САРАТОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМ. Н.Г. ЧЕРНЫШЕВСКОГО

Экономический факультет

Кафедра менеджмента и маркетинга

Леванова Л.Н.

Учебно-методическое пособие по курсу

ЭКОНОМЕТРИКА

Саратов 2015

УДК 330.43(072.8)
ББК 65в6я73

Л-34

Л-34 Леванова Л.Н. Эконометрика. Учебно-методическое пособие. Саратов: Амирит, 2015. С.144.

ISBN 978-5-9907420-3-1

Учебно-методическое пособие подготовлено в соответствии с положениями и требованиями Государственного образовательного стандарта высшего профессионального образования, включает основные вопросы лекций, терминологический словарь и основные теоретические положения эконометрики, планы семинарских занятий, методические указания по выполнению всех практических и самостоятельных работ, текущие семестровые задания, тестовые задания для оценки остаточных знаний.

Для студентов заочной формы обучения.

Составитель

Кандидат экономических наук, доцент Л.Н. Леванова

Рекомендуют к печати:

Кафедра менеджмента и маркетинга Саратовского государственного университета им. Н.Г. Чернышевского

Кандидат экономических наук, доцент Л.И. Дорофеева

ISBN 978-5-9907420-3-1

УДК 330.43(072.8)
ББК 65в6я73

СОДЕРЖАНИЕ

Введение

В современных условиях развития экономики принятие управленческих решений невозможно без количественного анализа, формирующего основу эконометрического моделирования. «Эконометрика» как дисциплина дает студентам навыки в моделировании, прогнозировании и выборе оптимальных путей развития экономических процессов и явлений на различных уровнях, в том числе при разработке стратегии управления предприятием.

Дисциплина «Эконометрика» изучает методы построения тех количественных взаимосвязей экономических процессов и явлений, которые студенты изучают в таких дисциплинах как «Микроэкономика», «Макроэкономика», «Стратегический менеджмент», «Теория отраслевых рынков», «Менеджмент», «Маркетинг», «Инвестиционная деятельность», «Финансовый менеджмент», «Управление ценовой политикой предприятия». В силу того, что «Эконометрика» является сплавом четырех компонент: экономической теории, статистических и математических методов, компьютерных вычислений, данная дисциплина является логическим продолжением таких модулей дисциплины «Математика» как «Математический анализ», «Линейная алгебра», «Теория вероятностей и математическая статистика», а также дисциплин «Информатика», «Статистика» математического цикла.

Знания, умения и навыки в решении задач в области высшей математики: дифференциального и интегрального исчисления, в решении систем линейных уравнений, в области теории вероятностей являются необходимыми для освоения эконометрических методов.

Целью освоения дисциплины «Эконометрика» являются:

- научить студентов строить количественные взаимосвязи в менеджменте и экономике, определять характер зависимости экономических параметров, а именно находить причинно-следственную связь явлений и процессов, рассматриваемых в экономике и управлении;
- научить студентов строить стандартные эконометрические модели исследуемых процессов, явлений и объектов, относящихся к области профессиональной деятельности, используя регрессионный анализ: модели парной и множественной регрессии; системы эконометрических уравнений; временные ряды;
- дать студентам знания математического аппарата, позволяющие анализировать и интерпретировать полученные модели, строить сценарии развития исследуемых процессов и выбирать оптимальный.

В результате освоения дисциплины обучающийся должен:

Знать:

- основные типы эконометрических моделей и методы их построения;

- основные методы и критерии оценки статистической значимости построенных эконометрических моделей;
- области применения различных типов эконометрических моделей в управленческой науке и экономике в целом.

Уметь:

- анализировать количественную и качественную информацию, выявлять причинно-следственные связи показателей, отражающих поведение и результативность организаций;
- строить, используя специальные математические методы, по собранным статистическим выборкам, описывающим различные аспекты деятельности компаний, стандартные эконометрические модели;
- давать экономическую интерпретацию полученных эконометрических моделей;
- доказывать статистическую значимость построенных эконометрических моделей и адекватность их рассматриваемым объектам – оригиналам;
- прогнозировать с помощью построенных эконометрических моделей значения показателей и поведение экономических субъектов, строить сценарии развития в результате принятия управленческих решений;
- использовать возможности Excel для построения эконометрических моделей.

Владеть:

- современной методикой построения эконометрических моделей и эконометрического прогнозирования;
- инструментарием оценки статистической значимости построенных эконометрических моделей и адекватности их рассматриваемым объектам – оригиналам;
- навыками факторного анализа экономических явлений и процессов, количественного и качественного анализа информации при принятии управленческих решений и построения причинно-следственных интересующих экономических связей;
- навыками применения компьютерной техники (приложением Excel) для решения эконометрических задач.

В пособии представлены основные теоретические аспекты эконометрического моделирования, позволяющие студенту изучить основы эконометрики, рассмотрены примеры построения эконометрических моделей с помощью средств Excel, что позволяет студентам заочной формы обучения выполнять задания в домашних условиях.

Также пособие содержит фонды оценочных средств в форме тестов и вопросов для обсуждения, а также задания для самостоятельной работы студентов.

ТЕМА 1. ВВЕДЕНИЕ В ЭКОНОМЕТРИКУ

Ключевые слова:

Эконометрика. Модель. Моделирование. Адекватность модели. Логические, геометрические и математические модели. Экономические и эконометрические модели. Модели микроэкономики, мезоэкономики и макроэкономики. Статические и динамические модели. Этапы эконометрического моделирования. Экзогенные переменные. Эндогенные переменные. Предопределенные переменные. Лаговые эндогенные переменные. Динамические и пространственные модели. Закрытые и открытее модели. Случайная величина. Испытание. Событие. Пространство элементарных событий. Генеральная совокупность. Выборка. Вероятность случайной величины. Дискретная и непрерывная случайные величины. Математическое ожидание. Дисперсия. Теоретическое стандартное отклонение. Функция плотности вероятности. Функция распределения случайной величины. Равномерное распределение. Теорема Ляпунова. Нормальное распределение. Стандартное нормальное распределение. Степень свободы. Распределение Стьюдента. Несмещенность. Эффективность. Состоятельность.

Основные теоретические аспекты темы:

1.1. Предмет и задача эконометрики

Немецкий философ Э.Кант говорил «Любая наука лишь постольку наука, поскольку она математика».

Впервые термин «эконометрика» был предложен в 1910 году бухгалтером П.Цьемпа. Предложенный термин звучал как «эконометрия». Эконометрика как наука стала формироваться в 30 –х годах XX века.

В развитии любой науки прослеживаются следующие основные этапы:

- наблюдение за поведением исследуемого объекта или явления;
- выявление основных качественных и количественных факторов, влияющих на его поведение;
- выявление и описание закономерностей влияния факторов на поведение объекта;
- формализованное представление накопленных знаний с помощью того или иного, как правило, математического аппарата с целью прогнозирования или управления поведением объекта.

Вступление экономики в стадию формализованного представления имеющихся результатов и широкого применения математических моделей для управления экономическими объектами поставило задачу разработки методических подходов для построения необходимых моделей. Таким образом, эконометрика как научное направление призвана обеспечить экономистов инструментарием для решения конкретных экономических задач.

В настоящее время нет единого определения эконометрики как науки. Термин «эконометрика» был введен Рагнером Фришем в 1926 г. и в дословном переводе означает «экономические измерения» или «измерения в экономике». Эконометрика – «экономика» + «метрика».

Рассмотрим ряд определений «эконометрики»:

Эконометрика – это совокупность методов анализа связей между различными экономическими показателями (факторами) на основании реальных статистических данных с использованием аппарата теории вероятностей и математической статистики.

Эконометрика - это раздел экономики, изучающий конкретные количественные закономерности и взаимосвязи между переменными экономических объектов с помощью математических методов и моделей.

В вышеприведенном определении обозначается место эконометрики как науки — это раздел экономики, а также определен методический подход *к* получению необходимых количественных закономерностей.

Эконометрика - это наука об измерении и анализе экономических явлений, о количественных выражениях тех связей и соотношений, которые раскрыты и обоснованы экономической теорией. Это сплав четырех компонент: экономической теории, статистических и математических методов, компьютерных вычислений.

Таким образом, *задача эконометрики* состоит в выявлении связей между количественными характеристиками экономических объектов или построения эконометрических моделей в целях осуществления прогнозов (вычисления приближенных значений). Практическим материалом для построения эконометрических моделей являются результаты наблюдений за изучаемыми экономическими объектами и значения количественных характеристик этих объектов.

Как отмечает американский экономист, лауреат Нобелевской премии 1985 г. Л. Клейн, «основная задача эконометрики — наполнить эмпирическим содержанием априорные экономические рассуждения». То есть, задачей эконометрики является дать количественные оценки выводам и закономерностям, сформулированным в общей экономической теории.

1.2. Эконометрическая модель и этапы построения модели

В основе эконометрики лежит эконометрическое моделирование, а результатом эконометрического моделирования является математическая модель экономического объекта, которая адекватно описывает его реальное поведение и в дальнейшем может использоваться с целью прогнозирования поведения объекта в интересующих исследователя условиях или с целью решения задачи оптимального управления объектом.

Экономический объект - это любой хозяйствующий субъект. Это может быть домашнее хозяйство, производство, отдельный регион, национальная или глобальная экономическая система, т.е. любой объект, который участвует в процессах производства, потребления или обмена.

Модель – объект любой природы, который создается исследователем с целью получения новых знаний об объекте-оригинале и отражает только существенные (с точки зрения разработчика) свойства оригинала.

Моделирование - процесс построения, изучения и применения моделей.

Модель адекватна объекту-оригиналу – если она с достаточной степенью точности приближения отражает закономерности процесса функционирования реального объекта.

Математическая модель – это абстракция реального мира, в которой интересующие исследователя отношения между реальными элементами заменены подходящими отношениями между математическими категориями. Эти отношения, как правило, представлены в форме уравнений или неравенств.

Вероятностная математическая модель – модель, имитирующая механизм функционирования гипотетического (не конкретного) явления или системы стохастической природы.

Вероятностно - статистическая математическая модель – модель, в которой значения отдельных характеристик (параметров) оцениваются по результатам наблюдений (исходным статистическим данным), то есть это модель, характеризующая механизм функционирования конкретного явления или системы стохастической природы.

Экономическая модель - вероятностная математическая модель, описывающая механизм функционирования экономического явления, экономической или социально – экономической системы.

Эконометрическая модель – вероятностно – статистическая модель, описывающая механизм функционирования экономического явления, экономической или социально – экономической системы.

Процесс эконометрического моделирования можно разбить на несколько этапов:

Этап 1. Постановка задачи, спецификация модели.

Необходимо четко сформулировать сущность экономической проблемы и принимаемые допущения. Этап включает выделение важнейших черт и свойств объекта и абстрагирование от второстепенных свойств, изучение основных закономерностей взаимодействия переменных, выявленных экономической теорией. На основе данной информации подбирается поведенческая функция, с помощью которой предполагается моделировать поведение изучаемого объекта.

Этап 2. Подготовка исходной информации.

Реальные возможности получения информации, а именно ограничения, связанные с возможностью получения данных по срокам и затраты на получение информации, ограничивают возможности выбора и построения модели. Но собранная информация должна обеспечить вычисление всех неизвестных параметров модели. Может оказаться, что невозможно собрать данные по всем переменным, включенным в спецификацию модели. Причины могут быть разными: длительность экономического процесса, неоправданно высокие затраты на сбор исходной информации или наличие в спецификации модели таких переменных, наблюдение которых невозможно. Выход из создавшейся ситуации один: вернуться к первому этапу и изменить спецификацию модели таким образом, чтобы сделать выполнимым этап сбора необходимой информации.

*Этап 3. Оценивание неизвестных параметров***.**

Этап включает выбор методов оценки неизвестных параметров модели.

Этап 4. Проверка гипотез статистической значимости параметров.

Проверяется ряд статистических гипотез с целью подтверждения выполнения условий или ограничений, наложенных на модель и ее основные элементы.

Этап 5. Проверка адекватности.

По результатам этапа делается заключение о практической пригодности модели, то есть для использования ее в целях прогнозирования.

1.3.Переменные в моделях и их типы

Каждому экономическому объекту присущ набор количественных показателей, которые характеризуют его состояние.

Переменными модели называются количественные показатели экономического объекта, которые могут принимать любое значение из их области определения, то есть количественные показатели, которые характеризуют состояние или поведение объекта.

Спецификация модели представляет собой подробную запись на математическом языке известных закономерностей поведения экономического объекта.

Под закономерностями поведения экономического объекта понимаются закономерности взаимосвязей между экономическими переменными. Источником закономерностей служат выводы, полученные в результате изучения особенностей поведения объекта. Они могут быть известны из общей экономической теории или в результате собственного анализа объекта. Очевидно, экономическая теория не может рассмотреть каждый конкретный экономический объект. Она рассматривает, как правило, общие закономерности поведения и развития отдельных экономических систем. Моделирование конкретного объекта по конкретным статистическим данным является задачей эконометрики.

Рассмотрим закономерности равновесного конкурентного рынка:

1. Спрос на товар падает с ростом цены на товар.
2. Предложение товара растет с ростом цены на него.
3. Равновесная цена на товар есть результат равновесия между уровнями спроса и предложения.

Исходя из приведенных закономерностей, спецификацию модели конкурентного рынка можно записать в виде системы линейных алгебраических уравнений:

$$\begin{cases} y^d = a_0 + a_1 p; \\ y^s = b_0 + b_1 p; \\ y^d = y^s. \end{cases}$$

Где: y^d, y^s, p – переменные модели (объекта);

$a_1<0$, a_0, b_0, $b_1>0$ – коэффициенты или параметры модели.

Из экономической теории известно, что уровень спроса не только падает с ростом цены, но возрастает с ростом располагаемого дохода потребителя. Если x- располагаемый доход потребителя, то спецификация модели конкурентного рынка получит вид:

$$\begin{cases} y^d = a_0 + a_1 p + a_2 x; \\ y^s = b_0 + b_1 p; \\ y^d = y^s \end{cases}$$

Значение переменной x (располагаемый доход потребителя) формируется вне конкурентного рынка, а переменные y^d, y^s, p наоборот, формируются внутри конкурентного рынка в результате их взаимодействия, как между собой, так и с переменной *x*. Эти различия находят отражение в следующих определениях.

Экзогенные переменные в модели (независимые, внешние) – переменные, значения которых формируются вне модели, задаются «извне», автономно от модели, управляемые и планируемые.

Эндогенные переменные в модели (зависимые, внутренние) – переменные, значения которых формируются в процессе и внутри функционирования анализируемой социально – экономической системы в существенной мере под воздействием экзогенных переменных и во взаимодействии друг с другом. В эконометрическом моделировании являются предметом объяснения.

И один из принципов спецификации модели заключается в том, что количество уравнений в модели равно количеству эндогенных переменных, участвующих в ней. Этот принцип спецификации модели служит одной из проверок правильности спецификации модели и устанавливает соотношение между количеством уравнений в модели и количеством текущих эндогенных переменных.

В модели конкурентного рынка три эндогенные переменные, соответственно, мдель содержит три уравнения.

В моделировании необходимо учитывать влияние времени на значения переменных, так как практически все переменные экономического объекта изменяются со временем. Отражение зависимости переменных от момента времени осуществляется в модели с помощью индекса *t* при переменных.

Следовательно, модель конкурентного рынка принимает вид:

$$\begin{cases} y_t^{\,d} = a_0 + a_1 p_t + a_2 x_t; \\ y_t^{\,s} = b_0 + b_1 p_t; \\ y_t^{\,d} = y_t^{\,s} \end{cases}$$

Индекс t может быть только при переменных, они изменяются со временем. Параметры – константы от времени не зависят.

Смысл индекса *t* — это интервал времени, внутри которого значение переменных модели можно считать практически неизменными. Длина этого интервала для каждой задачи своя. Это может быть день (курс валют), месяц, квартал, год. Но во всех случаях индекс *t* при переменной указывает на то, что рассматривается значение переменной в текущий момент времени.

Кроме того, еще одна особенность экономических объектов заключается в их инертности. С одной стороны, для того чтобы изменить объемы производства в ответ на изменение цен на продукцию на рынке, производителю нужно время. С другой стороны, планируя производство, производитель еще не знает, какая цена установится на рынке. Ему приходится ориентироваться на ту цену, которая была в предыдущий период времени. Поэтому в моделях могут присутствовать как переменные, отнесенные к текущему моменту времени, так и переменные, отнесенные к предыдущим моментам времени. Исходя из сказанного, спецификацию модели конкурентного рынка можно записать в виде:

$$\begin{cases} y_t^{\,d} = a_0 + a_1 p_t + a_2 x_t; \\ y_t^{\,s} = b_0 + b_1 p_{t-1}; \\ y_t^{\,d} = y_t^{\,s} \end{cases}$$

Лаговые переменные - переменные модели, отнесенные к предшествующим моментам времени.

Лаговыми могут быть как эндогенные, так и экзогенные переменные. Лаг (задержка) может быть как в один, так и в несколько периодов. Появление лаговых переменных в моделях делает необходимым внесение дополнений в классификацию переменных.

Датированные переменные (от слова «дата»). - переменные модели, отнесенные ко времени.

В последнем виде модели конкурентного рынка все переменные отнесены к определенному моменту времени: переменные y_t^d, y_t^s, p_t, x_t соответствуют текущему моменту (периоду) времени, а переменная p_{t-1} соответствует предыдущему моменту (периоду) времени

Предопределенные переменные модели – все экзогенные переменные модели и лаговые эндогенные переменные. Это переменные, значения которых предварительно (до текущего момента времени) определены.

Так, в модели переменные p_{t-1} и x_t составляют группу предопределенных переменных.

Введение в спецификацию датированных переменных позволяет выделить виды моделей.

Динамические модели - модели, содержащие датированные переменные.

Пространственные модели – модели, в которых все переменные не зависят от времени.

Замкнутые модели (закрытые) – модели, в состав которых входят только текущие эндогенные переменные.

Открытые модели – модели, в составе которых присутствует хотя бы одна предопределенная переменная.

1.4. Экономические показатели как случайные величины

Рыночная экономика в отличие от административно-хозяйственной системы обладает свойством стохастичности или случайности. То есть в условиях рыночных отношений на формирование значений экономических показателей влияет множество факторов, часть из которых известна исследователю, а часть – нет. В результате при моделировании наблюдаются случайные возмущения и все переменные, используемые в эконометрическом моделировании, являются случайными величинами.

Рассмотрим ряд понятий из теории вероятностей, необходимые в процессе построения эконометрических моделей, оценки значимости их коэффициентов и проверки адекватности моделей в целом.

Случайная величина – действительная переменная, которая в зависимости от исхода опыта, т.е. в зависимости от случая принимает различные значения, (значение не может быть точно предсказано).

$$X = x_1;\ X = x_2.$$

Испытание – реализация определенного комплекса условий, который может воспроизводиться неограниченное число раз.

Событие – результат испытания: достоверное (всегда происходит), невозможное (никогда не происходит), случайное (может произойти или не произойти).

Пространство элементарных событий – совокупность всех возможных, конкретных исходов.

Генеральная совокупность – совокупность всех возможных значений наблюдений интересующего показателя.

Выборка – множество наблюдений, составляющих часть генеральной совокупности.

Под вероятностью некоторого события (события, состоящего в том, что случайная переменная приняла определенное значение**)** понимается доля числа исходов, благоприятствующих данному событию, в общем числе возможных исходов.

$$P\{X = x_1\}.$$

Дискретная случайная величина – случайная величина, принимающая в результате испытания значение из конечного либо счетного множества возможных чисел.

Математическое ожидание дискретной случайной величины – это взвешенное среднее всех ее возможных значений, где в качестве весового коэффициента берется вероятность соответствующего исхода.

$$\mathrm{E(x)} = \mu = \sum_{i=1}^{n} x_i * P_i .$$

Теоретическая дисперсия дискретной случайной величины – мера разброса случайной величины вокруг среднего значения, определяется как математическое ожидание квадрата разности между величиной x и ее средним.

$$\sigma_x^2 = \sum_{i=1}^{n} (x-\mu)^2 * P_i = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} .$$

Теоретическое стандартное отклонение – мера разброса случайной величины вокруг среднего значения, имеющая размерность данной случайной величины. Это среднее квадратическое разброса случайной величины m.c.o σ_x

Коэффициент вариации случайной величины – мера относительного разброса случайной величины, показывает какую долю среднего значения случайной величины составляет ее средний разброс.

$$V_x = \frac{\sigma_x}{\mu} .$$

Непрерывная случайная величина – случайная величина, принимающая значение из непрерывного диапазона значений (их нельзя пересчитать, ставя в соответствие им натуральные числа).

Функция плотности вероятности h=f(x) - функция значений случайной переменной, показывает вероятность нахождения случайной переменной внутри единичного интервала вокруг данной точки (вероятность попадания случайной величины x в бесконечно малый интервал).

Функция распределения случайной величины $F_x(x)$- вероятность того, что случайная величина принадлежит интервалу [a,b].

$$F_x(x) = P\{X \in [a,b]\} .$$

Математическое ожидание непрерывной случайной величины -

$$\mu = \int_a^b x * f(x) dx .$$

Теоретическая дисперсия непрерывной случайной величины -

$$\sigma_x^2 = \int_a^b (x-\mu)^2 * f(x) dx.$$

Равномерное распределение – это такое распределение вероятности, плотность которого постоянна в заданном интервале изменения случайной величины X. $a \le X \le b$.

Плотность вероятности на [a,b]: $f(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0 & x < a, x > b \end{cases}$

*Функция распределения***:** $F(x) = \begin{cases} 0, & x \le a \\ \dfrac{x-a}{b-a}, & a < x \le b \\ 1. & x > b \end{cases}$

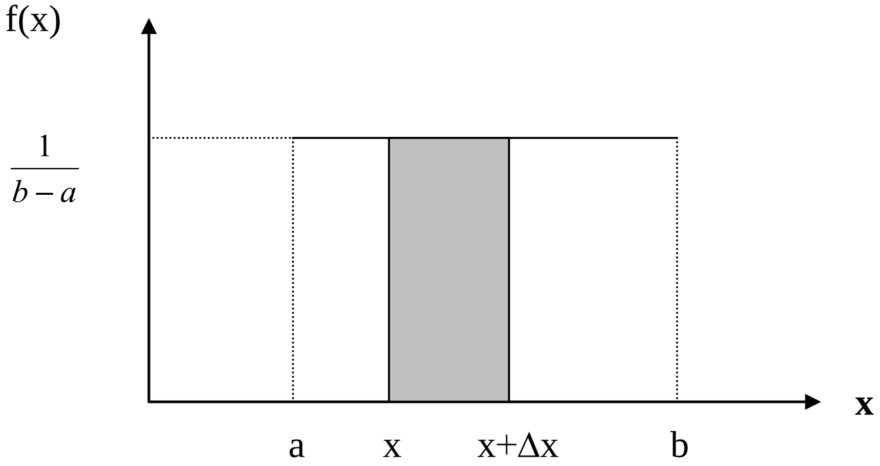


Рис. 1.4.1. Плотность вероятности равномерно распределенной случайной величины.

Вероятность попадания равномерно распределенной случайной величины в заданный интервал равна произведению плотности вероятности на длину интервала $\Delta P = \dfrac{\Delta x}{b-a}$.

Теорема Ляпунова (центральная предельная теорема).

Распределение суммы n произвольно распределенных и взаимно независимых случайных величин при n→∞ стремится к нормальному распределению, если вклад отдельных слагаемых в сумму равномерно мал.

То есть если случайная величина является общим результатом взаимодействия большого числа других случайных величин, ни одна из которых не является доминирующей, то она будет иметь приблизительно нормальное распределение, даже если отдельные составляющие не имеют нормального распределения.

Нормальное распределение случайной величины X характеризуется лишь двумя параметрами: средним значением μ и стандартным отклонением σ. X=N(μ, σ).

Плотность вероятности нормального распределения на [-∞,+∞]:

$$f_N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}.$$

Функция распределения:

$$F_N(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(t-\mu)^2}{2\sigma^2}} dt.$$

Стандартное нормальное распределение (z – распределение)– это нормальное распределение случайной величины $z = \frac{x-\mu}{\sigma}$ с параметрами μ =0 и σ=1. Z=N(0, 1).

Ее распределение может быть протабулировано. [Приложение А, Таблица А 1.].

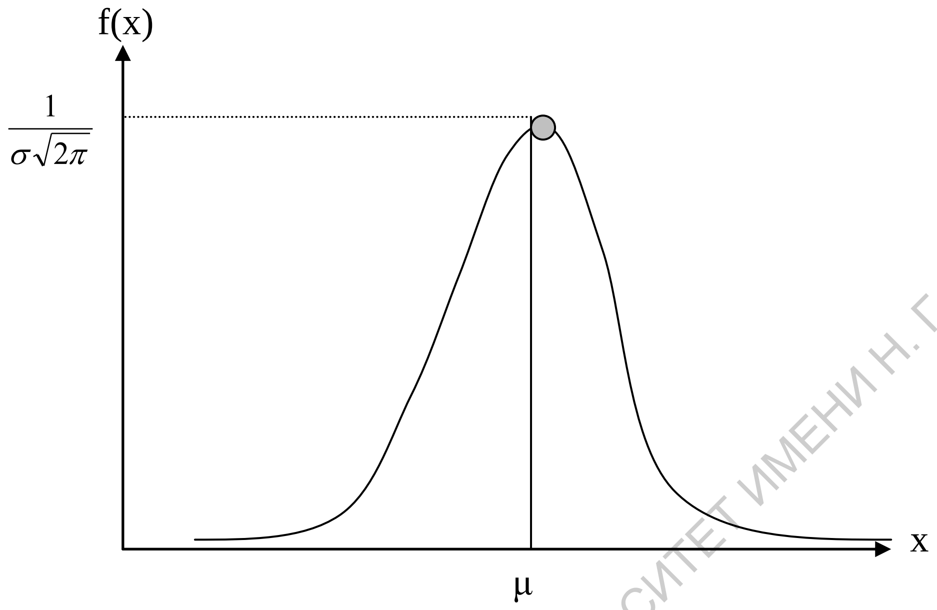


Рис.1.4.2. Плотность вероятности нормально распределенной случайной величины

Распределение Стьюдента (t -статистика) – выборочный аналог нормального распределения, переходящее в него при бесконечно большом числе наблюдений, это распределение случайной величины $t = \frac{x-\bar{x}}{s}$; где s – выборочное стандартное отклонение, определяемое по данным наблюдений и меняющееся от выборки к выборке.

На практике обычно используют не таблицы функции распределения Стьюдента, а таблицы критических точек, то есть точек с заданной вероятностью попадания в начинающиеся от них хвосты распределения. [5, Таблица А 2.].

1.5. Оценки и их свойства

Оценкой параметра называется приближенное значение этого параметра, вычисленное по результатам выборки.

В отличие от параметра его оценка является величиной случайной. Очевидно, можно предложить некоторое количество процедур, с помощью которых можно по результатам наблюдений вычислить значение оценки параметра.

Таблица 1.5.1.

Характеристики генеральной совокупности и оценки

Характеристики генеральной совокупности	Формулы оценивания
Математическое ожидание $\mu = \sum_{i=1}^{n} x_i * P_i$	*Выборочное среднее* $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
Теоретическая дисперсия $\sigma_x^2 = \sum_{i=1}^{n} (x-\mu)^2 * P_i$	*Выборочная дисперсия* $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$

Рассмотрим ряд свойств оценок:

Оценка называется несмещенной если математическое ожидание оценки равнялось соответствующей характеристике генеральной совокупности. Если это не так, то оценка называется смещенной, и разница между ее математическим ожиданием и соответствующей характеристикой генеральной совокупности называется *смещением*.

Эффективная оценка – эта та оценка, у которой дисперсия минимальна.

Состоятельная оценка - если предел оценки по вероятности равен истинному значению характеристики генеральной совокупности, то есть оценка, которая дает точное значение для большой выборки независимо от входящих в нее конкретных наблюдений. Оценка называется *несостоятельной*, если распределение вероятностей при увеличении размера выборки, либо вообще не концентрируется, либо концентрируется около точки, отличной от истинного значения.

Оценка в виде определенного числа называется в статистике *точечной.* Наряду с точечной на практике часть пользуются интервальной оценкой параметра.

Интервальной оценкой параметра а называется числовой интервал (m^-, m^+), который с заданной вероятностью (доверительной вероятностью) $P_{дов}$ «накрывает» неизвестное значение параметра а.

Такой интервал называется доверительным. Размер доверительного интервала существенно зависит от объема выборки и уровня доверительной вероятности. Размер доверительного интервала уменьшается с увеличением объема выборки и сокращается с ростом доверительной вероятности.

При построении модели необходимо не только вычислять значения оценок параметров, но и проверить, насколько они отвечают требованиям, предъявляемым к ним. Последняя задача относится к проблеме проверки статистических гипотез и решается на основе методов проверки статистических гипотез.

Вопросы для обсуждения:

1. Каково место эконометрики в современной науки?
2. Каковы задачи эконометрики?
3. Почему дисциплина «эконометрика» появилась в российском образовании сравнительно недавно?
4. Назовите плюсы и минусы моделирования как инструмента исследования экономических процессов и явлений.
5. Перечислите этапы моделирования.
6. Чем отличаются переменные модели от ее параметров?
7. Какова роль спецификации модели в процессе моделирования?
8. В чем различие экономических, математических и эконометрических моделей?
9. Может ли выходная переменная модели быть одновременно и входной переменной? Если да, то в каких случаях?
10. Если результаты (эндогенные переменные, выходные параметры) модели явно неверные, в чем может быть причина неудачного моделирования?
11. Каковы минусы агрегирования при макроэкономическом моделировании?
12. Почему экономические показатели, рассчитанные на данных современных экономик, носят случайный характер?
13. Приведите примеры экономических показателей, к которым можно применить теорему Ляпунова.
14. Приведите примеры экономических показателей, которые можно рассматривать как дискретные и непрерывные случайные величины.
15. Объясните влияние количества наблюдений и σ на график нормального распределения.
16. Почему необходимо рассчитывать z – статистику?
17. Почему исследователи вынуждены рассчитывать оценки параметров?

Тест:

1 Впервые термин «эконометрика» был предложен:
   a. в 1910 году;
   b. в 30 –х годах XX века;
   c. в 90 –х годах XX века.

2 Эконометрика – это:
   a. сплав трех компонент: экономической теории, математических методов, компьютерных вычислений.
   b. Это сплав четырех компонент: экономической теории, статистических и математических методов, компьютерных вычислений

c. Это сплав трех компонент: статистических и математических методов, компьютерных вычислений.

3 Практическим материалом для построения эконометрических моделей являются:

a. качественные показатели**;**
b. теоретические предположения;
c. результаты наблюдений за изучаемыми экономическими объектами.

4 Задачей эконометрики является:

a. дать теоретические заключения по закономерностям, сформулированным в общей экономической теории;
b. дать количественные оценки выводам и закономерностям, сформулированным в общей экономической теории;
c. дать качественные оценки выводам и закономерностям, сформулированным в общей экономической теории.

5 Модель адекватна объекту-оригиналу:

a. если она с достаточной степенью точности приближения отражает закономерности процесса функционирования реального объекта;
b. если она с минимальной степенью точности приближения отражает закономерности процесса функционирования реального объекта;
c. если она не отражает закономерности процесса функционирования реального объекта.

6 Вероятностно - статистическая математическая модель:

a. это абстракция реального мира, в которой интересующие исследователя отношения между реальными элементами заменены подходящими отношениями между математическими категориями;
b. модель, имитирующая механизм функционирования гипотетического (не конкретного) явления или системы стохастической природы;
c. модель, в которой значения отдельных характеристик (параметров) оцениваются по результатам наблюдений (исходным статистическим данным), то есть это модель, характеризующая механизм функционирования конкретного явления или системы стохастической природы.

