \hat{a}_1$ функция ошибок имеет минимум.

Предположим, что решив систему уравнений (1.38), мы определили оценки $\hat{a}_0$ и $\hat{a}_1$. Тогда выравненные значения $\hat{y}_i$ уровней ряда *(i = 1, 2, ..., n)* находятся на прямой

$$\hat{y} = \hat{a}_0 + \hat{a}_1 X . \qquad (1.42)$$

Положим

$$\overline{X} = \tfrac{1}{n}\sum X_i , \quad \overline{Y} = \tfrac{1}{n}\sum Y_i$$

и разделим первое из уравнений (1.38) на n. Получим

$$\overline{Y} = \hat{a}_o + \hat{a}_1 \overline{X}. \qquad (1.43)$$

Сравнивая (1.43) и (1.42), заключаем, что оцененная линия регрессии (1.42) проходит через центр рассеивания $(\overline{X}, \overline{Y})$ и ее уравнение можно записать в виде

$$\hat{y} - \overline{Y} = \hat{a}_1 (X - \overline{X}), \qquad (1.44)$$

если вычесть (1.43) из (1.42). Из (1.42) и первого уравнения (1.38), кроме того, находим

$$\sum \hat{U}_i = \sum (Y_i - \hat{y}_i) = 0 , \qquad (1.45)$$

т. е. *МНК*-оценки таковы, что сумма отклонений всех точек от прямой (1.42) равна нулю. Это свойство позволяет простым способом проверить, были ли сделаны ошибки в расчетах при вычислении $\hat{y}_i$ и $\hat{U}_i = Y_i - \hat{y}_i$ $(i = 1,2, ..., n)$.

МНК при оценивании параметров многочленов. Предположим, что экономический анализ изучаемого процесса и учет характера скопления точек на диаграмме рассеяния позволил выдвинуть гипотезу о полиномиальной связи между переменными Y и X (1.26). В этом случае функция ошибок (1.35) является квадратичной функцией параметров a_0, a_1, ..., a_m, а частные производные - линейны относительно параметров. Условие равенства нулю частных производных позволяет получить следующую систему нормальных уравнений:

$$
\begin{cases}
n\hat{a}_0 + \hat{a}_1\sum X_i + \hat{a}_2\sum X_i^2 + \ldots + \hat{a}_m\sum X_i^m = \sum Y_i, \\
\hat{a}_0\sum X_i + \hat{a}_1\sum X_i^2 + \hat{a}_2\sum X_i^3 + \ldots + \hat{a}_m\sum X_i^{m+1} = \sum Y_i X_i, \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
\hat{a}_0\sum X_i^m + \hat{a}_1\sum X_i^{m+1} + \hat{a}_2\sum X_i^{m+2} + \ldots + \hat{a}_m\sum X_i^{2m} = \sum Y_i X_i^m.
\end{cases}
\tag{1.46}
$$

Система (1.46) содержит $m + 1$ линейное уравнение с $m + 1$ неизвестными - оценками $\hat{a}_0$, $\hat{a}_1$, ..., $\hat{a}_m$ параметров a_0, a_1, ... , a_m. Можно показать, что как и в случае $m = 1$, она имеет единственное решение и обеспечивает минимум функции ошибок для любого значения $m < n - 1$.

В случае параболической функции регрессии ($\breve{y} = a_0 + a_1 X + a_2 X^2$) система уравнений (1.46) приобретает вид

$$
\begin{cases}
n\hat{a}_0 + \hat{a}_1\sum X_i + \hat{a}_2\sum X_i^2 = \sum Y_i, \\
\hat{a}_0\sum X_i + \hat{a}_1\sum X_i^2 + \hat{a}_2\sum X_i^3 = \sum Y_i X_i, \\
\hat{a}_0\sum X_i^2 + \hat{a}_1\sum X_i^3 + \hat{a}_2\sum X_i^4 = \sum Y_i X_i^2.
\end{cases}
\tag{1.47}
$$

При сглаживании уровней временных рядов существует упрощенный прием расчета значений оценок $\hat{a}_0$, $\hat{a}_1$, ..., $\hat{a}_m$ параметров a_0, a_1, ..., a_m полинома (1.26), применение которого дает ощутимую экономию труда без какой-либо потери точности результатов. Упрощенный расчет параметров системы (1.46) заключается в переносе начала координат в середину ряда динамики. В этом случае упрощаются сами нормальные уравнения, кроме того, уменьшаются абсолютные значения коэффициентов системы. В самом деле, если до переноса начала координат объясняющая переменная t (время) была равна 1, 2, ..., n, то после переноса условное время τ = ..., -3, -2, 0, 1, 2, 3,..., если число членов ряда нечетное и τ = ..., -5, -3, -1, 1, 3, 5,..., если n - четно. Следовательно, $\sum\tau$ и все $\sum\tau^k$, у которых k – нечетное число, равны нулю, и все члены уравнений с такими суммами могут быть исключены из рассмотрения.

Упрощенная система нормальных уравнений для прямой (1.38) имеет вид

$$\begin{cases} n\widehat{a}_0 = \sum Y_t, \\ \widehat{a}_1 \sum \tau^2 = \sum \tau Y_t, \end{cases} \tag{1.48}$$

для параболы второго порядка (1.47) получаем

$$\begin{cases} n\widehat{a}_0 + \widehat{a}_2 \sum \tau^2 = \sum Y_t, \\ \widehat{a}_1 \sum \tau^2 \qquad = \sum \tau Y_t, \\ \widehat{a}_0 \sum \tau^2 + \widehat{a}_2 \sum \tau^4 = \sum \tau^2 Y_t. \end{cases} \tag{1.49}$$

Пример 1 Выберем для описания динамики роста курса акций (табл.1.6) линейную модель.

Т а б л и ц а 1.8 - *Оценивание параметров уравнения прямой*

Время t	Y_t	Условное время τ	τ^2	τY_t	Расчетные значения $\hat{y}_t$	Отклонение $\hat{u}_t$
1	25	- 4	16	- 100	27,5	-2,5
2	34	- 3	9	- 102	34,6	-0,6
3	42	-2	4	- 84	41,7	0,3
4	51	-1	1	- 51	48,9	2,1
5	55	0	0	0	56,0	-1,0
6	67	1	1	67	63,1	3,9
7	73	2	4	146	70,3	2,7
8	76	3	9	228	77,4	-1,4
9	81	4	16	324	84,5	-3,5
Σ	504	0	60	428	504	0

Оценки $\hat{a}_0$ и $\hat{a}_1$ параметров a_0 и a_1 определяются из системы уравнений (1.48). Имеем

$$\begin{cases} 9\hat{a}_0 = 504, \\ 60\hat{a}_1 = 428 \end{cases} \quad => \quad \begin{cases} \hat{a}_0 = 56, \\ \hat{a}_1 = 7{,}13. \end{cases}$$

Таким образом, линейная модель имеет вид

$$\hat{y}_t = 56{,}0 + 7{,}13\tau, \quad \tau = t - 5. \tag{1.50}$$

Оцененное значение $\hat{a}_1 = 7{,}13$ коэффициента регрессии a_1 в данном случае показывает, что еженедельно курс акций увеличивается в среднем приблизительно на 7 пунктов.

Пример 2 Пусть исследуется зависимость выполнения плана по выпуску товарной продукции (в %) от удельного веса потерь от брака по 8 цехам (в % к общим потерям по заводу). Данные приведены в табл. 1.9.

Т а б л и ц а 1.9

i	Y_i	X_i	X_i^2	$Y_i X_i$	$\hat{y}_i$	$\hat{u}_i$
1	108,5	8,2	67,24	889,70	110,2	-1,7
2	96,1	14,3	204,49	1374,23	94,9	1,2
3	102,4	12,4	153,76	1269,76	99,7	2,7
4	109,6	8,4	70,56	920,64	109,7	-0,1
5	104,2	11,2	125,44	1167,04	102,7	1,5
6	82,4	18,3	334,89	1507,92	84,9	-2,5
7	104,0	10,0	100,00	1040,00	105,7	-1,7
8	88,2	17,2	295,84	1517,04	87,6	0,6
Σ	795,4	100,0	1352,22	9686,33	795,4	0

Анализ диаграммы рассеяния по выборочным данным позволяет предположить линейную связь между планом по выпуску товарной продукции (Y) и удельным весом потерь от брака (X). Поэтому имеет смысл искать зависимость в виде линейной функции регрессии (1.24). Табл. 1.9 содержит промежуточные результаты, необходимые для вычисления оценок $\hat{a}_0$ и $\hat{a}_1$ параметров a_0 и a_1.

Система уравнений (1.38) принимает вид

$$\begin{cases} 8\hat{a}_0 \quad +100\hat{a}_1 \qquad = 795{,}4\,, \\ 100\hat{a}_0 + 1352{,}22\hat{a}_1 = 9686{,}33, \end{cases}$$

откуда находим $\hat{a}_0 = 130{,}751$, $\hat{a}_1 = -2{,}506$. Таким образом, функция регрессии имеет вид

$$\hat{y} = 130{,}751 - 2{,}506X.$$

Равенство нулю суммы остатков свидетельствует о правильности вычислений.

В данном случае коэффициент регрессии a_1 является показателем влияния изменения удельного веса потерь от брака на выполнение плана по выпуску товарной продукции в предположении, что влияние прочих факторов не учитывается. Его оценка $\hat{a}_1 = -2{,}506$ показывает, что выполнение плана по выпуску продукции возрастет в среднем на 2,5%, если удельный вес потерь уменьшится на 1%. Найдем также коэффициент эластичности (1.25). Имеем

$$\overline{X} = \tfrac{100}{8} = 12{,}5, \quad \overline{Y} = \tfrac{795{,}4}{8} = 99{,}425,$$

$$E = \hat{a}_1 \cdot \frac{\overline{X}}{\overline{Y}} = -2{,}506 \cdot \frac{12{,}5}{99{,}425} = -0{,}315.$$

Итак, выполнение плана по выпуску товарной продукции уменьшится в среднем на 0,315% (с y до y - 0,00315y), если удельный вес потерь от брака увеличится на 1% (с x до x+0,01x) при условии, что другие факторы не учитываются.

Пример 3 В 80-е годы уровень дефицита бюджета в СССР (Н.И. Коршунова, В.С.Плясунов, 1996) складывался следующим образом (табл.1.10).

Т а б л и ц а 1.10 - *Уровень дефицита бюджета в СССР в 80-е годы (в процентах к вал. нац. продукту)*

t	1980	1981	1982	1983	1984	1985	1986	1987	1988
Y_t	2,9	2,3	3,1	2,2	2,0	2,7	6,5	8,0	9,1

Анализ диаграммы рассеяния в данном случае позволяет выдвинуть гипотезу о параболической функции регрессии $(\breve{y} = a_0 + a_1 t + a_2 t^2)$.

Для неизвестных оценок $\hat{a}_0$, $\hat{a}_1$ и $\hat{a}_2$ система уравнений (1.49) приобретает вид

$$\begin{cases} 9\hat{a}_0 + 60\hat{a}_2 = 38{,}8, \\ 60\hat{a}_1 = 49{,}2, \\ 60\hat{a}_0 + 708\hat{a}_2 = 328, \end{cases}$$

откуда находим $\hat{a}_0 = 2{,}81$, $\hat{a}_1 = 0{,}82$, $\hat{a}_2 = 0{,}225$. Следовательно,

$\hat{y}_t = 2{,}81 + 0{,}82\tau + 0{,}225\tau^2$, $\tau = t - 1984$. По оценкам экспертов, дефицит бюджета в СССР в 1989 году должен был составить 13,0%. Полученное уравнение регрессии дает оценку

$$\hat{y}(1989) = 2{,}81 + 0{,}82 \cdot 5 + 0{,}225 \cdot 5^2 = 12{,}54\%,$$

т. е. прогноз, основанный на параболической функции регрессии, дает вполне удовлетворительный результат.

Т а б л и ц а 1.11 - *Оценивание параметров уравнения параболы*

t	Y_t	τ	τ^2	τ^4	τY_t	$\tau^2 Y_t$	$\hat{y}_t$	$\hat{u}_t$
1980	2,9	- 4	16	256	-11,6	46,4	3,13	-0,23
1981	2,3	- 3	9	81	- 6,9	20,7	2,38	-0,08
1982	3,1	- 2	4	16	- 6,2	12,4	2,07	1,03
1983	2,2	- 1	1	1	- 2,2	2,2	2,22	-0,02
1984	2,0	0	0	0	0	0	2,81	-0,81
1985	2,7	1	1	1	2,7	2,7	3,86	-1,16
1986	6,5	2	4	16	13,0	26,0	5,35	1,15
1987	8,0	3	9	81	24,0	72,0	7,29	0,71
1988	9,1	4	16	256	36,4	145,6	9,69	-0,59
$\sum$	38,8	0	60	708	49,2	328	38,80	0

1.6 Определение оценок параметров эмпирических функций регрессии, моделирующих монотонный процесс

В § 1.4 приведены девять функций, зависящих от двух параметров, которые могут быть использованы при модели-ровании монотонных процессов при условии, что число наблю-дений n невелико. Там же приведен критерий (1.33), позволя-ющий, не зная значений оценок, выбрать более подходящие из них в качестве функции регрессии. Если провести еще анализ диаграммы рассеяния, то ошибка определения оптимальной спецификации формы связи между изучаемыми переменными Y и X может быть сведена до минимума.

Опираясь на метод наименьших квадратов, мы можем получить численные значения оценок параметров для всех зависимостей и вычислить в каждом случае значение функции ошибок, т. е. сумму квадратов разностей наблюдаемых и рассчитанных по соответствующей формуле значений результативной переменной. Наиболее адекватной исследуемому процессу естественно считать ту модель, для которой эта сумма будет минимальной.

Особенностью регрессионных моделей является то, что параметры моделей a_0, a_1, ..., a_m и возмущение U входят в них линейно. Из девяти упомянутых зависимостей только первые три удовлетворяют этому требованию. Остальные модели нелинейны относительно параметров и их необходимо путем преобразований (см. §1.1) сводить к линейным моделям. В данном случае такими преобразованиями являются логарифмирование и нахождение обратного значения.

Продемонстрируем методику применения *МНК* для всех девяти монотонных зависимостей на примере моделирования динамики роста курса акций (табл. 1.6) с последующим анализом построенных моделей и сравнением полученных результатов с выводами, основанными на использовании критерия (1.33).

