ФГБОУ ВПО
«Брянский государственный аграрный университет»

Кафедра коммерции и экономического анализа

А.В. Раевская

ЭКОНОМЕТРИКА

Учебно-методическое пособие для бакалавров, обучающихся по направлению подготовки 09.03.03 Прикладная информатика

Брянская область
2016

УДК 330.4(075)
ББК 65в631я73
Э40

Раевская А.В. **Эконометрика:** Учебно-методическое пособие для бакалавров, обучающихся по направлению подготовки 09.03.03 Прикладная информатика. – Брянск: Изд-во БГАУ, 2016. – 147 с.

Рецензент: к.э.н., доцент кафедры коммерции и экономического анализа **Дьяченко О.В.**

Учебно-методическое пособие содержит курс лекций по основным разделам дисциплины, а также включает в себя тестовые задания, контрольные вопросы, задания для лабораторно-практических занятий и самостоятельной работы студентов.

Целью данного пособия является обзор основных эконометрических методов и понятий и демонстрация их использования для решения реальных экономических задач.

Пособие предназначено для студентов дневной формы обучения, а также для студентов заочной и дистанционной форм обучения в качестве пособия для самостоятельного изучения дисциплины.

Рекомендовано к изданию учебно-методическим советом экономического факультета (протокол № 9 от 30 июня 2016 г.).

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

В современной экономической науке статистические методы исследования, моделирования и проектирования играют все большую роль. Это обусловлено совершенствованием вычислительной техники, благодаря которой существенно расширяется возможность успешного применения статистики и математического аппарата при решении экономических задач. В результате выделилось и сформировалось одно из направлений экономических исследований – эконометрика.

Эконометрика означает измерения в экономике, но область исследований этой дисциплины гораздо шире. Эконометрика – это наука, в которой на базе реальных статистических данных строятся, анализируются и отлаживаются математические модели реальных экономических процессов. Одним из важнейших направлений эконометрики является построение прогнозов по различным экономическим показателям.

Целями освоения дисциплины являются:

- формирование у студентов знаний и умений, связанных с эконометрическим моделированием;
- выявление закономерностей функционирования экономических систем разного уровня;
- изучение методов оценки и прогнозирования экономических показателей, характеризующих состояние и развитие анализируемых экономических систем;
- освоение современных компьютерных технологии эконометрического анализа и возможности их применения для решения прикладных экономических задач, а также формирование социально-личностных качеств студентов:
- организованности, гражданственности, повышение общей культуры, готовности к деятельности в профессиональной среде.

В результате освоения дисциплины у обучающихся должны сформироваться способности:

- использовать основы экономических знаний в различных сферах деятельности (ОК-3);
- проектировать ИС в соответствии с профилем подготовки по видам обеспечения (ПК-3);
- применять системный подход и математические методы в формализации решения прикладных задач (ПК-23).

Знания и навыки, получаемые студентами в результате изучения дисциплины, необходимы для успешного применения эконометрических моделей для описания, анализа и прогнозирования реальных экономических процессов.

Учебный материал в пособии условно разбит на четыре части и приложения.

В первой части рассмотрены модели парной регрессии (линейная и нелинейные модели).

Во второй части достаточно подробно разбирается модель множественной линейной регрессии и кратко обсуждается проблемы гомоскедастичности и автокоррелированности остатков.

Третья часть посвящена системам одновременных эконометрических уравнений.

В четвертой части рассматриваются модели временных рядов.

В приложениях содержатся варианты самостоятельных работ по всем темам курса и статистико-математические таблицы распределений Фишера, Стьюдента и Дарбина-Уотсона.

1. ВВЕДЕНИЕ В ЭКОНОМЕТРИКУ

1.1. Определение эконометрики

Эконометрика как наука возникла в первой половине 20-го века в результате активного использования для решения задач экономической теории математических и статистических методов.

Термин эконометрика введен в научную литературу в 1930 году норвежским статистиком Рагнаром Фришем. Он первым определил эконометрику, как научную дисциплину, базирующуюся на синтезе экономической теории, статистики и математики.

В дословном переводе слово эконометрика означает «экономические измерения». Это очень широкое толкование данного понятия. Как правило, термин эконометрика применяется в более узком смысле. А именно, под **эконометрикой понимается раздел науки, изучающий конкретные количественные и качественные взаимосвязи экономических объектов и процессов с помощью математических и статистических методов и моделей.**

Можно сказать, что главной задачей эконометрики является количественная оценка имеющихся взаимосвязей между экономическими явлениями и процессами.

Экономические явления взаимосвязаны и взаимообусловлены. Следствием этого является то, что значения соответствующих экономических показателей изменяются во времени с учетом этих взаимосвязей. Так, например, известно, что совокупный спрос зависит от уровня цен, потребление – от располагаемого дохода, инвестиции – от процентной ставки и так далее. Перед исследователем стоит задача выявления таких связей, количественная их оценка и изучение возможности использования выявленных связей в экономическом анализе и прогнозировании. **Разработкой соответствующего инструментария и его применением для решения конкретных практических экономических задач как раз и занимается эконометрика.**

В основе любого эконометрического исследования лежит построение экономико-математической модели, адекватной изучаемым реальным экономическим явлениям и процессам.

Процесс построения эконометрических моделей начинается с качественного исследования проблемы методами экономической теории, формулируются цели исследования, выделяются факторы, влияющие на изучаемый показатель, и формулируются предположения о характере предполагаемой зависимости.

На этой основе изучаемые зависимости выражаются в виде математических формул и соотношений.

Следует отметить, что ввиду невозможности одновременно учесть большое количество факторов, влияющих на изучаемый показатель, предполагаемые зависимости между переменными будут выполняться не точно, а с определенной погрешностью. Кроме того, экономическим явлениям присуща внутренняя неопределенность, связанная с целенаправленной деятельностью субъектов экономики.

Вышесказанное обуславливает применение статистических методов, с помощью которых осуществляется отбор значимых факторов, определяется наличие и степень тесноты связи между изучаемыми показателями, дается количественная оценка параметров предполагаемых зависимостей и исследуется степень их соответствия реальной действительности.

Основным инструментом математической статистики, используемым для построения эконометрических моделей, являются методы корреляционного и регрессионного анализа.

Корреляционный анализ ставит своей целью проверку наличия и значимости линейной зависимости между переменными без разделения переменных на зависимые и объясняющие. Ответ на эти вопросы дается с помощью вычисления показателей (коэффициентов) корреляции.

Регрессионный анализ направлен на выражение изучаемой зависимости в виде аналитической формулы с предварительным выделением зависимых и объясняющих переменных.

Регрессионный анализ призван ответить на такие вопросы, как:

– какие переменные определяют поведение других величин и, следовательно, могут использоваться как объясняющие переменные?

– какова формула зависимости и каков экономический смысл ее коэффициентов?

Результатом проведения регрессионного анализа является построение, так называемого, уравнения регрессии.

1.2. Характеристика взаимосвязей

Основная задача эконометрики заключается в исследовании и количественной оценке объективно существующих взаимосвязей и зависимостей между экономическими явлениями. Наибольший интерес для исследователя представляют причинно-следственные отношения между явлениями, что позволяет выявлять факторы, оказывающие основное влияние на вариацию изучаемых явлений и процессов.

Причинно-следственное отношение – это такая связь между явлениями, при которой изменение одного из них, называемого причиной, ведет к изменению другого, называемого следствием. Следовательно, причина всегда предшествует следствию.

Причинно-следственные связи в социально-экономических явлениях обладают следующими особенностями. Во-первых, причина X и следствие Y взаимодействуют не непосредственно, а через промежуточные факторы, которые, как правило, при анализе опускаются. Формально это может быть выражено с помощью схемы X—>X'—>X''—>Y, где X' и X'' изображают такие промежуточные факторы.

Во-вторых, социально-экономические явления развиваются и формируются в результате одновременного воздействия большого числа факторов. Поэтому одной из главных проблем при изучении этих явлений становится задача выявления главных, существенных причин и абстрагирование от второстепенных.

Признаки по их роли в изучаемой взаимосвязи делятся на два класса: *факторные* и *результативные*.

Факторными признаками (факторами) называются признаки, обусловливающие изменения других, связанных с ними признаков. Факторные признаки называются также независимыми, объясняющими или входными переменными.

Результативными называются признаки, изменяющиеся под действием факторных признаков. Результативные признаки называются также зависимыми, объясняемыми или выходными переменными.

По направлению изменения связи подразделяются на *прямые* (когда изменение результативного и факторного признаков происходит в одном направлении) и *обратные* (когда изменение результативного и факторного признаков происходит в противоположных направлениях).

По характеру проявления различают *функциональную* связь и *стохастическую* зависимость. Функциональной называют такую связь, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака. Функциональная связь проявляется во всех случаях наблюдения и для каждой конкретной единицы исследуемой совокупности. Такие связи изучаются в основном в естественных науках.

В эконометрике в основном изучаются причинные зависимости, которые проявляется не в каждом отдельном случае, а в общем, среднем при большом числе наблюдений. То есть одним и тем же значениям факторных признаков, как правило, соответствуют различные значения результативного признака.

Но, тем не менее, рассматривая всю совокупность наблюдений можно отметить наличие определенной зависимости между значениями признаков. Такие причинные зависимости называются *стохасти-*

ческими. Частным случаем стохастической связи является *корреляционная* связь, при которой изменение среднего значения результативного признака обусловлено изменением факторных признаков.

По аналитическому выражению выделяют связи *линейные* и *нелинейные*. Линейной называется связь, в которой изменение результативного признака прямо пропорционально изменению факторных признаков. В противном случае связь называется нелинейной. Аналитически линейная стохастистическая связь между явлениями может быть представлена уравнением прямой линии на плоскости, либо уравнением гиперплоскости в n-мерном пространстве (при наличии n факторных переменных).

1.3. Классификация видов эконометрических переменных и типов данных. Проблемы, связанные с данными

В эконометрических моделях в основном используются данные трёх типов:

1) пространственные данные (cross-sectional data);
2) временные ряды (time-series data);
3) панельные данные (panel data).

Пространственными данными называется совокупность экономической информации, которая характеризует различные объекты, однако полученной за один и тот же период или момент времени.

Пространственные данные являются выборочной совокупностью из некоторой генеральной совокупности. Примером пространственных данных может служить комплекс экономической информации по какому-либо предприятию (численность работников, объём производства, размер основных фондов), объёмах потребления продукции определённого вида, данные о ВВП различных стран в каком-либо конкретном году и т.д.

Временными данными называется совокупность экономической информации, которая характеризует один и тот же объект, но за разные периоды времени.

Отдельно взятый временной ряд можно рассматривать как выборку из бесконечного ряда значений показателей во времени. Примером временных данных могут служить данные о динамике индекса потребительских цен, ежедневные обменные курсы валют.

Отличия временных данных от пространственных данных:

1) единицы временных рядов подвержены явлению автокорреляции (зависимости между прошлыми и текущими наблюдениями временного ряда), т. е. они не являются статистически независимыми в отличие

от единиц случайной пространственной выборки;

2) единицы временных рядов не являются одинаково распределёнными величинами;

3) в отличие от пространственных данных временные данные естественным образом упорядочены во времени.

Панельными данными называются данные, содержащие сведения об одном и том же множестве объектов за ряд последовательных периодов времени.

Панельные данные являются обобщением или комбинацией пространственных и временных данных. Примером панельных данных могут служить показатели хозяйственной деятельности совокупности предприятий, которые собираются каждый год. В этом случае мы получим массив данных, в котором содержатся и данные об однородных объектах за один и тот же период времени, и последовательные значения одной экономической переменной в различные периоды времени. Но если совокупность предприятий из года в год будет различна, то такие данные уже не будут панельными.

Набором признаков называется совокупность экономической информации, которая характеризует изучаемый процесс или объект.

Признаки взаимосвязаны между собой, и при этом они могут выступать в одной из двух ролей:

1) в роли результативного или зависимого признака;

2) в роли факторного или независимого признака.

В эконометрических моделях результативный признак называется объясняемой переменной, а факторный признак называется объясняющей переменной.

В эконометрическом моделировании выделяют следующие виды экономических переменных:

1) экзогенные или независимые переменные (x), значения которых задаются извне. В определённой степени экзогенные переменные поддаются управлению;

2) эндогенные или зависимые переменные (y), значения которых определяются внутри модели;

3) лаговые переменные – это экзогенные или эндогенные переменные, которые относятся к предыдущим моментам времени и находятся в эконометрической модели одновременно с переменными, относящимися к текущему моменту времени. Например, x_{t-1} – это лаговая экзогенная переменная, а y_{t-1} – это лаговая эндогенная переменная;

4) предопределённые или объясняющие переменные – это лаговые (x_{t-1}) и текущие (x) экзогенные переменные, а также лаговые эндогенные переменные (y_{t-1}).

5) фиктивные переменные используются в эконометрических моделях для характеристики явления или процесса, в отношении которого нет данных по качественному признаку;

6) переменные-заместители искусственно вводятся в эконометрическую модель для характеристики явления или процесса, который не может быть количественно охарактеризован. При этом переменная-заместитель тесно коррелирует с этим явлением.

В эконометрических исследованиях большое внимание уделяется проблеме данных, т.е. специальным методам работы при наличии данных с пропусками, влиянию агрегирования данных на эконометрические измерения. Зачастую по единицам исследуемой совокупности информация отсутствует, а в наличии имеются данные, характеризующие более крупные единицы (агрегаты). Следует отметить, что при агрегировании временных данных опасность искажения результатов измерений гораздо больше, чем при агрегировании пространных данных, потому что с одной стороны, добавляется эффект автокорреляции, а с другой – происходит погашение случайной компоненты.

1. 4. Измерения в экономике

Понятие эконометрика включает экономические измерения. Измерение – это, прежде всего, <u>получение, сравнение и упорядочение информации</u>. Это определение исходит из того, что измерение предполагает выделение некоторого свойства, по которому производится сравнение объектов в определенном отношении.

Другое понимание измерения исходит из числового выражения результата, т.е. **измерение** трактуется как **операция, в результате которой получается численное значение величины.**

Третий подход к **измерению** связан с обязательным наличием **единицы измерения** (эталона). Любому измерению предшествует качественный анализ, учитывающий цели исследования качественный анализ необходим и после выполнения измерения.

Специфика экономических измерений состоит в наличии большого числа разнородных данных – разнородных ресурсов, разнородных результатов (например, товаров и услуг).

Нередко в экономических измерениях возникает задача отражения иерархии измерителей, которая выражается в выделении интегрального и частных показателей. **Точность измерения – это его адекватность.**

Для социально экономических измерений характерны специфические представления о точности. Экономику относят к неточным наукам, т.к. невозможно провести измерение с произвольно малой погрешностью.

По объективным причинам для социально-экономических измерений характерна низкая контролируемость их точности.

Основной базой данных для эконометрических исследований служат данные официальной статистики либо данные бухгалтерского учета. Таким образом, проблемы экономического измерения – это проблемы статистики и учета.

1.5. Классификация эконометрических моделей

Общая классификация эконометрических или экономико-математических моделей включает более десяти основных признаков, но с развитием экономико-математических исследований проблема классификации данных моделей всё более усложняется. Помимо появления новых типов моделей (особенно смешанных типов) и новых признаков их классификаций, также идёт процесс интеграции моделей различных типов в более сложные, комбинированные модельные конструкции.

Рассмотрим несколько ключевых **классификаций эконометрических моделей:**

1) **по целевому назначению**:

а) теоретико-аналитические модели, которые используются при исследовании общих свойств и закономерностей экономических процессов;

б) прикладные модели, которые используются при решении конкретных экономических задач (модели экономического анализа, прогнозирования, управления);

Также эконометрические модели могут быть использованы при исследовании различных сторон народного хозяйства и его отдельных частей.

2) **по исследуемым экономическим процессам и содержательной проблематике**. При этом выделяются модели:

а) народного хозяйства в целом и его отдельных подсистем-отраслей, регионов и т.д.;

б) производства и потребления;

в) формирования и распределения доходов;

г) трудовых ресурсов;

д) ценообразования;

е) финансовых связей и др.

3) **деление на дескриптивные и нормативные модели:**

а) дескриптивные модели предназначены для объяснения наблюдаемых фактов или для построения вероятностного прогноза. В качестве примера дескриптивной модели можно привести производственные функции и функции покупательного спроса, построенные на основе обработки статистических данных;

б) нормативные модели отвечают на вопрос «как это должно быть?», т.е. предполагают целенаправленную деятельность. В качестве примера нормативной модели можно привести модели оптимального планирования, характеризующие тем или иным образом цели экономического развития, возможности и средства их достижения;

4) **по характеру отражения причинно-следственных связей**. При этом выделяют:

а) модели жестко детерминистские;

б) модели, в которых учитываются факторы случайности и неопределенности.

Вследствие перехода от жёстко детерминированных моделей к моделям второго типа, были разработаны реальные возможности успешного применения более совершенной методологии моделирования экономических процессов, учитывающих факторы случайности и неопределённости, а именно:

а) проведение многовариантных расчетов и модельных экспериментов с вариацией конструкции модели и ее исходных данных;

б) изучение устойчивости и надежности получаемых решений;

в) выделение зоны неопределенности;

г) включение в модель резервов;

д) применение приемов, повышающих приспособляемость (адаптивность) экономических решений к вероятным и непредвиденным ситуациям

В последнее время широко применяются эконометрические модели, непосредственно отражающие стохастичность и неопределенность экономических процессов. Данные модели используют соответствующий математический аппарат: теорию вероятностей и математическую статистику, теорию игр и статистических решений, теорию массового обслуживания, теорию случайных процессов.

5) **по способам отражения фактора времени:**

а) статические модели, характеризующие исследуемую зависимость между переменными на определённый момент времени;

б) динамические модели, характеризующие изменение экономических процессов во времени.

1.6. Эконометрические модели

Многообразие и сложность экономических процессов предопределяет многообразие моделей, используемых для эконометрического анализа. Правильный выбор вида экономической модели является отправной точкой для качественного ее анализа. Безусловно, на практике неизвестно, какая модель является верной, и зачастую подбирают такую модель, которая наиболее точно соответствует реальным данным. При этом необходимо учитывать, что идеальной модели не существует.

Признаки «хорошей» модели:

1. *Скупость (простота)*. Модель должна быть максимально простой. Данное свойство определяется тем фактом, что модель не отражает действительность идеально, а является ее упрощением. Поэтому из двух моделей, приблизительно одинаково отражающих реальность, предпочтение отдается модели, содержащей меньшее число объясняющих переменных.

2. *Единственность.* Для любого набора статистических данных определяемые коэффициенты должны вычисляться однозначно.

3. *Максимальное соответствие*. Уравнение тем лучше, чем большую часть разброса зависимой переменной оно может объяснить.

4. *Согласованность с теорией*. Никакое уравнение не может быть признано качественным, если оно не соответствует известным теоретическим предпосылкам. Другими словами, модель обязательно должна опираться на теоретический фундамент, так как в противном случае результат использования регрессионного уравнения может быть весьма плачевным.

5. *Прогнозные качества*. Модель может быть признана качественной, если полученные на ее основе прогнозы подтверждаются реальностью.

Поскольку не существует какого-либо единого правила построения моделей, анализ перечисленных свойств позволяет строить более качественные эконометрические модели.

Одно из главных направлений эконометрического анализа – постоянное совершенствование моделей. Здесь следует отметить, что какого-то глобального подхода, определяющего заранее возможные пути совершенствования, нет и, скорее всего, быть не может. Исследователь должен помнить, что совершенной модели не существует. В силу постоянно изменяющихся условий протекания экономических процессов не может быть и постоянно качественных моделей. Новые условия требуют пересмотра даже весьма устойчивых моделей.

До сих пор достаточно спорным является вопрос, как строить модели: начинать с самой простой и постоянно усложнять ее или начинать с максимально сложной модели и упрощать ее на основе проводимых исследований.

Выделяют три основных класса эконометрических моделей.

1. Модели временных рядов.

Модель представляет собой зависимость результативного признака от переменной времени или переменных, относящихся к другим моментам времени. К моделям временных рядов, в которых результативный признак зависит от времени, относятся:

1) модель тренда (модель зависимости результативного признака от трендовой компоненты);

2) модель сезонности (модель зависимости результативного признака от сезонной компоненты);

3) модель тренда и сезонности.

К моделям временных рядов, в которых результативный признак зависит от переменных, датированных другими моментами времени, относятся:

1) модели с распределенным лагом, которые объясняют вариацию результативного признака в зависимости от предыдущих значений факторных переменных;

2) модели авторегрессии, которые объясняют вариацию результативного признака в зависимости от предыдущих значений результативных переменных;

3) модели ожидания, объясняющие вариацию результативного признака в зависимости от будущих значений факторных или результативных переменных.

Модели временных рядов делятся на модели, построенные по стационарным и нестационарным временным рядам.

Стационарные временные ряды характеризуются постоянными во времени средней, дисперсией и автокорреляцией, т. е. данный временной ряд не содержит трендового и сезонного компонента.

Если временной ряд не отвечает перечисленным условиям, то он является нестационарным (т. е. содержит трендовую и сезонную компоненты).

2. Регрессионные модели с одним уравнением.

В подобных моделях зависимая (результативная) переменная, обозначаемая обычно y, представляется в виде функции факторных или независимых признаков $x_1 ... x_n$:

$$\tilde{y} = a + b_1 x_1 + b_1 x_2 + ... + x_n b_n$$

где $a, b_1, ..., b_n$ - параметры регрессионного уравнения.

Регрессионные модели делятся на парные (с одним факторным признаком) и множественные регрессии.

В зависимости от вида функции $f(x, b)$ модели делятся на линейные и нелинейные регрессии.

3. Системы одновременных уравнений.

Данные модели описываются системами взаимозависимых регрессионных уравнений. Системы могут состоять из тождеств и регрессионных уравнений, каждое из которых может включать в себя не только факторные переменные, но и результативные переменные из других уравнений системы.

Для тождеств характерно то, что их вид и значения параметров известны.

Регрессионные уравнения, из которых состоит система, называются пове**денческими** уравнениями. В поведенческих уравнениях значения параметров являются неизвестными и подлежат оцениванию.

1.7. Этапы эконометрического моделирования

В качестве цели эконометрического моделирования обычно рассматривают анализ исследуемого экономического объекта (процесса); прогноз его экономических показателей, имитацию развития объекта при различных значениях экзогенных переменных (отражая их случайный характер, изменение во времени), выработку управленческих решений.

Можно выделить несколько этапов эконометрического моделирования.

1. Постановочный.

Формируется цель исследования, набор участвующих в модели экономических переменных. На данном этапе определяются конечные цели и задачи исследования и набор участвующих в модели факторных и результативных экономических переменных.

При выборе экономических переменных необходимо теоретическое обоснование каждой переменной (при этом рекомендуется, чтобы число их было не очень большим и, как минимум, в несколько раз меньше числа наблюдений). Объясняющие переменные не должны быть связаны функциональной или тесной корреляционной зависимостью, так как это может привести к невозможности оценки параметров модели или к получению неустойчивых, не имеющим реального смысла оценок, т.е. к явлению мультиколлинеарности.

Для отбора переменных могут быть использованы различные методы, в частности процедуры пошагового отбора переменных. А для оценки влияния качественных признаков (например, пол, образование) могут быть использованы фиктивные переменные. Но в любом случае определяющим при включении в модель тех или иных переменных является экономический (качественный) анализ исследуемого объекта.

Включение в эконометрическую модель той или иной переменной должно быть теоретически обосновано. Число переменных не должно быть слишком большим. Факторные переменные не должны быть связаны функциональной или тесной корреляционной связью, присутствие в модели условия мультиколлинеарности может привести к негативным последствиям всего процесса моделирования.

2. Априорный. На этом этапе проводятся теоретический анализ сущности изучаемого процесса, а также формирование и формализация известной до моделирования (априорной) информации.

3. Параметризация. Осуществляется выбор общего вида модели и выявление состава и формы входящих в нее связей, т.е. происходит непосредственно моделирование.

Основная задача этапа моделирования заключается в выборе наиболее оптимального вида функции зависимости результативной переменной от факторных признаков. Если возникает возможность выбора между нелинейной и линейной формой зависимости, то предпочтение всегда отдается линейной форме как наиболее простой и надежной.

Помимо этого, на этапе моделирования решается задача спецификации модели путем:

1) аппроксимации математической формой выявленных связей и соотношений между переменными;

2) определения зависимых и независимых переменных;

3) формулировки исходных предпосылок и ограничений модели.

Успех эконометрического моделирования во многом зависит от правильного решения проблемы спецификации модели.

4. Информационный. Происходит сбор необходимой статистической базы данных, т. е. эмпирических (наблюдаемых) значений экономических переменных, анализ качества собранной информации.

5. Идентификация модели. На данном этапе осуществляются статистический анализ модели и оценка ее параметров.

6. Оценка качества (верификация) модели. Проверяются достоверность и адекватность модели, т.е. определяется, насколько успешно решены задачи спецификации и идентификации модели, какова точность расчетов, полученных на ее основе. Построенная модель должна быть адекватна реальному экономическому процессу. Если качество модели оказалось неудовлетворительным, то вновь возвращаются ко второму этапу моделирования.

7. **Интерпретация результатов эконометрического моделирования.** На этом этапе делаются выводы по полученным результатам. Осуществляется прогнозирование.

Вопросы для самоконтроля

1. Охарактеризуйте предмет эконометрики.
2. Укажите основные этапы эконометрического исследования.
3. Каковы особенности причинно-следственных отношений в социально-экономических явлениях?
4. Какие типы данных используются в эконометрическом исследовании?
5. Опишите основные этапы построения эконометрической модели.
6. Какие виды аналитических зависимостей, наиболее часто используются при построении моделей?
7. Какие методы используются для отбора факторов?
8. Охарактеризуйте принципы и критерии эконометрики.

Тестовые задания

1. Под эконометрикой в широком смысле слова понимается …

а) совокупность теоретических результатов

б) совокупность различного рода экономических исследований, проводимых с использованием математических методов

в) самостоятельная научная дисциплина

г) применение статистических методов

2. Переменные, используемые в эконометрике называются …

а) факторные, результативные

б) линейные, нелинейные

в) экзогенные, эндогенные, предопределенные

г) внешние, внутренние

3. Этапы построения эконометрической модели …

а) постановочный, априорный, параметризация, идентификация модели

б) постановочный, информационный, априорный, верификация модели

в) постановочный, априорный, параметризация, информационный, идентификация модели, верификация модели

г) параметризация, информационный, идентификация модели

4. Эндогенные переменные – это …

а) лаговые переменные

б) внешние переменные

в) автономные переменные

г) внутренние переменные

5. Верификация модели – это …

а) статистический анализ модели
б) определение конечных целей моделирования
в) сбор необходимой статистической информации
г) сопоставление реальных и модельных данных, проверка адекватности модели

6. Эконометрическая модель – это …

а) приближенное описание объекта моделирования, выраженное с помощью математической символики
б) статистическая модель
в) описание экономического объекта
г) математическая модель, которая отражает влияние факторов на результат функционирования экономической системы

7. Типы данных, существующих в эконометрике …

а) регрессионные, временные
б) пространственные, временные, пространственно-временные
в) экзогенные, эндогенные, предопределенные
г) факторные, результативные

8. Предопределенные переменные – это …

а) внутренние переменные
б) автономные переменные
в) переменные, которые задаются из вне модели
г) лаговые эндогенные переменные

9. Основные типы эконометрических моделей …

а) модели тренда, модель сезонности
б) модель временных рядов, регрессионные модели, системы одновременных уравнений
в) аддитивные и мультипликативные модели
г) модели сезонности

10. Эконометрика – это наука, предметом изучения которой является …

а) количественная сторона массовых социально-экономических явлений и процессов в конкретных условиях места и времени
б) количественное выражение взаимосвязей в экономике
в) общие закономерности случайных явлений и методы количественной оценки влияния случайных факторов

11. При моделировании экономических процессов используют типы данных …

а) временные
б) пространственные
в) экономические
г) статистические
д) панельные

2. ПАРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

2.1. Понятие парного регрессионного анализа

Парная регрессия представляет собой регрессию между двумя переменными – y и x, т. е. модель вида:

$$\tilde{y}_x = f(x),$$

где y – зависимая переменная (результативный признак); x – независимая, или объясняющая, переменная (признак-фактор). Знак «^» означает, что между переменными x и y нет строгой функциональной зависимости, поэтому практически в каждом отдельном случае величина y складывается из двух слагаемых:

$$y = y_x + \varepsilon,$$

где y – фактическое значение результативного признака; $\tilde{y}_x$ – теоретическое значение результативного признака, найденное исходя из уравнения регрессии; ε – случайная величина, характеризующая отклонения реального значения результативного признака от теоретического, найденного по уравнению регрессии.

Случайная величина ε называется также возмущением. Она включает влияние не учтенных в модели факторов, случайных ошибок и особенностей измерения. Ее присутствие в модели порождено тремя источниками: спецификацией модели, выборочным характером исходных данных, особенностями измерения переменных.

От правильно выбранной спецификации модели зависит величина случайных ошибок: они тем меньше, чем в большей мере теоретические значения результативного признака $\tilde{y}_x$, подходят к фактическим данным y.

К ошибкам спецификации относятся неправильный выбор той или иной математической функции для $\tilde{y}_x$ и недоучет в уравнении регрессии какого-либо существенного фактора, т. е. использование парной регрессии вместо множественной.

Наряду с ошибками спецификации могут иметь место ошибки выборки, которые имеют место в силу неоднородности данных в исходной статистической совокупности, что, как правило, бывает при изучении экономических процессов. Если совокупность неоднородна, то уравнение регрессии не имеет практического смысла. Для получения хорошего результата обычно исключают из совокупности единицы с аномальными значениями исследуемых признаков. И в этом случае результаты регрессии представляют собой выборочные характеристики.

Использование временной информации также представляет собой выборку из всего множества хронологических дат. Изменив временной интервал, можно получить другие результаты регрессии.

Наибольшую опасность в практическом использовании методов регрессии представляют ошибки измерения. Если ошибки спецификации можно уменьшить, изменяя форму модели (вид математической формулы), а ошибки выборки – увеличивая объем исходных данных, то ошибки измерения практически сводят на нет все усилия по количественной оценке связи между признаками.

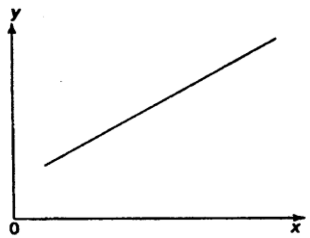
Особенно велика роль ошибок измерения при исследовании на макроуровне. Так, в исследованиях спроса и потребления в качестве объясняющей переменной широко используется «доход на душу населения». Вместе с тем, статистическое измерение величины дохода сопряжено с рядом трудностей и не лишено возможных ошибок, например, в результате наличия скрытых доходов.

Предполагая, что ошибки измерения сведены к минимуму, основное внимание в эконометрических исследованиях уделяется ошибкам спецификации модели.

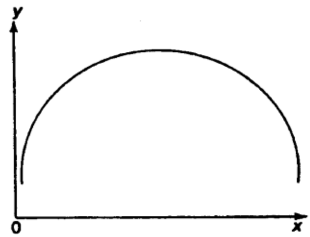
В парной регрессии выбор вида математической функции $\tilde{y}_x = f(x)$ может быть осуществлен тремя методами:

1) графическим;
2) аналитическим, т.е. исходя из теории изучаемой взаимосвязи;
3) экспериментальным.