7 Эконометрическая модель - это модель:

a. гипотетического экономического объекта;

b. конкретно-существующего экономического объекта, построенная на гипотетических данных;
c. конкретно-существующего экономического объекта, построенная на статистических данных.

8 Модель, отражающая положительную зависимость предложения денег от ставки процента, является:
a. мезомоделью;
b. макромоделью;
c. микромоделью.

9 Выделение важнейших черт и свойств объекта и абстрагирование от второстепенных свойств, изучение основных закономерностей взаимодействия переменных, выявленных экономической теорией происходит в рамках этапа моделирования:
a. постановка задачи, спецификация модели;
b. подготовка исходной информации;
c. проверка адекватности.

10 Выбор методов оценки неизвестных параметров модели происходит в рамках этапа моделирования:
a. проверка адекватности;
b. проверка гипотез статистической значимости параметров;
c. оценивание неизвестных параметров.

11 Экономическая теория:
a. может рассмотреть каждый конкретный экономический объект;
b. рассматривает, как правило, общие закономерности поведения и развития отдельных экономических систем;
c. моделирует поведение конкретного объекта по конкретным статистическим данным.

12 Эндогенные переменные в модели:
a. переменные, значения которых формируются в процессе и внутри функционирования анализируемой социально – экономической системы;
b. переменные, значения которых формируются вне модели, задаются «извне», автономно от модели;
c. переменные управляемые и планируемые.

13 Предопределенные переменные включают:
a. все экзогенные и эндогенные переменные;
b. все экзогенные переменные и лаговые эндогенные переменные;
c. лаговые экзогенные и эндогенные переменные.

14 Принцип спецификации модели заключается в том, что:

a. количество уравнений в модели больше количества эндогенных переменных, участвующих в ней;
b. количество уравнений в модели меньше количества эндогенных переменных, участвующих в ней;
c. количество уравнений в модели равно количеству эндогенных переменных, участвующих в ней.

15 Лаговые переменные**:**
a. переменные модели, отнесенные к текущему моменту времени;
b. переменные модели, отнесенные к предшествующим моментам времени.
c. переменные модели, отнесенные к будущим моментам времени.

16 Модели, содержащие датированные переменные:
a. динамические модели;
b. пространственные модели;
c. открытые модели**.**

17 Модели, в состав которых входят только текущие эндогенные переменные:
a. динамические модели;
b. пространственные модели;
c. замкнутые модели (закрытые).

18 Чем точнее информация об исследуемом объекте, тем:
a. больше доля «черного ящика»;
b. меньше доля «черного ящика»;
c. качество информации не влияет на долю «черного ящика» в моделировании.

19 Степени свободы в наборе данных определяют число единиц данных:
a. независимых друг от друга, которые могут нести отдельные элементы информации;
b. зависимых друг от друга, которые могут нести отдельные элементы информации;
c. независимых друг от друга, которые могут нести общие элементы информации.

20 Оценочные значения характеристик рассчитываются по данным:
a. генеральной совокупности;
b. выборки;
c. как по выборочным данным, так и по данным генеральной совокупности.

21 Чем больше σ, тем выборка:

a. лучше;
b. хуже;
c. σ не влияет на качество выборки.

22 Оценка называется несмещенной если:
a. математическое ожидание оценки больше соответствующей характеристике генеральной совокупности;
b. математическое ожидание оценки меньше соответствующей характеристике генеральной совокупности;
c. математическое ожидание оценки равняется соответствующей характеристике генеральной совокупности.

23 Эффективная оценка:
a. эта та оценка, у которой дисперсия минимальна;
b. эта та оценка, у которой дисперсия максимальна;
c. эта та оценка, у которой дисперсия находится в определенном диапазоне.

24 Оценка в виде определенного числа называется:
a. точечной;
b. интервальной.

25 Размер доверительного интервала:
a. увеличивается с увеличением объема выборки;
b. увеличивается с ростом доверительной вероятности;
c. уменьшается с увеличением объема выборки.

Задания для самостоятельной работы:

№ 1.

В таблице приведены данные чистого дохода как процента от стоимости акционерного капитала для 42 - х компаний. Рассчитайте выборочные среднюю и дисперсию для приведенных данных компаний**.**

17	14	15	14	11	12	9	18	14	7	17	14	15	20	12	14	9	1	18	27	11
11	23	36	25	10	18	14	23	13	2	6	15	14	10	7	13	8	11	16	44	1

№ 2.

Рассчитайте вероятность попадания величины z в конечный интервал [0,32; 2,27].

№ 3.

Найдите интервал, в который попадает случайная величина z с вероятностью 0,5557

№ 4.

Анализируются объемы ежедневных продаж некторого товара за 60 дней. Получены следующие данные:

5,6,3,2,7,7,,6,6,10,11,6,4,5,6,3,12,9,10,7,4,6,7,8,8,10,5,5,4,3,6,6,7,7,8,8,10,6,4,5,6, 127,7,8,11,9,10,5,6,4,2,7,11,8,7,9,5,6,9,5.

Необходимо: Построить статистический ряд. Определить размер выборки. Построить эмпирическую функцию распределения и ее график. Определить выборочное среднее, дисперсию, среднее квадратическое отклонение.

№ 5.

Модель динамики цен и капитала. Экономический процесс в текущий момент времени описывается следующими переменными:

y_{1t}— темп изменения ежемесячной заработной платы;

y_{2t}— темп изменения цен;

x_{1t}— процент безработных;

x_{2t}— темп изменения постоянного капитала;

x_{3t}— темп изменения цен на импорт сырья.

Требуется составить спецификацию модели товарно-денежного рынка.

При составлении спецификации учесть следующие экономические закономерности:

- темп изменения ежемесячной заработной платы в текущем периоде зависит от темпа изменения цен и процента безработных в том же периоде;
- текущее значение темпа изменения цен определяется текущим темпом изменения цен и зависит от темпа изменения постоянного капитала и темпа изменения цен на импорт в том же периоде.

№ 6.

Экономическим объектом служит товарно-денежный рынок, состояние которого в текущий момент времени описывается следующими количественными переменными:

R_t— процентная ставка;

Y_t— объем ВВП;

M_t— объем денежной массы;

I_t— объем внутренних инвестиций*;*

G_t— объем государственных расходов.

Требуется составить спецификацию модели товарно-денежного рынка;

При составлении спецификации учесть следующие экономические закономерности:

- размер процентной ставки в текущем периоде возрастает с ростом текущего значения ВВП и ростом текущего объема денежной массы;
- текущее значение объема ВВП возрастает с ростом текущих внутренних инвестиций и ростом процентной ставки;

- текущий объем внутренних инвестиций падает с ростом процентной ставки.

№ 7.

Экономическим объектом служит закрытая национальная экономика без государственного вмешательства. Ее состояние в текущий момент времени описывается следующими переменными:

M_t — доля импорта в текущем объеме ВВП;

N_t — количество поданных прошений об освобождении от таможенных пошлин;

S_t — количество удовлетворенных прошений об освобождении от таможенных пошлин;

X_t — реальный объем чистого экспорта;

$У_t$ — объем ВВП.

Требуется:

1. Составить спецификацию упрощенной модели протекционизма Сальватора;
2. Найти количество экзогенных переменных, лаговых переменных и предопределенных переменных.

   При составлении спецификации учесть следующие экономические закономерности:

- текущая доля импорта в ВВП зависит от ее доли в предыдущем периоде, а так же от количества поданных и удовлетворенных заявок на освобождение от таможенных пошлин в текущем периоде;
- текущее количество поданных прошений зависит от текущего значения доли импорта в ВВП, количества удовлетворенных прошений в прошедшем периоде и объема ВВП в текущем периоде;
- количество удовлетворенных прошений определяется текущими значениями доли импорта в ВВП, числа поданных прошений и объема чистого экспорта.

№ 8.

Экономическим объектом служит закрытая национальная экономика без государственного вмешательства. Состояние объекта определяется следующими переменными:

$У_t$ — объем ВВП;

I_t — объем инвестиций;

C_t — объем потребления;

R_t —банковская процентная ставка;

G_t — объем государственных расходов;

M_t — объем денежной массы, находящейся в обороте.

Требуется составить спецификацию упрощенной модели протекционизма Сальватора.

Переменные объекта взаимодействуют между собой следующим образом:

- текущий объем потребления растет с ростом текущего объема ВВП и зависит от объема потребления в предшествующем периоде;
- текущий объем инвестиций зависит от объема ВВП в предшествующем периоде и падает с ростом банковского процента в текущем периоде;
- размер банковской процентной ставки в текущем периоде определяется объемами ВВП и денежной массы в том же периоде;
- объем ВВП в текущем периоде складывается из текущих объемов инвестиций, потребления и государственных расходов.

.

ТЕМА 2. МОДЕЛЬ ПАРНОЙ РЕГРЕССИИ

Ключевые слова:

Спецификация модели парной регрессии. Результативный признак, признак-фактор и стохастическая переменная в модели. Графический, аналитический и экспериментальный способы задания функции. Коэффициент вариации случайной величины. Коэффициент корреляции. Метод наименьших квадратов**.** Стандартная ошибка коэффициентов регрессии. t критическое. Доверительный интервал. Коэффициент детерминации. Общая сумма квадратов отклонений. Факторная сумма квадратов отклонений. Остаточная сумма квадратов отклонений. Дисперсии на одну степень свободы. F- критерий Фишера. Средняя ошибка аппроксимации Прогнозное значение. Интервалы прогноза.

Основные теоретические аспекты темы:

2.1. Спецификация модели парной регрессии: понятие и способы задания функций.

Можно указать два варианта рассмотрения связей между переменными x и y. В первом случае обе переменные считаются равноценными в том смысле, что они не подразделяются на первичную и вторичную (независимую и зависимую). Основным в этом случае является вопрос о силе связи между этими переменными. При исследовании силы линейной зависимости между такими переменными обращаются к корреляционному анализу, основной мерой которого является коэффициент корреляции.

Коэффициент корреляции величин x и y (r_{xy})– свидетельствует о наличии или отсутствии линейной связи между переменными:

$$r_{xy}=\frac{cov(x,y)}{s_x s_y}=\frac{\overline{yx}-\overline{y}*\overline{x}}{s_x\ s_y};\qquad r_{xy}\in[-1;1].$$

Если: $r_{xy}=-1$, то наблюдается строгая отрицательная связь;
$r_{xy}=1$, то наблюдается строгая положительная связь;

$r_{xy} = 0$, то линейная связь отсутствует.

Другой вариант рассмотрения взаимосвязей выделяет одну из переменных как независимую (объясняющую), а другую как зависимую (объясняемую). В этом случае изменение первой служит причиной для изменения другой, а связь называют регрессионной.

Парная регрессия – регрессия между двумя переменными y и x, то есть модель вида: $y=f(x)+\varepsilon$;

где: y - зависимая переменная, фактическое значение результативного признака;

x – независимая, объясняющая, переменная, признак- фактор;

ε - возмущение, случайная, стохастическая переменная.

Причинами существования стохастической переменной являются:

1. Не учтенные в модели объясняющие переменные.
2. Агрегирование переменных.
3. Неправильное описание структуры модели.
4. Неправильная функциональная спецификация.
5. Ошибки измерения

Основные типы функций, используемые при количественной оценке связей:

Линейная функция: $y=a+bx$;

Нелинейные функции: $y= a+b/x$ - гипербола;

$y=a+bx+cx^2$ – парабола;

$y=a+bx+cx^2+dx^3$-кубический многочлен;

$y=ax^b$ –степенная функция;

$y=ab^x$-показательная функция;

$y=a+b\lg x$-логарифмическая функция;

$y= 1/(a+bx)$;

$y=a+bx+c(1/x)$;

$y=1/(a+bx+cx^2)$;

$\lg y=a+bx+cx^2$

$y=a/(1+be^{-cx})$.

В парной регрессии выбор вида математической функции $y_x=f(x)$ может быть осуществлен тремя методами:

- Графическим. При изучении зависимости между двумя признаками графический метод подбора вида уравнения регрессии достаточно нагляден. Он базируется на поле корреляции.
- Аналитическим, то есть исходя из теории изучаемой взаимосвязи. Метод основан на изучении материальной природы связи исследуемых признаков.
- Экспериментальным, то есть путем сравнения величины остаточной дисперсии, рассчитанной при различных моделях:

$$D_{ост} = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_x)^2 .$$

Чем меньше величина остаточной дисперсии, тем в меньшей мере наблюдается влияние прочих, не учитываемых в уравнении регрессии факторов, и тем лучше уравнение регрессии подходит к исходным данным. При обработке статистических данных на компьютере перебираются разные математические функции в автоматическом режиме и из них выбирается та, для которой остаточная дисперсия является наименьшей.

Если остаточная дисперсия оказывается примерно одинаковой, то на практике предпочтение отдается более простым видам функций, так как они в большей степени поддаются интерпретации и требуют меньшего объема наблюдений.

Результаты многих исследований подтверждают, что число наблюдений должно в 6-7 раз превышать число рассчитываемых параметров при переменной x.

2.2. Параметризация модели: оценка параметров уравнения линейной регрессии. Метод наименьших квадратов.

Дано: $y = \alpha + \beta x + u$ - уравнение парной регрессии для генеральной совокупности,

где: α, β – истинные коэффициенты модели;

u– стохастическая переменная модели.

В силу того, что генеральную совокупность собрать и проанализировать в большинстве случаев невозможно, исследователи собирают и оценивают выборочную совокупность данных.

Необходимо по выборке, состоящей из n статистических значений независимой, объясняющей переменной x: $x_1, \ldots x_n$ и n статистических значений зависимой переменной y: $y_1, \ldots y_n$ построить уравнение парной регрессии $y=a+bx+\varepsilon$,

где**:** a,b - оценки α, β соответствующие генеральной совокупности;

ε - оценка u.

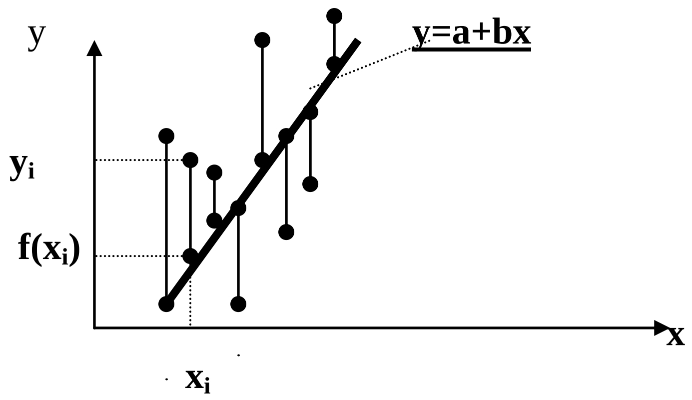


Рис.2.2.1. Корреляционное поле и линия парной регрессии.

Стандартное отклонение случайной величины x (выборочная дисперсия s_x) - мера разброса случайной величины вокруг среднего значения выборки.

$$s_x = \sqrt{\overline{x^2} - \bar{x}^2}\ .$$

Коэффициент вариации случайной величины x (V_x.) – мера относительного разброса случайной величины. Показывает, какую долю среднего значения случайной величины составляет ее средний разброс.

$$V_x = \frac{s_x}{\bar{x}}.$$

Метод наименьших квадратов – метод оценивания параметров линейной регрессии, минимизирующий сумму квадратов отклонений наблюдений зависимой переменной от искомой линейной функции.

$$RSS = \sum_{i=1}^{n} e_i^2 = \sum_i (y_i - f(x_i))^2 = \sum_{i=1}^{n} (y_i - (a + bx_i))^2 \to \min$$

где: y_i – статистические значения зависимой переменной;
$y_x = f(x_i)$ - теоретические значения зависимой переменной, рассчитанные с помощью уравнения регрессии.

Исследуя RSS как функцию двух переменных a и b, на экстремум, строим систему уравнений:

$$\begin{cases} \dfrac{dRSS}{da} = 0 \\ \dfrac{dRSS}{db} = 0 \end{cases}.$$

Решением данной системы уравнений будут значения коэффициентов регрессии a и b, вычисляемы следующим образом:

$$b = \frac{\overline{yx} - \bar{y} * \bar{x}}{\overline{x^2} - \bar{x}^2}; \quad a = \bar{y} - b * \bar{x};$$

где: $\overline{yx}$ - среднее значение y*x;
$\bar{y}$ - среднее значение y;
$\bar{x}$ - среднее значение x.

2.3. Интерпретация уравнения парной регрессии: экономический смысл параметров регрессии.

Существуют два этапа интерпретации уравнения регрессии. Первый этап состоит в словесном истолковании уравнения так, чтобы это было понятно человеку, не являющемуся специалистом в области эконометрики. На втором этапе необходимо решить, следует ли ограничиться этим или следует провести более детальное исследование зависимости.

Параметр b показывает среднее изменение результата y с изменением фактора x на единицу.

Знак при коэффициенте регрессии b показывает направление связи: при b>0 – связь прямая, а при b<0 – связь обратная.

Формально, параметр a=y, когда x=0. Если x не может быть равен 0 исходя из экономической сущности x, то a не имеет экономического смысла. Параметр a может и не иметь экономического содержания. Попытки экономически интерпретировать параметр a могут привести к абсурду, особенно при a<0.

Интерпретировать можно только знак при a: если a>0, то относительное изменение результата происходит медленнее, чем изменение фактора, то есть вариация результата меньше вариации фактора: $V_y<V_x$, и наоборот.

2.4. Эксперимент Монте – Карло. Свойства коэффициентов регрессии

Коэффициенты регрессии, вычисленный методом наименьших квадратов, - это особая форма случайной величины, свойства которой зависят от свойств остаточного члена в уравнении.

Эксперимент Монте – Карло – это искусственный контролируемый эксперимент, дающий возможность проверки хорошие или плохие оценки дает метод наименьших квадратов. Эксперимент Монте-Карло состоит из трех частей:

Этап 1:

1. Выбираются истинные значения α и β;
2. В каждом наблюдении выбирается значение x;
3. Используется процесс генерирования случайных чисел для получения значений случайного фактора u

Этап 2:

В каждом наблюдении генерируется значение y с использованием соотношения $y = \alpha + \beta x + u$.

Этап 3:

Применяется регрессионный анализ для получения оценок a и b с использованием полученных значений y и x.

Вывод по эксперименту:

Свойства коэффициентов регрессии существенным образом зависят от свойств случайной переменной u. Чтобы регрессионный анализ, основанный на обычном методе наименьших квадратов, давал наилучшие из всех возможных результаты, случайная переменная должна удовлетворять четырем условиям, известным как условия Гаусса – Маркова. Если эти условия не выполнены, исследователь должен это осознавать. Если корректирующие действия возможны, то аналитик должен быть в состоянии их выполнить. Если ситуацию исследовать невозможно, исследователь должен быть способен судить, насколько серьезно это может повлиять на результаты.

Условия Гаусса –Маркова:

1 – е условие Гаусса – Маркова: $E(u_i)=0$ для всех наблюдений.

Математическое ожидание случайной переменной в любом наблюдении должно быть равно нулю. Иногда случайная переменная будет положительной, иногда отрицательной, но он не должен иметь систематического смещения ни в каком из двух возможных направлений.

Фактически, если уравнение регрессии включает постоянный член, то обычно бывает разумно предположить, что это условие выполняется автоматически, так как роль постоянного члена состоит в отражении любой систематической, но постоянной составляющей в y, которую не учитывают объясняющие переменные.

2 – е условие Гаусса – Маркова: $\sigma^2(u_i) = \sigma^2_u$ *для всех наблюдений.*

Теоретическая дисперсия случайной переменной постоянна для всех наблюдений. Не должно быть априорной причины для того, чтобы случайная переменная порождала большую ошибку в одних наблюдениях, чем в других.

Величина σ^2_u, конечно неизвестна. Одна из задач регрессионного анализа состоит в оценке стандартного отклонения случайного члена. Если рассматриваемое условие не выполняется, то коэффициенты регрессии, найденные по МНК, будут неэффективны.

3 – е условие Гаусса – Маркова: $cov(u_i, u_j) = 0$.

Отсутствие систематической связи между значениями случайной переменной в любых двух наблюдениях. То есть, если случайная переменная велика и положительна в одном наблюдении, это не должно обусловливать систематическую тенденцию к тому, что она будет большой и положительной в следующем наблюдении (или большой и отрицательной, или малой и положительной, или малой и отрицательной).

Значения случайной переменной должны быть абсолютно независимой друг от друга. Если рассматриваемое условие не выполняется, то коэффициенты регрессии, найденные по МНК, будут неэффективны.

4 – е условие Гаусса – Маркова: $cov(u_i, x_j) = 0$

Случайная переменная должна быть распределена независимо от объясняющих переменных.

Предположение о нормальности

Наряду с условиями Гаусса – Маркова обычно также предполагается нормальность распределения случайной переменной. Если случайная переменная u нормально распределена, то нормально будут распределены и коэффициенты регрессии.

Теорема Гаусса – Маркова

Если условия Гаусса – Маркова для случайной переменной выполнены, то коэффициенты регрессии, построенные обычным МНК, будут наилучшими линейными несмещенными оценками:

- несмещенными ($E(a) = \alpha$; $E(b) = \beta$);
- линейными, так как они являются линейными функциями значений y;
- наилучшими, так как они являются наиболее эффективными в классе всех несмещенных линейных оценок.

2.5. Точность коэффициентов регрессии

Рассмотрим теоретические дисперсии оценок a и b. Они задаются следующими выражениями:

$$\sigma_a^2 = \frac{\sigma_u^2}{n}\left[1+\frac{\overline{x^2}}{\sigma_x}\right] \text{ и } \sigma_b^2 = \frac{\sigma_u^2}{n\sigma_x}.$$

Из выражений можно сделать следующие выводы:

1. Дисперсии a и b обратно пропорциональны числу наблюдений в выборке. Чем больше информации у исследователя, тем более точными, скорее всего, будут оценки.
2. Дисперсии оценок пропорциональны дисперсии случайного члена. Чем больше дисперсия случайного члена в зависимости, тем менее точными, скорее всего, будут оценки при прочих равных условиях.
3. Дисперсии коэффициентов регрессии обратно пропорциональны дисперсии x. Коэффициенты регрессии вычисляются на основании предположения, что наблюдаемые изменения y происходят вследствие изменений x, но в действительности они лишь отчасти вызваны изменениями в x, а отчасти вариациями в u. Чем меньше дисперсия x, тем больше будет относительное влияние фактора случайности при определении отклонений в y, и тем более вероятно, что регрессионный анализ даст неточные результаты.

На практике невозможно вычислить теоретические дисперсии a и b, так как σ_u^2 неизвестно. Но можно получить оценку σ_u^2 на основе остатков. Выборочная дисперсия остатков, которую можно измерить, может быть использована для определения σ_u^2, которую измерить невозможно. Выборочная дисперсия s_u^2 будет представлять собой несмещенную оценку σ_u^2

$$s_u^2 = \frac{n}{n-2}\sigma_e.$$

Можно получить оценки теоретических дисперсий для a и b и после извлечения квадратных корней - оценки их стандартных отклонений. Вместо громоздкого термина – «оценка стандартного отклонения функции плотности вероятности» коэффициента регрессии будем использовать термин *«стандартная ошибка»* коэффициента регрессии, которую в сокращении записывают с.о. (а) и с.о.(b), или m_a и m_b.

Таким образом, для парного регрессионного анализа с*тандартные ошибки коэффициентов регрессии* имеют вид:

$$\text{с. о. (а)} = m_a = \sqrt{\frac{s_u^2}{n}\left[1+\frac{\overline{x^2}}{\sigma_x}\right]} = \sqrt{\frac{s_\varepsilon^2}{(n-2)}\left(1+\frac{\overline{x^2}}{s_x^2}\right)} = \sqrt{\frac{\sum_{i=1}^{n}(y_i-y_x)^2 * \sum_i x^2}{(n-2)*n*\sum_i (x_i-\overline{x})^2}};$$

$$\textbf{с}.\ \text{o. (b)} = m_b = \sqrt{\frac{s_u^2}{n\sigma_x}} = \sqrt{\frac{s_\varepsilon^2}{(n-2)s_x^2}} = m_b = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_x)^2/(n-2)}{\sum_i (x_i - \overline{x})^2}}\ .$$

2.6. Оценка значимости коэффициентов линейной регрессии: проверка гипотез, относящихся к коэффициентам регрессии.

Большинство эконометрических моделей требуют многократного улучшения и уточнения. Для этого необходимо проведение соответствующих расчетов, связанных с установлением выполнимости или невыполнимости тех или иных предпосылок, анализом качества найденных оценок, достоверностью полученных выводов. При построении модели необходимо не только вычислить значения оценок параметров, но и проверить, насколько они отвечают требованиям, предъявляемым к ним. Эта задача относится к проблеме проверки статистических гипотез и решается на основе методов проверки статистических гипотез.

Статистическая гипотеза – любое априорное предположение о значении параметра или вида закона распределения.

Приняты следующие обозначения. Основную гипотезу обозначают символом H_0, за которым следует математическая формулировка гипотезы. Альтернативные гипотезы в обозначении имеют символ, отличный от нуля, например:

$$H_0:\quad a=0;$$
$$H_1:\quad a\neq 0.$$

Наряду с основной гипотезой могут быть выдвинуты и альтернативные к ней гипотезы.

Проверку статистической гипотезы осуществляют на основании данных выборки. Для этого используют специально подобранную случайную величину (статистику, критерий), точное или приближенное значение которой известно. Эту величину обозначают:

- Z – если она имеет стандартизированное нормальное распределение;
- T – если она распределена по закону Стьюдента;
- χ^2 -если она распределена по закону χ^2.
- F- если она имеет распределение Фишера.

Статистическим критерием называют случайную величину K, которая служит для проверки нулевой гипотезы. После выбора определенного критерия множество всех его возможных значений разбивают на два непересекающихся подмножества: одно из них содержит значения критерия, при которых нулевая гипотеза откланяется, другое – при которых она не отклоняется.

Критической областью называют совокупность значений критерия, при которых нулевую гипотезу отклоняют.

Областью принятия гипотезы называют совокупность значений критерия, при которых нулевую гипотезу не отклоняют.

Основной принцип проверки статистических гипотез: если наблюдаемое значение критерия К (вычисленное по выборке) принадлежит критической области, то нулевую гипотезу отклоняют. Если же наблюдаемое значение критерия К принадлежит области принятия гипотезы, то нулевую гипотезу не отклоняют (принимают).

Точки, разделяющие критическую область и область принятия гипотезы, называют критическими.

Общая схема проверки гипотез:

1. Формулировка проверяемой (нулевой H_0) и альтернативной H_1 гипотез. Формулировка делается, как в описательной (вербальной) форме, так и в математическом виде.
2. Искусственно создается случайная величина z (*статистика* критерия гипотезы H_0), тесно связанная с выдвинутой гипотезой и с известным законом распределения. Закон распределения случайной переменной, которая содержится в сформулированной основной гипотезе, может быть неизвестен, а, следовательно, ничего нельзя сказать о ее поведении. Поэтому, создается случайная переменная, о поведении которой можно судить по ее закону распределения.
3. Задается значение доверительной вероятности Рдов. Область определения созданной случайной переменной разбивается на две непересекающиеся области: область, где выдвинутая гипотеза H_0 принимается с вероятностью Рдов и критическую область.

   Граница, разделяющая область принятия гипотезы и критическую область называется *критическим значением* распределения.
4. Проверяется появление случайного события $z \in Z(H_0)$. Если событие появилось, то гипотеза H_0 принимается как не противоречащая опытным данным. Если оно не появилось, то гипотеза H_0 отклоняется.

При проверке статистических гипотез, связанных с анализом статистической значимости коэффициентов регрессии используют дробь Стьюдента:

$$t = \frac{Разница\_между\_оценкой\_регрессии\_и\_гипотетическим\_значением}{Стандартное\_отклонение\_(стандартная\_ошибка)\_оценки};$$

$$t_a = \frac{a - \alpha_0}{m_a}; \qquad t_b = \frac{b - \beta_0}{m_b}.$$

Дробь Стьюдента подчиняется закону распределения Стьюдента и применяется для тестирования статистической гипотезы о равенстве оценки истинному значению параметра:

Нулевая гипотеза: H_0: $\beta = \beta_0$;

Альтернативная гипотеза: H_1: $\beta \neq \beta_0$.

Методические указания по нахождению критических значений t – распределения Стьюдента.

Работа с таблицами t- распределения Стьюдента [Приложение А, Таблица А 2.].

В таблице функции распределения Стьюдента приводятся для различных чисел степеней свободы ν критические точки, соответствующие приведенным в верхней строке таблицы уровням значимости критерия α.

Степени свободы в наборе данных – определяют число единиц данных, независимых друг от друга, которые могут нести отдельные элементы информации. Оценивание каждого параметра в уравнении регрессии поглощает одну степень свободы в выборке. Следовательно, число степеней свободы равно количеству наблюдений в выборке минус количество оцениваемых параметров.

Уровень значимости критерия α – это вероятность превышения t – статистикой приведенного в таблице критического значения при соответствующем числе степеней свободы ν (вероятность попадания в правый «хвост» распределения). α =1-Рдов.

Работа в EXCEL:

Критические значения дроби Стьюдента находятся с помощью функции СТЬЮДРАСПОБР. Параметрами этой функции являются уровень значимости и степени свободы.

Тогда, условие того, что оценка регрессии не должна приводить к отказу от нулевой гипотезы $H_0: \beta = \beta_0$ имеет вид**:**

$$-t_{крит} \le \frac{b-\beta_0}{m_b} \le t_{крит};$$

а от нулевой гипотезы H_0: $\alpha = \alpha_0$**:** $-t_{крит} \le \frac{a-\alpha_0}{m_a} \le t_{крит}$.

При проверке гипотез могут возникнуть *ошибки I и II рода:*

Ошибка I рода имеет место в том случае, когда отвергается истинная нулевая гипотеза.

Ошибка II рода возникает, когда не отвергается ложная гипотеза.

Чем ниже критическая вероятность, тем меньше риск получения ошибок I рода. Если используйте уровень значимости в 5%, то отвергается истинная гипотеза в 5% случаев. Если уровень значимости составляет 1%, то ошибка I рода совершается в 1% случаев. Следовательно, 1% уровень значимости более надежен.

Если нулевая гипотеза ложна, то чем выше уровень значимости, тем шире область принятия гипотезы, тем выше вероятность того, что вы не отвергаете ее, тем выше риск допущения ошибки II рода.

Если настаивайте на очень высоком уровне значимости, то столкнетесь с относительно высоким риском допущения ошибки II рода, если гипотеза окажется ложной. Если выбираете низкий уровень значимости, то оказываетесь перед относительно высоким риском допущения ошибки I рода, если гипотеза истинна.

Большинство исследователей выбирают достаточно простую форму обеспечения гарантий и осуществляют проверку на обоих уровнях значимости, представляя результаты каждой такой проверки.

Часто нет необходимости непосредственно ссылаться на оба результата. Если отклоняется гипотеза при 1% уровне значимости, то автоматически она отклоняется и при уровне значимости в 5%. Если не отвергается гипотеза при уровне значимости в 5%, то из этого следует, что она не отвергается и при 1% уровне значимости.

Только в одном случае необходимо представить оба результата: если гипотеза отвергается на 5%, но не на 1% уровне значимости.

Доверительные интервалы отражают все гипотезы совместимые с результатом оценивания регрессии

Доверительный интервал для β:

$$b - t^* m_b \le \beta \le b + t^* m_b.$$

Данное неравенство отражает, что любое гипотетическое значение для β, которое удовлетворяет данному соотношению, будет автоматически совместимо с оценкой b.

Доверительный интервал для α:

$$a - t^* m_a \le \alpha \le a + t^* m_a.$$

Данное неравенство отражает, что любое гипотетическое значение для α, которое удовлетворяет данному соотношению, будет автоматически совместимо с оценкой a.