1 Согласно (1.50) линейная модель имеет вид

$$\hat{y}_t = 20{,}35 + 7{,}13\, t, \quad t = 1, 2, ..., 9. \tag{1.51}$$

2 В случае логарифмической функции регрессии $\breve{y} = a_0 + a_1 \ln t$ получаем

$$\begin{cases} n\widehat{a}_0 + \widehat{a}_1 \sum \ln t = \sum Y_t, \\ \widehat{a}_0 \sum \ln t + \widehat{a}_1 \sum \ln^2 t = \sum Y_t \ln t \end{cases} \Rightarrow \begin{cases} 9\widehat{a}_0 + 12{,}802\widehat{a}_1 = 504, \\ 12{,}802\widehat{a}_0 + 22{,}348\widehat{a}_1 = 827{,}04. \end{cases}$$

Решение системы уравнений дает $\widehat{a}_0 = 18{,}15$, $\widehat{a}_1 = 26{,}61$, и

$$\widehat{y}_t = 18{,}15 + 26{,}61 \ln t, \quad t = 1, 2, ..., 9. \qquad (1.52)$$

3 Если выбрана гиперболическая зависимость $\breve{y} = a_o + \dfrac{a_1}{t}$, то, положив в (1.38) $X_t = \frac{1}{t}$ (t = 1, 2, ..., 9), для определения оценок $\hat{a}_0$ и $\hat{a}_1$ приходим к системе уравнений

$$\begin{cases} n\widehat{a}_0 + \widehat{a}_1 \sum \frac{1}{t} = \sum Y_t, \\ \widehat{a}_0 \sum \frac{1}{t} + \widehat{a}_2 \sum \frac{1}{t^2} = \sum \frac{Y_t}{t} \end{cases} \Rightarrow \begin{cases} 9\widehat{a}_0 + 2.829 a_1 = 504, \\ 2.829\widehat{a}_o + 1.540\widehat{a}_1 = 119{,}848, \end{cases}$$

откуда находим $\hat{a}_0 = 74{,}64$, $\hat{a}_1 = -59{,}31$, т. е.

$$\widehat{y}_t = 74{,}64 - \frac{59{,}31}{t}, \quad t = 1, 2, ..., 9. \qquad (1.53)$$

4 Если исходная модель характеризуется показательной функцией $\breve{y} = a_0 \cdot a_1^t$, то полагаем $Y = a_0 \cdot a_1^t \cdot e^U$ и после логарифмирования получаем линейную регрессионную модель

$$Y^* = a_0^* + a_1^* t + U \quad (Y^* = \ln Y, \quad a_0^* = \ln a_0, \quad a_1^* = \ln a_1).$$

Далее имеем

$$\begin{cases} n\widehat{a}_0^* + \widehat{a}_1^* \sum t = \sum \ln Y_t, \\ \widehat{a}_0^* \sum t + \widehat{a}_1^* \sum t^2 = \sum t \ln Y_t \end{cases} \Rightarrow \begin{cases} 9\widehat{a}_0^* + 45\widehat{a}_1^* = 35{,}642, \\ 45\widehat{a}_0^* + 285\widehat{a}_1^* = 186{,}705. \end{cases}$$

Значения оценок $\widehat{a}_0^* = 3{,}253$, $\widehat{a}_1^* = 0{,}140$, $\widehat{a}_0 = e^{\widehat{a}_0^*} = e^{3,253} = 25{,}85$, $\widehat{a}_1 = e^{\widehat{a}_1^*} = e^{0,140} = 1{,}15$. Искомая показательная функция имеет вид

$$\widehat{y}_t = 25{,}85 \cdot 1{,}15^t, \quad t = 1, 2, ..., 9. \tag{1.54}$$

5 Пусть исходная модель описывается степенной функцией $\breve{y} = a_0 \cdot t^{a_1}$. С целью получения линейной регрессионной модели положим $Y = a_0 \cdot t^{a_1} \cdot e^U$. Следовательно, $\ln Y = \ln a_0 + a_1 \ln t + U$, и уравнение $Y^* = a_0^* + a_1 \ln t + U$ $(Y^* = \ln Y,\ a_0^* = \ln a_0)$ становится линейным относительно параметров a_0^*, a_1 и возмущения U. Система уравнений (1.38) для оценок $\widehat{a}_0^*$ и $\widehat{a}_1$ приобретает вид

$$\begin{cases} n\widehat{a}_0^* + \widehat{a}_1 \sum \ln t = \sum \ln Y_t, \\ \widehat{a}_0^* \sum \ln t + \widehat{a}_1 \sum \ln^2 t = \sum \ln Y_t \cdot \ln t \end{cases} \Rightarrow \begin{cases} 9\widehat{a}_0^* + 12{,}802\widehat{a}_1 = 35{,}642, \\ 12{,}802\widehat{a}_0^* + 22{,}348\widehat{a}_1 = 52{,}992. \end{cases}$$

Из последней системы находим $\widehat{a}_0^* = 3{,}171$, $\hat{a}_1 = 0{,}56$. Теперь $\widehat{a}_0 = e^{\widehat{a}_0^*} = e^{3,171} = 23{,}84$. Таким образом,

$$\widehat{y}_t = 23{,}84 \cdot t^{0,56}, \quad \mathrm{t} = 1, 2, ..., 9. \tag{1.55}$$

6 Если исходная модель описывается функцией $\breve{y} = e^{a_0 + \frac{a_1}{t}}$, то с целью получения линейной регрессионной модели полагаем $Y = e^{a_0 + \frac{a_1}{t} + U}$. После логарифмирования получаем $Y^* = a_0 + \dfrac{a_1}{t} + U$, где $Y^* = \ln Y$. Далее имеем

$$\begin{cases} n\widehat{a}_0 + \widehat{a}_1 \sum \dfrac{1}{t} = \sum \ln Y_t, \\ \widehat{a}_0 \sum \dfrac{1}{t} + \widehat{a}_1 \sum \dfrac{1}{t^2} = \sum \dfrac{1}{t} \cdot \ln Y_t \end{cases} \Rightarrow \begin{cases} 9\widehat{a}_0 + 2{,}829\widehat{a}_1 = 35{,}642, \\ 2{,}829\widehat{a}_0 + 1{,}540\widehat{a}_1 = 10{,}355. \end{cases}$$

Значения оценок $\widehat{a}_0 = 4{,}368$, $\widehat{a}_1 = -1{,}301$. А искомая функция имеет вид

$$\widehat{y}_t = e^{4,368-\frac{1,301}{t}}, \quad t = 1, 2, ..., 9. \qquad (1.56)$$

7 Предположим, что моделирующая функция $\breve{y} = \frac{1}{a_0 + a_1 t}$, тогда связь между переменными Y и t будем искать в виде $Y = \frac{1}{a_0 + a_1 t + U}$, что приводит к следующей преобразованной (линейной) модели $\frac{1}{Y} = a_0 + a_1 t + U$. Теперь

$$\begin{cases} n\widehat{a}_0 + \widehat{a}_1 \sum t = \sum \frac{1}{Y_t}, \\ \widehat{a}_0 \sum t + \widehat{a}_1 \sum t^2 = \sum \frac{t}{Y_t} \end{cases} \Rightarrow \begin{cases} 9\hat{a}_0 + 45\hat{a}_1 = 0,185, \\ 45\hat{a}_0 + 285\hat{a}_1 = 0,714, \end{cases}$$

откуда получаем $\hat{a}_0 = 0,03593$, $\hat{a}_1 = -0,00307$, т. е.

$$\hat{y}_t = \frac{1000}{35,93 - 3,07t}, \quad t = 1, 2, ..., 9. \qquad (1.57)$$

8 Если моделирующая функция $\breve{y} = \frac{1}{a_0 + a_1 \ln t}$, то полагаем $Y = \frac{1}{a_0 + a_1 \ln t + U}$. И соответственно линейная модель $\frac{1}{Y} = a_0 + a_1 \ln t + U$. Далее имеем

$$\begin{cases} n\widehat{a}_0 + \widehat{a}_1 \sum \ln t = \sum \frac{1}{Y_t}, \\ \widehat{a}_0 \sum \ln t + \widehat{a}_1 \sum \ln^2 t = \sum \frac{\ln t}{Y_t} \end{cases} \Rightarrow \begin{cases} 9\widehat{a}_0 + 12,802\,\widehat{a}_1 = 0,185, \\ 12,802\,\widehat{a}_0 + 22,348\,\widehat{a}_1 = 0,211, \end{cases}$$

Следовательно, $\hat{a}_0 = 0,03862$, $\widehat{a}_1 = 0,01269$ и

$$\widehat{y}_t = \frac{1000}{38,62 - 12,69 \ln t}, \quad t = 1, 2, ..., 9. \qquad (1.58)$$

9 В случае зависимости $\breve{y} = \frac{t}{a_0 + a_1 t}$ или $\breve{y} = \frac{1}{a_1 + \frac{a_0}{t}}$ полагаем $Y = \frac{1}{a_1 + \frac{a_0}{t} + U}$, $\frac{1}{Y} = a_1 + \frac{a_0}{t} + U$. Система нормальных уравнений, соответствующая линейной модели, такова:

$$\begin{cases} n\widehat{a}_1 + \widehat{a}_0 \sum \frac{1}{t} = \sum \frac{1}{Y_t}, \\ \widehat{a}_1 \sum \frac{1}{t} + \widehat{a}_0 \sum \frac{1}{t^2} = \sum \frac{1}{tY_t} \end{cases} \Rightarrow \begin{cases} 9\hat{a}_1 + 2{,}829\hat{a}_0 = 0{,}185, \\ 2{,}829\hat{a}_1 + 1{,}540\hat{a}_0 = 0{,}079. \end{cases}$$

Следовательно, $\hat{a}_0 = 0{,}03143$, $\hat{a}_1 = 0{,}01069$ и

$$\hat{y}_t = \frac{1000t}{31{,}43 + 10{,}69t}, \quad t = 1, 2, ..., 9 . \qquad (1.59)$$

Итак, конкретный вид девяти предполагаемых монотонно возрастающих функциональных зависимостей, моделирующих динамику роста курса акций по данным табл. 1.6, построен.

Т а б л и ц а 1.12 - *Выбор оптимальной функции регрессии по данным табл.1.6*

Номер функции	Вид эмпирической функции	Оценки параметров	$\sum_{y=1}^{n}(Y_t - \widehat{y}_t)^2$
1	$\breve{y} = a_0 + a_1 t$	$\widehat{y} = 20{,}35 + 7{,}13t$	64
2	$\breve{y} = a_0 + a_1 \ln t$	$\widehat{y} = 18{,}15 + 26{,}61 \ln t$	171
3	$\breve{y} = a_0 + \frac{a_1}{t}$	$\hat{y} = 74{,}64 - \frac{59{,}31}{t}$	814
4	$\breve{y} = a_0 \cdot a_1^t$	$\widehat{y} = 25{,}85 \cdot 1{,}15^t$	268
5	$\breve{y} = a_0 \cdot t^{a_1}$	$\widehat{y} = 23{,}84 \cdot t^{0{,}56}$	31
6	$\breve{y} = e^{a_0 + \frac{a_1}{t}}$	$\widehat{y} = e^{4{,}368 - \frac{1{,}301}{t}}$	524
7	$\breve{y} = \frac{1}{a_0 + a_1 t}$	$\hat{y} = \frac{1000}{35{,}93 - 3{,}07t}$	1998
8	$\breve{y} = \frac{1}{a_0 + a_1 \ln t}$	$\widehat{y} = \frac{1000}{38{,}62 - 12{,}69 \ln t}$	223
9	$\breve{y} = \frac{t}{a_0 + a_1 t}$	$\hat{y} = \frac{1000t}{31{,}43 + 10{,}69t}$	304

Какая из них лучше других подходит в данной ситуации? Поскольку для каждого из уравнений (1.51) - (1.59) может быть вычислена сумма квадратов отклонений наблюдаемых значений объясняемой переменной от теоретических, т. е. рассчитанных по оцененному уравнению регрессии, то наиболее адекватной исследуемому процессу будем считать ту, для которой эта сумма наименьшая.

Анализ результатов, представленных в табл. 1.12, показывает, что наиболее подходящей оцененной функцией регрессии при моделировании рассматриваемого процесса является степенная функция $\widehat{y} = 23.84 \cdot t^{0.56}$. Вполне удовлетворительной может быть и линейная модель $\widehat{y} = 20.35 + 7.13t$. Видно также, что выбор в качестве модели гиперболических функций регрессии $\widehat{y} = 74.64 - \frac{59.31}{t}$ или $\widehat{y} = \frac{1000}{35.93-3.07t}$ является неудачным. Сопоставление с результатами, приведенными в табл. 1.7, подтверждает в данном случае корректность критерия (1.33) при выборе оптимальной эмпирической функции без знания численных значений оценок параметров моделей.

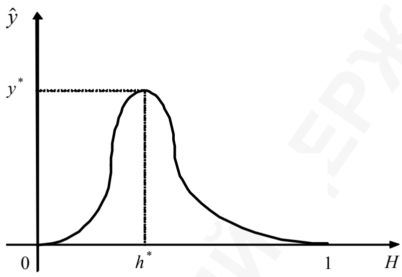
1.7 Оценивание параметров функции налоговых поступлений

В настоящее время хорошо осознаётся тот факт, что налоговая система оказывает большое влияние на развитие народного хозяйства. Необоснованно завышенный или, наоборот, заниженный уровень налоговых отчислений приводит к спаду производства, упадку бюджетной сферы и другим негативным последствиям. В этой связи следует отметить, что недостаточно, например, утверждать, что совокупный налог должен равняться 40%. Требуется еще указать, к каким последствиям это приведет или какие показатели будут наилучшими при 40%, и дать научное обоснование этому. Иначе возникает вопрос: почему именно 40%, а не 38% или 21.6%, или 58.14%?

Поставим следующую задачу: какие должны быть налоговые ставки, чтобы общие налоговые поступления принимали наибольшее значение. Считаем, что налоговые ставки постоянны для всех товаров и услуг.