При изучении зависимости между двумя признаками графический метод подбора вида уравнения регрессии достаточно нагляден. Он основан на поле корреляции. Основные типы кривых, используемые при количественной оценке связей, представлены на рис. 1.1:



$y_x = a + b \cdot x$



$y_x = a + b \cdot x + c \cdot x^2$

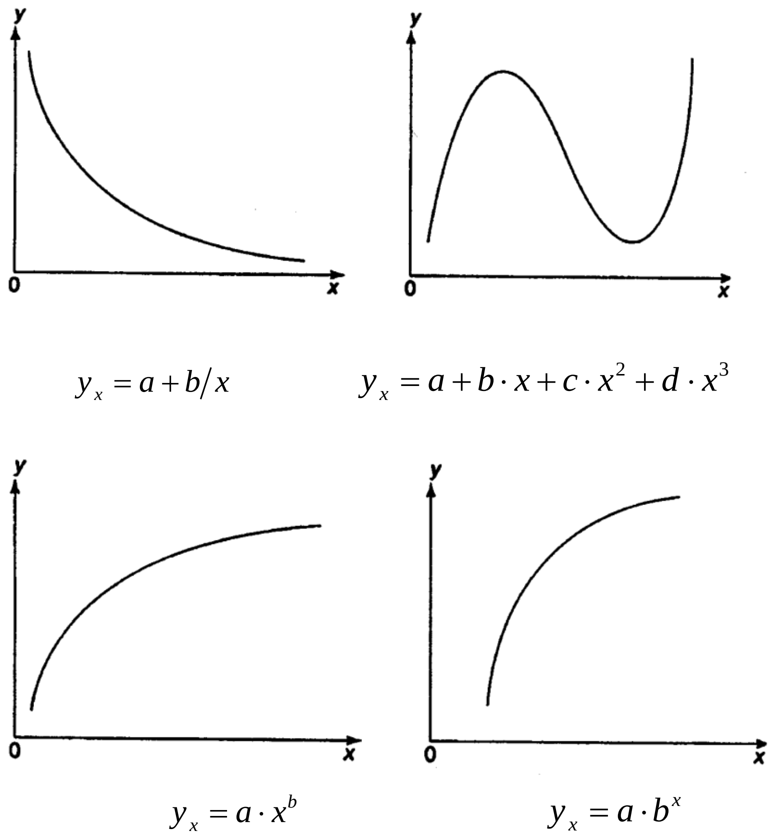


Рис. 2.1. Основные типы кривых, используемые приколичественной оценке связей между двумя переменными

Значительный интерес представляет аналитический метод выбора типа уравнения регрессии. Он основан на изучении материальной природы связи исследуемых признаков.

При обработке информации на компьютере выбор вида уравнения регрессии обычно осуществляется экспериментальным методом, т.е. путем сравнения величины остаточной дисперсии $\sigma^2_{\text{ост}}$, рассчитанной при разных моделях.

Если уравнение регрессии проходит через все точки корреляционного поля, что возможно только при функциональной связи, когда все точки лежат на линии регрессии $\tilde{y}_x = f(x)$, то фактические значения результативного признака совпадают с теоретическими $y = y_x$, т.е. они полностью обусловлены влиянием фактора X. В этом случае остаточная дисперсия $\sigma^2_{\text{ост}} = 0$.

В практических исследованиях, как правило, имеет место некоторое рассеяние точек относительно линии регрессии. Оно обусловлено влиянием прочих, не учитываемых в уравнении регрессии, факторов. Иными словами, имеют место отклонения фактических данных от теоретических $\left(y - y_x\right)$. Величина этих отклонений и лежит в основе расчета остаточной дисперсии:

$$\sigma^2_{\text{ост}} = \frac{1}{n}\sum\left(y - y_x\right)^2.$$

Чем меньше величина остаточной дисперсии, тем меньше влияние не учитываемых в уравнении регрессии факторов и тем лучше уравнение регрессии подходит к исходным данным.

Считается, что число наблюдений должно в 7-8 раз превышать число рассчитываемых параметров при переменной *x*. Это означает, что искать линейную регрессию, имея менее 7 наблюдений, вообще не имеет смысла. Если вид функции усложняется, то требуется увеличение объема наблюдений, ибо каждый параметр при *x* должен рассчитываться хотя бы по 7 наблюдениям. Значит, если мы выбираем параболу второй степени $y_x = a + b \cdot x + c \cdot x^2$, то требуется объем информации уже не менее 14 наблюдений.

2.2 Линейная модель парной регрессии

Рассмотрим простейшую модель парной регрессии – линейную регрессию. Линейная регрессия находит широкое применение в эконометрике ввиду четкой экономической интерпретации ее параметров.

Линейная регрессия сводится к нахождению уравнения вида

$$y_x = a + b \cdot x \text{ или } y = a + b \cdot x + \varepsilon. \qquad (2.1)$$

Уравнение вида $y_x = a + b \cdot x$ позволяет по заданным значениям фактора x находить теоретические значения результативного признака, подставляя в него фактические значения фактора x.

Построение линейной регрессии сводится к оценке ее параметров – a и b. Классический подход к оцениванию параметров линейной регрессии основан на методе наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров a и b, при которых сумма квадратов отклонений фактических значений результативного признака y от теоретических $\tilde{y}_x$ минимальна:

$$\sum_{i=1}^{n}\left(y_i - y_{x_i}\right)^2 = \sum_{i=1}^{n}\varepsilon_i^2 \quad \to \quad \min \tag{2.2}$$

Т.е. из всего множества линий линия регрессии на графике выбирается так, чтобы сумма квадратов расстояний по вертикали между точками и этой линией была бы минимальной (рис. 2):

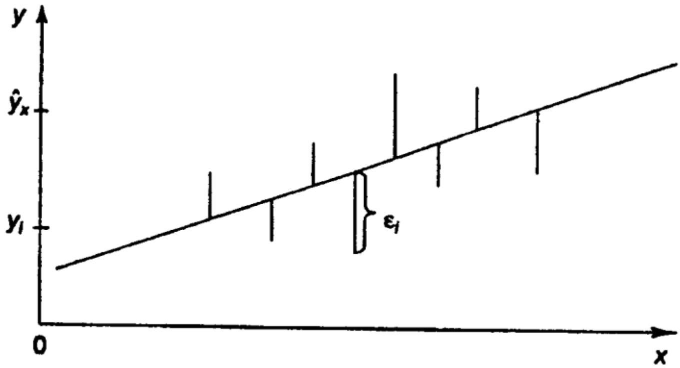


Рис. 2.2. Линия регрессии с минимальной дисперсией остатков

Как известно из курса математического анализа, чтобы найти минимум функции (2.2), надо вычислить частные производные по каждому из параметров a и b и приравнять их к нулю. Обозначим $\sum\varepsilon_i^2$ через $S(a,b)$, тогда:

$$S(a,\ b) = \sum\left(y - a - b\cdot x\right)^2 \cdot$$
$$\begin{cases} \dfrac{\partial S}{\partial a} = -2\sum\left(y - a - b\cdot x\right) = 0; \\ \dfrac{\partial S}{\partial b} = -2\sum x\left(y - a - b\cdot x\right) = 0. \end{cases} \tag{2.3}$$

После несложных преобразований, получим следующую систему линейных уравнений для оценки параметров a и b:

$$\begin{cases} a\cdot n \quad + b\cdot\sum x = \sum y; \\ a\cdot\sum x + b\cdot\sum x^2 = \sum x\cdot y. \end{cases} \qquad (2.4)$$

Решая систему уравнений (2.4), найдем искомые оценки параметров a и b. Можно воспользоваться следующими готовыми формулами, которые следуют непосредственно из решения системы (2.4):

$$a = \bar{y} - b\cdot\bar{x}, \qquad b = \frac{\operatorname{cov}(x, y)}{\sigma_x^2}, \qquad (2.5)$$

$\operatorname{cov}(x,y) = \overline{y\cdot x} - \bar{y}\cdot\bar{x}$ – ковариация признаков x и y,

$\sigma_x^2 = \overline{x^2} - \bar{x}^2$ – дисперсия признака x и

$$\bar{x} = \frac{1}{n}\sum x, \quad \bar{y} = \frac{1}{n}\sum y, \quad \overline{y\cdot x} = \frac{1}{n}\sum y\cdot x, \quad \overline{x^2} = \frac{1}{n}\sum x^2.$$

Ковариация – числовая характеристика совместного распределения двух случайных величин, равная математическому ожиданию произведения отклонений этих случайных величин от их математических ожиданий. Дисперсия – характеристика случайной величины, определяемая как математическое ожидание квадрата отклонения случайной величины от ее математического ожидания. Математическое ожидание – сумма произведений значений случайной величины на соответствующие вероятности.

Параметр b называется коэффициентом регрессии. Его величина показывает среднее изменение результата с изменением фактора на одну единицу.

Возможность четкой экономической интерпретации коэффициента регрессии сделала линейное уравнение регрессии достаточно распространенным в эконометрических исследованиях.

Формально a – значение y при $x = 0$. Если признак-фактор x не может иметь нулевого значения, то вышеуказанная трактовка свободного члена a не имеет смысла, т.е. параметр a может не иметь экономического содержания.

Уравнение регрессии всегда дополняется показателем тесноты связи. При использовании линейной регрессии в качестве такого показателя выступает линейный коэффициент корреляции r_{xy}, который можно рассчитать по следующим формулам:

$$r_{xy} = b\cdot\frac{\sigma_x}{\sigma_y} = \frac{\operatorname{cov}(x, y)}{\sigma_x\cdot\sigma_y}. \qquad (2.6)$$

Линейный коэффициент корреляции находится в пределах: $-1 \le r \le 1$. Чем ближе абсолютное значение r_{xy} к единице, тем сильнее линейная связь между факторами (при $r = \pm 1$) имеем строгую функциональную зависимость). Но следует иметь в виду, что близость абсолютной величины линейного коэффициента корреляции к нулю еще не означает отсутствия связи между признаками. При другой (нелинейной) спецификации модели связь между признаками может оказаться достаточно тесной.

Для оценки качества подбора линейной функции рассчитывается квадрат линейного коэффициента корреляции r_{xy}^2, называемый коэффициентом детерминации. Коэффициент детерминации характеризует долю дисперсии результативного признака y, объясняемую регрессией, в общей дисперсии результативного признака:

$$r_{xy}^2 = 1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2}, \qquad (2.7)$$

$$\sigma_{\text{ост}}^2 = \frac{1}{n}\sum\left(y - y_x\right)^2, \quad \sigma_y^2 = \frac{1}{n}\sum\left(y - \bar{y}\right)^2 = \overline{y^2} - \bar{y}^2.$$

Соответственно величина $1 - r_{xy}^2$ характеризует долю дисперсии y, вызванную влиянием остальных, не учтенных в модели, факторов.

После того как найдено уравнение линейной регрессии, проводится оценка значимости как уравнения в целом, так и отдельных его параметров.

Проверить значимость уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных (одной или нескольких) для описания зависимой переменной.

Чтобы иметь общее суждение о качестве модели из относительных отклонений по каждому наблюдению, определяют среднюю ошибку аппроксимации:

$$\bar{A} = \frac{1}{n}\sum\left|\frac{y - y_x}{y}\right| \cdot 100\%. \qquad (2.8)$$

Средняя ошибка аппроксимации не должна превышать 8–10%.

Оценка значимости уравнения регрессии в целом производится на основе F-критерия Фишера, которому предшествует дисперсионный анализ. В математической статистике дисперсионный анализ рассматривается как самостоятельный инструмент статистического анализа. В эконометрике он применяется как вспомогательное средство для изучения качества регрессионной модели.

Согласно основной идее дисперсионного анализа, общая сумма квадратов отклонений переменной y от среднего значения $\bar{y}$ раскладывается на две части – «объясненную» и «необъясненную»:

$$\sum(y-\bar{y})^2 = \sum\left(y_x-\bar{y}\right)^2 + \sum\left(y-y_x\right)^2,$$

где $\sum(y-\bar{y})^2$ – общая сумма квадратов отклонений; $\sum\left(y_x-\bar{y}\right)^2$ – сумма квадратов отклонений, объясненная регрессией (или факторная сумма квадратов отклонений); $\sum\left(y-y_x\right)^2$ – остаточная сумма квадратов отклонений, характеризующая влияние неучтенных в модели факторов.

Схема дисперсионного анализа имеет вид, представленный в таблице 1 (n – число наблюдений, m – число параметров при переменной x).

Таблица 2.1

Компоненты дисперсии	Сумма квадратов	Число степеней свободы	Дисперсия на одну степень свободы
Общая	$\sum(y-\bar{y})^2$	$n-1$	$S^2_{\text{общ}} = \dfrac{\sum(y-\bar{y})^2}{n-1}$
Факторная	$\sum\left(y_x-\bar{y}\right)^2$	m	$S^2_{\text{факт}} = \dfrac{\sum\left(y_x-\bar{y}\right)^2}{m}$
Остаточная	$\sum\left(y-y_x\right)^2$	$n-m-1$	$S^2_{\text{ост}} = \dfrac{\sum\left(y-y_x\right)^2}{n-m-1}$

Определение дисперсии на одну степень свободы приводит дисперсии к сравнимому виду. Сопоставляя факторную и остаточную дисперсии в расчете на одну степень свободы, получим величину F-критерия Фишера:

$$F = \frac{S^2_{\text{факт}}}{S^2_{\text{ост}}}. \qquad (2.9)$$

Фактическое значение F-критерия Фишера (2.9) сравнивается с табличным значением $F_{\text{табл}}(\alpha; k_1; k_2)$ при уровне значимости α и степенях свободы $k_1 = m$ и $k_2 = n - m - 1$. При этом, если фактическое значение F-критерия больше табличного, то признается статистическая значимость уравнения в целом.

Для парной линейной регрессии $m = 1$, поэтому

$$F = \frac{S^2_{\text{факт}}}{S^2_{\text{ост}}} = \frac{\sum\left(y_x - \bar{y}\right)^2}{\sum\left(y - y_x\right)^2} \cdot (n-2). \qquad (2.10)$$

Величина F-критерия связана с коэффициентом детерминации r^2_{xy}, и ее можно рассчитать по следующей формуле:

$$F = \frac{r^2_{xy}}{1 - r^2_{xy}} \cdot (n-2). \qquad (2.11)$$

В парной линейной регрессии оценивается значимость не только уравнения в целом, но и отдельных его параметров. С этой целью по каждому из параметров определяется его стандартная ошибка: m_b и m_a.

Стандартная ошибка коэффициента регрессии определяется по формуле:

$$m_b = \sqrt{\frac{S^2_{\text{ост}}}{\sum\left(x - \bar{x}\right)^2}} = \frac{S_{\text{ост}}}{\sigma_x \cdot \sqrt{n}}, \qquad (2.12)$$

где $S^2_{\text{ост}} = \dfrac{\sum\left(y - y_x\right)^2}{n-2}$ – остаточная дисперсия на одну степень свободы.

Величина стандартной ошибки совместно с t-распределением Стьюдента при $n-2$ степенях свободы применяется для проверки существенности коэффициента регрессии и для расчета его доверительного интервала.

Для оценки существенности коэффициента регрессии его величина сравнивается с его стандартной ошибкой, т.е. определяется фактическое значение t-критерия Стьюдента: $t_b = \dfrac{b}{m_b}$ которое затем сравнивается с табличным значением при определенном уровне значимости α и числе степеней свободы $(n-2)$. Доверительный интервал для коэффициента регрессии определяется как $b \pm t_{\text{табл}} \cdot m_b$. Поскольку знак коэффициента регрессии указывает на рост результативного признака y при увеличении признака-фактора x ($b > 0$), уменьшение результативного признака при увеличении признака-фактора ($b < 0$) или его независимость от независимой переменной ($b = 0$) (см. рис. 1.3), то границы доверительного интервала для коэффициента регрессии не должны содержать противоречивых результатов, например, $-1{,}5 \le b \le 0{,}8$. Такого рода запись указывает, что истинное значение коэффициента регрессии одновременно содержит положительные и отрицательные величины и даже ноль, чего не может быть.

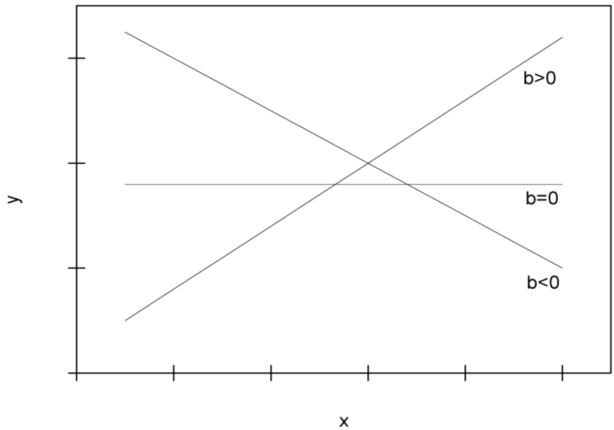


Рис.2.3. Наклон линии регрессии в зависимости от значения параметра *b*

Стандартная ошибка параметра *a* определяется по формуле:

$$m_a = \sqrt{S_{\text{ост}}^2 \cdot \frac{\sum x^2}{n \cdot \sum \left(x - \overline{x}\right)^2}} = S_{\text{ост}} \cdot \frac{\sqrt{\sum x^2}}{\sigma_x \cdot n}. \tag{2.13}$$

Процедура оценивания существенности данного параметра не отличается от рассмотренной выше для коэффициента регрессии. Вычисляется t-критерий: $t_a = \dfrac{a}{m_a}$, его величина сравнивается с табличным значением при $n-2$ степенях свободы.

Значимость линейного коэффициента корреляции проверяется на основе величины ошибки коэффициента корреляции m_r:

$$m_r = \sqrt{\frac{1-r^2}{n-2}}. \qquad (2.14)$$

Фактическое значение t-критерия Стьюдента определяется как

$$t_r = \frac{r}{m_r}.$$

Существует связь между t -критерием Стьюдента и F -критерием Фишера:

$$t_b = t_r = \sqrt{F}. \qquad (2.15)$$

В прогнозных расчетах по уравнению регрессии определяется предсказываемое $\tilde{y}_p$ значение как точечный прогноз $\tilde{y}_x$ при $x_p = x_k$, т.е. путем подстановки в уравнение регрессии $y_x = a + b \cdot x$ соответствующего значения x. Однако точечный прогноз явно не реален. Поэтому он дополняется расчетом стандартной ошибки $\tilde{y}_p$, т.е. m_{y_p}, и соответственно интервальной оценкой прогнозного значения $\tilde{y}_p$:

$$y_p - \Delta_{y_p} \le y_p \le y_p + \Delta_{y_p},$$

где $\Delta_{y_p} = m_{y_p} \cdot t_{\text{табл}}$, а m_{y_p} – средняя ошибка прогнозируемого индивидуального значения:

$$m_{y_p} = S_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{\left(x_p - \bar{x}\right)^2}{n \cdot \sigma_x^2}}. \qquad (2.16)$$

ПРАКТИЧЕСКОЕ ЗАДАНИЕ. По данным проведенного опроса восьми групп семей известны данные связи расходов населения на продукты питания с уровнем доходов семьи.

Таблица 2.2

Данные опроса семей

Расходы на продукты питания, y, тыс. руб.	0,9	1,2	1,8	2,2	2,6	2,9	3,3	3,8
Доходы семьи, x, тыс. руб.	1,2	3,1	5,3	7,4	9,6	11,8	14,5	18,7

Предположим, что связь между доходами семьи и расходами на продукты питания линейная. Для подтверждения нашего предположения построим поле корреляции.

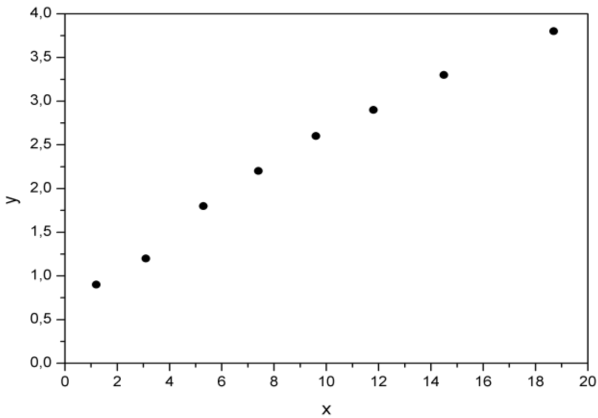


Рис. 2.4. Поле корреляции

По графику видно, что точки выстраиваются в некоторую прямую линию.

Для удобства дальнейших вычислений составим таблицу.

Таблица 2.3

Расчетная таблица

	x	y	$x \cdot y$	x^2	y^2	y_x	$y - y_x$	$\left(y - y_x\right)^2$	A_i, %
1	1,2	0,9	1,08	1,44	0,81	1,038	–0,138	0,0190	15,33
2	3,1	1,2	3,72	9,61	1,44	1,357	–0,157	0,0246	13,08
3	5,3	1,8	9,54	28,09	3,24	1,726	0,074	0,0055	4,11
4	7,4	2,2	16,28	54,76	4,84	2,079	0,121	0,0146	5,50
5	9,6	2,6	24,96	92,16	6,76	2,449	0,151	0,0228	5,81
6	11,8	2,9	34,22	139,24	8,41	2,818	0,082	0,0067	2,83
7	14,5	3,3	47,85	210,25	10,89	3,272	0,028	0,0008	0,85
8	18,7	3,8	71,06	349,69	14,44	3,978	–0,178	0,0317	4,68
Итого	71,6	18,7	208,71	885,24	50,83	18,717	–0,017	0,1257	52,19
Среднее значение	8,95	2,34	26,09	110,66	6,35	2,34	–	0,0157	6,52
σ	5,53	0,935	–	–	–	–	–	–	–
σ^2	30,56	0,874	–	–	–	–	–	–	–

Рассчитаем параметры линейного уравнения парной регрессии $y_x = a + b \cdot x$. Для этого воспользуемся формулами (1.5):

$$b = \frac{\operatorname{cov}(x,\, y)}{\sigma_x^2} = \frac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{26,09 - 8,95 \cdot 2,34}{30,56} = 0,168;$$

$$a = \overline{y} - b \cdot \overline{x} = 2,34 - 0,168 \cdot 8,95 = 0,836.$$

Получили уравнение: $y_x = 0,836 + 0,168 \cdot x$. Т.е. с увеличением дохода семьи на 1000 руб. расходы на питание увеличиваются на 168 руб.

Как было указано выше, уравнение линейной регрессии всегда дополняется показателем тесноты связи – линейным коэффициентом корреляции r_{xy}:

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = 0,168 \cdot \frac{5,53}{0,935} = 0,994.$$

Близость коэффициента корреляции к 1 указывает на тесную линейную связь между признаками.

Коэффициент детерминации $r_{xy}^2 = 0,987$ (примерно тот же результат получим, если воспользуемся формулой (1.7)) показывает, что уравнением регрессии объясняется 98,7% дисперсии результативного признака, а на долю прочих факторов приходится лишь 1,3%.

Оценим качество уравнения регрессии в целом с помощью F-критерия Фишера. Сосчитаем фактическое значение F-критерия:

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n-2) = \frac{0,987}{1 - 0,987} \cdot 6 = 455,54.$$

Табличное значение ($k_1 = 1$, $k_2 = n - 2 = 6$, $\alpha = 0,05$): $F_{\text{табл}} = 5,99$. Так как $F_{\text{факт}} > F_{\text{табл}}$, то признается статистическая значимость уравнения в целом.

Для оценки статистической значимости коэффициентов регрессии и корреляции рассчитаем *t*-критерий Стьюдента и доверительные интервалы каждого из показателей. Рассчитаем случайные ошибки параметров линейной регрессии и коэффициента корреляции

$$\left(S_{\text{ост}}^2 = \frac{\sum \left(y - y_x \right)^2}{n-2} = \frac{0,1257}{8-2} = 0,021 \right):$$

$$m_b = \frac{S_{\text{ост}}}{\sigma_x \cdot \sqrt{n}} = \frac{\sqrt{0,021}}{5,53 \cdot \sqrt{8}} = 0,0093,$$

$$m_a = S_{\text{ост}} \cdot \frac{\sqrt{\sum x^2}}{\sigma_x \cdot n} = \frac{\sqrt{0,021 \cdot 885,24}}{5,53 \cdot 8} = 0,0975,$$

$$m_r = \sqrt{\frac{1 - r^2}{n-2}} = \sqrt{\frac{1 - 0,987}{6}} = 0,0465.$$

Фактические значения t-статистик:

$$t_b = \frac{0,168}{0,0093} = 18,065, \quad t_a = \frac{0,836}{0,0975} = 8,574, \quad t_r = \frac{0,994}{0,0465} = 21,376.$$

Табличное значение t-критерия Стьюдента при $\alpha = 0,05$ и числе степеней свободы $\nu = n - 2 = 6$ есть $t_{\text{табл}} = 2,447$. Так как $t_b > t_{\text{табл}}$, $t_a > t_{\text{табл}}$ и $t_r > t_{\text{табл}}$, то признаем статистическую значимость параметров регрессии и показателя тесноты связи. Рассчитаем доверительные интервалы для параметров регрессии a и b: $a \pm t \cdot m_a$ и $b \pm t \cdot m_b$. Получим, что $a \in [0,597; \quad 1,075]$ и $b \in [0,145; \quad 0,191]$.

Средняя ошибка аппроксимации (находим с помощью столбца 10 таблицы 1.3; $A_i = \left| \frac{y_i - y_{x_i}}{y_i} \right| \cdot 100\%$) $\bar{A} = 6,52\%$ говорит о хорошем качестве уравнения регрессии, т.е. свидетельствует о хорошем подборе модели к исходным данным.

И, наконец, найдем прогнозное значение результативного фактора y_p при значении признака-фактора, составляющем 110% от среднего уровня $x_p = 1,1 \cdot \bar{x} = 1,1 \cdot 8,95 = 9,845$, т.е. найдем расходы на питание, если доходы семьи составят 9,85 тыс. руб.

$$y_p = 0,836 + 0,168 \cdot 9,845 = 2,490 \text{ (тыс. руб.)}$$

Значит, если доходы семьи составят 9,845 тыс. руб., то расходы на питание будут 2,490 тыс. руб.

Найдем доверительный интервал прогноза. Ошибка прогноза

$$m_{y_p} = S_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n \cdot \sigma_x^2}} = \sqrt{0,021 \cdot \left(1 + \frac{1}{8} + \frac{(9,845 - 8,95)^2}{8 \cdot 30,56}\right)} = 0,154,$$

а доверительный интервал ($y_p - \Delta_{y_p} \le y_p \le y_p + \Delta_{y_p}$):

$2{,}113 < y_p < 2{,}867\cdot$

Т.е. прогноз является статистически надежным.

Теперь на одном графике изобразим исходные данные и линию регрессии:

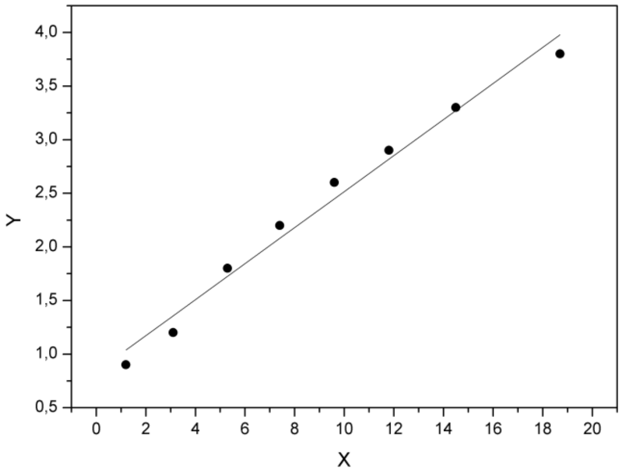


Рис. 2.5. Исходные данные и линию регрессии

2.3. Нелинейные модели парной регрессии и корреляции

Если между экономическими явлениями существуют нелинейные соотношения, то они выражаются с помощью соответствующих нелинейных функций.

Различают два класса нелинейных регрессий:

1. Регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, например

– полиномы различных степеней – $y_x = a + b \cdot x + c \cdot x^2$, $y_x = a + b \cdot x + c \cdot x^2 + d \cdot x^3$;

– равносторонняя гипербола – $y_x = a + b/x$;

– полулогарифмическая функция – $y_x = a + b \cdot \ln x \cdot$

2. Регрессии, нелинейные по оцениваемым параметрам, например

– степенная – $y_x = a \cdot x^b$;

– показательная – $y_x = a \cdot b^x$;

– экспоненциальная – $y_x = \mathrm{e}^{a+b\cdot x}$.

Регрессии нелинейные по включенным переменным приводятся к линейному виду простой заменой переменных, а дальнейшая оценка параметров производится с помощью метода наименьших квадратов. Рассмотрим некоторые функции.

Парабола второй степени $y_x = a + b \cdot x + c \cdot x^2$ приводится к линейному виду с помощью замены: $x = x_1$, $x^2 = x_2$. В результате приходим к двухфакторному уравнению $y_x = a + b \cdot x_1 + c \cdot x_2$, оценка параметров которого при помощи МНК, приводит к системе следующих нормальных уравнений:

$$\begin{cases} a \cdot n + b \cdot \sum x_1 + c \cdot \sum x_2 = \sum y; \\ a \cdot \sum x_1 + b \cdot \sum x_1^2 + c \cdot \sum x_1 \cdot x_2 = \sum x_1 \cdot y; \\ a \cdot \sum x_2 + b \cdot \sum x_1 \cdot x_2 + c \cdot \sum x_2^2 = \sum x_2 \cdot y. \end{cases}$$

А после обратной замены переменных получим

$$\begin{cases} a \cdot n + b \cdot \sum x + c \cdot \sum x^2 = \sum y; \\ a \cdot \sum x + b \cdot \sum x^2 + c \cdot \sum x^3 = \sum x \cdot y; \\ a \cdot \sum x^2 + b \cdot \sum x^3 + c \cdot \sum x^4 = \sum x^2 \cdot y. \end{cases} \qquad (2.17)$$

Парабола второй степени обычно применяется в случаях, когда для определенного интервала значений фактора меняется характер связи рассматриваемых признаков: прямая связь меняется на обратную или обратная на прямую.

Равносторонняя гипербола $y_x = a + b/x$ может быть использована для характеристики связи удельных расходов сырья, материалов, топлива от объема выпускаемой продукции, времени обращения товаров от величины товарооборота, процента прироста заработной платы от уровня безработицы (например, кривая А.В. Филлипса), расходов на непродовольственные товары от доходов или общей суммы расходов (например, кривые Э. Энгеля) и в других случаях. Гипербола приводится к линейному уравнению простой заменой: $z = 1/x$. Система линейных уравнений при применении МНК будет выглядеть следующим образом:

$$\begin{cases} a \cdot n \quad + b \cdot \sum \frac{1}{x} = \sum y; \\ a \cdot \sum \frac{1}{x} + b \cdot \sum \frac{1}{x^2} = \sum \frac{1}{x} \cdot y. \end{cases} \qquad (2.18)$$

Аналогичным образом приводятся к линейному виду зависимости $y_x = a + b \cdot \ln x$, $y_x = a + b \cdot \sqrt{x}$ и другие.

Несколько иначе обстоит дело с регрессиями нелинейными по оцениваемым параметрам, которые делятся на два типа: нелинейные модели внутренне линейные (приводятся к линейному виду с помощью соответствующих преобразований, например, логарифмированием) и нелинейные модели внутренне нелинейные (к линейному виду не приводятся).

К внутренне линейным моделям относятся, например, степенная функция – $y_x = a \cdot x^b$, показательная – $y_x = a \cdot b^x$, экспоненциальная – $y_x = \mathrm{e}^{a+b\cdot x}$, логистическая – $y_x = \dfrac{a}{1 + b \cdot \mathrm{e}^{-c\cdot x}}$, обратная – $y_x = \dfrac{1}{a + b \cdot x}$.

К внутренне нелинейным моделям можно, например, отнести следующие модели: $y_x = a + b \cdot x^c$, $y_x = a \cdot \left(1 - \dfrac{1}{1 - x^b}\right)$.

Среди нелинейных моделей наиболее часто используется степенная функция $y = a \cdot x^b \cdot \varepsilon$, которая приводится к линейному виду логарифмированием:

$$\ln y = \ln\left(a \cdot x^b \cdot \varepsilon\right);$$

$$\ln y = \ln a + b \cdot \ln x + \ln \varepsilon;$$

$$Y = A + b \cdot X + \mathrm{E},$$

где $Y = \ln y,\ X = \ln x,\ A = \ln a,\ \mathrm{E} = \ln \varepsilon$. Т.е. МНК мы применяем для преобразованных данных:

$$\begin{cases} A \cdot n \quad + b \cdot \sum X = \sum Y, \\ A \cdot \sum X + b \cdot \sum X^2 = \sum X \cdot Y, \end{cases}$$

а затем потенцированием находим искомое уравнение.

