ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ОРЕНБУРГСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Б1.Б.07 Эконометрика

Направление подготовки 38.03.01 Экономика

Профиль подготовки Бухгалтерский учет, анализ и аудит

Форма обучения заочная

1. КОНСПЕКТ ЛЕКЦИЙ

1.1 Лекция № 1 (2 часа)

Тема: «Введение в эконометрику»

1.1.1 Вопросы лекции:

1. Эконометрика как наука. Предмет и задачи эконометрики
2. Информационные технологии на базе ПВЭМ в эконометрических исследованиях
3. Классификация переменных в эконометрических исследованиях
4. Понятие спецификации и идентифицируемости модели

1.1.2 Краткое содержание вопросов:

1. Эконометрика как наука. Предмет и задачи эконометрики

В журнале «Эконометрика», основанном в 1933 г. Р. Фришем (1895–1973), он дал следующее определение эконометрики: «Эконометрика – это не то же самое, что экономическая статистика. Она не идентична и тому, что мы называем экономической теорией, хотя значительная часть этой теории носит количественный характер. Эконометрика не является синонимом приложений математики к экономике. Как показывает опыт, каждая из трех отправных точек – статистика, экономическая теория и математика - необходимое, но не достаточное условие для понимания количественных соотношений в современной экономической жизни. Это – единство всех трех составляющих. И это единство образует эконометрику».

Таким образом, эконометрика – это наука, которая дает количественное выражение взаимосвязей экономических явлений и процессов. Нельзя утверждать, что достигнуто однозначное определение эконометрики. Так, Э. Маленво придерживался широкого понимания, интерпретируя эконометрику как «любое приложение математики или статистических методов к изучению экономических явлений».

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

С помощью эконометрики решается очень широкий круг задач. Наиболее общими задачами эконометрики являются:

1) обнаружение и анализ статистических закономерностей в экономике;

2) построение на базе выявленных эмпирических экономических зависимостей эконометрических моделей.

Данные задачи делятся на более конкретные подзадачи, которые можно классифицировать по трём признакам:

1) классификация задач по конечным прикладным целям:

а) прогноз социально-экономических показателей, определяющих состояние и развитие изучаемой системы;

б) моделирование возможных вариантов социально-экономического развития системы для выявления факторов, изменение которых оказывает наиболее мощное влияние на состояние системы в целом;

2) классификация задач по уровню иерархии:

а) задачи, решаемые на макроуровне (страна в целом);

б) задачи, решаемые на мезоуровне (уровень отраслей, регионов);

в) задачи, решаемые на микроуровне (уровень фирмы, семьи, предприятия);

3) классификация задач по профилю изучаемой экономической системы:

а) рынок;
б) инвестиционная, социальная, финансовая политика;
в) ценообразование;
г) распределительные отношения;
д) спрос и потребление;
е) отдельно выделенный комплекс проблем.

2. Информационные технологии на базе ПВЭМ в эконометрических исследованиях

Определение ПППП. Виды ППП. Проблемно-ориентированные ППП. Автоматизации проектирования (или САПР). ППП общего назначения. Офисные ППП. Настольные издательские системы. Системы искусственного интеллекта. Статистический пакет. STATGRAPHICS, SPSS, SYSTAT, BMDP,SAS, CSS, STATISTICA, S-plus, и др. STADIA, ЭВРИСТА, МЕЗОЗАВР, ОЛИМП: Стат-Эксперт, Статистик-Консультант, САНИ, КЛАСС-МАСТЕР и др. Специализированные пакеты. Пакеты общего назначения.

3. Классификация переменных в эконометрических исследованиях

В эконометрических моделях в основном используются данные трёх типов:
1) пространственные данные (cross-sectional data);
2) временные ряды (time-series data);
3) панельные данные (panel data).

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

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

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

4. Понятие спецификации и идентифицируемости модели

Помимо выбора спецификации модели не менее важно также правильное описание структуры модели, в частности, для зависимости данных о временном ряде. Значение результативного признака может зависеть не от фактического значения объясняющей переменной, а от значения, которое ожидалось в предыдущем периоде. Тогда, если ожидаемое и фактическое значения тесно связаны, то будет казаться, что между результативным признаком и объясняющей переменной имеется зависимость, хотя в действительности это всего лишь приближение (аппроксимация) и расхождение опять будет связано с наличием остаточного случайного члена.

1.2 Лекция № 2 (2 часа)

Тема: «Парная линейная регрессия»

1.2.1 Вопросы лекции:

1. Метод наименьших квадратов для построения линейной модели
2. Линейный коэффициент корреляции. Коэффициент детерминации
3. Проверка значимости регрессионной модели, проверка качества параметров уравнения и построение доверительных интервалов коэффициентов регрессии

4. Интерпретация уравнения регрессии
5. Средняя ошибка аппроксимации

1.2.2 Краткое содержание вопросов:

1. Метод наименьших квадратов для построения линейной модели

Классический подход к оцениванию параметров линейной регрессии основан на *методе наименьших квадратов* (МНК), разработанный Гауссом.

Метод наименьших квадратов позволяет получить такие оценки параметров a и b, при которых сумма квадратов отклонений фактических значений результативного признака у от расчетных (теоретических) минимальна:

$$S = \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2 \longrightarrow \min \quad \text{т.е} \quad \sum \varepsilon_i^2 \to \min$$

Другими словами, из множества линий линия регрессии выбирается так, чтобы сумма квадратов расстояний по вертикали между точками и этой линией была бы минимальной.

Для линейной парной зависимости:

$$S = \sum_{i=1}^{n} \left[y_i - (a + b \cdot x)\right]^2 \longrightarrow \min .$$

2. Линейный коэффициент корреляции. Коэффициент детерминации

Уравнение регрессии всегда дополняется показателем тесноты связи. Для линейной регрессии таким показателем является линейный коэффициент корреляции.

Был впервые введен Карлом Пирсоном (1857 - 1936).

В теории разработаны и на практике применяются различные модификации формул расчета:

$$r = b\frac{\sigma_x}{\sigma_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y},$$

$$\bar{x} = \frac{\sum x_i}{n} = \frac{\sum x_i \cdot f_i}{\sum f_i}; \quad \bar{y} = \frac{\sum y_i}{n} = \frac{\sum y_i \cdot f_i}{\sum f_i};$$

$$\overline{x \cdot y} = \frac{\sum x_i \cdot y_i}{n};$$

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x}) \cdot f_i}{\sum f_i}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{\sum x_i^2}{n} - (\bar{x})^2}$$

$$\sigma_y = \sqrt{\frac{\sum (y_i - \bar{y}) \cdot f_i}{\sum f_i}} = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}} = \sqrt{\frac{\sum y_i^2}{n} - (\bar{y})^2}$$

Линейный коэффициент корреляции находится в пределах:

$-1 \le r \le 1$

Чем ближе r по абсолютной величине к 1, тем теснее связь между признаками.

Знак указывает на направление связи

Если $|r| \ge 0{,}7$, то считают связь сильной;

$0{,}5 \le |r| \le 0{,}7$ – средней тесноты;

$|r| < 0{,}5$ – слабой.

Квадрат линейного коэффициента корреляции называется ***коэффициентом детерминации***. Этот коэффициент характеризует долю общей вариации результативного признака, которая объяснятся вариацией факторного признака. (Например, рассмотрели группировку производительности труда по квалификации рабочих и получили r^2=45,4%.

Фактор квалификации рабочих объясняет 45,4% вариации производительности труда, а неучтенные факторы – 54,6%.).

3. Проверка значимости регрессионной модели, проверка качества параметров уравнения и построение доверительных интервалов коэффициентов регрессии

Оценка значимости параметров уравнения регрессии дается с помощью *t*-критерия Стьюдента. Выдвигается нулевая гипотеза о равенстве оцениваемого параметра нулю.

- коэффициента регрессии в

H_0: b=0

1. Стандартная ошибка коэффициента регрессии

$$m_b = \sqrt{\frac{\sum (y-\tilde{y})^2}{\sum (x-\bar{x})^2 \cdot (n-2)}}$$

2. Фактическое значение t-критерия Стьюдента

$$t_b = \frac{b}{m_b}$$

3. Определяется $t_{табл}$ ($\alpha,\ df = n\text{-}2$) при определенном уровне значимости и числе степеней свободы.