Очевидно, что при нулевой налоговой ставке (H) налоговые поступления (Y) будут равны нулю. При стопроцентной ставке все оборотные средства производителей изымаются, закупка ресурсов становится невозможной, производство останавливается и налоговые поступления резко снижаются к нулевым. Пусть $\hat{y}(H)$ - кривая налоговых поступлений. Тогда $\hat{y}(0)=\hat{y}(1)=0$. Учитывая, что $\hat{y}(H)>0$ при $0 < H < 1$, заключаем, что в некоторой внутренней точке $h^* \in (0, 1)$ кривая $\hat{y}(H)$ будет принимать наибольшее значение. Лаффер выдвинул гипотезу, что кривая $\hat{y}(H)$ имеет один максимум и приблизительно симметрична относительно прямой $H = h^*$ вблизи точки максимума (см. рис.1.15*а*). Поскольку представляет интерес поведение кривой $\hat{y}(H)$ вблизи точки h^*, то в качестве функции налоговых поступлений можно взять функцию

$$\hat{y}(H) = y^* e^{-c(H-h^*)^2},\ c > 0. \qquad (1.60)$$



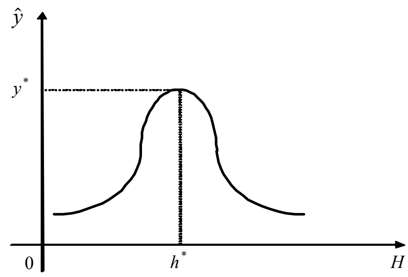


Рисунок 1.15

Кривая Лаффера (1.60) имеет вид, представленный на рис.1.15 *б*. Точка (h^*, y^*) является оптимальной: при ставке налога $H = h^*$ налоговые поступления принимают наибольшее значение, равное y^*. Однако значения параметров h^*, y^*, c неизвестны и подлежат оцениванию.

Составим регрессионную модель (U - случайное возмущение)

$$Y = y^* e^{-c(H-h^*)^2+U}, \ c > 0. \tag{1.61}$$

Прологарифмировав, находим

$$\ln Y = \ln y^* - cH^2 + 2cHh^* - ch^{*2} + U.$$

Замена

$$Z = \ln Y, \quad a = \ln y^* - ch^{*2}, \quad b = 2ch^* \tag{1.62}$$

позволяет получить параболическую регрессию

$$Z = a + bH - cH^2 + U, \tag{1.63}$$

линейную относительно оцениваемых параметров и возмущения.

Система нормальных уравнений (1.47) для определения оценок $\hat{a}$, $\hat{b}$, $\hat{c}$ параметров a, b, c в данном случае приобретает вид

$$\begin{cases} n\hat{a} + \hat{b}\Sigma H_i - \hat{c}\Sigma H_i^2 = \Sigma \ln Y_i, \\ \hat{a}\Sigma H_i + \hat{b}\Sigma H_i^2 - \hat{c}\Sigma H_i^3 = \Sigma H_i \ln Y_i, \\ \hat{a}\Sigma H_i^2 + \hat{b}\Sigma H_i^3 - \hat{c}\Sigma H_i^4 = \Sigma H_i^2 \ln Y_i. \end{cases} \tag{1.64}$$

Если известны значения $\hat{a}$, $\hat{b}$, $\hat{c}$, то, учитывая (1.62), найдём оценки параметров h^*, y^*. Имеем

$$\widehat{h}^* = \frac{\widehat{b}}{2\widehat{c}}, \quad \widehat{y}^* = e^{\widehat{a} + \widehat{c}\widehat{h}^{*2}}. \tag{1.65}$$

Итак, налоговая политика государства при постоянных налоговых ставках для всех товаров и услуг была бы обоснованной, если ставка налога равнялась величине $\hat{h}^{*2}$. В этом случае будут достигаться наибольшие налоговые поступления в размере $\hat{y}^*$.

1.8 Практическое задание Моделирование роста прибыли фирмы

В таблице 1.13 приведена динамика роста прибыли некоторой фирмы за последние n лет в процентах к базовому (нулевому) году. Необходимо:

1 Определить наилучшую функцию регрессии методом характерных средних из девяти возможных функций.

2 Для установленной зависимости роста прибыли от номера года методом наименьших квадратов определить оценки неизвестных параметров.

3 Оценить средний рост прибыли фирмы в год.

4 Сделать оценку величины прибыли через 3,5 года с начала работы.

5 Спрогнозировать величину прибыли на конец текущего и середину следующего годов.

Т а б л и ц а 1.13

Номер	**Год**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	104,3	112,8	118,4	130,1	148,7	172,5	202,6	220,8	250,4	295,9
2	110,1	130,4	158,3	170,7	198,5	230,9	280,6	320,6	360,4	392,8
3	115,4	132,7	161,2	181,4	214,5	241,2	272,9	301,4	320,5	350,4
4	109,2	122,8	141,5	160,6	187,4	222,5	270,4	304,5	331,7	362,8
5	112,0	133,4	149,7	162,5	183,2	201,9	232,6	259,6	280,4	312,0
6	121,5	142,4	159,1	173,6	190,9	212,2	237,8	263,4	287,5	-
7	115,0	132,5	146,4	159,8	172,4	191,6	210,3	228,4	250,0	-
8	110,5	124,6	142,4	162,5	181,7	198,9	215,6	240,4	265,5	-
9	114,0	128,5	151,2	173,4	195,8	220,6	248,7	280,0	310,4	-
10	125,4	141,2	168,4	181,3	197,6	211,9	228,3	248,1	266,0	-
11	118,6	132,4	152,7	168,9	181,4	196,8	213,5	235,5	-	-
12	122,0	141,6	156,4	179,8	192,1	216,7	248,0	270,4	-	-
13	128,1	154,2	182,5	203,4	221,9	243,4	272,5	300,0	-	-
14	120,0	135,6	151,2	170,4	188,2	206,5	225,0	250,5	-	-
15	114,1	128,3	150,4	166,2	183,0	200,6	215,3	240,0	-	-
16	120,5	132,4	151,2	172,8	193,4	212,8	240,1	272,4	300,0	-
17	130,4	151,2	174,4	196,8	220,3	252,4	274,5	290,4	315,0	-

Продолжение таблицы 1.13

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
18	141,2	170,4	191,3	220,8	254,2	282,1	305,4	332,9	370,5	-
19	118,3	132,6	149,4	170,2	190,5	221,4	248,3	275,0	300,0	-
20	122,8	145,6	164,3	179,7	196,4	212,5	234,6	252,7	284,5	-
21	115,4	132,6	148,3	164,8	185,1	201,4	220,3	241,7	270,0	296,5
22	109,3	124,5	140,0	157,3	172,7	191,9	207,3	223,9	240,5	260,0
23	114,2	128,3	147,4	160,7	175,0	194,8	210,0	228,4	245,7	265,5
24	118,0	132,1	145,2	165,3	190,4	221,7	243,4	270,5	294,8	320,4
25	132,4	150,2	171,8	198,3	230,0	261,2	293,4	321,2	348,3	375,2
26	122,8	141,2	157,4	172,8	194,7	218,3	245,8	270,1	297,3	330,0
27	111,0	132,4	152,7	170,6	188,5	199,9	220,0	241,4	262,5	-
28	117,4	138,1	160,2	181,4	203,0	228,1	255,3	282,7	320,0	-
29	124,2	150,0	178,3	210,7	251,2	290,7	331,8	372,4	410,8	-
30	119,4	140,3	170,5	190,8	214,1	242,7	273,2	302,6	350,4	-
31	115,6	126,2	138,1	150,4	163,5	180,2	200,0	-	-	-
32	120,0	141,2	160,3	188,4	210,3	235,3	260,0	-	-	-
33	112,4	117,4	130,3	150,2	177,3	204,0	223,7	-	-	-
34	107,2	121,3	140,4	162,5	180,7	201,2	225,6	-	-	-
35	115,6	137,2	155,8	174,6	198,9	215,4	240,3	-	-	-
36	112,8	123,4	137,7	154,0	171,2	193,6	210,8	-	-	-
37	113,4	132,5	154,6	178,2	199,4	230,4	262,5	-	-	-
38	107,3	120,5	134,6	152,7	165,4	184,6	202,5	-	-	-
39	114,1	123,7	138,6	154,2	170,6	190,2	210,6	-	-	-
40	110,5	122,7	137,4	150,2	164,3	180,2	197,5	-	-	-
41	112,4	120,8	132,4	141,2	150,4	162,3	177,5	-	-	-
42	108,4	121,2	137,3	152,8	171,4	192,3	215,6	-	-	-
43	120,4	148,5	172,4	198,5	227,4	258,7	286,5	-	-	-
44	127,5	141,3	158,6	177,4	196,8	213,4	240,1	-	-	-
45	121,4	147,3	173,5	199,6	227,3	252,4	280,0	-	-	-
46	115,3	128,3	142,7	160,4	180,5	199,6	227,4	-	-	-
47	112,8	127,3	141,4	159,7	173,8	197,6	215,0	-	-	-
48	116,4	127,4	143,7	162,5	187,4	205,6	230,4	-	-	-
49	108,6	121,6	136,4	158,3	176,4	198,2	218,6	-	-	-
50	110,4	127,6	142,4	160,2	182,6	202,7	225,0	-	-	-

ГЛАВА 2 МОДЕЛИРОВАНИЕ ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ С ПОМОЩЬЮ ОРТОГОНАЛЬНОЙ ПОЛИНОМИАЛЬНОЙ РЕГРЕССИИ

2.1 Ортогональная полиномиальная регрессия

Полиномиальной регрессией называется модель вида

$$f_m(x) = a_0 + a_1 x + \ldots + a_m x^m \,, \qquad (2.1)$$

где $f_m(x) = M(Y)$. Она линейна относительно параметров a_0, a_1, …, a_m и нелинейна относительно переменной *x*.

В случае, если степень *m* многочлена (2.1) выбрана заранее, методом наименьших квадратов могут быть найдены *МНК*-оценки $\hat{a}_0$, $\hat{a}_1, \ldots,$ $\hat{a}_m$ неизвестных параметров по методике, изложенной в предыдущей главе, можно провести анализ модели. Так что здесь не возникает никаких трудностей. Однако степень многочлена, как правило, неизвестна, и при переходе от *m* степени к (*m*+1) путем добавления слагаемого вида $a_{m+1} x^{m+1}$ *МНК*-оценки параметров модели нужно вычислять заново, а те, что были найдены ранее, для меньших степеней, оказываются абсолютно бесполезными. Заново также придется проводить анализ модели. Кроме того, при выборе степени многочлена следует обратить внимание на следующее противоречие. С одной стороны, желательно рассматривать более широкий класс функций, т.е. многочлены более высокой степени. С другой стороны, такое расширение ведет к увеличению количества неизвестных коэффициентов и соответственно их оценок (числа степеней свободы $\ell = m + 1$) и, следовательно, к возможному увеличению дисперсии.

Другой подход заключается в постепенном уточнении уравнения полиномиальной модели и основан на применении ортогональных многочленов. Пусть

$$f_m(x) = в_0 P_0(x) + в_1 P_1(x) + \ldots + в_m P_m(x), \qquad (2.2)$$

где $P_0(x) = 1$, $P_1(x), \ldots, P_m(x)$ – многочлены 1-й, ..., *m*-й степени (такие многочлены линейно независимы). Уравнение оцененной регрессии для модели (2.2) следующее:

$$\hat{y}_m(x) = \hat{в}_0 P_0(x) + \hat{в}_1 P_1(x) + \ldots + \hat{в}_m P_m(x), \qquad (2.3)$$

причем *МНК*-оценки $\hat{в}_0$, $\hat{в}_1$, …, $\hat{в}_m$ определяются из системы линейных уравнений:

$$\begin{cases} \hat{в}_0\sum P_0^2(x_i)+\hat{в}_1\sum P_0(x_i)P_1(x_i)+...+\hat{в}_m\sum P_0(x_i)P_m(x_i)=\sum y_iP_0(x_i), \\ \hat{в}_0\sum P_0(x_i)P_1(x_i)+\hat{в}_1\sum P_1^2(x_i)+...+\hat{в}_m\sum P_1(x_i)P_m(x_i)=\sum y_iP_1(x_i), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{в}_0\sum P_0(x_i)P_m(x_i)+\hat{в}_1\sum P_1(x_i)P_m(x_i)+...+\hat{в}_m\sum P_m^2(x_i)=\sum y_iP_m(x_i). \end{cases} \quad (2.4)$$

Потребуем в системе нормальных уравнений (2.4) выпол-нение условий при $k \neq \ell$:

$$\sum_{i=1}^{n} P_k(x_i)P_\ell(x_i)=0, \quad k,\ell=0,\ 1,\ \ldots,\ m. \quad (2.5)$$

Многочлены $P_0(x)$, $P_1(x)$, ..., $P_m(x)$, удовлетворяющие услови-ям (2.5), называются ортогональными. Полиномиальная регрес-сионная модель (2.2) в этом случае называется ***ортогональной***.

Нетрудно заметить, что для ортогональных моделей каж-дое уравнение в (2.4) содержит лишь одно неизвестное (матри-ца системы диагональная), и мы сразу находим значения всех оценок:

$$\hat{в}_k=\frac{\sum y_iP_k(x_i)}{\sum P_k^2(x_i)}, \quad k=0,\ 1,\ ...,\ m. \quad (2.6)$$

Можно показать, что МНК-оценки $\hat{в}_0$, $\hat{в}_1$, …, $\hat{в}_m$ имеют нормальное распределение, не коррелируют между собой, и корреляционная матрица, соответствующая им, имеет диагональный вид. Это свойство *МНК*-оценок $\hat{в}_0, \hat{в}_1, \ldots, \hat{в}_m$ значительно упрощает дисперсионный анализ ортогональной модели (2.2).

Из соотношений (2.5) вытекает, что ортогональные многочлены $P_0(x)$, $P_1(x)$, …, $P_m(x)$ определяются неоднозначно. Ортогональными будут также многочлены $c_0P_0(x)$, $c_1P_1(x)$, …, $c_mP_m(x)$, где c_0, c_1, ..., c_m -произвольные, отличные от нуля, постоянные. Однако можно показать, что выбор этих постоянных не влияет на свойства ортогональных моделей. Значения множителей-констант выбираются из соображений удобства.