Широкое использование степенной функции связано с тем, что параметр b в ней имеет четкое экономическое истолкование – он является коэффициентом эластичности. (Коэффициент эластичности показывает, на сколько процентов измениться в среднем результат, если фактор изменится на 1%.) Формула для расчета коэффициента эластичности имеет вид:

$$Э = f'(x) \cdot \frac{x}{y}. \qquad (2.19)$$

Так как для остальных функций коэффициент эластичности не является постоянной величиной, а зависит от соответствующего значения фактора x, то обычно рассчитывается средний коэффициент эластичности:

$$\bar{Э} = f'(\bar{x}) \cdot \frac{\bar{x}}{\bar{y}}. \qquad (2.20)$$

Приведем формулы для расчета средних коэффициентов эластичности для наиболее часто используемых типов уравнений регрессии:

Таблица 2.5

Вид функции, y	Первая производная, y'	Средний коэффициент эластичности, $\bar{Э}$
$y = a + b \cdot x + \varepsilon$	b	$\dfrac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$
$y = a + b \cdot x + c \cdot x^2 + \varepsilon$	$b + 2c \cdot x$	$\dfrac{(b + 2c \cdot \bar{x}) \cdot \bar{x}}{a + b \cdot \bar{x} + c \cdot \bar{x}^2}$
$y = a + \dfrac{b}{x} + \varepsilon$	$-\dfrac{b}{x^2}$	$-\dfrac{b}{a \cdot \bar{x} + b}$
$y = a \cdot x^b \cdot \varepsilon$	$a \cdot b \cdot x^{b-1}$	b
$y = a \cdot b^x \cdot \varepsilon$	$a \cdot \ln b \cdot b^x$	$\bar{x} \cdot \ln b$
$y = a + b \cdot \ln x + \varepsilon$	$\dfrac{b}{x}$	$\dfrac{b}{a + b \cdot \ln \bar{x}}$
$y = \dfrac{1}{a + b \cdot x + \varepsilon}$	$-\dfrac{b}{(a + b \cdot x)^2}$	$-\dfrac{b \cdot \bar{x}}{a + b \cdot \bar{x}}$

Возможны случаи, когда расчет коэффициента эластичности не имеет смысла. Это происходит тогда, когда для рассматриваемых признаков бессмысленно определение изменения в процентах.

Уравнение нелинейной регрессии, так же, как и в случае линейной зависимости, дополняется показателем тесноты связи. В данном случае это индекс корреляции:

$$\rho_{xy} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}}, \qquad (2.21)$$

где $\sigma^2_y = \frac{1}{n}\sum\left(y - \bar{y}\right)^2$ – общая дисперсия результативного признака y, $\sigma^2_{\text{ост}} = \frac{1}{n}\sum\left(y - y_x\right)^2$ – остаточная дисперсия.

Величина данного показателя находится в пределах: $0 \le \rho_{xy} \le 1$. Чем ближе значение индекса корреляции к единице, тем теснее связь рассматриваемых признаков, тем более надежно уравнение регрессии.

Квадрат индекса корреляции носит название индекса детерминации и характеризует долю дисперсии результативного признака y, объясняемую регрессией, в общей дисперсии результативного признака:

$$\rho^2_{xy} = 1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y} = \frac{\sigma^2_{\text{объясн}}}{\sigma^2_y}, \qquad (2.22)$$

т.е. имеет тот же смысл, что и в линейной регрессии; $\sigma^2_{\text{объясн}} = \frac{1}{n}\sum\left(y_x - \bar{y}\right)^2$.

Индекс детерминации ρ^2_{xy} можно сравнивать с коэффициентом детерминации r^2_{xy} для обоснования возможности применения линейной функции. Чем больше кривизна линии регрессии, тем величина r^2_{xy} меньше ρ^2_{xy}. А близость этих показателей указывает на то, что нет необходимости усложнять форму уравнения регрессии и можно использовать линейную функцию.

Индекс детерминации используется для проверки существенности в целом уравнения регрессии по F-критерию Фишера:

$$F = \frac{\rho^2_{xy}}{1 - \rho^2_{xy}} \cdot \frac{n - m - 1}{m}, \qquad (2.23)$$

где ρ^2_{xy} – индекс детерминации, n – число наблюдений, m – число параметров при переменной x. Фактическое значение F-критерия (1.23) сравнивается с табличным при уровне значимости α и числе степеней свободы $k_2 = n - m - 1$ (для остаточной суммы квадратов) и

$k_1 = m$ (для факторной суммы квадратов).

О качестве нелинейного уравнения регрессии можно также судить и по средней ошибке аппроксимации, которая, так же как и в линейном случае, вычисляется по формуле (2.8).

ПРАКТИЧЕСКОЕ ЗАДАНИЕ. Связь между признаками носит нелинейный характер, и найдем параметры следующих нелинейных уравнений: $y = a + b \cdot \ln x + \varepsilon$, $y = a + b \cdot \sqrt{x} + \varepsilon$ и $y = a \cdot x^b \cdot \varepsilon$.

Для нахождения параметров регрессии $y_x = a + b \cdot \ln x$ делаем замену $z = \ln x$ и составляем вспомогательную таблицу ($\varepsilon = y - y_x$).

Найдем уравнение регрессии:

$$b = \frac{\operatorname{cov}(z,\, y)}{\sigma_z^2} = \frac{5,240 - 1,914 \cdot 2,34}{0,716} = 1,063,$$

$$a = \overline{y} - b \cdot \overline{z} = 2,34 - 1,063 \cdot 1,914 = 0,305,$$

т.е. получаем следующее уравнение регрессии: $y_x = 0,305 + 1,063 \cdot \ln x$

Теперь заполняем столбцы 8-11 нашей таблицы.

Индекс корреляции находим по формуле (2.21):

$$\rho_{xy} = \sqrt{1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2}} = \sqrt{1 - \frac{0,0711}{0,874}} = 0,958$$

а индекс детерминации $\rho_{xy}^2 = 0,918$, который показывает, что 91,8% вариации результативного признака объясняется вариацией признака-фактора, а 8,2% приходится на долю прочих факторов.

Средняя ошибка аппроксимации: $\overline{A} = 14,51\%$, что недопустимо велико.

F-критерий Фишера:

$$F = \frac{\rho_{xy}^2}{1 - \rho_{xy}^2} \cdot \frac{n - m - 1}{m} = \frac{0,919}{1 - 0,919} \cdot \frac{8 - 1 - 1}{1} = 68,07,$$

значительно превышает табличное $F_{\text{табл}} = 5,99$.

Таблица 2.5

Расчетная таблица

	x	z	y	$z \cdot y$	z^2	y^2	y_x	ε	ε^2	A_i
1	1,2	0,182	0,9	0,164	0,033	0,81	0,499	0,401	0,1610	44,58
2	3,1	1,131	1,2	1,358	1,280	1,44	1,508	-0,308	0,0947	25,64
3	5,3	1,668	1,8	3,002	2,781	3,24	2,078	-0,278	0,0772	15,43
4	7,4	2,001	2,2	4,403	4,006	4,84	2,433	-0,233	0,0541	10,57
5	9,6	2,262	2,6	5,881	5,116	6,76	2,709	-0,109	0,0119	4,20
6	11,8	2,468	2,9	7,157	6,092	8,41	2,929	-0,029	0,0008	0,99
7	14,5	2,674	3,3	8,825	7,151	10,89	3,148	0,152	0,0232	4,62
8	18,7	2,929	3,8	11,128	8,576	14,44	3,418	0,382	0,1459	10,05
Итого	71,6	15,315	18,7	41,918	35,035	50,83	18,720	-0,020	0,5688	116,08
Среднее значение	8,95	1,914	2,34	5,240	4,379	6,35	–	–	0,0711	14,51
σ	–	0,846	0,935	–	–	–	–	–	–	–
σ^2	–	0,716	0,874	–	–	–	–	–	–	–

Изобразим на графике исходные данные и линию регрессии:

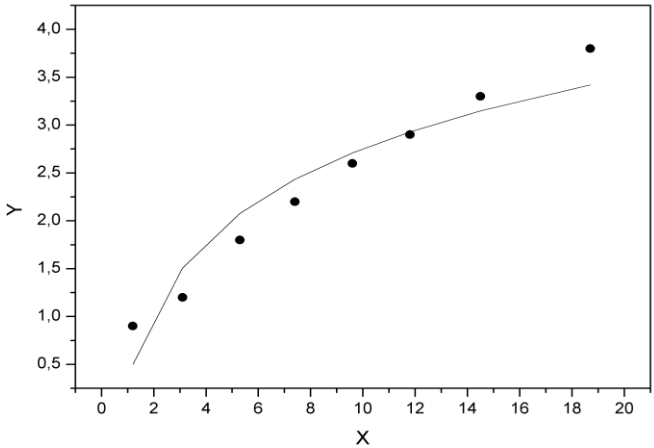


Рис. 2.6. Линия регрессии

Для нахождения параметров регрессии $y_x = a + b \cdot \sqrt{x}$ делаем замену $z = \sqrt{x}$ и составляем вспомогательную таблицу $\varepsilon = y - y_x$.

Найдем уравнение регрессии:

$$b = \frac{\operatorname{cov}(z,\, y)}{\sigma_z^2} = \frac{7{,}53 - 2{,}82 \cdot 2{,}34}{1{,}00} = 0{,}931,$$

$$a = \overline{y} - b \cdot \overline{z} = 2{,}34 - 0{,}931 \cdot 2{,}82 = -0{,}286.$$

Получаем следующее уравнение регрессии: $y_x = -0{,}286 + 0{,}931 \cdot \sqrt{x}$.

Теперь заполняем столбцы 8-11 нашей таблицы.

Индекс корреляции находим по формуле (2.21):

$$\rho_{xy} = \sqrt{1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2}} = \sqrt{1 - \frac{0{,}0077}{0{,}874}} = 0{,}996,$$

Индекс детерминации $\rho^2 = 0{,}991$, который показывает, что 99,1% вариации результативного признака объясняется вариацией признака-фактора, а 0,9% приходится на долю прочих факторов.

Средняя ошибка аппроксимации: $\overline{A} = 0{,}0498 \cdot 100\% = 4{,}98\%$ показывает, что линия регрессии хорошо приближает исходные данные.

Таблица 2.6

Расчетная таблица

	x	z	y	$z \cdot y$	z^2	y^2	y_x	ε	ε^2	A_i
1	1,2	1,10	0,9	0,99	1,2	0,81	0,734	0,166	0,0276	18,46
2	3,1	1,76	1,2	2,11	3,1	1,44	1,353	-0,153	0,0235	12,77
3	5,3	2,30	1,8	4,14	5,3	3,24	1,857	-0,057	0,0033	3,19
4	7,4	2,72	2,2	5,98	7,4	4,84	2,247	-0,047	0,0022	2,12
5	9,6	3,10	2,6	8,06	9,6	6,76	2,599	0,001	0,0000	0,05
6	11,8	3,44	2,9	9,96	11,8	8,41	2,912	-0,012	0,0001	0,42
7	14,5	3,81	3,3	12,57	14,5	10,89	3,259	0,041	0,0017	1,20
8	18,7	4,32	3,8	16,43	18,7	14,44	3,740	0,060	0,0036	1,58
Итого	71,6	22,54	18,7	60,24	71,6	50,83	18,700	-0,001	0,0619	39,82
Среднее значение	8,95	2,82	2,34	7,53	8,95	6,35	–	–	0,0077	4,98
σ	–	1,00	0,935	–	–	–	–	–	–	–
σ^2	–	1,00	0,874	–	–	–	–	–	–	–

F -критерий Фишера:

$$F = \frac{\rho_{xy}^2}{1-\rho_{xy}^2} \cdot \frac{n-m-1}{m} = \frac{0,991}{1-0,991} \cdot \frac{8-1-1}{1} = 660,67,$$

значительно превышает табличное $F_{\text{табл}} = 5,99$.

Изобразим на графике исходные данные и линию регрессии:

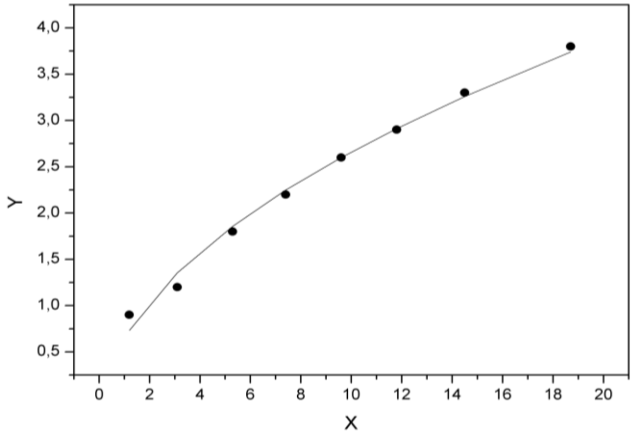


Рис. 2.7. Исходные данные и линия регрессии

Для нахождения параметров регрессии $y = a \cdot x^b \cdot \varepsilon$ необходимо провести ее линеаризацию, как было показано выше:

$$Y = A + b \cdot X + \mathrm{E},$$

где $Y = \ln y$, $X = \ln x$, $A = \ln a$, $\mathrm{E} = \ln \varepsilon$.

Составляем вспомогательную таблицу для преобразованных данных:

Найдем уравнение регрессии:

$$b = \frac{\operatorname{cov}(X,\ Y)}{\sigma_X^2} = \frac{1,830 - 1,914 \cdot 0,750}{0,716} = 0,551,$$

$$A = \overline{Y} - b \cdot \overline{X} = 0,750 - 0,551 \cdot 1,914 = -0,305.$$

Т.е. получаем следующее уравнение регрессии: $Y_x = -0,305 + 0,551 \cdot X$. После потенцирования находим искомое уравнение регрессии:

$$y_x = 0,737 \cdot x^{0,551}.$$

Таблица 2.7

Расчетная таблица

	X	Y	$X \cdot Y$	X^2	Y^2	y_x	ε	ε^2	A_i
1	0,182	-0,105	-0,019	0,033	0,011	0,8149	0,0851	0,0072	9,46
2	1,131	0,182	0,206	1,280	0,033	1,3747	-0,1747	0,0305	14,56
3	1,668	0,588	0,980	2,781	0,345	1,8473	-0,0473	0,0022	2,63
4	2,001	0,788	1,578	4,006	0,622	2,2203	-0,0203	0,0004	0,92
5	2,262	0,956	2,161	5,116	0,913	2,5627	0,0373	0,0014	1,43
6	2,468	1,065	2,628	6,092	1,134	2,8713	0,0287	0,0008	0,99
7	2,674	1,194	3,193	7,151	1,425	3,2165	0,0835	0,0070	2,53
8	2,929	1,335	3,910	8,576	1,782	3,7004	0,0996	0,0099	2,62
Итого	15,315	6,002	14,637	35,035	6,266	18,608	0,0919	0,0595	35,14
Среднее значение	1,914	0,750	1,830	4,379	0,783	–	–	0,0074	4,39
σ	0,846	0,470	–	–	–	–	–	–	–
σ^2	0,716	0,221	–	–	–	–	–	–	–

Теперь заполняем столбцы 7-10 нашей таблицы.

Индекс корреляции находим по формуле (2.21):

$$\rho_{xy} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}} = \sqrt{1 - \frac{0,0074}{0,221}} = 0,983,$$

а индекс детерминации $\rho^2 = 0,967$, который показывает, что 96,7% вариации результативного признака объясняется вариацией признака-фактора, а 3,3% приходится на долю прочих факторов.

Средняя ошибка аппроксимации: $\overline{A} = 4,39\%$ показывает, что линия регрессии хорошо приближает исходные данные.

F-критерий Фишера:

$$F = \frac{\rho^2_{xy}}{1 - \rho^2_{xy}} \cdot \frac{n - m - 1}{m} = \frac{0,967}{1 - 0,967} \cdot \frac{8 - 1 - 1}{1} = 175,82,$$

значительно превышает табличное $F_{\text{табл}} = 5,99$.

Изобразим на графике исходные данные и линию регрессии:

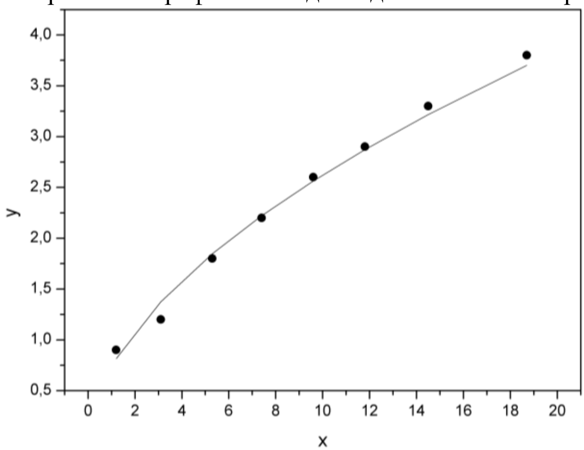


Рис. 2.8. Исходные данные и линия регрессии

Сравним построенные модели по индексу детерминации и средней ошибке аппроксимации:

Таблица 2.8

Модель	Индекс детерминации, R^2 (r_{xy}^2, ρ_{xy}^2)	Средняя ошибка аппроксимации, $\bar{A}$, %
Линейная модель, $y_x = a + b \cdot x$	0,987	6,52
Полулогарифмическая модель, $y_x = a + b \cdot \ln x$	0,918	14,51
Модель с квадратным корнем, $y_x = a + b \cdot \sqrt{x}$	0,991	4,98
Степенная модель, $y = a \cdot x^b \cdot \varepsilon$	0,967	4,39

Наиболее хорошо исходные данные аппроксимирует модель с квадратным корнем. Но в данном случае, так как индексы детерминации линейной модели и модели с квадратным корнем отличаются всего на 0,004, то вполне можно обойтись более простой линейной функцией.

Вопросы для самоконтроля

1. Построение уравнения парной регрессии: постановка задачи, спецификация модели
2. Оценка параметров линейной парной регрессии
3. Качество оценок МНК линейной регрессии. Теорема Маркова-Гаусса
4. Проверка качества уравнения парной регрессии. F-критерий Фишера
5. Коэффициенты корреляции. Оценка тесноты связи. Коэффициент эластичности
6. Точность коэффициентов парной регрессии. Проверка значимости параметров линейной парной регрессии
7. Точечный и интервальный прогноз по уравнению линейной регрессии

Тестовые задания

1. Для оценки значимости парного коэффициента корреляции используется ...

а) критерий Фишера

б) критерий Стьюдента

в) коэффициент эластичности
г) средняя ошибка аппроксимации

2. Для оценки значимости уравнения регрессии используется …
а) критерий Фишера
б) критерий Стьюдента
в) коэффициент эластичности
г) средняя ошибка аппроксимации

3. Для оценки значимости параметров уравнения регрессии используется …
а) критерий Фишера
б) критерий Стьюдента
в) коэффициент эластичности
г) средняя ошибка аппроксимации

4. Для оценки качества уравнения регрессии используется …
а) критерий Фишера
б) критерий Стьюдента
в) коэффициент эластичности
г) средняя ошибка аппроксимации

5. Коэффициент парной корреляции может принимать значения …
а) от 1 до 100
б) от 0 до 10
в) от $-\infty$ до $+\infty$
г) от – 1 до + 1
д) от – 10 до + 10
е) от 0 до + 1

6. Между y и x существует прямая связь, если
а) $a > 0$, $b > 0$
б) $a < 0$, $b > 0$
в) $a = 0$, $b > 0$
г) $a = 0$, $b < 0$
д) $a = 0$, $b = 0$
е) $a < 0$, $b < 0$

7. Между y и x существует обратная связь, если
а) $a > 0$, $b > 0$
б) $a < 0$, $b > 0$
в) $a = 0$, $b > 0$
г) $a = 0$, $b < 0$
д) $a = 0$, $b = 0$
е) $a < 0$, $b < 0$

8. Коэффициент детерминации равен …, если коэффициент корреляции составляет 0,845
а) 0,845
б) 0,919
в) 0,714
г) 1,000

9. В уравнении регрессии параметр b характеризует …
а) среднее изменение y при изменении x на одну единицу
б) среднее изменение x при изменении y на одну единицу

в) среднее значение y при $x = 0$

г) среднее значение x при $y = 0$

10. Средняя ошибка аппроксимации характеризует ...

а) качество параметров уравнения

б) качество связи между x и y

в) качество уравнения регрессии

г) значение коэффициента корреляции

д) значение коэффициента детерминации

11. Критерий Фишера применяется для проверки статистической гипотезы о ...

а) значимости параметров уравнения регрессии

б) значимости коэффициента регрессии

в) значимости всего уравнения регрессии

г) угле наклона линии регрессии

12. Зная уровень значимости, объем выборки и число объясняющих переменных, можно вычислить ...

а) величину F для критерия Фишера

б) величину t для критерия Фишера

в) среднюю ошибку аппроксимации

г) $\beta -$ коэффициент

д) средний коэффициент эластичности

13. Метод подбора вида уравнения регрессии, основанный на поле корреляции, называется ...

а) графический

б) аналитический

в) экспериментальный

3. МНОЖЕСТВЕННЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

3.1. Понятие множественного регрессионного анализа. Спецификация модели. Отбор факторов

Парная регрессия может дать хороший результат при моделировании, если влиянием других факторов, воздействующих на объект исследования, можно пренебречь. Если же этим влиянием пренебречь нельзя, то в этом случае следует попытаться выявить влияние других факторов, введя их в модель, т.е. построить уравнение множественной регрессии $y = f\left(x_1, x_2, ..., x_m\right)$,

где y – зависимая переменная (результативный признак), x_i – независимые, или объясняющие, переменные (признаки-факторы).

Множественная регрессия широко используется в решении проблем спроса, доходности акций, при изучении функции издержек производства, в макроэкономических расчетах и целом ряде других вопросов эконометрики. В настоящее время множественная регрессия – один из наиболее распространенных методов в эконометрике. Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное их воздействие на моделируемый показатель.

Построение уравнения множественной регрессии начинается с решения вопроса о спецификации модели. Он включает в себя два круга вопросов: отбор факторов и выбор вида уравнения регрессии.

Включение в уравнение множественной регрессии того или иного набора факторов связано прежде всего с представлением исследователя о природе взаимосвязи моделируемого показателя с другими экономическими явлениями. Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям.

1. Они должны быть количественно измеримы. Если необходимо включить в модель качественный фактор, не имеющий количественного измерения, то ему нужно придать количественную определенность.

2. Факторы не должны быть интеркоррелированы и тем более находиться в точной функциональной связи.

Включение в модель факторов с высокой интеркорреляцией, может привести к нежелательным последствиям – система нормальных уравнений может оказаться плохо обусловленной и повлечь за собой неустойчивость и ненадежность оценок коэффициентов регрессии.

Если между факторами существует высокая корреляция, то нельзя определить их изолированное влияние на результативный показатель и параметры уравнения регрессии оказываются неинтерпретируемыми.

Включаемые во множественную регрессию факторы должны объяснить вариацию независимой переменной. Если строится модель с набором m факторов, то для нее рассчитывается показатель детерминации R^2, который фиксирует долю объясненной вариации результативного признака за счет рассматриваемых в регрессии m факторов. Влияние других, не учтенных в модели факторов, оценивается как $1-R^2$ с соответствующей остаточной дисперсией S^2.

При дополнительном включении в регрессию $m+1$ фактора коэффициент детерминации должен возрастать, а остаточная дисперсия уменьшаться:

$$R_{m+1}^2 \geq R_m^2 \qquad \text{и} \qquad S_{m+1}^2 \leq S_m^2 .$$

Если же этого не происходит и данные показатели практически не отличаются друг от друга, то включаемый в анализ фактор x_{m+1} не улучшает модель и практически является лишним фактором.

Насыщение модели лишними факторами не только не снижает величину остаточной дисперсии и не увеличивает показатель детерминации, но и приводит к статистической незначимости параметров регрессии по критерию Стьюдента.

Таким образом, хотя теоретически регрессионная модель позволяет учесть любое число факторов, практически в этом нет необходимости. Отбор факторов производится на основе качественного теоретико-экономического анализа. Однако теоретический анализ часто не позволяет однозначно ответить на вопрос о количественной взаимосвязи рассматриваемых признаков и целесообразности включения фактора в модель. Поэтому отбор факторов обычно осуществляется в две стадии: на первой подбираются факторы исходя из сущности проблемы; на второй – на основе матрицы показателей корреляции определяют статистики для параметров регрессии.

Коэффициенты интеркорреляции (т.е. корреляции между объясняющими переменными) позволяют исключать из модели дублирующие факторы. Считается, что две переменные явно коллинеарны, т.е. находятся между собой в линейной зависимости, если $r_{x_i x_j} \geq 0,7$. Если факторы явно коллинеарны, то они дублируют друг друга и один из них рекомендуется исключить из регрессии. Предпочтение при этом отдается не фактору, более тесно связанному с результатом, а тому фактору, который при достаточно тесной связи с результатом имеет наименьшую тесноту связи с другими факторами. В этом требовании проявляется специфика множественной регрессии как метода исследования комплексного воздействия факторов в условиях их независимости друг от друга.

Пусть, например, при изучении зависимости $y = f(x_1, x_2, x_3)$ матрица парных коэффициентов корреляции оказалась следующей:

Таблица 3.1

	y	x_1	x_2	x_3
y	1	0,8	0,7	0,6
x_1	0,8	1	0,8	0,5
x_2	0,7	0,8	1	0,2
x_3	0,6	0,5	0,2	1

Очевидно, что факторы x_1 и x_2 дублируют друг друга. В анализ целесообразно включить фактор x_2, а не x_1, хотя корреляция x_2 с результатом y слабее, чем корреляция фактора x_1 с y $\left(r_{yx_2} = 0,7 < r_{yx_1} = 0,8\right)$, но зато значительно слабее межфакторная корреляция $r_{x_2x_3} = 0,2 < r_{x_1x_3} = 0,5$. Поэтому в данном случае в уравнение множественной регрессии включаются факторы x_2, x_3.

По величине парных коэффициентов корреляции обнаруживается лишь явная коллинеарность факторов. Наибольшие трудности в использовании аппарата множественной регрессии возникают при наличии мультиколлинеарности факторов, когда более чем два фактора связаны между собой линейной зависимостью, т.е. имеет место совокупное воздействие факторов друг на друга. Наличие мультиколлинеарности факторов может означать, что некоторые факторы будут всегда действовать в унисон. В результате вариация в исходных данных перестает быть полностью независимой и нельзя оценить воздействие каждого фактора в отдельности.

Включение в модель мультиколлинеарных факторов нежелательно в силу следующих последствий:

1. Затрудняется интерпретация параметров множественной регрессии как характеристик действия факторов в «чистом» виде, ибо факторы коррелированы; параметры линейной регрессии теряют экономический смысл.

2. Оценки параметров ненадежны, обнаруживают большие стандартные ошибки и меняются с изменением объема наблюдений (не только по величине, но и по знаку), что делает модель непригодной для анализа и прогнозирования.

Для оценки мультиколлинеарности факторов может использоваться определитель матрицы парных коэффициентов корреляции между факторами.

Если бы факторы не коррелировали между собой, то матрица парных коэффициентов корреляции между факторами была бы единичной матрицей, поскольку все недиагональные элементы $r_{x_i x_j}$ $(i \neq j)$ были бы равны нулю. Так, для уравнения, включающего три объясняющих переменных

$$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3$$

Матрица коэффициентов корреляции между факторами имела бы определитель, равный единице:

$$\operatorname{Det} \mathbf{R} = \begin{vmatrix} r_{x_1x_1} & r_{x_1x_2} & r_{x_1x_3} \\ r_{x_2x_1} & r_{x_2x_2} & r_{x_2x_3} \\ r_{x_3x_1} & r_{x_3x_2} & r_{x_3x_3} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1 .$$

Если же, наоборот, между факторами существует полная линейная зависимость и все коэффициенты корреляции равны единице, то определитель такой матрицы равен нулю:

$$\operatorname{Det} \mathbf{R} = \begin{vmatrix} r_{x_1x_1} & r_{x_1x_2} & r_{x_1x_3} \\ r_{x_2x_1} & r_{x_2x_2} & r_{x_2x_3} \\ r_{x_3x_1} & r_{x_3x_2} & r_{x_3x_3} \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0 .$$

Чем ближе к нулю определитель матрицы межфакторной корреляции, тем сильнее мультиколлинеарность факторов и ненадежнее результаты множественной регрессии. И, наоборот, чем ближе к единице определитель матрицы межфакторной корреляции, тем меньше мультиколлинеарность факторов.

Существует ряд подходов преодоления сильной межфакторной корреляции. Самый простой путь устранения мультиколлинеарности состоит в исключении из модели одного или нескольких факторов. Другой подход связан с преобразованием факторов, при котором уменьшается корреляция между ними.

Одним из путей учета внутренней корреляции факторов является переход к совмещенным уравнениям регрессии, т.е. к уравнениям, которые отражают не только влияние факторов, но и их взаимодействие. Так, если $y = f(x_1, x_2, x_3)$, то возможно построение следующего совмещенного уравнения:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + \varepsilon \cdot$$

Рассматриваемое уравнение включает взаимодействие первого порядка (взаимодействие двух факторов). Возможно включение в модель и взаимодействий более высокого порядка, если будет доказана их статистическая значимость по F-критерию Фишера, но, как правило, взаимодействия третьего и более высоких порядков оказываются статистически незначимыми.

Отбор факторов, включаемых в регрессию, является одним из важнейших этапов практического использования методов регрессии. Подходы к отбору факторов на основе показателей корреляции могут быть разные. Они приводят построение уравнения множественной регрессии соответственно к разным методикам. В зависимости от того, какая методика построения уравнения регрессии принята, меняется алгоритм ее решения на ЭВМ.

Наиболее широкое применение получили следующие методы построения уравнения множественной регрессии:

1. Метод исключения – отсев факторов из полного его набора.
2. Метод включения – дополнительное введение фактора.
3. Шаговый регрессионный анализ – исключение ранее введенного фактора.

При отборе факторов также рекомендуется пользоваться следующим правилом: число включаемых факторов обычно в 6–7 раз меньше объема совокупности, по которой строится регрессия. Если это соотношение нарушено, то число степеней свободы остаточной дисперсии очень мало. Это приводит к тому, что параметры уравнения регрессии оказываются статистически незначимыми, а F-критерий меньше табличного значения.

3.2. Метод наименьших квадратов (МНК). Свойства оценок на основе МНК

Возможны разные виды уравнений множественной регрессии: линейные и нелинейные.

Ввиду четкой интерпретации параметров наиболее широко используется линейная функция. В линейной множественной регрессии $y_x = a + b_1x_1 + b_2x_2 + ... + b_mx_m$ параметры при x называются коэффициентами «чистой» регрессии. Они характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизмененном значении других факторов, закрепленных на среднем уровне.

Рассмотрим линейную модель множественной регрессии

$$y = a + b_1x_1 + b_2x_2 + ... + b_mx_m + \varepsilon. \qquad (3.1)$$

Классический подход к оцениванию параметров линейной модели множественной регрессии основан на методе наименьших квадратов (МНК). МНК позволяет получить такие оценки параметров, при которых сумма квадратов отклонений фактических значений результативного признака y от расчетных y минимальна:

$$\sum_i \left(y_i - y_{x_i}\right)^2 \to \min. \qquad (3.2)$$

Как известно из курса математического анализа, для того чтобы найти экстремум функции нескольких переменных, надо вычислить частные производные первого порядка по каждому из параметров и приравнять их к нулю.