$t_b > t_{табл}$, H_0 отклоняется и параметр b неслучайно отличается от нуля и сформировался под влиянием систематически действующего фактора *x* (статистически значим с заданной вероятностью).

$t_b < t_{табл}$ – H_0 принимается и признается случайная природа формирования b, т.е. связь надежно не установлена, статистически не значима.

Доверительный интервал для коэффициента регрессии определяется как $b \pm t \cdot m_b$. Поскольку коэффициент регрессии имеет четкую экономическую интерпретацию, доверительные границы интервала не должны содержать противоречивой информации, в частности включать значение 0 в этот интервал.

4. Интерпретация уравнения регрессии

Для прогнозирования возможных значений результативного признака

- авторегрессионное прогнозирование по тренду и колеблемости;

- факторное прогнозирование, основанное на изучении и количественном измерении взаимосвязи между признаками.

Основным условием прогнозирования на основании регрессионного уравнения является стабильность или, по крайней мере, малая изменчивость других факторов и условий изучаемого процесса, не связанных с ними. Если резко изменится «внешняя среда» протекающего процесса, прежнее уравнение регрессии потеряет свое значение.

Прогнозирование по уравнению регрессии проводится в два этапа:

1. Вычисляется «точечный прогноз».
2. Определяется доверительный интервал с достаточно большой вероятностью (интервальный прогноз).

5. Средняя ошибка аппроксимации

Важнейшей характеристикой качества модели является средняя относительная ошибка аппроксимации:

$$\bar{A} = \frac{1}{n}\sum\left|\frac{\hat{y}_i - y_i}{y_i}\right| \cdot 100$$

$\hat{y}_i$, y_i - соответственно расчетное и фактическое значения;

n – число наблюдений.

$\bar{A}$ < 10-13% - высокая точность модели;

10% < $\bar{A}$ < 20% - хорошая точность модели;

20% < $\bar{A}$ < 50% - удовлетворительная.

1.3 Лекция № 3 (1 час)

Тема: «Нелинейные модели регрессии и линеаризация»

1.3.1 Вопросы лекции:

1. Классификация нелинейной регрессии
2. Линеаризация нелинейной регрессии
3. Коэффициенты эластичности и аппроксимации
4. Корреляция для нелинейной регрессии

1.3.2 Краткое содержание вопросов:

1. Классификация нелинейной регрессии

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

Различают *два класса нелинейных регрессий:*

- регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам;
- регрессии, нелинейные по оцениваемым параметрам.

Примером нелинейной регрессии по включаемым в нее объясняющим переменным могут служить следующие функции:

- полиномы разных степеней

$\tilde{y}_x = a + bx + cx^2 + \varepsilon$

- равносторонняя гипербола $\tilde{y}_x = a + \frac{b}{x} + \varepsilon$

К нелинейным регрессиям по оцениваемым параметрам относятся функции:

- степенная - $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$
- показательная - $\tilde{y}_x = a \cdot b^x \cdot \varepsilon$
- экспоненциальная $\tilde{y}_x = e^{a+bx} \cdot \varepsilon$

-

2. Линеаризация нелинейной регрессии

Нелинейная регрессия по включенным переменным не таит каких-либо сложностей в оценке ее параметров. Она определяется, как и в линейной регрессии, методом наименьших квадратов (МНК), ибо эти функции линейны по параметрам. Так, в параболе второй степени $\tilde{y}_x = a + bx + cx^2 + \varepsilon$ заменяя переменные $x = x_1$, $x^2 = x_2$, получим двухфакторное уравнение линейной регрессии: $\tilde{y}_x = a + bx_1 + cx_2 + \varepsilon$.

Применение МНК для оценки параметров параболы второй степени приводит к следующей системе нормальных уравнений:

$$\begin{cases} \sum y = na + b\sum x + c\sum x^2 \\ \sum yx = a\sum x + b\sum x^2 + c\sum x^3 \\ \sum yx^2 = a\sum x^2 + b\sum x^3 + c\sum x^4 \end{cases}$$

Решение ее возможно методом определителей:

$a = \frac{\Delta a}{\Delta}$, $b = \frac{\Delta b}{\Delta}$, $c = \frac{\Delta c}{\Delta}$

где Δ - определитель системы;

Δa, Δb, Δc – частные определители для каждого из параметров.

При $b > 0$ и $c < 0$ кривая симметрична относительно высшей точки, т. е. точки перелома кривой, изменяющей направление связи, а именно рост на падение.

При b < 0 и $c > 0$ парабола второго порядка симметрична относительно своей низшей точки, что позволяет определять минимум функции в точке, меняющей направление связи, т. е. снижение на рост.

Такого рода функцию можно наблюдать в экономике труда при изучении зависимости заработной платы работников физического труда от возраста – с увеличением возраста повышается заработная плата ввиду одновременного увеличения опыта и повышения квалификации работника. Однако с определенного возраста ввиду старения организма и снижения производительности труда дальнейшее повышение возраста может приводить к снижению заработной платы работника.

Для равносторонней гиперболы вида $\tilde{y}_x = a + \frac{b}{x} + \varepsilon$ заменив $\frac{1}{x}$ на z, получим линейное уравнение регрессии $\tilde{y}_z = a + bz + \varepsilon$, оценка параметров которого может быть дана МНК. Система нормальных уравнений составит:

$$\begin{cases} \sum y = na + b\sum \frac{1}{x} \\ \sum \frac{y}{x} = a\sum \frac{1}{x} + b\sum \frac{1}{x^2} \end{cases}$$

При b>0 имеем обратную зависимость, которая при $x \to \infty$ характеризуется минимальным предельным значением y, оценкой которого служит параметр a.

При b < 0 имеем медленно повышающуюся функцию.

В отдельных случаях может использоваться и нелинейная модель вида $\tilde{y}_x = \frac{1}{a + bx + \varepsilon}$, так называемая обратная модель, являющаяся разновидностью гиперболы. Но если в равносторонней гиперболе $\tilde{y}_x = a + \frac{b}{x} + \varepsilon$ преобразованию подвергаются объясняющие переменные $\frac{1}{x} = z$, то для получения линейной формы зависимости в обратной модели преобразовывается y, а именно: $1/y$ = z и $z = a + bx + \varepsilon$. В результате обратная модель оказывается внутренне нелинейной и требование МНК выполняется не для фактических значений признака y, а для их обратных величин 1/y , а именно:

$$\sum \left(z - \tilde{z}_x\right)^2 \to \min$$

Соответственно получаем систему нормальных уравнений

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

Иначе обстоит дело с регрессией, нелинейной по оцениваемым параметрам. Данный класс нелинейных моделей подразделяется на два типа: нелинейные модели внутренне линейные и нелинейные модели внутренне нелинейные. Если нелинейная модель внутренне линейна, то она с помощью соответствующих преобразований может быть приведена к линейному виду. Если же нелинейная модель внутренне нелинейна, то она не может быть сведена к линейной функции. Например, в эконометрических исследованиях при изучении эластичности спроса от цен широко используется степенная функция: $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$,

где y - спрос (количество);

x - цена;

ε - случайная ошибка.

Данная модель нелинейна относительно оцениваемых параметров. Однако ее можно считать внутренне линейной, ибо логарифмирование данного уравнения по основанию ε приводит его к линейному виду:

$\ln y = \ln a + b \ln x + \ln \varepsilon$.

Соответственно оценки параметров a и b могут быть найдены МНК. Если же модель представить в виде $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$, то она становится внутренне нелинейной, ибо ее невозможно превратить в линейный вид.

В специальных исследованиях по регрессионному анализу часто к нелинейным относят модели, только внутренне нелинейные по оцениваемым параметрам, а все другие модели, которые внешне нелинейны, но путем преобразований параметров могут быть приведены к линейному виду, относятся к классу линейных моделей.

Модели внутренне нелинейные по параметрам могут иметь место в эконометрических исследованиях. Однако гораздо большее распространение получили модели, приводимые к линейному виду. Решение такого типа моделей реализовано в стандартных пакетах прикладных программ.

Среди нелинейных функций, которые могут быть приведены к линейному виду, в эконометрических исследованиях очень широко используется степенная функция $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$. Связано это с тем, что параметр b в ней имеет четкое экономическое истолкование, т. е. он является коэффициентом эластичности. Это значит, что величина коэффициента b показывает, на сколько процентов изменится в среднем результат, если фактор изменится на 1 %.