Заметим, что если выбирается 5 –процентный уровень значимости, то соответствующим доверительным интервалом считается 95-процентный интервал. Если выбирается 1-процентный уровень, то можно получить доверительный интервал с 99% и т.д. Так как $t_{крит}$ будет больше для 1-процентного уровня значимости, чем для 5-процентного, то интервал в 99% будет шире интервала в 95%.

2.7. Качество оценки: коэффициент детерминации. F – критерий Фишера для проверки качества оценивания

Тестирование качества спецификации модели направлено на выявление факторов, не оказывающих влияния на формирование эндогенной переменной. После проверки значимости каждого коэффициента регрессии обычно проверяется общее качество уравнения регрессии, которое оценивается по тому, как хорошо эмпирическое уравнение согласуется со статистическими данными, то есть, на сколько широко рассеяны точки наблюдений относительно линии регрессии. Очевидно, если все точки лежат на построенной прямой, то регрессия y от x «идеально» объясняет поведение зависимой переменной. В реальной жизни такая ситуация практически не встречается. Обычно поведение y лишь частично объясняется влиянием переменной x, что объясняет существование случайного члена в уравнении регрессии:

$$y = a + bx + \varepsilon = y_x + \varepsilon .$$

В данном уравнении y_x -это вклад в значение y, вызванный изменением x, а случайный член – влияние случайных факторов, которые не связаны с изменением x.

Отсюда вытекает идея тестирования. Необходимо установить, какое из слагаемых вносит наибольший вклад в общий разброс наблюдаемых значений эндогенной переменной. Характеристикой разброса случайной переменной служит дисперсия. Следовательно, необходимо определить, какое из слагаемых превалирует в функции дисперсии эндогенной переменной.

Введем следующие обозначения:

Сумма квадратов ошибок или факторная сумма квадратов (Error Sum Squares, ESS):

$$\text{ESS} = \sum_i (y_x - \bar{y})^2 = b^2 * (\sum_i x^2 - n * \bar{x}^2).$$

Регрессионная сумма квадратов или остаточная сумма квадратов отклонений (Regression Sum Squares, RSS):

$$\text{RSS} = \sum_{i=1}^{n} (y_i - y_x)^2$$

Общая сумма квадратов (Total Sum Squares, TSS)

$$\text{TSS} = \text{ESS} + \text{RSS}$$

В качестве меры влияния фактора x на формирование значения эндогенной переменной y вводится коэффициент детерминации.

Коэффициент детерминации (R^2_{xy} , r_{xy}^2) – характеризует долю дисперсии результативного признака y, объясняемую регрессией, в общей дисперсии результативного признака. Чем ближе r_{xy}^2 к 1, тем качественнее регрессионная модель, то есть исходная модель хорошо аппроксимирует исходные данные и спецификация качественная. Чем ближе r_{xy}^2 к 0, тем некачественная спецификация модели.

$$R^2_{xy} = r_{xy}^2 = \frac{ESS}{TSS}.$$

Коэффициент детерминации – величина случайная**,** так как его значение вычислено по случайной выборке. Следовательно, для тестирования гипотезы о том, что выбранный фактор x не оказывает влияния на формирование значения эндогенной переменной, согласно алгоритму проверки статистических гипотез, необходимо создать случайную переменную, связанную с гипотезой, закон распределения которой был бы известен.

Выдвигаем нулевую и альтернативную гипотезы:

Нулевая гипотеза: H_0: $R^2_{xy} = 0$ или b=0;

Альтернативная гипотеза: H_1: $R^2_{xy} \neq 0$ или b≠0.

Рассчитываем дисперсии на одну степень свободы:

$$TMS = \frac{TSS}{n-1}; \; EMS = \frac{ESS}{1}; \; RMS = \frac{RSS}{n-1-1};$$

где: $v_1=df_{ESS}=1$ -степень свободы факторной суммы квадратов отклонений;

$v_2=d_{fRSS}=n-1-1$ -степень свободы остаточной суммы квадратов отклонений.

Если нулевая гипотеза H_0 справедлива, то факторная и остаточная дисперсии не отличаются друг от друга. Если H_0 не справедлива, то факторная дисперсия превышает остаточную в несколько раз.

Для проверки гипотезы в качестве статистики принимается переменная F, которая вычисляется по формуле:

$$F_{вычисляемое} = \frac{EMS}{RMS}.$$

Переменная F подчиняется закону распределения Фишера с параметрами m и n-m-1,

где: m -количество факторов в уравнении;

n -количество наблюдений в выборке.

Английским статистиком Снедекером разработаны таблицы критических значений F – отношений при разных уровнях существенности нулевой гипотезы и различном числе степеней свободы $v_1=df_{ESS}$; $v_2=df_{RSS}$ [Приложение А, Таблица А 3.].

Табличное значение F – критерия – это максимальная величина отношения дисперсий, которая может иметь место при случайном расхождении их для данного уровня вероятности наличия нулевой гипотезы.

Если $F_{\textbf{вычисляемое}} > F_{табличное,}$ то уравнение регрессии значимо, то есть фактор x существенно влияет на формирование величины y. Следовательно, H_0 отклоняется.

Часто нет необходимости проверять гипотезу для обоих уровней значимости. Если отклоняется гипотеза при 1% уровне значимости, то автоматически она отклоняется и при уровне значимости в 5%. Если не отвергается гипотеза при уровне значимости в 5%, то из этого следует, что она не отвергается и при 1% уровне значимости. Проверять гипотезу для обоих уровней значимости стоит только в том случае, когда нулевая гипотеза отвергается при уровне значимости в 5%, но не отвергается при уровне значимости в 1%.

2.8. Прогнозирование на основе линейного уравнения регрессии. Интервальный прогноз.

В прогнозных расчетах по уравнению регрессии определяется предсказываемое y_P значение как точечный прогноз y_x при $x_P=x_k$, то есть путем подстановки в линейное уравнение регрессии $y_x = a + b * x$ соответствующего значения x.

Точечное прогнозное значение: $y_P = a + b * x_P$.

Однако точечный прогноз явно нереален, поэтому он дополняется расчетом стандартной ошибки m_y:

$$m_y = \sqrt{RMS*(\frac{1}{n}+\frac{(x_P-\overline{x})^2}{\sum(x_i-\overline{x})^2})}\text{; } x_P=x_k.$$

Тогда интервальная оценка прогнозного значения имеет вид:

$$y_P - t_{крит} * m_y \le y^* \le y_P + t_{крит} * m_y.$$

Рассмотренная формула стандартной ошибки предсказываемого среднего значения у при заданном значении x_k характеризует ошибку положения линии регрессии. Величина стандартной ошибки достигает минимума при $x_k = \bar{x}$ и возрастает по мере того, как «удаляется» от $\bar{x}$ в любом направлении. То есть , чем больше разность между x_k и $\bar{x}$, тем больше ошибка m_y, с которой предсказывается среднее значение у для заданного значения x_k.

Можно ожидать наилучшие результаты прогноза, если признак фактор х находится в центре области наблюдений х, и нельзя ожидать хороших результатов прогноза при удалении x_k от $\bar{x}$. Если же значение x_k оказывается за пределами наблюдаемых значений х, используемых при построении линейной регрессии, то результаты прогноза ухудшаются в зависимости от того, насколько x_k отклоняется от области наблюдаемых значений фактора х.

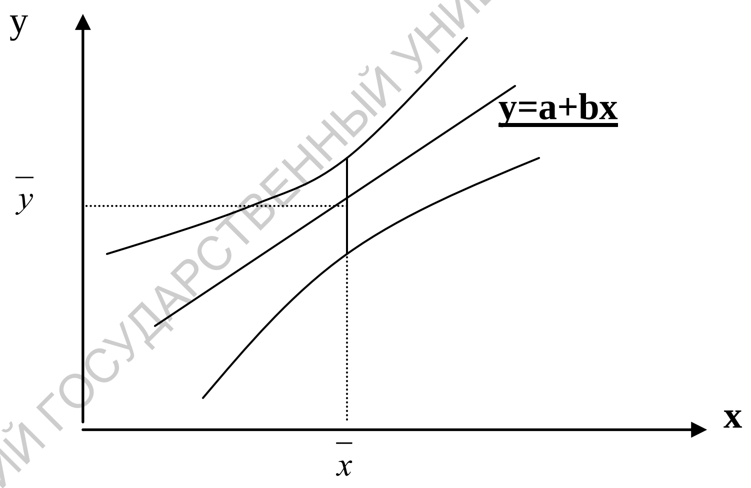


Рис.2.8.1. Доверительный интервал линии регрессии.

На графике, представленном на рис.2.8.1. доверительные границы представляют собой гиперболы, расположенные по обе стороны от линии регрессии. Рисунок показывает, как изменяются пределы в зависимости от изменения x_k : две гиперболы по обе стороны от линии регрессии определяют 95% -ные доверительные интервалы для среднего значения у при заданном значении х (верхняя доверительная граница и нижняя доверительная граница).

Фактические значения у варьируют около среднего значения y_x. Индивидуальные значения у могут отклоняться от y_x на величину случайной ошибки ε, дисперсия которой оценивается как остаточная дисперсия на одну степень свободы RMS. Поэтому ошибка предсказываемого индивидуального значения у должна включать не только стандартную ошибку m_y, но и случайную ошибку $\sqrt{RMS}$.

Средняя ошибка прогнозируемого индивидуального значения у составит:

$$m_y = \sqrt{RMS * (1 + \frac{1}{n} + \frac{(x_P - \overline{x})^2}{\sum (x_i - \overline{x})^2}}.$$

2.9. Методические указания построения моделей регрессии с помощью ППП Excel

Современный уровень развития и доступность вычислительной техники вызвали потребность к разработке различных пакетов прикладных программ для ПЭВМ, позволяющих упростить процесс вычислений. Статистические расчеты были одной из первых областей применения средств ЭВТ. В настоящее время существует достаточно большой набор программных продуктов, ориентированных на решение задач эконометрики, это E-VIEWS, STATA и др. Как правило, это дорогостоящие программные продукты, которые далеко не всегда доступны для использования. Поэтому можно воспользоваться доступным всем даже в домашних условиях пакетом Микрософт-офис и в частности программой EXCEL.

В этом приложении есть несколько функций, которыми можно воспользоваться для построения и последующего анализа эконометрических моделей. Рассмотрим наиболее простую, а именно, функцию *«ЛИНЕЙН»*.

Рассмотрим применение этой функции в виде последовательности действий специалиста.

Шаг 1. Работа начинается с того, что на рабочем листе создается набор данных, т.е. вводится содержимое выборки наблюдений за поведением объекта. Это таблица, содержащая значения эндогенной и экзогенных переменных.

Замечание. Существуют определенные требования: столбец значений эндогенной переменной должен быть крайним; между строками данных не должно быть пустых ячеек. Между столбцами со значениями экзогенных переменных также не должно быть пустых ячеек.

Шаг 2. На листе выделяется область размером в пять строк (всегда!) и числом столбцов равным числу столбцов с данными (эндогенная + экзогенные). При работе с данными рекомендуется не скупиться на поясняющие комментарии. Если комментарий занимает целый столбец, то в расчет количества столбцов при выделении он не входит. Рекомендуется выделять названную область непосредственно под данными. Это подсказывает необходимое количество столбцов. Выделенная область будет использована функцией «ЛИНЕЙН» для размещения результатов своей работы.

<u>*Шаг 3.*</u> На панели инструментов в разделе «функция» «f» вызывается функция «ЛИНЕЙН»: В окне «Категория» выбирается «Статистические», в окне «Функция» выбирается «ЛИНЕЙН». Необходимо заполнить аргументы функции:

- *Известные значения y* – диапазон, содержащий данные результативного признака;
- *Известные значения x* – диапазон, содержащий данные факторов независимого признака;
- *Константа* – логическое значение, которое указывает на наличие или на отсутствие свободного члена в уравнении. Если *константа* =1, то свободный член рассчитывается обычным образом. Если *константа* = 0, то свободный член =0.
- *Статистика* – логическое значение, которое указывает, выводить дополнительную информацию по регрессионному анализу или нет. Если *Статистика* =1, то дополнительная информация выводится, если *статистика* = 0, то выводятся только оценки параметров уравнения.

Шаг 4. Для завершения процесса ввода исходной информации необходимо набрать на клавиатуре комбинацию клавиш: <CTRL+SHIFT+ENTER>. После этого произойдет заполнение выделенной ранее области на листе. Дополнительная регрессионная статистика будет выводиться в порядке, указанном в таблице 2.9.1.

Таблица 2.9.1.

Регрессионная статистика функции «ЛИНЕЙН»

Значение коэффициента b	*Значение коэффициента a*
Стандартная ошибка m_b	*Стандартная ошибка* m_a
Коэффициент детерминации R^2_{xy}	*Стандартная ошибка* m_y, $x_P = \bar{x}$
F - статистика	*Число степеней свободы (n-m-1)*
ESS	*RSS*

Рассмотрим применение *инструмента анализа данных «Регрессия».*

<u>Шаг 1.</u> Необходимо проверить доступ к пакету анализа. В меню «Файл» последовательно выберите: «Параметры» - «Надстройки» - «Управление» - «Надстройки» - «Управление надстройки Excel» - «Перейти» - «Пакет анализа» - «ОК». После установления программного обеспечения в меню «Данные» появится «Пакет анализа». Необходимая функция – «Регрессия».

<u>Шаг 2.</u> После вызова «Регрессия» необходимо заполнить диалоговое окно ввода данных и параметров вывода.

- *Входной интервал y* – диапазон, содержащий данные результативного признака;

- *Входной интервал x* – диапазон, содержащий данные факторов независимого признака;
- *Метки* – флажок, который указывает, содержит ли первая строка названия столбцов или нет;
- *Константа* – ноль – флажок, указывающий на наличие или отсутствие свободного члена в уравнении;
- *Новый рабочий лист* – данные будут занесены на новый рабочий лист;
- *Если необходимо получить информацию и графики остатков.* Установите соответствующие флажки в диалоговом окне.

2.10. Пример построения модели парной регрессии с помощью пакета Excel и оценка ее значимости

Задание

Даны статистические данные, описывающие зависимость удельного веса бракованной продукции от удельного веса рабочих со специальной подготовкой на предприятиях.

№	Удельный вес рабочих со специальной подготовкой, %	Удельный вес бракованной продукции, %
1	15	18
2	25	12
3	35	10
4	45	8
5	55	6
6	65	5
7	70	3

Необходимо:

1. Построить модель парной регрессии.
2. Дать экономическую интерпретацию модели.
3. Оцените значимость коэффициентов линейной регрессии. Построить доверительные интервалы для коэффициентов линейной регрессии.
4. Оценить качество модели, рассчитав коэффициент детерминации. Применить F – критерий Фишера для проверки качества оценивания
5. Спрогнозировать значение y для какого – либо x_k, осуществив интервальный прогноз.
6. Постройте графики статистических и теоретических значений эндогенной переменной.

Решение:

По результатам применения встроенной функции «ЛИНЕЙН» и пакета анализа «Регрессия» (Приложение В Пример 1.), можно сделать следующие выводы:

1. Модель имеет вид: y=19,41-0,24x. Коэффициент корреляции R_{xy}=-0,97 свидетельствует о сильной линейной обратной связи между экзогенной и эндогенной переменными.
2. Удельный вес бракованной продукции=19,41-0,24*Удельный вес рабочих со специальной подготовкой.
   Экономический смысл коэффициента a.
   Удельный вес бракованной продукции =19,41% в случае полного отсутствия рабочих со специальной подготовкой, то есть при x=0.
   Экономический смысл коэффициента b.
   При увеличении рабочих со специальной подготовкой на 1%, количество брака на предприятии уменьшается на 0,24%.
3. Выдвигаются гипотезы:
   H_0: a=0;
   H_1: a≠0.
   Стандартная ошибка коэффициента a m_a=1,34; t-статистика коэффициента a: t_a =14,5.
   $t_{крит}$ =6,869 при уровне значимости 0,05% , то есть, $t_a > t_{крит}$.
   Следовательно, гипотеза H_0:a=0 отклоняется и принимается альтернативная гипотеза H_1:a≠0. Коэффициент a является статистически значимым.
   Выдвигаются гипотезы:
   H_0: b=0;
   H_1: b≠0.
   Стандартная ошибка коэффициента b m_b=0,028; t-статистика коэффициента b: t_b=8,57.
   $t_{крит}$ =6,869 при уровне значимости 0,05% , то есть, $t_b > t_{крит}$.
   Следовательно, гипотеза H_0:b=0 отклоняется и принимается альтернативная гипотеза H_1:b≠0. Коэффициент b является статистически значимым.
   95% -Доверительные интервалы:
   Все теоретические α, удовлетворяющие неравенству 15,97 <α<22,85, могут быть оценены a=19,41.
   Все теоретические β, удовлетворяющие неравенству -0,3<β<-0,167, могут быть оценены b=-0,24.
4. Коэффициент детерминации R_{xy}^2=0,94, то есть 94% разброса эндогенной переменной вокруг среднего значения объясняется экзогенной переменной, а остальные 6% приходится на объяснение случайной переменной.
   Выдвигаем нулевую и альтернативную гипотезы:
   Нулевая гипотеза: H_0: $R_{xy}^2 = 0$ или b=0;
   Альтернативная гипотеза: H_1: $R_{xy}^2 \neq 0$ или; b≠0.
   $F_{вычисляемое}$=73,5;
   $F_{табличное}$ =6,61 при 5% уровне значимости;
   $F_{табличное}$ =16,26 при 1% уровне значимости.

Следовательно, $F_{табличное} < F_{вычисляемое}$ при 1% уровне значимости, и гипотеза о незначимости коэффициента детерминации отклоняется. Можно заключить, что уравнение регрессии значимо, и фактор x существенно влияет на формирование величины y.

5. Построим точечный прогноз для x_k=30%.
   y=19,4-0,24*30=12,2%.
   Построим доверительный интервал для прогнозного значения:
   m_y=1,5; $t_{крит}$ =2,57 при уровне значимости 5%, тогда 95% доверительный интервал для прогнозного значения имеет вид:
   $12,2-1,5*2,57 \le y_P \le 1,5*2,57+12,2$ или $8,3 \le y_P \le 16,1$.

Вопросы для обсуждения:

1. Объясните, чем вызвано появление в модели регрессии стохастической переменной ε?
2. Почему перед построением модели парной линейной регрессии необходимо рассчитывать коэффициент корреляции?
3. В чем идея метода наименьших квадратов?
4. Сформулируйте теорему Гаусса –Маркова. Объясните смысл каждой предпосылки теоремы Гаусса –Маркова.
5. Объясните смысл проверки гипотез, относящихся к коэффициентам регрессии.
6. Зачем необходимо рассчитывать t-критерий Стьюдента?
7. В чем смысл доверительных интервалов?
8. По каким вычислениям можно судить о значимости модели в целом?
9. Объясните экономический смысл TSS, ESS, RSS.
10. Почему необходимо проверять статистическую значимость коэффициента детерминации?
11. Когда необходимо оценивать значимость модели и параметров регрессии, как при 1%, так и при 5% уровне значимости?.
12. Объясните смысл понятия «число степеней свободы».
13. Зачем необходимо оценивать интервалы прогноза по линейному уравнению регрессии?
14. Что происходит с интервалами прогноза по мере удаления от среднего значения выборки?

Тест:

1 Стохастические модели характеризуются:

a. наличием случайной составляющей;
b. отсутствием случайной составляющей;
c. случайная составляющая не играет роли в определении типа модели.

2 Причинно – следственная связь между двумя переменными в экономике называется:

a. регрессионной;
b. корреляционной.

3 Если коэффициент корреляции равен 0, 24, то наблюдается:

a. положительная сильная линейная связь;
b. отрицательная слабая линейная связь;
c. положительная слабая линейная связь.

4 Если коэффициент корреляции равен –0,977, то наблюдается:

a. положительная сильная линейная связь;
b. отрицательная слабая линейная связь;
c. положительная слабая линейная связь;
d. отрицательная сильная линейная связь.

5 Парная регрессия – регрессия между:

a. двумя переменными y и x;
b. переменной y и двумя факторами x_1 и x_2;
c. связь между двумя факторами x_1 и $x_{2.}$

6 Экспериментальный метод выбора вида математической функции $y_x=f(x)$ заключается в:

a. анализе поля корреляции;
b. анализе теории изучаемой взаимосвязи;
c. в сравнении величины остаточной дисперсии, рассчитанной при различных моделях.

7 Чем меньше величина остаточной дисперсии, тем:

a. в меньшей мере наблюдается влияние прочих, не учитываемых в уравнении регрессии факторов;
b. тем хуже уравнение регрессии подходит к исходным данным;
c. в большей мере наблюдается влияние прочих, не учитываемых в уравнении регрессии факторов.

8 Метод наименьших квадратов:

a. метод оценивания параметров линейной регрессии, максимизирующий сумму квадратов отклонений наблюдений зависимой переменной от искомой линейной функции;
b. метод оценивания параметров линейной регрессии, минимизирующий сумму квадратов отклонений наблюдений зависимой переменной от искомой линейной функции.

9 Коэффициенты регрессии, найденные методом наименьших квадратов

являются:

a. истинными коэффициентами регрессии;
b. оценками истинных коэффициентов регрессии;
c. не являются ни теми ни другими.

10 Если модель парной регрессии описывает зависимость спроса товара от цены, то коэффициент а:

a. имеет экономический смысл;
b. не имеет экономического смысла;
c. определить невозможно.

11 Если a<0, то:

a. относительное изменение результата происходит медленнее, чем изменение фактора;
b. относительное изменение результата происходит быстрее, чем изменение фактора;
c. относительное изменение результата происходит с той же скоростью что и изменение фактора.

12 В модели y=15+6x коэффициент 6 показывает:

a. среднее изменение фактора y при изменении x на единицу;
b. общее изменение фактора y при изменении x на единицу;
c. коэффициент 6 не имеет экономического смысла.

13 В эксперименте Монте – Карло:

a. неизвестны истинные значения коэффициентов регрессии;
b. известны заранее истинные коэффициенты регрессии;
c. истинные коэффициенты регрессии находятся методом наименьших квадратов.

14 Одним из условий Гаусса – Маркова является условие:

a. зависимости случайных компонент для каждого наблюдения;
b. независимость случайных компонент для каждого наблюдения;
c. неравенство математического ожидания случайной компоненты нулю.

15 В соответствии теореме Гаусса – Маркова, коэффициенты регрессии, построенные МНК, являются:

a. смещенными оценками;
b. несмещенными оценками;
c. не являются оценками.

16 Какое из утверждений не является условием Гаусса-Маркова?

a. Математическое ожидание случайной переменной в любом наблюдении должно быть равно нулю.

b. Теоретическая дисперсия случайной переменной различна для всех наблюдений.
c. Отсутствие систематической связи между значениями случайной переменной в любых двух наблюдениях.
d. Случайная переменная должна быть распределена независимо от объясняющих переменных.

17 Чем больше информации у исследователя, тем:
a. более точными, скорее всего, будут оценки;
b. менее точными, скорее всего, будут оценки;
c. качество оценок не зависит от объема информации.

18 Чем больше дисперсия случайного члена в зависимости, тем:
a. более точными, скорее всего, будут оценки при прочих равных условиях;
b. менее точными, скорее всего, будут оценки при прочих равных условиях;
c. качество оценок не зависит от дисперсии случайного члена.

19 Чем меньше дисперсия x, тем:
a. меньше будет относительное влияние фактора случайности при определении отклонений в y;
b. тем менее вероятно, что регрессионный анализ даст неточные результаты.
c. больше будет относительное влияние фактора случайности при определении отклонений в y.

20 Критическая область:
a. совокупность значений критерия, при которых нулевую гипотезу отклоняют;
b. совокупность значений критерия, при которых нулевую гипотезу не отклоняют;
c. невозможно сделать однозначный вывод.

21 Если наблюдаемое значение критерия К (вычисленное по выборке) принадлежит критической области, то:
a. нулевую гипотезу не отклоняют;
b. нулевую гипотезу отклоняют;
c. невозможно сделать однозначный вывод.

22 Оценивание каждого параметра в выборке поглощает:
a. три степени свободы в выборке;
b. две степени свободы;
c. одну степень свободы.

23 Ошибка первого рода имеет место когда:

a. отвергается ложная нулевая гипотеза;
b. принимается ложная нулевая гипотеза;
c. отвергается истинная нулевая гипотеза.

24 Условие того, что оценка регрессии не должна приводить к отказу от нулевой гипотезы $H_0 : \beta = \beta_0$ имеет вид**:**

a. $-t_{крит} \le \frac{b-\beta_0}{m_b} \le t_{крит}$;
b. $-t_{крит} \ge \frac{b-\beta_0}{m_b} \ge t_{крит}$;
c. $\frac{b-\beta_0}{m_b} = t_{крит}$.

25 Если отклоняется гипотеза при 1% уровне значимости, то:

a. она не отклоняется при уровне значимости в 5%;
b. автоматически она отклоняется и при уровне значимости в 5%;
c. невозможно ответить однозначно.

26 Если не отвергается гипотеза при уровне значимости в 5%, то:

a. она не отвергается и при 1% уровне значимости;
b. она отвергается и при 1% уровне значимости;
d. невозможно ответить однозначно.

27 Необходимо представить оба результата:

a. если гипотеза не отвергается на 5%, но принимается на 1% уровне значимости;
b. если гипотеза отвергается на 5%, и принимается на 1% уровне значимости;
c. если гипотеза отвергается на 5%, но не на 1% уровне значимости.

28 Интервал в 99% будет:

a. шире интервала в 95%;
b. уже интервала в 95%;
c. совпадать с интервалом в 95%.

29 Доверительные интервалы в регрессионном анализе строятся для:

a. истинных значений параметров модели;
b. оценок истинных значений параметров модели;
c. для статистических данных y и x.

30 Факторная сумма квадратов отклонения вычисляется по формуле:

a. $\sum_i (y_i - \bar{y})^2 = \sum_i y^2 - n * \bar{y}^2$;

b. $\sum_i (y_x - \bar{y})^2 = b^2 * (\sum_i x^2 - n * \bar{x}^2)$;

c. $\sum_{i=1}^{n} (y_i - y_x)^2$.

31 Если коэффициент детерминации равен 0,65, то модель описывает:
a. 65% вариации признака y, объясненную влиянием фактора x;
b. 35% вариации признака y объясненную влиянием фактора x;
c. по значению коэффициента детерминации невозможно ответить на данный вопрос.

32 F – критерий рассчитывается по формуле:
a. $F_{вычисляемое} = \frac{RMS}{EMS}$
b. $F_{вычисляемое} = \frac{EMS}{RMS}$;
c. $F_{вычисляемое} = \frac{EMS}{TMS}$.

33 Табличное значение F – критерия – это:
a. минимальная величина отношения дисперсий, которая может иметь место при наличии нулевой гипотезы;
b. максимальная величина отношения дисперсий, которая может иметь место при наличии нулевой гипотезы.

34 Если $F_{\textbf{вычисляемое}} > F_{табличное,}$ то уравнение регрессии:
a. не значимо;
b. фактор x не влияет на формирование величины y;
c. H_0 о незначимости отклоняется.

35 Интервальная оценка прогнозного значения принимает наименьшее значение:
a. в средней точке выборки;
b. в дальних точках от среднего значения;
c. в ближних точках к среднему значению.

<u>Задания для самостоятельной работы:</u>

Задание

Даны статистические данные, описывающие зависимость y от x.

Необходимо:

1. Построить модель парной регрессии.
2. Дать экономическую интерпретацию модели.
3. Оцените значимость коэффициентов линейной регрессии. Построить доверительные интервалы для коэффициентов линейной регрессии.
4. Оценить качество модели, рассчитав коэффициент детерминации. Применить F – критерий Фишера для проверки качества оценивания
5. Спрогнозировать значение y для какого – либо x_k, осуществив интервальный прогноз.
6. Постройте графики статистических и теоретических значений эндогенной переменной.

№1.

Статистические данные, описывающие зависимость уровня рентабельности на предприятии от скорости товарооборота.

№	1	2	3	4	5	6	7	8
Число оборотов	5,49	4,68	4,67	4,54	4,56	6,02	5,72	5,43
Уровень рентабельности, %	7,8	3,8	2,1	5,1	9,5	10,5	8,3	9,8

№2.

Статистические данным, описывающим зависимость индекса Лернера от рыночной доли фирмы.

	1	2	3	4	5	6	7
Рыночная доля фирмы, s_i	0,064	0,223	0,273	0,182	0,073	0,05	0,04
Индекс Лернера L	0,1	0,2	0,35	0,15	0,11	0,045	0,038

№3.

Статистические данные, описывающие зависимость уровня рентабельности на предприятии от удельного веса продовольственных товаров в товарообороте.

№	1	2	3	4	5	6	7
Удельный вес продовольственных товаров в товарообороте, %.	74,2	73,5	77	84,3	67,3	70,1	83,1
Уровень рентабельности, %	3,62	3,8	2,77	2,12	4,33	4,01	2,01

№4.

Статистические данные, описывающие зависимость объема спроса на товар от его цены.

№	1	2	3	4	5	6	7	8	9	10
Цена товара, руб.	99	82	77	69	52	44	31	29	28	27,5
Спрос на товар, шт.	100	115	210	270	323	478	544	564	570	574

№5.

В таблице приведены значения выручки от экспорта 1 тонны синтетического каучука за 10 кварталов и цены его на внутреннем рынке.

Период	Выручка от экспорта 1 тонны, долл.	Цена внутреннего рынка, долл. за 1 тонну
1–й квартал	1090	1090
2-й квартал	1190	1550
3-й квартал	1320	2180
4-й квартал	1430	2370
5-й квартал	1470	2440
6-й квартал	1510	2560
7-й квартал	1535	2570
8-й квартал	1570	2700
9-й квартал	1600	2759
10-й квартал	1615	2820

№6.

В таблице представлены средние расходы на потребление Y и агрегированный располагаемый доход X в некоторой национальной экономике.

Год	Y_t, $	X_t, $
1986	152	170
1987	159	179
1988	162	187
1989	165	189
1990	170	193
1991	172	199
1992	177	200
1993	179	207
1994	184	215
1995	186	216
1996	190	220
1997	191	225

№7.

В таблице приведены данные о реальной стоимости нескольких конструкторских проектов (млн. $) конструкторской фирмы и ранее выполненные оценки данных проектов. Какова будет действительная стоимость проекта, оценочная стоимость которого 35, 200 млн. $.

№	1	2	3	4	5	6	7	8	9	10
Действительная стоимость проекта	0,9	7,2	14,5	30,2	38,1	15,3	14,8	51,2	34,1	2
Оценка стоимости проекта	0,5	6,1	11,2	28,1	30,1	21,1	8,6	40,6	37,8	1

№8.

Статистические данные о годовых расходах на техническое обслуживание автобусов и возраста автобусов. Спрогнозируйте расходы на содержание автобуса возрастом 10 лет.

№ автобуса	1	2	3	4	5	6	7	8	9
Расходы на содержание ($)	859	682	471	708	1094	224	320	651	1049
Возраст (годы)	8	5	3	9	11	2	1	8	12

№9.

Статистические данные количества проданных книг в сети книжных магазинов и объем демонстрационного пространства.

№	1	2	3	4	5	6	7	8	9	10
Количество проданных книг, шт	275	142	168	197	215	188	241	295	125	266
Объем демонстрационного пространства (в кв.м)	68	33	41	42	48	39	49	77	31	59

№10.

Статистические данные количества заказов на товары по почте и количества распространенных каталогов.

Город	1	2	3	4	5	6	7	8	9	10	11	12
Количество заказов на товары в каждом городе	24	16	23	15	32	25	18	18	35	34	15	32
Количество распространенных каталогов	6	2	5	1	10	7	15	3	11	13	2	12

№11.

Статистические данные количества выданных инвестиционных кредитов в банке и банковской учетной ставки.