Итак. Имеем функцию $m+1$ аргумента:

$$S(a,\ b_1,\ b_2,\ ...,\ b_m) = \sum \left(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m\right)^2.$$

Находим частные производные первого порядка:

$$\begin{cases} \dfrac{\partial S}{\partial a} = -2\sum \left(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m\right) = 0; \\ \dfrac{\partial S}{\partial b_1} = -2\sum x_1\left(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m\right) = 0; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \dfrac{\partial S}{\partial b_m} = -2\sum x_m\left(y - a - b_1x_1 - b_2x_2 - ... - b_mx_m\right) = 0. \end{cases}$$

После элементарных преобразований приходим к системе линейных нормальных уравнений для нахождения параметров линейного уравнения множественной регрессии (2.1):

$$\begin{cases} na + b_1\sum x_1 + b_2\sum x_2 + \ldots + b_m\sum x_m = \sum y, \\ a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 + \ldots + b_m\sum x_1x_m = \sum yx_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a\sum x_m + b_1\sum x_1x_m + b_2\sum x_2x_m + \ldots + b_m\sum x_m^2 = \sum yx_m. \end{cases} \quad (3.3)$$

Для двухфакторной модели данная система будет иметь вид:

$$\begin{cases} na + b_1\sum x_1 + b_2\sum x_2 = \sum y, \\ a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 = \sum yx_1, \\ a\sum x_2 + b_1\sum x_1x_2 + b_2\sum x_2^2 = \sum yx_2. \end{cases}$$

Метод наименьших квадратов применим и к уравнению множественной регрессии в стандартизированном масштабе:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \ldots + \beta_m t_{x_m} + \varepsilon, \quad (3.4)$$

где $t_y,\ t_{x_1},\ \ldots,\ t_{x_m}$ – стандартизированные переменные: $t_y = \dfrac{y - \bar{y}}{\sigma_y}$, $t_{x_i} = \dfrac{x_i - \bar{x}_i}{\sigma_{x_i}}$, для которых среднее значение равно нулю: $\bar{t}_y = \bar{t}_{x_i} = 0$, а среднее квадратическое отклонение равно единице: $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$; β_i – стандартизированные коэффициенты регрессии.

Стандартизованные коэффициенты регрессии показывают, на сколько единиц изменится в среднем результат, если соответствующий фактор x_i изменится на одну единицу при неизменном среднем уровне других факторов. В силу того, что все переменные заданы как центрированные и нормированные, стандартизованные коэффициенты регрессии β_i можно сравнивать между собой. Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат. В этом основное достоинство стандартизованных коэффициентов регрессии в отличие от коэффициентов «чистой» регрессии, которые несравнимы между собой.

Применяя МНК к уравнению множественной регрессии в стандартизированном масштабе, получим систему нормальных уравнений вида

$$\begin{cases} r_{yx_1} = \beta_1 \qquad + \beta_2 r_{x_1x_2} + \beta_3 r_{x_1x_3} + \ldots + \beta_m r_{x_1x_m}, \\ r_{yx_2} = \beta_1 r_{x_1x_2} + \beta_2 \qquad + \beta_3 r_{x_1x_3} + \ldots + \beta_m r_{x_1x_m}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ r_{yx_m} = \beta_1 r_{x_1x_m} + \beta_2 r_{x_2x_m} + \beta_3 r_{x_3x_m} + \ldots + \beta_m, \end{cases}$$

(2.5)

где r_{yx_i} и $r_{x_ix_j}$ – коэффициенты парной и межфакторной корреляции.

Коэффициенты «чистой» регрессии b_i связаны со стандартизованными коэффициентами регрессии β_i следующим образом:

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}. \qquad (3.6)$$

Поэтому можно переходить от уравнения регрессии в стандартизованном масштабе (2.4) к уравнению регрессии в натуральном масштабе переменных (2.1), при этом параметр a определяется как $a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 - \ldots - b_m\bar{x}_m$.

Рассмотренный смысл стандартизованных коэффициентов регрессии позволяет их использовать при отсеве факторов – из модели исключаются факторы с наименьшим значением β_i.

На основе линейного уравнения множественной регрессии

$$y = a + b_1x_1 + b_2x_2 + \ldots + b_mx_m + \varepsilon \qquad (3.7)$$

могут быть найдены частные уравнения регрессии:

$$\begin{cases} y_{x_1 \cdot x_2, x_3, \ldots, x_m} = f(x_1), \\ y_{x_2 \cdot x_1, x_3, \ldots, x_m} = f(x_2), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_{x_m \cdot x_1, x_2, \ldots, x_{m-1}} = f(x_m), \end{cases} \qquad (3.8)$$

т.е. уравнения регрессии, которые связывают результативный признак с соответствующим фактором x_i при закреплении остальных факторов на среднем уровне. В развернутом виде систему (2.8) можно переписать в виде:

$$\begin{cases} y_{x_1 \cdot x_2, x_3, \ldots, x_m} = a + b_1 x_1 + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \ldots + b_m \bar{x}_m + \varepsilon, \\ y_{x_2 \cdot x_1, x_3, \ldots, x_m} = a + b_1 \bar{x}_1 + b_2 x_2 + b_3 \bar{x}_3 + \ldots + b_m \bar{x}_m + \varepsilon, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_{x_m \cdot x_1, x_2, \ldots, x_{m-1}} = a + b_1 \bar{x}_1 + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \ldots + b_m x_m + \varepsilon. \end{cases}$$

При подстановке в эти уравнения средних значений соответствующих факторов они принимают вид парных уравнений линейной регрессии, т.е. имеем

$$\begin{cases} y_{x_1 \cdot x_2, x_3, \ldots, x_m} = A_1 + b_1 x_1, \\ y_{x_2 \cdot x_1, x_3, \ldots, x_m} = A_2 + b_2 x_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_{x_m \cdot x_1, x_2, \ldots, x_{m-1}} = A_m + b_m x_m, \end{cases} \qquad (3.9)$$

где

$$\begin{cases} A_1 = a + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \ldots + b_m \bar{x}_m, \\ A_2 = a + b_1 \bar{x}_1 + b_3 \bar{x}_3 + \ldots + b_m \bar{x}_m, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ A_m = a + b_1 \bar{x}_1 + b_2 \bar{x}_2 + b_3 \bar{x}_3 + \ldots + b_{m-1} x_{m-1}. \end{cases}$$

В отличие от парной регрессии частные уравнения регрессии характеризуют изолированное влияние фактора на результат, ибо другие факторы закреплены на неизменном уровне. Эффекты влияния других факторов присоединены в них к свободному члену уравнения множественной регрессии. Это позволяет на основе частных уравнений регрессии определять частные коэффициенты эластичности:

$$Э_{y_{x_i}} = b_i \cdot \frac{x_i}{y_{x_i \cdot x_1, x_2, \ldots x_{i-1}, x_{i+1}, \ldots, x_m}}, \qquad (3.10)$$

где b_i – коэффициент регрессии для фактора x_i в уравнении множественной регрессии, $y_{x_i \cdot x_1, x_2, \ldots x_{i-1}, x_{i+1}, \ldots, x_m}$ – частное уравнение регрессии.

Наряду с частными коэффициентами эластичности могут быть найдены средние по совокупности показатели эластичности:

$$\bar{Э}_i = b_i \cdot \frac{\bar{x}_i}{\bar{y}_{x_i}}, \qquad (2.11)$$

которые показывают на сколько процентов в среднем изменится результат, при изменении соответствующего фактора на 1%. Средние показатели эластичности можно сравнивать друг с другом и соответственно ранжировать факторы по силе их воздействия на результат.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ. Пусть имеются следующие данные (условные) о сменной добыче угля на одного рабочего y (т), мощности пласта x_1 (м) и уровне механизации работ x_2 (%), характеризующие процесс добычи угля в 10 шахтах.

Таблица 3.2

№	1	2	3	4	5	6	7	8	9	10
x_1	8	11	12	9	8	8	9	9	8	12
x_2	5	8	8	5	7	8	6	4	5	7
y	5	10	10	7	5	6	6	5	6	8

Предполагая, что между переменными y, x_1, x_2 существует линейная корреляционная зависимость, найдем уравнение регрессии y по x_1 и x_2.

Для удобства дальнейших вычислений составляем таблицу $\varepsilon = y - y_x$ (табл. 2.3).

Для нахождения параметров уравнения регрессии в данном случае необходимо решить следующую систему нормальных уравнений:

$$\begin{cases} 10a + 94b_1 + 63b_2 = 68, \\ 94a + 908b_1 + 603b_2 = 664, \\ 63a + 603b_1 + 417b_2 = 445. \end{cases}$$

Откуда получаем, что $a = -3{,}54$, $b_1 = 0{,}854$, $b_2 = 0{,}367$. Т.е. получили следующее уравнение множественной регрессии:

$$y_x = -3{,}54 + 0{,}854 \cdot x_1 + 0{,}367 \cdot x_2 \cdot$$

Оно показывает, что при увеличении только мощности пласта x_1 (при неизменном x_2) на 1 м добыча угля на одного рабочего y увеличится в среднем на 0,854 т, а при увеличении только уровня механизации работ x_2 (при неизменном x_1) на 1% – в среднем на 0,367 т.

Найдем уравнение множественной регрессии в стандартизованном масштабе:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \varepsilon,$$

при этом стандартизованные коэффициенты регрессии будут

$$\beta_1 = b_1 \frac{\sigma_{x_1}}{\sigma_y} = 0,854 \cdot \frac{1,56}{1,83} = 0,728,$$

$$\beta_2 = b_2 \frac{\sigma_{x_2}}{\sigma_y} = 0,367 \cdot \frac{1,42}{1,83} = 0,285.$$

Т.е. уравнение будет выглядеть следующим образом:

$$\hat{t}_y = 0,728 \cdot t_{x_1} + 0,285 \cdot t_{x_2}.$$

Так как стандартизованные коэффициенты регрессии можно сравнивать между собой, то можно сказать, что мощность пласта оказывает большее влияние на сменную добычу угля, чем уровень механизации работ.

Сравнивать влияние факторов на результат можно также при помощи средних коэффициентов эластичности (3.11):

$$\bar{Э}_i = b_i \cdot \frac{\bar{x}_i}{\bar{y}_{x_i}}.$$

Вычисляем:

$$\bar{Э}_1 = 0,854 \cdot \frac{9,4}{6,8} = 1,18, \qquad \bar{Э}_2 = 0,367 \cdot \frac{6,3}{6,8} = 0,34.$$

Т.е. увеличение только мощности пласта (от своего среднего значения) или только уровня механизации работ на 1% увеличивает в среднем сменную добычу угля на 1,18% или 0,34% соответственно. Таким образом, подтверждается большее влияние на результат y фактора x_1, чем фактора x_2.

Таблица 2.3

№	x_1	x_2	y	x_1^2	x_2^2	y^2	$x_1 \cdot x_2$	$x_1 \cdot y$	$x_2 \cdot y$	y_x	ε^2
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	8	5	5	64	25	25	40	40	25	5,13	0,016
2	11	8	10	121	64	100	88	110	80	8,79	1,464
3	12	8	10	144	64	100	96	120	80	9,64	0,127
4	9	5	7	81	25	49	45	63	35	5,98	1,038
5	8	7	5	64	49	25	56	40	35	5,86	0,741
6	8	8	6	64	64	36	64	48	48	6,23	0,052
7	9	6	6	81	36	36	54	54	36	6,35	0,121
8	9	4	5	81	16	25	36	45	20	5,61	0,377
9	8	5	6	64	25	36	40	48	30	5,13	0,762
10	12	7	8	144	49	64	84	96	56	9,28	1,631
Сумма	94	63	68	908	417	496	603	664	445	68	6,329
Среднее значение	9,4	6,3	6,8	90,8	41,7	49,6	60,3	66,4	44,5	–	–
σ^2	2,44	2,01	3,36	–	–	–	–	–	–	–	–
σ	1,56	1,42	1,83	–	–	–	–	–	–	–	–

3.3. Проверка существенности факторов и показатели качества регрессии

Практическая значимость уравнения множественной регрессии оценивается с помощью показателя множественной корреляции и его квадрата – показателя детерминации.

Показатель множественной корреляции характеризует тесноту связи рассматриваемого набора факторов с исследуемым признаком или, иначе, оценивает тесноту совместного влияния факторов на результат.

Независимо от формы связи показатель множественной корреляции может быть найден как индекс множественной корреляции:

$$R_{yx_1x_2\ldots x_m} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}}, \qquad (3.12)$$

где σ^2_y – общая дисперсия результативного признака; $\sigma^2_{\text{ост}}$ – остаточная дисперсия.

Границы изменения индекса множественной корреляции от 0 до 1. Чем ближе его значение к 1, тем теснее связь результативного признака со всем набором исследуемых факторов. Величина индекса множественной корреляции должна быть больше или равна максимальному парному индексу корреляции:

$$R_{yx_1x_2\ldots x_m} \geq r_{yx_i(\max)} \quad \left(i = \overline{1,m}\right).$$

При правильном включении факторов в регрессионную модель величина индекса множественной корреляции будет существенно отличаться от индекса корреляции парной зависимости. Если же дополнительно включенные в уравнение множественной регрессии факторы третьестепенны, то индекс множественной корреляции может практически совпадать с индексом парной корреляции (различия в третьем, четвертом знаках). Отсюда ясно, что сравнивая индексы множественной и парной корреляции, можно сделать вывод о целесообразности включения в уравнение регрессии того или иного фактора.

Расчет индекса множественной корреляции предполагает определение уравнения множественной регрессии и на его основе остаточной дисперсии:

$$\sigma^2_{\text{ост}} = \frac{1}{n}\sum\left(y - y_{x_1x_2\ldots x_m}\right)^2. \qquad (3.13)$$

Можно пользоваться следующей формулой индекса множественной детерминации:

$$R^2_{yx_1x_2...x_m} = 1 - \frac{\sum\left(y - y_{x_1x_2...x_m}\right)^2}{\sum\left(y - \bar{y}\right)^2}. \qquad (3.14)$$

При линейной зависимости признаков формула индекса множественной корреляции может быть представлена следующим выражением:

$$R_{yx_1x_2...x_m} = \sqrt{\sum \beta_i \cdot r_{yx_i}}, \qquad (3.15)$$

где β_i – стандартизованные коэффициенты регрессии; r_{yx_i} – парные коэффициенты корреляции результата с каждым фактором.

Формула индекса множественной корреляции для линейной регрессии получила название *линейного коэффициента множественной корреляции*, или, что то же самое, *совокупного коэффициента корреляции*.

Возможно также при линейной зависимости определение совокупного коэффициента корреляции через матрицу парных коэффициентов корреляции:

$$R_{yx_1x_2,...,x_p} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}}, \qquad (3.16)$$

где

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & ... & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1x_2} & ... & r_{x_1x_p} \\ r_{yx_2} & r_{x_2x_1} & 1 & ... & r_{x_2x_p} \\ ... & ... & ... & ... & ... \\ r_{yx_p} & r_{x_px_1} & r_{x_px_2} & ... & 1 \end{vmatrix}$$

– определитель матрицы парных коэффициентов корреляции;

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & ... & r_{x_1x_p} \\ r_{x_2x_1} & 1 & ... & r_{x_2x_p} \\ ... & ... & ... & ... \\ r_{x_px_1} & r_{x_px_2} & ... & 1 \end{vmatrix}$$

– определитель матрицы межфакторной корреляции.

Как видим, величина множественного коэффициента корреляции зависит не только от корреляции результата с каждым из факторов, но и от межфакторной корреляции. Рассмотренная формула позволяет определять совокупный коэффициент корреляции, не обращаясь при этом к уравнению множественной регрессии, а используя лишь парные коэффициенты корреляции.

В рассмотренных показателях множественной корреляции (индекс и коэффициент) используется остаточная дисперсия, которая имеет систематическую ошибку в сторону преуменьшения, тем более значительную, чем больше параметров определяется в уравнении регрессии при заданном объеме наблюдений n. Если число параметров при x_i равно m и приближается к объему наблюдений, то остаточная дисперсия будет близка к нулю и коэффициент (индекс) корреляции приблизится к единице даже при слабой связи факторов с результатом. Для того чтобы не допустить возможного преувеличения тесноты связи, используется скорректированный индекс (коэффициент) множественной корреляции.

Скорректированный индекс множественной корреляции содержит поправку на число степеней свободы, а именно остаточная сумма квадратов $\sum\left(y - y_{x_1 x_2 \dots x_m}\right)^2$ делится на число степеней свободы остаточной вариации $(n-m-1)$, а общая сумма квадратов отклонений $\sum\left(y-\bar{y}\right)^2$ на число степеней свободы в целом по совокупности $(n-1)\cdot$

Формула скорректированного индекса множественной детерминации имеет вид:

$$R^{2} = 1 - \frac{\sum\left(y-y\right)^2 \Big/ (n-m-1)}{\sum\left(y-\bar{y}\right) \big/ (n-1)}, \qquad (3.17)$$

где m – число параметров при переменных x; n – число наблюдений.

Поскольку $\dfrac{\sum\left(y - y_{x_1 x_2 \dots x_m}\right)^2}{\sum\left(y-\bar{y}\right)^2} = 1 - R^2$, то величину скорректированного индекса детерминации можно представить в виде:

$$R^{2} = 1 - \left(1 - R^2\right) \cdot \frac{n-1}{n-m-1}. \qquad (3.17а)$$

Чем больше величина m, тем сильнее различия R^{2} и R^2.

Как было показано выше, ранжирование факторов, участвующих во множественной линейной регрессии, может быть проведено через стандартизованные коэффициенты регрессии (β-коэффициенты). Эта же цель может быть достигнута с помощью частных коэффициентов корреляции (для линейных связей). Кроме того, частные показатели корреляции широко используются при решении проблемы отбора факторов: целесообразность включения того или иного фактора в модель можно доказать величиной показателя частной корреляции.

Частные коэффициенты корреляции характеризуют тесноту связи между результатом и соответствующим фактором при элиминировании (устранении влияния) других факторов, включенных в уравнение регрессии.

Показатели частной корреляции представляют собой отношение сокращения остаточной дисперсии за счет дополнительного включения в анализ нового фактора к остаточной дисперсии, имевшей место до введения его в модель.

В общем виде при наличии m факторов для уравнения

$y = a + b_1x_1 + b_2x_2 ... + b_mx_m + \varepsilon$

коэффициент частной корреляции, измеряющий влияние на y фактора x_i, при неизменном уровне других факторов, можно определить по формуле:

$$r_{yx_i \cdot x_1x_2...x_{i-1}x_{i+1}...x_m} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2...x_i...x_m}}{1 - R^2_{yx_1x_2...x_{i-1}x_{i+1}...x_m}}}, \qquad (3.18)$$

где $R^2_{yx_1x_2...x_i...x_m}$ – множественный коэффициент детерминации всех m факторов с результатом; $R^2_{yx_1x_2...x_{i-1}x_{i+1}...x_m}$ – тот же показатель детерминации, но без введения в модель фактора x_i.

При двух факторах формула (2.18) примет вид:

$$r_{yx_1 \cdot x_2} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_2}}}; \qquad r_{yx_2 \cdot x_1} = \sqrt{1 - \frac{1 - R^2_{yx_1x_2}}{1 - r^2_{yx_1}}}. \qquad (3.18а)$$

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. Например, $r_{yx_1 \cdot x_2}$ – коэффициент частной корреляции первого порядка. Соответственно коэффициенты парной корреляции называются коэффициентами нулевого порядка. Коэффициенты частной корреляции более высоких порядков можно определить через коэффициенты частной корреляции более низких порядков по рекуррентной формуле:

$$r_{yx_i \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_m} = \frac{r_{yx_i \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_{m-1}} - r_{yx_m \cdot x_1 x_2 \ldots x_{m-1}} \cdot r_{x_i x_m \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_{m-1}}}{\sqrt{\left(1 - r^2_{yx_m \cdot x_1 x_2 \ldots x_{m-1}}\right) \cdot \left(1 - r^2_{x_i x_m \cdot x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_{m-1}}\right)}}$$
(3.19).

При двух факторах данная формула примет вид:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{\left(1 - r^2_{yx_2}\right) \cdot \left(1 - r^2_{x_1 x_2}\right)}}$$

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{\left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{x_1 x_2}\right)}} \qquad (3.19а)$$

Для уравнения регрессии с тремя факторами частные коэффициенты корреляции второго порядка определяются на основе частных коэффициентов корреляции первого порядка. Так, по уравнению $y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + \varepsilon$ возможно исчисление трех частных коэффициентов корреляции второго порядка:

$$r_{yx_1 \cdot x_2 x_3},\ r_{yx_2 \cdot x_1 x_3},\ r_{yx_3 \cdot x_1 x_2},$$

каждый из которых определяется по рекуррентной формуле. Например, при $i = 1$ имеем формулу для расчета $r_{yx_1 \cdot x_2 x_3}$:

$$r_{yx_1 \cdot x_2 x_3} = \frac{r_{yx_1 \cdot x_2} - r_{yx_3 \cdot x_2} \cdot r_{x_1 x_3 \cdot x_2}}{\sqrt{\left(1 - r^2_{yx_3 \cdot x_2}\right)\left(1 - r^2_{x_1 x_3 \cdot x_2}\right)}}. \qquad (3.20)$$

Рассчитанные по рекуррентной формуле частные коэффициенты корреляции изменяются в пределах от –1 до +1, а по формулам через множественные коэффициенты детерминации – от 0 до 1. Сравнение их друг с другом позволяет ранжировать факторы по тесноте их связи с результатом. Частные коэффициенты корреляции дают меру тесноты связи каждого фактора с результатом в чистом виде. Если из стандартизованного уравнения регрессии $t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \beta_3 t_{x_3} + \varepsilon$ следует, что $\beta_1 > \beta_2 > \beta_3$, т.е. по силе влияния на результат порядок факторов таков: x_1, x_2, x_3, то этот же порядок факторов определяется и по соотношению частных коэффициентов корреляции, $r_{yx_1 \cdot x_2 x_3} > r_{yx_2 \cdot x_1 x_3} > r_{yx_3 \cdot x_1 x_2}$.

В эконометрике частные коэффициенты корреляции обычно не имеют самостоятельного значения. Их используют на стадии формирования модели. Так, строя многофакторную модель, на первом шаге определяется уравнение регрессии с полным набором факторов и рассчитывается матрица частных коэффициентов корреляции. На втором шаге отбирается фактор с наименьшей и несущественной по t-критерию Стьюдента величиной показателя частной корреляции. Исключив его из модели, строится новое уравнение регрессии. Процедура продолжается до тех пор, пока не окажется, что все частные коэффициенты корреляции существенно отличаются от нуля. Если исключен несущественный фактор, то множественные коэффициенты детерминации на двух смежных шагах построения регрессионной модели почти не отличаются друг от друга, $R^2_{m+1} \approx R^2_m$, где m – число факторов.

Из приведенных выше формул частных коэффициентов корреляции видна связь этих показателей с совокупным коэффициентом корреляции. Зная частные коэффициенты корреляции (последовательно первого, второго и более высокого порядка), можно определить совокупный коэффициент корреляции по формуле:

$$R_{yx_1x_2\ldots x_m} = \sqrt{1-\left(1-r^2_{yx_1}\right)\cdot\left(1-r^2_{yx_2\cdot x_1}\right)\cdot\left(1-r^2_{yx_3\cdot x_1x_2}\right)\cdot\ldots\cdot\left(1-r^2_{yx_m\cdot x_1x_2\ldots x_{m-1}}\right)}\,. \quad (3.21)$$

В частности, для двухфакторного уравнения формула (3.21) принимает вид:

$$R_{yx_1x_2\ldots x_m} = \sqrt{1-\left(1-r^2_{yx_1}\right)\cdot\left(1-r^2_{yx_2\cdot x_1}\right)}\,. \quad (3.21а)$$

При полной зависимости результативного признака от исследуемых факторов коэффициент совокупного их влияния равен единице. Из единицы вычитается доля остаточной вариации результативного признака $\left(1-r^2\right)$, обусловленная последовательно включенными в анализ факторами. В результате подкоренное выражение характеризует совокупное действие всех исследуемых факторов.

Значимость уравнения множественной регрессии в целом, так же как и в парной регрессии, оценивается с помощью F-критерия Фишера:

$$F = \frac{S_{\text{факт}}}{S_{\text{ост}}} = \frac{R^2}{1-R^2}\cdot\frac{n-m-1}{m}, \quad (3.22)$$

где $S_{\text{факт}}$ – факторная сумма квадратов на одну степень свободы; $S_{\text{ост}}$ – остаточная сумма квадратов на одну степень свободы; R^2 – коэффициент (индекс) множественной детерминации; m – число параметров при переменных x (в линейной регрессии совпадает с числом включенных в модель факторов); n – число наблюдений.

Оценивается значимость не только уравнения в целом, но и фактора, дополнительно включенного в регрессионную модель. Необходимость такой оценки связана с тем, что не каждый фактор, вошедший в модель, может существенно увеличивать долю объясненной вариации результативного признака. Кроме того, при наличии в модели нескольких факторов они могут вводиться в модель в разной последовательности. Ввиду корреляции между факторами значимость одного и того же фактора может быть разной в зависимости от последовательности его введения в модель. Мерой для оценки включения фактора в модель служит частный F-критерий, т.е. F_{x_i}.

Частный F-критерий построен на сравнении прироста факторной дисперсии, обусловленного влиянием дополнительно включенного фактора, с остаточной дисперсией на одну степень свободы по регрессионной модели в целом. В общем виде для фактора x_i частный F-критерий определится как

$$F_{x_i} = \frac{R^2_{yx_1 \ldots x_i \ldots x_m} - R^2_{yx_1 \ldots x_{i-1} x_{i+1} \ldots x_m}}{1 - R^2_{yx_1 \ldots x_i \ldots x_m}} \cdot \frac{n - m - 1}{1}, \quad (3.23)$$

где $R^2_{yx_1 \ldots x_i \ldots x_m}$ – коэффициент множественной детерминации для модели с полным набором факторов, $R^2_{yx_1 \ldots x_{i-1} x_{i+1} \ldots x_m}$ – тот же показатель, но без включения в модель фактора x_i, n – число наблюдений, m – число параметров в модели (без свободного члена).

Фактическое значение частного F-критерия сравнивается с табличным при уровне значимости α и числе степеней свободы: 1 и $n-m-1$. Если фактическое значение F_{x_i} превышает $F_{\text{табл}}(\alpha,\, k_1,\, k_2)$, то дополнительное включение фактора x_i в модель статистически оправданно и коэффициент чистой регрессии b_i при факторе x_i статистически значим. Если же фактическое значение F_{x_i} меньше табличного, то дополнительное включение в модель фактора x_i не увеличивает существенно долю объясненной вариации признака y, следовательно, нецелесообразно его включение в модель; коэффициент регрессии при данном факторе в этом случае статистически незначим.

Для двухфакторного уравнения частные F-критерии имеют вид:

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot (n-3), \quad F_{x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot (n-3) \quad (3.23a)$$

С помощью частного F-критерия можно проверить значимость всех коэффициентов регрессии в предположении, что каждый соответствующий фактор x_i вводился в уравнение множественной регрессии последним.

Частный F-критерий оценивает значимость коэффициентов чистой регрессии. Зная величину F_{x_i}, можно определить и t-критерий для коэффициента регрессии при i-м факторе, t_{b_i}, а именно:

$$t_{b_i} = \sqrt{F_{x_i}}. \quad (3.24)$$

Оценка значимости коэффициентов чистой регрессии по t-критерию Стьюдента может быть проведена и без расчета частных F-критериев. В этом случае, как и в парной регрессии, для каждого фактора используется формула:

$$t_{b_i} = \frac{b_i}{m_{b_i}}, \quad (3.25)$$

где b_i – коэффициент чистой регрессии при факторе x_i, m_{b_i} – средняя квадратическая (стандартная) ошибка коэффициента регрессии b_i.

Для уравнения множественной регрессии $y = a + b_1 x_1 + b_2 x_2 + ... + b_m x_m$ средняя квадратическая ошибка коэффициента регрессии может быть определена по следующей формуле:

$$m_{b_i} = \frac{\sigma_y \sqrt{1 - R^2_{yx_1...x_m}}}{\sigma_{x_i} \sqrt{1 - R^2_{x_i x_1...x_m}}} \cdot \frac{1}{\sqrt{n-m-1}}, \quad (3.26)$$

где σ_y – среднее квадратическое отклонение для признака y, σ_{x_i} – среднее квадратическое отклонение для признака x_i, $R^2_{yx_1...x_m}$ – коэффициент детерминации для уравнения множественной регрессии, $R^2_{x_i x_1...x_m}$ – коэффициент детерминации для зависимости фактора x_i со всеми другими факторами уравнения множественной регрессии; $n-m-1$ – число степеней свободы для остаточной суммы квадратов отклонений.

Как видим, чтобы воспользоваться данной формулой, необходимы матрица межфакторной корреляции и расчет по ней соответствующих коэффициентов детерминации $R^2_{x_i x_1...x_m}$. Так, для уравнения $y = a + b_1 x_1 + b_2 x_2 + b_3 x_3$ оценка значимости коэффициентов регрессии

b_1, b_2, b_3 предполагает расчет трех межфакторных коэффициентов детерминации: $R^2_{x_1 \cdot x_2 x_3}$, $R^2_{x_2 \cdot x_1 x_3}$, $R^2_{x_3 \cdot x_1 x_2}$.

Взаимосвязь показателей частного коэффициента корреляции, частного F-критерия и t-критерия Стьюдента для коэффициентов чистой регрессии может использоваться в процедуре отбора факторов. Отсев факторов при построении уравнения регрессии методом исключения практически можно осуществлять не только по частным коэффициентам корреляции, исключая на каждом шаге фактор с наименьшим незначимым значением частного коэффициента корреляции, но и по величинам t_{b_i} и F_{x_i}. Частный F-критерий широко используется и при построении модели методом включения переменных и шаговым регрессионным методом.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ. Оценим качество уравнения, полученного в предыдущем параграфе. Сначала найдем значения парных коэффициентов корреляции:

$$r_{yx_1} = \frac{\overline{y \cdot x_1} - \bar{y} \cdot \bar{x}_1}{\sigma_y \cdot \sigma_{x_1}} = \frac{66,4 - 6,8 \cdot 9,4}{1,83 \cdot 1,56} = 0,869;$$

$$r_{yx_2} = \frac{\overline{y \cdot x_2} - \bar{y} \cdot \bar{x}_2}{\sigma_y \cdot \sigma_{x_2}} = \frac{44,5 - 6,8 \cdot 6,3}{1,83 \cdot 1,42} = 0,639;$$

$$r_{x_1 x_2} = \frac{\overline{x_1 \cdot x_2} - \bar{x}_1 \cdot \bar{x}_2}{\sigma_{x_1} \cdot \sigma_{x_2}} = \frac{60,3 - 9,4 \cdot 6,3}{1,56 \cdot 1,42} = 0,488.$$

Значения парных коэффициентов корреляции указывают на достаточно тесную связь сменной добычи угля на одного рабочего y с мощностью пласта x_1 и на умеренную связь с уровнем механизации работ x_2. В то же время межфакторная связь $r_{x_1 x_2}$ не очень сильная ($r_{x_1 x_2} = 0,49 < 0,7$), что говорит о том, что оба фактора являются информативными, т.е. и x_1, и x_2 необходимо включить в модель.

Теперь рассчитаем совокупный коэффициент корреляции $R_{yx_1 x_2}$ Для этого сначала найдем определитель матрицы парных коэффициентов корреляции:

$$\Delta r=\begin{vmatrix}1 & 0,87 & 0,64\\ 0,87 & 1 & 0,49\\ 0,64 & 0,49 & 1\end{vmatrix}=0,139064,$$

и определитель матрицы межфакторной корреляции:

$$\Delta r_{11}=\begin{vmatrix}1 & 0,49\\ 0,49 & 1\end{vmatrix}=0,7599.$$

Тогда коэффициент множественной корреляции по формуле (2.16):

$$R_{yx_1x_2}=\sqrt{1-\frac{\Delta r}{\Delta r_{11}}}=\sqrt{1-\frac{0,139064}{0,7599}}=0,904.$$

Т.е. можно сказать, что 81,7% (коэффициент детерминации $R^2_{yx_1x_2}=0,817$) вариации результата объясняется вариацией представленных в уравнении признаков, что указывает на весьма тесную связь признаков с результатом.

Примерно тот же результат (различия связаны с ошибками округлений) для коэффициента множественной регрессии получим, если воспользуемся формулами (2.12) и (2.15):

$$R_{yx_1x_2}=\sqrt{1-\frac{\sigma^2_{\text{ост}}}{\sigma^2_y}}=\sqrt{1-\frac{0,6329}{3,36}}=0,901;$$

$$R_{yx_1x_2}=\sqrt{\sum\beta_i\cdot r_{yx_i}}=\sqrt{0,728\cdot 0,87+0,285\cdot 0,64}=0,903.$$

Скорректированный коэффициент множественной детерминации $R=1-\left(1-R^2\right)\cdot\frac{n-1}{n-m-1}=1-\left(1-0,817\right)\cdot\frac{10-1}{10-2-1}=0,765$ указывает на умеренную связь между результатом и признаками. Это связано с малым количеством наблюдений.