Для оценки параметров степенной функции $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$ применяется МНК к линеаризованному (логарифмическому) уравнению $\ln y = \ln a + b \ln x + \ln \varepsilon$, т.е. решается система нормальных уравнений:

$$\begin{cases} \sum \ln y = n \ln a + b \sum \ln x \\ \sum \ln y \ln x = \ln a \sum \ln x + b \sum (\ln x)^2 \end{cases}$$

Параметр b определяется непосредственно из системы, а параметр a — косвенным путем после потенцирования величины *ln*a.

3. Коэффициент эластичности

Коэффициент эластичности является показателем силы связи. На его основе делают выбор наиболее значимого фактора. Рассмотрим формулу расчета коэффициента эластичности

$$\acute{Y} = f'(x) \frac{x}{y},$$

где f′*(x)* — первая производная, характеризующая соотношение приростов результата и фактора для соответствующей формы связи.

Во всех функциях, кроме степенной, коэффициент эластичности зависит от значений фактора *x*.

В силу того, что коэффициент эластичности для линейной функции не является величиной постоянной, а зависит от соответствующего значения *x,* то обычно рассчитывается средний показатель эластичности по формуле

$$\bar{\acute{Y}} = b \frac{\bar{x}}{\bar{y}}.$$

Средний коэффициент эластичности $\bar{\acute{Y}}$ показывает, на сколько процентов в среднем по совокупности измениться результат y от своей средней величины при изменении фактора x на 1 % от своего среднего значения.

Поскольку коэффициенты эластичности представляют экономический интерес, а виды моделей не ограничиваются только степенной функцией, приведем формулы расчета коэффициентов эластичности для наиболее распространенных типов уравнений регрессии (табл. 1).

Таблица 1. – Коэффициенты эластичности для ряда математических функций

Вид функции, $\tilde{y}_x$	Первая производная, $f'(x)$	Коэффициент эластичности, $Э = f'(x)\frac{x}{y}$
Линейная $\tilde{y}_x = a + bx + \varepsilon$	b	$Э = \frac{bx}{a+bx}$
Параболическая $\tilde{y}_x = a + bx + cx^2 + \varepsilon$	$b + 2cx$	$Э = \frac{(b+2cx)x}{a+bx+cx^2}$
Гипербола $\tilde{y}_x = a + \frac{b}{x} + \varepsilon$	$\frac{-b}{x^2}$	$Э = \frac{-b}{ax+b}$
Степенная $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$	$a \cdot b \cdot x^{b-1}$	$Э = a \cdot b \cdot x^{b-1} \frac{x}{a \cdot x^b} = b$
Показательная $\tilde{y}_x = a \cdot b^x \cdot \varepsilon$	$\ln b \cdot a \cdot b^x$	$Э = \ln x \cdot b$
Обратная $\tilde{y}_x = \frac{1}{a+bx+\varepsilon}$	$-\frac{b}{(a+bx)^2}$	$Э = \frac{-bx}{a+bx}$

Несмотря на широкое использование в эконометрике коэффициентов эластичности, возможны случаи, когда их расчет экономического смысла не имеет. Это происходит тогда, когда для рассматриваемых признаков бессмысленно определение изменения значений в процентах. Например, вряд ли кто будет определять, на сколько процентов может измениться заработная плата с ростом стажа работы на 1 %. Или, например, на сколько процентов изменится урожайность пшеницы, если качество почвы, измеряемое в баллах, изменится на 1 %. В такой ситуации степенная функция, даже если она оказывается наилучшей по формальным соображениям (исходя из наименьшего значения остаточной вариации), не может быть экономически интерпретирована.

4.Корреляция для нелинейной регрессии

Уравнение нелинейной регрессии, также, как и в линейной зависимости, дополняется показателем корреляции, а именно индексом корреляции (R):

$$R = \sqrt{1 - \frac{\sum\left(y - \acute{y}_x\right)^2}{\sum\left(y - \bar{y}\right)^2}}$$

Величина данного показателя находится в границах: $0 \le R \le 1$, чем ближе к единице, тем теснее связь рассматриваемых признаков, тем более надежно найденное уравнение регрессии.

Парабола второй степени, как и полином более высокого порядка, при линеаризации принимает вид уравнения множественной регрессии. Если же нелинейное относительно объясняемой переменной уравнение регрессии при линеаризации принимает форму линейного уравнения парной регрессии, то для оценки тесноты связи может быть использован линейный коэффициент корреляции, величина которого в этом случае совпадет с индексом корреляции $R_{yx} = r_{yz}$, где z - преобразованная величина признака-фактора, например $z = \frac{1}{x}$ или $z = \ln x$.

Иначе обстоит дело, когда преобразования уравнения в линейную форму связаны с зависимой переменной. В этом случае линейный коэффициент корреляции по преобразованным значениям признаков дает лишь приближенную оценку тесноты связи и численно не совпадает с индексом корреляции. Так, для степенной функции $\tilde{y}_x = a \cdot x^b \cdot \varepsilon$ после перехода к логарифмически линейному уравнению $\ln y = \ln a + b \ln x + \ln \varepsilon$ может быть найден линейный коэффициент корреляции не для фактических значений переменных *x* и *y,* а для их логарифмов, т. е. $r_{\ln y \ln x}$. Соответственно квадрат его значения будет характеризовать отношение факторной суммы квадратов отклонений к общей, но не для *y,* а для его логарифмов:

$$r^2_{\ln y \ln x} = \frac{\sum(\ln \tilde{y} - \overline{\ln y})^2}{\sum\left(\ln y - \overline{\ln y}\right)^2} = 1 - \frac{\sum(\ln y - \ln \tilde{y})^2}{\sum\left(\ln y - \overline{\ln y}\right)^2}$$

Поскольку в расчете индекса корреляции используется соотношение факторной и общей суммы квадратов отклонений, то R^2 имеет тот же смысл, что и коэффициент детерминации. В специальных исследованиях величину R^2 для нелинейных связей называют индексом детерминации.

Оценка существенности индекса корреляции проводится, так же как и оценка надежности коэффициента корреляции.

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

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - m}{m - 1}$$

где R^2 - индекс детерминации;

n - число наблюдений;

m — число параметров.

Величина (*m*-1*)* характеризует число степеней свободы для факторной суммы квадратов, а *(n - m*) — число степеней свободы для остаточной суммы квадратов.

1.4 Лекция №4 (1 час)

Тема «Линейная модель множественной регрессии»

1.4.1 Вопросы лекции:

1. Предпосылки МНК.
2. Множественная линейная регрессия и оценка ее параметров.
3. Проверка статистической значимости уравнения регрессии в целом и его параметров.
4. Показатели корреляции и оценка их значимости.
5. Мультиколлинеарность.

1.4.2. Краткое содержание вопросов

1. Предпосылки МНК.

Математическое ожидание случайного отклонения ε_i равно нулю: $M(\varepsilon_i) = 0$, для всех наблюдений.

Дисперсия случайных отклонений ε_i постоянна: $D(\varepsilon_i) = D(\varepsilon_j) = \sigma^2$ для любого наблюдения i = j.

3. Случайные отклонения ε_i и ε_j являются независимыми друг от друга для $i \neq j$.

Предполагается, что отсутствует систематическая связь между любыми случайными отклонениями: $\sigma_{\varepsilon_i\varepsilon_j}=\text{cov}(\varepsilon_i\varepsilon_j)=\begin{cases}0, & i\neq j\\ \sigma^2, & i=j\end{cases}$. Если данное условие выполняется, то говорят об отсутствии *автокорреляции*.

4. Случайное отклонение должно быть независимо от объясняющих переменных.

Обычно это условие выполняется автоматически, если объясняющие переменные не являются случайными в данной модели: $\sigma_{\varepsilon_i x_i}=0$. Выполнимость данной предпосылки не столь критична для эконометрической модели.

5. Модель является линейной относительно параметров.

Для случая множественной регрессии существенными являются еще две предпосылки.

6. Отсутствие мультиколлинеарности.

Между объясняющими переменными отсутствует строгая (сильная) линейная зависимость.

7. Ошибки ε_i , i=1,2,… , n, имеют нормальное распределение (ε_i~N(0, σ)).

Выполнимость данной предпосылки важна для проверки статистических гипотез и построения интервальных оценок.

Теорема (Гаусса-Маркова) Если предпосылки 1-7 выполняются, то оценки полученные по МНК обладают свойствами несмещенности, состоятельности и эффективности.