	1	2	3	4	5	6	7	8	9
Количество выданных инвестиционных кредитов, шт	786	494	289	892	343	888	509	987	187
Банковская учетная ставка	10,2	12,6	13,5	9,7	10,8	9,5	10,9	9,2	14,2

№12.

Статистические данные цены товара компании АВС и цены товара конкурента.

	1	2	3	4	5	6	7	8	9	10
Цена товара компании АВС	99	104	99,5	99,9	98,9	101	100	98,2	93,8	99,4

Цена товара конкурента	100	105	99,5	95,9	98,8	101,5	101,2	99,1	94,8	100

№13.

Для 14 однотипных предприятий имеются данные за год о производительности труда и уровне механизации работ, %.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Производительность труда, %	20	24	28	30	31	33	34	37	38	40	41	43	45	48
Уровень механизации, %	32	30	36	40	41	47	56	54	60	55	61	67	69	76

№14.

По территориям некоторых регионов страны известны данные за год по среднедневной заработной плате y и среднедушевому прожиточному минимуму x.

№	1	2	3	4	5	6	7	8	9	10
Среднедневная заработная плата, руб.	162	151	190	178	161	175	144	191	160	161
Среднедушевой прожиточный минимум, руб.	95	107	125	111	89	97	95	131	92	102

№15.

Имеются статистические данные за 10 лет по прибылям в % двух компаний.

№	1	2	3	4	5	6	7	8	9	10
X	19,2	15,8	12,5	10,3	5,7	-5,8	-3,5	5,2	7,3	6,7
Y	20,1	18	10,3	12,5	6	-6,8	-2,8	3	8,5	8

№16.

Для прогноза возможного объема экспорта на основе ВНП предложено использовать линейную регрессионную модель. При этом используются данные за 1999-2008 гг.

Годы	99	00	01	02	03	04	05	06	07	08
ВНП	1000	1090	1150	1230	1300	1360	1400	1470	1500	1580
Экспорт	190	220	240	240	260	250	280	290	310	350

№17.

Проводится анализ взаимосвязи количества населения и количества практикующих врачей

Годы	01	02	03	04	05	06	07	08	09	10
Население	10	10,3	10,4	10,55	10,6	10,7	10,75	10,9	10,9	11
Врачи	12,1	12,6	13	13,8	14,9	16	18	20	21	22

ТЕМА 3. МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕСИИ

Ключевые слова:

Спецификация модели. Результативный признак, признак-факторы и стохастическая переменная в модели. Параметры регрессии. Интеркорреляция факторов модели. Мультиколлинеарность факторов. Определитель матрицы межфакторной корреляции. Индекс множественной корреляции. Метод наименьших квадратов. Стандартизованное уравнение. Стандартизованные коэффициенты. Стандартная ошибка коэффициента множественной регрессии. F- критерий Фишера модели множественной регрессии. Скорректированные индексы множественной корреляции и детерминации. Фиктивные переменные. Частные уравнения регрессии Частные коэффициенты эластичности Коэффициент частной корреляции. Частный F- критерий Фишера.

Основные теоретические аспекты темы:

3.1. Спецификация модели множественной регрессии: понятие и способы задания функций

На любой экономический показатель чаще всего оказывает влияние не один, а несколько факторов. Например, спрос на некоторое благо определяется не только ценой данного блага, но и ценами на замещающие и дополняющие блага, доходом потребителей и многими другими факторами. Поэтому вместо парной регрессии строится модель множественной регрессии. То есть, множественный регрессионный анализ является развитием парного регрессионного анализа.

Множественная регрессия – регрессия между переменными y и $x_1, x_2, \ldots, x_m$ то есть модель вида: $\mathbf{y}=f(x_1, x_2, \ldots, x_m)+\varepsilon$,

где: y - зависимая переменная, фактическое значение результативного признака,

$x_1, x_2, \ldots, x_m$ – независимые, объясняющие переменные, признак-факторы,

ε - возмущение, случайная, стохастическая переменная.

Основная цель множественной регрессии – построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное воздействие их на моделируемый показатель.

Причинами существования ε являются:

1. Невключение объясняющих переменных.
2. Агрегирование переменных.
3. Неправильное описание структуры модели.
4. Неправильная функциональная спецификация.

5. Ошибки измерения

Основными типами функций, используемых при количественной оценке связей при построении модели множественной регрессии являются:

Линейная функция: $y=a_0+a_1 x_1+a_2 x_2+, \dots, +a_m x_m$; или
$y=a+b_1 x_1+b_2 x_2+, \dots, +b_m x_m$

Нелинейные функции: $y=a\, x_1^{b1} x_2^{b2}, \dots, x_m^{bm}$ – степенная функция;
$y= 1/(a_0+a_1 x_1+a_2 x_2+, \dots, +a_m x_m)$ – гипербола;
$y= e^{a0+a1\ x1\ +a2\ x2\ +, \dots, +am\ xm}$ – экспонента.

Индекс множественной корреляции – характеризует тесноту связи рассматриваемого набора факторов с исследуемым результативным признаком или оценивает тесноту совместного влияния факторов на результат. Чем ближе R к 1, тем теснее связь результативного признака со всем набором исследуемых факторов.

$$R=\sqrt{1-\frac{\sum_{i=1}^{n}(y_i-y_x)^2}{\sum_{i}(y_i-\overline{y})^2}}=\sqrt{1-\frac{RSS}{TSS}}=\sqrt{\frac{ESS}{TSS}};\ R\in[0;1].$$

3.2. Отбор факторов при построении модели множественной регрессии. Мультиколлинеарность и методы ее преодоления

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. Быть количественно измеримы. Если необходимо включить в модель качественный фактор, не имеющий количественного измерения, то нужно придать ему количественную определенность.
2. Не должны быть коррелированы между собой и тем более находиться в точной функциональной связи.

Включение в модель факторов с высокой интеркорреляцией, когда $R_{yx1}<R_{x1x2}$ в уравнении с двумя факторами, может привести к нежелательным последствиям – неустойчивости и ненадежности оценок коэффициентов. 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 зависимой переменной. Если строится модель с набором p факторов, то для нее рассчитывается показатель детерминации R^2, который фиксирует долю объясненной вариации результативного. признака за счет рассматриваемых в регрессии p факторов. Влияние других не учтенных в модели факторов оценивается как (1 — R^2) с соответствующей остаточной дисперсией.

При дополнительном включении в регрессию (p+1) фактора коэффициент детерминации должен возрастать, а остаточная дисперсия уменьшаться. То есть $R^2_{p+1} \geq R^2_p$; $RSS_{p+1} \leq RSS_p$.

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 t-критерию Стьюдента.

Несмотря на то, что теоретически регрессионная модель позволяет учесть любое число факторов, практически в этом нет необходимости. Отбор факторов проводится 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 проводится в две стадии:

- на первой отбираются факторы исходя из сути проблемы;
- на второй — на основе матрицы показателей корреляции и определения t-статистики для параметров регрессии.

Две переменные явно коллинеарны, т. е. находятся между собой в линейной зависимости, если $r_{xixj} \geq 0,7$.

Коэффициенты интеркорреляции позволяют исключить из модели дублирующие факторы. 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.

Наибольшие трудности в использовании аппарата множественной регрессии возникают при наличии мультиколлинеарности факторов.

Мультиколлинеарность факторов – когда, более чем два фактора связаны между собой линейной зависимостью, то есть имеет место совокупное воздействие факторов друг на друга и некоторые факторы действуют в унисон.

Следствие мультиколлинеарности факторов:

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.

Но высокая коррелированность не всегда ведет к несостоятельным оценкам. Если все остальные факторы, определяющие дисперсию коэффициентов регрессии, благоприятствуют, т.е. число наблюдений и выборочные дисперсии объясняющих переменных велики и дисперсия случайной переменной мала, можно получить, тем не менее, хорошие оценки.

Мультиколлинеарность должна вызываться сочетанием высокой коррелированности и других неблагоприятных условий. Это вопрос степени выраженности явления, а не его сущности. Любая регрессия будет «страдать» от нее в определенной степени, если только все независимые переменные не будут абсолютно некоррелированными.

Число включаемых факторов должно быть в 6-7 раз меньше объема совокупности, по которой строится регрессия. Если это нарушается, то степень свободы df RSS очень мало, и параметры будут незначимыми, а F-критерий меньше табличного.

Оценка мультиколлинеарности факторов:

1. Расчет определителя матрицы межфакторной корреляции (на примере модели $y=a_0+a_1 x_1+a_2 x_2+a_3 x_3+\varepsilon$)

$$Det|R| = \begin{vmatrix} r_{x1x1} & r_{x2x1} & r_{x3x1} \\ r_{x1x2} & r_{x2x2} & r_{x3x2} \\ r_{x1x3} & r_{x2x3} & r_{x3x3} \end{vmatrix}.$$

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.

$Det|R| = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0$ - полная мультиколлинеарность факторв;

$Det|R| = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1$ - полное отсутствие мультиколлинеарности.

Оценка значимости мультиколлинеарности факторов может быть проведена методом испытания гипотезы о независимости переменных $H_0 : Det|R| = 1$.

Доказано, что величина $\left[n - 1 - \frac{1}{6}(2m+5)\lg DetR\right]$ имеет приближенное распределение χ^2 с $df = \frac{1}{2}m(m-1)$ степенями свободы. Если фактическое, рассчитываемое значение χ^2 превосходит табличное (критическое) [Приложение А, Таблица А 4.], то гипотеза H_0 отклоняется. Это означает, что $Det|R| \neq 1$, недиагональные ненулевые коэффициенты корреляции указывают на коллинеарность факторов. Мультиколлинеарность считается доказанной.

2. Расчет коэффициентов множественной детерминации. Можно найти переменные, ответственные за мультиколлинеарность факторов. Для этого в качестве зависимой переменной рассматривается каждый из факторов.

Чем ближе значение коэффициента множественной детерминации к единице, тем сильнее проявляется мультиколлинеарность факторов. Сравнивая между собой коэффициенты множественной детерминации факторов $R^2_{x1|x2x3...xm\rangle}$, $R^2_{x2|x1x3...xm\rangle}$ и т.д., можно определить переменные, ответственные за мультиколлинеарность, оставляя в уравнении факторы с минимальной величиной коэффициента множественной детерминации.

Имеется ряд подходов преодоления сильной межфакторной корреляции:

Прямые: улучшение условий Гаусса – Маркова:

- Уменьшение дисперсии случайного параметра.
- Увеличение числа наблюдений.
- Увеличение дисперсии объясняющих переменных.
- Выбор переменных, наименее связанных между собой (начальный этап моделирования).

Косвенные

1. Исключение из модели одного или несколько факторов.
2. Преобразование факторов, при котором уменьшается корреляция между ними.
    - Если коррелированные переменные связаны между собой концептуально, то может быть рационально объединить их в совокупный индекс (взвешенное среднее).
    - При построении модели на основе рядов динамики переходят от первоначальных данных к первым разностям уровней ряда $\Delta y = y_t - y_{t-1}$, чтобы исключить влияние тенденции.
    - Переход к уравнениям приведенной формы: $y = a + b_1x_1 + b_2x_2 + \varepsilon$; $x_1 = c + dx_2$; $y = a + b_1(c + dx_2) + b_2x_2 + \varepsilon$.
    - Переход к совмещенным уравнениям регрессии, т.е. к уравнениям, которые отражают не только влияние факторов, но и их взаимодействие:
      $y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_1b_2x_1x_2 + b_1b_3x_1x_3 + b_2b_3x_2x_3 + \varepsilon$

3.3. Параметризация модели множественной линейной регрессии. Метод наименьших квадратов для модели множественной регрессии

Дано уравнение множественной регрессии для генеральной совокупности: $y = \alpha + \beta_1x + \beta_2x_2 + ... + \beta_mx_m + u$;

где: α, $\beta_1 \dots \beta_m$– истинные коэффициенты модели;

u– стохастическая переменная модели.

Необходимо по выборке, состоящей из m факторов $x_1, \dots x_m$, каждый из которых имеет n статистических значений и n статистических значений

зависимой переменной y: y_1, … y_n построить уравнение множественной регрессии $y=a+b_1 x_1 +b_2 x_2 + \ldots +b_m x_m + \varepsilon$,
где: a, b_1, … b_m - оценки α, $\beta_1 \ldots \beta_m$, соответствующие генеральной совокупности,

ε - оценка u.

Метод наименьших квадратов для уравнения множественной регрессии - метод оценивания параметров множественной линейной регрессии, минимизирующий сумму квадратов отклонений наблюдений зависимой переменной от искомой линейной функции.

$$RSS = \sum_{i=1}^{n} e_i^2 = \sum_i (y_i - f(x_{ij}))^2 = \sum_{i=1}^{n} (y_i - (a + b_1 x_i + b_2 x_2 + \ldots + b_m x_m))^2 \to \min$$

где: y_i – статистические значения зависимой переменной;

$f(x_{ij})$ - теоретические значения зависимой переменной, рассчитанные с помощью уравнения множественной регрессии.

Исследуя RSS как функцию m+1 переменной (a и b_1, …b_m) на экстремум, строим систему уравнений:

$$\begin{cases} \dfrac{dRSS}{da} \\ \dfrac{dRSS}{db_1} \\ \dfrac{dRSS}{db_2} \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{dRSS}{db_m} \end{cases} = 0;$$

Y=XA+E; где: $X = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{1m} \\ 1 & x_{21} & x_{22} & x_{2m} \\ . & . & . & . \\ 1 & x_{n1} & x_{n2} & x_{nm} \end{pmatrix}$; $Y = \begin{pmatrix} y_1 \\ y_2 \\ . \\ y_n \end{pmatrix}$; $E = \begin{pmatrix} E_1 \\ E_2 \\ . \\ E_n \end{pmatrix}$.

Решением данной системы уравнений будут значения коэффициентов регрессии a и b_1, …b_m, вычисляемы следующим образом:

$$A = (a, b_1, \ldots, b_m) = (X^T X)^{-1} X^T Y.$$

3.4. Интерпретация уравнения множественной линейной регрессии: экономический смысл параметров регрессии Стандартизованное уравнение множественной регрессии

В линейной множественной регрессии $y=a+b_1x_1 +b_2x_2 + \ldots +b_mx_m$ параметры при x называются *коэффициентами «чистой» регрессии* и характеризуют среднее изменение результата с изменением соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне.

Свободный член уравнения множественной регрессии (параметр a) вбирает в себя информацию о прочих не учитываемых в модели факторах.

Его величина экономической интерпретации не имеет. Формально его значение предполагает то значение y, когда все x=0, что практически не бывает.

Стандартизованное уравнение множественной регрессии имеет вид**:**

$$t_y = \eta_1 x_1 + \eta_2 x_2 + ... + \eta_m x_m \text{; где: } \quad \eta_j = b_j * \frac{s_{xj}}{s_{yj}}.$$

Стандартизованные коэффициенты множественной регрессии η_j показывают, что с изменением фактора x_j на одно выборочное стандартное отклонение s_j, при неизменном среднем уровне других факторов, y изменяется на η_j выборочных стандартных отклонений.

В силу того, что все переменные заданы как центрированные и нормированные, стандартизованные коэффициенты регрессии $\eta_1 ... \eta_m$ сравнимы 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.

3.5. Свойства коэффициентов множественной регрессии. Оценка значимости коэффициентов множественной регрессии: проверка гипотез, относящихся к коэффициентам регрессии

Как и в случае парного регрессионного анализа, коэффициенты множественной регрессии должны рассматриваться как случайные переменные специального вида, случайные компоненты которых обусловлены наличием в модели случайного члена. Поэтому, коэффициенты множественной регрессии определяются значениями независимых переменных и случайным параметром, и их свойства существенно зависят от последнего. Продолжаем считать, что выполняются условия Гаусса – Маркова, а именно:

1. Математическое ожидание случайной переменной в любом наблюдении должно быть равно нулю.
2. Теоретическая дисперсия случайной переменной постоянна для всех наблюдений.
3. Отсутствие систематической связи между значениями случайной переменной в любых двух наблюдениях**.**
4. Случайная переменная должна быть распределена независимо от объясняющих переменных.

Но существует еще два практических требования:

1. Необходимо иметь достаточное количество данных.
2. Отсутствие строгой линейной зависимости между признак – факторами.

В теореме Гаусса – Маркова для множественной регрессии доказывается, что как и для парной регрессии, обычный МНК дает наиболее эффективные линейные оценки в том смысле, что на основе той же самой выборочной информации невозможно найти другие несмещенные оценки с меньшими дисперсиями при выполнении условий Гаусса – Маркова.

Коэффициенты регрессии будут более точными:

- чем больше число наблюдений в выборке;
- чем больше дисперсия объясняющих переменных;
- чем меньше теоретическая дисперсия случайного параметра;
- чем меньше связаны между собой объясняющие переменные.

При проверке статистических гипотез, связанных с анализом статистической значимости коэффициентов регрессии, как и в парной регрессии, используют дробь Стьюдента:

$$t_a = \frac{a - \alpha_0}{m_a}; \quad t_b = \frac{b_j - \beta_{0j}}{m_{bj}};$$

Стандартные ошибки коэффициентов множественной регрессии рассчитываются по формуле**:**

$$m_{bj} = \frac{s_{yj}\sqrt{1 - R^2_{yx1...xm}}}{s_{xj}\sqrt{1 - R^2_{xjx1...xm}}} * \frac{1}{\sqrt{n - m - 1}};$$

где: s_y –среднее кавдратическое отклонение для признака y;

s_x – среднее квадратическое отклонение для признака x_j;

$R^2_{yx1....xm}$ – коэффициент детерминации для уравнения множественной регрессии;

$R^2_{xj\ x1....xm}$ – коэффициент детерминации для зависимости фактора x_j со всеми другими факторами уравнения множественной регрессии;

n-m-1 – число степеней свободы для остаточной суммы квадратов отклонений

Аналогично парной регрессии выдвигается нулевая и альтернативная ей гипотезы:

Нулевая гипотеза: H_0: $\beta_j = \beta_0$;

Альтернативная гипотеза**:** H_1: $\beta_j \neq \beta_0$.

Заметим, что во множественной регрессии при проверке значимости коэффициентов, как и в парной регрессии, $\beta_0 = 0$.

Рассчитываются фактические значения t-критерия Стьюдента для каждого коэффициента регрессии по соответствующим формулам.

Критические значения $t_{крит}$ определяются аналогично $t_{крит}$ модели парной регрессии величиной имеющее t распределение [Приложение А, Таблица А 2.].

Условием того, что оценка регрессии не должна приводить к отказу от нулевой гипотезы H_0: $\beta_j = \beta_{0j}$ имеет вид:

$$-t_{крит} \leq \frac{b_j - \beta_{0j}}{m_{bj}} \leq t_{крит}.$$

Доверительные интервалы для каждого коэффициента множественной регрессии строятся с помощью неравенства:

$$b_j - t^*m_{bj} \le \beta_j \le b_j + t^* m_{bj}.$$

Данное неравенство отражает то, что любое гипотетическое значение для β_j, которое удовлетворяет данному соотношению, будет автоматически совместимо с оценкой b_j.

3.6. Качество оценки: коэффициент детерминации. F – критерий Фишера для проверки качества оценивания модели множественной регрессии

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

Как и в парном регрессионном анализе, *коэффициент детерминации* R^2 определяет долю дисперсии y, объясненную регрессией и эквивалентно определяется как квадрат коэффициента корреляции

$$\mathbf{R^2_{xy}} = \frac{ESS}{TSS}.$$

Для оценки статистической значимости коэффициента детерминации выдвигается нулевая и альтернативная гипотезы. Нулевая гипотеза заключается в том, что коэффициент детерминации незначим и модель не обладает объясняющей способностью.

Нулевая гипотеза: H_0: $\beta_j = 0$;

Альтернативная гипотеза: H_1: $\beta_j \neq 0$.

Альтернативная гипотеза заключается в том, что по крайней мере один из коэффициентов $\beta_j \neq 0$.

F-тест проверяет совместную объясняющую способность переменных множественной регрессии.

F статистика записывается как:

$$F_{вычисляемое} = \frac{EMS}{RMS} = \frac{R^2/m}{(1-R^2)/(n-1-m)};$$

$$TMS = \frac{TSS}{n-1};\quad EMS = \frac{ESS}{m};\quad RMS = \frac{RSS}{n-1-m};\ \text{где:}$$

где: n – число наблюдений;

m – количество факторов в уравнении регрессии;

$v_1 = df_{ESS} = m$- степень свободы факторной суммы квадратов отклонений

$v_2 = d_{fRSS} = n-1-m$ - степень свободы остаточной суммы квадратов отклонений.

Тест выполняется путем сравнения этой величины с критическим значением F.

Критические значения F – отношений при разных уровнях существенности нулевой гипотезы и различном числе степеней свободы

$v_1=df_{ESS}$; $v_2=df_{RSS}$ определяются аналогично критическим значениям F – отношений модели парной регрессии [Приложение А, Таблица А 3.].

.

Скорректированный индекс множественной корреляции.

В модели множественной регрессии в расчете показателя множественной корреляции используется остаточная дисперсия, которая имеет систематическую ошибку в сторону преуменьшения, тем более значительную, чем больше параметров определяется в уравнении регрессии при заданном объеме n. Если число параметров приближается к объему наблюдений, то остаточная дисперсия будет близка к нулю и индекс корреляции приблизится к единице даже при слабой связи факторов с результатом.

Для того, чтобы не допустить возможного преувеличения тесноты связи, применяется скорректированный индекс (коэффициент) множественной корреляции. Скорректированный индекс множественной корреляции содержит поправку на число степеней свободы, а именно остаточная сумма квадратов RSS делится на число степеней свободы остаточной вариации (n-m-1) , а общая сумма квадратов отклонений – на число степеней свободы в целом по совокупности (n-1):

$$\overline{R} = \sqrt{1 - \frac{\sum_{i=1}^{n} (y_i - y_x)^2 / (n-m-1)}{\sum_{i} (y_i - \overline{y^2}) / (n-2)}};$$

Тогда скорректированный *индекс множественной детерминации* имеет вид:

$$\overline{R}^2 = 1 - (1 - R^2) * \frac{(n-1)}{n-m-1}.$$

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

3.7.Фиктивные переменные в модели множественной регрессии

Часто случается так, что некоторые из факторов, которые необходимо ввести в регрессионную модель, являются качественными по своей природе, и, следовательно, не измеряются в числовой шкале. Это могут быть разного рода атрибутивные признаки, такие, например, как профессия, пол, образование, климатические условия, принадлежность к определенному региону.

Качественные признаки могут приводить к неоднородности исследуемой совокупности, что может быть уточнено при моделировании двумя путями:

1. Регрессия строится для каждой качественно отличной группы единиц совокупности, то есть для каждой группы в отдельности, с последующим выяснением, различаются ли полученные коэффициенты.

2. Оценивание единой регрессии с использованием вей совокупности наблюдений и измерением степени влияния качественного фактора посредством так называемой фиктивной переменной.

Фиктивные переменные (структурные переменные) – переменные, отражающие атрибутивные признаки (пол, профессия, образование, климатические условия, принадлежность к региону и др.)

Второй подход имеет два важных преимущества:

- имеется простой способ проверки, является воздействие качественного фактора значимым;
- при условии выполнения определенных предположений регрессионные оценки оказываются более эффективными.

Чтобы ввести такие переменные в регрессионную модель, им должны быть присвоены цифровые метки, то есть качественные переменные необходимо преобразовать в количественные.

Фиктивная переменная z рассматривается как дихотомическая переменная, принимающая всего два значения: 1 и 0. Так, если качественный фактор имеет два состояния z_1 и z_2 (например женский и мужской пол), то фиктивная переменная будет принимать значения:

$$z = \begin{cases} 1, z = z_1 \\ 0, z = z_2 \end{cases}.$$

Если же число градаций качественного признака –фактора превышает два, то в модель вводится несколько фиктивных переменных, число которых должно быть меньше числа качественных градаций на 1.

3.8. Частная корреляция

Ранжирование факторов, участвующих в множественной линейной регрессии, может быть проведено через стандартизированные коэффициенты регрессии. Эту же цель можно достичь с помощью частных коэффициентов корреляции для линейных связей. При нелинейной взаимосвязи исследуемых признаков эту функцию выполняют частные индексы детерминации. Кроме того, частные показатели корреляции широко используются при отборе факторов: целесообразность включения того или иного фактора в модель доказывается величиной показателя частной корреляции.

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.

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. Например, $R_{yx1\ x2}$ коэффициент частной корреляции первого порядка. Если рассматривается регрессия с числом факторов m, то возможны частные коэффициенты корреляции не только первого, но и второго, третьего,..., m-1, то есть влияние

фактора x_1 можно оценить при разных уровнях независимости действия других факторов:
$R_{yx1\,x2}$- при постоянном действии фактора x_2;
$R_{yx1\,x2x3}$- при постоянном действии фактора x_2 и x_3;
$R_{yx1\,x2xm}$ - при неизменном действии всех факторов, включенных в уравнение регрессии.

Сопоставление коэффициентов частной корреляции разных порядков по мере увеличения числа включаемых факторов показывает процесс «очищения» связи результативного признака с исследуемым фактором.

Коэффициент частной корреляции – измеряет влияние на результат фактора x_i при неизменном уровне других факторов**:**

$$r_{yxix1x2...x(i-1)x(i+1)...xm} = \sqrt{1 - \frac{1 - R^2_{yx1x2...xi...xm}}{1 - R_{yx1x2...x(i-1)x(i+1)...xm}}}\ ;$$

где: $R^2_{yx1x2\ldots xi\ldots xm}$ – множественный коэффициент детерминации всего комплекса m факторов с результатом;

$R^2_{yx1x2\ldots x(i-1)\,x(i+1)\,\ldots xm}$ – тот же показатель детерминации, но без введения в модель фактора x_i.

Частный F-критерий Фишера модели множественной регрессии для фактора x_i:

$$F_{xi} = \frac{R^2_{yx1...xi...xm} - R^2_{yx1...x(i-1)x(i+1)...xm}}{1 - R^2_{yx1...xi...xm}} * \frac{n - m - 1}{1}\ ;$$

где в числителе показан прирост доли объясненной вариации y за счет дополнительного включения в модель соответствующего фактора. А в знаменателе – доля остаточной вариации по регрессионной модели, включающей полный набор факторов.

Частным уравнением регрессии модели $y=a_0+a_1 x_1 + a_2 x_2 +, \ldots, +a_m x_m + \varepsilon$ называется уравнение вида $y_{xi\ \ x1,\ldots,\ x(i-1),\ \ x(i+1),\ \ldots,\ xm} = f(\bar{x}_1,\ldots,\bar{x}_{i-1},x_i,\bar{x}_{ii+1},\ldots\bar{x}_m)$, то есть уравнения регрессии, которые связывают результативный признак с соответствующими факторами x при закреплении других учитываемых в уравнении регрессии факторов на среднем уровне.

При подстановки в уравнение средних значений соответствующих факторов уравнение принимает вид парных уравнений линейной регрессии.

3.9.Пример построения модели множественной регрессии с помощью пакета Excel

Задание

Даны статистические данные, описывающие зависимость цены от потребительских свойств станка (модель Ланкастера),
где: P – основной размер станка;
N – мощность главного привода;
n - максимальная частота вращения шпинделя;

УА – уровень автоматизации;
Т – класс точности
Y – цена станка.

№	Р, мм Размер	N, кВТ Мощность	n, об/мин Скорость вращения шпинделя	УА Уровень автоматизации	Т, Точность	Y Цена станка, тыс.руб.
1	400	11	2000	3	1	53,6
2	400	10	1500	1	1	43,6
3	400	8	2000	1	1	35
4	400	8	2500	1	1,6	39
5	320	8	2500	1	1	27
6	320	8	3000	3	1,6	44
7	320	6,3	2000	1	1	26
8	250	8	3000	3	1,6	32,7
9	250	6,3	3000	1	1,6	22,5
10	250	5,5	2500	1	1	20
11	320	8	2500	1	1,6	29,8
12	400	8	2000	1	1,6	37,5

Необходимо:

1. Построить модель парной регрессии.
2. Дать экономическую интерпретацию модели.
3. Провести анализ факторов на предмет мультиколлинеарности.
4. Построить стандартизованное уравнение регрессии.
5. Оцените значимость коэффициентов линейной регрессии. Построить доверительные интервалы для коэффициентов линейной регрессии.
6. Оценить качество модели, рассчитав коэффициент детерминации. Применить F – критерий Фишера для проверки качества оценивания
7. Спрогнозировать значение y для какого – либо набора x_{j}..

Решение (Приложение Пример 2.).

Вопросы для обсуждения:

1. Почему необходимо часто строить модель множественной регрессии; приведите примеры экономических процессов и явлений, в которых вы бы применяли данную модель?
2. В чем отличие целей построения модели парной регрессии и модели множественной регрессии?
3. Объясните, почему в эконометрическом моделировании возникает проблема мультиколлинеарности?
4. Каким свойствам должны отвечать параметры модели множественной регрессии и почему?
5. Почему параметры «чистой» регрессии не сопоставимы между собой? Каковы методы разрешения данной проблемы?

6. Как должны соотноситься коэффициенты детерминации для m и m+1 факторов модели?
7. В чем заключается смысл расчета скорректированного индекса корреляции, и какова связь его с индексом корреляции при различных количествах вводимых в модель факторах?
8. Каков экономический смысл применения фиктивных переменных в модели множественной регрессии?
9. Сколько фиктивных переменных должно быть в модели, если необходимо оценить влияние двух качественных признаков, один из которых принимает три характеристики, а другой две?
10. Объясните необходимость построения частных коэффициентов корреляции.

<u>Тест:</u>

1 Основная цель множественной регрессии:
   a. построить модель с двумя факторами, не определяя при этом влияние каждого из них в отдельности;
   b. построить модель с большим числом факторов, определив при этом влияние каждого из них в отдельности, а также совокупное воздействие их на моделируемый показатель;
   c. определить совокупное воздействие факторов на моделируемый показатель.

2 Факторы, включаемые во множественную регрессию, должны отвечать следующим требованиям:
   a. быть количественно измеримы;
   b. быть качественно определены;
   c. быть взаимосвязанными.

3 При дополнительном включении в регрессию (p+1) фактора:
   a. коэффициент детерминации должен уменьшаться, а остаточная дисперсия возрастать;
   b. коэффициент детерминации должен уменьшаться и остаточная дисперсия уменьшаться;
   c. коэффициент детерминации должен возрастать, а остаточная дисперсия уменьшаться.

4 Мультиколлинеарность факторов в модели множественной регрессии:
   a. не всегда ведет к несостоятельным оценкам;
   b. всегда ведет к несостоятельным оценкам;
   c. не позволяет применять МНК вообще.

5 Две переменные коллинеарны, если:
   a. $r_{xixj} \geq 0,7$;

b. $r_{xixj} < 0,7$.

6 Если определитель матрицы межфактороной корреляции равен 1, то это свидетельствует о:

a. полной мультиколлинеарности факторов;
b. полном отсутствии мультиколлинеарности факторов;
c. о высокой мультиколлинеарности факторов.

7 К методам преодоления межфакторной корреляции относится:

a. увеличение дисперсии случайного параметра;
b. уменьшение числа наблюдений;
c. увеличение дисперсии объясняющих переменных.

8 Особенностью коэффициентов «чистой» регрессии является:

a. сравнимость между собой;
b. несравнимость между собой;
c. отсутствие экономической интерпретации.

9 Стандартизованные коэффициенты множественной регрессии:

a. показывают, что с изменением фактора x_j на одно выборочное стандартное отклонение s_j, при неизменном среднем уровне других факторов, у изменяется на η_j выборочных стандартных отклонений;
b. среднее изменение результата с изменением соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне.

10 Коэффициенты регрессии будут более неточными:

a. чем больше число наблюдений в выборке;
b. чем меньше дисперсия объясняющих переменных;
c. чем меньше теоретическая дисперсия случайного параметра.