Теперь найдем частные коэффициенты корреляции по формулам (2.18а) и (2.19а):

$$r_{yx_1\cdot x_2}=\sqrt{1-\frac{1-R^2_{yx_1x_2}}{1-r^2_{yx_2}}}=\sqrt{1-\frac{1-0,817}{1-0,408}}=0,831;$$

$$r_{yx_2\cdot x_1}=\sqrt{1-\frac{1-R^2_{yx_1x_2}}{1-r^2_{yx_1}}}=\sqrt{1-\frac{1-0,817}{1-0,755}}=0,503.$$

$$r_{yx_1\cdot x_2}=\frac{r_{yx_1}-r_{yx_2}\cdot r_{x_1x_2}}{\sqrt{\left(1-r_{yx_2}^2\right)\cdot\left(1-r_{x_1x_2}^2\right)}}=\frac{0,869-0,639\cdot 0,488}{\sqrt{\left(1-0,489^2\right)\left(1-0,639^2\right)}}=0,830;$$

$$r_{yx_2\cdot x_1}=\frac{r_{yx_2}-r_{yx_1}\cdot r_{x_1x_2}}{\sqrt{\left(1-r_{yx_1}^2\right)\cdot\left(1-r_{x_1x_2}^2\right)}}=\frac{0,639-0,869\cdot 0,488}{\sqrt{\left(1-0,488^2\right)\left(1-0,869^2\right)}}=0,498.$$

Т.е. можно сделать вывод, что фактор x_1 оказывает более сильное влияние на результат, чем признак x_2.

Оценим надежность уравнения регрессии в целом и показателя связи с помощью F-критерия Фишера. Фактическое значение F-критерия (2.22)

$$F_{\text{факт}}=\frac{R^2}{1-R^2}\cdot\frac{n-m-1}{m}=\frac{0,817}{1-0,817}\cdot\frac{10-2-1}{2}=15,63.$$

Табличное значение F-критерия при пятипроцентном уровне значимости ($\alpha=0,05$, $k_1=2$, $k_2=10-2-1=7$): $F_{\text{табл}}=4,74$. Так как $F_{\text{факт}}=15,63>F_{\text{табл}}=4,10$, то уравнение признается статистически значимым.

Оценим целесообразность включения фактора x_1 после фактора x_2 и x_2 после x_1 с помощью частного F-критерия Фишера (2.23а):

$$F_{x_1}=\frac{R_{yx_1x_2}^2-r_{yx_2}^2}{1-R_{yx_1x_2}^2}\cdot(n-3)=\frac{0,817-0,408}{1-0,817}\cdot 7=15,65;$$

$$F_{x_2}=\frac{R_{yx_1x_2}^2-r_{yx_1}^2}{1-R_{yx_1x_2}^2}\cdot(n-3)=\frac{0,817-0,755}{1-0,817}\cdot 7=2,37.$$

Табличное значение частного F-критерия при пятипроцентном уровне значимости ($\alpha=0,05$, $k_1=1$, $k_2=10-2-1=7$): $F_{\text{табл}}=5,59$. Так как $F_{x_1}=15,65>F_{\text{табл}}=5,59$, а $F_{x_2}=2,37<F_{\text{табл}}=5,59$, то включение фактора x_1 в модель статистически оправдано и коэффициент чистой регрессии b_1 статистически значим, а дополнительное включение фактора x_2, после того, как уже введен фактор x_1, нецелесообразно.

Уравнение регрессии, включающее только один значимый аргумент x_2: $y=-2,754+1,016x_1$.

3.4. Линейные регрессионные модели с гетероскедастичными остатками

При оценке параметров уравнения регрессии применяется метод наименьших квадратов (МНК). При этом делаются определенные предпосылки относительно случайной составляющей ε. В модели

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_m x_m + \varepsilon$$

случайная составляющая ε представляет собой ненаблюдаемую величину. После того как произведена оценка параметров модели, рассчитывая разности фактических и теоретических значений результативного признака y, можно определить оценки случайной составляющей $y - y_x$. Поскольку они не являются реальными случайными остатками, их можно считать некоторой выборочной реализацией неизвестного остатка заданного уравнения, т.е. ε_i.

При изменении спецификации модели, добавлении в нее новых наблюдений выборочные оценки остатков ε_i могут меняться. Поэтому в задачу регрессионного анализа входит не только построение самой модели, но и исследование случайных отклонений ε_i, т.е. остаточных величин.

При использовании критериев Фишера и Стьюдента делаются предположения относительно поведения остатков ε_i – остатки представляют собой независимые случайные величины и их среднее значение равно 0; они имеют одинаковую (постоянную) дисперсию и подчиняются нормальному распределению.

Статистические проверки параметров регрессии, показателей корреляции основаны на непроверяемых предпосылках распределения случайной составляющей ε_i. Они носят лишь предварительный характер. После построения уравнения регрессии проводится проверка наличия у оценок ε_i (случайных остатков) тех свойств, которые предполагались. Связано это с тем, что оценки параметров регрессии должны отвечать определенным критериям. Они должны быть несмещенными, состоятельными и эффективными. Эти свойства оценок, полученных по МНК, имеют чрезвычайно важное практическое значение в использовании результатов регрессии и корреляции.

Несмещенность оценки означает, что математическое ожидание остатков равно нулю. Если оценки обладают свойством несмещенности, то их можно сравнивать по разным исследованиям.

Оценки считаются *эффективными*, если они характеризуются наименьшей дисперсией. В практических исследованиях это означает возможность перехода от точечного оценивания к интервальному.

Состоятельность оценок характеризует увеличение их точности с увеличением объема выборки. Большой практический интерес представляют те результаты регрессии, для которых доверительный интервал ожидаемого значения параметра регрессии b_i имеет предел значений вероятности, равный единице. Иными словами, вероятность получения оценки на заданном расстоянии от истинного значения параметра близка к единице.

Указанные критерии оценок (несмещенность, состоятельность и эффективность) обязательно учитываются при разных способах оценивания. Метод наименьших квадратов строит оценки регрессии на основе минимизации суммы квадратов остатков. Поэтому очень важно исследовать поведение остаточных величин регрессии ε_i. Условия, необходимые для получения несмещенных, состоятельных и эффективных оценок, представляют собой предпосылки МНК, соблюдение которых желательно для получения достоверных результатов регрессии.

Исследования остатков ε_i предполагают проверку наличия следующих пяти предпосылок МНК:

1) случайный характер остатков;
2) нулевая средняя величина остатков, не зависящая от x_i;
3) гомоскедастичность – дисперсия каждого отклонения ε_i, одинакова для всех значений x;
4) отсутствие автокорреляции остатков – значения остатков ε_i распределены независимо друг от друга;
5) остатки подчиняются нормальному распределению.

Если распределение случайных остатков ε_i не соответствует некоторым предпосылкам МНК, то следует корректировать модель.

Прежде всего, проверяется случайный характер остатков ε_i – первая предпосылка МНК. С этой целью стоится график зависимости остатков ε_i от теоретических значений результативного признака (рис. 3.1). Если на графике получена горизонтальная полоса, то остатки ε_i представляют собой случайные величины и МНК оправдан, теоретические значения y_x хорошо аппроксимируют фактические значения y.

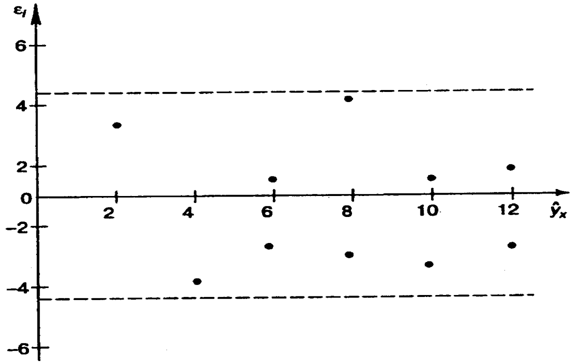
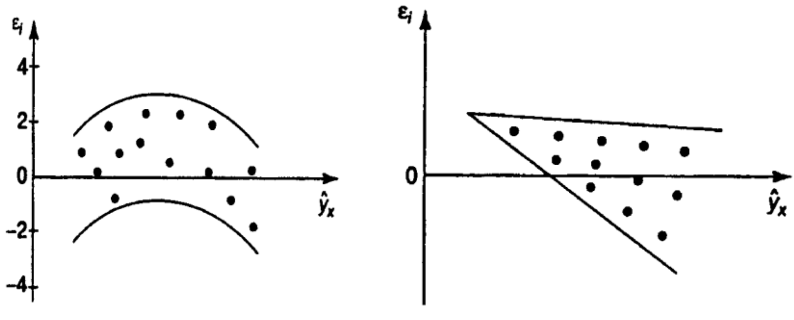


Рис. 3.1. Зависимость случайных остатков ε_i от теоретических значений y_x.

Возможны следующие случаи, если ε_i зависит от y_x то:

1) остатки ε_i не случайны (рис. 3.2а);
2) остатки ε_i не имеют постоянной дисперсии (рис. 3.2б);
3) остатки ε_i носят систематический характер (рис. 3.2в).



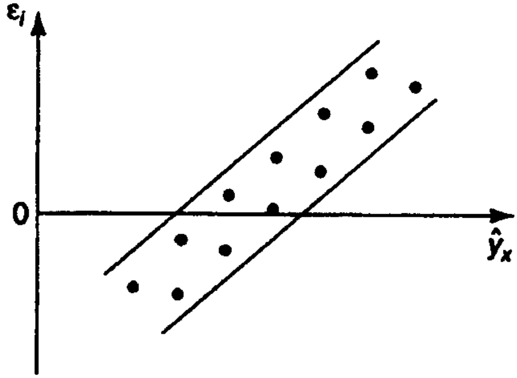


Рис. 3.2. Зависимость случайных остатков ε_i от теоретических значений y_x

В этих случаях необходимо либо применять другую функцию, либо вводить дополнительную информацию и заново строить уравнение регрессии до тех пор, пока остатки ε_i не будут случайными величинами.

Вторая предпосылка МНК относительно нулевой средней величины остатков означает, что $\sum\left(y - y_x\right) = 0$. Это выполнимо для линейных моделей и моделей, нелинейных относительно включаемых переменных.

Вместе с тем, несмещенность оценок коэффициентов регрессии, полученных МНК, зависит от независимости случайных остатков и величин x, что также исследуется в рамках соблюдения второй предпосылки МНК. С этой целью наряду с изложенным графиком зависимости остатков ε_i от теоретических значений результативного признака y_x строится график зависимости случайных остатков ε_i от факторов, включенных в регрессию x_j (рис. 3.3).

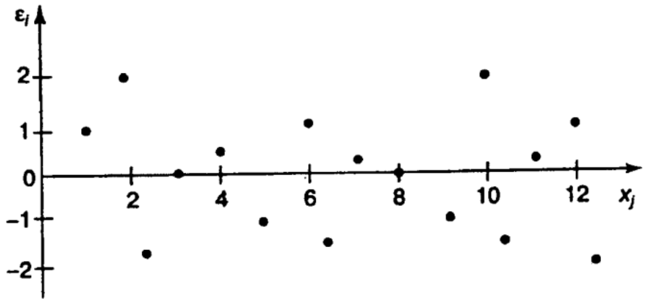


Рис. 3.3. Зависимость величины остатков от величины фактора x_j

Если остатки на графике расположены в виде горизонтальной полосы, то они независимы от значений x_j. Если же график показывает наличие зависимости ε_i и x_j, то модель неадекватна. Причины неадекватности могут быть разные. Возможно, что нарушена третья предпосылка МНК и дисперсия остатков не постоянна для каждого значения фактора x_j. Может быть неправильна спецификация модели и в нее необходимо ввести дополнительные члены от x_j, например x_j^2. Скопление точек в определенных участках значений фактора x_j говорит о наличии систематической погрешности модели.

Предпосылка о нормальном распределении остатков позволяет проводить проверку параметров регрессии и корреляции с помощью F - и t-критериев. Вместе с тем, оценки регрессии, найденные с применением МНК, обладают хорошими свойствами даже при отсутствии нормального распределения остатков, т.е. при нарушении пятой предпосылки МНК.

Совершенно необходимым для получения по МНК состоятельных оценок параметров регрессии является соблюдение третьей и четвертой предпосылок.

В соответствии с третьей предпосылкой МНК требуется, чтобы дисперсия остатков была *гомоскедастичной*. Это значит, что для каждого значения фактора x_j остатки ε_i имеют одинаковую дисперсию. Если это условие применения МНК не соблюдается, то имеет место *гетероскедастичность*. Наличие гетероскедастичности можно наглядно видеть из поля корреляции (рис. 3.4).

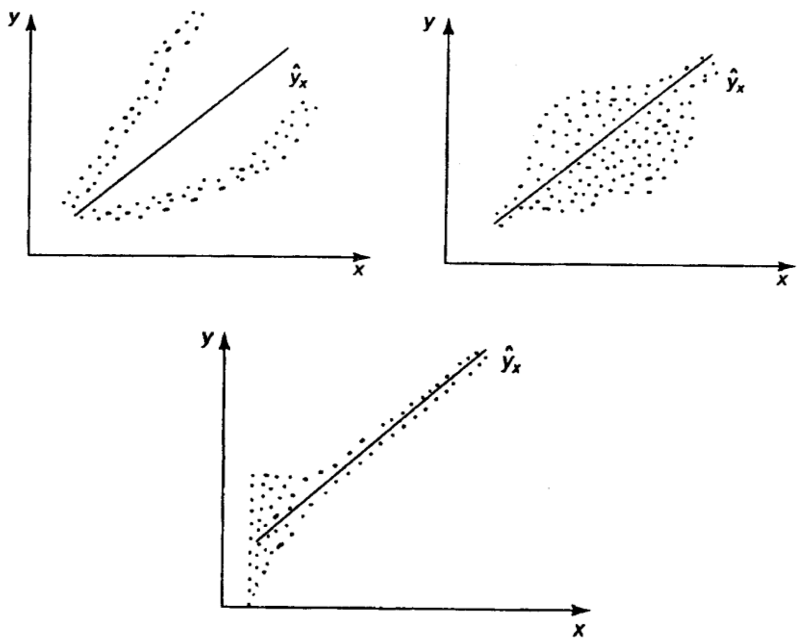


Рис. 3.4. Примеры гетероскедастичности

На рис. 3.4 изображено: а – дисперсия остатков растет по мере увеличения x; б – дисперсия остатков достигает максимальной величины при средних значениях переменной x и уменьшается при минимальных и максимальных значениях x; в – максимальная дисперсия остатков при малых значениях x и дисперсия остатков однородна по мере увеличения значений x.

Наличие гомоскедастичности или гетероскедастичности можно видеть и по рассмотренному выше графику зависимости остатков ε_i от теоретических значений результативного признака y_x. Так, для рис. 3.4а зависимость остатков от y_x представлена на рис. 3.5.

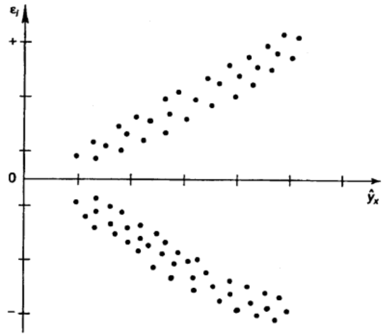


Рис. 3.5. Гетероскедастичность: большая дисперсия ε_i для больших значений y_x

Соответственно для зависимости, изображенной на полях корреляции рис. 3.4б и 3.4в гетероскедастичность остатков представлена на рис. 3.6 и 3.7.

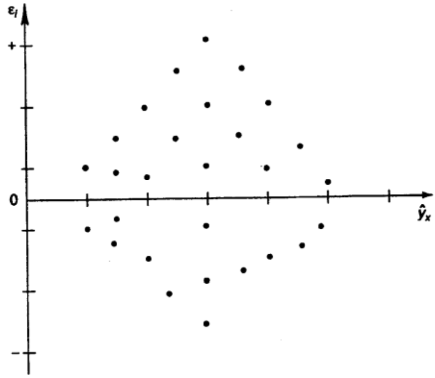


Рис. 3.6. Гетероскедастичность, соответствующая полю корреляции на рис. 3.4б.

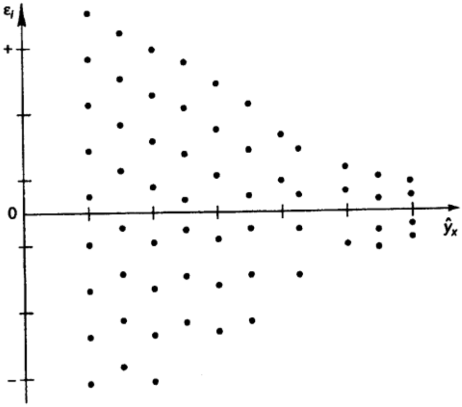


Рис. 3.7. Гетероскедастичность, соответствующая полю корреляции на рис. 3.4в.

Для множественной регрессии данный вид графиков является наиболее приемлемым визуальным способом изучения гомо- и гетероскедастичности.

При построении регрессионных моделей чрезвычайно важно соблюдение четвертой предпосылки МНК – отсутствие автокорреляции остатков, т.е. значения остатков ε_i, распределены независимо друг от друга. Автокорреляция остатков означает наличие корреляции между остатками текущих и предыдущих (последующих) наблюдений. Коэффициент корреляции между ε_i и ε_j, где ε_i – остатки текущих наблюдений, ε_j – остатки предыдущих наблюдений (например, $j = i - 1$), может быть определен как

$$r_{\varepsilon_i \varepsilon_j} = \frac{\operatorname{cov}\left(\varepsilon_i,\ \varepsilon_j\right)}{\sigma_{\varepsilon_i} \cdot \sigma_{\varepsilon_j}},$$

т.е. по обычной формуле линейного коэффициента корреляции. Если этот коэффициент окажется существенно отличным от нуля, то остатки автокоррелированы и функция плотности вероятности $F\left(\varepsilon\right)$ зависит от j-й точки наблюдения и от распределения значений остатков в других точках наблюдения.

Отсутствие автокорреляции остаточных величин обеспечивает состоятельность и эффективность оценок коэффициентов регрессии. Особенно актуально соблюдение данной предпосылки МНК при построении регрессионных моделей по рядам динамики, где ввиду наличия тенденции последующие уровни динамического ряда, как правило, зависят от своих предыдущих уровней.

При несоблюдении основных предпосылок МНК приходится корректировать модель, изменяя ее спецификацию, добавлять (исключать) некоторые факторы, преобразовывать исходные данные для того, чтобы получить оценки коэффициентов регрессии, которые обладают свойством несмещенности, имеют меньшее значение дисперсии остатков и обеспечивают в связи с этим более эффективную статистическую проверку значимости параметров регрессии.

3.5. Обобщенный метод наименьших квадратов (ОМНК)

При нарушении гомоскедастичности и наличии автокорреляции ошибок рекомендуется традиционный метод наименьших квадратов (известный в английской терминологии как метод OLS – Ordinary Least Squares) заменять *обобщенным методом*, т.е. *методом GLS* (Generalized Least Squares).

Обобщенный метод наименьших квадратов применяется к преобразованным данным и позволяет получать оценки, которые обладают не только свойством несмещенности, но и имеют меньшие выборочные дисперсии. Остановимся на использовании ОМНК для корректировки гетероскедастичности.

Как и раньше, будем предполагать, что среднее значение остаточных величин равно нулю. А вот дисперсия их не остается неизменной для разных значений фактора, а пропорциональна величине K_i, т.е. $\sigma^2_{\varepsilon_i} = \sigma^2 \cdot K_i$,

где $\sigma^2_{\varepsilon_i}$ – дисперсия ошибки при конкретном i-м значении фактора; σ^2 – постоянная дисперсия ошибки при соблюдении предпосылки о гомоскедастичности остатков; K_i – коэффициент пропорциональности, меняющийся с изменением величины фактора, что и обусловливает неоднородность дисперсии.

При этом предполагается, что σ^2 неизвестна, а в отношении величин K_i выдвигаются определенные гипотезы, характеризующие структуру гетероскедастичности.

В общем виде для уравнения $y_i = a + bx_i + \varepsilon_i$ при $\sigma^2_{\varepsilon_i} = \sigma^2 \cdot K_i$ модель примет вид: $y_i = a + bx_i + \sqrt{K_i}\varepsilon_i$. В ней остаточные величины гетероскедастичны. Предполагая в них отсутствие автокорреляции, можно перейти к уравнению с гомоскедастичными остатками, поделив все переменные, зафиксированные в ходе i-го наблюдения, на $\sqrt{K_i}$. Тогда дисперсия остатков будет величиной постоянной, т. е. $\sigma^2_{\varepsilon_i} = \sigma^2$.

Иными словами, от регрессии y по x перейдем к регрессии на новых переменных: $y/\sqrt{K}$ и $x/\sqrt{K}$. Уравнение регрессии примет вид: $\dfrac{y_i}{\sqrt{K_i}} = \dfrac{a}{\sqrt{K_i}} + b \cdot \dfrac{x_i}{\sqrt{K_i}} + \varepsilon_i$,

а исходные данные для данного уравнения будут иметь вид:

$$y = \begin{vmatrix} \dfrac{y_1}{\sqrt{K_1}} \\ \dfrac{y_2}{\sqrt{K_2}} \\ \dots\dots \\ \dfrac{y_n}{\sqrt{K_n}} \end{vmatrix}, \qquad x = \begin{vmatrix} \dfrac{x_1}{\sqrt{K_1}} \\ \dfrac{x_2}{\sqrt{K_2}} \\ \dots\dots \\ \dfrac{x_n}{\sqrt{K_n}} \end{vmatrix}.$$

По отношению к обычной регрессии уравнение с новыми, преобразованными переменными представляет собой взвешенную регрессию, в которой переменные y и x взяты с весами $1/\sqrt{K}$.

Оценка параметров нового уравнения с преобразованными переменными приводит к взвешенному методу наименьших квадратов, для которого необходимо минимизировать сумму квадратов отклонений вида

$$S(a, b) = \sum_{i=1}^{n} \frac{1}{K_i} (y_i - a - bx_i)^2.$$

Соответственно получим следующую систему нормальных уравнений:

$$\begin{cases} \sum \frac{y}{K} = a \cdot \sum \frac{1}{K} + b \cdot \sum \frac{x}{K}, \\ \sum \frac{y \cdot x}{K} = a \cdot \sum \frac{x}{K} + b \cdot \sum \frac{x^2}{K}. \end{cases}$$

Если преобразованные переменные x и y взять в отклонениях от средних уровней, то коэффициент регрессии b можно определить как

$$b = \frac{\sum \frac{1}{K} \cdot x \cdot y \cdot}{\sum \frac{1}{K} \cdot x^2}$$

При обычном применении метода наименьших квадратов к уравнению линейной регрессии для переменных в отклонениях от средних уровней коэффициент регрессии b определяется по формуле:

$$b = \frac{\sum x \cdot y}{\sum x^2}.$$

Как видим, при использовании обобщенного МНК с целью корректировки гетероскедастичности коэффициент регрессии b представляет собой взвешенную величину по отношению к обычному МНК с весом $1/K$.

Аналогичный подход возможен не только для уравнения парной, но и для множественной регрессии. Предположим, что рассматривается модель вида

$$y = a + b_1 x_1 + b_2 x_2 + \varepsilon,$$

для которой дисперсия остаточных величин оказалась пропорциональна K_i^2. K_i представляет собой коэффициент пропорциональности, принимающий различные значения для соответствующих i значений факторов x_1 и x_2. Ввиду того, что

$$\sigma_{\varepsilon_i}^2 = \sigma^2 \cdot K_i^2,$$

Рассматриваемая модель примет вид

$$y_i = a + b_1 x_{1i} + b_2 x_{2i} + K_i \varepsilon_i,$$ где ошибки гетероскедастичны.

Для того чтобы получить уравнение, где остатки ε_i гомоскедастичны, перейдем к новым преобразованным переменным, разделив все члены исходного уравнения на коэффициент пропорциональности K. Уравнение с преобразованными переменными составит

$$\frac{y_i}{K_i}=\frac{a}{K_i}+b_1\frac{x_{1i}}{K_i}+b_2\frac{x_{2i}}{K_i}+\varepsilon_i \cdot$$

Это уравнение не содержит свободного члена. Вместе с тем, найдя переменные в новом преобразованном виде и применяя обычный МНК к ним, получим иную спецификацию модели:

$$\frac{y_i}{K_i}=A+b_1\frac{x_{1i}}{K_i}+b_2\frac{x_{2i}}{K_i}+\varepsilon_i \cdot$$

Параметры такой модели зависят от концепции, принятой для коэффициента пропорциональности K_i. В эконометрических исследованиях довольно часто выдвигается гипотеза, что остатки ε_i пропорциональны значениям фактора. Так, если в уравнении

$$y=a+b_1x_1+b_2x_2+...+b_mx_m+e$$

предположить, что $e=\varepsilon\cdot x_1$, т.е. $K=x_1$ и $\sigma^2_{\varepsilon_i}=\sigma^2\cdot x_1$, то обобщенный МНК предполагает оценку параметров следующего трансформированного уравнения:

$$\frac{y}{x_1}=b_1+b_2\frac{x_2}{x_1}+...+b_m\frac{x_m}{x_1}+\varepsilon \cdot$$

Применение в этом случае обобщенного МНК приводит к тому, что наблюдения с меньшими значениями преобразованных переменных x/K имеют при определении параметров регрессии относительно больший вес, чем с первоначальными переменными. Вместе с тем, следует иметь в виду, что новые преобразованные переменные получают новое экономическое содержание и их регрессия имеет иной смысл, чем регрессия по исходным данным.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ. Пусть y – издержки производства, x_1 – объем продукции, x_2 – основные производственные фонды, x_3 – численность работников, тогда уравнение

$$y=a+b_1x_1+b_2x_2+b_3x_3+e$$

является моделью издержек производства с объемными факторами. Предполагая, что $\sigma^2_{\varepsilon_i}$ пропорциональна квадрату численности работников x_3, мы получим в качестве результативного признака затраты на одного работника y/x_3, а в качестве факторов следующие показатели: производительность труда x_1/x_3 и фондовооруженность труда x_2/x_3. Соответственно трансформированная модель примет вид

$$\frac{y}{x_3}=b_3+b_1\frac{x_1}{x_3}+b_2\frac{x_2}{x_3}+\varepsilon,$$

где параметры b_1, b_2, b_3 численно не совпадают с аналогичными параметрами предыдущей модели. Кроме этого, коэффициенты регрессии меняют экономическое содержание: из показателей силы связи, характеризующих среднее абсолютное изменение издержек производства с изменением абсолютной величины соответствующего фактора на единицу, они фиксируют при обобщенном МНК среднее изменение затрат на работника; с изменением производительности труда на единицу при неизменном уровне фовдовооруженности труда; и с изменением фондовооруженности труда на единицу при неизменном уровне производительности труда.

Если предположить, что в модели с первоначальными переменными дисперсия остатков пропорциональна квадрату объема продукции, $\sigma^2_{\varepsilon_i}=\sigma^2\cdot x_1^2$, можно перейти к уравнению регрессии вида

$$\frac{y}{x_1}=b_1+b_2\frac{x_2}{x_1}+b_3\frac{x_3}{x_1}+\varepsilon.$$

В нем новые переменные: y/x_1 – затраты на единицу (или на 1 руб. продукции), x_2/x_1 – фондоемкость продукции, x_3/x_1 – трудоемкость продукции.

Гипотеза о пропорциональности остатков величине фактора может иметь реальное основание: при обработке недостаточно однородной совокупности, включающей как крупные, так и мелкие предприятия, большим объемным значениям фактора может соответствовать большая дисперсия результативного признака и большая дисперсия остаточных величин.

При наличии одной объясняющей переменной гипотеза $\sigma^2_{\varepsilon_i}=\sigma^2x^2$ трансформирует линейное уравнение $y=a+bx+e$ в уравнение $\frac{y}{x}=b+\frac{a}{x}+\varepsilon$, в котором параметры a и b поменялись местами, константа стала коэффициентом наклона линии регрессии, а коэффициент регрессии – свободным членом.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ. Рассматривая зависимость сбережений y от дохода x, по первоначальным данным было получено уравнение регрессии

$$y = -1{,}081 + 0{,}1178 \cdot x.$$

Применяя обобщенный МНК к данной модели в предположении, что ошибки пропорциональны доходу, было получено уравнение для преобразованных данных:

$$\frac{y}{x} = 0{,}1026 - 0{,}8538 \cdot \frac{1}{x}.$$

Коэффициент регрессии первого уравнения сравнивают со свободным членом второго уравнения, т.е. 0,1178 и 0,1026 – оценки параметра b зависимости сбережений от дохода.

Переход к относительным величинам существенно снижает вариацию фактора и соответственно уменьшает дисперсию ошибки. Он представляет собой наиболее простой случай учета гетероскедастичности в регрессионных моделях с помощью обобщенного МНК. Процесс перехода к относительным величинам может быть осложнен выдвижением иных гипотез о пропорциональности ошибок относительно включенных в модель факторов. Использование той или иной гипотезы предполагает специальные исследования остаточных величин для соответствующих регрессионных моделей. Применение обобщенного МНК позволяет получить оценки параметров модели, обладающие меньшей дисперсией.

3.6. Регрессионные модели с переменной структурой (фиктивные переменные)

До сих пор в качестве факторов рассматривались экономические переменные, принимающие количественные значения в некотором интервале. Вместе с тем может оказаться необходимым включить в модель фактор, имеющий два или более качественных уровней. Это могут быть разного рода атрибутивные признаки, такие, например, как профессия, пол, образование, климатические условия, принадлежность к определенному региону. Чтобы ввести такие переменные в регрессионную модель, им должны быть присвоены те или иные *цифровые метки,* т.е. качественные переменные преобразованы в количественные. Такого вида сконструированные переменные в эконометрике принято называть *фиктивными переменными.*

Рассмотрим применение фиктивных переменных для функции спроса. Предположим, что по группе лиц мужского и женского пола изучается линейная зависимость потребления кофе от цены. В общем виде для совокупности обследуемых уравнение регрессии имеет вид:

$$y = a + bx + \varepsilon,$$

где y – количество потребляемого кофе; x – цена.

Аналогичные уравнения могут быть найдены отдельно для лиц мужского пола: $y_1 = a_1 + b_1 x_1 + \varepsilon_1$ и женского пола: $y_2 = a_2 + b_2 x_2 + \varepsilon_2$.

Различия в потреблении кофе проявятся в различии средних $\bar{y}_1$ и $\bar{y}_2$. Вместе с тем сила влияния x на y может быть одинаковой, т.е. $b \approx b_1 \approx b_2$. В этом случае возможно построение общего уравнения регрессии с включением в него фактора «пол» в виде фиктивной переменной. Объединяя уравнения y_1 и y_2 и, вводя фиктивные переменные, можно прийти к следующему выражению:

$$y = a_1 z_1 + a_2 z_2 + bx + \varepsilon,$$

где z_1 и z_2 – фиктивные переменные, принимающие значения:

$$z_1 = \begin{cases} 1 \ - \ \text{мужской пол,} \\ 0 \ - \ \text{женский пол;} \end{cases} \qquad z_2 = \begin{cases} 0 \ - \ \text{мужской пол,} \\ 1 \ - \ \text{женский пол.} \end{cases}$$

В общем уравнении регрессии зависимая переменная y рассматривается как функция не только цены x но и пола $(z_1,\ z_2)$. Переменная z рассматривается как дихотомическая переменная, принимающая всего два значения: 1 и 0. При этом когда $z_1 = 1$, то $z_2 = 0$, и наоборот.

Для лиц мужского пола, когда $z_1 = 1$ и $z_2 = 0$, объединенное уравнение регрессии составит: $y = a_1 + bx$, а для лиц женского пола, когда $z_1 = 0$ и $z_2 = 1$: $y = a_2 + bx$. Иными словами, различия в потреблении для лиц мужского и женского пола вызваны различиями свободных членов уравнения регрессии: $a_1 \neq a_2$. Параметр b является общим для всей совокупности лиц, как для мужчин, так и для женщин.