(Бородич С.А. Эконометрика / Учебное пособие, стр. 124)

2. Множественная линейная регрессия и оценка ее параметров

Как и в парной зависимости, возможны разные виды уравнений множественной регрессии: линейные и нелинейные.

Ввиду четкой интерпретации параметров наиболее широко используются линейная функция: $\tilde{y}_x=a+b_1x_1+b_2x_2+...+b_nx_n$. (*)

В линейной множественной регрессии параметры при x называются коэффициентами «чистой» регрессии. Они характеризуют среднее изменение результата с увеличением соответствующего фактора на единицу при неизменном значении других факторов, закрепленных на среднем уровне.

Основная задача регрессионного анализа заключается в нахождении по выборке объемом n, оценки неизвестных коэффициентов регрессии b_0, b_1, …, b_k модели (*).

Для оценки параметров уравнения множественной регрессии применяют метод наименьших квадратов. Для линейных уравнений строиться система нормальных уравнений, решение которой позволяет получить оценки параметров регрессии:

$$\begin{cases}\sum y=na+b_1\sum x_1+b_2\sum x_2+...+b_n\sum x_n\\ \sum yx_1=a\sum x_1+b_1\sum x_1^2+b_2\sum x_2x_1+...+b_n\sum x_nx_1\\ ....\\ \sum yx_n=a\sum x_n+b_1\sum x_1x_n+b_2\sum x_2x_n+...+b_n\sum x_n^2\end{cases}$$

Для ее решения может быть применен метод определителей:

$$a=\frac{\Delta a}{\Delta},\quad b_1=\frac{\Delta b_1}{\Delta},\quad ...,\quad b_n=\frac{\Delta b_n}{\Delta},$$

где Δa, Δb_1,…, Δb_p – частные определители, которые получаются путем замены соответствующего столбца матрицы определителя системы данными левой части системы.

Δ - определитель системы.

Возможна также и другая запись уравнения (*), в так называемом стандартизованном масштабе: $t_Y = \beta_1 t_{x_1} + \ldots + \beta_k t_{x_k} + \varepsilon$,

где $t_y, t_{x_1}, \ldots, t_{x_k}$ - стандартизованные переменные: $t_y = \dfrac{y - \bar{y}}{\sigma_y}$, $t_{x_j} = \dfrac{x_j - \bar{x}_j}{\sigma_{x_j}}$, *j=1, 2, ..., k,* для которых среднее значение равно нулю, а среднее квадратическое отклонение равно единице;

β_j – стандартизованные коэффициенты регрессии.

$$\beta_j = b_j \frac{\sigma_{x_j}}{\sigma_y},\ j=1, 2, \ldots, k.$$

Данное соотношение позволяет переходить от уравнения (*) к уравнению в стандартизованном масштабе.

Стандартизованные коэффициенты регрессии показывают на сколько «сигм» изменится в среднем результат (У), если соответствующий фактор X_j изменится на одну «сигму» при неизменном среднем уровне других факторов.

В силу того, что все переменные центрированы и нормированы, коэффициенты β_j $j=l,2,\ldots,k$, сравнимы между собой (в этом их отличие от b). Сравнивая их друг с другом, можно ранжировать факторы по силе их воздействия на результат, что позволяет произвести отсев факторов — исключить из модели факторы с наименьшими значениями β_j.

3. Проверка статистической значимости уравнения регрессии в целом и его параметров.

Значимость уравнения регрессии, т. е. гипотеза H_0: b=0 ($b_0 = b_1 = \ldots = b_k = 0$), проверяется по *F-критерию*, наблюдаемое значение которого определяется по формуле:

$$F = \frac{R^2}{1 - R^2} \cdot \frac{n - m}{m - 1}$$

R^2 - коэффициент (индекс) множественной детерминации;

m - число параметров;

n - число наблюдений.

По таблице F-распределения для заданных α, V_1= к + 1, V_2 = n – k – 1 находят $F_{кр}$. Гипотеза H_0 отклоняется с вероятностью α, если $F_{набл} > F_{кр}$. Из этого следует, что уравнение является значимым, т. е. хотя бы один из коэффициентов регрессии отличен от нуля.

Для проверки значимости отдельных коэффициентов регрессии, т. е. гипотез H_0: b_j = 0, где $j=l,2,\ldots k$, используют t-критерий и вычисляют:

$t_{набл}(b_j) = b_j / m_{bj}$.

$$m_{b_j} = \frac{\sigma_y \cdot \sqrt{1 - R_y^2}}{\sigma_{x_j} \cdot \sqrt{1 - R_{x_j}^2}} \cdot \frac{1}{\sqrt{n - m - 1}},$$

где R_y – коэффициент детерминации уравнения множественной регрессии;

R_{xj} – коэффициент детерминации для зависимости фактора x_j со всеми другими факторами уравнения регрессии.

По таблице t - распределения для заданного α и ν = n – k – 1, находят $t_{кр}$.

Гипотеза H_0 отвергается с вероятностью α, если $t_{набл} > t_{кр}$

Из этого следует, что соответствующий коэффициент регрессии b_j значим, т. е. $b_j \neq 0$. В противном случае коэффициент регрессии незначим и соответствующая переменная в модель не включается.

4. Показатели корреляции и оценка их значимости.

Прежде чем определять вид зависимости и описывать ее уравнением регрессии необходимо ответить на вопрос о наличии или отсутствии зависимостей между анализируемыми показателями.

Если распределение значений каждого факторного признака $X_j = (x_1, x_2, \dots, x_k)_j$ подчиняется k-мерному нормальному закону распределения, то степень линейной зависимости между признаками величинами с помощью парных, частных и множественных коэффициентов корреляции и детерминации.

Парный коэффициент корреляции характеризует тесноту линейной зависимости между двумя переменными на фоне действия всех остальных показателей, входящих в модель.

$$r_{jl} = \frac{1/n \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{il} - \bar{x}_l)}{\sigma_j \cdot \sigma_l}, \ -1 \le r_{jl} \le 1.$$

Если r_{jl} близок к ± 1, то связь между переменными сильная, если $r_{jl} = 0$ – линейная связь отсутствует.

$r_{jl} > 0$ – связь между переменными прямая;

$r_{jl} < 0$ – связь обратная.

Для многомерной корреляционной модели важную роль играют частные и множественные коэффициенты корреляции, детерминации (квадраты соответствующих коэффициентов корреляции).

Частный коэффициент корреляции характеризует тесноту линейной связи между двумя переменными при исключении влияния всех остальных признаков, входящих в модель.

Обладает всеми свойствами парного коэффициента корреляции.

Частный коэффициент корреляции k-2-го порядка между признаками x_1 и x_2 при фиксированном воздействии переменных x_3, x_4, …, x_k может быть определен по следующей формуле:

$$r_{x_1 x_2}(x_3, x_4, \dots, x_k) = r_{12/3,4,\dots,k} = \frac{-R_{12}}{\sqrt{R_{11} R_{22}}}$$

R_{jl} — алгебраическое дополнение матрицы R к элементу r_{jl}.

$$r_{yx_j}(x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k) = r_{yx_j/1,\dots,j-1,j+1,\dots,k} = \sqrt{1 - \frac{1 - R^2_{yx_1 \dots x_k}}{R^2_{yx_1 \dots x_{j-1}, x_{j+1} \dots x_k}}},$$

$R^2_{yx_1 \dots x_k}$ - множественный коэффициент детерминации всего комплекса k факторов с результатом;

$R^2_{yx_1 \dots x_{j-1}, x_{j+1, \dots} x_k}$ - тот же показатель детерминации, но без введения в модель фактора x_j.

Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается. В практических исследованиях предпочтение отдают показателям частной корреляции самого высокого порядка.

Частный коэффициент корреляции между x_1 и x_2 при фиксированном воздействии переменной x_3 рассчитывается по формуле:

$$r_{x_1 x_2}(x_3) = \frac{-R_{12}}{\sqrt{R_{11} R_{22}}} = \frac{r_{12} - r_{13} r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

Остальные частные коэффициенты определяются аналогично, путем замены соответствующих индексов.

Если частный коэффициент корреляции меньше парного, т.е. $r_{x_1x_2}(x_3) < r_{x_1x_2}$, то взаимодействие между x_1 и x_2 обусловлено частично (или полностью, если $r_{x_1x_2}(x_3) = 0$) воздействием фиксируемых прочих переменных, т.е. - x_3. Если частный коэффициент корреляции $r_{x_1x_2}(x_3) > r_{x_1x_2}$, то фиксируемые прочие переменные ослабляют линейную связь.