Построим пример линейной ортогональной модели. Модель $f(x)=a_0+a_1x$ - неортогональна, так как многочлены 1 и x не удовлетворяют условию $\sum 1\cdot x_i=0$ для произвольной выборки. А вот альтернативная модель $f(x)=в_0+в_1(x-\bar{x})$ оказывается ортогональной, поскольку $\sum 1 \cdot (x_i - \bar{x}) = 0$ для любой выборки. Из (2.6) находим значения оценок

$$\widehat{в}_0 = \frac{1}{n}\sum y_i = \bar{y}, \quad \widehat{в}_1 = \frac{\sum y_i (x_i - \bar{x})}{\sum (x_i - \bar{x})^2}.$$

Предположим, что для заданной выборки объема n система ортогональных многочленов известна заранее. Тогда при фиксированном m для ортогональной модели (2.2), используя формулы (2.6), можно сразу записать оцененную модель (2.3). Для перехода к ортогональной модели (m+1)-й степени

$$f_{m+1}(x) = в_0 P_0(x) + в_1 P_1(x) + \ldots + в_m P_m(x) + в_{m+1} P_{m+1}(x) \qquad (2.7)$$

необходимо вычислить только значение оценки $\hat{в}_{m+1}$ параметра $в_{m+1}$ в дополнительном слагаемом $\varphi_{m+1}(x) = в_{m+1} P_{m+1}(x)$. Все остальные слагаемые уже оценены в предыдущей модели m-й степени. Возникает естественный вопрос: нуждается ли модель (2.2) в улучшении, т.е. нужен ли переход к ортогональной модели более высокой степени?

Качество регрессионной модели, мы знаем, определяется коэффициентом детерминации $R^2 = \theta^2$, который тем больше, чем меньше сумма квадратов отклонений. Учитывая условия ортогональности (2.5) и выражения (2.6) для *МНК*-оценок пара-метров модели (2.2), вычислим сумму квадратов отклонений S_m указанной модели. Имеем

$$S_m = \sum \left[y_i - \hat{в}_0 P_0(x_i) - \hat{в}_1 P_1(x_i) - \ldots - \hat{в}_m P_m(x_i) \right]^2 =$$
$$= \sum y_i^2 - 2\hat{в}_0 \sum y_i P_0(x_i) - 2\hat{в}_1 \sum y_i P_1(x_i) - \ldots - 2\hat{в}_m \sum y_i P_m(x_i) +$$
$$+ \widehat{в}_0^2 \sum P_0^2(x_i) + \widehat{в}_1^2 \sum P_1^2(x_i) + \ldots + \widehat{в}_m^2 \sum P_m^2(x_i) =$$
$$= \sum y_i^2 - \hat{в}_0^2 \sum P_0^2(x_i) - \hat{в}_1^2 \sum P_1^2(x_i) - \ldots - \hat{в}_m^2 \sum P_m^2(x_i). \quad (2.8)$$

Так как $P_0(x) = 1$ и $\hat{в}_0 = \bar{y}$, то, очевидно,

$$\sum (y_i - \bar{y})^2 = \hat{в}_1^2 \sum P_1^2(x_i) + \ldots + \hat{в}_m^2 \sum P_m^2(x_i) + S_m. \qquad (2.9)$$

Общая вариация $v_{общ} = \sum (y_i - \bar{y})^2$ выборочных данных около их среднего разлагается на две составляющие: вариацию $v_м$ зна-чений, обусловленных моделью, и остаточную вариацию $v_{ост}$, связанную со случайностью модели. В нашем случае

$$v_{общ} = v_M + v_{ост},$$

$$\nu_м = \hat{в}_1^2 \sum P_1^2(x_i) + \ldots + \hat{в}_m^2 \sum P_m^2(x_i), \;\; \nu_{ост} = S_m. \qquad (2.10)$$

Из (2.10) вытекает, что с увеличением m сумма $\nu_м$ возрастает и, как следствие, $\nu_{ост}$ уменьшается, что ведет к увеличению коэффициента детерминации:

$$\theta^2 = 1 - \frac{\nu_{ост}}{\nu_{общ}}. \qquad (2.11)$$

С другой стороны, с ростом порядка m модели может возрастать выборочная дисперсия возмущений:

$$D_m = \frac{S_m}{n-m-1}. \qquad (2.12)$$

Более того, очевидно, $m<n$-1.

Приведенные рассуждения указывают на то, что поправка $\varphi_{m+1} = в_{m+1} P_{m+1}(x)$ может быть принята, но может быть и отклонена.

Принятие поправки означает уменьшение остаточной суммы квадратов отклонений, причем

$$S_{m+1} = S_m - \hat{в}_{m+1}^2 \sum P_{m+1}^2(x_i), \qquad (2.13)$$

а также то, что дисперсия $D_{m+1} < D_m$, причем значимо. Отклонение поправки можно рассматривать как следствие спра-ведливости нулевой гипотезы $H_0 : в_{m+1} = 0$. Альтернативная ги-потеза $H_1 : в_{m+1} \neq 0$ и ее подтверждение будут свидетельствовать о значимости поправки.

Таким образом, принятие или отклонение поправки связано со значимостью оценки $\hat{в}_{m+1}$. Двусторонний критерий значимости оценки $\hat{в}_{m+1}$ имеет вид

$$\left| \frac{\hat{в}_{m+1}}{\hat{\sigma}_{\hat{в}_{m+1}}} \right| > t_{кр}, \qquad (2.14)$$

$$\hat{\sigma}_{\hat{в}_{m+1}} = \sqrt{\frac{S_{m+1}}{(n-m-2)\sum P_{m+1}^2(x_i)}} = \frac{\hat{\sigma}_u}{\sqrt{\sum P_{m+1}^2(x_i)}},$$

где $\hat{\sigma}_{\hat{в}_{m+1}}$ - стандартная ошибка оценки $\hat{в}_{m+1}$, $t_{кр} = t_{1-\frac{\alpha}{2}}$ - квантиль распределения Стьюдента с n-m-2 степенями свободы при уров-не значимости α.

Другой подход к выяснению значимости поправки связан со сравнением дисперсий D_{m+1} и D_m. Поправка должна быть безусловно отвергнута, если окажется, что $D_{m+1} \geq D_m$. Но даже и уменьшение дисперсии не означает, что поправку $\varphi_{m+1}(x)$ нужно принять - это уменьшение должно быть значимым. Значимость уменьшения дисперсии проверяется по одно-стороннему критерию Фишера:

$$\frac{D_m}{D_{m+1}} > F_{кр}, \tag{2.15}$$

где критическое значение $F_{кр} = F_{1-\alpha}$ - квантиль F-распределения с числами степеней свободы n-m-1 и n-m-2 при уровне значимости α. Итак, для оценки вклада в ортогональную модель того члена, который включен последним, можно использовать критерии (2.14) и (2.15). На практике решение обычно заканчивают, если считается, что два последних члена оказались равными нулю. Это необходимое условие, поскольку если подгонять к выборочным данным полином нечетной степени, то члены четных степеней (не считая свободного члена) скорее всего будут вносить малый вклад и наоборот.

Ортогональные многочлены степени выше первой в общем случае, когда значения x расположены произвольно, приходится строить специально для каждого набора данных, что в значительной степени снижает преимущество ортогональных моделей, обусловленных легкостью решения системы нормальных уравнений. Если же узлы x_1, x_2, ..., x_n равномерно расположены, т.е. находятся друг от друга на одинаковом расстоянии ($x_{i+1} = x_i + h$), то ортогональные многочлены зависят только от объема выборки n и являются стандартными, для которых существуют специальные таблицы.

Если $x_i = x_1 + (i-1)h$, то замена

$$z = \frac{x - x_1}{h} + 1 \tag{2.16}$$

позволяет каждое значение x_i заменить своим номером $z_i = i$. Поскольку ортогональные многочлены определяются с точностью до множителей-констант, то в случае целочисленного аргумента удобно эти множители выбрать так, чтобы старшие коэффициенты всех ортогональных многочленов были равны 1. В этом случае система ортогональных многочленов определяется однозначно следующим образом :

$$P_0(i) = 1,\ P_1(i) = i - \frac{n+1}{2},$$

$$P_{m+1}(i) = P_1(i)P_m(i) - \frac{m^2(n^2 - m^2)}{4(4m^2 - 1)}P_{m-1}(i),\ \ m = 1, 2, ... \tag{2.17}$$

Например

$$P_2(i) = i^2 - (n+1)i + \frac{(n+1)(n+2)}{6},$$

$$P_3(i) = i^3 - \frac{3(n+1)}{2} i^2 + \frac{6n^2+15n+11}{10} i - \frac{(n+1)(n+2)(n+3)}{20}.$$

При целочисленных значениях аргумента для фигурирующих в (2.6) сумм $\sum P_k^2(i)$ можно использовать сокращенную формулу

$$\sum P_k^2(i) = \frac{(k!)^2 n(n^2-1)(n^2-4)...(n^2-k^2)}{\left[(2k-1)!!\right]^2 2^{2k}(2k+1)}.$$

Здесь выражение (2k-1)!! означает произведение всех нечетных целых чисел от 1 до 2k-1. В практических расчетах последнюю формулу удобно применять в рекуррентной форме:

$$\sum P_k^2(i) = \frac{k^2(n^2-k^2)}{4(4k^2-1)} \sum P_{k-1}^2(i), \quad k = 1,2,... \; , \qquad (2.18)$$

где $\sum P_0^2(i) = n.$

После того как уравнение регрессии будет найдено и исследовано, переменную z (2.16) опять заменяют перво-начальной переменной x ($x = x_1 + (t - 1)h$).

Применим изложенный метод к исследованию изменения уровня дефицита бюджета в СССР в 80-е годы по данным табл.1.7. Сделаем замену $z = t - 1979$, тогда $z_i = i$, $\bar{z} = 5$ и, следо-вательно, $P_0(z) = 1$, $P_1(z) = z - 5$. По второму столбцу табл.2.1 на-ходим $\hat{в}_0 = \bar{y} = \frac{38,8}{9} = 4,31.$ По формуле (2.18)

$$\sum P_1^2(i) = \frac{1 \cdot 80}{4 \cdot 3} \cdot 9 = 60, \qquad \hat{в}_1 = \frac{49,2}{60} = 0,82.$$

Таким образом, ортогональная модель первого порядка имеет вид

$$\hat{y}_1(z) = 4,31 + 0,82(z-5).$$

Вычисляем дисперсию D_1. Имеем

$$S_0 = \sum (y_i - \bar{y})^2 = 61,23,$$

$$S_1 = S_0 - \hat{в}_1^2 \sum P_1^2(i) = 61,23 - 40,34 = 20,89.$$

$$D_1 = \frac{S_1}{n-2} = \frac{20,89}{7} = 2,98.$$

Коэффициент детерминации линейной модели

$$\theta^2 = r^2 = 1 - \frac{v_{ост}}{v_{общ}} = 1 - \frac{S_1}{S_0} = 1 - \frac{20,89}{61,23} = 0,659.$$

Значение $\theta^2 = 0,659$ означает, что линейная модель объясняет 65,9% дисперсии Y, остальные 34,1% не обусловлены линейной моделью. Проверяем критерий значимости (2.14) оценки $\hat{в}_1$ коэффициента регрессии $в_1$. При уровне значимости $\alpha = 0{,}05$ и числе степеней свободы n-2=7 квантиль $t_{0,975}$ =2,36.

$$\hat{\sigma}_{\hat{b}_1} = \sqrt{\frac{S_1}{(n-2)\sum P_1^2(i)}} = \sqrt{\frac{20{,}89}{7 \cdot 60}} = 0{,}223\,, \quad \frac{\hat{b}_1}{\hat{\sigma}_{\hat{b}_1}} = \frac{0{,}82}{0{,}223} = 3{,}68\,.$$

Так как 3,68 > 2,36, то $в_1 \neq 0$.

Переходим к вычислению и оценке ортогональной ре-грессии второго порядка. В нашем случае $P_2(i) = i^2 - 10i + \frac{55}{3}$, от-сюда находим числа $P_2(i)$, приведенные в пятом столбце табл.2.1. Далее

$$\sum P_2^2(i) = \frac{4 \cdot 77}{60} \cdot 60 = 308, \quad \hat{в}_2 = \frac{69{,}26}{308} = 0{,}225,$$

$$\hat{y}_2(z) = \hat{y}_1(z) + 0{,}225(z^2 - 10z + 18{,}33).$$

Так как $z^2 - 10z = (z-5)^2 - 25,$ то полагая $\tau = z - 5,$ получаем

$$\hat{y}_2(\tau) = 0{,}225\tau^2 + 0{,}82\tau + 2{,}81,$$

что, как и следовало ожидать, совпадает с уравнением параболической регрессии (см. § 1.3), полученной методом наименьших квадратов. Теперь

$$S_2 = S_1 - \hat{в}_2^2 \sum P_2^2(i) = 20{,}89 - 15{,}59 = 5{,}30,$$

$$\theta^2 = 1 - \frac{S_2}{S_0} = 1 - \frac{5{,}30}{61{,}23} = 0{,}913, \quad D_2 = \frac{S_2}{n-3} = 0{,}88.$$

Проверим, является ли уравнение второго порядка существенно лучшим приближением регрессии, чем уравнение первого порядка. При уровне значимости $\alpha = 0{,}05$ и числе степеней свободы n-3=6 $t_{0,975} = 2{,}45$.

$$\hat{\sigma}_{\hat{b}_2} = \sqrt{\frac{S_2}{(n-3)\sum P_2^2(i)}} = \sqrt{\frac{5{,}30}{6 \cdot 308}} = 0{,}054\,, \quad \frac{\hat{b}_2}{\hat{\sigma}_{\hat{b}_2}} = \frac{0{,}225}{0{,}054} = 4{,}17\,.$$

Видим, что 4,17 > 2,45, следовательно, $в_2 \neq 0.$ Параболическая модель объясняет 91,3% дисперсии *Y*.

Для регрессии третьего порядка

$$P_3(i) = i^3 - 15i^2 + 63{,}2i - 66, \quad \sum P_3^2(i) = \frac{9 \cdot 72}{4 \cdot 35} \cdot 308 = 1425{,}6,$$

$$\hat{в}_3 = \frac{-2{,}30}{1425{,}6} = -0{,}0016,\ S_3 = S_2 - \hat{в}_3^2 \sum P_3^2(i) = 5{,}30 - 0{,}0032 = 5{,}2968,$$

$$D_3 = \frac{S_3}{n-4} = \frac{5{,}2968}{5} = 1{,}06, \quad D_3 > D_2.$$

Таким образом, переход к уравнению третьего порядка ухудшает регрессионную модель. Считаем, что $\hat{в}_3 = 0$.