Однако при введении двух фиктивных переменных z_1 и z_2 в модель $y = a_1 z_1 + a_2 z_2 + bx + \varepsilon$ применение МНК для оценивания параметров a_1 и a_2 приведет к вырожденной матрице исходных данных, а следовательно, и к невозможности получения их оценок. Объясняется это тем, что при использовании МНК в данном уравнении появляется свободный член, т.е. уравнение примет вид

$$y = A + a_1 z_1 + a_2 z_2 + bx + \varepsilon \cdot$$

Предполагая при параметре A независимую переменную, равную 1, имеем следующую матрицу исходных данных:

$$\begin{bmatrix} 1 & 1 & 0 & x_1 \\ 1 & 1 & 0 & x_2 \\ 1 & 0 & 1 & x_3 \\ 1 & 1 & 0 & x_4 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & 1 & x_n \end{bmatrix}.$$

В рассматриваемой матрице существует линейная зависимость между первым, вторым и третьим столбцами: первый равен сумме второго и третьего столбцов. Поэтому матрица исходных факторов вырождена. Выходом из создавшегося затруднения может явиться переход к уравнениям

$$y = A + A_1 z_1 + bx + \varepsilon \text{ или}$$

$$y = A + A_2 z_2 + bx + \varepsilon,$$

т.е. каждое уравнение включает только одну фиктивную переменную z_1 или z_2.

Предположим, что определено уравнение

$$y = A + A_1 z_1 + bx + \varepsilon,$$

где z_1 принимает значения 1 для мужчин и 0 для женщин.

Теоретические значения размера потребления кофе для мужчин будут получены из уравнения

$$y = A + A_1 + bx \cdot$$

Для женщин соответствующие значения получим из уравнения

$$y = A + bx \cdot$$

Сопоставляя эти результаты, видим, что различия в уровне потребления мужчин и женщин состоят в различии свободных членов данных уравнений: A – для женщин и $A + A_1$ – для мужчин.

Теперь качественный фактор принимает только два состояния, которым соответствуют значения 1 и 0. Если же число градаций качественного признака-фактора превышает два, то в модель вводится несколько фиктивных переменных, число которых должно быть меньше числа качественных градаций. Только при соблюдении этого положения матрица исходных фиктивных переменных не будет линейно зависима и возможна оценка параметров модели.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ. Проанализируем зависимость цены двухкомнатной квартиры от ее полезной площади. При этом в модель могут быть введены фиктивные переменные, отражающие тип дома: «хрущевка», панельный, кирпичный.

При использовании трех категорий домов вводятся две фиктивные переменные: z_1 и z_2. Пусть переменная z_1 принимает значение 1 для панельного дома и 0 для всех остальных типов домов; переменная z_2 принимает значение 1 для кирпичных домов и 0 для остальных; тогда переменные z_1 и z_2 принимают значения 0 для домов типа «хрущевки».

Предположим, что уравнение регрессии с фиктивными переменными составило:

$$y = 320 + 500x + 2200z_1 + 1600z_2.$$

Частные уравнения регрессии для отдельных типов домов, свидетельствуя о наиболее высоких ценах квартир в панельных домах, будут иметь следующий вид: «хрущевки» – $y = 320 + 500x$; панельные – $y = 2520 + 500x$; кирпичные – $y = 1920 + 500x$.

Параметры при фиктивных переменных z_1 и z_2 представляют собой разность между средним уровнем результативного признака для соответствующей группы и базовой группы. В рассматриваемом примере за базу сравнения цены взяты дома «хрущевки», для которых $z_1 = z_2 = 0$. Параметр при Z_1, равный 2200, означает, что при одной и той же полезной площади квартиры цена ее в панельных домах в среднем на 2200 долл. США выше, чем в «хрущевках». Соответственно параметр при Z_2 показывает, что в кирпичных домах цена выше в среднем на 1600 долл. при неизменной величине полезной площади по сравнению с указанным типом домов.

В отдельных случаях может оказаться необходимым введение двух и более групп фиктивных переменных, т.е. двух и более качественных факторов, каждый из которых может иметь несколько градаций. Например, при изучении потребления некоторого товара наряду с факторами, имеющими количественное выражение (цена, доход на одного члена семьи, цена на взаимозаменяемые товары и др.), учитываются и качественные факторы. С их помощью оцениваются различия в потреблении отдельных социальных групп населения, дифференциация в потреблении по полу, национальному составу и др. При построении такой модели из каждой группы фиктивных переменных следует исключить по одной переменной. Так, если модель будет включать три социальные группы, три возрастные категории и ряд экономических переменных, то она примет вид:

$$y = a + b_1 s_1 + b_2 s_2 + b_3 z_1 + b_4 z_2 + b_5 x_1 + b_6 x_2 + \ldots + b_{m+4} x_m + \varepsilon,$$

где y – потребление;

$$s_i = \begin{cases} 1 - \text{ если наблюдения относятся к } i\text{-й социальной группе } (i = 1,\ 2), \\ 0 - \text{ в остальных случаях;} \end{cases}$$

$$z_i = \begin{cases} 1 - \text{ если наблюдения относятся к } j\text{-й возрастной группе } (j = 1,\ 2), \\ 0 - \text{ в остальных случаях;} \end{cases}$$

$x_1,\ x_2,\ \ldots,\ x_m$ – экономические (количественные) переменные.

До сих пор мы рассматривали фиктивные переменные как факторы, которые используются в регрессионной модели наряду с количественными переменными. Вместе с тем возможна регрессия только на фиктивных переменных. Например, изучается дифференциация заработной платы рабочих высокой квалификации по регионам страны. Модель заработной платы может иметь вид:

$$y = a + b_1 z_1 + b_2 z_2 + \ldots + b_m z_m,$$

где y – средняя заработная плата рабочих высокой квалификации по отдельным предприятиям;

$$z_1 = \begin{cases} 1 - \text{ если предприятие находится в Северо-Западном районе;} \\ 0 - \text{ если предприятие находится в остальных районах;} \end{cases}$$

$$z_2 = \begin{cases} 1 - \text{ если предприятие находится в Волго-Вятском районе;} \\ 0 - \text{ если предприятие находится в остальных районах;} \end{cases}$$

…………………………………………………………………………..

$$z_m = \begin{cases} 1 - \text{ если преприятие находится в Дальневосточном районе;} \\ 0 - \text{ если предприятие находится в остальных районах.} \end{cases}$$

Поскольку последний район, указанный в модели, обозначен z_m, то в исследование включено $m+1$ район.

Мы рассмотрели модели с фиктивными переменными, в которых последние выступают факторами. Может возникнуть необходимость построить модель, в которой дихотомический признак, т.е. признак, который может принимать только два значения, играет роль результата. Подобного вида модели применяются, например, при обработке данных социологических опросов. В качестве зависимой переменной y рассматриваются ответы на вопросы, данные в альтернативной форме: «да» или «нет». Поэтому зависимая переменная имеет два значения: 1, когда имеет место ответ «да», и 0 – во всех остальных случаях. Модель такой зависимой переменной имеет вид:

$$y = a + b_1 x_1 + ... + b_m x_m + \varepsilon \cdot$$

Модель является вероятностной линейной моделью. В ней y принимает значения 1 и 0, которым соответствуют вероятности p и $1-p$. Поэтому при решении модели находят оценку условной вероятности события y при фиксированных значениях x. Для оценки параметров линейно-вероятностной модели применяются методы Logit-, Probit- и Tobit-анализа. Такого рода модели используют при работе с неколичественными переменными. Как правило, это модели выбора из заданного набора альтернатив. Зависимая переменная y представлена дискретными значениями (набор альтернатив), объясняющие переменные x_j – характеристики альтернатив (время, цена), z_j – характеристики индивидов (возраст, доход, уровень образования). Модель такого рода позволяет предсказать долю индивидов в генеральной совокупности, которые выбирают данную альтернативу.

Среди моделей с фиктивными переменными наибольшими прогностическими возможностями обладают модели, в которых зависимая переменная y рассматривается как функция ряда экономических факторов x_i и фиктивных переменных z_j. Последние обычно отражают различия в формировании результативного признака по отдельным группам единиц совокупности, т.е. в результате неоднородной структуры пространственного или временного характера.

Вопросы для самоконтроля

1. Что понимается под множественной регрессией?
2. Какие задачи решаются при построении уравнения регрессии?
3. Какие задачи решаются при спецификации модели?
4. Какие требования предъявляются к факторам, включаемым в уравнение регрессии?
5. Что понимается под коллинеарностью и мультиколлинеарностью факторов?
6. Как проверяется наличие коллинеарности и мультиколлинеарности?
7. Какие подходы применяются для преодоления межфакторной корреляции?
8. Что означает низкое значение коэффициента множественной корреляции?
9. Как проверяется значимость уравнения регрессии и его коэффициентов?
10. Как оценивается значимость факторов?
11. Что понимается под гомоскедастичностью остатков?
12. Какие переменные называются фиктивными?
13. Цель применения теста Чоу.

Тестовые задания

1. Коэффициент множественной корреляции может принимать значения …

а) от 1 до 100
б) от 0 до 10
в) от - ∞ до + ∞
г) от – 1 до + 1
д) от – 10 до + 10
е) от 0 до + 1

2. Для выявления мультиколлинеарности используются …

а) частные коэффициенты корреляции
б) парные коэффициенты корреляции
в) множественный коэффициент детерминации
г) коэффициент регрессии

3. Коэффициент множественной корреляции используется для …

а) оценки тесноты связи между y и всеми факторами модели
б) оценки тесноты связи между y и отдельным фактором
в) оценки тесноты связи между факторами
г) оценки влияния на y неучтенных в модели факторов

4. Тесная зависимость между факторными признаками, включенными в модель, называется …

а) регрессия
б) мультиколлинеарность
в) корреляция

г) коллинеарность

5. Тест Фишера является …

а) двусторонним

б) односторонним

в) многосторонним

г) многокритериальным

6. Коэффициенты парной корреляции между тремя регрессорами следующие: $r_{x_1x_2} = 0,903$*;* $r_{x_1x_3} = 0,618$*;* $r_{x_2x_3} = 0,586$*;* $r_{yx_1} = -0,812$*;* $r_{yx_2} = 0,456$ *и* $r_{yx_3} = -0,348$. *Между какими переменными наблюдается мультиколлинеарность?*

а) y и x_1	г) x_1 и x_2
б) y и x_2	д) x_2 и x_3
в) y и x_3	е) x_1 и x_3

7. Коэффициенты парной корреляции между тремя регрессорами следующие: $r_{x_1x_2} = 0,903$*;* $r_{x_1x_3} = 0,618$*;* $r_{x_2x_3} = 0,586$*;* $r_{yx_1} = -0,812$*;* $r_{yx_2} = 0,456$ *и* $r_{yx_3} = -0,348$. *Какой из факторов следует исключить из анализа?*

а) x_1 б) x_2 в) x_3

8. МНК автоматически дает для данной выборки … значение коэффициента детерминации:

а) минимальное

б) максимальное

в) среднее

г) средневзвешенное

д) случайное

9. Число степеней свободы для критерия Фишера:

а) $n-m$

б) $n-m+1$

в) $n-m-1$

г) $n-1$

д) m

10. Число степеней свободы для критерия Стьюдента:

а) $n-m$

б) $n-m+1$

в) $n-m-1$

г) $n-1$

д) $n+1$

11. Фиктивная переменная – переменная, принимающая в каждом наблюдении:

а) ряд значений от 0 до 1

б) только отрицательные значения

в) только два значения 0 или 1

г) только положительные значения

д) случайные

12. При добавлении еще одной переменной в уравнение регрессии коэффициент детерминации …

а) остается неизменным

б) уменьшается

в) увеличивается

13. В модели множественной регрессии за изменение … отвечает несколько объясняющих переменных …

а) свободного члена уравнения

б) зависимой переменной

в) независимой переменной

г) случайной составляющей

14. В регрессионном анализе коэффициент детерминации определяет

а) долю вариации факторного признака

б) долю вариации результативного и факторного признаков

в) долю вариации результативного признака

г) долю вариации случайной ошибки

15. Зная уровень значимости, объем выборки и число объясняющих переменных, можно вычислить …

а) величину F для критерия Фишера

б) величину t для критерия Фишера

в) среднюю ошибку аппроксимации

г) $\beta -$ коэффициент

д) средний коэффициент эластичности

16. В каком из методов построения уравнения множественной регрессии используется отсев факторов из полного его набора?

а) пошаговый регрессионный анализ

б) метод включения

в) метод исключения

4. СИСТЕМЫ ЭКОНОМЕТРИЧЕСКИХ УРАВНЕНИЙ

4.1. Понятие СОУ

При использовании отдельных уравнений регрессии, например для экономических расчетов, в большинстве случаев предполагается, что аргументы (факторы) можно изменять независимо друг от друга. Однако это предположение является очень грубым: практически изменение одной переменной, как правило, не может происходить при абсолютной неизменности других. Ее изменение повлечет за собой изменения во всей системе взаимосвязанных признаков. Следовательно, отдельно взятое уравнение множественной регрессии не может характеризовать истинные влияния отдельных признаков на вариацию результирующей переменной. Именно поэтому в последние десятилетия в экономических исследованиях важное место заняла проблема описания структуры связей между переменными **системой** так называемых **одновременных эконометрических уравнений (СОУ)**, называемых также структурными уравнениями.

Система уравнений в эконометрических исследованиях может быть построена по-разному.

Возможна *система независимых уравнений,* когда каждая зависимая переменная y рассматривается как функция одного и того же набора факторов x:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\ y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_m. \end{cases} \tag{4.1}$$

Набор факторов x_j в каждом уравнении может варьировать. Каждое уравнение системы независимых уравнений может рассматриваться самостоятельно. Для нахождения его параметров используется метод наименьших квадратов. По существу, каждое уравнение этой системы является уравнением регрессии. Так как фактические значения зависимой переменной отличаются от теоретических на величину случайной ошибки, то в каждом уравнении присутствует величина случайной ошибки ε_i.

Если зависимая переменная y одного уравнения выступает в виде фактора x в другом уравнении, то исследователь может строить модель в виде *системы рекурсивных уравнений:*

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = b_{m1}y_1 + \ldots + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_m. \end{cases} \tag{4.2}$$

В данной системе зависимая переменная y включает в каждое последующее уравнение в качестве факторов все зависимые переменные предшествующих уравнений наряду с набором собственно факторов x. Каждое уравнение этой системы может рассматриваться самостоятельно, и его параметры определяются методом наименьших квадратов (МНК).

Наибольшее распространение в эконометрических исследованиях получила *система взаимозависимых уравнений*. В ней одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других уравнениях – в правую часть системы:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \ldots + b_{1m}y_m \qquad + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\ y_2 = b_{21}y_1 + b_{23}y_3 + \ldots + b_{2m}y_m \qquad + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + \ldots + b_{3m}y_m \qquad + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = b_{m1}y_1 + b_{m2}y_2 + \ldots + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_n. \end{cases} \tag{4.3}$$

Система взаимозависимых уравнений получила название *системы совместных, одновременных уравнений*. Тем самым подчеркивается, что в системе одни и те же переменные одновременно рассматриваются как зависимые в одних уравнениях и как независимые в других. В эконометрике эта система уравнений называется также *структурной формой модели*. В отличие от предыдущих систем каждое уравнение системы одновременных уравнений не может рассматриваться самостоятельно, и для нахождения его параметров традиционный МНК неприменим. С этой целью используются специальные приемы оценивания.

4.2. Структурная и приведенная формы модели

Система совместных, одновременных уравнений (или структурная форма модели) обычно содержит эндогенные и экзогенные переменные.

Эндогенные переменные – это зависимые переменные, число которых равно числу уравнений в системе и которые обозначаются через y. *Экзогенные переменные* – это предопределенные переменные, влияющие на эндогенные переменные, но не зависящие от них. Обозначаются через x.

Классификация переменных на эндогенные и экзогенные зависит от теоретической концепции принятой модели. Экономические переменные могут выступать в одних моделях как эндогенные, а в других как экзогенные переменные. Внеэкономические переменные (например, климатические условия, социальное положение, пол, возрастная категория) входят в систему только как экзогенные переменные. В качестве экзогенных переменных могут рассматриваться значения эндогенных переменных за предшествующий период времени (*лаговые переменные*).

Структурная форма модели позволяет увидеть влияние изменений любой экзогенной переменной на значения эндогенной переменной. Целесообразно в качестве экзогенных переменных выбирать такие переменные, которые могут быть объектом регулирования. Меняя их и управляя ими, можно заранее иметь целевые значения эндогенных переменных.

Структурная форма модели в правой части содержит при эндогенных переменных коэффициенты b_{ik} и экзогенных переменных – коэффициенты a_{ij}, которые называются *структурными коэффициентами* модели. Все переменные в модели выражены в отклонениях от среднего уровня, т.е. под x подразумевается $x-\bar{x}$, а под y – соответственно $y-\bar{y}$. Поэтому свободный член в каждом уравнении системы (4.3) отсутствует.

Использование МНК для оценивания структурных коэффициентов модели дает, как принято считать в теории, смещенные и несостоятельные оценки. Поэтому обычно для определения структурных коэффициентов модели структурная форма модели преобразуется в *приведенную форму* модели.

Приведенная форма модели представляет собой систему линейных функций эндогенных переменных от экзогенных:

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + ... + \delta_{1n}x_n + u_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + ... + \delta_{2n}x_n + u_2, \\ .... \\ y_m = \delta_{m1}x_1 + \delta_{m2}x_2 + ... + \delta_{mn}x_n + u_m, \end{cases} \quad (4.4)$$

где δ_{ij} – коэффициенты приведенной формы модели, u_i – остаточная величина для приведенной формы.

По своему виду приведенная форма модели ничем не отличается от системы независимых уравнений, параметры которой оцениваются традиционным МНК. Применяя МНК, можно оценить δ_{ij}, а затем оценить значения эндогенных переменных через экзогенные.

Коэффициенты приведенной формы модели представляют собой нелинейные функции коэффициентов структурной формы модели. Рассмотрим это положение на примере простейшей структурной модели, выразив коэффициенты приведенной формы модели через коэффициенты структурной модели.

Для структурной модели вида

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_2 \end{cases} \quad (4.5)$$

приведенная форма модели имеет вид

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + u_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + u_2. \end{cases} \quad (4.6)$$

Из первого уравнения (4.5) можно выразить y_2 следующим образом (ради упрощения опускаем случайную величину):

$$y_2 = \frac{y_1 - a_{11}x_1}{b_{12}}.$$

Подставляя во второе уравнение (4.5), имеем

$$\frac{y_1 - a_{11}x_1}{b_{12}} = b_{21}y_1 + a_{22}x_2,$$

откуда

$$y_1 = \frac{a_{11}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}b_{12}}{1 - b_{12}b_{21}}x_2.$$

Поступая аналогично со вторым уравнением системы (4.5), получим

$$y_2 = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}}{1 - b_{12}b_{21}}x_2,$$

т.е. система (3.5) принимает вид

$$\begin{cases} y_1 = \dfrac{a_{11}}{1 - b_{12}b_{21}}x_1 + \dfrac{a_{22}b_{12}}{1 - b_{12}b_{21}}x_2, \\ y_2 = \dfrac{a_{11}b_{21}}{1 - b_{12}b_{21}}x_1 + \dfrac{a_{22}}{1 - b_{12}b_{21}}x_2. \end{cases}$$

Таким образом, можно сделать вывод о том, что коэффициенты приведенной формы модели будут выражаться через коэффициенты структурной формы следующим образом:

$$\delta_{11} = \frac{a_{11}}{1 - b_{12}b_{21}}, \quad \delta_{12} = \frac{a_{22}b_{12}}{1 - b_{12}b_{21}},$$

$$\delta_{21} = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}}, \quad \delta_{22} = \frac{a_{22}}{1 - b_{12}b_{21}}.$$

Следует заметить, что приведенная форма модели хотя и позволяет получить значения эндогенной переменной через значения экзогенных переменных, но аналитически она уступает структурной форме модели, так как в ней отсутствуют оценки взаимосвязи между эндогенными переменными.

4.3. Проблема идентификации

При переходе от приведенной формы модели к структурной эконометрист сталкивается с проблемой идентификации. Идентификация – это единственность соответствия между приведенной и структурной формами модели.

Структурная модель (4.3) в полном виде содержит

$m \cdot (m+n-1)$ параметров, а приведенная форма модели в полном виде содержит $m \cdot n$ параметров. Т.е. в полном виде структурная модель содержит большее число параметров, чем приведенная форма модели. Соответственно $m \cdot (m+n-1)$ параметров структурной модели не могут быть однозначно определены из $m \cdot n$ параметров приведенной формы модели.

Чтобы получить единственно возможное решение для структурной модели, необходимо предположить, что некоторые из структурных коэффициентов модели ввиду слабой взаимосвязи признаков с эндогенной переменной из левой части системы равны нулю. Тем самым уменьшится число структурных коэффициентов модели. Уменьшение числа структурных коэффициентов модели возможно и другим путем: например, путем приравнивания некоторых коэффициентов друг к другу, т.е. путем предположений, что их воздействие на формируемую эндогенную переменную одинаково. На структурные коэффициенты могут накладываться, например, ограничения вида $b_{ik} + a_{ij} = 0 \cdot$

С позиции идентифицируемости структурные модели можно подразделить на три вида:

1) идентифицируемые;
2) неидентифицируемые;
3) сверхидентифицируемые.

Модель *идентифицируема*, если все структурные ее коэффициенты определяются однозначно, единственным образом по коэффициентам приведенной формы модели, т. е. если число параметров структурной модели равно числу параметров приведенной формы модели. В этом случае структурные коэффициенты модели оцениваются через параметры приведенной формы модели и модель идентифицируема.

Модель *неидентифицируема*, если число приведенных коэффициентов меньше числа структурных коэффициентов, и в результате структурные коэффициенты не могут быть оценены через коэффициенты приведенной формы модели.

Модель *сверхидентифицируема*, если число приведенных коэффициентов больше числа структурных коэффициентов. В этом случае на основе коэффициентов приведенной формы можно получить два или более значений одного структурного коэффициента. В этой модели число структурных коэффициентов меньше числа коэффициентов приведенной формы. Сверхидентифицируемая модель в отличие от неидентифицируемой модели практически решаема, но требует для этого специальных методов исчисления параметров.

Структурная модель всегда представляет собой систему совместных уравнений, каждое из которых требуется проверять на идентификацию. Модель считается идентифицируемой, если каждое уравнение системы идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то и вся модель считается неидентифицируемой. Сверхидентифицируемая модель содержит хотя бы одно сверхидентифицируемое уравнение.

Выполнение условия идентифицируемости модели проверяется для каждого уравнения системы. Чтобы уравнение было идентифицируемо, необходимо, чтобы число предопределенных переменных, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении без одного.

Если обозначить число эндогенных переменных в i-м уравнении системы через H, а число экзогенных (предопределенных) переменных, которые содержатся в системе, но не входят в данное уравнение, — через D, то условие идентифицируемости модели может быть записано в виде следующего счетного правила:

Таблица 4.1

$D+1=H$	уравнение идентифицируемо
$D+1<H$	уравнение неидентифицируемо
$D+1>H$	уравнение сверхидентифицируемо

Для оценки параметров структурной модели система должна быть идентифицируема или сверхидентифицируема.

Рассмотренное счетное правило отражает необходимое, но недостаточное условие идентификации. Более точно условия идентификации определяются, если накладывать ограничения на коэффициенты матриц параметров структурной модели. Уравнение идентифицируемо, если по отсутствующим в нем переменным (эндогенным и экзогенным) можно из коэффициентов при них в других уравнениях системы получить матрицу, определитель которой не равен нулю, а ранг матрицы не меньше, чем число эндогенных переменных в системе без одного.

Целесообразность проверки условия идентификации модели через определитель матрицы коэффициентов, отсутствующих в данном уравнении, но присутствующих в других, объясняется тем, что возможна ситуация, когда для каждого уравнения системы выполнено счетное правило, а определитель матрицы названных коэффициентов равен нулю. В этом случае соблюдается лишь необходимое, но недостаточное условие идентификации.

В эконометрических моделях часто наряду с уравнениями, параметры которых должны быть статистически оценены, используются балансовые тождества переменных, коэффициенты при которых равны

± 1. В этом случае, хотя само тождество и не требует проверки на идентификацию, ибо коэффициенты при переменных в тождестве известны, в проверке на идентификацию собственно структурных уравнений системы тождества участвуют.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ. Изучается модель вида

$$\begin{cases} C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + \varepsilon_1, \\ I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + \varepsilon_2, \\ r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + \varepsilon_3, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

где C_t – расходы на потребление в период t, Y_t – совокупный доход в период t, I_t – инвестиции в период t, r_t – процентная ставка в период t, M_t – денежная масса в период t, G_t – государственные расходы в период t, C_{t-1} – расходы на потребление в период $t-1$, I_{t-1} инвестиции в период $t-1$. Первое уравнение – функция потребления, второе уравнение – функция инвестиций, третье уравнение – функция денежного рынка, четвертое уравнение – тождество дохода.

Модель представляет собой систему одновременных уравнений. Проверим каждое ее уравнение на идентификацию.

Модель включает четыре эндогенные переменные $\left(C_t, I_t, Y_t, r_t\right)$ и четыре предопределенные переменные (две экзогенные переменные – M_t и G_t и две лаговые переменные – C_{t-1} и I_{t-1}).

Проверим необходимое условие идентификации для каждого из уравнений модели.

Первое уравнение: $C_t = a_1 + b_{11} \cdot Y_t + b_{12} \cdot C_{t-1} + \varepsilon_1$. Это уравнение содержит две эндогенные переменные C_t и Y_t и одну предопределенную переменную C_{t-1}. Таким образом, $H = 2$, а $D = 4 - 1 = 3$, т.е. выполняется условие $D + 1 > H$. Уравнение сверхидентифицируемо.

Второе уравнение: $I_t = a_2 + b_{21} \cdot r_t + b_{22} \cdot I_{t-1} + \varepsilon_2$. Оно включает две эндогенные переменные I_t и r_t и одну экзогенную переменную I_{t-1}. Выполняется условие $D+1=3+1>H=2$. Уравнение сверхидентифицируемо.

Третье уравнение: $r_t = a_3 + b_{31} \cdot Y_t + b_{32} \cdot M_t + \varepsilon_3$. Оно включает две эндогенные переменные Y_t и r_t и одну экзогенную переменную M_t. Выполняется условие $D+1=3+1>H=2$. Уравнение сверхидентифицируемо.

Четвертое уравнение: $Y_t = C_t + I_t + G_t$. Оно представляет собой тождество, параметры которого известны. Необходимости в идентификации нет.

Проверим для каждого уравнения достаточное условие идентификации. Для этого составим матрицу коэффициентов при переменных модели.

	C_t	I_t	r_t	Y_t	C_{t-1}	I_{t-1}	M_t	G_t
I уравнение	–1	0	0	b_{11}	b_{12}	0	0	0
II уравнение	0	–1	b_{21}	0	0	b_{22}	0	0
III уравнение	0	0	–1	b_{31}	0	0	b_{32}	0
Тождество	1	1	0	–1	0	0	0	1

В соответствии с достаточным условием идентификации ранг матрицы коэффициентов при переменных, не входящих в исследуемое уравнение, должен быть равен числу эндогенных переменных модели без одного.

Первое уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

	I_t	r_t	I_{t-1}	M_t	G_t
II уравнение	–1	b_{21}	b_{22}	0	0
III уравнение	0	–1	0	b_{32}	0
Тождество	1	0	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3не равен нулю:

$$\begin{vmatrix} b_{22} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{22}b_{32} \neq 0.$$

Достаточное условие идентификации для данного уравнения выполняется.

Второе уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

	C_t	Y_t	C_{t-1}	M_t	G_t
I уравнение	–1	b_{11}	b_{12}	0	0
III уравнение	0	b_{31}	0	b_{32}	0
Тождество	1	–1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3не равен нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12}b_{32} \neq 0.$$

Достаточное условие идентификации для данного уравнения выполняется.

Третье уравнение. Матрица коэффициентов при переменных, не входящих в уравнение, имеет вид

	C_t	I_t	C_{t-1}	I_{t-1}	G_t
I уравнение	–1	0	b_{12}	0	0
II уравнение	0	–1	0	b_{22}	0
Тождество	1	1	0	0	1

Ранг данной матрицы равен трем, так как определитель квадратной подматрицы 3×3 не равен нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12}b_{22} \neq 0.$$

Достаточное условие идентификации для данного уравнения выполняется.

Таким образом, все уравнения модели сверхидентифицируемы.

Приведенная форма модели в общем виде будет выглядеть следующим образом:

$$\begin{cases} C_t = A_1 + \delta_{11}C_{t-1} + \delta_{12}I_{t-1} + \delta_{13}M_t + \delta_{14}G_t + u_1, \\ I_t = A_2 + \delta_{21}C_{t-1} + \delta_{22}I_{t-1} + \delta_{23}M_t + \delta_{24}G_t + u_2, \\ r_t = A_3 + \delta_{31}C_{t-1} + \delta_{32}I_{t-1} + \delta_{33}M_t + \delta_{34}G_t + u_3, \\ Y_t = A_4 + \delta_{41}C_{t-1} + \delta_{42}I_{t-1} + \delta_{43}M_t + \delta_{44}G_t + u_1. \end{cases}$$

4.4. Методы оценки параметров структурной формы модели

Коэффициенты структурной модели могут быть оценены разными способами в зависимости от вида системы одновременных уравнений. Наибольшее распространение в литературе получили следующие методы оценивания коэффициентов структурной модели:

1) косвенный метод наименьших квадратов;
2) двухшаговый метод наименьших квадратов;
3) трехшаговый метод наименьших квадратов;
4) метод максимального правдоподобия с полной информацией;
5) метод максимального правдоподобия при ограниченной информации.

Рассмотрим вкратце сущность каждого из этих методов.

Косвенный метод наименьших квадратов (КМНК) применяется в случае точно идентифицируемой структурной модели. Процедура применения КМНК предполагает выполнение следующих этапов работы.

1. Структурная модель преобразовывается в приведенную форму модели.

2. Для каждого уравнения приведенной формы модели обычным МНК оцениваются приведенные коэффициенты δ_{ij}.

3.Коэффициенты приведенной формы модели трансформируются в параметры структурной модели.

Если система сверхидентифицируема, то КМНК не используется, ибо он не дает однозначных оценок для параметров структурной модели. В этом случае могут использоваться разные методы оценивания, среди которых наиболее распространенным и простым является двухшаговый метод наименьших квадратов (ДМНК).

Основная идея ДМНК – на основе приведенной формы модели получить для сверхидентифицируемого уравнения теоретические значения эндогенных переменных, содержащихся в правой части уравнения.

Далее, подставив их вместо фактических значений, можно применить обычный МНК к структурной форме сверхидентифицируемого уравнения. Метод получил название двухшагового МНК, ибо дважды используется МНК: на первом шаге при определении приведенной формы модели и нахождении на ее основе оценок теоретических значений эндогенной переменной $y_i = \delta_{i1}x_1 + \delta_{i2}x_2 + \ldots + \delta_{in}x_n$ и на втором шаге применительно к структурному сверхидентифицируемому уравнению при определении структурных коэффициентов модели по данным теоретических (расчетных) значений эндогенных переменных.

Сверхидентифицируемая структурная модель может быть двух типов:

1) все уравнения системы сверхидентифицируемы;
2) система содержит наряду со сверхидентифицируемыми точно идентифицируемые уравнения.

Если все уравнения системы сверхидентифицируемые, то для оценки структурных коэффициентов каждого уравнения используется ДМНК. Если в системе есть точно идентифицируемые уравнения, то структурные коэффициенты по ним находятся из системы приведенных уравнений.

Для примера, рассмотренного в предыдущем параграфе, необходимо применить именно двухшаговый метод наименьших квадратов. Но можно сделать следующее замечание. Если из модели исключить тождество дохода, число эндогенных переменных модели снизится на единицу – переменная Y_t станет экзогенной. А число предопределенных переменных модели не изменится, т.к. из модели будет исключена эндогенная переменная G_t, но ее место займет переменная Y_t. В правых частях функции потребления и функции денежного рынка будут находиться только предопределенные переменные. Функция инвестиций постулирует зависимость эндогенной переменной I_t от эндогенной переменной r_t (которая зависит только от предопределенных переменных) и предопределенной переменной I_{t-1}. Таким образом, мы получим рекурсивную систему. Ее параметры можно оценивать обычным МНК, и нет необходимости исследования уравнения на идентификацию.