Множественный коэффициент корреляции характеризует степень линейной связи между одной переменной (результативной) и остальными, входящими в модель, и может быть рассчитан по формуле:

$$r_y = r_y(x_1, x_2, ..., x_k) = \sqrt{1 - \frac{\det R}{R_{yy}}}$$

$$r_y = r_y(x_1, x_2) = \sqrt{1 - \frac{\det R}{R_{yy}}} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{x_1x_2}r_{yx_1}r_{yx_2}}{1 - r_{x_1x_2}^2}}$$

Если $r_y = 1$, то связь между y и двумерной переменной (x_1, x_2) является функциональной, линейной. Если $r_{x_3} = 0$, то; линейной связи нет.

Из формулы r_y следует, что $r_y > |r_{yx1}|$, $r_y > |r_{yx2}|$, $r_y > |r_{yx1|x2}|$, $r_y > |r_{yx2|x1}|$.

Отсюда можно заметить, что коэффициент множественной корреляции может только увеличиваться, если в модель включать дополнительные признаки – случайные величины, и не увеличиться, если из имеющихся признаков производить исключение.

Если $r_3 = 0$, то $r_{31} = r_{32} = r_{31|2} = r_{32|1} = 0$.

Наибольшему множественному коэффициенту детерминации соответствуют большие частные коэффициенты детерминации (например, r_1^2 соответствуют $r_{12|3}^2$ и $r_{13|2}^2$)

Множественный коэффициент детерминации (квадрат соответствующего множественного коэффициента корреляции) характеризует долю дисперсии, например, случайной величины y, обусловленную изменением остальных случайных величин x_1 и x_2, входящих в модель.

Множественный коэффициент детерминации характеризует качество регрессионной модели.

Однако использование R^2 в случае множественной регрессии является *не вполне корректным,* так как коэффициент детерминации возрастает при добавлении регрессоров в модель. Это происходит потому, что остаточная дисперсия уменьшается при введении дополнительных переменных. И если число факторов приблизится к числу наблюдений, то остаточная дисперсия будет равна нулю, и коэффициент множественной корреляции, а значит и коэффициент детерминации, приблизятся к единице, хотя в действительности связь между факторами и результатом и объясняющая способность уравнения регрессии могут быть значительно ниже.

Для того чтобы получить адекватную оценку того, насколько хорошо вариация результирующего признака объясняется вариацией нескольких факторных признаков, применяют скорректированный коэффициент детерминации

$$R_{скор}^2 = 1 - (1 - R^2) \cdot \frac{n-1}{n-k-1}.$$

Скорректированный коэффициент детерминации всегда меньше R^2. Кроме того, в отличие от R^2, который всегда положителен, $R_{с\,корр}^2$ может принимать и отрицательное значение. Устраняет эффект, связанный с ростом R^2 при возрастании числа регрессоров, являясь коррекцией R^2 на число регрессоров. Использование $R_{скор}^2$ более корректно для сравнения регрессий при изменении количества регрессоров.

Полученные коэффициенты корреляции определяются по выборочной совокупности и являются точечными оценками. Поэтому необходима статистическая проверка значимости коэффициентов корреляции.

Проверка статистической значимости парного и частного коэффициентов корреляции.

Назовем параметр связи в генеральной совокупности значимо отличающимся от нуля (значимым), если гипотеза о равенстве нулю этого параметра отвергается с заданным уровнем значимости α. Если же эта гипотеза принимается, генеральный параметр связи называется незначимым. В корреляционной модели соответствующая связь между величинами считается недоказанной или отсутствующей.

Проверка статистической значимости парного и частного коэффициентов корреляции осуществляется с помощью t – критерия Стьюдента.

H_0: ρ=0, H_1: ρ≠0

$$t_{расч} = \frac{r}{\sqrt{1-r^2}}\sqrt{n-l-2} \quad ,$$

где *r* – частный или парный коэффициент корреляции;

l – порядок частного коэффициента корреляции;

Если $|t_{расч}| > t_{табл}(\alpha,\ v=n\text{-}l\text{-}2)$, то гипотеза H_0 с вероятностью ошибки 5% отвергается, проверяемый коэффициент корреляции считается значимым с вероятностью 95%.

Если же $|t_{расч}| < t_{табл}(\alpha,\ v=n\text{-}l\text{-}2)$, то гипотеза H_0 не отвергается.

С помощью таблиц Фишера-Иейтса.

По таблице Фишера–Иейтса определяется $r_{кр}(\alpha,\ v=n\text{-}l\text{-}2)$.

Если $|r_{расч}| > r_{кр}(\alpha,\ v=n\text{-}l\text{-}2)$, то гипотеза H_0 с вероятностью ошибки 5% отвергается, проверяемый коэффициент корреляции считается значимым с вероятностью 95%.

Если же $|r_{расч}| < r_{кр}(\alpha,\ v=n\text{-}l\text{-}2)$, то гипотеза H_0 не отвергается.

Проверка статистической значимости множественного коэффициента корреляции и коэффициента детерминации проводится с помощью F-критерияФишера.

Нулевая гипотеза: отсутствует линейная связь между переменной x: и остальными переменными, образующими многомерный признак, H_0: R_y=0.

Рассчитываем статистику

$$F = \frac{R_y^2 / k}{(1-R_y^2)/(n-k-1)}$$

Если $F_{расч} > F_{табл}$ (α, $v_1=k$, $v_2=n\text{-}k\text{-}1$), то гипотеза H_0 отвергается, т.е. r_y значимо отличается от нуля, следовательно, линейная связь есть и она статистически значима с вероятностью γ=1-α.

Иначе связь между случайной величиной *y* и остальными случайными величинами отсутствует.

Конечно, проверку значимости коэффициентов связи начинать с частных коэффициентов корреляции не обязательно. Можно в некоторых случаях сократить такую проверку, например, если r_1 незначим, то коэффициенты $r_{12|3}$ и $r_{13|2}$ становятся незначимыми. Далее, если $r_{12|3}$ незначим, то $r_1=|r_{13}|$ (множественный коэффициент корреляции незначимо отличается от абсолютной величины парного коэффициента корреляции). Для значимых параметров корреляции можно найти интервальные оценки.

5. Мультиколлинеарность.

Одним из вопросов спецификации модели является отбор факторов, включаемых в модель.

Включение в уравнение множественной регрессии того или иного набора факторов связано, прежде всего, с представлениями исследователя о природе взаимосвязи моделируемого показателя с другими экономическими явлениями.

Факторы, включаемые во множественную регрессию должны отвечать следующим требованиям:

1. Они должны быть количественно измеримы.

2. Отсутствие коллинеарности и мультиколлинеарности факторов.

Одним из основных препятствий эффективного применения множественного регрессионного анализа, является мультиколлинеарность.

Мультиколлинеарность - линейная взаимосвязь двух или нескольких объясняющих переменных.

Последствия мультиколлинеарности:

1. система нормальных уравнений может оказаться плохо обусловленной и повлечь за собой неустойчивость и ненадежность оценок коэффициентов регрессии. (Матрица (X^TX) становится слабообусловленной, т.е. ее определитель близок к нулю).

2. большие дисперсии $S_{b_j}^2$ (стандартные ошибки) оценок коэффициентов регрессии. Это расширяет интервальные оценки, ухудшая их точность;

3. заниженные значения $t(b_j)$, что может привести к неоправданному выводу о существенности влияния соответствующего фактора на результат;

4. затрудняется определение влияния каждой из объясняющих переменных на результативный показатель, и параметры параметры уравнения регрессии оказываются неинтерпретируемыми;

5. возможно получение неверного знака у коэффициента регрессии;

6. завышенные значения множественного коэффициента корреляции

На практике, о наличии мультиколлинеарности, обычно судят по матрице парных эффициентов корреляции. Если один из элементов матрицы R больше 0,7 - 0,8, т. е. r_{jl} >0,7-0,8 , то считают что имеет место мультиколлинеарность. А также о ее проявлениях, указанных в «последствиях мультиколлинеарности».

Устранять мультиколлинеарность или нет – зависит от цели исследования. Если целью является прогноз значений зависимых переменных, то наличие мультиколлинеарности не сказывается на прогнозных значениях, при условии, что между коррелированными переменными будут сохраняться те же отношения. Если цель – определение влияния каждой из объясняющей переменной на зависимую переменную, то мультиколлинеарность – серьезная проблема.