11 Скорректированный индекс множественной регрессии:

a. завышает обычный индекс множественной регрессии;
b. занижает обычный индекс множественной регрессии;
c. равен обычному индексу множественной регрессии.

12 Фиктивные переменные отражают в модели:

a. количественные показатели;
b. качественные показатели;
c. как те, так и другие.

Задания для самостоятельной работы:

Задание I

Даны статистические данные, описывающие зависимость y от $x_1, \ldots x_m$.

Необходимо:

1. Построить модель парной регрессии.
2. Дать экономическую интерпретацию модели.
3. Провести анализ факторов на предмет мультиколлинеарности.
4. Построить стандартизованное уравнение регрессии.
5. Оцените значимость коэффициентов линейной регрессии. Построить доверительные интервалы для коэффициентов линейной регрессии.
6. Оценить качество модели, рассчитав коэффициент детерминации. Применить F – критерий Фишера для проверки качества оценивания
7. Спрогнозировать значение y для какого – либо набора x_{j}..

№1.

№ торговых предприятий	Факторы		Уровень рентабельности, %
	Удельный вес продовольственных товаров в товарообороте, %	Среднемесячная оплата труда, руб.	
1	74,2	1560	3,62
2	73,5	1620	3,8
3	77	1490	2,77
4	84,3	1330	2,12
5	67,3	1970	4,33
6	70,1	1820	4,01
7	83,1	1270	2,01

№2.

№	Факторы					Уровень рентабельности, %
	Среднемесячный товарооборот в расчете на душу населения	Удельный вес продовольственных товаров в товарообороте, %	Время обращения товаров, дней	Среднемесячная оплата труда	Трудоемкость товаро-оборота (численность работников на 100000 ед. товарооборота)	
1	27	74,2	35	1560	11	3,62
2	29	73,5	32	1620	12	3,8
3	28	77	33	1490	13	2,77
4	21	84,3	41	1330	17	2,12
5	35	67,3	29	1970	9	4,33
6	33	70,1	31	1820	10	4,01
7	21	83,1	39	1270	18	2,01

№3.

№ завода	Факторы		Производительность труда
	Удельный вес рабочих с технической подготовкой, %	Удельный вес механизированных работ, %	
1	64	84	4300
2	61	83	4150
3	47	67	3000
4	46	63	3420
5	49	69	3300
6	54	70	3400
7	53	73	3420
8	61	81	4100
9	57	77	3700
10	54	72	3500
11	60	80	4000
12	67	85	4450
13	63	83	4270
14	50	70	3300
15	67	87	4500

№4.

Статистические данным, описывающие зависимость накопления пяти случайно выбранных семей от дохода и размера имущества. Спрогнозируйте накопления семьи, имеющей доход 40 тыс. руб. и имущество стоимостью 25 тыс. руб.

Семья	Накопления, S	Доход, Y	Имущество, W
1	3,0	40	60
2	6,0	55	36
3	5,0	45	36
4	3,5	30	15
5	1,5	30	90

№5.

В кейнсианской теории спрос на деньги зависит от доходов и процентных ставок. Рассмотрим модель: $m_t=a_0+ a_1y_t+a_2i_t+\varepsilon_t$; где
m_t – агрегат денежной массы М1 (млрд. долл.),
y_t- валовой национальный доход (млрд. долл.);
i_t – процентные ставки по государственным облигациям (%).
Оцените спрос на деньги при 1) ВНД = 1000 и i = 10%; 2) ВНД= 2500 и i = 5%..

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14
m_t	141	146	149	154	161	173	199	216	251	277	310	363	414	478
y_t	506	524	565	597	637	756	873	992	1185	1434	1718	2164	2631	3073
i_t	3,3	2,6	2,9	3,2	3,7	5	5,5	6,6	4,5	7,9	5,3	7,6	11,4	11,1

№ 6.

Статистические данные реального дохода на душу населения y (тыс. долл.) процента рабочей силы, занятой в сельском хозяйстве – x_1, и среднего уровня образования населения в возрасте после 25 лет x_2 (число лет, проведенных в учебных заведениях) для 15 развитых стран с 1983 г.

Страна	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y	7	9	9	8	8	14	9	8	10	11	11	12	9	10	12
x_1	8	9	7	6	10	4	5	5	6	7	6	4	8	5	8
x_2	9	13	11	11	12	16	11	11	12	14	11	15	15	10	13

№7

Данные о величинах объема реализации продукции y некоторой фирмы, зависящие от цены x_1 и расходов на рекламу x_2.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Y, шт.	23	18	27	29	43	23	55	47	35	38	14	51	20	39	35
x_1, у.е.	37	33	15	36	26	24	15	33	44	34	63	8	44	43	31
x_2,у.е.	39	40	35	48	53	42	54	54	50	53	46	50	43	55	51

№8.

Данные о зависимости цены товара y, руб. от дальности его перевозки x_1, км и расходов на рекламу x_2., тыс руб.

	1	2	3	4	5	6	7
Y, руб	46,72	51,01	49,39	71,71	65,16	67,27	40,09
x_1, км.	10	17	15	25	19	20	8
x_2, тыс. руб	9	13	9	10	5	6	11

№9.

Зависимость выработки продукции на одного работника y, тыс. руб. от ввода в действие новых основных фондов x_1, % от стоимости фондов на конец года и от удельного веса рабочих высокой квалификации в общей численности рабочих x_2, %.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
y	23,1	27,5	23,8	28,9	28,1	33,5	31,7	32,8	32,4	39,2	34,7	36,8	38,7	41,3	46,8	51,7
X_1	5,6	4,4	2,2	7	3,7	6,2	7,5	5,3	5,2	6,7	5,9	7,3	7,6	5,9	7,2	9,1
X_2	10	14	15	16	17	19	19	20	20	20	21	22	22	25	28	29

№10.

По 20 крупнейшим компаниям имеются данные за год. Y – чистая прибыль, млрд. долл., x1 – оборот капитала, млрд. долл., x2 – использованный капитал, млрд. долл., x3 – численность служащих, тыс.чел.

№	Чистая прибыль y	Оборот капитала	Использованный капитал	Численность служащих
1	7,9	6,9	83,6	222
2	2,4	18	6,5	32

3	9,2	107	50,4	82
4	2,8	16,7	15,4	45,2
5	4,7	79,6	29,6	299,3
6	1,9	16,2	13,3	41,6
7	0,5	5,9	5,9	17,8
8	4,5	53,1	27,1	151
9	3,2	18,8	11,2	82,3
10	3,8	35,3	16,4	103
11	4	71,9	32,5	225,4
12	1,6	93,6	25,4	675
13	1,7	10	6,4	43,8
14	3,7	31,5	12,5	102,3
15	2,9	36,7	14,3	105
16	2,7	13,8	6,5	49,1
17	4,1	64,8	22,7	50,4
18	2,1	30,4	15,8	480
19	2,3	12,1	9,3	71
20	4	31,3	18,9	43

№11.

Менеджер отдела кадров компании заинтересован в получении обоснованного прогноза, сможет ли определенный кандидат стать хорошим продавцом. Для этой цели в качестве зависимой переменной у он выбрал данные об объеме продаж за месяц работы и решил принять к рассмотрению следующие независимые переменные:

X1 – результат теста способностей к продаже;

X2 – возраст;

X3 – результат теста тревожности;

X4 - опыт работы;

X5 – средний балл школьного аттестата.

№	Объем продаж за месяц, шт.	Результат теста способностей	Возраст	Результат теста тревожности	Опыт работы	Средний балл школьного аттестата
1	44	10	22,1	4,9	0	2,4
2	47	19	22,5	3	1	2,6
3	60	27	23,1	1,5	0	2,8
4	71	31	24	0,6	3	2,7
5	61	64	22,6	1,8	2	2
6	60	81	21,7	3,3	1	2,5
7	58	42	23,8	3,2	0	2,5
8	56	67	22	2,1	0	2,3
9	66	48	22,4	6	1	2,8
10	61	64	22,6	1,8	1	3,4
11	51	57	21,1	3,8	0	3
12	47	10	22,5	4,5	1	2,7
13	53	48	22,2	4,5	0	2,8
14	74	96	24,8	0,1	3	3,8
15	65	75	22,6	0,9	0	3,7

16	33	12	20,5	4,8	0	2,1
17	54	47	21,9	2,3	1	1,8
18	39	20	20,5	3	2	1,5
19	52	73	20,8	0,3	2	1,9
20	30	4	20	2,7	0	2,2
21	58	9	23,3	4,4	1	2,8
22	59	98	21,3	3,9	1	2,9
23	52	27	22,9	1,4	2	3,2
24	56	59	22,3	2,7	1	2,4
25	49	23	22,6	2,7	1	2,4
26	63	90	22,4	2,2	2	2,6
27	61	34	23,8	0,7	1	3,4
28	39	16	20,6	3,1	1	2,3
29	62	32	24,4	0,6	3	4
30	78	94	25	4,6	5	3,6

№12.

Менеджер по продажам фирмы, занимающейся реализацией запчастей к автомобилям, хотел бы найти модель, с помощью которой уже в мае можно было бы спрогнозировать годовой объем продаж в регионе. Если этот объем можно спрогнозировать для каждого региона, то можно будет составить прогноз продаж и для всей компании в целом. Количество пунктов розничной торговли данной компании в регионе и количество автомобилей, зарегистрированных в регионе на 1 мая, - это две независимые переменные.

Регион	1	2	3	4	5	6	7	8	9	10	11
Годовой объем продаж, (млн. долл.)	52,3	26	20,2	16	30	46,2	35	3,5	33,1	25,2	38,2
Количество пунктов обслуживания	2011	2850	650	480	1694	2302	2214	125	1840	1233	1699
Количество, зарегистрированных автомобилей, млн. шт.	24,6	22,1	7,9	12,5	9	11,5	20,5	4,1	8,9	6,1	9,5

Задание II.

Даны статистические данные, среди которых присутствует качественный фактор.

Необходимо:

Построить уравнение множественной регрессии с применением фиктивных переменных и оценить ее значимсоть.

№1.

Исследователю необходимо изучить, насколько некоторый тест способностей сможет предсказать будущую производительность труда работника. Восемь женщин и семь мужчин выполнили предусмотренные тестом задания, предназначенные для оценки ловкости рук при работе с мелкими предметами. После этого каждый из протестированных прошел месячный курс подготовки к работе сборщиком электронных плат, а затем в течение месяца выполнял соответствующую работу, после чего его производительность была оценена индексом, принимающим значение от 0 до 10.

Работник	Оценка производительности	Данные теста способностей	Пол
1	5	60	Ж
2	4	55	Ж
3	3	35	Ж
4	10	96	Ж
5	2	35	Ж
6	7	81	Ж
7	6	65	Ж
8	9	85	Ж
9	9	99	М
10	2	43	М
11	8	98	М
12	6	91	М
13	7	95	М
14	3	70	М
15	6	85	М

№2.

Необходимо исследовать зависимость между результатами письменных вступительных y и курсовых экзаменов x по математике. Получены данные о числе решенных задач (задание – 10 задач) на вступительных и курсовых экзаменах 16 студентами, а также распределение этих студентов по категории «пол».

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y	9	8	2	10	4	4	9	2	10	4	9	3	9	1	8
x	10	8	5	10	6	6	10	4	10	6	9	6	10	4	9
пол	м	м	ж	м	ж	ж	м	ж	м	ж	м	ж	м	ж	м

№3.

Зависимость средней заработной платы y, долл. от возраста рабочего x, лет.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
y	338	323	299	334	305	327	336	319	340	315	323	321	293	313	352
X	35	55	32	44	33	58	38	50	45	47	38	35	35	48	58
пол	м	ж	ж	м	ж	ж	м	ж	м	ж	м	м	ж	ж	м

№4.

Зависимость текущих расходов школы от численности учащихся и типа школы.

Школа	Тип	Расходы	Численность учащихся
1	Профессиональная	345000	623

2	Профессиональная	537000	653
3	Обычная	170000	400
4	Профессиональная	526000	663
5	Обычная	100000	563
6	Обычная	28000	236
7	Обычная	160000	307
8	Профессиональная	45000	173
9	Профессиональная	120000	146
10	Профессиональная	61000	99

№5.

Зависимость текущих расходов школы от численности учащихся и типа школы.

Школа	Тип	Расходы	Численность учащихся
1	Техническая	345000	623
2	Техническая	537000	653
3	Общеобразовательная	170000	400
4	Подготовки квалифицированных рабочих	526000	663
5	Общеобразовательная	100000	563
6	Специализированная	28000	236
7	Специализированная	160000	307
8	Техническая	45000	173
9	Техническая	120000	146
10	Подготовки квалифицированных рабочих	61000	99

ТЕМА 4. НЕЛИНЕЙНЫЕ МОДЕЛИ РЕГРЕССИИ

Ключевые слова:

Нелинейные модели внутренне линейные. Нелинейные модели внутренне нелинейные. Коэффициенты эластичности. Коэффициент роста. Уровень насыщения. Замена переменных. Логарифмирование. Потенциирование. Производственная функция Кобба Производительность факторов производства.– Дугласа. Кривые Энгеля. Функции спроса.

Основные теоретические аспекты темы

4.1. Модели нелинейные по переменным

Во многих практических случаях моделирование экономических зависимостей с помощью линейных уравнений дает вполне удовлетворительные результаты и может использоваться для анализа и прогнозирования поведения экономических объектов и систем. Однако в силу многообразия и сложности экономических объектов и процессов описание их поведения только с помощью линейных моделей невозможно. Многие экономические зависимости не являются линейными по своей природе, и поэтому их моделирование возможно лишь на основе нелинейных уравнений регрессии,

Так, зависимость объема выпуска продукции с основными факторами производства — трудом и капиталом (производственная функция Кобба — Дугласа), зависимость спроса на товары и услуги от цены и располагаемого дохода и многие другие зависимости являются нелинейными.

Выбор вида зависимости осуществляется на основании экономического анализа исследуемого объекта, а также по результатам изучения характера взаимосвязей между переменными, входящими в модель.

Достоинством процедуры МНК применительно к линейным моделям является то, что система нормальных уравнений для нахождения оценок параметров модели получается линейной, решение которой относительно просто. В случае с нелинейными моделями процедура МНК приводит к нелинейной системе нормальных уравнений, многообразие которых не позволяет предложить универсальную процедуру их решения.

Основной прием, который используется для построения нелинейных регрессионных моделей, — линеаризация, который заключается в искусственном преобразовании исходной спецификации нелинейной модели к линейному виду.

Различают два вида нелинейных моделей, внутренне линейных: нелинейные модели по переменным и нелинейные модели по параметрам. Типичным представителем модели нелинейной одновременно по переменным и по параметрам является производственная функция Кобба Дугласа. В ней аргументы присутствуют в виде сомножителей, а параметрами служат показатели степени при аргументах.

Регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам имеют вид:

$y= a+b/x+\varepsilon$ - гипербола;
$y=a+bx+cx^2 +\varepsilon$– парабола;
$y=a+bx+cx^2+dx^3+\varepsilon$ - кубический многочлен;
$y= a+b\ln x+\varepsilon$ – полулогарифмическая функция;
$y= 1/(a+bx)+ \varepsilon$;
$\lg y=a+bx+cx^2+\varepsilon$
$y=a+bx+c(1/x)+ \varepsilon$;
$y=1/(a+bx+cx^2)+ \varepsilon$

Данные функции сводятся к линейным путем замены переменных. Тогда относительно новых переменных модель принимает вид уравнения множественной линейной регрессии (кроме гиперболы). Исходя из замены переменных, следует построить новую выборку наблюдений и воспользоваться имеющимися знаниями по оценке и анализу таких моделей.

Полиномы второй степени нашли применение:

- при моделировании средних и предельных издержек в зависимости от обема выпуска продукции;
- при моделировании зависимости прибыли предприятия от расходов на рекламу.

Полиномы третьей степени применяются при моделирования общих издержек в зависимости от объема выпуска продукции.

Модели гиперболического типа успешно применяются в экономике при моделировании:

- зависимости спроса от цен;
- зависимости спроса от дохода (кривые Эйнгеля);
- спроса на предметы роскоши от дохода (функции Торсвиста);
- уровня относительного изменения заработной платы в зависимости от относительного изменения уровня безработицы (кривая Филипса).

Гиперболические модели используются в тех случаях, когда при неограниченном увеличении фактора значение эндогенной переменной ассимптотически стремится к некоторому значению. При стремлении фактора к бесконечности значение эндогенной переменной стремится к значению параметра *a,* который экономически интерпретируется как предельный уровень эндогенной переменной. Параметр b характеризует скорость приближения эндогенной переменной к этому уровню.

Модели полулогарифмические. К данному типу моделей, как правило, относят уравнения парной регрессии, в которых объясняющая переменная используется в логарифмическом виде. Полулогарифмические модели используются в случаях, когда ставится задача определить темп роста или темп прироста каких-либо экономических показателей, например, при анализе банковского вклада по его первоначальному значению и процентной ставки, при исследовании зависимости прироста объема выпуска от относительного увеличения затрат ресурсов, бюджетного дефицита от темпа роста ВНП, темпа роста инфляции от объема денежной массы в обращении и т.п. Спецификация полулогарифмической модели вид: $y= a+b\ln x+\varepsilon$.

4.2. Модели нелинейные по оцениваемым параметрам

К нелинейным по параметрам моделям, оценка которых возможна методом линеаризации, относятся следующие: степенная модель, показательная модель, логистическая модель и некоторые мультипликативные комбинации из этих моделей.

Степенная модель:

$$y=ax^b\varepsilon.$$

Степенные модели достаточно широко используются в экономике. К классу степенных функций относятся, в частности, модели спроса и предложения (кривые Эйнгеля), производственные функции, кривые освоения для характеристики связи между трудоемкостью продукции и масштабами производства в период освоения и выпуска нового вида продукции, а также зависимость валового национального дохода от уровня занятости.

Метод линеаризации степенных моделей заключается в логарифмировании обеих частей уравнения по натуральному или десятичному основанию и осуществления замены. Именно это послужило основанием к тому, чтобы случайное возмущение в модели записать не в виде слагаемого, а в виде сомножителя. Значение случайного возмущения в этом случае показывает долю случайного возмущения в значении эндогенной переменной.

В результате логарифмирования и заены строится уравнение множественной линейной регрессии относительно новых переменных. Для спецификации формируется выборка наблюдений в соответствии с правилами и производится оценка и анализ модели. Чтобы получить модель в исходном виде , достаточно вычислить значение параметра a и подставить оценки всех параметров в выражение.

Интерес представляет экономическая интерпретация параметров модели, а именно коэффициента b. b - называется *коэффициентом эластичности*. Он показывает, на сколько процентов изменится в среднем результат при изменении соответствующего фактора x на 1%.

Множественная степенная модель:

$$y=a\, x_1^{b1}\, x_2^{b2}\, ,\ldots ,\, x_m^{\ bm}$$

Экономический смысл коэффициентов b_1, b_2 , b_m: b_1, b_2 , b_m в том, что это коэффициенты эластичности. Они показывают на сколько процентов изменится в среднем результат при изменении соответствующего фактора на 1% при неизменности действия других факторов.

Показательная функция или функция постоянного роста:

$$y=ab^x\varepsilon;$$

$$y=e^{\ a+bx}\varepsilon;$$

$$y=e^{\ a+b/x}\varepsilon;\ s – \text{образная кривая.}$$

В случае, когда в качестве основания степени используется константа e, модель называют экспоненциальной, а exp(b) определяет уровень насыщения.

Показательные модели характеризуются постоянным темпом относительного прироста эндогенной переменной, поэтому экономическая интерпретация коэффициента b: b– коэффициент роста.

Линеаризация модели производится также с помощью логарифмирования.

Множественная модель:

$$y = ab_1^{x1}b_2^{x2}...b_m^{xm}\varepsilon$$

Это функция постоянного роста, а коэффициенты - коэффициенты роста по фактору x_i.

Логистическая модель:

Данную модель можно отнести к комбинации гиперболической и показательных моделей. Ее спецификация имеет вид:

$$y = \frac{1}{a + be^{-x} + \varepsilon}.$$

С одной стороны функция гипербола, с другой- объясняющая переменная участвует в ней в виде показательной степени при экспоненте. График функции имеет две горизонтальные асимптоты y=0 и y=1/a и точку перегиба:($x = \ln\left(\frac{b}{a}\right)$; $y = \frac{1}{2a}$).

Функция представляет собой частный случай логистической кривой, которая впервые была применена А.Кетле.

Логистические функции используются для описания поведения экономических показателей, имеющих уровни «насыщения», например, для описания зависимости спроса на товар от дохода, развитие производства нового товара от роста численности населения и т.п.

Тест Бокса – Кокса – тест сравнения коэффициента R^2 для линейной и логарифмической вариантов модели.

Тест предполагает такое преобразование масштаба наблюдений y, при котором обеспечивалась бы возможность непосредственного сравнения RSS в линейной и логарифмических моделях.

1. Вычисляется среднее геометрическое значений y в выборке. Оно совпадает с экспонентой среднего арифметического lny, которое легко рассчитать: $e^{\frac{1}{n}\sum \ln y} = (y_1 * ... * y_n)^{\frac{1}{n}}$;
2. Пересчитываются наблюдения y, они делятся на это значение: $y_i^* = \frac{y_i}{среднеегеометрическое\_y}$;
3. Оценивается регрессия для линейной модели с использованием y^* вместо y в качестве зависимой переменной и для логарифмической модели с использованием $\ln y^*$ вместо lny; во всех других отношениях модели должны оставаться неизменными. Теперь значения RSS для двух регрессий сравнимы, и, следовательно, модель с меньшей суммой квадратов отклонений обеспечивает лучшее соответствие.
4. Для того чтобы проверить, обеспечивает ли одна из моделей значимо лучшее соответствие, можно вычислить величину $-\frac{n}{2}\ln z$, где z – отношение RSS в пересчитанных регрессиях, а n – число наблюдений, и взять ее абсолютное значение. Эта статистика имеет распределение χ^2 (хи - квадрат) с одной степенью свободы. Если она превышает

критическое значение χ^2 при выбранном уровне значимости, то делается вывод о наличии значимой разницы в качестве оценивания.

Вопросы для обсуждения:

1. Объясните необходимость построения нелинейных моделей парной и множественной регрессии в экономике.
2. В чем принципиальное отличие нелинейных, внутренне линейных, функций и нелинейных, внутренне нелинейных, функций?
3. Приведите области в экономике, в которых можно использовать: гиперболу, степенную функцию, показательную, параболу, кубический многочлен.
4. Какие методы используются для сведения нелинейных функций к линейному виду?
5. Какой способ линеаризации наиболее удобен для моделей линейных по параметрам?
6. Какой способ линеаризации используется для нелинейных мультипликативных моделей?
7. Почему не сравнимы между собой коэффициенты детерминации линейной модели и модели, построенной при использовании логарифмов этих же данных?
8. В чем особенность оценки статистической значимости нелинейных моделей парной регрессии?

Тест:

1 МНК не применим к моделям:
   a. линейным;
   b. нелинейным, внутренне нелинейным;
   c. нелинейным, внутренне линейным.

2 Линеаризацию методом замены можно осуществить в модели:
   a. $y=a+bx+cx^2+\varepsilon$ – парабола;
   b. $y=ax^b\varepsilon$;
   c. $y=e^{a+bx}\varepsilon$.

3 Линеаризацию методом логарифмирования можно осуществить в модели:
   a. $y= a+b/x+\varepsilon$ - гипербола;
   b. $y=ax^b\varepsilon$;
   c. $y=1/(a+bx+cx^2)+ \varepsilon$.

4 Показатель степени в степенной функции является:

a. показателем чистой регрессии;
b. показателем эластичности;
c. показателем постоянного роста.

5 Основание в показательной функции является:
a. показателем чистой регрессии;
b. показателем эластичности;
c. показателем постоянного роста.

6 Коэффициенты детерминации для линейного и логарифмического уравнения:
a. не сравнимы между собой;
b. сравнимы между собой;
c. не возникает необходимости их сравнивать.

7 Если сумма показателей степени в производственной функции Кобба – Дугласа равны 1, то наблюдается:
a. положительный эффект масштаба;
b. отрицательный эффект масштаба;
c. постоянный эффект масштаба.

8 Для построения степенной модели МНК необходимо статистические данные:
a. потенциировать;
b. логарифмировать;
c. не изменять, использовать в единицах измерения.

Задания для самостоятельной работы:

№1.

Кривая Филипса описывает связь темпа роста зарплаты и уровня безработицы. А именно: $\delta\omega_t = \beta_1 + \beta_2 * \frac{1}{u_t} + \varepsilon_t$, где ω_t –уровень заработной платы, $\delta\omega_t = 100(\omega_t - \omega_{t-1})/\omega_{t-1}$ - темп роста зарплаты (в процентах) и u_t – процент безработных в год t. Используя данные для некоторой страны постройте уравнение парной регрессии и проверьте наличие значимой связи между $\delta\omega$ и u. Найдите «естественный уровень безработицы», т. е. такой уровень безработицы, при котором $\delta\omega=0$.

Год t	ω_t	u_t
1	1.62	1
2	1.65	1.4
3	1.79	1.1
4	1.94	1.5

5	2.03	1.5
6	2.12	1.2
7	2.26	1.0
8	2.44	1.1
9	2.57	1.3
10	2.66	1.8
11	2.73	1.9
12	2.8	1.5
13	2.92	1.4

№2.

Менеджер новой чебуречной не уверен в правильности выбранной цены на чебуреки, поэтому в течение 12 недель он варьирует цену и записывает количество проданных чебуреков. Получены следующие данные.

1. Постройте линейную модель парной регрессии. Найдите оптимальную, в смысле максимума выручки от продаж цену чебурека. Какую ценовую политику следует предпринять менеджеру чебуречной?
2. Постройте модель степенной функции и оцените, эластичен ли спрос на чебуреки? Совпадает ли ваш вывод о ценовой политике с предыдущим?
3. Проведите тест Бокса – Кокса.

Неделя	1	2	3	4	5	6	7	8	9	10	11	12
Количество проданных чебуреков q, шт.	795	915	965	892	585	644	714	1180	851	779	625	1001
Цена чебуреков, p, руб.	12,3	11,5	11	12	13,5	12,5	12,8	9,9	12,2	12,5	13	10,5

№3.

После финансового кризиса спрос на чебуреки упал, и менеджер был вынужден тратить часть средств на рекламу. Для изучения зависимости объема продаж от цены и расходов на рекламу менеджер использует следующую модель: $q_t=a_0+ a_1p_t+a_2c_t+a_3c_t^2+\varepsilon_t$.

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
q_t	525	567	396	726	265	615	370	789	513	661	407	608	399	631	545	512	845	571
p_t	5,9	6,5	6,5	6,1	6,6	5,2	5,1	5,1	6,7	5,5	6,6	6,9	6,9	6,5	6,5	6,8	5,1	6,1
c_t	479	361	549	278	574	134	581	339	374	359	519	327	469	379	429	271	221	309

1. Пусть себестоимость производства одного чебурека = 2 руб. Найдите формулу выручки и прибыли.
2. Найдите оптимальную цену, при которой прибыль достигает максимального значения, при расходах на рекламу = 280 руб.
3. Найдите оптимальный уровень расходов на рекламу, при котором прибыль достигает максимального значения, при цене чебурека =6 руб.

4. Помогите менеджеру найти оптимальное решение для цены и объема расходов на рекламу, при которых прибыль оптимальна (достигает максимального значения).

№4.

Дана зависимость выпуска Q от трудозатрат L и капиталовложений K 15 фирм некоторой отрасли. Оцените по этим данным производственную функцию Кобба-Дугласа $Q = \alpha L^{\beta 1} K^{\beta 2}$.

Фирма №	Q	L	K
1	2350	2334	1570
2	2470	2425	1850
3	2110	2230	1150
4	2560	2463	1940
5	2650	2565	2450
6	2240	2278	1340
7	2430	2380	1700
8	2530	2437	1860
9	2550	2446	1880
10	2450	2403	1790
11	2290	2301	1480
12	2160	2253	1240
13	2400	2367	1660
14	2490	2430	1850
15	2590	2470	2000

Рассчитайте объем выпуска при L = 2500 и K = 1800.

№5.

Выберите модель нелинейной парной регрессии, описывающей зависимость между ежегодным потреблением бананов y, кг. и годовым доходом, x, тыс. руб.

Семья	1	2	3	4	5	6	7	8	9	10
Y	1,93	7,14	8,78	9,69	10,09	10,42	10,62	10,71	10.79	11,13
x	10	20	30	40	50	60	70	80	90	100

№6.

Выберите модель нелинейной парной регрессии, описывающей зависимость между ежегодным потреблением апельсинов y, кг. и годовым доходом, x, тыс. руб. для 10 семей.

Семья	1	2	3	4	5	6	7	8	9	10
Y	1,15	4,71	5,59	7,43	7,47	7,33	7,7	8,15	8,35	8,59
x	10	20	30	40	50	60	70	80	90	100

№7.

Динамика выпуска продукции y, млн. долл. некоторой страны за 30 лет характеризуется данными. Постройте экспоненциальную модель.

Год, t	1	2	3	4	5	6	7	8	9	10
y	1089	1006	1450	1273	1983	2076	2017	2193	2170	2378
Год, t	11	12	13	14	15	16	17	18	19	20
y	2914	3100	3741	3286	3706	5608	4326	6400	6373	9995

Год, t	21	22	23	24	25	26	27	28	29	30
y	7709	7977	9634	9818	9453	12553	9445	18299	12888	18581

№8.

Выберите модель нелинейной парной регрессии, описывающей зависимость между рентабельностью продукции y, %, от ее трудоемкости x, чел. на ед. продукции для 15 предприятий.

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Y	58,8	65,9	23,5	15,4	13,9	12,3	19,9	13,6	9,7	11,2	10,3	10,5	9,13	7,5	6,4
x	1	1,2	1,5	1,8	2	2,1	2,3	2,4	2,8	3	3,1	3,3	3,7	4,1	4,6

№9.

Имеются данные об индексах реального объема производства, реальных капитальных затрат и реальных затрат труда в промышленности некоторой страны за 20 лет. Постройте функцию Кобба – Дугласа.

Год	Y	K	L	Год	Y	K	L
1	100	100	100	11	169	197	143
2	110	109	109	12	175	210	147
3	109	119	108	13	180	217	149
4	115	120	116	14	191	225	153
5	126	135	120	15	195	238	155
6	133	139	122	16	192	247	151
7	142	152	127	17	199	268	153
8	152	166	135	18	244	299	176
9	154	178	139	19	268	337	189
10	161	189	138	20	282	370	195

№10.

По данным зависимости прибыли фирмы от затрат на рекламу (млн. у.е.) требуется:

1. Построить диаграмму рассеяния.
2. Построить и оценить линейную модель парной регрессии.
3. Построить и оценить полиномиальную модель второй степени.
4. Построить и оценить модель гиперболического типа.

Прибыль	5	7	13	15	20	25	25	24	25	25,5	24,5
Расходы на рекламу	0,8	1	1,8	2,4	4,1	5,6	7,3	8,2	8,8	9,2	10

№11.

В таблице приведены поквартальные данные динамики сбора налогов (трлн. руб) в Российской Федерации в период 1995-2013 гг.

Необходимо:

1. Построить диаграмму рассеяния и визуально оценить возможные виды моделей, подходящих для описания объекта.

2. Оценить модели полиномиальную, степенную и показательную.
3. Из числа оцененных моделей выбрать наилучшую и указать критерий отбора.