Косвенный и двухшаговый методы наименьших квадратов подробно описаны в литературе и рассматриваются как традиционные методы оценки коэффициентов структурной модели. Эти методы достаточно легко реализуемы.

Метод максимального правдоподобия рассматривается как наиболее общий метод оценивания, результаты которого при нормальном распределении признаков совпадают с МНК. Однако при большом числе уравнений системы этот метод приводит к достаточно сложным вычислительным процедурам. Поэтому в качестве модификации используется метод максимального правдоподобия при ограниченной информации (метод наименьшего дисперсионного отношения), разработанный в 1949 г. Т.Андерсоном и Н.Рубиным.

В отличие от метода максимального правдоподобия в данном методе сняты ограничения на параметры, связанные с функционированием системы в целом. Это делает решение более простым, но трудоемкость вычислений остается достаточно высокой. Несмотря на его значительную популярность, к середине 60-х годов он был практически вытеснен двухшаговым методом наименьших квадратов (ДМНК) в связи с гораздо большей простотой последнего.

Дальнейшим развитием ДМНК является трехшаговый МНК (ТМНК), предложенный в 1962 г. А.Зельнером и Г.Тейлом. Этот метод оценивания пригоден для всех видов уравнений структурной модели. Однако при некоторых ограничениях на параметры более эффективным оказывается ДМНК.

Вопросы для самоконтроля

1. В каких случаях модель строится в виде систем эконометрических уравнений?
2. Какие проблемы возникают при оценке параметров систем эконометрических уравнений?
3. Какие переменные называются эндогенными, экзогенными, лаговыми и предопределенными?
4. Что представляет собой структурная форма модели?
5. Что представляет собой приведенная форма модели?
6. В чем заключается проблема идентифицируемости модели?
7. Как проверяется идентифицируемость уравнений модели?
8. Какие методы применяются для нахождения структурных коэффициентов модели для различных видов систем уравнений?
9. Что представляет собой косвенный МНК?
10. Что представляет собой двухшаговый МНК?
11. Что представляет собой трехшаговый МНК?

Тестовые задания

1. В системе независимых уравнений каждое уравнение представлено

а) изолированным уравнением регрессии

б) поведенческим уравнением

в) уравнением, не содержащим параметров и определяющим фиксированные отношения между переменными

2. Относительно системы $\begin{cases} y_1 = a_{12}y_2 + b_{11}x_1 + \varepsilon_1 \\ y_2 = a_{21}y_1 + b_{21}x_1 + \varepsilon_2 \end{cases}$ *верно следующее утверждение: система записана в … форме*

а) приведенной

б) рекурсивной

в) структурной

г) нормальной

3. Число приведенных коэффициентов системы одновременных уравнений равно числу структурных коэффициентов, тогда модель …

а) сверхидентифицируема

б) идентифицируема

в) неидентифицируема

г) не существует

4. Оценки параметров сверхидентифицируемой системы экономических уравнений могут быть найдены с помощью … метода наименьших квадратов…

а) обычного

б) взвешенного

в) двухшагового

г) косвенного

5. Уравнение в котором H – число эндогенных переменных, D – число отсутствующих экзогенных переменных, идентифицируемо, если …

а) $D + 1 = H$

б) $D + 1 > H$

в) $D + 1 < H$

6. Уравнение в котором H число эндогенных переменных, D число отсутствующих экзогенных переменных, неидентифицируемо, если …

а) $D + 1 = H$

б) $D + 1 > H$

в) $D + 1 < H$

7. Уравнение в котором H число эндогенных переменных, D число отсутствующих экзогенных переменных, сверхидентифицируемо, если …

а) $D + 1 = H$

б) $D + 1 > H$

в) $D + 1 < H$

8. Структурной формой модели называется система ...

а) уравнений с фиксированным набором факторов;

б) взаимосвязанных уравнений;

в) независимых уравнений;

г) рекурсивных уравнений.

9. Оценки параметров точно идентифицируемой системы экономических уравнений могут быть найдены с помощью ... метода наименьших квадратов...

а) обычного

б) взвешенного

в) двухшагового

г) косвенного

10. Сверхидентифицируемая система экономических уравнений ...

а) не имеет решений

б) имеет множество решений

в) имеет одно решение

11. Точно идентифицируемая система экономических уравнений ...

а) не имеет решений

б) имеет множество решений

в) имеет одно решение

12. Неидентифицируемая система экономических уравнений ...

а) не имеет решений

б) имеет множество решений

в) имеет одно решение

13. Система эконометрических уравнений является сверхидентифицируемой, если ...

а) в системе присутствуют неидентифицируемые и сверхидентифицируемые уравнения

б) сверхидентифицируемо хотя бы одно уравнение системы

в) в системе присутствуют идентифицируемые и сверхидентифицируемые уравнения

14. Структурная форма модели используется, если ...

а) есть взаимное влияние между всеми переменными модели

б) зависимая переменная является функцией только предопределенных переменных

в) каждая зависимая переменная является функцией только предопределенных переменных и зависимых переменных, определенных в предыдущих уравнениях системы

5. МОДЕЛИРОВАНИЕ ОДНОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ И ПРОГНОЗИРОВАНИЕ

5.1. Понятие временных рядов

При построении эконометрической модели используются два типа данных:

1) данные, характеризующие совокупность различных объектов в определенный момент времени;

2) данные, характеризующие один объект за ряд последовательных моментов времени.

Модели, построенные по данным первого типа, называются *пространственными моделями*. Модели, построенные на основе второго типа данных, называются *моделями временных рядов*.

Временной ряд (ряд динамики) – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени. Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы:

1) факторы, формирующие тенденцию ряда;
2) факторы, формирующие циклические колебания ряда;
3) случайные факторы.

Рассмотрим воздействие каждого фактора на временной ряд в отдельности.

Большинство временных рядов экономических показателей имеют тенденцию, характеризующую совокупное долговременное воздействие множества факторов на динамику изучаемого показателя. Все эти факторы, взятые в отдельности, могут оказывать разнонаправленное воздействие на исследуемый показатель. Однако в совокупности они формируют его возрастающую или убывающую тенденцию.

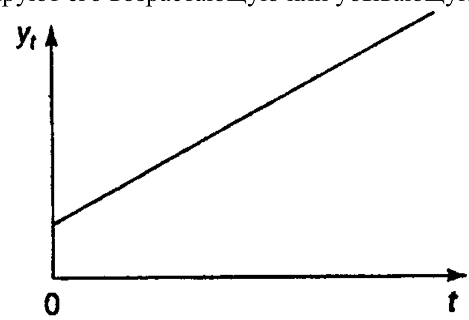


Рис. 5.1. Временной ряд, содержащий возрастающую тенденцию

На рис. 5.1 показан гипотетический временной ряд, содержащий возрастающую тенденцию.

Также изучаемый показатель может быть подвержен циклическим колебаниям. Эти колебания могут носить сезонный характер, поскольку экономическая деятельность ряда отраслей экономики зависит от времени года (например, цены на сельскохозяйственную продукцию в летний период выше, чем в зимний; уровень безработицы в курортных городах в зимний период выше по сравнению с летним). При наличии больших массивов данных за длительные промежутки времени можно выявить циклические колебания, связанные с общей динамикой конъюнктуры рынка. На рис. 4.2 представлен гипотетический временной ряд, содержащий только сезонную компоненту.

Рис. 5.2. Временной ряд, содержащий только сезонную компоненту

Некоторые временные ряды не содержат тенденции и циклической компоненты, а каждый следующий их уровень образуется как сумма среднего уровня ряда и некоторой (положительной или отрицательной) случайной компоненты. Пример ряда, содержащего только случайную компоненту, приведен на рис. 5.3.

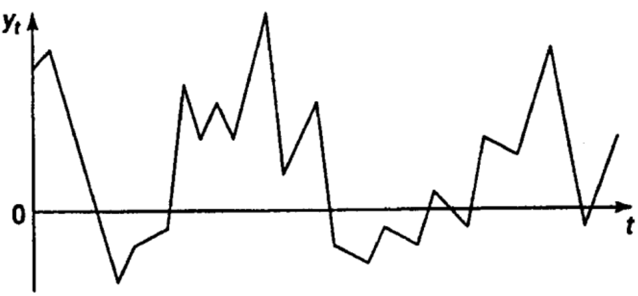


Рис. 5.3. Временной ряд, содержащий только случайную компоненту

Очевидно, что реальные данные не следуют целиком и полностью из каких-либо описанных выше моделей. Чаще всего они содержат все три компоненты. Каждый их уровень формируется под воздействием тенденции, сезонных колебаний и случайной компоненты.

В большинстве случаев фактический уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент. Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется *аддитивной моделью* временного ряда. Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется *мультипликативной моделью* временного ряда. Основная задача эконометрического исследования отдельного временного ряда – выявление и придание количественного выражения каждой из перечисленных выше компонент с тем, чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

5.2. Автокорреляция уровней временного ряда

При наличии во временном ряде тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих. Корреляционную зависимость между последовательными уровнями временного ряда называют автокорреляцией уровней ряда.

Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Формула для расчета коэффициента автокорреляции имеет вид:

$$r_1 = \frac{\sum_{t=2}^{n}\left(y_t - \bar{y}_1\right)\left(y_{t-1} - \bar{y}_2\right)}{\sqrt{\sum_{t=2}^{n}\left(y_t - \bar{y}_1\right)^2 \sum_{t=2}^{n}\left(y_{t-1} - \bar{y}_2\right)^2}}, \tag{5.1}$$

$$\bar{y}_1 = \frac{1}{n-1}\sum_{t=2}^{n} y_t, \qquad \bar{y}_2 = \frac{1}{n-1}\sum_{t=2}^{n} y_{t-1}.$$

Эту величину называют коэффициентом автокорреляции уровней ряда первого порядка, так как он измеряет зависимость между соседними уровнями ряда t и y_{t-1}.

Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков. Так, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями y_t и y_{t-2} и определяется по формуле:

$$r_2 = \frac{\sum_{t=3}^{n}\left(y_t - \bar{y}_3\right)\left(y_{t-2} - \bar{y}_4\right)}{\sqrt{\sum_{t=3}^{n}\left(y_t - \bar{y}_3\right)^2 \sum_{t=3}^{n}\left(y_{t-2} - \bar{y}_4\right)^2}}, \qquad (5.2)$$

$$\bar{y}_3 = \frac{1}{n-2}\sum_{t=3}^{n} y_t, \qquad \bar{y}_4 = \frac{1}{n-2}\sum_{t=3}^{n} y_{t-2}.$$

Число периодов, по которым рассчитывается коэффициент автокорреляции, называют *лагом*. С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается. Считается целесообразным для обеспечения статистической достоверности коэффициентов автокорреляции использовать правило – максимальный лаг должен быть не больше $n/4$.

Свойства коэффициента автокорреляции.

1. Он строится по аналогии с линейным коэффициентом корреляции и таким образом характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. Поэтому по коэффициенту автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию (например, параболу второго порядка или экспоненту), коэффициент автокорреляции уровней исходного ряда может приближаться к нулю.
2. По знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных содержат положительную автокорреляцию уровней, однако при этом могут иметь убывающую тенденцию.

Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют *автокорреляционной функцией* временного ряда. График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется *коррелограммой*.

Анализ автокорреляционной функции и коррелограммы позволяет определить лаг, при котором автокорреляция наиболее высокая, а следовательно, и лаг, при котором связь между текущим и предыдущими уровнями ряда наиболее тесная, т.е. при помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда.

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка τ, то ряд содержит циклические колебания с периодичностью в τ моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать одно из двух предположений относительно структуры этого ряда: либо ряд не содержит тенденции и циклических колебаний, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ. Поэтому коэффициент автокорреляции уровней и автокорреляционную функцию целесообразно использовать для выявления во временном ряде наличия или отсутствия трендовой компоненты и циклической (сезонной) компоненты.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ. Пусть имеются некоторые условные данные об общем количестве правонарушений на таможне одного из субъектов РФ.

Таблица 5.1

Год	Квартал	t	Количество возбужденных дел, y_t
1	I	1	375
	II	2	371
	III	3	869
	IV	4	1015
2	I	5	357
	II	6	471
	III	7	992
	IV	8	1020
3	I	9	390
	II	10	355
	III	11	992
	IV	12	905
4	I	13	461
	II	14	454
	III	15	920
	IV	16	927

Построим поле корреляции:

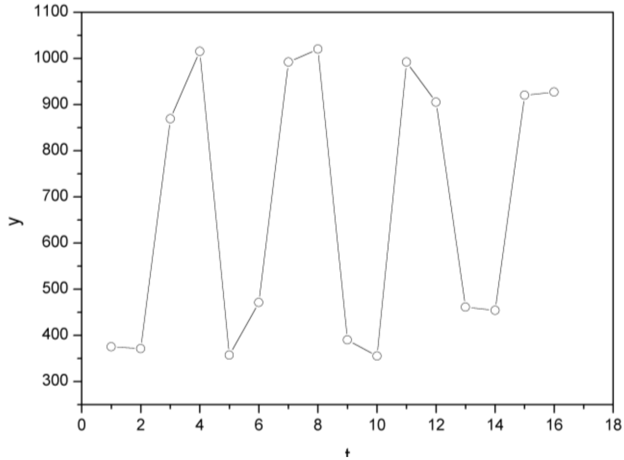


Рис. 5.4. Поле корреляции

Уже исходя из графика видно, что значения y образуют пилообразную фигуру. Рассчитаем несколько последовательных коэффициентов автокорреляции. Для этого составляем первую вспомогательную таблицу.

Следует заметить, что среднее значение получается путем деления не на 16, а на 15, т.к. у нас теперь на одно наблюдение меньше.

Составляем вспомогательную таблицу (5.2) для расчета коэффициента автокорреляции первого порядка.

Теперь вычисляем коэффициент автокорреляции первого порядка по формуле (5.1):

$$r_1 = \frac{74085,16}{\sqrt{1153756,39 \cdot 1187469,73}} = 0,063294 .$$

Составляем вспомогательную таблицу (5.3) для расчета коэффициента автокорреляции второго порядка.

Следовательно

$$r_2 = \frac{-1034792,71}{\sqrt{1037835,43 \cdot 1116776,36}} = -0,961183$$

Все полученные значения заносим в сводную таблицу (5.4).

Таблица 5.2

t	y_t	y_{t-1}	$y_t - \overline{y}_1$	$y_{t-1} - \overline{y}_2$	$(y_t - \overline{y}_1)\times$ $\times(y_{t-1} - \overline{y}_2)$	$(y_t - \overline{y}_1)^2$	$(y_{t-1} - \overline{y}_2)^2$
1	375	–	–	–	–	–	–
2	371	375	-328,33	-288,13	94601,72	107800,59	83018,90
3	869	371	169,67	-292,13	-49565,70	28787,91	85339,94
4	1015	869	315,67	205,87	64986,98	99647,55	42382,46
5	357	1015	-342,33	351,87	-120455,66	117189,83	123812,50
6	471	357	-228,33	-306,13	69898,66	52134,59	93715,58
7	992	471	292,67	-192,13	-56230,69	85655,73	36913,94
8	1020	992	320,67	328,87	105458,74	102829,25	108155,48
9	390	1020	-309,33	356,87	-110390,60	95685,05	127356,20
10	355	390	-344,33	-273,13	94046,85	118563,15	74600,00
11	992	355	292,67	-308,13	-90180,41	85655,73	94944,10
12	905	992	205,67	328,87	67638,69	42300,15	108155,48
13	461	905	-238,33	241,87	-57644,88	56801,19	58501,10
14	454	461	-245,33	-202,13	49588,55	60186,81	40856,54
15	920	454	220,67	-209,13	-46148,72	48695,25	43735,36
16	927	920	227,67	256,87	58481,59	51833,63	65982,20
Сумма	10499	9947	9,05	0,05	74085,16	1153766,39	1187469,73
Среднее значение	699,33	663,13	–	–	–	–	–

Таблица 5.3

t	y_t	y_{t-2}	$y_t - \bar{y}_3$	$y_{t-2} - \bar{y}_4$	$(y_t - \bar{y}_3)\times$ $\times(y_{t-2} - \bar{y}_4)$	$(y_t - \bar{y}_3)^2$	$(y_{t-2} - \bar{y}_4)^2$
1	375	–	–	–	–	–	–
2	371	–	–	–	–	–	–
3	869	375	145,57	-269,79	-39273,33	21190,62	72786,64
4	1015	371	291,57	-273,79	-79828,95	85013,06	74960,96
5	357	869	-366,43	224,21	-82157,27	134270,94	50270,12
6	471	1015	-252,43	370,21	-93452,11	63720,90	137055,44
7	992	357	268,57	-287,79	-77291,76	72129,84	82823,08
8	1020	471	296,57	-173,79	-51540,90	87953,76	30202,96
9	390	992	-333,43	347,21	-115770,23	111175,56	120554,78
10	355	1020	-368,43	375,21	-138238,62	135740,66	140782,54
11	992	390	268,57	-254,79	-68428,95	72129,84	64917,94
12	905	355	181,57	-289,79	-52617,17	32967,66	83978,24
13	461	992	-262,43	347,21	-91118,32	68869,50	120554,78
14	454	905	-269,43	260,21	-70108,38	72592,52	67709,24
15	920	461	196,57	-183,79	-36127,60	38639,76	33778,76
16	927	454	203,57	-190,79	-38839,12	41440,74	36400,82
Сумма	10128	9027	-0,02	-0,06	-1034792,71	1037835,43	1116776,36
Среднее значение	723,43	644,79	–	–	–	–	–

Таблица 5.4

Лаг	Коэффициент автокорреляции уровней
1	0,063294
2	–0,961183
3	–0,036290
4	0,964735
5	0,050594
6	–0,976516
7	–0,069444
8	0,964629
9	0,162064
10	-0,972918
11	-0,065323
12	0,985761

Строим график.

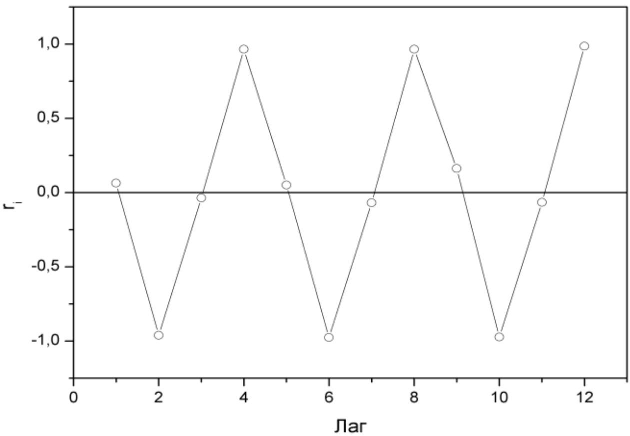


Рис. 5.5. Коррелограмма

Анализ коррелограммы и графика исходных уровней временного ряда позволяет сделать вывод о наличии в изучаемом временном ряде сезонных колебаний периодичностью в четыре квартала.

5.3. Моделирование тенденции временного ряда

Распространенным способом моделирования тенденции временного ряда является построение аналитической функции, характеризующей зависимость уровней ряда от времени, или тренда. Этот способ называют *аналитическим выравниванием временного ряда*.

Поскольку зависимость от времени может принимать разные формы, для ее формализации можно использовать различные виды функций. Для построения трендов чаще всего применяются следующие функции:

линейный тренд: $y_t = a + b \cdot t$;

гипербола: $y_t = a + b/t$;

экспоненциальный тренд: $y_t = \mathrm{e}^{a+b\cdot t}$ (или $y_t = a \cdot b^t$);

степенная функция: $y_t = a \cdot t^b$;

полиномы различных степеней:

$y_t = a + b_1 \cdot t + b_2 \cdot t^2 + ... + b_m \cdot t^m$.

Параметры каждого из перечисленных выше трендов можно определить обычным МНК, используя в качестве независимой переменной время $t = 1, 2, ..., n$, а в качестве зависимой переменной – фактические уровни временного ряда y_t. Для нелинейных трендов предварительно проводят стандартную процедуру их линеаризации.

Существует несколько способов определения типа тенденции. К числу наиболее распространенных способов относятся качественный анализ изучаемого процесса, построение и визуальный анализ графика зависимости уровней ряда от времени. В этих же целях можно использовать и коэффициенты автокорреляции уровней ряда. Тип тенденции можно определить путем сравнения коэффициентов автокорреляции первого порядка, рассчитанных по исходным и преобразованным уровням ряда. Если временной ряд имеет линейную тенденцию, то его соседние уровни y_t и y_{t-1} тесно коррелируют. В этом случае коэффициент автокорреляции первого порядка уровней исходного ряда должен быть высоким. Если временной ряд содержит нелинейную тенденцию, например, в форме экспоненты, то коэффициент автокорреляции первого порядка по логарифмам уровней исходного ряда будет выше, чем соответствующий коэффициент, рассчитанный по уровням ряда. Чем сильнее выражена нелинейная тенденция в изучаемом временном ряде, тем в большей степени будут различаться значения указанных коэффициентов.

Выбор наилучшего уравнения в случае, когда ряд содержит нелинейную тенденцию, можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации и средней ошибки аппроксимации. Этот метод легко реализуется при компьютерной обработке данных.

Простейший подход к моделированию сезонных колебаний – это расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда.

Общий вид аддитивной модели следующий:

$$Y = T + S + E. \qquad (5.3)$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как сумма трендовой (T), сезонной (S) и случайной (E) компонент.

Общий вид мультипликативной модели выглядит так:

$$Y = T \cdot S \cdot E. \qquad (5.4)$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как произведение трендовой (T), сезонной (S) и случайной (E) компонент.

Выбор одной из двух моделей осуществляется на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Построение аддитивной и мультипликативной моделей сводится к расчету значений T, S и E для каждого уровня ряда.

Процесс построения модели включает в себя следующие шаги.

1) Выравнивание исходного ряда методом скользящей средней.

2) Расчет значений сезонной компоненты S.

3) Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных ($T + E$) в аддитивной или ($T \cdot E$) в мультипликативной модели.

4) Аналитическое выравнивание уровней ($T + E$) или ($T \cdot E$) и расчет значений T с использованием полученного уравнения тренда.

5) Расчет полученных по модели значений ($T + E$) или ($T \cdot E$).

6) Расчет абсолютных и/или относительных ошибок. Если полученные значения ошибок не содержат автокорреляции, ими можно

заменить исходные уровни ряда и в дальнейшем использовать временной ряд ошибок E для анализа взаимосвязи исходного ряда и других временных рядов.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ. Построение аддитивной модели временного ряда. Обратимся к данным об объеме правонарушений на таможне за четыре года, представленным в табл. 5.5.

Было показано, что данный временной ряд содержит сезонные колебания периодичностью 4, т.к. количество правонарушений в первый-второй кварталы ниже, чем в третий-четвертый. Рассчитаем компоненты аддитивной модели временного ряда.

Шаг 1. Проведем выравнивание исходных уровней ряда методом скользящей средней. Для этого:

1.1. Просуммируем уровни ряда последовательно за каждые четыре квартала со сдвигом на один момент времени и определим условные годовые объемы потребления электроэнергии (гр. 3 табл. 5.5).

1.2. Разделив полученные суммы на 4, найдем скользящие средние (гр. 4 табл. 5.5). Полученные таким образом выровненные значения уже не содержат сезонной компоненты.

1.3. Приведем эти значения в соответствие с фактическими моментами времени, для чего найдем средние значения из двух последовательных скользящих средних – центрированные скользящие средние (гр. 5 табл. 5.5).

Шаг 2. Найдем оценки сезонной компоненты как разность между фактическими уровнями ряда и центрированными скользящими средними (гр. 6 табл. 5.5). Используем эти оценки для расчета значений сезонной компоненты S (табл. 5.6). Для этого найдем средние за каждый квартал (по всем годам) оценки сезонной компоненты S_i. В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются. В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна нулю.

Для данной модели имеем:

$-289{,}667 - 264 + 271{,}417 + 293{,}375 = 11{,}125$.

Корректирующий коэффициент: $k = 11{,}125/4 = 2{,}781$.

Рассчитываем скорректированные значения сезонной компоненты ($S_i = \bar{S}_i - k$) и заносим полученные данные в таблицу 5.6.

Проверим равенство нулю суммы значений сезонной компоненты: $-292{,}448 - 266{,}781 + 268{,}636 + 290{,}593 = 0$.

Таблица 5.5

№ квартала, t	Количество правонарушений, y_t	Итого за четыре квартала	Скользящая средняя за четыре квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	375	–	–	–	–
2	371	2630	657,5	–	–
3	869	2612	653	655,25	213,75
4	1015	2712	678	665,5	349,5
5	357	2835	708,75	693,75	-336,75
6	471	2840	710	709,375	-238,375
7	992	2873	718,25	714,125	277,875
8	1020	2757	689,25	703,75	316,25
9	390	2757	689,25	689,25	-299,25
10	355	2642	660,5	674,875	-319,875
11	992	2713	678,25	669,375	322,625
12	905	2812	703	690,625	214,375
13	461	2740	685	694	-233
14	454	2762	690,5	687,75	-233,75
15	920	–	–	–	–
16	927	–	–	–	–

Таблица 5.6

Показатели	Год	№ квартала, i			
		I	II	III	IV
	1	–	–	213,75	349,5
	2	-336,75	-238,375	277,875	316,25
	3	-299,25	-319,875	322,625	214,375
	4	-233	-233,75	–	–
Всего за i-й квартал		-869	-792	814,25	880,125
Средняя оценка сезонной компоненты для i-го квартала, $\bar{S}_i$		-289,667	-264	271,417	293,375
Скорректированная сезонная компонента, S_i		-292,448	-266,781	268,636	290,593

Шаг 3. Исключим влияние сезонной компоненты, вычитая ее значение из каждого уровня исходного временного ряда. Получим величины $T+E=Y-S$ (гр. 4 табл. 5.7). Эти значения рассчитываются за каждый момент времени и содержат только тенденцию и случайную компоненту.

Шаг 4. Определим компоненту T данной модели. Для этого проведем аналитическое выравнивание ряда ($T+E$) с помощью линейного тренда. Результаты аналитического выравнивания следующие: $T=671{,}777+0{,}9233\cdot t$.

Подставляя в это уравнение значения $t=1, 2, ..., 16$, найдем уровни T для каждого момента времени (гр. 5 табл. 5.7).

Шаг 5. Найдем значения уровней ряда, полученные по аддитивной модели. Для этого прибавим к уровням T значения сезонной компоненты для соответствующих кварталов (гр. 6 табл. 5.7).

На одном графике отложим фактические значения уровней временного ряда и теоретические, полученные по аддитивной модели.

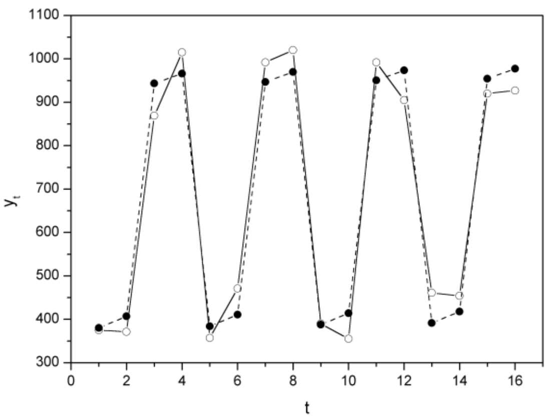


Рис. 5.6. Коррелограмма

Для оценки качества построенной модели применим сумму квадратов полученных абсолютных ошибок.

$$R^2=1-\frac{E^2}{\left(y_t-\bar{y}\right)^2}=1-\frac{37911{,}973}{1252743{,}75}=0{,}970\,.$$

Таблица 5.7

t	y_t	S_i	$y_t - S_i$	T	$T+S$	$E = y_t - (T+S)$	E^2
1	375	-292,448	667,448	672,700	380,252	-5,252	27,584
2	371	-266,781	637,781	673,624	406,843	-35,843	1284,721
3	869	268,636	600,364	674,547	943,183	-74,183	5503,117
4	1015	290,593	724,407	675,470	966,063	48,937	2394,830
5	357	-292,448	649,448	676,394	383,946	-26,946	726,087
6	471	-266,781	737,781	677,317	410,536	60,464	3655,895
7	992	268,636	723,364	678,240	946,876	45,124	2036,175
8	1020	290,593	729,407	679,163	969,756	50,244	2524,460
9	390	-292,448	682,448	680,087	387,639	2,361	5,574
10	355	-266,781	621,781	681,010	414,229	-59,229	3508,074
11	992	268,636	723,364	681,933	950,569	41,431	1716,528
12	905	290,593	614,407	682,857	973,450	-68,450	4685,403
13	461	-292,448	753,448	683,780	391,332	69,668	4853,630
14	454	-266,781	720,781	684,703	417,922	36,078	1301,622
15	920	268,636	651,364	685,627	954,263	-34,263	1173,953
16	927	290,593	636,407	686,550	977,143	-50,143	2514,320

Следовательно, можно сказать, что аддитивная модель объясняет 97% общей вариации уровней временного ряда количества правонарушений по кварталам за 4 года.

Шаг 6. Прогнозирование по аддитивной модели. Предположим, что по нашему примеру необходимо дать прогноз об общем объеме правонарушений на I и II кварталы 5 года. Прогнозное значение F_t уровня временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент. Для определения трендовой компоненты воспользуемся уравнением тренда

$$T = 671,777 + 0,9233 \cdot t \cdot$$

Получим

$$T_{17} = 671,777 + 0,9233 \cdot 17 = 687,473;$$

$$T_{18} = 671,777 + 0,9233 \cdot 18 = 688,396 \cdot$$

Значения сезонных компонент за соответствующие кварталы равны: $S_1 = -292,448$ и $S_2 = -266,781$. Таким образом,

$$F_{17} = T_{17} + S_1 = 687,473 - 292,448 \approx 395;$$

$$F_{18} = T_{18} + S_2 = 688,396 - 266,781 \approx 422 \cdot$$

Т.е. в первые два квартала следующего года следует ожидать порядка 395 и 422 правонарушений соответственно.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ. Построение мультипликативной модели рассмотрим на данных предыдущего примера.

Шаг 1. Методика, применяемая на этом шаге, полностью совпадает с методикой построения аддитивной модели.

Шаг 2. Найдем оценки сезонной компоненты как частное от деления фактических уровней ряда на центрированные скользящие средние (гр. 6 табл. 5.8). Эти оценки используются для расчета сезонной компоненты S (табл. 5.9). Для этого найдем средние за каждый квартал оценки сезонной компоненты S_i. Так же как и в аддитивной модели считается, что сезонные воздействия за период взаимопогашаются. В мультипликативной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна числу периодов в цикле. В нашем случае число периодов одного цикла равно 4.

Имеем $0,5816 + 0,6167 + 1,3991 + 1,4283 = 4,0257 \cdot$

Определяем корректирующий коэффициент:

$$k = 4/4,0257 = 0,9936 \cdot$$

Таблица 5.8

№ квар-тала, t	Количество правонару-шений, y_t	Итого за четы-ре квар-тала	Сколь-зящая средняя за четы-ре квар-тала	Центриро-ванная скользящая средняя	Оценка сезон-ной компо-ненты
1	375	–	–	–	–
2	371	2630	657,5	–	–
3	869	2612	653	655,25	1,3262
4	1015	2712	678	665,5	1,5252
5	357	2835	708,75	693,75	0,5146
6	471	2840	710	709,375	0,6640
7	992	2873	718,25	714,125	1,3891
8	1020	2757	689,25	703,75	1,4494
9	390	2757	689,25	689,25	0,5658
10	355	2642	660,5	674,875	0,5260
11	992	2713	678,25	669,375	1,4820
12	905	2812	703	690,625	1,3104
13	461	2740	685	694	0,6643
14	454	2762	690,5	687,75	0,6601
15	920	–	–	–	–
16	927	–	–	–	–

Таблица 5.9

Показатели	Год	№ квартала, i			
		I	II	III	IV
	1	–	–	1,3262	1,5252
	2	0,5146	0,6640	1,3891	1,4494
	3	0,5658	0,5260	1,4820	1,3104
	4	0,6643	0,6601	–	–
Всего за i-й квартал		1,7447	1,8501	4,1973	4,2850
Средняя оценка сезонной компоненты для i-го квартала, $\overline{S}_i$		0,5816	0,6167	1,3991	1,4283
Скорректированная сезонная компонента, S_i		0,5779	0,6128	1,3901	1,4192

Скорректированные значения сезонной компоненты S_i получаются при умножении ее средней оценки $\overline{S}_i$ на корректирующий коэффициент k.