Чтобы избавиться от этого негативного явления при построении эконометрических моделей:

1. используют метод исключения переменных: из уравнения регрессии исключают один или несколько факторных признаков. Предпочтение отдается тому фактору, который при достаточно тесной связи с Y имеет наименьшую тесноту связи с другими факторами. Исключение факторов может проводиться и по t-критерию Стьюдента. Из уравнения исключаются факторы с величиной t-критерия меньше табличного.

Однако необходимо учитывать, что если переменные были включены в модель на теоретической основе, то неправомочно исключать их только ради «улучшения» статистических результатов.

2. используют алгоритм пошагового включения переменных: первоначально строится парное уравнение регрессии с фактором, наиболее тесно связанным с результатом, проверяется значимость уравнения в целом и его параметров. Затем вводится следующая переменная, по силе связи с зависимой переменной. Если вновь вводимая переменная улучшает модель, т.е. значение множественного коэффициента детерминации увеличивается, то ее включают в модель.

3. Сроят уравнение регрессии на главных компонентах.

1.5 Лекция №5 (2 часа)
Тема «Модели стационарных временных рядов»

1.5.1 Вопросы лекции:

1. Стационарные временные ряды и их основные характеристики
2. Модели стационарных временных рядов

1.5.2. Краткое содержание вопросов

1. Стационарные временные ряды и их основные характеристики

Стохастический процесс Y_t называется стационарным в сильном смысле (строго стационарным или стационарным в узком смысле), если совместное распределение вероятностей всех переменных $y_{t1}, y_{t2}, ..., y_{tn}$ точно то же самое, что и для переменных $y_{t1+\tau}, y_{t2+\tau}, ..., y_{tn+\tau}$.

Под стационарным процессом *в слабом смысле* (в широком смысле) понимается стохастический процесс, для которого среднее и дисперсия независимо от рассматриваемого периода времени имеют постоянное значение, а автокорреляция зависит только от длины лага между рассматриваемыми переменными:

$$\mu(y_t) = \mu(y_{t+\tau}) = \mu;$$

$$D(y_t) = D(y_{t+\tau}) = \mu(y_t - \mu)^2 = \mu(y_{t+\tau} - \mu)^2 = D_0 = const;$$

$$\operatorname{cov}(y_t, y_{t+\tau}) = \mu[(y_t - \mu)(y_{t+\tau} - \mu)] = \operatorname{cov}(\tau).$$

Из этого следует, что автоковариация будет зависеть только от сдвига по времени и не будет зависеть от t.

При анализе изменения $\operatorname{cov}(\tau)$ в зависимости от временного сдвига принято говорить об автоковариационной функции.

С понятием автоковариационной функции тесно связано понятие автокорреляционной функции (АКФ):

$$\rho(\tau) = \frac{\operatorname{cov}(y_t, y_{t+\tau})}{D(y_t)} = \frac{\operatorname{cov}(\tau)}{D_0}$$

Выборочная оценка коэффициента автокорреляции $r(\tau)$ может быть определена следующим образом:

$$r(\tau) = \frac{\frac{1}{n-\tau}\sum_{t=1}^{n-\tau}(y_t - \bar{y})(y_{t+\tau} - \bar{y})}{\frac{1}{n}\sum_{t=1}^{n}(y_t - \bar{y})},$$

где n - длина временного ряда;

- временной сдвиг (лаг);

$\bar{y}$ - среднее значение временного ряда.

Формула для расчета выборочной оценки частного коэффициента автокорреляции:

$$r_{ij.k} = \frac{r_{ij} - r_{ik} \cdot r_{jk}}{\sqrt{(1 - r_{ik}^2)(1 - r_{jk}^2)}}.$$

Например, $r_ч$ 1- го порядка между y_t и y_{t+2} при устранении влияния y_{t+1}:

$$r_ч(2) = r_{02.1} = \frac{r(2) - r(1) \cdot r(1,2)}{\sqrt{(1 - r^2(1))(1 - r^2(1,2))}},$$

где $r(2)$ - коэффициент корреляции между y_t и y_{t+2};

$r(1)$- коэффициент корреляции между y_t и y_{t+1};

$r(1,2)$- коэффициент корреляции между y_{t+1} и y_{t+2}.

В практической аналитической работе стационарность временного ряда означает отсутствие:

- тренда;
- систематических изменений дисперсии;
- строго периодичных флуктуаций;
- систематически изменяющихся взаимосвязей между элементами временного ряда.

2. Модели стационарных временных рядов

Наиболее распространенные модели стационарных рядов – модели авторегрессии и модели скользящего среднего.

В авторегрессии каждое значение ряда находится в линейной зависимости от предыдущих значений. Если анализируемый динамический процесс зависит от значений, отстоящих на *p* временных лагов назад, то авторегрессионный процесс порядка *p*, т.е. AR (*p*):

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + ... + \alpha_p y_{t-p} + \varepsilon_t ,$$

где ε_t - «белый шум» с $\mu_\varepsilon = 0$.

α_0 - свободный член (часто приравнивается к нулю (опускается)).

Для выполнения условия стационарности все корни характеристического уравнения $1 - \alpha_1 z - \alpha_2 z^2 - ... - \alpha_p z^p = 0$ должны быть по модулю больше 1 и различны, т.е. $|z| > 1$. Если $|z| = 1$, процесс называется процессом *единичного корня* и является нестационарным.

Рассмотрим простейший вариант линейного авторегрессионного процесса – *модель* авторегрессии 1-го порядка – AR(1), или марковский процесс.

Эта модель может быть представлена в виде:

$$y_t = \alpha y_{t-1} + \varepsilon_t ,$$

где - числовой коэффициент, $|\alpha| < 1$,

ε_t - последовательность случайных величин, образующих белый шум.

Основные свойства Марковского процесса:

$$\mu y_t = 0$$

$$D(y_t) = \frac{\sigma_\varepsilon^2}{1-\alpha^2}$$

$$\operatorname{cov}(y_t, y_{t\pm k}) = \alpha^k D(y_t)$$

$$\rho(y_t, y_{t\pm k}) = \alpha^k$$

Поэтому степень тесноты корреляционной связи между членами последовательности экспоненциально убывает по мере их взаимного удаления друг от друга во времени.

Таким образом, $\alpha = \dfrac{\operatorname{cov}(y_t, y_{t-1})}{D(y_t)}$

Условие стационарности ряда для AR(1) определяется требованием к коэффициенту : $|\alpha| < 1$.

Практические рекомендации по идентификации авторегрессионных моделей опираются на изучение АКФ и ЧАКФ:

1. У моделей AR(p) значения АКФ экспоненциально затухают (либо монотонно, либо попеременно меняя знак);

2. ЧАКФ для моделей AR(p) будет иметь ненулевые значения лишь при $k \le p$, а начиная с лага $k = p + 1$ теоретическая ЧАКФ равна нулю. Это свойство становится ключевым при подборе порядка *p* авторегрессионной модели для конкретных экономических временных рядов.

2. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ ЛАБОРАТОРНЫХ РАБОТ

2.1 Лабораторная работа № 1 (2 часа).
Тема: «Парная линейная регрессия»

2.1.1 Цель работы: отработать навыки практических расчетов по построению линейного уравнения парной регрессии, оценки значимости его параметров, точности и надежности.

2.1.2 Задачи работы:

1. Построить поле корреляции и сформулируйте гипотезу о форме связи между темпом прироста затрат электроэнергии и темпом прироста выпуска продукции предприятия.
2. Рассчитать оценки параметров уравнения парной линейной регрессии.
3. Оценить тесноту связи между темпом прироста затрат электроэнергии и темпом прироста выпуска продукции с помощью выборочного коэффициента корреляции. Проверьте значимость коэффициента корреляции.
4. Рассчитать выборочный коэффициент детерминации. Сделайте экономический вывод.
5. Проверить значимость оценки коэффициента регрессии с помощью критерия Стьюдента при уровне значимости $\alpha = 0{,}05$.
6. Построить доверительный интервал для коэффициента регрессии. Дайте экономическую интерпретацию.
7. Составить таблицу дисперсионного анализа.
8. Оценить с помощью F- критерия Фишера - Снедекора значимость уравнения линейной регрессии ($\alpha = 0{,}05$).
9. Рассчитать темп прироста выпуска продукции, если темп прироста затрат электроэнергии будет равен 39%. Постройте доверительный интервал для прогнозного значения объясняемой переменной. Сделайте экономический вывод.
10. Рассчитать средний коэффициент эластичности (Э). Сделайте экономический вывод.
11. Определить среднюю ошибку аппроксимации.
12. На поле корреляции построить линию регрессии.