Год/квартал	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**
I	26,5	43,4	66,2	76	110,1	234,3
II	40,6	65,7	106,4	124,6	223,7	329,9
III	49,5	73,6	78,9	89,9	195,3	326,1
IV	53,2	103,1	104,7	129,4	229	358,2
Год/квартал	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**
I	283,8	356,1	440,7	518,7	870,7	1207,4
II	378,7	488,9	547,0	675,9	1004,5	1313,7
III	326,2	492,0	559,1	757,4	1145,7	1426,9
IV	413,9	508,7	565,7	923,4	1227,5	1438,7
Год/квартал	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**
I	1311,6	1741,6	1219,6	1760,7	2247,7	2809,4
II	1474,1	2115,8	1464,4	2004,7	2724,9	3159,8
III	1618,7	2332,3	1842,7	2114,5	2793,8	3066,4
IV	1929,9	2028,7	1947,9	2339,3	3113,6	3373,9

№12.

По данным таблицы оценить и проанализировать логистическую модель: $y = \dfrac{1}{a + be^{-x} + \varepsilon}$.

x	0,2	0,8	2	3	4	5	6	7
y	0,01	0,02	0,07	0,2	0,35	1,08	1,95	3
x	7,5	8	9	10	11	12	13	15
y	3,28	3,45	3,81	3,93	3,97	3,99	14,05	14

№13.

Анализируется индекс потребительских цен по объему денежной массы x (млрд у.е.) на основании данных с 1995 по 2011 г.

Год	95	96	97	98	99	00	01	02	03
y	65	68	72,5	77,5	82	85,5	88,5	91	95
x	110	125	132	137	160	177	192	215	235
Год	04	05	06	07	08	09	10	11	
y	100	106,5	112	115,5	118,5	120	120,5	121	
x	240	245	250	275	285	295	320	344	

1. Построить диаграмму рассеяния и визуально оценить возможные виды моделей, подходящих для описания объекта.
2. Оценить модели полиномиальную, степенную и показательную.
3. Из числа оцененных моделей выбрать наилучшую и указать критерий отбора.

ТЕМА 5. ГЕТЕРОСКЕДАСТИЧНОСТЬ И АВТОКОРРЕЛЯЦИИЯ

<u>Ключевые слова:</u>

Гомоскедастичность. Гетероскедастичность. Тест ранговой корреляции Спирмена. Тест Гольдфельда Квандта. Тест Глейзера. Взвешенный метод наименьших квадратов. Автокорреляция. Ложная автокорреляция. Положительная автокорреляция. Отрицательная автокорреляция. Авторегрессионные модели. Критерий Дарбина-Уотсона. Метод Кохрейна – Оркатта.

<u>Основные теоретические аспекты темы:</u>

5.1. Понятие геторoскедастичности: ее последствия и причины

Гетероскедастичность и автокорреляция представляют собой недостатки («патологию») регрессионного анализа, основанного на методе наименьших квадратов.

Во втором условии Гаусса – Маркова утверждается, что дисперсия случайного члена в каждом наблюдении должна быть постоянной. Случайные члены $u_1 \dots u_n$ в n наблюдениях появляются на основе вероятностных распределений, имеющих нулевое математическое ожидание и одну и ту же дисперсию. Их фактические значения в выборке иногда будут положительными, иногда – отрицательными, иногда – относительно далекими от нуля, иногда – относительно близкими к нему, но нет причин ожидать появления особенно больших отклонений в любом данном наблюдении. То есть, вероятность того, что величина u примет какое-то данное положительное или отрицательное значение, будет одинаковой во всех наблюдений. Это условие известно как *гомоскедастичность*, что означает одинаковый разброс (Рис.5.1.1.).

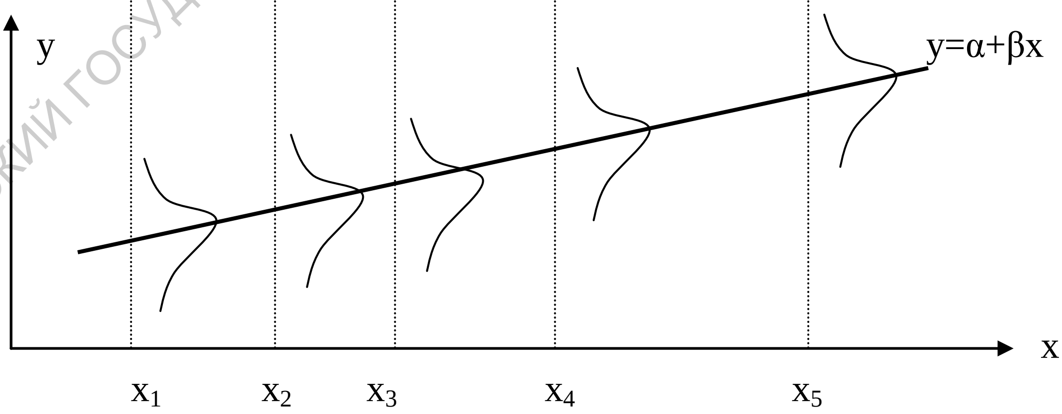


Рис.5.1.1. Гомоскедастичность.

Хотя гомоскедастичность в регрессионном анализе часто рассматривается как данная, в некоторых случаях более реалистичным оказывается предположение, что потенциальное распределение случайного члена в разных наблюдениях выборки различно. Это явление называется *гетероскедастичностью*, что отражает «неодинаковый разброс» (Рис. 5.1.2).

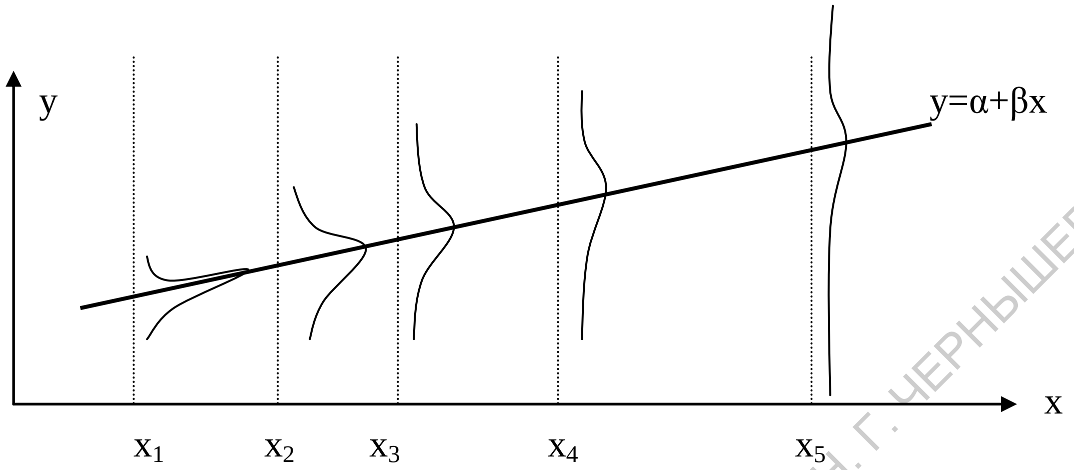


Рис.5.1.2. Гетероскедастичность.

Математически гомоскедастичность и гетероскедастичность могут определяться следующим образом:

Гомоскедастичность: $\sigma_{u_i}^2 = \sigma_u^2$, величина одинакова для всех наблюдений.

Гетеороскедастичность: $\sigma_{u_i}^2$ не одинакова для всех наблюдений.

Последствия гетероскедастичности:

Если имеет место гетероскедастичность, то МНК – оценки неэффективны, поскольку можно найти другие оценки, которые имеют меньшую дисперсию и, тем не менее, являются несмещенными.

Оценки стандартных ошибок коэффициентов регрессии будут неверны. Они вычисляются на основе предположения о том, что распределение случайного члена гомоскедастично. Если это не так, то они оказываются смещены, и вследствие этого t – критерии и обычный F- критерий неприменимы. Вероятно, что стандартные ошибки будут занижены, а следовательно t-статистика – завышена, и будет получено неправильное представление о точности коэффициентов регрессии. Можно решить, что коэффициент значимо отличается от нуля при данном уровне значимости, тогда как в действительности это не так.

Причины гетероскедастичности:

Гетероскедастичность становится проблемой, когда значения переменных, входящих в уравнение регрессии, значительно различаются в разных наблюдениях. Если в уравнении парной регрессии экономические переменные меняют свой масштаб одновременно, то вариации значений невключенных переменных и ошибки измерения, определяющие совместно значение случайного члена, часто сравнительно малы при малых y и x и сравнительно велики – при больших y и x.

5.2. Обнаружение гетероскедастичности

Рассмотрим три обычно используемых теста (критерия), в которых делаются различные предположения о зависимости между дисперсией случайного члена и величиной объясняющей переменной (переменных).

Тест ранговой корреляции Спирмена.

Предполагается, что дисперсия случайного члена будет либо увеличиваться, либо уменьшаться по мере увеличения x, и поэтому в регрессии, оцениваемой с помощью МНК, абсолютные величины остатков и значения x будут коррелированны.

Данные по x и абсолютные величины остатков упорядочиваются, и коэффициент ранговой корреляции определяется как:

$$r_{x_e} = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)};$$

где: D_i - разность между рангом x и рангом e в наблюдении i.

Если предположить, что соответствующий коэффициент корреляции для генеральной совокупности равен нулю, то коэффициент ранговой корреляции имеет нормальное распределением с математическим ожиданием 0 и дисперсией 1/(n-1) в больших выборках. Следовательно, соответствующая тестовая t- статистика равна $r_{x_e}\sqrt{n-1}$, и при использовании двустороннего критерия нулевая гипотеза о гомоскедастичности будет отклонена при уровне значимости 5%, если ее абсолютная величина превысит 1,96, и при уровне значимости в 1%, если ее абсолютная величина превысит 2,58.

Если в модели регрессии имеется более одной объясняющей переменной, то проверка гипотезы может выполняться с использованием любой из них.

Тест Гольдфельда - Квандта.

Наиболее популярным формальным критерием является критерий, предложенный С. Голдфельдом и Р.Квандтом в 1965 г. при проведении проверки по этому критерию предполагается, что стандартное отклонение распределения вероятностей случайного члена в наблюдении i пропорционально значению x_i. Предполагается также, что случайный член нормально распределен и удовлетворяет другим условиям Гаусса-Маркова.

Все n наблюдений в выборке упорядочиваются по величине x, после чего оцениваются отдельные регрессии для первых n’ и для последних n’ наблюдений; средние (n-2 n’) наблюдений отбрасываются.

Если имеет место гетероскедастичность и если предположение относительно ее природы верно, то дисперсия u в последних n’ наблюдениях будет больше, чем в первых n’, и это будет отражено в сумме квадратов остатков в двух указанных «частных» регрессиях.

Обозначая суммы квадратов остатков в регрессиях для первых n’ и последних n’ наблюдений соответственно через RSS_1 и RSS_2, рассчитывают отношение RSS_2/ RSS_1, которое при выполнении нулевой гипотезы о гомоскедастичности имеет F- распределение с (n’-k) и (n’-k) степенями свободы, где k- число объясняющих переменных в регрессионном уравнении.

Мощность критерия зависит от выбора n’ по отношению к n. Голдфелд и Квандт, основываясь на результатах некоторых проведенных ими экспериментов, утверждают, что n’ должно составлять порядка 3/8 от n, в частности около 11, если n=30, и около 22, если n=60.

Нулевая гипотеза для данного теста состоит в том, что RSS_2 не превышает значимо RSS_1, а альтернативная гипотеза – значимо превышает. Если величина RSS_2 оказалась меньшей, чем RSS_1, то мы не можем отвергнуть нулевую гипотезу, и вычислять тестовую статистику RSS_2/ RSS_1 нет нужды.

Данный метод может также использоваться для проверки на гетероскедастичность при предположении, что стандартное отклонение случайного члена обратно пропорционально x_i. При этом используется та же процедура, что и описанная выше, но тестовой статистикой является RSS_1/ RSS_2, который имеет F – распределение с (n’-k) и (n’-k) степенями свободы при выполнении нулевой гипотезы о гомоскедастичности.

Тест Глейзера.

Данный тест позволяет более тщательно рассмотреть характер гетероскедастичности. Снимается предположение о том, что стандартное отклонение случайного члена пропорционально x.

Следует оценить регрессионную зависимость y от x с помощью обычного МНК, а затем оценить регрессию абсолютных величин остатков вида: $|e| = a + bx^{\gamma}$ для данного значения γ.

Можно оценить несколько таких уравнений регрессии, изменяя значение γ.

В каждом случае нулевая гипотеза о гомоскедастичности будет отклонена, если оценка b значимо отличается от нуля. Если при оценивании более чем одной функции получается значимая оценка b, то ориентиром при определении характера гетероскедастичности может служить наилучшая из них.

5.3. Оценивание модели в условиях гетероскедастичности случайных возмущений

Подход к решению проблемы устранения гетероскедастичности сводится к искусственному преобразованию спецификации модели таким образом, чтобы условие гомоскедастичности выполнялось тождественно.

Необходимо задать правило вычисления стандартных ошибок случайных возмущений, разделить на эти ошибки переменные модели и сделать замену переменных. В результате появляется возможность получить модель с гомоскедастичными остатками.

Используется предположение тестов Голдфельда –Квандта и Спирмена о том, что ошибки случайных возмущений связаны с абсолютными значениями факторов x_i. Предположим, что стандартную ошибку случайных возмущений, можно представить в виде:

$$\sigma(u) = p = (1 + \sum_{j=1}^{k} x_j)^{\mu};$$

где: μ – показатель степени, с помощью которого учитывается возможность нелинейной связи между ошибкой остатка и абсолютным весом регрессоров.

Разделим на эту величину переменные модели:

$$\frac{y}{p} = a\frac{1}{p} + b_1\frac{x_1}{p} + ... + b_k\frac{x_k}{p} + \frac{u}{p}.$$

И сделав соответствующую замену, вновь получим модель в виде линейного алгебраического уравнения с гомоскедастичными остатками.

Начинают процесс устранения гетероскедастичности со значения $\mu=1$. Если при $\mu=1$ модель остается гетероскедастичной, то вводится приращение $\Delta\mu$ (например, $\Delta\mu=0{,}5$) и модель проверяется на гетероскедастичность при $\mu = \mu + \Delta\mu$.

Меняя знак и абсолютное значение приращения $\Delta\mu$, добиваются выполнения условия гомоскедастичности.

Функцию $\sigma(u) = p = (1 + \sum_{j=1}^{k} x_j)^{\mu}$ называют весовой функцией.

Формализуя данные процедуры получим алгоритм *взвешенного метода наименьших квадратов (ВМНК):*

1. Подбирается функция p, с помощью которой моделируется зависимость дисперсии случайных возмущений от суммарного веса факторов в уравнениях наблюдений;
2. Все переменные в каждом наблюдении умножаются на 1/p;
3. К преобразованной таким образом выборке применяется метод наименьших квадратов для получения оценок параметров.

5.4. Понятие автокорреляции и ее причины

Если третье условие Гаусса-Маркова не выполняется, то есть что теоретические ковариации u_i и u_j не равны нулю для любых $i \neq j$, то говорят, что случайный член подвержен *автокорреляции.*

Автокорреляция обычно встречается только в регрессионном анализе при использовании данных временных рядов. Случайный член в уравнении регрессии подвергается воздействию тех переменных, влияющих на зависимую переменную, которые не включены в уравнение регрессии. Если значение u в любом наблюдении должно быть независимым от его значения в предыдущем наблюдении, то и значение любой переменной, «скрытой» в u, должно быть некоррелированным с ее значением в предыдущем наблюдении.

Постоянная направленность воздействия невключенных в уравнение переменных является наиболее частой причиной положительной автокорреляции (Рис. 5.4.1.).

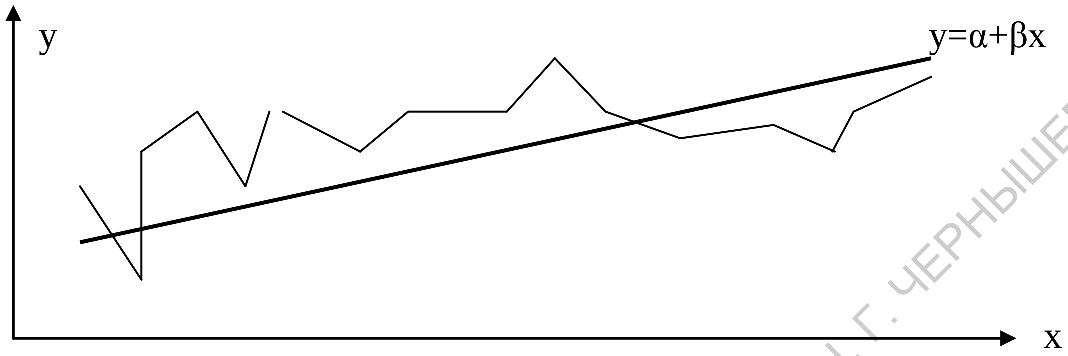


Рис.5.4.1. Положительная автокорреляция.

Если причиной автокорреляции является ошибка в спецификации модели (пропуск важного фактора, неправильный вид объясняющей части модели), то такую автокорреляцию называют *ложной*. Также в качестве причин автокорреляции могут быть ошибки измерения переменных модели и характер наблюдений.

Автокорреляция в целом представляет тем более существенную проблему, чем меньше интервал между наблюдениями. Чем больше этот интервал, тем менее правдоподобно, что при переходе от одного наблюдения к другому характер влияния неучтенных переменных будет сохраняться.

Автокорреляция может быть и отрицательной, то есть корреляция между последовательными значениями случайного члена отрицательна. В этом случае, скорее всего, за положительным значением в одном наблюдении идет отрицательное значение в следующем, и наоборот (Рис. 5.4.2.).

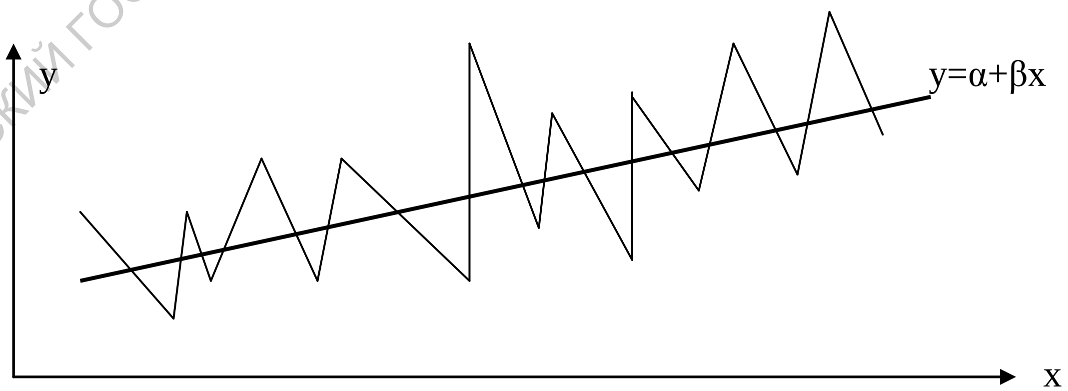


Рис.5.4.2. Отрицательная автокорреляция.

Модели с автокоррелированными остатками называются *авторегрессионными*.

Последствия автокорреляции для оценивания с помощью обычного МНК в некоторой степени сходны с последствиями гетероскедастичности. Коэффициенты регрессии остаются несмещенными, но становятся неэффективными, так как можно найти альтернативные несмещенные оценки с меньшей дисперсией. Стандартные ошибки оцениваются неправильно (чаще всего они смещаются вниз, то есть занижаются).

5.5. Обнаружение автокорреляции: критерий Дарбина -Уотсона

В данном случае рассматриваем автокорреляцию, подчиняющуюся авторегрессионной схеме первого порядка:

$$u_t = \rho u_{t-1} + \varepsilon_t .$$

Данный тип автокорреляции называется *авторегрессионным*, поскольку u_t определяется значениями этой же самой величины с запаздыванием, с добавлением нового элемента случайности ε_t. Эта схема называется схемой первого порядка AR (1).

Автокорреляция бывает как положительной, так и отрицательной в соответствии со знаком ρ в уравнении.

Наибольшее распространение при тестировании моделей на отсутствие автокорреляции между случайными возмущениями получил тест Дарбина – Уотсона. Он позволяет идентифицировать, как ложную, так и истинную автокорреляцию.

В основе теста лежат следующие предположения:

- Случайные возмущения подчиняются нормальному закону распределения.
- Тип авторегрессии AR (1).

Статистика Дарбина-Уотсона, с помощью которой тестируется модель на автокорреляцию, рассчитывается по величинам отклонений и имеет вид:

$$d = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$$

На больших выборках: $d \to 2 - 2\rho$.

Если автокорреляция отсутствует, то ρ=0, и, таким образом, d должна быть близкой к двум.

Если имеется положительная автокорреляция, то d, в тенденции будет меньше двух.

Если есть отрицательная автокорреляция, то она в тенденции будет больше двух.

При выполнении теста предполагается, что ρ лежит в интервале -1> ρ >1 и, следовательно, d лежит между 4 и 0.

Выдвигается нулевая гипотеза: H_0: $\rho=0$. Даже если нулевая гипотеза истина, то величина d не будет в точности равняться 2.

Проверка гипотезы осуществляется на основе установления критического значения $d_{крит}$, ниже которого d может оказаться с вероятностью 5%. Если $d<d_{крит}$, то гипотеза H_0 отвергается при уровне значимости 5%.

Критическое значение d при любом данном уровне значимости зависит от числа объясняющих переменных в уравнении регрессии и от числа наблюдений в выборке. Но оно также зависит от конкретных значений, принимаемых объясняющими переменными. Поэтому невозможно составить таблицу с указанием точных критических значений для всех возможных выборок, как это можно сделать для t- и F-статистик. Но можно вычислить верхнюю и нижнюю границы для критического значения d (Приложение А. Таблица А.5.).

Для положительной автокорреляции они обычно обозначаются как d_U и d_L. На рис.5.5.1. данная ситуация представлена в виде схемы; стрелка указывает критический уровень d, который обозначается как $d_{крит}$.

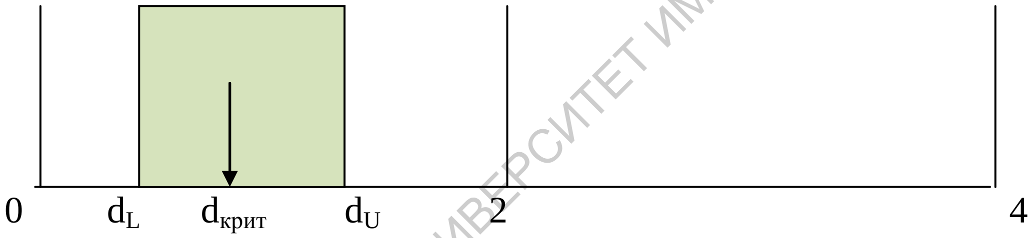


Рис.5.5.1. Тест Дарбина –Уотсона. Зона неопределенности в случае предполагаемой положительной автокорреляции.

Если бы знать значение $d_{крит}$, то можно сравнить с ним значение d, рассчитанное для данной регрессии. Но возможно знать, что $d_{крит}$ лежит где-то между d_L и d_U. Это предполагает три возможных исхода для теста:

1. Величина d меньше, чем d_L. В этом случае она будет также меньше, чем $d_{крит}$. Отвергается нулевая гипотеза и делается вывод о наличии положительной автокорреляции.
2. Величина d больше, чем d_U. В этом случае она также больше критического уровня, и поэтому нельзя отклонить нулевую гипотезу.
3. Величина d находится между d_L и d_U. В этом случае она может быть больше или меньше критического уровня $d_{крит}$. Поскольку нельзя определить, какая из двух возможностей имеет место, невозможно, ни отклонить, ни принять нулевую гипотезу.

В таблице А.5 видно, что чем больше число наблюдений, тем уже зона неопределенности, представленная отрезком между d_L и d_U.

Проверка на отрицательную автокорреляцию проводится по аналогичной схеме, причем зона, содержащая критический уровень, расположена симметрично справа от значения 2. Так как отрицательная автокорреляция встречается относительно редко, предполагается, что можно самостоятельно вычислить границы зоны на основе соответствующих значений для положительной автокорреляции при данном числе наблюдений и объясняющих переменных. Как показано на рис.5.5.2., величина 4-d_U есть нижний предел, ниже которого признается отсутствие автокорреляции, а 4-d_L - верхний предел, выше которого делается вывод о наличии отрицательной автокорреляции.

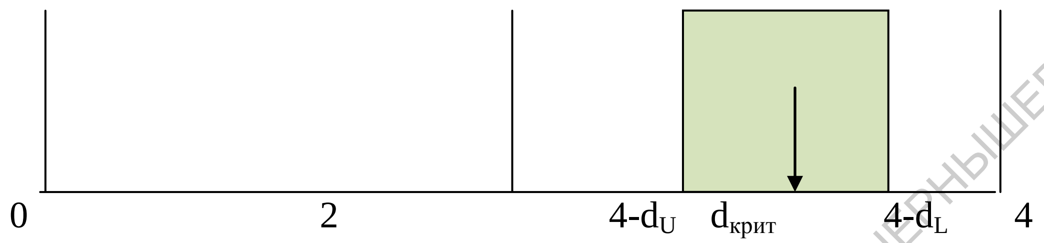


Рис.5.5.2. Тест Дарбина –Уотсона. Зона неопределенности в случае предполагаемой отрицательной автокорреляции.

5.6. Оценивание линейных моделей в условиях автокорреляйции случайных возмущений

Рассмотрим одну из процедур оценивания линейных моделей в условиях автокорреляции случайных возмущений – итеративную процедуру Кокрана - Оркатта для устранения автокорреляции вида AR (1).

Отправной точкой для итеративной процедуры Кокрана – Оркатта является уравнение:

$$\tilde{y}_t = a' + b\tilde{x}_t + \varepsilon_t;$$

где: $\tilde{y}_t = y_t - \rho y_{t-1}$; $\tilde{x}_t = x_t - \rho x_{t-1}$; $a' = a(1-\rho)$.

Если известна величина ρ, то нужно лишь рассчитать $\tilde{y}_t$ и $\tilde{x}_t$ по данным y и x и оценить парную регрессию между $\tilde{y}_t$ и $\tilde{x}_t$. Коэффициент при $\tilde{x}_t$ был бы непосредственной оценкой b, а свободный член мог бы быть использован для расчета оценки a.

Конечно, в действительности значение ρ неизвестно и его нужно оценить наряду с остальными параметрами модели. Итеративная процедура Кокрана – Оркатта позволяет сделать это в предположении, что если случайный член соответствует процессу AR(1), то отклонения от линии регрессии тоже (приближенно) ему соответствуют, и, следовательно, оценивание регрессии между e_t и e_{t-1} даст оценку ρ.

Данная процедура включает следующие шаги:

1. Оценивается регрессия между y и x по первоначальным, не преобразованным данным.
2. Рассчитываются отклонения от линии регрессии.
3. Оценивается регрессия между e_t и e_{t-1} для получения оценки ρ.

4. С использованием оценки ρ рассчитываются $\tilde{y}_t$ и $\tilde{x}_t$ и оценивается исходное уравнение. Коэффициент при $\tilde{x}_t$ обеспечивает новую оценку b, а оценка a' дает новую оценку a.
5. Отклонения пересчитываются, и процесс возвращается к шагу 3.

Процесс переключается между пересмотром оценок a и b и корректировкой оценки ρ до тех пор, пока не будет достигнута устойчивость оценок, то есть до тех пор, пока оценки на последнем цикле не совпадут с оценками предыдущего цикла с точностью до заданного числа десятичных знаков.

Помимо данного метода существуют еще ряд методов преодоления автокорреляции, такие как метод Прайса – Уинстона, процедура Холдрета – Ли, процедура Дарбина.[1]

Вопросы для обсуждения:

1. Каково содержание понятия «гомоскедастичность случайных возмущений»?
2. Каково содержание понятия «гетероскедастичность случайных возмущений»?
3. Какие последствия вызывает гетероскедастичность случайных возмущений?
4. Какие особенности экономического объекта являются причиной возникновения гетероскедастичности?
5. В чем содержание теста проверки на гетероскедастичность Спирмена?
6. В чем содержание теста проверки на гетероскедастичность Глейзера?
7. В чем содержание теста проверки на гетероскедастичность Гольдфельда - Квандта?
8. Что такое весовая функция?
9. В чем идея методики исправления гетероскедастичности случайных возмущений в линейных моделях?
10. Каков порядок действий при оценке параметров линейной модели взвешенным методом наименьших квадратов?
11. Каково содержание автокорреляции случайных возмущений?
12. К каким последствиям приводит автокорреляция случайных возмущений в линейных моделях?
13. Какие причины вызывают автокорреляцию случайных возмущений?
14. Какие предположения лежат в основе теста Дарбина – Уотсона на отсутствие автокорреляции?
15. В чем суть статистики Дарбина – Уотсона?
16. Назовите область возможных значений Дарбина –Уотсона.

[1] Кастюнин В.И. Эконометрика: учебник и практикум для прикладного бакалавриата. М.: Издательство Юрайт, 2015.С.141-150.

17. Опишите алгоритм проверки модели на автокоррелируемость случайных возмущений с помощью теста Дарбина-Уотсона в случае положительной автокорреляции.
18. Опишите алгоритм проверки модели на автокоррелируемость случайных возмущений с помощью теста Дарбина-Уотсона в случае положительной автокорреляции.
19. Опишите алгоритм проверки модели на автокоррелируемость случайных возмущений с помощью теста Дарбина-Уотсона в случае отрицательной автокорреляции.
20. В чем заключается метод исправления автокорреляции случайных возмущений?
21. Опишите итерационную процедуру Кохрейна – Оркатта нахождения состоятельных оценок параметров линеной регрессии в условиях автокорреляции.

Тест:

1 Гомоскедастичность – это:

a. условие, при котором вероятность того, что величина u примет какое-то данное положительное или отрицательное значение, будет одинаковой во всех наблюдений;
b. условие, при котором вероятность того, что величина u примет какое-то данное положительное или отрицательное значение, будет неодинаковой во всех наблюдений.

2 Математически гетероскедастичность может определяться следующим образом:

a. $\sigma_{u_i}^2 = \sigma_u^2$, величина одинакова для всех наблюдений;
b. $\sigma_{u_i}^2$ не одинакова для всех наблюдений.

3 В тесте ранговой корреляции Спирмена рассчитывают:

a. $r_{x_e} = 1 - \dfrac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)}$;
b. $RSS_2/\ RSS_{1;}$
c. $|e| = a + bx^{\gamma}$.

4 В тесте тесте Глейзера рассчитывают:

a. $r_{x_e} = 1 - \dfrac{6\sum_{i=1}^{n} D_i^2}{n(n^2-1)}$;
b. $RSS_2/\ RSS_{1;}$
c. $|e| = a + bx^{\gamma}$.

5 При проведении проверки по критерию Гольдфельда - Квандта предполагается:

a. что стандартное отклонение распределения вероятностей случайного члена в наблюдении i пропорционально значению x_i;
b. что стандартное отклонение распределения вероятностей случайного члена в наблюдении i не пропорционально значению x_i.

6 Голдфелд и Квандт утверждали:
a. что n’ должно составлять порядка 1/8 от n;
b. что n’ должно составлять порядка 5/8 от n;
c. что n’ должно составлять порядка 3/8 от n.

7 Случайный член подвержен автокорреляции, когда:
a. теоретические ковариации u_i и u_j равны нулю для любых $i \neq j$;
b. теоретические ковариации u_i и u_j не равны нулю для любых $i \neq j$;
c. $\sigma_{u_i}^2 = \sigma_u^2$, величина одинакова для всех наблюдений;

8 При положительной автокорреляции наблюдается:
a. постоянная направленность воздействия невключенных в уравнение переменных;
b. за положительным значением в одном наблюдении идет отрицательное значение в следующем.

9 Если автокорреляция отсутствует, то:
a. ρ=0, и, таким образом, d должна быть близкой к двум;
b. d, в тенденции будет меньше двух;
c. d в тенденции будет больше двух.