Т а б л и ц а 2.1 - *Оценивание параметров ортогональных полиномиальных моделей*

I	y_i	$P_1(i)$	$y_iP_1(i)$	$P_2(i)$	$y_iP_2(i)$	$P_3(i)$	$y_iP_3(i)$	$P_4(i)$	$y_iP_4(i)$
1	2,9	- 4	-11,6	9, 33	27,06	-16,79	-48,69	23,98	69,54
2	2,3	- 3	- 6,9	2,33	5,36	8,41	19,34	-36,01	-82,82
3	3,1	- 2	- 6,2	-2,67	-8,23	15,61	48,40	-18,85	-58,44
4	2,2	- 1	- 2,2	-5,67	-12,47	10,80	23,76	15,44	33,97
5	2,0	0	0	-6,67	-13,34	0	0	30,88	61,76
6	2,7	1	2,7	-5,67	-15,31	-10,80	-29,16	15,44	41,69
7	6,5	2	13,0	-2,67	-17,35	-15,61	-101,46	-18,85	-122,53
8	8,0	3	24,0	2,33	18,64	-8,41	-67,28	-36,01	-288,08
9	9,1	4	36,4	9,33	84,90	16,79	152,79	23,98	218,22
Σ	38,8	0	49,2	0	69,26	0	-2,30	0	-126,69

Построим регрессию четвертого порядка. Значения $P_4(i)$ находим по рекуррентной формуле (2.17):

$$P_4(i) = P_1(i)P_3(i) - \tfrac{162}{35} P_2(i),$$

используя известные значения $P_1(i)$, $P_2(i)$ и $P_3(i)$, полу чим

$$\sum P_4^2(i) = \frac{16 \cdot 65}{4 \cdot 63} \cdot 1425{,}6 = 5883{,}4,\ \hat{в}_4 = \frac{-126{,}69}{5883{,}4} = -0{,}0215,$$

$$S_4 = S_3 - \hat{в}_4^2 \sum P_4^2(i) = 2{,}5687,$$

$$D_4 = \frac{S_4}{n-5} = \frac{2{,}5687}{4} = 0{,}642.$$

Делаем проверку значимости оценки $\hat{в}_4$. При $\alpha = 0{,}05$ и n-5=4 $t_{0{,}975} = 2{,}78$.

$$\widehat{\sigma}_{\widehat{в}_4}=\sqrt{\frac{S_4}{(n-5)\sum P_4^2(i)}}=\sqrt{\frac{2{,}5687}{4\cdot 5883{,}4}}=0{,}0104\,,\quad \left|\frac{\widehat{в}_4}{\widehat{\sigma}_{\widehat{в}_4}}\right|=2{,}07<2{,}78\,.$$

Критерий значимости $\hat{в}_4$ не выполняется, считаем, что $\hat{в}_4=0$. Используем критерий Фишера (2.15). По таблице приложения Б находим $F_{0,95}(6;4)=6{,}16$, $\frac{D_2}{D_4}=\frac{0{,}88}{0{,}642}=1{,}37$. Так как 1,37 < 6,16, то критерий не выполняется, и мы отвергаем поправку $\varphi_4(z)$. Таким образом, аргументы против $в_4=0$ слабы, и мы приходим к выводу, что имеющиеся данные вполне приемлемо аппроксимируются квадратичной моделью $\widehat{y}_2(\tau)=0{,}225\tau^2+0{,}82\tau+2{,}81$.

2.2 Практическое задание Определение оптимальной полиномиальной регрессии

Пусть для некоторого сельхозпредприятия известны средние урожайности зерновых культур за последние n лет. Необходимо:

1. Методом ортогональных многочленов построить полиномиальные регрессии 1-го, 2-го, …, m-го порядков зависимости средней урожайности зерновых от номера года.

2. Для каждой регрессии вычислить коэффициент детерминации и при уровне значимости α =0,05 проверить значимость нового коэффициента с помощью
а) критерия Стьюдента; б) критерия Фишера.

3. Определить оптимальную степень полиномиальной регрессии, считая, что улучшение результата не происходит, если два последних коэффициента оказались незначимыми. Оцените качество оптимальной регрессии.

4. Сделать прогноз урожайности зерновых на текущий и последующий годы.

Т а б л и ц а 2.2

Но-мер	**Год**								
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1	23,5	26,4	28,2	27,3	25,6	22,4	28,3	30,5	32,8
2	28,4	27,1	25,4	23,2	25,8	27,6	29,2	31,4	33,5
3	32,1	34,2	31,3	30,0	28,3	26,2	28,4	31,2	33,7
4	26,2	28,4	29,3	27,2	25,4	28,1	29,6	32,7	34,0
5	21,3	22,8	25,2	23,6	22,5	24,0	26,3	27,8	29,6
6	27,2	26,1	25,3	27,4	28,5	29,6	30,4	32,7	34,0
7	30,6	32,4	33,8	31,4	29,6	30,8	32,7	34,5	36,2
8	27,8	26,4	25,2	27,3	28,7	29,3	30,6	32,4	35,3
9	30,4	31,2	32,6	30,8	29,7	28,0	29,4	30,7	33,6
10	24,7	23,2	22,3	24,1	25,4	26,7	27,8	28,4	29,6
11	28,0	28,8	29,7	30,3	32,4	31,2	30,1	28,3	27,0
12	31,2	30,4	28,3	27,6	26,4	28,5	30,2	33,3	36,1
13	35,4	34,2	33,6	32,1	32,7	33,8	35,6	37,2	38,8
14	19,5	19,2	18,6	18,0	18,5	19,4	21,3	23,8	25,6
15	22,4	23,8	24,4	26,0	27,1	26,2	25,4	24,3	23,5
16	28,3	29,2	30,5	31,3	32,4	31,5	30,2	28,3	27,0
17	24,7	23,9	23,4	22,5	22,9	24,0	25,6	27,2	29,4
18	29,1	29,6	30,5	31,2	32,4	31,7	31,1	30,2	28,8
19	30,0	29,1	28,4	27,2	28,3	29,6	30,7	32,8	35,0
20	19,2	18,6	18,1	17,5	18,2	18,8	20,0	22,3	24,4
21	27,3	26,2	24,3	23,5	24,5	26,0	27,4	28,1	30,0
22	29,6	30,2	30,9	31,4	32,8	31,7	30,8	29,8	28,4
23	21,7	21,2	20,2	19,8	20,4	21,3	23,0	25,2	27,5
24	27,6	28,1	28,9	29,4	30,2	29,2	28,0	26,3	25,0
25	24,2	26,3	27,5	28,6	29,9	29,2	28,3	27,0	24,5
26	37,3	36,2	34,3	32,2	34,2	35,1	37,2	38,3	40,2
27	40,3	38,1	37,2	35,4	36,8	37,9	39,6	40,8	43,0
28	33,2	36,4	38,3	38,8	38,0	36,4	35,0	31,1	32,2
29	26,5	24,2	22,7	20,4	23,2	24,8	27,4	29,9	32,6
30	17,8	19,4	22,3	24,1	25,8	24,3	23,0	22,1	20,8
31	24,8	26,1	28,3	29,4	28,2	27,1	26,2	25,0	24,0
32	25,6	24,3	23,4	22,0	23,0	24,5	25,9	26,7	28,2
33	22,1	24,3	26,2	27,7	26,5	25,1	23,7	22,8	22,0
34	19,8	22,4	24,2	25,6	23,1	22,6	22,1	20,4	18,6

Продолжение таблицы 2.2

	1	2	3	4	5	6	7	8	9
35	31,4	33,2	34,9	36,4	35,2	33,3	31,2	30,1	28,4
36	32,7	30,4	29,3	28,7	29,5	31,4	32,5	30,0	36,0
37	34,2	35,3	35,8	36,4	35,3	34,1	33,7	32,8	31,8
38	28,8	24,3	26,7	28,0	29,4	30,0	28,7	27,2	26,0
39	19,4	20,6	21,8	23,2	22,4	21,8	21,0	20,1	19,4
40	30,5	29,4	28,3	27,6	28,7	28,0	27,1	25,6	24,5
41	32,4	30,6	28,4	27,0	28,6	29,8	32,5	34,0	36,2
42	27,4	28,6	30,5	31,4	30,6	29,2	27,1	25,8	24,5
43	28,6	27,2	26,3	25,5	27,6	29,8	32,4	35,6	37,0
44	25,6	27,2	27,7	29,2	28,4	27,7	26,4	25,2	24,0
45	27,4	28,1	30,0	32,5	33,0	32,1	30,6	28,4	27,2
46	25,1	24,2	23,4	22,6	24,2	26,7	28,1	29,6	31,0
47	24,6	25,6	27,2	28,3	29,6	28,4	27,5	26,1	24,8
48	28,4	27,1	26,2	25,4	26,7	28,3	29,4	32,1	34,5
49	35,6	33,2	32,1	30,4	31,8	32,4	33,6	35,6	37,0
50	32,1	30,6	29,2	28,4	27,2	26,1	28,4	31,2	33,2

ГЛАВА3 *МНОЖЕСТВЕННЫЙ РЕГРЕССИОННЫЙ АНАЛІЗ*

3.1 Оценивание множественной регрессионной модели

Определение *МНК*-оценок

Будем исходить из того, что между объясняемой и объясняющими переменными выбрана линейная связь. Если исходная модель нелинейна, то предполагаем, что с помощью некоторого преобразования она сведена к линейной модели. Имеем

$$Y = a_0x_0 + a_1x_1 + \ldots + a_mx_m + U. \qquad (3.1)$$

Здесь $x_0 \equiv 1$ - фиктивная переменная, введённая для удобства. Слагаемое U отражает влияние на Y других факторов, ошибки измерений, ошибки выбора модели.

Пусть с целью исследования линейной связи (3.1) проведена выборка объёма *n*. Тогда для наблюдаемых величин можно записать

$$y_i = a_0 x_{i0} + a_1 x_{i1} + \ldots + a_m x_{im} + u_i, \quad i = 1, 2, \ldots, n. \qquad (3.2)$$

В системе уравнений (3.2) постоянные коэффициенты $a_0, a_1, ..., a_m$ неизвестны и должны быть оценены. Если $\breve{a}_0, \breve{a}_1, ..., \breve{a}_m$ - возможные оценки параметров a_0, a_1, ..., a_m, то функция регрессии, соответствующая модели (3.1), имеет вид

$$\breve{y} = \breve{a}_0 x_0 + \breve{a}_1 x_1 + ... + \breve{a}_m x_m. \qquad (3.3)$$

Отклонения выборочных данных y_i от неё определяются величинами ($i = 1, 2, ..., n$):

$$\breve{u}_i = y_i - \breve{y}_i, \; \breve{y}_i = \breve{a}_0 x_{i0} + \breve{a}_1 x_{i1} + ... + \breve{a}_m x_{im}.$$

Отметим, что несмещенной оценкой неизвестной дисперсии возмущений σ_u^2 является выборочная дисперсия

$$\breve{\sigma}_u^2 = \frac{1}{n-l} \sum \left(\breve{u}_i - \bar{\breve{u}} \right)^2,$$

где l - число связей, накладываемых функцией регрессии на выборку. Тогда заключаем, что $l = m + 1$, т.е. общее число связей равно числу оценок, от которых зависит функция регрессии (3.3). Учитывая равенство

$$\bar{\breve{u}} = \frac{1}{n} \sum \breve{u}_i = 0,$$

получаем следующее выражение для выборочной дисперсии возмущений:

$$\breve{\sigma}_u^2 = \frac{1}{n-m-1} \sum \breve{u}_i^2 = \frac{1}{n-m-1} \sum \left(y_i - \breve{y}_i \right)^2.$$

Критерием выбора оценок $\breve{a}_0, \breve{a}_1, ..., \breve{a}_m$ в математической статистике является условие минимума дисперсии, которое при фиксированном значении *n-m-1* эквивалентно условию минимума функции ошибок

$$S = \sum\left(y_i - \breve{y}_i\right)^2 = \sum\left(y_i - \breve{a}_0 x_{i0} - \breve{a}_1 x_{i1} - \ldots - \breve{a}_m x_{im}\right)^2. \quad (3.4)$$

Имеем

$$\frac{\partial S}{\partial \breve{a}_0} = 0, \quad \frac{\partial S}{\partial \breve{a}_1} = 0, \quad \ldots, \quad \frac{\partial S}{\partial \breve{a}_m} = 0.$$

В результате для определения МНК-оценок $\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_m$ приходим к системе *m+1* линейных уравнений с *m+1* неизвестными

$$\begin{cases} \hat{a}_0 \sum x_{i0}^2 + \hat{a}_1 \sum x_{i0} x_{i1} + \ldots + \hat{a}_m \sum x_{i0} x_{im} = \sum y_i x_{i0}, \\ \hat{a}_0 \sum x_{i1} x_{i0} + \hat{a}_1 \sum x_{i1}^2 + \ldots + \hat{a}_m \sum x_{i1} x_{im} = \sum y_i x_{i1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\ \hat{a}_0 \sum x_{im} x_{i0} + \hat{a}_1 \sum x_{im} x_{i1} + \ldots + \hat{a}_m \sum x_{im}^2 = \sum y_i x_{im}. \end{cases} \quad (3.5)$$

Предполагая, что определитель системы уравнений (3.5) отличен от нуля, из нее находим единственные значения оценок $\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_m$, которые и обеспечивают минимальные значения функции ошибок (3.4) и выборочной дисперсии возмущений $\breve{\sigma}_u^2$.

Таким образом, функция регрессии, соответствующая *МНК*-оценкам $\hat{a}_0, \hat{a}_1, \ldots, \hat{a}_m$, имеет вид

$$\hat{y} = \hat{a}_0 x_0 + \hat{a}_1 x_1 + \ldots + \hat{a}_m x_m. \quad (3.6)$$

Несмещенной оценкой неизвестной дисперсии σ_u^2 является МНК-оценка

$$\hat{\sigma}_u^2 = \frac{1}{n-m-1} \sum \hat{u}_i^2 = \frac{1}{n-m-1} \sum\left(y_i - \hat{y}_i\right)^2. \quad (3.7)$$

Матричный способ оценки

Процесс оценивания регрессионной модели при $m \geq 2$ является довольно громоздким, поскольку приходится вычислять большое число сумм в (3.5) и решать системы уравнений с тремя и более неизвестными, что без использования ЭВМ весьма затруднительно. Если же в распоряжении пользователя ЭВМ имеются стандартные программы, позволяющие осуществлять действия над матрицами, то регрессионный анализ значительно упрощается.