Проверяем условие равенство 4 суммы значений сезонной компоненты:

$$0,5779+0,6128+1,3901+1,4192=4\cdot$$

Шаг 3. Разделим каждый уровень исходного ряда на соответствующие значения сезонной компоненты. В результате получим величины $T\cdot E=Y/S$ (гр. 4 табл. 5.10), которые содержат только тенденцию и случайную компоненту.

Шаг 4. Определим компоненту T в мультипликативной модели. Для этого рассчитаем параметры линейного тренда, используя уровни $T\cdot E$. В результате получим уравнение тренда:

$$T=651,6364+3,2809\cdot t\cdot$$

Подставляя в это уравнение значения $t=1, 2, ..., 16$, найдем уровни T для каждого момента времени (гр. 5 табл. 5.10).

Шаг 5. Найдем уровни ряда, умножив значения T на соответствующие значения сезонной компоненты (гр. 6 табл. 4.10). На одном графике откладываем фактические значения уровней временного ряда и теоретические, полученные по мультипликативной модели.

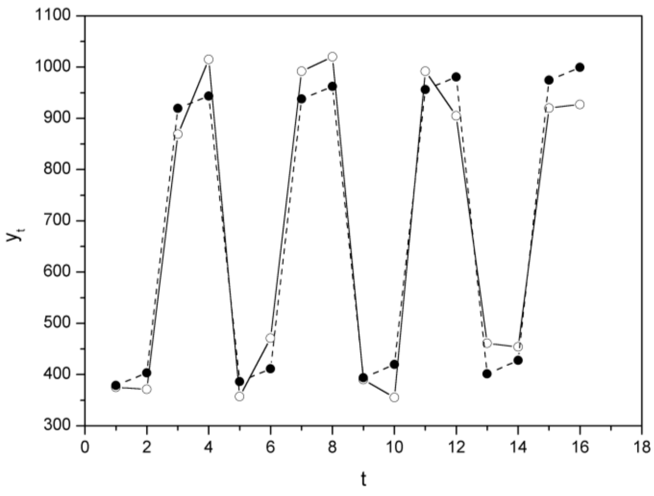


Рис. 5.7. Коррелограмма

Таблица 5.10

t	y_t	S_i	y_t/S_i	T	$T \cdot S$	$E = y_t/(T \cdot S)$
1	**2**	**3**	**4**	**5**	**6**	**7**
1	375	0,5779	648,9012	654,9173	378,4767	0,9908
2	371	0,6128	605,4178	658,1982	403,3439	0,9198
3	869	1,3901	625,1349	661,4791	919,5221	0,9451
4	1015	1,4192	715,1917	664,7600	943,4274	1,0759
5	357	0,5779	617,7539	668,0409	386,0608	0,9247
6	471	0,6128	768,6031	671,3218	411,3860	1,1449
7	992	1,3901	713,6177	674,6027	937,7652	1,0578
8	1020	1,4192	718,7148	677,8836	962,0524	1,0602
9	390	0,5779	674,8572	681,1645	393,6450	0,9907
10	355	0,6128	579,3081	684,4454	419,4281	0,8464
11	992	1,3901	713,6177	687,7263	956,0083	1,0377
12	905	1,4192	637,6832	691,0072	980,6774	0,9228
13	461	0,5779	797,7159	694,2881	401,2291	1,1490
14	454	0,6128	740,8616	697,5690	427,4703	1,0621
15	920	1,3901	661,8229	700,8499	974,2515	0,9443
16	927	1,4192	653,1849	704,1308	999,3024	0,9277

Расчет ошибки в мультипликативной модели производится по формуле: $E = Y/(T \cdot S)$.

Для сравнения мультипликативной модели и других моделей временного ряда можно, по аналогии с аддитивной моделью, использовать сумму квадратов абсолютных ошибок $(y_t - T \cdot S)^2$:

$$R^2 = 1 - \frac{(y_t - T \cdot S)^2}{(y_t - \bar{y})^2} = 1 - \frac{43065{,}02}{1252743{,}75} = 0{,}966.$$

Сравнивая показатели детерминации аддитивной и мультипликативной моделей, делаем вывод, что они примерно одинаково аппроксимируют исходные данные.

Шаг 6. Прогнозирование по мультипликативной модели. Если предположить, что по нашему примеру необходимо дать прогноз об общем объеме правонарушений на I и II кварталы 2003 года, прогнозное значение F_t уровня временного ряда в мультипликативной модели есть произведение трендовой и сезонной компонент. Для определения трендовой компоненты воспользуемся уравнением тренда

$$T = 651{,}6364 + 3{,}2809 \cdot t.$$

Получим

$$T_{17} = 651{,}6364 + 3{,}2809 \cdot 17 = 707{,}4117;$$

$$T_{18} = 651{,}6364 + 3{,}2809 \cdot 18 = 710{,}6926.$$

Значения сезонных компонент за соответствующие кварталы равны: $S_1 = 0{,}5779$ и $S_2 = 0{,}6128$. Таким образом

$$F_{17} = T_{17} \cdot S_1 = 707{,}4117 \cdot 0{,}5779 \approx 409;$$

$$F_{18} = T_{18} \cdot S_2 = 710{,}6926 \cdot 0{,}6128 \approx 436.$$

Т.е. в первые два квартала 2003 г. следовало ожидать порядка 409 и 436 правонарушений соответственно.

Таким образом, аддитивная и мультипликативная модели дают примерно одинаковый результат по прогнозу.

5.4. Автокорреляция в остатках. Критерий Дарбина-Уотсона

Автокорреляция в остатках может быть вызвана несколькими причинами, имеющими различную природу.

1. Она может быть связана с исходными данными и вызвана наличием ошибок измерения в значениях результативного признака.
2. В ряде случаев автокорреляция может быть следствием неправильной спецификации модели. Модель может не включать фактор, который оказывает существенное воздействие на результат и влияние которого отражается в остатках, вследствие чего последние могут оказаться автокоррелированными. Очень часто этим фактором является фактор времени t.

От истинной автокорреляции остатков следует отличать ситуации, когда причина автокорреляции заключается в неправильной спецификации функциональной формы модели. В этом случае следует изменить форму модели, а не использовать специальные методы расчета параметров уравнения регрессии при наличии автокорреляции в остатках.

Один из более распространенных методов определения автокорреляции в остатках – это расчет критерия Дарбина-Уотсона:

$$d = \frac{\sum_{t=2}^{n}\left(\varepsilon_t - \varepsilon_{t-1}\right)^2}{\sum_{t=1}^{n}\varepsilon_t^2}. \qquad (5.5)$$

Т.е. величина d есть отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели регрессии.

Можно показать, что при больших значениях n существует следующее соотношение между критерием Дарбина-Уотсона d и коэффициентом автокорреляции остатков первого порядка r_1:

$$d = 2\cdot\left(1 - r_1\right)\cdot \qquad (5.6)$$

Таким образом, если в остатках существует полная положительная автокорреляция и $r_1 = 1$, то $d = 0$. Если в остатках полная отрицательная автокорреляция, то $r_1 = -1$ и, следовательно, $d = 4$. Если автокорреляция остатков отсутствует, то $r_1 = 0$ и $d = 2$. Т.е. $0 \le d \le 4$.

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина-Уотсона следующий. Выдвигается гипотеза H_0 об отсутствии автокорреляции остатков. Альтернативные гипотезы H_1 и H_1^* состоят, соответственно, в наличии положительной или отрицательной автокорреляции в остатках. Далее по специальным таблицам (см. приложение 4) определяются критические значения критерия Дарбина-Уотсона d_L и d_U для заданного числа наблюдений n, числа независимых переменных модели m и уровня значимости α. По этим значениям числовой промежуток $[0;\ 4]$ разбивают на пять отрезков. Принятие или отклонение каждой из гипотез с вероятностью $1-\alpha$ осуществляется следующим образом:

$0<d<d_L$ – есть положительная автокорреляция остатков, H_0 отклоняется, с вероятностью $P=1-\alpha$ принимается H_1;

$d_L<d<d_U$ – зона неопределенности;

$d_U<d<4-d_U$ – нет оснований отклонять H_0, т.е. автокорреляция остатков отсутствует;

$4-d_U<d<4-d_L$ – зона неопределенности;

$4-d_L<d<4$ – есть отрицательная автокорреляция остатков, H_0 отклоняется, с вероятностью $P=1-\alpha$ принимается H_1^*.

Если фактическое значение критерия Дарбина-Уотсона попадает в зону неопределенности, то на практике предполагают существование автокорреляции остатков и отклоняют гипотезу H_0.

ПРАКТИЧЕСКОЕ ЗАДАНИЕ. Проверим гипотезу о наличии автокорреляции в остатках для аддитивной модели нашего временного ряда. Исходные данные и промежуточные расчеты заносим в таблицу:

Таблица 5.11

t	y_t	$\varepsilon_t = E$	ε_{t-1}	$\left(\varepsilon_t - \varepsilon_{t-1}\right)^2$	ε_t^2
1	375	-5,252	–	–	27,584
2	371	-35,843	-5,252	935,8093	1284,7
3	869	-74,183	-35,843	1469,956	5503,1
4	1015	48,937	-74,183	15158,53	2394,8
5	357	-26,946	48,937	5758,23	726,09
6	471	60,464	-26,946	7640,508	3655,9
7	992	45,124	60,464	235,3156	2036,2
8	1020	50,244	45,124	26,2144	2524,5
9	390	2,361	50,244	2292,782	5,574
10	355	-59,229	2,361	3793,328	3508,1
11	992	41,431	-59,229	10132,44	1716,5
12	905	-68,450	41,431	12073,83	4685,4
13	461	69,668	-68,45	19076,58	4853,6
14	454	36,078	69,668	1128,288	1301,6
15	920	-34,263	36,078	4947,856	1174
16	927	-50,143	-34,263	252,1744	2514,3
Сумма		-0,002	50,141	84921,85	37911,97

Фактическое значение критерия Дарбина-Уотсона для данной модели составляет:

$$d = \frac{84921{,}85}{37911{,}97} = 2{,}24.$$

Сформулируем гипотезы: H_0 – в остатках нет автокорреляции; H_1 – в остатках есть положительная автокорреляция; H_1^* – в остатках есть отрицательная автокорреляция. Зададим уровень значимости $\alpha = 0{,}05$. По таблице значений критерия Дарбина-Уотсона определим для числа наблюдений $n = 16$ и числа независимых параметров модели $k = 1$ (мы рассматриваем только зависимость от времени t) критические значения $d_L = 1{,}10$ и $d_U = 1{,}37$. Фактическое значение d-критерия Дарбина-Уотсона попадает в интервал $d_U < d < 4 - d_U$ (1,37<2,24<2,63). Следовательно, нет основания отклонять гипотезу

H_0 об отсутствии автокорреляции в остатках.

Существует несколько ограничений на применение критерия Дарбина-Уотсона.

1. Он неприменим к моделям, включающим в качестве независимых переменных лаговые значения результативного признака.
2. Методика расчета и использования критерия Дарбина-Уотсона направлена только на выявление автокорреляции остатков первого порядка.
3. Критерий Дарбина-Уотсона дает достоверные результаты только для больших выборок.

Вопросы для самоконтроля

1. Что называют временным рядом?
2. Какие компоненты выделяют в составе экономического временного ряда?
3. В чем заключается основная задача эконометрического исследования временного ряда?
4. Охарактеризуйте понятие автокорреляции уровней временного ряда.
5. Какие методы применяются для проверки наличия тенденции временного ряда?
6. Как осуществляется сглаживание временного ряда по методу скользящей средней?
7. Что понимается под аналитическим выравниванием временного ряда?
8. Какие методы применяются для определения вида тенденции временного ряда?
9. Для чего применяется критерий Дарбина-Уотсона?
10. Как осуществляется моделирование сезонных колебаний?
11. В чем заключаются особенности адаптивных методов прогнозирования?

Тестовые задания

1. Временным рядом является совокупность значений ...

а) экономического показателя за несколько последовательных моментов (периодов) времени

б) экономических однотипных объектов по состоянию на определенный момент времени

в) последовательных моментов (периодов) времени и соответствующих им значений экономического показателя

г) экономического показателя для однотипных объектов на определенный момент времени

2. Проверка является ли временной ряд «белым шумом» осуществляется с помощью…
а) Q-статистики Бокса-Пирса
б) Критерия Дарбина-Уотсона
в) Коэффициента детерминации
г) F-критерия Фишера
3. Выберите основные модели временных рядов …
а) полиномиальная
б) мультипликативная
в) обратная
г) аддитивная
4. Автокорреляционная функция и коррелограмма используются для выявления во временном ряде наличия или отсутствия ...
а) только случайной компоненты
б) тренда, циклической или сезонной компонент
в) только тренда
г) только циклической компоненты
д) только сезонной компоненты
5. Какой из перечисленных методов не может быть применен для обнаружения автокорреляции?
а) метод рядов
б) критерий Дарбина-Уотсона
в) тест ранговой корреляции Спирмена
г) тест Уайта
6. Временной ряд – это …
а) последовательность упорядоченных во времени числовых показателей, характеризующих уровень состояния и изменения изучаемого явления;
б) последовательность числовых показателей, характеризующих уровень состояния и изменения изучаемого явления;
в) последовательность упорядоченных временных интервалов, или моментов времени.
7. Для выявления основной тенденции развития явления используются
а) метод укрупнения интервалов
б) метод скользящей средней
в) расчет средней хронологической
г) расчет средней гармонической
д) аналитическое выравнивание
8. Периодические колебания, возникающие под влиянием смены времени года, называются …
а) хронологическими

б) сезонными
в) тенденцией
г) случайными

9. Автокорреляцией называется ...

а) зависимость вариации значений одного показателя от вариации значений другого
б) зависимость между цепными уровнями
в) отклонения от тенденции
г) зависимость последующего уровня временного ряда от предыдущего

10. Составляющие временного ряда ...

а) тренд
б) периодическая компонента
в) сезонная компонента
г) случайная составляющая
д) систематическая составляющая

11. Автокорреляцией уровней временного ряда называют ...

а) корреляционную зависимость между трендовой и сезонной компонентами ряда
б) корреляционную зависимость между наблюдаемыми и расчетными значениями исследуемого временного показателя
в) корреляционную зависимость между последовательными уровнями ряда
г) автокорреляцию остатков временного ряда

12. Под экстраполяцией понимают нахождение неизвестных уровней

а) за пределами временного ряда
б) внутри временного ряда
в) в середине временного ряда

13. Аддитивная модель ...

а) представляет собой сумму компонент
б) представляет собой произведение компонент
в) представляет собой сумму и произведение компонент

14. Мультипликативная модель ...

а) представляет собой сумму компонент
б) представляет собой произведение компонент
в) представляет собой сумму и произведение компонент

Приложение 1

ЗАДАНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ (ИНДИВИДУАЛЬНОЙ) РАБОТЫ

Задание 1. На основании имеющихся данных:

1. Постройте поле корреляции.
2. Составьте уравнение линейной парной регрессии. С помощью $t-$критерия Стьюдента оцените статистическую значимость параметров регрессии.
3. Оцените тесноту связи с помощью показателей корреляции и детерминации.
4. С помощью $F-$критерия Фишера оцените статистическую надежность результатов регрессионного моделирования.
5. С помощью средней ошибки аппроксимации оцените качество полученного уравнения.
6. Рассчитайте средний коэффициент эластичности.
7. Рассчитайте прогнозное значение результативного показателя, если прогнозное значение фактора увеличится на 10% от его среднего уровня.
8. Подберите наилучшую нелинейную функцию.

Все расчеты провести при уровне значимости $\alpha = 0,05$ (с вероятностью 95%).

Сделайте выводы по полученным результатам.

Для проведения расчетов используйте ППП Excel.

№ п/п	Продуктивность коров, ц	Расход кормов на 1 корову, ц к.ед.
1	43,3	54,2
2	26,6	46,4
3	34,8	53,9
4	26,1	40,2
5	32,8	48,2
6	32,3	53,4
7	29,9	45,5
8	51,4	60,7
9	31,0	47,8
10	46,3	60,1
11	29,6	47,1
12	29,6	41,4
13	29,1	43,1
14	23,7	39,4
15	27,2	46,3
16	49,5	59,4
17	41,6	53,2
18	34,8	49,4
19	50,0	52,5

№ п/п	Трудообеспеченность (приходится работников на 100 га пашни), чел.	Трудоемкость производства зерна, чел.-час
1	3,6	1,2
2	17,8	0,7
3	3,6	2,6
4	5	3,7
5	4,2	2,2
6	10,5	2,1
7	2,1	1,9
8	5,5	1,8
9	2,9	4,3
10	5,6	1,9
11	2,9	5,5
12	6,1	1,5
13	7,8	1,6
14	5,4	0,7
15	5,3	3,6
16	5,5	2,4

№ п/п	Производительность труда 1 работника, тыс. руб.	Энерговооруженность, л.с.
1	147	291
2	474	621
3	157	366
4	134	261
5	126	381
6	191	408
7	82	181
8	210	414
9	73	206
10	141	266
11	97	193
12	91	387
13	105	259
14	202	470
15	98	235
16	85	219
17	145	355
18	186	403

Задание 2. Постройте модель множественной регрессии, используя метод включения и исключения факторов. Сделайте выводы по полученным результатам.

№ п/п	Наименование района	Производительность труда, тыс. руб.	Фондообеспеченность, тыс. руб.	Энерговооруженность, л.с.	Нагрузка коров на 1 доярку, гол.	Урожайность зерна, ц/га	Нагрузка пашни на 1 трактор, га
1	Брасовский	2,13	11,6	54,5	16	8,5	158,3
2	Брянский	29,6	39,5	55,4	22,9	12,5	58
3	Выгоничский	9,64	12,4	61,5	17,3	6,6	165,2
4	Гордеевский	3,27	14,5	39,4	17,6	5,8	139,3
5	Дубровский	3,66	5,1	56,2	21,2	2,9	180,1
6	Дятьковский	11,27	21,2	53	24,6	5	137,2
7	Жирятинский	57,84	6,4	40,1	18,4	4,1	242,4
8	Жуковский	5,4	19,8	55,8	19	8,5	129,3
9	Злынковский	9,16	21,4	53,2	24	3,7	207,7
10	Карачевский	8,98	9,1	43,8	23,9	8,6	134
11	Клетнянский	10,87	8,8	37,9	15,3	3,7	159,9
12	Климовский	1,06	15,8	62,7	21,7	5,8	118,8
13	Клинцовский	2,67	23,1	44,7	19	8,7	85,5
14	Комаричский	3,13	8,9	5	19,7	8,7	144,9
15	Красногорский	4,27	19,8	56,9	24,8	4,8	122,5
16	Мглинский	4,62	5	39,2	21,7	5,7	141,2
17	Навлинский	10,64	9,6	58,2	15,9	5,4	215
18	Новозыбковский	3,91	19,9	48,2	25,8	8,7	85,3
19	Погарский	2,32	13,3	49,1	18,6	8,9	108,6
20	Почепский	3	14,1	53,7	22,2	6,5	155,2
21	Рогнединский	3,72	7,9	52,5	17,6	2,8	123,8
22	Севский	3,24	11,7	43,6	19,7	8,4	163
23	Стародубский	1,32	18,3	46,1	18,8	8,1	82,2

№ п/п	Наименование района	Себестоимость 1 ц зерна, руб.	Урожайность зерна, ц/га	Затраты труда на 1 га посева зерновых, чел.-час	Трудообеспеченность, чел.	Энергообеспеченность, л.с.	Нагрузка пашни на 1 трактор, га
1	Брасовский	252	8,5	14,2	3,9	2,1	158,3
2	Брянский	212	12,5	8,1	15,5	8,6	58
3	Выгоничский	332,3	6,6	20,1	3,5	2,2	165,2
4	Гордеевский	389	5,8	32,2	5,7	2,3	139,3
5	Дубровский	579,2	2,9	9,8	3,6	2	180,1
6	Дятьковский	407,1	5	10,3	8	4,2	137,2
7	Жирятинский	715,9	4,1	15,1	2,8	1,1	242,4
8	Жуковский	332,6	8,5	12	5,9	3,3	129,3
9	Злынковский	463,2	3,7	10	4,4	2,4	207,7
10	Карачевский	226,2	8,6	16,3	6,6	2,9	134
11	Клетнянский	962	3,7	27,5	4	1,5	159,9
12	Климовский	246	5,8	12,1	5,7	3,6	118,8
13	Клинцовский	221,6	8,7	22,7	9	4	85,5
14	Комаричский	228,5	8,7	8,6	4,6	0,2	144,9
15	Красногорский	381,1	4,8	16,3	5,9	3,3	122,5
16	Мглинский	258,6	5,7	23,7	6,5	2,6	141,2
17	Навлинский	408,6	5,4	19,8	2,4	1,4	215
18	Новозыбковский	241,2	8,7	10,6	8	3,8	85,3
19	Погарский	214,2	8,9	1,3	7,6	3,8	108,6
20	Почепский	311	6,5	11,1	4,1	2,2	155,2
21	Рогнединский	753,5	2,8	9	3,7	2	123,8
22	Севский	230,1	8,4	11,9	3,8	1,7	163
23	Стародубский	225,4	8,1	14,3	6,5	3	82,2
24	Суземский	419,6	5,5	18,6	4,7	2,7	125,8
25	Суражский	368,2	5	20,9	7,1	2,7	122,7

Задание 3.

ВАРИАНТ 1

Задача 1.

Дана система одновременных уравнений:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2 \\ y_2 = b_{23}y_3 + a_{21}x_1 + a_{23}x_3 + a_{24}x_4 \\ y_3 = b_{31}y_1 + a_{33}x_3 + a_{34}x_4 \end{cases}$$

Задание. Проверить каждое уравнение системы на необходимое и достаточное условия идентификации.

Задача 2. Дана упрощенная модель Менгеса:

$$\begin{cases} Y_t = a_1 + b_{11}Y_{t-1} + b_{12}I_t + \varepsilon_1 \\ I_t = a_2 + b_{21}Y_t + b_{22}Q_t + \varepsilon_2 \\ C_t = a_3 + b_{31}Y_t + b_{32}C_{t-1} + b_{33}P_t + \varepsilon_3 \end{cases},$$

где Y – национальный доход; C – расходы на личное потребление; I – инвестиции; Q – валовая прибыль экономики; P – индекс стоимости жизни; t – текущий период; $t-1$ – предыдущий период.

Задание. 1. Применив необходимое и достаточное условия идентификации, определите, идентифицировано ли каждое из уравнений модели; 2. Определите метод оценки параметров модели.

ВАРИАНТ 2

Задача 1.

Дана система одновременных уравнений:

$$\begin{cases} y_1 = b_{12}y_2 + b_{14}y_4 + a_{11}x_1 + a_{13}x_3 \\ y_2 = b_{21}y_1 + b_{23}y_3 + a_{22}x_2 \\ y_4 = b_{32}y_2 + a_{32}x_2 + a_{33}x_3 \end{cases}$$

Задание. Проверить каждое уравнение системы на необходимое и достаточное условия идентификации.

Задача 2. Дана упрощенная конъюнктурная модель:

$$\begin{cases} C_t = a_1 + b_{11}Y_t + b_{12}C_{t-1} + \varepsilon_1 \\ I_t = a_2 + b_{21}R_t + b_{22}I_{t-1} + \varepsilon_2 \\ Y_t = C_t + I_t + G_t \end{cases},$$

где C – расходы на потребление; Y – совокупный доход; I – инвестиции; R – процентная ставка банка; G – государственные расходы; t – текущий период; $t-1$ – предыдущий период.

Задание. 1. Применив необходимое и достаточное условия идентификации, определите, идентифицировано ли каждое из уравнений модели; 2. Определите метод оценки параметров модели.

ВАРИАНТ 3

Задача 1.

Дана система одновременных уравнений:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2 \\ y_2 = b_{23}y_3 + a_{22}x_2 + a_{23}x_3 \\ y_3 = b_{31}y_1 + a_{31}x_1 + a_{33}x_3 \end{cases}$$

Задание. Проверить каждое уравнение системы на необходимое и достаточное условия идентификации.

Задача 2. Дана упрощенная модель Клейна:

$$\begin{cases} C_t = a_1 + b_{12}Y_t + b_{13}T_t + \varepsilon_1 \\ I_t = a_2 + b_{21}Y_t + b_{22}K_{t-1} + \varepsilon_2 \\ Y_t = C_t + I_t \end{cases},$$

где C – расходы на потребление; Y – совокупный доход; I – инвестиции; T – налоги; K – запас капитала; t – текущий период; $t-1$ – предыдущий период.

Задание. 1. Применив необходимое и достаточное условия идентификации, определите, идентифицировано ли каждое из уравнений модели; 2. Определите метод оценки параметров модели.

Задание 4. Постройте мультипликативную модели временного ряда … Сделайте выводы по полученным результатам.

Вариант 1. Удельного веса простоев оборудования, %

Базисный год												Отчетный год			
январь	февраль	март	апрель	май	июнь	июль	август	сентябрь	октябрь	ноябрь	декабрь	январь	февраль	март	апрель
18,1	7,8	17,4	6,4	7,8	17,1	10,2	14,1	20,0	16,7	16,0	20,4	16,2	14,8	20,1	16,0

Вариант 2. Цены реализации свеклы, руб.

Базисный год												Отчетный год			
январь	февраль	март	апрель	май	июнь	июль	август	сентябрь	октябрь	ноябрь	декабрь	январь	февраль	март	апрель
41,6	34,6	50,1	51,6	55,2	53,5	28,6	44,8	30,4	48,8	56,4	40,5	30,4	85,3	91,3	104,5

Вариант 6. Удельного веса в товарообороте предприятия продукции собственного производства, %

январь	февраль	март	апрель	май	июнь	июль	август	сентябрь	октябрь	ноябрь	декабрь	январь	февраль	март	апрель
18,6	19,2	23,4	20,1	19,8	20,3	29,3	19,7	20,1	18,9	18,0	18,6	19,2	21,3	19,6	18,5

Приложение 2

Таблица значений F-критерия Фишера при уровне значимости
$\alpha = 0,05$

k_2 \ k_1	1	2	3	4	5	6	8
1	161,5	199,5	215,7	224,6	230,2	233,9	238,9
2	18,51	19,00	19,16	19,25	19,30	19,33	19,37
3	10,13	9,55	9,28	9,12	9,01	8,94	8,84
4	7,71	6,94	6,59	6,39	6,26	6,16	6,04
5	6,61	5,79	5,41	5,19	5,05	4,95	4,82
6	5,99	5,14	4,76	4,53	4,39	4,28	4,15
7	5,59	4,74	4,35	4,12	3,97	3,87	3,73
8	5,32	4,46	4,07	3,84	3,69	3,58	3,44
9	5,12	4,26	3,86	3,63	3,48	3,37	3,23
10	4,96	4,10	3,71	3,48	3,33	3,22	3,07
11	4,84	3,98	3,59	3,36	3,20	3,09	2,95
12	4,75	3,88	3,49	3,26	3,11	3,00	2,85
13	4,67	3,80	3,41	3,18	3,02	2,92	2,77
14	4,60	3,74	3,34	3,11	2,96	2,85	2,70
15	4,54	3,68	3,29	3,06	2,90	2,79	2,64
16	4,49	3,63	3,24	3,01	2,85	2,74	2,59
17	4,45	3,59	3,20	2,96	2,81	2,70	2,55
18	4,41	3,55	3,16	2,93	2,77	2,66	2,51
19	4,38	3,52	3,13	2,90	2,74	2,63	2,48
20	4,35	3,49	3,10	2,87	2,71	2,60	2,45
21	4,32	3,47	3,07	2,84	2,68	2,57	2,42
22	4,30	3,44	3,05	2,82	2,66	2,55	2,40
23	4,28	3,42	3,03	2,80	2,64	2,53	2,38
24	4,26	3,40	3,01	2,78	2,62	2,51	2,36
25	4,24	3,38	2,99	2,76	2,60	2,49	2,34
26	4,22	3,37	2,98	2,74	2,59	2,47	2,32
27	4,21	3,35	2,96	2,73	2,57	2,46	2,30
28	4,20	3,34	2,95	2,71	2,56	2,44	2,29
29	4,18	3,33	2,93	2,70	2,54	2,43	2,28

Приложение 3

Критические точки распределения t – критерия Стьюдента

Число степеней свободы k	Уровень значимости α			
	0,1	**0,05**	**0,02**	**0,01**
1	6,3138	12,7062	31,8205	63,6567
2	2,9200	4,3027	6,9646	9,9248
3	2,3534	3,1824	4,5407	5,8409
4	2,1318	2,7764	3,7469	4,6041
5	2,0150	2,5706	3,3649	4,0321
6	1,9432	2,4469	3,1427	3,7074
7	1,8946	2,3646	2,9980	3,4995
8	1,8595	2,3060	2,8965	3,3554
9	1,8331	2,2622	2,8214	3,2498
10	1,8125	2,2281	2,7638	3,1693
11	1,7959	2,2010	2,7181	3,1058
12	1,7823	2,1788	2,6810	3,0545
13	1,7709	2,1604	2,6503	3,0123
14	1,7613	2,1448	2,6245	2,9768
15	1,7531	2,1314	2,6025	2,9467
16	1,7459	2,1199	2,5835	2,9208
17	1,7396	2,1098	2,5669	2,8982
18	1,7341	2,1009	2,5524	2,8784
19	1,7291	2,0930	2,5395	2,8609
20	1,7247	2,0860	2,5280	2,8453
21	1,7207	2,0796	2,5176	2,8314
22	1,7171	2,0739	2,5083	2,8188
23	1,7139	2,0687	2,4999	2,8073
24	1,7109	2,0639	2,4922	2,7969
25	1,7081	2,0595	2,4851	2,7874
26	1,7056	2,0555	2,4786	2,7787
27	1,7033	2,0518	2,4727	2,7707
28	1,7011	2,0484	2,4671	2,7633
29	1,6991	2,0452	2,4620	2,7564
30	1,6973	2,0423	2,4573	2,7500
35	1,6896	2,0301	2,4377	2,7238
40	1,6839	2,0211	2,4233	2,7045
45	1,6794	2,0141	2,4121	2,6896
50	1,6759	2,0086	2,4033	2,6778

Приложение 4

Значения статистик Дарбина-Уотсона при 5%-ном уровне значимости

n	$k=1$		$k=2$		$k=3$		$k=4$		$k=5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0,61	1,40								
7	0,70	1,36	0,47	1,90						
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1,05	1,35	0,91	1,55	0,77	1,78				
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,85	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,99
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83

Библиографический список

Основная литература

1. Новиков А.И. Эконометрика: учебное пособие для бакалавров[Электронный ресурс]. Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=5670. М.: Дашков и К^{o}, 2013.
2. Уткин В.Б. Эконометрика: учебник [Электронный ресурс]. Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=56347. Загл. с экрана. М.: Дашков и К^{o}, 2013.

Дополнительная литература

1. Валентинов В.А. Эконометрика: практикум. М.: Дашков и К, 2008.
2. Валентинов В.А. Эконометрика: учеб. для вузов. М.: Дашков и К, 2007.
3. Елисеева И.И. Эконометрика. М.: Финансы и статистика, 2003.
4. Елисеева И.И. Практикум по эконометрике. М.: Финансы и статистика, 2003.
5. Яновский Л.П. Введение в эконометрику: электронный учеб. для вузов. М.: КноРус, 2009.

Учебное издание

Анна Васильевна Раевская

ЭКОНОМЕТРИКА:

Учебно-методическое пособие

Редактор Лебедева Е.М.

Подписано к печати 07.07.2016 г. Формат 60х84 Бумага печатная
Усл. п.л. 4,53. Тираж 50 экз. Издат. № 5175.

Издательство Брянского государственного аграрного университета
243365 Брянская обл., Выгоничский район, с. Кокино, Брянский ГАУ