10 Если имеется положительная автокорреляция, то:
a. ρ=0, и, таким образом, d должна быть близкой к двум;
b. d, в тенденции будет меньше двух;
c. d в тенденции будет больше двух.

11 Если есть отрицательная автокорреляция, то:
a. ρ=0, и, таким образом, d должна быть близкой к двум;
b. d, в тенденции будет меньше двух;
c. d в тенденции будет больше двух.

12 Отвергается нулевая гипотеза и делается вывод о наличии положительной автокорреляции, если:
a. величина d меньше, чем $d_{L;}$
b. величина d больше, чем $d_{U;}$
c. величина d находится между d_L и d_U.

Задания для самостоятельной работы:

№1.

Построить и протестировать на отсутствие гетероскедастичности и автокорреляцию модель «государственные расходы на образование в зависимости от объема ВВП».

№	Страна	Гос. расходы на образование	ВВП	№	Страна	Гос. расходы на образование	ВВП
1	Люксембург	0,34	5,67	18	Турция	1,6	66,97
2	Уругвай	0,22	10,13	19	Саудовская Аравия	6,4	115,97
3	Сингапур	0,32	11,34	20	Бельгия	7,15	119,49
4	Ирландия	1,23	18,88	21	Швеция	11,22	124,15
5	Израиль	1,81	20,94	22	Австралия	8,66	1490,98
6	Новая Зеландия	1,27	23,83	23	Аргентина	5,56	153,85
7	Гонконг	0,67	27,56	24	Нидерланды	13,41	169,38
8	Венгрия	1,02	22,16	25	Испания	4,79	211,78
9	Португалия	1,07	24,67	26	Мексика	5,46	186,33
10	Чили	1,25	27,57	27	Канада	18,9	261,41
11	Греция	0,75	40,15	28	Бразилия	8,92	249,72
12	Финляндия	2,8	51,62	29	Италия	15,95	395,52
13	Норвегия	4,9	57,71	30	Великобритания	29.9	534,97
14	Дания	4,45	66,32	31	Франция	33,59	655,29
15	Австрия	4,26	76,88	32	германия	38,62	815,00
16	Югославия	3,5	63,03	33	Япония	61,61	1040,45
17	Швейцапия	5,31	101,65	34	США	181,30	2586,40

№2.

Оценить и проанализировать на присутствие гетероскедастичности и автокорреляцию модель зависимости расходов на жилье в зависимости от распологаемого дохода и индекса цен на жилье.

№	Расходы на жилье	Доход	Индекс цен	№	Расходы на жилье	Доход	Индекс цен
1	60,9	479,7	104,5	14	124,2	858,4	95,1
2	64	489,7	104,5	15	118,2	865,3	99,1
3	67	503,8	105,1	16	128,3	875,8	93,3
4	70,7	524,9	105	17	89,1	873,5	102,2
5	74	542,3	104,8	18	134,9	906,8	93,7
6	77,4	580,8	104,5	19	141,3	942,9	94,5

7	81,6	616,3	104	20	148,5	988,8	94,7
8	85,3	646,8	102,6	21	154,8	1015,5	93,8
9	93,5	701,3	100,9	22	159,8	1021,6	93
10	98,4	722,5	100	23	164,8	1049,3	94,2
11	102	751,6	99,6	24	167,5	1058,3	96,7
12	106,4	779,2	100	25	171,3	1095,4	99,7
13	112,5	810,3	100				

№3.

По данным таблицы проверьте модель на гомоскедастичность и автокорреляцию. В случае необходимости оцените параметры модели с помощью процедуры Кохрейна – Оркатта.

№	x_1	x_2	x_3	y
1	7,59	12,2	14,32	25,73
2	6,32	12,67	9,07	11,97
3	-0,45	8,49	15,66	10,13
4	8,14	18,69	14,85	16,92
5	4,05	18,36	14,69	7,22
6	6,7	14,72	7,83	9,03
7	7,25	2,79	11,3	34,29
8	3,53	8,74	3,78	2,07
9	7,64	8,55	13,24	29,79
10	13,58	8,21	11,69	42,99
11	2,33	14,2	1,92	-11,79
12	9,96	6,25	7,19	27,16
13	0,29	7,25	14,79	15,55
14	11,47	20,44	13,95	18,73
15	2,17	10,35	-0,48	-9,6
16	13,92	7,86	6,17	33,43
17	5,59	7,69	9,9	17,42
18	13,72	3,31	15,38	55,41
19	6,39	15,25	13,77	15,93
20	6,26	2,85	13,95	34,31
21	7,54	3,58	6,57	26,25
22	2,6	16,51	11,45	2,45
23	0,38	3,88	6,81	9,66
24	4,03	3,82	12,86	23,37
25	9,24	7,84	0,05	12,7
26	3,17	3,01	2,5	11,88
27	2,42	16,32	16,34	6,28
28	-1,04	5,3	0,99	-4,83
29	5,93	12,99	12,59	16,46
30	13,89	17,96	11,88	26,18

ТЕМА 6. МОДЕЛИРОВАНИЕ ОДНОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ

Ключевые слова:

Временной ряд. Трендовая, циклическая и случайные компоненты. Аддитивная модель. Мультипликативная модель. Автокорреляция. Лаг. Автокорреляционная функция. Коррелограмма. Структура временного ряда. Аналитическое выравнивание временного ряда. Метод скользящей средней. Структурные изменения. Кусочно-линейная модель регрессии. Тест Чоу.

Основные теоретические аспекты темы:

6.1. Понятие временного ряда и его компонент

Эконометрическую модель можно построить, используя два типа исходных данных:

- данные, характеризующие совокупность различных объектов в определенны момент (период) времени;
- данные, характеризующие один объект за ряд последовательных моментов (периодов) времени.

Модели, построенные по данным первого типа, называются *пространственными моделями*.

Модели, построенные по данным второго типа, называются моделями *временных рядов.*

Временной ряд – совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени.

Уровень ряда – каждое наблюдение значения временного ряда, который формируется под действием большого числа факторов, которые можно разделить на три группы:

- факторы, формирующие тенденцию ряда;
- факторы, формирующие циклические колебания;
- случайные факторы.

При различных сочетаниях этих факторов зависимость уровней ряда от времени может принимать разные формы.

В большинстве случаев фактический уровень временного ряда можно представить как сумму или произведение треднговой, циклической и случайной компонент.

Аддитивная модель – модель вида: Y=T+S+E;

Мультипликативная модель – модель вида: Y=T*S*E;

где T - трендовая компонента;

S – циклическая компонента;

E – случайная компонента.

Аддитивная модель, как правило, используется в случаях, когда амплитуда циклической составляющей не зависит от времени. Мультипликативная модель хорошо описывает процессы, когда амплитуда циклической составляющей с ходом времени изменяется в том же направлении, что и тренд.

Присутствие в моделях первых двух составляющих не обязательно: модель может иметь тренд, но не иметь циклической составляющей и наоборот. Но наличие случайной составляющей является обязательным.

Основная задача эконометрического исследования отдельного временного ряда – выявление и придание количественного выражения каждой из компонент, с тем, чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

6.2. Автокорреляция временного ряда и выявление его структуры

При наличии тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих значений.

Автокорреляция временного ряда – корреляционная зависимость между последовательными уровнями временного ряда. Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов времени.

Лаг – число периодов, по которым рассчитывается коэффициент автокорреляции.

Коэффициент автокорреляции уровней ряда первого порядка – коэффициент корреляции, измеряющий зависимость между соседними уровнями ряда y_t и y_{t-1}, то есть при лаге 1:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \overline{y_1}) * \left(y_{t-1} - \overline{y_2}\right)}{\sqrt{\sum_{t=2}^{n}\left(y_t - \overline{y_1}\right)^2 * \sum_{t=2}^{n}\left(y_{t-1} - \overline{y_2}\right)^2}}; \quad \overline{y_1} = \frac{\sum_{t=2}^{n} y_t}{n-1}; \quad \overline{y_2} = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1}.$$

Соответственно, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями ряда y_t и y_{t-2}.

С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается. Считается, что для обеспечения статистической достоверности коэффициентов автокорреляции использовать правило «максимальный лаг должен быть не больше n/4».

Свойства коэффициента автокорреляции:

1. Коэффициент автокорреляции строится по аналогии с линейным коэффициентом корреляции, и, таким образом, характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. По коэффициенту автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции. Для рядов, имеющих сильную нелинейную тенденцию (параболу, экспоненту), коэффициент автокорреляции может приближаться к нулю.

2. По знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда.

Автокорреляционная функция временного ряда – последовательность коэффициентов автокорреляции уровней временного ряда.

Коррелограмма – график зависимости значений автокорреляционной функции от величины лага.

При помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда, а именно определить лаг, при котором автокорреляция наиболее высокая, следовательно, лаг, при котором связь между текущими и последующими уровнями ряда наиболее тесная.

Анализ структуры ряда:

- Если коэффициент автокорреляции первого порядка (r_1) наиболее высокий, то исследуемый ряд содержит только тенденцию;
- Если наиболее высоким оказался коэффицент автокорреляции порядка τ, то ряд содержит циклические колебания с периодичностью в τ моментов времени;
- Если ни один из коэффициентов автокорреляции не является значимым, то либо ряд не содержит тенденции и циклических колебаний и содержит только случайную компоненту, либо ряд содержит сильную нелинейную тенденцию.

Например, если при анализе временного ряда наиболее высоким оказался коэффициент автокорреляции второго порядка, то ряд содержит циклические колебания с циклом, равным двум периодам времени, то есть имеет пилообразную структуру.

6.3. Моделирование тенденции временного ряда и случайной компоненты

Аналитическое выравнивание временного ряда – способ моделирования тенденции временного ряда: построение аналитической функции, характеризующей зависимость уровней ряда от времени, или тренда.

Для построения трендов чаще всего применяются следующие функции:

- линейный тренд: $y_t=a+b*t$;
- гипербола $y_t=a+b/t$;
- степенная функция $y=a*t^b$;
- экспоненциальная функция $y_t=a*b^t$;
- параболы разных порядков$y= a_0+a_1 t +a_2 t^2 +, . . . , +a_m t^{m.}$.

*Способы определения типа тенденции***:**

- Качественный анализ.
- Визуальный анализ графика.
- Коэффициенты автокорреляции: Если временной ряд имеет линейную тенденции., то его соседние уровни y_t и y_{t-1} тесно коррелируют. Коэффициент автокорреляции первого порядка 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.

- Перебор основных форм тренда осуществляется с помощью расчета скорректированного коэффициента детерминации R^2.

.

6.4. Моделирование сезонных и циклических колебаний: метод скользящей средней

Существуют несколько подходов к анализу структуры временных рядов, содержащих сезонные или циклические колебания. Простейший подход – расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда. Построение аддитивной и мультипликативной моделей сводится к расчету значений T, S и E для каждого уровня ряда.

Алгоритм метода скользящей средней

<u>Шаг1. Выравнивание исходных уровней ряда методом скользящей средней:</u>

1. Суммируем уровни ряда последовательно за каждый промежуток времени, в котором наблюдаются колебания со сдвигом на один момент времени и определяем условные величины показателя Y.
2. Делим полученные величины на число моментов времени в промежутке и находим скользящие средние.
3. Находим средние значения из двух последовательных скользящих

<u>Шаг 2. Оценка сезонной компоненты:</u>

1. Находим оценку сезонной компоненты, как разность между фактическими уровнями ряда и центрированными скользящими средними. В мультипликативной модели – как частное от деления фактических уровней ряда на центрированные скользящие средние.
2. Находим средние оценки сезонной компоненты за каждый промежуток времени, в котором наблюдаются колебания $\overline{S_i}$.
3. Исходя из условия взаимопогашения сезонных воздействий определяем корректирующий коэффициент k.
   - В аддитивной модели сумма значений сезонной компоненты по всем кварталам должна быть равна нулю, тогда $k = \frac{\sum S_i}{n}$.
   - В мультипликативной модели сумма значений сезонной компоненты по всем кварталам должна быть равна числу

периодов в цикле, т.е. четырем в случае четырех кварталов: $k = \frac{n}{\sum S_i}$; где n – период колебаний.

4. Рассчитываем скорректированные значения сезонных компонент: в аддитивной модели: $S_i = \overline{S_i}$ - k; в мультипликативной модели: $S_i = \overline{S_i}$ *k;

Шаг 3. Элиминирование влияния сезонной компоненты:

- Находим значения T+E как Y-S – в аддитивной модели.
- Находим значения T*E как Y/S – в мультипликативной модели.

Шаг 4. Определение трендовой компоненты ряда.

1. Трендовая компонента ряда определяется с помощью построения регрессионной модели, параметры которой находятся методом наименьших квадратов.
2. С помощью уравнения регрессии находим теоретические уровни трендовой компоненты T для каждого момента времени t.

Шаг 5. Находим значения T+S в аддитивной модели или T*S в мультипликативной модели.

Шаг 6. Находим случайную компоненту E= Y-(T+S) в аддитивной модели и E= Y/(T*S) в мультипликативной модели

Шаг 7. Оценка качества модели.

1. Находим сумму квадратов случайной компоненты.
2. Находим отношение суммы квадратов случайной компоненты к общей сумме квадратов отклонений уровней ряда от его среднего значения: $\frac{\sum E^2}{\sum (y_t - \overline{y})^2} * 100\%$

6.5. Применение фиктивных переменных для моделирования сезонных колебаний

Еще одним методом моделирования временного ряда, содержащего сезонные колебания, является построение модели множественной регрессии с включением фактора времени и фиктивных переменных.

Количество фиктивных переменных в такой модели должно быть на единицу меньше числа моментов (периодов) времени внутри одного цикла колебаний. Так, при моделировании поквартальных данных модель должна включать четыре независимые переменные – фактор времени и три фиктивные переменные. Каждая фиктивная переменная отражает сезонную компоненту временного ряда для какого-либо одного периода. Она равна единице для данного периода и нулю для всех остальных периодов.

Таким образом, модель примет вид:

$$y = a + bt + c_1 d_1 + c_2 d_2 + c_3 d_3 + \varepsilon;$$

где: a, b, c_1, c_2, c_3 - параметры модели;

d_1, d_2, d_3 - фиктивные переменные;

$$d_1 = \begin{cases} 1, для\_первого\_квартала \\ 0, в\_остальных\_случаях \end{cases};$$

$$d_2 = \begin{cases} 1, для\_второго\_квартала \\ 0, в\_остальных\_случаях \end{cases};$$

$$d_3 = \begin{cases} 1, для\_третьего\_квартала \\ 0, в\_остальных\_случаях \end{cases}.$$

Параметр b в этой модели характеризует среднее абсолютное изменение уровней ряда под воздействием тенденции.

Данная модель есть аналог аддитивной модели временного ряда, так как фактический уровень временного ряда – это сумма трендовой, сезонной и случайной компонент.

Недостатком модели с фиктивными переменными для описания сезонных и циклических колебаний является наличие большого количества переменных, что требует большого массива данных для исследования.

6.6. Моделирование тенденции временного ряда при наличии структурных изменений

От сезонных и циклических колебаний следует отличать единовременные изменения характера тенденции временного ряда, вызванные структурными изменениями в экономике или иными факторами. В этом случае, начиная с некоторого момента времени t^*, происходит изменение характера динамики изучаемого показателя, что приводит к изменению параметров тренда, описывающего эту динамику.

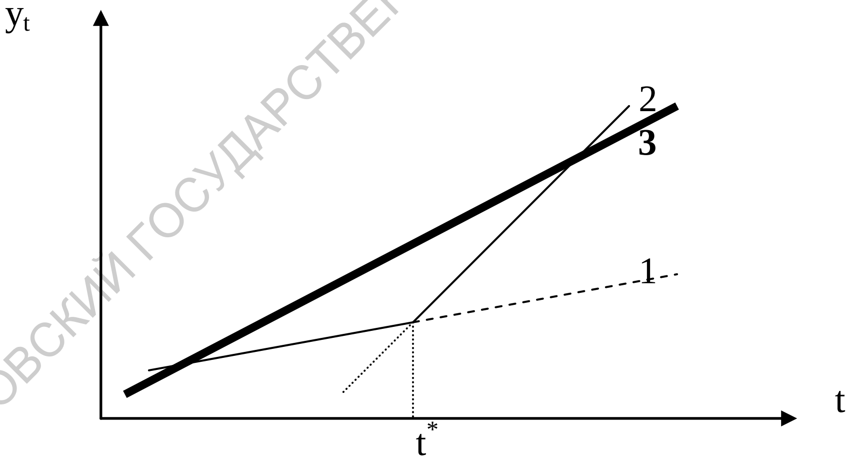


Рис. 6.6.1. Изменение характера тенденции временного ряда.

Чаще всего эти изменения вызваны факторами глобального характера (начало крупных экономических реформ, ресурсные кризисы, скачки валютных курсов, политические события).

Если временной ряд включает в себя соответствующий момент (период) времени, то одной из задач его изучения становится выяснение вопроса о том, значимо ли повлияли общие структурные изменения на характер этой тенденции.

Если структурное изменение значимы, то для моделирования тенденции данного временного ряда следует использовать кусочно – линейные модели регрессии, то есть разделить исходную совокупность на две подсовокупности (до момента времени t^* и после момента времени t^*) и построить отдельно по каждой подсовокупности уравнения линейной регрессии (на рис.6.6.1. это линии 1и 2).

Если структурные изменения незначительно повлияли на характер тенденции ряда, то ее можно описать с помощью единого для всей совокупности данных уравнения тренда (на рис.6.6.1. это линии 3).

Каждый подход имеет свои положительные и отрицательные стороны:

Последствия при построении кусочно - линейной модели:

«+» - снижение остаточной суммы квадратов по сравнению с единым для всей совокупности уравнением тренда;

«-» - потеря числа наблюдений и, следовательно, снижение числа степеней свободы в каждом уравнении кусочно – линейной модели.

Последствия при построении единой модели:

«+» - сохранение числа наблюдений исходной совокупности;

«-» - остаточная сумма квадратов будет выше по сравнению с кусочно – линейной моделью.

Выбор одной из двух моделей (кусочно – линейной или единого уравнения тренда) будет зависеть от соотношения между снижением остаточной дисперсии и потерей числа степеней свободы при переходе от единого уравнения регрессии к кусочно – линейной модели.

Формальный статистический тест для оценки этого соотношения был предложен Грегори Чоу в 1960 г .

Тест Чоу: Анализ значимости структурных изменений:

Выдвигается гипотеза H_0: о структурной стабильности тенденции изучаемого временного ряда и рассчитываются параметры по кусочно – линейной модели и по единой модели ряда.

Таблица 6.6.1.

Расчет параметров по кусочно –линейной модели и по единой модели ряда

Номер уравнения	Вид уравнения	Число наблюдений в совокупности	Остаточная сумма квадратов	Число параметров в уравнении	Число степеней свободы остаточной дисперсии
Кусочно – линейная модель					
1	$Y_1=a_1+b_1t$	n_1	RSS_1	k_1 (2)	n_1-k_1
2	$Y_2=a_2+b_2t$	n_2	RSS_2	k_2 (2)	n_2-k_2
Уравнение тренда по всей совокупности					
3	$Y_3=a_3+b_3t$	n	RSS_3	k_3(2)	n-k_3

Остаточная сумма квадратов по кусочно – линейной модели:

$$RSS_{кл} = RSS_1 + RSS_2.$$

Соответствующее ей число степеней свободы составит:

$$(n - k_1 - k_2) = (n_1 - k_1) + (n_2 - k_2).$$

Сокращение остаточной дисперсии при переходе от единого уравнения тренда к кусочно – линейной модели:

$$\Delta RSS = RSS_3 - RSSкл.$$

Число степеней свободы, соответствующее ΔRSS:

$$(n - k_3) - (n - k_1 - k_2) = (k_1 + k_2 - k_3)$$

В соответствии с предложенной Г.Чоу методикой определяется фактическое значение F-критерия по следующим дисперсиям на одну степень свободы:

$$F - критерий = \frac{\Delta RSS / (k_1 + k_2 + k_3)}{RSSкл / (n - k_1 - k_2)}.$$

Найденное значение F – критерия сравнивают с табличным, полученным по таблицам распределения Фишера для определенного уровня значимости и числа степеней свободы $(k_1 + k_2 - k_3)$ и $(n - k_1 - k_2)$.

Если Fфакт > Fтабл, то гипотеза о структурной стабильности тенденции отклоняется, а влияние структурных изменений на динамику изучаемого показателя признают значимым. В этом случае моделирование тенденции временного ряда следует осуществлять с помощью кусочно – линейной модели.

Причины структурных изменений обуславливают различия в оценках параметров уравнений (1) и (2):

- Изменение численной оценки a_1 и a_2 при условии, что различия между b_1 и b_2 статистически незначимы. Геометрически это означает, что прямые параллельны. В данной ситуации можно говорить о скачкообразном изменении уровней ряда в момент времени t^* при неизменном среднем абсолютном приросте за период;
- Изменение численной оценки b_1 и b_2 при условии, что различия между a_1 и a_2 статистически незначимы. Геометрически это означает, что прямые пересекают ось ординат в одной точке. В данной ситуации изменение тенденции связано с изменением среднего абсолютного прироста временного ряда t^*.
- Изменение численной оценки a_1 и a_2 и b_1 и b_2 . Геометрически эта ситуация означает, что изменение характера тенденции сопровождается изменением как начального уровня ряда, так и среднего за период абсолютного прироста.

6.7. Пример построения аддитивной модели временного ряда с помощью пакета Excel

Задание

Собраны статистические данные потребления электроэнергии за 4 года.

№ квартала t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
y_t	6	4,4	5	9	7,2	4,8	6	10	8	5,6	6,4	11	9	6,6	7	10,8

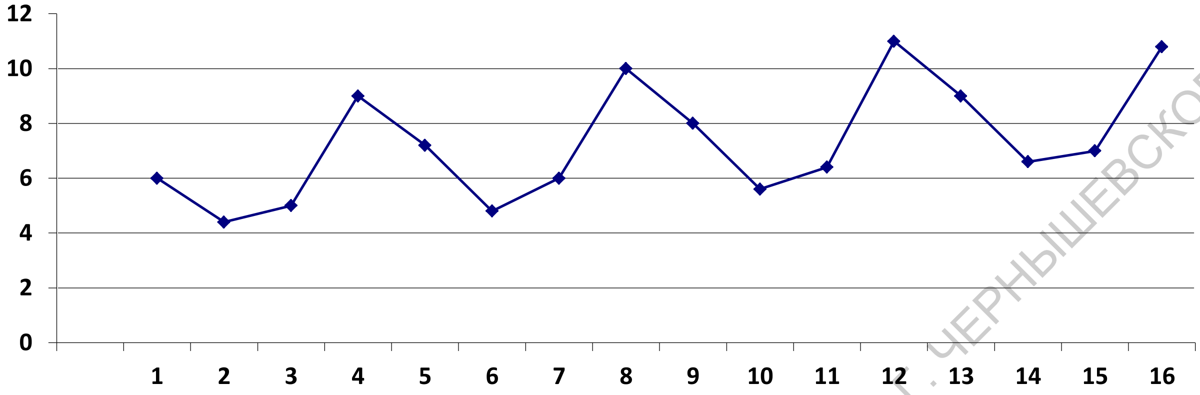


Рис. 6.7.1. Потребление электроэнергии за 4 года

Необходимо:

1. Построить коррелограмму временного ряда.
2. Найти значения трендовой, сезонной и случайной компонент.
3. Применить фиктивные переменные для моделирования сезонных колебаний.

Решение.

1. Коррелограмма временного ряда потребления электроэнергии:

Лаг	Коэффициент автокорреляции уровней
1	0,165
2	0,5666
3	0,113
4	0,983
5	0,1187
6	0,722
7	0,003
8	0,97

2. Метод скользящей средней

Шаг 1. Выравнивание исходных уровней ряда методом скользящей средней

Таблица 1.

Выравнивание исходных уровней ряда методом скользящей средней

№ квартала	Yt	Итого за 4 квартала	Скользя-щая средняя за 4 квартала	Центриро-ванная скользя-щая средняя	Оценка сезонной компонен-ты
1	6				
2	4,4	24,4	6,1		
3	5	25,6	6,4	6,25	-1,25
4	9	26	6,5	6,45	2,55
5	7,2	27	6,75	6,625	0,575
6	4,8	28	7	6,875	-2,075
7	6	28,8	7,2	7,1	-1,1
8	10	29,6	7,4	7,3	2,7
9	8	30	7,5	7,45	0,55
10	5,6	31	7,75	7,625	-2,025
11	6,4	32	8	7,875	-1,475
12	11	33	8,25	8,125	2,875
13	9	33,6	8,4	8,325	0,675
14	6,6	33,4	8,35	8,375	-1,775
15	7	24,4			
16	10,8	17,8			

Шаг 2. Оценка сезонной компоненты:

Таблица 2.

Оценка сезонной компоненты

Показатели	Год	№ квартала i			
		1	2	3	4
	1	-	-	-1,25	2,55
	2	0,575	-2,075	-1,1	2,7
	3	0,55	-2,025	-1,475	2,875
	4	0,675	-1,775	-	-
Итого за i квартал		1,8	-5,875	-3,825	8,125
Средняя оценка сезонной компоненты для i – го квартала		0,6	-1,958	-1,275	2,708
Скорректированная сезонная компонента Si		0,58125	-1,978	-1,294	2,69

Корректирующий коэффициент $k = \frac{0{,}6 - 1{,}958 - 1{,}275 + 2{,}71}{4} = 0{,}0185$

Шаг 3. Элиминирование влияния сезонной компоненты: (Y-S) (столбец.4 в таблице 3).

Шаг 4. Определение трендовой компоненты ряда. (столбец 5 в таблице 3).

Шаг 6. Находим значения T+S (столбец 6 в таблице 3)

Шаг 7. Находим случайную компоненту E= Y-(T+S) (столбец 7 в таблице 3).

Таблица 3.

Нахождение трендовой и случайной компонент

№ t	Y_t	S_i	T+E=Y-S	T	T+S	E=Y-(T+S)	E^2
1	6	0,581	5,419	5,9019	6,4829	-0,4829	0,2332
2	4,4	-1,977	6,377	6,0883	4,1113	0,2887	0,0833
3	5	-1,294	6,294	6,2747	4,9807	0,0193	0,0004
4	9	2,69	6,31	6,4611	9,1511	-0,1511	0,0228
5	7,2	0,581	6,619	6,6476	7,2286	-0,0286	0,0008
6	4,8	-1,977	6,777	6,834	4,857	-0,057	0,0032
7	6	-1,294	7,294	7,0204	5,7264	0,2736	0,0749
8	10	2,69	7,31	7,2068	9,8968	0,1032	0,0107
9	8	0,581	7,419	7,3932	7,9742	0,0258	0,0007
10	5,6	-1,977	7,577	7,5796	5,6026	-0,0026	7E-06
11	6,4	-1,294	7,694	7,766	6,472	-0,072	0,0052
12	11	2,69	8,31	7,9524	10,642	0,3576	0,1279
13	9	0,581	8,419	8,1388	8,7198	0,2802	0,0785
14	6,6	-1,977	8,577	8,3252	6,3482	0,2518	0,0634
15	7	-1,294	8,294	8,5117	7,2177	-0,2176	0,0474
16	10,8	2,69	8,11	8,6981	11,388	-0,5881	0,3458

Нахождение модели Тренда T=a+b*t 1,098126

0,186412	5,7155	
0,015189	0,146868	T=5,7155+0,18641*t
0,914959	0,280067	
150,6265	14	
11,81478	1,098126	

Сумма квадратов абсолютных ошибок = 1,0981

Шаг 8. Оценка качества модели.

Сумма квадратов абсолютных ошибок = 1,0981

Отношение суммы квадратов случайной компоненты к общей сумме квадратов отклонений уровней ряда от его среднего значения:

$$\frac{\sum E^2}{\sum (y_t - \overline{y})^2} * 100\% = 1{,}5\%.$$

Вывод: Построенная аддитивная модель объясняет 98,5% общей вариации уровней временного ряда потребления электроэнергии за 16 кварталов исследуемых 4 – х лет и ее можно использовать в прогнозах будущего потребления электроэнергии**.**

Вопросы для обсуждения:

1. Объясните, почему временной ряд представляет собой совокупность трендовой, циклической и случайной компоненты?
2. Какой вид связи между соседними уровнями ряда характеризует коэффициент автокорреляции?

3. В чем сходство и различие коэффициента корреляции в регрессионном анализе и коэффициента автокорреляции?
4. Объясните, что представляет собой структура временного ряда? Какой анализ позволяет ее определять?
5. Как регрессионный анализ применяется в моделировании одномерных временных рядов?
6. Какой критерий лежит при выборе построения аддитивной или мультипликативной модели временного ряда?
7. Назовите положительные и отрицательные моменты в построении кусочно-линейных и единого уравнения тренда при наличии структурных изменений в динамике переменных.
8. Каков критерий выбора построения модели временного ряда при наличии структурных изменений в динамике переменных?

Тест:

1 Модели временных рядов строятся на базе:
   a. данных, характеризующих совокупность различных объектов в определенны момент (период) времени;
   b. данных, характеризующих один объект за ряд последовательных моментов (периодов) времени.

2 В случае неравномерной амплитуды колебаний необходимо строить:
   a. аддитивную модель временного ряда;
   b. мультипликативную модель временного ряда;
   c. не имеет значения.

3 Высокий коэффициент автокорреляции первого порядка свидетельствует о наличии:
   a. сезонности;
   b. линейной связи;
   c. нелинейной связи.

4 С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции:
   a. уменьшается;
   b. увеличивается;
   c. не зависит от числа пар значений ряда.

5 Для обеспечения статистической достоверности коэффициентов автокорреляции использовать правило:
   a. «максимальный лаг должен быть не больше n/3»;
   b. «максимальный лаг должен быть не больше n/5»;
   c. «максимальный лаг должен быть не больше n/4».

6 По коэффициенту автокорреляции можно судить о наличии:
   a. линейной (или близкой к линейной) тенденции;
   b. нелинейной тенденции;
   c. линейной и нелинейной тенденции.

7 По знаку коэффициента автокорреляции:
   a. можно делать вывод о возрастающей или убывающей тенденции в уровнях ряда;
   b. нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда.

8 Коррелограмма - это:
   a. последовательность коэффициентов автокорреляции уровней временного ряда;
   b. график зависимости значений автокорреляционной функции от величины лага.

9 В аддитивной модели:
   a. сумма значений сезонной компоненты по всем кварталам должна быть равна нулю;
   b. сумма значений сезонной компоненты по всем кварталам должна быть равна числу периодов в цикле.

10 В мультипликативной модели оценка сезонной компоненты, находится как:
   a. как разность между фактическими уровнями ряда и центрированными скользящими средними;
   b. как частное от деления фактических уровней ряда на центрированные скользящие средние.

11 Для отражения сезонности для четырех времен года необходимо использовать:
   a. две фиктивных переменных;
   b. три фиктивных переменных;
   c. четыре фиктивных переменных.

12 Положительным моментом построения кусочно – линейной модели является:
   a. снижение остаточной суммы квадратов по сравнению с единым для всей совокупности уравнением тренда;
   b. потеря числа наблюдений и, следовательно, снижение числа степеней свободы в каждом уравнении кусочно – линейной модели;
   c. сохранение числа наблюдений исходной совокупности.
   d. остаточная сумма квадратов будет выше по сравнению с единой

моделью.