Поступим точно так же, как и при матричном анализе однофакторных регрессионных моделей. Запишем регрессионное соотношение (3.2) для наблюдаемых величин в развёрнутом виде:

$$\begin{cases} y_1 = a_0 x_{10} + a_1 x_{11} + \ldots + a_m x_{1m} + u_1 , \\ y_2 = a_0 x_{20} + a_1 x_{21} + \ldots + a_m x_{2m} + u_2 , \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots , \\ y_n = a_0 x_{n0} + a_1 x_{n1} + \ldots + a_m x_{nm} + u_n \ . \end{cases} \tag{3.8}$$

Здесь $x_{i0} \equiv 1$ для всех $i = 1, 2, \ldots, n$. Матричная запись системы уравнений (3.8) такова:

$$\vec{y} = \vec{X}\vec{a} + \vec{u} , \tag{3.9}$$

где $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$, $\vec{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}$, $\vec{X} = \begin{bmatrix} x_{10} & x_{11} \ldots x_{1m} \\ x_{20} & x_{21} \ldots x_{2m} \\ \ldots\ldots & \ldots\ldots\ldots \\ x_{n0} & x_{n1} \ldots x_{nm} \end{bmatrix}$.

В линейных регрессионных моделях предполагается, что выборочные наблюдения должны быть такими, чтобы число степеней свободы $l = n - m - 1$ было больше 0, и чтобы матрица $\vec{X}$ имела полный столбцевой ранг $m + 1$. Из курса линейной алгебры известно, что в этом случае ранг транспонированной матрицы также равен $m + 1$, а симметричная матрица размерности $(m + 1) \times (m + 1)$

$$\vec{X}'\vec{X} = \begin{bmatrix} \sum x_{i0}^2 & \sum x_{i0}x_{i1} & \ldots & \sum x_{i0}x_{im} \\ \sum x_{i1}x_{i0} & \sum x_{i1}^2 & \ldots & \sum x_{i1}x_{im} \\ \ldots & \ldots & \ldots & \ldots \\ \sum x_{im}x_{i0} & \sum x_{im}x_{i1} & \ldots & \sum x_{im}^2 \end{bmatrix} \tag{3.10}$$

имеет ранг, равный $m+1$, и, следовательно, существует обратная матрица $\left(\vec{X}'\vec{X}\right)^{-1}$.

Нетрудно заметить, что система линейных уравнений (3.5), из которой определяются *МНК*-оценки $\hat{a}_0, \hat{a}_1, ..., \hat{a}_m$, может быть записана в виде

$$\vec{X}'\vec{X}\vec{\hat{a}} = \vec{X}'\vec{y}, \tag{3.11}$$

откуда находим вектор-столбец искомых *МНК*-оценок. Имеем

$$\vec{\hat{a}} = \left(\vec{X}'\vec{X}\right)^{-1}\vec{X}'\vec{y}. \tag{3.12}$$

Таким образом, вектор-оценку $\vec{\hat{a}}$ можно определять двумя способами: либо решая систему линейных уравнений (3.11), либо пользуясь формулой (3.12). Далее получаем

$$\vec{\hat{a}} = \vec{a} + \left(\vec{X}'\vec{X}\right)^{-1}\vec{X}'\vec{u}, \tag{3.13}$$

$$M(\vec{\hat{a}}) = \vec{a}\,, \tag{3.14}$$

$$\vec{\Sigma} = D(\vec{\hat{a}}) = \sigma_u^2 \left(\vec{X}'\vec{X}\right)^{-1}, \tag{3.15}$$

$$\vec{\hat{\Sigma}} = \hat{D}(\vec{\hat{a}}) = \hat{\sigma}_u^2 \left(\vec{X}'\vec{X}\right)^{-1}, \tag{3.16}$$

$$\hat{\sigma}_u^2 = \frac{\vec{\hat{u}}'\vec{\hat{u}}}{n - m - 1}\,,\ \vec{\hat{u}} = \vec{y} - \vec{X}\vec{\hat{a}}\,. \tag{3.17}$$

Коэффициент детерминации

Качество регрессионной модели, как и в случае однофакторных функций регрессии, будем характеризовать коэффициентом детерминации θ^2, который в случае линейной регрессии обозначается R^2 и равен квадрату выборочного коэффициента корреляции между двумя рядами наблюдений - экспериментальными значениями регрессанда (y_i) и его расчётными значениями ($\hat{y}_i$).

$$R^2 = \frac{\left[\sum (y_i - \bar{y})(\hat{y}_i - \bar{y})\right]^2}{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{y})^2}\,. \tag{3.18}$$

Здесь учтено, что для классической регрессионной модели $\bar{\hat{y}} = \bar{y}$.

Равноценные определения коэффициента детерминации

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} \tag{3.19}$$

или

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (y_i - \bar{y})^2}. \qquad (3.20)$$

3.2 Пример множественной регрессии

В качестве примера рассмотрим зависимость прибыли (Y в млн. грн.) от затрат на 1 грн. произведённой продукции (x_1 в коп.) и стоимости основных фондов (x_2 в млн. грн.) по данным десяти однотипных предприятий (табл. 3.1).

Выберем в качестве функции регрессии линейную

$$\breve{y} = a_0 x_0 + a_1 x_1 + a_2 x_2, \ x_0 \equiv 1.$$

Система нормальных уравнений (3.5) для определения МНК-оценок $\hat{a}_0$, $\hat{a}_1$, $\hat{a}_2$, соответствующая данным табл. 3.1, приобретает вид

$$\begin{cases} 10\hat{a}_0 + 847\hat{a}_1 + 2562\hat{a}_2 = 414{,}9, \\ 847\hat{a}_0 + 72113\hat{a}_1 + 216324\hat{a}_2 = 34834{,}9, \\ 2562\hat{a}_0 + 216324\hat{a}_1 + 670706\hat{a}_2 = 109051{,}8. \end{cases}$$

Решая полученную систему уравнений, находим $\hat{a}_0 = 42{,}5663$,

$\hat{a}_1 = -0{,}5201$, $\hat{a}_2 = 0{,}1677$.

Таким образом, оцененное уравнение регрессии имеет вид

$$\hat{y} = 42{,}5663 - 0{,}5201 x_1 + 0{,}1677 x_2. \qquad (3.21)$$

Т а б л и ц а 3.1 - *Расчётная таблица для определения параметров уравнения регрессии*

Y	x_1	x_2	x_1^2	x_1x_2	x_1Y	x_2^2	x_2Y	Y^2
26,1	96	215	9216	20640	2505,6	46225	5511,5	618,21
30,3	89	195	7921	17355	2696,7	38025	5908,5	918,09
38,9	82	215	6724	17630	3189,8	46225	8363,5	1513,21
39,5	81	245	6561	19845	3199,5	60025	9677,5	1560,25
40,2	80	246	6400	19680	3216,0	60516	9889,2	1616,04
42,4	88	262	7744	23056	3731,2	68644	11108,8	1797,76
45,7	85	280	7225	23800	3884,5	78400	12796,0	2088,49
48,2	92	314	8464	28888	4434,4	98596	15134,8	2323,24
50,1	77	295	5929	22715	3857,7	87025	14779,5	2510,01
53,5	77	295	5929	22715	4119,5	87025	15782,5	2862,25
414,9	847	2562	72113	216324	34834,9	670706	109051,8	17870,55
41,49	84,7	256,2	7211,3	21632,4	3483,49	67070,6	10905,18	1787,055

Рассмотрим теперь матричный способ. Имеем

$$\vec{X} = \begin{bmatrix} 1 & 96 & 215 \\ 1 & 89 & 195 \\ 1 & 82 & 215 \\ 1 & 81 & 245 \\ 1 & 80 & 246 \\ 1 & 88 & 262 \\ 1 & 85 & 280 \\ 1 & 92 & 314 \\ 1 & 77 & 295 \\ 1 & 77 & 295 \end{bmatrix}, \qquad \vec{X}'\vec{X} = \begin{bmatrix} 10 & 847 & 2562 \\ 847 & 72113 & 216324 \\ 2562 & 216324 & 670706 \end{bmatrix},$$

$$\vec{X}\,\vec{y} = \begin{bmatrix} 414,9 \\ 34834,9 \\ 109051,8 \end{bmatrix}, \quad (\vec{X}'\vec{X})^{-1} = \begin{bmatrix} 32,248160 & -0,284709 & -0,031356 \\ -0,284709 & 0,002941 & 0,000139 \\ -0,031356 & 0,000139 & 0,000076 \end{bmatrix},$$

$$\vec{a} = (\vec{X}'\vec{X})^{-1}\vec{X}\vec{y} = \begin{bmatrix} 4 \\ -0,5201 \\ 0,1677 \end{bmatrix}.$$

Видим, что полученное решение совпадает с предыдущим, что свидетельствует о правильности вычислений.

Сделаем анализ регрессионной модели (3.21).

1 Отрицательное значение коэффициента $\hat{a}_1$ свидетельствует о том, что с ростом затрат на 1 грн. произведенной продукции прибыль предприятия уменьшается, причем при прочих равных условиях увеличение (уменьшение) затрат на 1 коп. ведет к уменьшению (увеличению) прибыли в среднем на 520,1 тыс. грн. Коэффициент эластичности

$$\widehat{E}_1 = a_1 \cdot \frac{\overline{x_1}}{\overline{\overline{Y}}} = -0,5201 \cdot \frac{84,7}{41,49} = -1,06$$

показывает, что если затраты увеличить (уменьшить) на 1%, то прибыль в результате этого уменьшится (увеличится) в среднем на 1,06% (при прочих равных условиях).

2 Положительное значение коэффициента $\widehat{a}_2$ говорит о том, что с ростом производственных фондов прибыль предприятия также растет, причем при прочих равных условиях увеличение (уменьшение) стоимости основных фондов на 1млн. грн. влечет за собой увеличение (уменьшение) прибыли в среднем на 167,7 тыс. грн. Показатель эластичности

$$\widehat{E}_2 = \widehat{a}_2 \cdot \frac{\overline{x_2}}{\overline{\overline{Y}}} = 0,1677 \cdot \frac{256,2}{41,49} = 1,04$$

показывает, что если стоимость основных фондов увеличить (уменьшить) на 1%, то это приведет к росту (спаду) прибыли в среднем на 1,04% (при прочих равных условиях).

Оценим теперь качество модели (3.21). С этой целью вычислим коэффициенты детерминации. Имеем

$$R^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{34,629}{679} = 0,949.$$

Здесь оцененные значения прибыли предприятий и их отклонения от наблюдаемых значений

$$\vec{\hat{y}} = \vec{X}\vec{\hat{a}} = \begin{bmatrix} 28,703 \\ 28,989 \\ 35,984 \\ 41,536 \\ 42,223 \\ 40,747 \\ 45,326 \\ 47,388 \\ 52,002 \\ 52,002 \end{bmatrix}, \qquad \vec{\hat{u}} = \vec{y} - \vec{\hat{y}} = \begin{bmatrix} -2,603 \\ 1,311 \\ 2,916 \\ -2,036 \\ -2,023 \\ 1,653 \\ 0,374 \\ 0,812 \\ -1,902 \\ 1,498 \end{bmatrix}.$$

Значение $R^2 = 0,949$ свидетельствует о том, что линейная модель (3.21) объясняет 94,9% всей дисперсии Y, остальные 5,1% связаны со случайностью модели.

На диагонали оцененной ковариационной матрицы (3.16)

$$\hat{\vec{\Sigma}} = \hat{\sigma}_u^2 \left(\vec{X}'\vec{X}\right)^{-1} = \begin{bmatrix} 159,543842 & -1,408564 & -0,155128 \\ -1,408564 & 0,014549 & 0,000688 \\ -0,155128 & 0,000688 & 0,000378 \end{bmatrix},$$

где $\hat{\sigma}_u^2 = \frac{\sum \hat{u}_i^2}{7} = \frac{34,629}{7} = 4,947$, находятся оценки $\hat{\sigma}_{\hat{a}_o}^2$, $\hat{\sigma}_{\hat{a}_1}^2$, $\hat{\sigma}_{\hat{a}_2}^2$ дисперсий параметров $\hat{a}_0, \hat{a}_1, \hat{a}_2$. Извлекая из них квадратные корни, находим стандартные ошибки оценок $\hat{a}_0, \hat{a}_1, \hat{a}_2$: $\hat{\sigma}_{\hat{a}_o}^2 = 12,6311$,

$\hat{\sigma}^2_{\hat{a}_1} = 0{,}1206$, $\hat{\sigma}^2_{\hat{a}_2} = 0{,}0194$. Доверительный коридор для значе-ний теоретической регрессии на базисных данных определим согласно

$$y_i^T \in \left[\hat{y}_i - t_{1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{y}_i}, \hat{y}_i + t_{1-\frac{\alpha}{2}} \cdot \hat{\sigma}_{\hat{y}_i}\right],\ i = 1, 2, \ldots, n. \quad (3.22).$$

При уровне значимости $\alpha=0{,}05$ и числе степеней свободы n-m-1=7 приходим к табл.3.2.