2.1.3 Перечень приборов, материалов, используемых в лабораторной работе:

1. Персональный компьютер

2.1.4 Описание (ход) работы: работа выполняется в рабочей тетради согласно предложенному алгоритму.

3. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

3.1 Практическое занятие № 1 (2 часа).
Тема: «Введение в эконометрику»

3.1.1. Задание для работы:

1. Отбор факторов и оценивание неизвестных параметров классической модели линейной регрессии
2. Фиктивные переменные во множественной регрессии
3. Виды нелинейной регрессии. Оценка параметров
4. Корреляция для нелинейной регрессии

3.1.2 Краткое описание проводимого занятия:

1) *Самостоятельная подготовка студентов в объеме учебных заданий*. Работа по данной теме предусматривает повторение лекционного материала, изучение материала, представленного в учебном пособии Евсеев, Е.А. Эконометрика: учебное пособие для академического бакалавриата / Е.А. Евсеев, В.М. Буре. – М.: Издательство «Юрайт», 2017. Кроме того, для более глубокого изучения темы рекомендуется изучить дополнительный материал (см. список дополнительной литературы). Простой же пересказ содержания лекций не является залогом получения высокой оценки.

2) *Проверка подготовленности студентов к занятиям путем проведения собеседования или дискуссии*. Содержание вопросов имеет преимущественно практическую направленность. Их обсуждение способствует выработке навыка у студентов применять те или иные теоретические положения по дисциплине к практическим ситуациям

3) *Решение задач* следующего типа:

Задача № 1.

По территориям региона приводятся данные за 199X г.:

Номер региона	Среднедушевой прожиточный минимум в день одного трудоспособного, руб., x	Среднедневная заработная плата, руб., y
1	x1	y1
2	x2	y2
3	x3	y3
4	x4	y4
5	x5	y5
6	x6	y6
7	x7	y7
8	x8	y8
9	x9	y9
10	x10	y10
11	x11	y11
12	x12	y12

1. Построить линейное уравнение парной регрессии y от x
2. Рассчитать линейный коэффициент парной корреляции и среднюю ошибку аппроксимации
3. Оценить статистическую значимость параметров регрессии и корреляции.
4. Выполнить прогноз заработной платы y при прогнозном значении среднедушевого прожиточного минимума x, составляющем 107% от среднего уровня.
5. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.

Задача № 2.

По 30 территориям России имеются данные, представленные в таблице.

По данным таблицы:

1. Построить уравнение множественной регрессии в стандартизованной и естественной форме; рассчитать частные коэффициенты эластичности, сравнить их с ß1 и ß2, пояснить различия между ними.

2. Рассчитать линейные коэффициенты частной корреляции и коэффициент множественной корреляции, сравнить их с линейными коэффициентами парной корреляции, пояснить различия между ними.

3. Рассчитать общий и частные F-критерии Фишера.

4) Для *самооценки* полученных знаний по теме *выполнение тестов*.

3.1.3 Результаты и выводы:

В результате проведения практического занятия созданы условия для восприятия темы, установлена связь с предыдущими темами курса. Поставлены задачи, создающие логическое мышление студентов. Тема практического занятия усвоена.

3.1 Практическое занятие № 2 (2 часа).
Тема: «Парная линейная регрессия»

3.1.1. Задание для работы:

1. Проверка общего качества регрессионной модели
2. Гетероскедастичность и автокорреляция регрессионных остатков
3. Определение доверительных интервалов для коэффициента и функции регрессии

3.2.2 Краткое описание проводимого занятия:

1) *Самостоятельная подготовка студентов в объеме учебных заданий*. Работа по данной теме предусматривает повторение лекционного материала, изучение материала, представленного в учебном пособии Евсеев, Е.А. Эконометрика: учебное пособие для академического бакалавриата / Е.А. Евсеев, В.М. Буре. – М.: Издательство «Юрайт», 2017. Кроме того, для более глубокого изучения темы рекомендуется изучить дополнительный материал (см. список дополнительной литературы). Простой же пересказ содержания лекций не является залогом получения высокой оценки.

2) *Проверка подготовленности студентов к занятиям путем проведения собеседования или дискуссии*. Содержание вопросов имеет преимущественно практическую направленность. Их обсуждение способствует выработке навыка у студентов применять те или иные теоретические положения по дисциплине к практическим ситуациям

3) *Решение задач* следующего типа:

Задача № 1.

По 20 предприятиям региона (см. таблицу) изучается зависимость выработки продукции на одного работника y (тыс. руб.) от ввода в действие новых основных фондов x1 (% от стоимости фондов на конец года) и от удельного веса рабочих высокой квалификации в общей численности рабочих x2 (%).

Номер предприятия	y	x1	x2	Номер предприятия	y	x1	x2
1	7.0	3.9	10.0	11	9.0	6.0	21.0
2	7.0	3.9	14.0	12	11.0	6.4	22.0
3	7.0	3.7	15.0	13	9.0	6.8	22.0
4	7.0	4.0	16.0	14	11.0	7.2	25.0
5	7.0	3.8	17.0	15	12.0	8.0	28.0
6	7.0	4.8	19.0	16	12.0	8.2	29.0
7	8.0	5.4	19.0	17	12.0	8.1	30.0
8	8.0	4.4	20.0	18	12.0	8.5	31.0
9	8.0	5.3	20.0	19	14.0	9.6	32.0
10	10.0	6.8	20.0	20	14.0	9.0	36.0

1. Оценить показатели вариации каждого признака и сделать вывод о возможностях применения МНК для их изучения.

2. Проанализировать линейные коэффициенты парной и частной корреляции.

3. Написать уравнение множественной регрессии, оценить значимость его параметров, пояснить их экономический смысл.

4. С помощью F-критерия Фишера оценить статистическую надежность уравнения регрессии и R2yx1x2. Сравнить значения скорректированного и нескорректированного линейных коэффициентов множественной детерминации.

5. С помощью частных F-критериев Фишера оценить целесообразность включения в уравнение множественной регрессии фактора x1 после x2 и фактора x2 после x1.

6. Рассчитать средние частные коэффициенты эластичности и дать на их основе сравнительную оценку силы влияния факторов на результат.

Задача № 2.

По данным за 18 месяцев построено уравнение регрессии зависимости прибыли предприятия y (млн. руб.) от цен на сырье x1 (тыс. руб. за 1 т) и производительности труда x2 (ед. продукции на 1 работника):

y = 200 - 1,5 * x1 +4,0 * x2.

При анализе остаточных величин были использованы значения, приведенные в табл.:

№	y	x1	x2
1	210	800	300
2	720	1000	500
3	300	1500	600

SUM E2t = 10500, SUM (Et - Et-1)2 = 40000

1. По трем позициям рассчитать y, Et, Et-1, E2t, (Et - Et-1)2.
2. Рассчитать критерий Дарбина - Уотсона.
3. Оценить полученный результат при 5%-ном уровне значимости.
4. Указать, пригодно ли уравнение для прогноза.

4) Для *самооценки* полученных знаний по теме *выполнение тестов*.

3.2.3 Результаты и выводы:

В результате проведения практического занятия созданы условия для восприятия темы, установлена связь с предыдущими темами курса. Поставлены задачи, создающие логическое мышление студентов. Тема практического занятия усвоена.

3.2 Практическое занятие № 3 (1 час).
Тема: «Нелинейные модели регрессии и линеаризация»

3.2.1. Задание для работы:

1. Освоить методику параметризации и построения регрессионного уравнения для отдельных видов нелинейных регрессий.

2. Получить опыт практических расчетов показателей тесноты связи и точности модели для отдельных видов нелинейных регрессий.

3.3.2 Краткое описание проводимого занятия:

1) *Самостоятельная подготовка студентов в объеме учебных заданий*. Работа по данной теме предусматривает повторение лекционного материала, изучение материала, представленного в учебном пособии Евсеев, Е.А. Эконометрика: учебное пособие для академического бакалавриата / Е.А. Евсеев, В.М. Буре. – М.: Издательство «Юрайт», 2017.