13 В случае структурных изменений изменение численной оценки a_1 и a_2 при условии, что различия между b_1 и b_2 статистически незначимы геометрически означает:

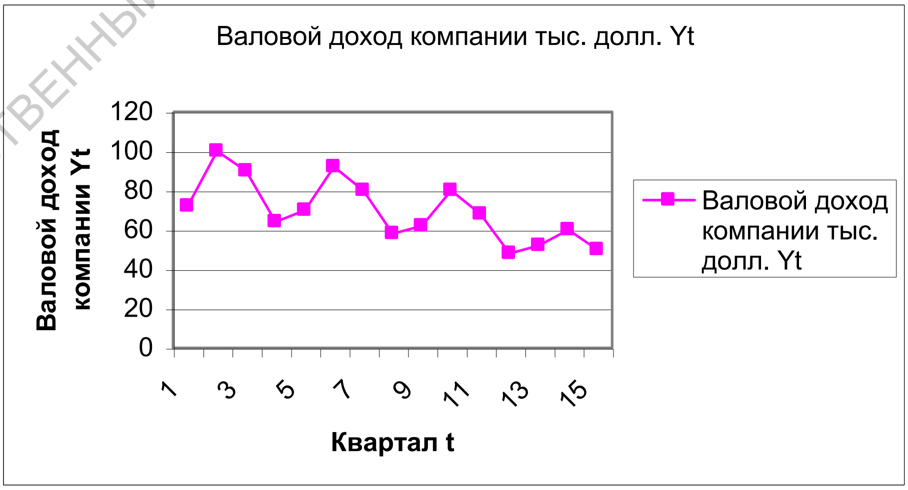
a. что прямые параллельны;
b. что прямые пересекают ось ординат в одной точке;
c. что изменение характера тенденции сопровождается изменением как начального уровня ряда, так и среднего за период абсолютного прироста.

Задания для самостоятельной работы:

№ 1 .

По статистическим данными постройте модель временного ряда. С помощью коэффициентов автокорреляции определите ее структуру и тип модели. Спрогнозируйте с помощью модели валовой доход в 5-м году работы компании. С помощью фиктивных переменных вычислить сезонную, трендовую и случайную компоненту аддитивной модели.

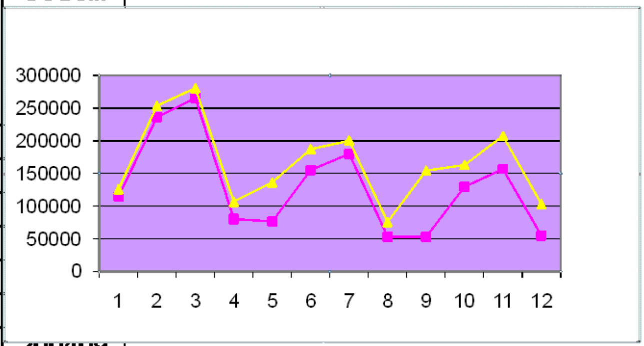
№ квартала t	Валовой доход компании тыс. долл. Yt
1	72
2	100
3	90
4	64
5	70
6	92
7	80
8	58
9	62
10	80
11	68
12	48
13	52
14	60
15	50
16	30



№2

По статистическим данным, описывающим поквартальное производство масла и объем его продаж на внутреннем рынке за 2 гола постройте модели временных рядов и спрогнозируйте по ним величины производства и объем продаж в следующие 3 года. С помощью фиктивных переменных вычислить сезонную, трендовую и случайную компоненту аддитивной модели

№ квартала t	Прода-жи на внутреннем рынке, тонн.	Объем
1	115007	[illegible]
2	236426	[illegible]
3	265466	[illegible]
4	80249	[illegible]
5	76681	[illegible]
6	155130	[illegible]
7	180000	200409
8	53461	76059
9	52845	154753
10	129759	163209
11	156879	207323
12	55145	103566



№3.

В таблице приведены данные динамики изменения индекса цен на продовольственные товары.

Год/квартал	2005	2006	2007	2008	2009	2010	2011	2012
I	104,8	105,5	102,8	105,5	104,4	103,1	105,7	102,3
II	104,5	102,3	102,8	105,9	103,1	102,1	101,3	101,9
III	100,2	100,5	102,8	101,6	100,2	102,1	98,1	102,1
IV	101,4	100,2	105,7	103,2	99	103,7	100,2	101

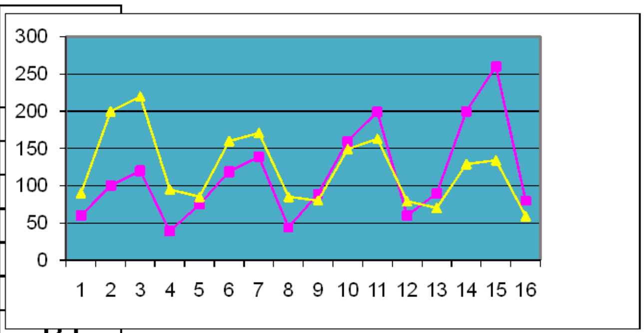
Необходимо:

1. Построить график уровней ряда.
2. С помощью коэффициентов автокорреляции определите ее структуру и тип модели.
3. Построить модель временного ряда.
4. С помощью фиктивных переменных вычислить сезонную, трендовую и случайную компоненту аддитивной модели

№4.

По статистическим данным, описывающим объем спроса на прохладительные напитки двух фирм в течение 4-х лет, постройте модели временных рядов, описывающих динамику спроса обеих фирм. С помощью фиктивных переменных вычислить сезонную, трендовую и случайную компоненту аддитивной модели. Спрогнозируйте квартал, когда одна из фирм покинет рынок. Каков будет объем спроса в этот момент у фирмы-конкурента?

№ квартала t	Спрос Y1t шт.	
1	60	
2	100	
3	120	
4	39	
5	75	
6	119	
7	139	
8	44	85
9	89	80
10	160	149
11	199	163
12	60	79
13	90	70
14	200	129
15	260	134
16	80	59



№5.

По данными постройте модель временного ряда продаж компании, млн. долл. Определите ее структуру. Спрогнозируйте с помощью модели продажи в 2007 - м году.

Год	Квартал	Продажи	Год	Квартал	Продажи
1999	1	2292	2003	1	2643
	2	2450		2	2811
	3	2363		3	2679
	4	2477		4	2736
2000	1	2063	2004	1	2692
	2	2358		2	2871
	3	2316		3	2900
	4	2366		4	2811
2001	1	2268	2005	1	2497
	2	2533		2	2792
	3	2479		3	2838
	4	2625		4	2780
2002	1	2616	2006	1	2778
	2	2793		2	3066
	3	2656		3	3213
	4	2746		4	2928

№6.

Даны поквартальные данные о прибыли компании за последние четыре года, тыс. долл. Постройте мультипликативную модель временного ряда. Спрогнозируйте прибыль компании во втором квартале шестого года работы компании.

	Год 1				Год 2				Год 3				Год 4			
Год	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Прибыль	72	70	62	52	100	92	80	60	90	80	68	50	64	58	48	30

№7.

Даны данные о квартальных продажах фирмы, тыс. долл. Постройте мультипликативную модель временного ряда. Спрогнозируйте объем продаж в 1996 г.

Год	Квартал	Продажи	Год	Квартал	Продажи
1990	1	232,7	1993	1	178,3
	2	309,2		2	274,5
	3	310,7		3	295,4
	4	293,0		4	286,4
1991	1	205	1994	1	190,8
	2	234,4		2	263,5
	3	285,4		3	318,8
	4	258,7		4	305,3
1992	1	193,2	1995	1	242,6
	2	263,7		2	318,8
	3	292,5		3	329,6
	4	315,2		4	338,2

ТЕМА 7. СИСТЕМЫ ЭКОНОМЕТРИЧЕСКИХ УРАВНЕНИЙ

<u>Ключевые слова:</u>

Система независимых уравнений. Система рекурсивных уравнений. Система взаимозависимых уравнений. Структурная форма модели. Структурные коэффициенты модели. Структурные коэффициенты модели. Приведенная форма модели. Идентификация. Идентифицируемые, неидентифицируемые и сверхидентифицируемые структурные модели. Необходимое и достаточное условия идентификации модели. Косвенный метод наименьших квадратов. Двухшаговый метод наименьших квадратов.

<u>Основные теоретические аспекты темы:</u>

7.1. Понятие системы эконометрических уравнений

Экономические объекты и явления зачастую представляют собой сложные системы. И для описания их функционирования построение изолированных эконометрических уравнений может быть недостаточно. При использовании отдельных уравнений регрессии предполагается, что аргументы (факторы) можно изменять независимо друг от друга. Но данное предположение условно, так как на практике изменение одной переменной, как правило, не может происходить при абсолютной неизменности других. И отдельно взятое уравнение множественной регрессии не может характеризовать истинные влияния отдельных признаков на вариацию эндогенной переменной. Поэтому в экономических исследованиях важное место заняла проблема описания структуры связей между переменными системой одновременных уравнений.

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

Система независимых уравнений – система, в которой каждая зависимая переменная y рассматривается как функция одного и того же набора факторов x то есть система вида:

$$\begin{aligned} Y_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1; \\ Y_2 &= a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2; \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ Y_n &= a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n. \end{aligned}$$

Набор факторов в каждом уравнении может варьироваться.

Система рекурсивных уравнений – система, в которой зависимая переменная одного уравнения выступает в виде фактора x в другом уравнении, то есть система вида:

$$\begin{aligned} Y_1 &= a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1; \\ Y_2 &= b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2; \\ Y_3 &= b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{32}x_2 + \ldots + a_{3m}x_m + \varepsilon_2; \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ Y_n &= b_{n1}y_1 + b_{n2}y_2 + \ldots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n. \end{aligned}$$

Наибольшее распространение в эконометрических исследованиях получила система взаимозависимых уравнений.

Система взаимозависимых уравнений (система совместных одновременных уравнений) – система в которой одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других уравнениях – в правую, то есть система вида:

$$\begin{aligned} Y_1 &= b_{12}y_2 + b_{13}y_3 + \ldots + b_{1n}y_n + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1; \\ Y_2 &= b_{21}y_1 + b_{23}y_3 + \ldots + b_{2n}y_n + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_2; \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ Y_n &= b_{n1}y_1 + b_{n2}y_2 + \ldots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n. \end{aligned}$$

В эконометрике эта система уравнений называется *структурной формой модели*. И 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.

7.2. Структурная и приведенная форма модели

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

В структурной форме модели b_i – коэффициенты при эндогенных переменных, a_j – коэффициент при экзогенных переменных называются структурными коэффициентами модели. Все переменные в модели выражены в отклонениях от среднего уровня, поэтому свободный член в каждом уравнении системы отсутствует.

В силу того, что использование МНК для оценивания структурных коэффициентов модели дает смещенные и несостоятельные оценки, для определения структурных коэффициентов модели структурная форма модели преобразуется в приведенную форму модели.

Приведенная форма модели – система линейных функций эндогенных переменных от экзогенных:

$$
\begin{aligned}
&Y_1=\delta_{11}x_1+\delta_{12}x_2+\ldots+\delta_{1m}x_m\,;\\
&Y_2=\delta_{21}x_1+\delta_{22}x_2+\ldots+\delta_{2m}x_m\,;\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&Y_n=\delta_{n1}x_1+\delta_{n2}x_2+\ldots+\delta_{nm}x_m\,.
\end{aligned}
$$

где: δ_{ij} – коэффициенты приведенной формы модели.

Применяя МНК, можно оценить δ, а затем оценить значения эндогенных переменных через экзогенные.

Коэффициенты приведенной формы модели представляют собой нелинейные функции коэффициентов структурной формы модели.

7.3.Идентификация. Необходимое и достаточное условия идентификации модели

При переходе от приведенной формы модели к структурной исследователь сталкивается с проблемой идентификации.

Идентификация – это единственность соответствия между приведенной и структурной формами модели.

Структурная модель в полном виде, состоящая в каждом уравнении системы из n эндогенных и m экзогенных переменных, содержит n(n-1+m) параметров. При n=2 и m=3 модель содержит восемь структурных коэффициентов.

Приведенная форма модели в полном виде содержит nm параметров. То есть при n=2 и m=3 количество коэффициентов в приведенной модели равно 6. Следовательно, определение восьми коэффициентов структурной модели на основе шести коэффициентов приведенной модели не может привести к единственности решения.

Для того, 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 в модели. Также можно предположить, что воздействие экзогенных переменных на результирующую переменную одинаково, что приведет к равенству структурных коэффициентов модели.

С позиции идентифицируемости структурные модели подразделяются на три вида:

- Идентифицируемые;
- Неидентифицируемые;
- Сверхидентифицируемые.

Модель идентифицируема – если все ее структурные коэффициенты определяются однозначно, единственным образом по коэффициентам приведенной формы модели, то есть если число параметров структурной формы модели равно числу параметров приведенной формы модели.

Модель неидентифицируема – если число структурных коэффициентов больше числа приведенных коэффициентов и, следовательно, структурные коэффициенты не могут быть оценены через коэффициенты приведенной формы модели.

Структурная модель в полном виде, состоящая в каждом уравнении системы из n эндогенных и m предопределенных переменных в каждом уравнении системы, всегда неидентифицируема.

Модель сверхидентифицируема – если число структурных коэффициентов меньше числа приведенных коэффициентов и, следовательно, на основе приведенных коэффициентов можно получить два или более значений одного структурного коэффициента.

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 необходимо 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.

Необходимое условие идентифицируемости модели:

Чтобы уравнение было идентифицируемо, необходимо, чтобы число предопределенных (экзогенных) переменных D, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении (H) без одного.

То есть, условие идентифицируемости модели может быть записано в виде следующего счетного правила:

D+1=H – уравнение идентифицируемо;

D+1<H – уравнение неидентифицируемо;

D+1>H – уравнение сверхидентифицируемо.

Так, например, если в модели одно уравнение идентифицируемо, второе – сверхидентифицируемо, а третье - неидентифицируемо, то вся модель является неидентифицируемой.

Достаточное условие идентифицируемости модели:

Уравнение идентифицируемо, если по отсутствующим в нем экзогенным и эндогенным переменным 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.

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.

7.4. Оценивание параметров структурной модели

Коэффициенты структурной модели могут быть оценены разными способами в зависимости от вида системы одновременных уравнений. Косвенный и двухшаговый методы наименьших квадратов подробно описаны в литературе и рассматриваются как традиционные методы оценки коэффициентов структурной модели. Эти методы достаточно легко реализуемые. Косвенный метод наименьших квадратов применяется для идентифицируемой системы одновременных уравнений, а двухшаговый метод наименьших квадратов – для оценки коэффициентов сверхидентифицируемой модели.

Алгоритм косвенного метода наименьших квадратов:

- Структурная модель преобразовывается в приведенную форму модели.
- Для каждого уравнения приведенной формы модели обычным МНК оцениваются приведенные коэффициенты δ_{ij}.
- Коэффициенты приведенной формы модели трансформируются в параметры структурной формы модели.

Алгоритм двухшагового метода наименьших квадратов:

- Определяется приведенная форма модели, и находятся на ее основе теоретические значения эндогенных переменных, содержащиеся в правой части уравнения.
- Теоретические значения эндогенных переменных подставляются вместо фактических значений, и применяется обычный МНК к структурной форме сверхидентифицируемого уравнения.

Метод получил название «двухшагового метода наименьших квадратов», так как МНК используется дважды. На первом шаге - при определении приведенной формы модели и нахождении на ее основе оценок теоретических значений эндогенной переменной 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:

- Все уравнения системы сверхидентифицируемы;
- 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.

Вопросы для обсуждения:

1. Объясните, почему построение систем эконометрических уравнений важно в экономических исследованиях?
2. В чем сходство и различие моделей эконометрических уравнений с простыми моделями множественной регрессии?
3. Приведите примеры экономических процессов и явлений, которые могут быть описаны системами независимых, рекурсивных и взаимозависимых уравнений.
4. Почему необходимо преобразовывать структурную форму модели в приведенную? В чем отличие структурной и приведенной форм модели?
5. Что представляет собой идентификация модели?
6. Чем отличаются модели идентифицируемые, неидентифицируемые и сверхидентифицируемые?
7. В каком случае вся модель является идентифицируемой и сверхидентифицируемой?
8. В чем заключается необходимое условие идентифицируемости модели?
9. В чем заключается достаточное условие идентифицируемости модели?
10. В чем заключается алгоритм косвенного метода наименьших квадратов**?**

11. В чем заключается алгоритм двухшагового метода наименьших квадратов? Почему он называется двухшаговым?

Тест:

1 Система независимых уравнений – это:
   a. система, в которой каждая зависимая переменная у рассматривается как функция одного и того же набора факторов х;
   b. система, в которой зависимая переменная одного уравнения выступает в виде фактора х в другом уравнении;
   c. система в которой одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других уравнениях – в правую.

2 Система взаимозависимых уравнений – это:
   d. система, в которой каждая зависимая переменная у рассматривается как функция одного и того же набора факторов х;
   e. система, в которой зависимая переменная одного уравнения выступает в виде фактора х в другом уравнении;
   f. система в которой одни и те же зависимые переменные в одних уравнениях входят в левую часть, а в других уравнениях – в правую.

3 Верно утверждение:
   a. коэффициенты структурной формы модели представляют собой нелинейные функции коэффициентов приведенной формы модели;
   b. коэффициенты приведенной формы модели представляют собой нелинейные функции коэффициентов структурной формы модели;
   c. данные коэффициенты не связаны между собой.

4 При n=3 и m=4 модель содержит:
   a. восемь структурных коэффициентов;
   b. двенадцать структурных коэффициентов;
   c. восемнадцать структурных коэффициентов.

5 Приведенная форма модели при n=3 и m=4 содержит:
   a. восемь параметров;
   b. двенадцать параметров;
   c. восемнадцать параметров.

6 Модель идентифицируема, если:

a. все ее структурные коэффициенты определяются однозначно, единственным образом по коэффициентам приведенной формы модели;
b. если число структурных коэффициентов больше числа приведенных коэффициентов;
c. если число структурных коэффициентов меньше числа приведенных коэффициентов.

7 Если число структурных коэффициентов меньше числа приведенных коэффициентов, и, следовательно, на основе приведенных коэффициентов можно получить два или более значений одного структурного коэффициента, то модель:

a. идентифицируема;
b. неидентифицируема;
c. сверхидентифицируема.

8 Модель считается идентифицируемой:

a. если каждое уравнение системы идентифицируемо;
b. если хотя бы одно из уравнений системы неидентифицируемо;
c. если модель содержит хотя бы одно сверхидентифицируемое уравнение.

9 Чтобы уравнение было идентифицируемо, необходимо, чтобы:

a. число предопределенных (экзогенных) переменных D, отсутствующих в данном уравнении, но присутствующих в системе, было больше числу эндогенных переменных в данном уравнении (H) без одного;
b. число предопределенных (экзогенных) переменных D, отсутствующих в данном уравнении, но присутствующих в системе, было меньше числу эндогенных переменных в данном уравнении (H) без одного;
c. число предопределенных (экзогенных) переменных D, отсутствующих в данном уравнении, но присутствующих в системе, было равно числу эндогенных переменных в данном уравнении (H) без одного.

10 Уравнение идентифицируемо, если:

a. по отсутствующим в нем экзогенным и эндогенным переменным можно из коэффициентов при них в других уравнениях системы получить матрицу, определитель которой не равен нулю;
b. по отсутствующим в нем экзогенным и эндогенным переменным можно из коэффициентов при них в других уравнениях системы получить матрицу, определитель которой равен нулю;
c. по отсутствующим в нем экзогенным и эндогенным переменным

можно из коэффициентов при них в других уравнениях системы получить матрицу, определитель которой равен единице.

11 Уравнение сверхидентифицируемо, если:

a. D+1=H;
b. D+1<H;
c. D+1>H.

Задания для самостоятельной работы:

№ 1.

Проверьте, идентифицируема ли эконометрическая модель**:**

$Y_1 = b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2$;
$Y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + a_{23}x_3$;
$Y_3 = b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{33}x_3 + a_{34}x_4$.

№ 2.

Проверьте, идентифицируема ли эконометрическая модель**:**

$Y_1 = b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2$;
$Y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3$;
$Y_3 = b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{33}x_3 + a_{34}x_{4.}$.

№3.

Проверьте, каждое уравнение системы на необходимое и достаточное условие идентификации.

$Y_1 = b_{12}y_2 + b_{13}y_3 + a_{11}x_1 + a_{12}x_2$;
$Y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 + a_{24}x_4$;
$Y_3 = b_{31}y_1 + b_{32}y_2 + a_{31}x_1 + a_{32}x_2$.

№4.

Постройте, используя статистику в таблице, эконометрическую модель косвенным методом наименьших квадратов**:**

$Y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1$;
$Y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_{2.}$

№ региона	Y_1	Y_2	x_1	x_2
1	2	5	1	3
2	3	6	2	1
3	4	7	3	2
4	5	8	2	5
5	6	5	4	6

№5.

Постройте, используя статистику в таблице, эконометрическую модель двухшаговым методом наименьших квадратов**:**

$Y_1 = b_{12}(y_2 + x_1) + \varepsilon_1$;

$Y_2 = b_{21}y_1 + a_{22}x_2 + +\varepsilon_2.$

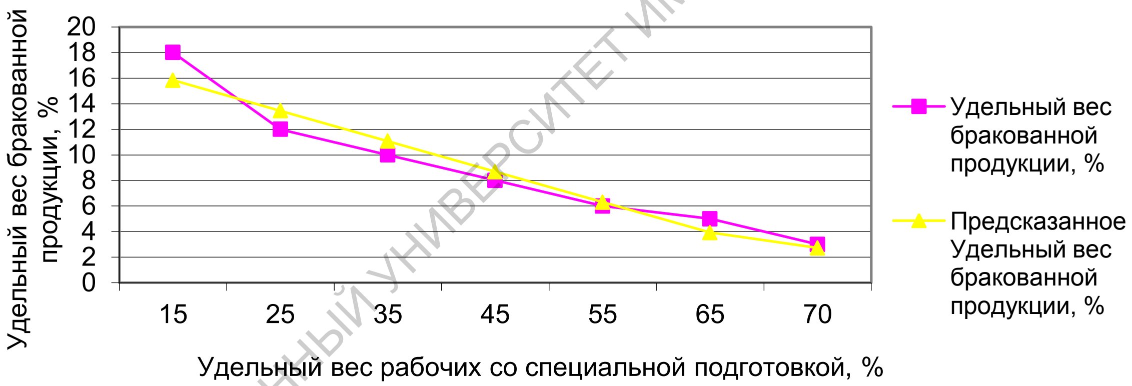
№ региона	Y_1	Y_2	x_1	x_2
1	2	5	1	3
2	3	6	2	1
3	4	7	3	2
4	5	8	2	5
5	6	5	4	6

Список литературы:

1. Аистов А.В., Максимов А.Г. Эконометрика шаг за шагом. – М.: Из – во ГУ ВШЭ, 2006.
2. Бородич С.А. Эконометрика.Практикум: учебное пособие. – М.:ИНФРА –М,2015.
3. Валентинов В.А. Эконометрика: Учебник – М.: Издательско – торговая корпорация «Дашков и К», 2006.
4. Гладилин А.В., Герасимов А.Н.. Громов Е.И. Эконометрика. Учебное пособие. – М.: Конкурс, 2006.
5. Доугерти Кристофер Введение в эконометрику: Учебник. 2-е изд./ Перевод с английского. М.: ИНФРА-М, 2007.
6. Костюнин В.И. Эконометрика: учебник и практикум для прикладного бакалавриата. - М.: Издательство Юрайт, 2015.
7. Колемаев В.А. Эконометрика: Учебник. – М.: ИНФРА – М, 2007.
8. Минько А.А. Прогнозирование в бизнесе с помощью Excel. Просто как дважды два. – М.: Эксмо, 2007.
9. Просветов Г.И. Эконометрика: задачи и решения: Учебно-практическое пообие. 5-е изд., доп. –М.: Издательство «Альфа-Пресс», 2008.
10. Эконометрика: Учебник. / И.И. Елисеева, С.В. Курышева, Т.В. Костеева и др.; Под ред. И.И. Елисеевой. – 2 –е изд., перераб. и доп. - М.: «Финансы и статистика», 2006.
11. Эконометрика: Учебник / Н.П. Тихомиров. Е.Ю. Дорохина – 2 –е изд. – М.: Из – во «Экзамен», 2007.

ПРИЛОЖЕНИЯ
Пример 1

Удельный вес рабочих со специальной подготовкой, %	Удельный вес бракованной продукции, %	Предсказанное Удельный вес бракованной продукции, %	Остатки
15	18	15,83427762	2,16572238
25	12	13,45184136	-1,45184136
35	10	11,0694051	-1,0694051
45	8	8,686968839	-0,68696884
55	6	6,304532578	-0,30453258
65	5	3,922096317	1,07790368
70	3	2,730878187	0,26912181



ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,967613138
R-квадрат	0,936275185
Нормированный R-квадрат	0,923530222
Стандартная ошибка	1,395764536
Наблюдения	7

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	1	143,1163497	143,1163497	73,46236939	0,000356286
Остаток	5	9,740793201	1,94815864		
Итого	6	152,8571429			

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	19,40793201	1,339264901	14,4914811	2,8238E-05	15,96524198	22,85062204
Удельный вес рабочих со специальной подготовкой, %	-0,238243626	0,027796417	-8,571019157	0,000356286	-0,309696592	-0,16679066

ВЫВОД ОСТАТКА

Наблюдение	*Предсказанное Удельный вес бракованной продукции, %*	*Остатки*
1	15,83427762	2,16572238
2	13,45184136	-1,45184136
3	11,0694051	-1,069405099
4	8,686968839	-0,686968839
5	6,304532578	-0,304532578
6	3,922096317	1,077903683
7	2,730878187	0,269121813

Пример 2

№	Цена станка Y	N, кВТ	n, об/мин	УА	Точность	Р,мм
1	53,6	11	2000	3	1	400
2	43,6	10	1500	1	1	400
3	35	8	2000	1	1	400
4	39	8	2500	1	1,6	400
5	27	8	2500	1	1	320
6	44	8	3000	3	1,6	320
7	26	6,3	2000	1	1	320
8	32,7	8	3000	3	1,6	250
9	22,5	6,3	3000	1	1,6	250
10	20	5,5	2500	1	1	250
11	29,8	8	2500	1	1,6	320
12	37,5	8	2000	1	1,6	400
Стандартные отклонения	9,883606721	1,506425631	482,6536	0,904534	0,313339781	63,02356462

Матрица коэффициентов парной корреляции

	Р,мм	*N, кВТ*	*n, об/мин*	*УА*	*Точность*	*Цена станка Y*
Р,мм	1					
N, кВТ	0,678177284	1				
n, об/мин	-0,735944385	-0,45168106	1			
УА	-0,119602623	0,430322863	0,364405	1		
Точность	-0,207157819	-0,144445942	0,631169	0,19245	1	
Цена станка Y	0,719398852	0,893602362	-0,35661	0,561823	0,002641915	1

Определитель матрицы коэффициентов парной корреляции (исключая r_{xy}) 0,041198237

1	0,678177284	-0,73594	-0,1196	-0,207157819
0,678177284	1	-0,45168	0,430323	-0,144445942
-0,735944385	-0,45168106	1	0,364405	0,631168744
-0,119602623	0,430322863	0,364405	1	0,19245009
-0,207157819	-0,144445942	0,631169	0,19245	1

Фактическое значение ХИ- квадрат = *14,4628*

Табличное значение ХИ-квадрат с 60 (66) степенями свободы=79,0819

	Коэффициенты	Стандартизованные коэффициенты регрессии
Y-	-18,54454098	

пересечение		
N, кВТ	1,714024417	0,261245756
n, об/мин	-0,001599577	-0,078113369
УА	5,741931188	0,52549361
Точность	3,361197928	0,106559989
Р,мм	0,089337289	0,569665968

Модель имеет вид
Y=-18,55+1,714*N-0,0016*n+5,7419*УА+3,36*Точность+0,0893*Р

Стандартизованное уранение множественной регрессии
t_Y=0,26*t_N-0,078*t_n+0,525*$t_{УА}$+0,1*$t_{Точность}$+0,57*$_P$

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,983177631
R-квадрат	0,966638253
Нормированный R-квадрат	0,938836798
Стандартная ошибка	2,444334082
Наблюдения	12

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	5	1038,693885	207,7387771	34,7693398	0,000234593
Остаток	6	35,84861464	5,974769106		
Итого	11	1074,5425			

F- критерий Фишера табличное с 6 и 5 степенями свободы (1% значимости) = 8,75

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*
Y-пересечение	-18,54454098	12,44261982	-1,490404855	0,18670494	-48,99055713	11,901475
N, кВТ	1,714024417	0,978855183	1,75105005	0,13050092	-0,681149682	4,1091985
n, об/мин	-0,001599577	0,003710633	-0,431079388	0,68146771	-0,010679175	0,0074800
УА	5,741931188	1,336397116	**4,296575561**	0,00511192	2,471882856	9,011979
Точность	3,361197928	3,553909218	0,945774842	0,38077674	-5,334911016	12,057306
Р,мм	0,089337289	0,024293487	**3,677417326**	0,01036244	0,029893225	0,1487813

t - Статистика табличное с 6 степенями свободы (5% значимости) = 1,9

№	Цена станка Y	N, кВТ	УА	Точность	Р,мм
1	53,6	11	3	1	400
2	43,6	10	1	1	400
3	35	8	1	1	400
4	39	8	1	1,6	400
5	27	8	1	1	320
6	44	8	3	1,6	320
7	26	6,3	1	1	320
8	32,7	8	3	1,6	250
9	22,5	6,3	1	1,6	250
10	20	5,5	1	1	250
11	29,8	8	1	1,6	320
12	37,5	8	1	1,6	400
Стандартные отклонения	9,883606721	1,506426	0,904534	0,31334	63,02356462

Матрица коэффициентов парной корреляции

	Цена станка Y	*N, кВТ*	*УА*	*Точность*	*Р,мм*
Цена станка Y	1	0,893602	0,561823	0,002642	0,719398852
N, кВТ	0,893602362	1	0,430323	-0,14445	0,678177284
УА	0,561822631	0,430323	1	0,19245	-0,119602623
Точность	0,002641915	-0,14445	0,19245	1	-0,207157819
Р,мм	0,719398852	0,678177	-0,1196	-0,20716	1

Определитель матрицы коэффициентов парной корреляции (исключая r_{xy}) 0,243286

Проверка H_0: о независимости переменных Det=1

Фактическое значение ХИ- квадрат = *7,886117*

Табличное значение ХИ-квадрат с 60 (66) степенями свободы=79,0819, следовательно

мультиколлиниарность не доказана

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,982652018
R-квадрат	0,965604988
Нормированный R-квадрат	0,945950695
Стандартная ошибка	2,297790946
Наблюдения	12

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	4	1037,583597	259,3958993	49,12946994	3,30499E-05
Остаток	7	36,95890262	5,279843232		
Итого	11	1074,5425			

F- критерий Фишера табличное с 7 и 4 степенями свободы (1% значимости) = 7,85

	Коэффициенты	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхни 95%*
Y-пересечение	-23,27443443	5,515878466	**-4,219533583**	0,003938883	-36,31740509	10,23146
N, кВТ	1,817560869	0,892041126	**2,037530351**	0,08100758	-0,291779702	3,92690
УА	5,477766749	1,116417631	**4,906557007**	0,00174007	2,837860433	8,117673
Точность	2,264460571	2,332662353	0,970762257	0,364006475	-3,251405454	7,780326
P,мм	0,095091263	0,019081095	**4,983532901**	0,001594612	0,049971676	0,140210

t - Статистика табличное с 7 степенями свободы (5% значимости) = 1,833

ВЫВОД ОСТАТКА

Наблюдение	*Предсказанное Цена станка Y*	*Остатки*
1	53,45300097	0,146999028
2	40,6799066	2,920093396
3	37,04478487	-2,044784866
4	38,40346121	0,596538791
5	29,43748386	-2,437483862
6	41,7516937	2,248306296
7	26,34763038	-0,347630385
8	35,09530532	-2,395305325
9	21,04991835	1,450081651
10	18,23719331	1,762806689
11	30,7961602	-0,996160205
12	38,40346121	-0,903461209