Т а б л и ц а 3.2 - *Нижняя и верхняя 95% доверительные границы для коридора значений регрессии* $\hat{y}_i = 42{,}5663 - 0{,}5201 x_{i1} + 0{,}1677 x_{i2}$

Но-мер	$\hat{y}_i$	$\hat{\sigma}_{\hat{y}_i}$	**Нижняя граница** $\hat{y}_i - 2.37\hat{\sigma}_{\hat{y}_i}$	**Верхняя граница** $\hat{y}_i + 2.37\hat{\sigma}_{\hat{y}_i}$
1	28,703	1,534	25,067	32,339
2	28,989	1,348	25,794	32,184
3	35,984	1,181	33,184	38,783
4	41,536	0,893	39,418	43,653
5	42,223	0,960	39,949	44,498
6	40,747	0,832	38,775	42,718
7	45,326	0,848	43,315	47,337
8	47,388	1,764	43,207	51,570
9	52,002	1,231	49,085	54,919
10	52,002	1,231	49,085	54,919

3.3 Парная и частная корреляции. Случай двух регрессоров

Рассмотрим теперь линейную регрессионную модель с двумя объясняющими переменными

$$Y = a_0 + a_1 X_1 + a_2 X_2 + U \quad (3.23)$$

в предположении, что произведена выборка объема n и получены соответствующие данные (m=2). Тогда можно вычислить выборочные коэффициенты обычной (парной) корреляции между Y и X_1, Y и X_2, X_1 и X_2:

$$r_{y1} = r(Y, X_1) = \frac{m_{y1}}{s_y s_1} = \frac{\overline{x_1 Y} - \overline{x_1}\,\overline{Y}}{s_y s_1},$$

$$r_{y2} = r(Y, X_2) = \frac{m_{y2}}{s_y s_2} = \frac{\overline{x_2 Y} - \overline{x_2}\,\overline{Y}}{s_y s_2}, \qquad (3.24)$$

$$r_{12} = r(X_1, X_2) = \frac{m_{12}}{s_1 s_2} = \frac{\overline{x_1 x_2} - \overline{x_1}\;\overline{x_2}}{s_1 s_2}.$$

Возникает следующий вопрос: наблюдаемая корреляция между зависимой переменной (Y) и какой-либо независимой переменной (например, X_1) обусловлена чистой зависимостью между ними или другая независимая переменная (X_2) оказывает на них влияние, что и служит причиной наблюдаемой корреляции между первыми двумя переменными (Y и X_1). Таким образом, мы приходим к понятию частной корреляции между Y и X_1, когда влияние X_2 на каждую из них устранено.

Таким образом, формула (3.24) приобретает вид

$$r_{y1,2} = \frac{r_{y1} - r_{y2} r_{12}}{\sqrt{1 - r_{y2}^2}\sqrt{1 - r_{12}^2}}.$$

Аналогично ($r_{21} = r_{12}$): (3.25)

$$r_{y2,1} = \frac{r_{y2} - r_{y1} r_{21}}{\sqrt{1 - r_{y1}^2}\sqrt{1 - r_{21}^2}}, \qquad r_{12,y} = \frac{r_{12} - r_{y1} r_{y2}}{\sqrt{1 - r_{y1}^2}\sqrt{1 - r_{y2}^2}}.$$

Из (3.25) вытекает, что выборочные частные коэф-фициенты корреляции лежат на отрезке [-1, 1], как и обычные коэффициенты корреляции. Значения выборочных парных (3.24) и частных (3.25) коэффициентов корреляции отличаются друг от друга: парный коэффициент характеризует связь между двумя признаками без учета влияния других признаков, частный – учитывает наличие и влияние других признаков.

Вычислим выборочные парные и частные коэффициенты корреляции по данным табл. 3.1, характеризующей зависимость прибыли (Y в млн. грн.) от затрат на 1 грн. произведённой продукции (x_1 в коп.) и стоимости основных фондов (x_2 в млн. грн.). Имеем:

$$s_1 = \sqrt{\overline{x_1^2} - \bar{x}_1^{\,2}} = \sqrt{7211{,}3 - 84{,}7^2} = 6{,}10\,,$$

$$s_2 = \sqrt{\overline{x_2^2} - \bar{x}_2^{\,2}} = \sqrt{67070{,}6 - 256{,}2^2} = 37{,}8439\,,$$

$$s_y = \sqrt{\overline{Y^2} - \overline{Y}^2} = \sqrt{1787{,}055 - 41{,}49^2} = 8{,}1015 \ ,$$

$$r_{y1} = \frac{\overline{x_1 Y} - \bar{x}_1 \overline{Y}}{s_y s_1} = \frac{3483.49 - 84.7 \cdot 41.49}{8.1015 \cdot 6.10} = -0.6218 \,,$$

$$r_{y2} = \frac{\overline{x_2 Y} - \bar{x}_2 \overline{Y}}{s_y s_2} = \frac{10905{,}18 - 256{,}2 \cdot 41{,}49}{8{,}1015 \cdot 37{,}8439} = 0{,}8984 \,,$$

$$r_{12} = \frac{\overline{x_1 x_2} - \bar{x}_1 \bar{x}_2}{s_1 s_2} = \frac{21632{,}4 - 84{,}7 \cdot 256{,}2}{6{,}10 \cdot 37{,}8439} = -0{,}2934 \,,$$

$$r_{y1,2} = \frac{r_{y1} - r_{y2} r_{12}}{\sqrt{1 - r_{y2}^2}\sqrt{1 - r_{12}^2}} = \frac{-0{,}6218 + 0{,}8984 \cdot 0{,}2934}{\sqrt{1 - 0{,}8984^2}\sqrt{1 - 0{,}2934^2}} = -0{,}8532 \,,$$

$$r_{y2,1} = \frac{r_{y2} - r_{y1} r_{21}}{\sqrt{1 - r_{y1}^2}\sqrt{1 - r_{21}^2}} = \frac{0{,}8984 - 0{,}6218 \cdot 0{,}2934}{\sqrt{1 - 0{,}6218^2}\sqrt{1 - 0{,}2934^2}} = 0{,}9561 \,,$$

$$r_{12,y} = \frac{r_{12} - r_{y1} r_{y2}}{\sqrt{1 - r_{y1}^2}\sqrt{1 - r_{y2}^2}} = \frac{-0{,}2934 + 0{,}6218 \cdot 0{,}8984}{\sqrt{1 - 0{,}6218^2}\sqrt{1 - 0{,}8984^2}} = 0{,}7711 \,.$$

Сделаем анализ полученных результатов.

1 Отрицательное значение r_{y1} указывает на то, что с ростом затрат прибыль Y будет уменьшаться. Однако небольшое абсолютное значение $|r_{y1}| = 0{,}6218$ означает слабое влияние затрат на прибыль. Устранение влияния признака x_2 существенным образом повышает корреляцию между Y и x_1 ($r_{y1,2} = -0{,}8532$).

2 Значение выборочного парного коэффициента корреляции между Y и x_2 ($r_{y2} = 0.8984$) высоко и указывает на тесную связь прибыли и основных фондов. Устранение влияния признака x_1 ещё более усиливает эту связь ($r_{y2,1} = 0{,}9561$).

3 Низкое значение $r_{12} = -0{,}2934$ означает, что признаки x_1 и x_2 практически не коррелируют. Этот факт очень важен в линейных регрессионных моделях. Однако устранение влияния прибыли Y значительно усиливает корреляцию между затратами и основными фондами ($r_{12,y} = 0{,}7711$).

Таким образом, применение обычных выборочных коэффициентов корреляции при установлении связи двух случайных величин может привести к неправильным выводам. Если при устранении влияния других факторов корреляция двух случайных величин возрастает, то это означает, что эти факторы скрывали истинную взаимозависимость этих двух случайных величин. Если же корреляция между двумя случайными величинами уменьшается или становится близкой к нулю при других фиксированных случайных величинах, то можно сказать, что взаимозависимость этих двух случайных величин в значительной мере имеет место благодаря другим факторам. В качестве проверки ранее полученных результатов вычислим значения оценок $\widehat{a}_1$, $\widehat{a}_2$ и коэффициента детерминации R^2 по формулам:

$$\widehat{a}_1 = \frac{s_y}{s_1}\frac{r_{y1} - r_{y2}r_{12}}{1 - r_{12}^2} = \frac{8{,}1015}{6{,}10}\cdot\frac{-0{,}6218 + 0{,}8984\cdot 0{,}2934}{1 - 0{,}2934^2} = -0{,}5201,$$

$$\widehat{a}_2 = \frac{s_y}{s_2}\frac{r_{y2} - r_{y1}r_{12}}{1 - r_{12}^2} = \frac{8{,}1015}{37{,}8439}\cdot\frac{0{,}8984 + 0{,}6218\cdot 0{,}2934}{1 - 0{,}2934^2} = 0{,}1677,$$

$$R^2 = \frac{0{,}6218^2 + 0{,}8984^2 - 2\cdot 0{,}6218\cdot 0{,}8984\cdot 0{,}2934}{1 - 0{,}2934^2} = 0{,}949.$$

Как и следовало ожидать, полученные значения совпадают с найденными в предыдущем параграфе.

5. Методом ортогональных многочленов построить полиномиальные регрессии 1-го, 2-го, …, m-го порядков зависимости средней урожайности зерновых от номера года.

6. Для каждой регрессии вычислить коэффициент детерминации и при уровне значимости α =0,05 проверить значимость нового коэффициента с помощью

а) критерия Стьюдента; б) критерия Фишера.

7. Определить оптимальную степень полиномиальной регрессии, считая, что улучшение результата не происходит, если два последних коэффициента оказались незначимыми. Оцените качество оптимальной регрессии.

8. Сделать прогноз урожайности зерновых на текущий и последующий годы.

9. Опеделите качество модели, вычислите коэффициент детерминации. Дайте интерпретацию его значению.

10. Найдите стандартные ошибки коэффициентов регрессии и проверьте их значимость при уровне значимости α=0.05.

11. Определите соответствующую *F*- статистику и проверьте значимость уравнения регрессии вцелом.

12. Определите нижнюю и верхнюю 95% доверительные границы для коридора значений регрессии на базисных данных

3.4 Практическое задание. Множественная регрессия

В таблице 3.3 приведены данные о прибыли (Y в млн. грн.), затратах на 1 грн. произведенной продукции (X_1 в коп.) и стоимости основных фондов (X_2 в млн. грн.) по данным 10 однотипных предприятий. Необходимо:

1 Вычислить выборочные парные коэффициенты корреляции r_{Y1}, r_{Y2}, r_{12} и выборочные частные коэффициенты корреляции $r_{Y1,*}$ и $r_{Y2,*}$, а также установить их значимость при уровне значимости $\alpha = 0{,}05$. Сделайте выводы о силе связи показателя Y и факторов X_1 и X_2.

2 Методом наименьших квадратов оценить коэффициенты линейной регрессии $Y = a_0 + a_1 X_1 + a_2 X_2 + U$.

Сделайте анализ оцененной регрессии, вычислив:

a) изменение прибыли при увеличении величины каждого из факторов на единицу;

b) средние коэффициенты эластичности для каждого фактора. Дайте экономическую интерпретацию полученных результатов.

3 Оценить качество полученной модели по рассчитанному коэффициенту детерминации R^2. Что означает вычисленное значение R^2?

Т а б л и ц а 3.3

Номер	1	2	3	4	5	6	7	8	9	10
1	25,0	26,4	27,3	29,2	30,7	31,8	33,4	35,6	30,8	32,4
	92	88	85	87	82	85	89	91	93	90
	202	215	220	230	210	217	220	217	215	207
2	24,6	28,7	29,3	27,6	32,4	35,3	37,6	38,4	29,1	28,5
	94	90	91	92	91	87	86	90	91	87
	207	213	215	208	217	224	250	244	210	231
3	26,2	30,0	38,7	39,3	40,4	45,2	48,0	51,2	52,3	55,4
	95	90	84	84	82	86	91	90	91	90
	210	198	212	214	238	270	300	310	320	340
4	27,2	30,7	32,3	35,7	38,3	42,3	44,7	52,1	58,4	60,0
	92	91	92	91	93	92	93	92	89	90
	202	220	225	231	235	240	241	258	273	277
5	28,4	29,1	30,2	32,8	34,3	36,4	40,3	45,7	48,4	50,0
	90	91	90	91	92	91	88	86	88	89
	213	217	223	230	242	253	270	293	299	315
6	21,3	23,7	28,7	31,7	38,4	40,3	47,4	51,3	58,3	60,4
	91	92	88	82	87	86	82	84	87	83
	203	208	230	247	255	263	277	290	295	320
7	22,7	24,4	28,3	32,5	37,4	40,3	44,7	48,4	50,3	55,0
	95	92	91	88	89	87	82	85	86	84
	200	212	230	250	264	280	297	310	323	340
8	28,3	29,4	30,7	35,8	37,2	38,4	41,2	44,3	46,7	48,4
	90	88	87	86	89	90	84	86	82	84
	210	221	232	250	261	270	288	300	330	350
9	21,3	27,4	28,3	29,4	30,4	32,1	35,4	38,4	38,5	42,3
	92	91	88	83	85	87	82	81	80	78
	200	210	231	252	261	270	291	300	300	322
10	28,4	29,3	31,2	33,4	35,2	39,7	43,6	45,7	47,3	50,0
	90	91	88	92	87	86	82	80	78	79
	210	212	223	228	235	240	253	260	271	280
11	31,2	33,7	36,3	39,2	41,3	43,7	48,4	50,4	52,4	55,4
	92	88	86	83	81	80	84	87	82	80
	208	220	228	237	244	250	263	280	291	300

Продолжение таблицы 3.3

Номер	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
12	30,4	32,4	33,7	35,6	38,4	40,2	42,4	44,7	48,4	50,2
	91	92	90	88	86	83	85	83	90	87
	211	220	228	237	244	250	263	280	291	300
13	28,1	28,7	29,2	29,8	30,4	31,2	31,8	32,4	33,2	34,0
	30,2	31,2	31,5	31,9	32,1	32,8	32,5	33,0	34,6	35,0
	215	220	222	225	228	235	240	245	250	255
14	30,2	30,8	31,4	31,8	32,0	32,5	33,0	33,5	33,8	34,5
	31,6	32,0	32,5	32,8	33,2	33,6	34,2	34,5	34,8	35,0
	218	225	230	233	235	240	242	248	250	255
15	31,2	31,5	31,8	32,0	32,4	32,8	33,2	33,6	34,2	35,0
	31,8	32,0	32,4	32,8	33,6	33,8	34,0	34,2	34,5	35,0
	220	225	228	232	240	244	248	252	255	258
16	31,4	31,7	31,8	32,4	32,6	33,0	33,2	33,6	33,8	34,0
	32,0	32,3	32,5	32,6	32,8	33,0	33,4	33,8	34,0	34,2
	225	228	230	232	235	238	240	244	248	250
17	31,8	31,9	32,4	32,8	33,0	33,5	34,0	34,2	34,6	35,0
	32,2	32,4	32,5	32,6	32,8	33,0	33,4	33,5	33,8	34,0
	230	232	235	238	240	244	248	250	254	260
18	32,0	32,4	32,6	33,0	33,4	33,8	33,9	34,0	34,2	34,5
	33,0	33,5	33,8	34,0	34,2	34,4	34,5	34,6	34,8	35,0
	235	238	240	242	245	248	249	250	252	255
19	33,0	33,4	33,8	34,0	34,5	34,8	35,0	35,8	36,0	37,0
	34,5	34,8	35,0	35,4	35,5	35,6	35,8	36,0	36,2	36,5
	240	244	246	248	250	252	254	260	264	270
20	35,4	36,2	36,8	37,0	37,5	38,5	38,8	39,4	39,8	40,0
	35,0	35,2	35,5	35,8	36,0	36,2	36,4	36,6	36,8	37,0
	250	252	254	255	258	260	262	264	265	266
21	35,8	36,0	37,2	37,5	38,0	38,5	39,0	39,5	39,7	40,0
	36,1	36,2	36,4	36,5	36,8	36,9	37,0	37,2	37,4	37,5
	252	253	260	262	264	266	268	270	272	275
22	36,0	36,5	37,0	37,8	38,0	38,4	39,2	39,5	39,7	40,0
	36,0	36,2	36,4	36,5	36,6	36,8	37,0	37,1	37,5	37,6
	255	258	260	264	265	268	272	274	275	277