Кроме того, для более глубокого изучения темы рекомендуется изучить дополнительный материал (см. список дополнительной литературы). Простой же пересказ содержания лекций не является залогом получения высокой оценки.

2) *Проверка подготовленности студентов к занятиям путем проведения собеседования или дискуссии*. Содержание вопросов имеет преимущественно практическую направленность. Их обсуждение способствует выработке навыка у студентов применять те или иные теоретические положения по дисциплине к практическим ситуациям

3) *Решение задач* следующего типа:

Задача № 1.

Имеются данные по 12 предприятиям: х (%) – темп прироста затрат электроэнергии, у (%) – темп прироста выпуска продукции. Признаки х и у имеют нормальный закон распределения.

x	6,4	12,3	13,5	19,7	20	27,1	42,5	44	56,9	66,5	72,9	90,2
y	31,7	53,3	46,2	65,7	63,3	70,2	75,1	76,3	83,2	83,2	85,1	87,8

1. Рассчитать параметры степенной функции.
2. Найти показатели тесноты связи по каждой модели.
3. Оценить каждую модель через показатель детерминации, F – критерий Фишера, ошибку аппроксимации и выбрать наилучшую из них.

Задача № 2.

По 13 районам области рассматривается взаимосвязь следующих признаков: х (тыс. руб.) - среднемесячный доход, у (тыс. руб.) - объем продаж строительных материалов. Признаки х и у имеют нормальный закон распределения.

x	8,7	10,1	13,3	15	17,4	19,7	4	7	5	7,7	6,4	10,6	5,4
y	7,2	6,9	8,2	7,5	8,4	9	2	4,7	3,5	6,2	3	8,4	5,6

1. Рассчитать параметры равносторонней гиперболы.
2. Найти показатели тесноты связи по каждой модели.
3. Оценить каждую модель через показатель детерминации, F – критерий Фишера, ошибку аппроксимации и выбрать наилучшую из них.

4) Для *самооценки* полученных знаний по теме *выполнение тестов*.

3.3.3 Результаты и выводы:

В результате проведения практического занятия созданы условия для восприятия темы, установлена связь с предыдущими темами курса. Поставлены задачи, создающие логическое мышление студентов. Тема практического занятия усвоена.

3.3 Практическое занятие № 4 (1 час).

Тема: «Линейная модель множественной регрессии»

3.3.1. Задание для работы:

1. Отбор факторов при построении множественной регрессии
2. Выбор формы уравнения регрессии
3. Оценка параметров уравнения множественной регрессии

3.4.2 Краткое описание проводимого занятия:

1) *Самостоятельная подготовка студентов в объеме учебных заданий*. Работа по данной теме предусматривает повторение лекционного материала, изучение материала, представленного в учебном пособии Евсеев, Е.А. Эконометрика: учебное пособие для академического бакалавриата / Е.А. Евсеев, В.М. Буре. – М.: Издательство «Юрайт», 2017. Кроме того, для более глубокого изучения темы рекомендуется изучить дополнительный материал (см. список дополнительной литературы). Простой же пересказ содержания лекций не является залогом получения высокой оценки.

2) *Проверка подготовленности студентов к занятиям путем проведения собеседования или дискуссии*. Содержание вопросов имеет преимущественно практическую направленность. Их обсуждение способствует выработке навыка у студентов применять те или иные теоретические положения по дисциплине к практическим ситуациям

3) *Решение задач* следующего типа:

1) Рассчитайте параметры линейного уравнения множественной регрессии с полным перечнем факторов по данным в соответствии с вариантом.

2) Дайте сравнительную оценку силы связи факторов с результатом с помощью средних (общих) коэффициентов эластичности.

3) Оцените с помощью F-критерия Фишера-Снедекора значимость уравнения линейной регрессии и показателя тесноты связи. Определите множественный коэффициент корреляции и детерминации и скорректированный коэффициент детерминации.

4) Оцените статистическую значимость коэффициентов регрессии с помощью t-критерия Стьюдента.

5) Оцените качество уравнения через среднюю ошибку аппроксимации.

6) Рассчитайте матрицу парных коэффициентов корреляции и отберите информативные факторы в модели. Укажите коллинеарные факторы.

7) Постройте модель в естественной форме только с информативными факторами и оцените ее параметры.

8) Постройте модель в стандартизованном масштабе и проинтерпретируйте ее параметры.

9) Рассчитайте прогнозное значение результата, если прогнозное значение факторов составляют 80% от их максимальных значений.

10) Рассчитайте ошибки и доверительный интервал прогноза для уровня значимости $\alpha = 0{,}05$.

11) По полученным результатам сделайте экономический вывод.

4) Для *самооценки* полученных знаний по теме *выполнение тестов*.

3.4.3 Результаты и выводы:

В результате проведения практического занятия созданы условия для восприятия темы, установлена связь с предыдущими темами курса. Поставлены задачи, создающие логическое мышление студентов. Тема практического занятия усвоена.

3.4 Практическое занятие № 5 (2 часа).
Тема: «Модели стационарных временных рядов»

3.4.1. Задание для работы:

1. Корреляция рядов динамики
2. Регрессия по рядам динамики и прогнозирование на ее основе

3.5.2 Краткое описание проводимого занятия:

1) *Самостоятельная подготовка студентов в объеме учебных заданий*. Работа по данной теме предусматривает повторение лекционного материала, изучение материала, представленного в учебном пособии Евсеев, Е.А. Эконометрика: учебное пособие для академического бакалавриата / Е.А. Евсеев, В.М. Буре. – М.: Издательство «Юрайт», 2017. Кроме того, для более глубокого изучения темы рекомендуется изучить дополнительный материал (см. список дополнительной литературы). Простой же пересказ содержания лекций не является залогом получения высокой оценки.

2) *Проверка подготовленности студентов к занятиям путем проведения собеседования или дискуссии*. Содержание вопросов имеет преимущественно практическую направленность. Их обсуждение способствует выработке навыка у студентов применять те или иные теоретические положения по дисциплине к практическим ситуациям

3) *Решение задач* следующего типа:

Задача № 1.

Имеются следующие данные о величине дохода на одного члена семьи и расхода на товар А:

Показатель	2006г.	2007г.	2008г.	2009г.	2010г.	2011г.
Расходы на товар А, руб.	30	35	39	44	50	53

Доход на одного члена семьи, % к 2006 г.	100	103	105	109	115	118

1. Определить ежегодные абсолютные приросты доходов и расходов и сделать выводы о тенденции развития каждого ряда.
2. Перечислить основные пути устранения тенденции для построения модели спроса на товар А в зависимости от дохода.
3. Построить линейную модель спроса, используя первые разности уровней исходных динамических рядов.
4. Пояснить экономический смысл коэффициента регрессии.
5. Построить линейную модель спроса на товар А, включив в нее фактор времени. Интерпретировать полученные параметры.

Задача № 2.

По данным машиностроительных предприятий, методами корреляционного анализа исследовать взаимосвязь между следующими показателями: X1- рентабельность (%); X2 - премии и вознаграждения на одного работника (млн. руб.); X3-фондоотдача

№ п/п	X1	X2	X3
1	13,26	1,23	1,45
2	10,16	1,04	1,3
3	13,72	1,8	1,37
4	12,82	0,43	1,65
6	9,12	0,57	1,68
7	25,83	1,72	1,94
8	23,39	1,7	1,89
9	14,68	0,84	1,94
10	10,05	0,6	2,06

1. Из предложенных данных вычеркните строчку с номером, соответствующим последней цифре номера зачетной книжки.
2. Рассчитайте вектора средних и среднеквадратических отклонений, матрицу парных коэффициентов корреляции
3. Рассчитайте частные коэффициенты корреляции r12/3 и r13/2
4. По корреляционной матрице R рассчитайте оценку множественного коэффициента корреляции r1/23
5. При a=0,05 проверьте значимость всех парных коэффициентов корреляции.
6. При a=0,05 проверьте значимость частных коэффициентов корреляции r12/3 и r13/2
7. При a=0,05 проверьте значимость множественного коэффициента корреляции.

4) Для *самооценки* полученных знаний по теме *выполнение тестов*.

3.5.3 Результаты и выводы:

В результате проведения практического занятия созданы условия для восприятия темы, установлена связь с предыдущими темами курса. Поставлены задачи, создающие логическое мышление студентов. Тема практического занятия усвоена.

Разработал ________________________________ И.Н. Выголова