Продолжение таблицы 3.3

Номер	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
23	37,0	37,8	38,0	38,8	40,0	40,5	41,0	41,5	42,0	43,0
	36,0	36,2	36,4	36,6	36,8	36,9	37,0	37,2	37,5	37,6
	256	260	261	264	268	270	272	275	276	280
24	36,5	36,8	37,8	38,0	38,5	39,0	39,6	40,0	40,5	41,0
	36,5	36,6	36,7	36,8	37,0	37,2	37,5	37,6	37,8	38,0
	250	252	255	257	260	263	265	267	270	275
25	34,2	36,1	39,4	44,7	50,4	54,7	58,3	60,0	65,6	70,0
	92	90	90	88	86	83	81	78	77	75
	134	140	148	156	164	172	181	192	207	220
26	37,1	40,2	45,4	50,0	54,8	60,1	64,7	68,6	72,4	75,0
	90	88	87	86	85	82	81	78	75	74
	150	162	175	186	195	210	221	232	245	250
27	20,4	24,1	28,6	33,2	35,4	38,4	44,2	48,3	50,7	55,5
	93	91	90	88	85	84	80	79	77	75
	141	150	160	169	175	183	194	204	210	220
28	28,6	30,4	35,6	38,7	43,5	48,6	50,0	55,3	58,6	60,0
	92	90	89	88	88	86	84	82	80	75
	150	161	175	183	200	215	222	235	242	245
29	19,6	24,1	28,6	30,0	35,4	39,7	45,4	52,3	58,6	60,5
	95	92	92	90	88	86	84	83	80	78
	130	141	150	155	164	170	183	196	210	215
30	24,0	29,4	34,8	40,0	45,6	49,3	53,4	55,1	59,3	65,0
	93	91	90	88	89	85	82	81	79	75
	180	195	206	215	222	237	249	256	270	295
31	13,6	14,5	18,7	22,4	31,2	35,0	39,4	45,6	49,4	55,2
	94	95	92	91	88	88	90	85	84	82
	140	144	156	164	190	202	210	233	247	266
32	21,2	25,4	29,3	35,6	40,4	46,3	55,2	59,1	66,2	70,0
	92	91	90	88	86	85	84	83	80	78
	140	156	168	182	198	214	233	251	269	290
33	30,0	35,6	38,4	42,6	46,4	49,3	55,4	60,0	65,2	70,5
	93	92	90	88	86	85	83	82	80	78
	150	166	175	188	198	213	228	240	251	265

Продолжение таблицы 3.3

Номер	1	2	3	4	5	6	7	8	9	10
34	32,4	63,1	41,2	47,3	52,8	55,7	59,1	65,9	70,7	75,0
	92	91	90	88	87	84	85	82	80	77
	148	160	168	182	200	208	214	221	230	241
35	44,7	46,1	49,7	58,3	62,4	68,3	72,4	75,0	79,6	85,0
	90	91	90	88	87	85	82	80	78	75
	162	165	173	188	196	207	213	218	227	239
36	27,2	28,3	30,1	32,4	35,6	36,2	36,8	37,4	39,6	40,2
	31,5	32,4	33,5	34,1	35,3	38,3	38,4	39,1	40,2	41,4
	210	215	218	220	228	232	235	238	241	245
37	28,4	29,1	30,3	31,5	32,3	33,4	34,2	37,1	38,3	39,6
	32,3	33,6	34,6	36,2	37,8	38,1	38,3	38,5	39,0	39,6
	218	220	227	232	238	243	245	247	252	260
38	27,3	28,6	29,1	29,6	32,1	34,2	25,3	38,2	38,8	39,2
	31,2	32,3	33,6	34,2	35,6	36,8	37,2	38,5	39,2	40,2
	240	246	252	258	262	265	268	272	277	280
39	27	29	30	32	34	35	37	39	42	45
	32	33	35	37	39	40	42	43	45	50
	215	218	220	228	230	235	238	242	243	245
40	28	29	32	34	36	39	41	44	48	50
	36	38	41	44	45	48	51	55	58	60
	210	215	218	224	225	228	231	236	238	240
41	29	32	35	37	39	42	46	47	49	52
	37	38	40	42	45	48	51	55	58	60
	218	220	228	230	235	240	242	243	245	255
42	32	35	38	39	44	48	52	56	58	60
	38	41	42	45	48	49	52	55	58	60
	210	215	220	225	230	238	240	245	248	250
43	35	37	38	40	42	44	48	50	52	55
	40	42	43	44	48	50	54	58	60	62
	220	222	225	228	230	235	240	245	248	250
44	40	42	44	45	47	49	49	52	55	57
	41	42	43	45	46	48	49	51	53	55
	225	228	230	235	238	240	241	247	250	255

Продолжение таблицы 3.3

Номер	1	2	3	4	5	6	7	8	9	10
45	42	44	46	48	50	53	56	58	59	60
	43	45	46	47	50	51	53	55	58	60
	228	230	232	235	238	240	245	248	249	250
46	44	46	48	50	52	55	58	60	62	64
	45	46	48	49	50	52	54	55	58	60
	230	232	235	238	240	244	250	254	255	258
47	48	49	50	52	55	58	60	62	64	66
	46	47	49	50	52	54	55	56	58	60
	232	235	238	240	244	248	250	252	255	260
48	15,4	18,6	23,7	28,2	34,1	42,4	48,6	54,5	60,2	65,6
	84	81	77	74	79	76	73	78	80	82
	172	184	195	208	224	241	256	274	281	296
49	12,6	15,4	18,3	21,4	28,6	34,5	37,3	39,0	45,4	51,6
	82	86	88	90	77	79	91	87	80	82
	140	151	160	172	175	188	212	221	234	250
50	10,8	14,6	18,4	25,6	28,2	35,5	39,4	45,4	52,7	59,8
	95	92	88	86	84	82	81	80	77	74
	110	121	128	137	145	156	165	177	188	205

ПРИЛОЖЕНИЕ А
(справочное)

Двусторонние квантили
распределения Стьюдента $t_{1-\frac{\alpha}{2}}$:

$$P(t \le t_{1-\frac{\alpha}{2}}) = 1 - \tfrac{\alpha}{2},$$

$$P(|t| > t_{1-\frac{\alpha}{2}}) = \alpha .$$

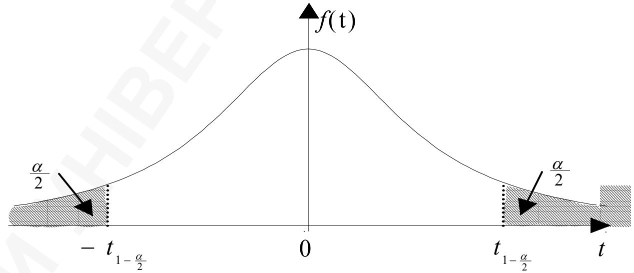


Таблица А.1

k \ α	***0.5***	***0.2***	***0.1***	***0.05***	***0.02***	***0.01***	***0.002***	***0.001***
1	1.00	3.078	6.314	12.706	31.821	63.657	318.3	636.619
2	0.816	1.886	2.920	4.303	6.965	9.925	22.33	31.598
3	0.765	1.638	2.353	3.182	4.541	5.841	10.210	12.941
4	0.741	1.533	2.132	2.776	3.747	4.604	1.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	3.93	4.318

Продолжение таблицы А.1

k \ α	***0.5***	***0.2***	***0.1***	***0.05***	***0.02***	***0.01***	***0.002***	***0.001***
13	0.694	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.689	1.333	1.74	2.110	2.567	2.898	3.646	3.965
18	0.688	1.33	1.734	2.101	2.552	2.878	3.610	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	0.686	1.321	1.717	2.074	2.506	2.819	3.505	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	0.685	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	0.683	1.313	1.701	2.048	2.467	2.763	3.408	3.674
23	0.683	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	0.683	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	0.681	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	0.679	1.299	1.676	2.009	2.403	2.678	3.261	3.496
100	0.677	1.290	1.660	1.984	2.364	2.626	3.174	3.390
∞	0.675	1.282	1.645	1.960	2.326	2.576	3.090	3.291

ПРИЛОЖЕНИЕ Б
(справочное)

Квантили распределения Фишера $\mathbf{F_{1-\alpha}(k_1,k_2)}$**:**

$$P(F \le F_{1-\alpha}) = 1-\alpha\,,$$

$$P(F > F_{1-\alpha}) = \alpha\,.$$

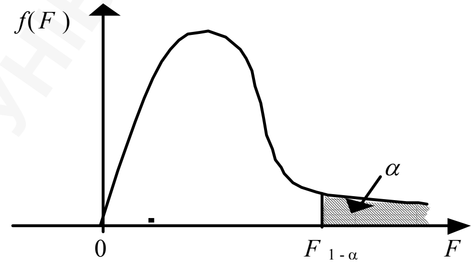


Таблица Б.1 - 95%-квантили (α=0,05) распределения Фишера *F(k₁,k₂)*

k_1 / k_2	1	2	3	4	5	6	7	8	9	10	11	12	13
1	162	200	216	225	230	234	237	239	241	242	243	244	245
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.20	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.86	5.94	5.91	5.89
5	6.61	579	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26
9	5.12	4.26	3.68	3.63	3.46	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05

Продолжение таблица Б.2

k_1 / k_2	1	2	3	4	5	6	7	8	9	10	11	12	13
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.60	2.75	2.72	2.69	2.66
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20
23	4.48	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.18
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.16	2.15
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15	2.12
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.17	2.13	2.10
25	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12	1.09
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.08
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06

ПРИЛОЖЕНИЕ В
(справочное)

Таблица В.1 **-** Квантили распределения Фишера-Иейтса выборочного коэффициента корреляции $r_{1-\frac{\alpha}{2}}$

k_1 / k_2	*0,10*	*0,05*	*0,02*	*0,01*	*0,001*
1	0,988	0,997	0,999	1,000	1,000
2	0,900	0,950	0,980	0,990	0,999
3	0,805	0,878	0,934	0,959	0,992
4	0,729	0,811	0,882	0,917	0,974
5	0,669	0,754	0,833	0,874	0,951
6	0,621	0,707	0,789	0,834	0,925
7	0,582	0,666	0,750	0,798	0,898
8	0,549	0,632	0,716	0,765	0,872
9	0,521	0,602	0,685	0,735	0,847
10	0,497	0,576	0,658	0,708	0,823
11	0,476	0,553	0,634	0,684	0,801
12	0,457	0,532	0,612	0,661	0,780
13	0,441	0,514	0,592	0,641	0,760
14	0,426	0,497	0,574	0,623	0,742

Продолжение таблицы В.1

k_1 / k_2	**0,10**	**0,05**	**0,02**	**0,01**	**0,001**
15	0,412	0,482	0,558	0,606	0,725
16	0,400	0,468	0,543	0,590	0,708
17	0,389	0,456	0,528	0,575	0,693
18	0,378	0,444	0,516	0,561	0,679
19	0,369	0,433	0,503	0,549	0,665
20	0,360	0,423	0,492	0,537	0,652
25	0,323	0,381	0,445	0,487	0,597
30	0,296	0,349	0,409	0,449	0,554
35	0,275	0,325	0,381	0,418	0,519
40	0,257	0,304	0,358	0,393	0,490
45	0,243	0,287	0,338	0,372	0,465
50	0,231	0,273	0,322	0,354	0,443
60	0,211	0,250	0,295	0,325	0,408
70	0,195	0,232	0,274	0,302	0,380
80	0,183	0,217	0,256	0,283	0,357
90	0,173	0,205	0,242	0,267	0,337
100	0,164	0,195	0,230	0,254	0,321

Список литературы

1. Доугерти К. Введение в эконометрику. – М.: ИНФРА-М, 1997. – 402 с.
2. Задорожный Г. В., Иващенко П. А. Эконометрика. - Харьков: Харьковский институт бизнеса и менеджмента, 1996. – Ч. 1. – 104 с.
3. Задорожный Г.В., Иващенко П.А. Эконометрика. - Харьков: Харьковский институт бизнеса и менеджмента, 1996. – Ч. 2. – 99 с.
4. Иванова В. М. Основы эконометрики: Учебное пособие. - М.: Моск. эконом.-стат. ин-т, 1995. – 145 с.
5. Коршунова Н. И., Плясунов В. С. Математика в экономике: Учебное пособие. – М.: Вита-Пресс, 1996. – 368 с.
6. Лившиц А.Я. Введение в рыночную экономику: Курс лекций. – М.: МП ТПО «Квадрат», 1991.- 255 с.
7. Лук`яненко І., Краснікова Л. Економетрика: Підручник. - К.:Товариство “Знання”, КОО, 1998.- 494 с.
8. Назаренко А. М. Эконометрика: Учебное пособие. – Сумы: Изд-во СумГУ, 2000. – 404 с.
9. Наконечний С.І., Терещенко Т.О., Романюк Т.П. Економетрія: Підручник.- К.:КНЕУ, 2000.- 296 с.
10. Толбатов Ю. А. Економетрика: Навчальний посібник. – К.: Четверта хвиля, 1997. – 320 с.